



Jan M. Boelmann
(Hrsg.)

Empirische Forschung in der Deutschdidaktik

Band 2: Erhebungs- und Auswertungsverfahren





Empirische Forschung in der Deutschdidaktik

Band 2

Erhebungs- und Auswertungsverfahren

Herausgegeben von

Jan M. Boelmann



Schneider Verlag Hohengehren GmbH

Umschlagidee:

Herausgeber

Bildquelle:

Fotolia: tablet computer mit daten für finanzanalyse/74033723

Die Veröffentlichung dieses Bandes als Open-Access-Veröffentlichung wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 16PGF0075 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.



Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

ISBN: 978-3-8340-1882-3

Schneider Verlag Hohengehren GmbH
Wilhelmstrasse 13
D-73666 Baltmannsweiler
Homepage: www.paedagogik.de

Das Werk und seine Teile sind in der Printversion (ISBN: 978-3-8340-1882-3) urheberrechtlich geschützt, jedoch in der vorliegenden Form unter der Creative Commons Lizenz *BY-NC-ND 4.0*: „Namensnennung-Nicht-kommerziell-Keine Bearbeitungen 4.0 International Public License“ veröffentlicht.



© Schneider Verlag Hohengehren, Baltmannsweiler 2018

Inhaltsverzeichnis

Jan M. Boelmann Vorwort.....	3
Jan M. Boelmann & Kathrin Pötter Zur Konzeption des Bandes.....	5
Jan M. Boelmann Von der Idee zum Forschungsprojekt.....	11

Erhebungsverfahren

Frederike Schmidt Interviewverfahren. Ein Überblick	23
Eva Maus Problemzentrierte Interviews	35
Frederike Schmidt Leitfadeninterviews.....	51
Ina Lindow Narrative Interviews.....	67
Daniel Scherf Gruppendiskussionen.....	81
Anica Betz Fragebögen	99
Juliane Dube Standardisierte Testverfahren	115
Wiebke Dannecker Lautes Denken	131
Diana Maak Videographische Datenerhebung.....	147
Ylva Schwinghammer Teilnehmende Beobachtung	165
Anja Binanzer Bildimpulse und Bildfolgen.....	179
Jasmin Benz Vignetten.....	203

Auswertungsverfahren

Vivien Heller & Miriam Morek Gesprächsanalyse	221
Silke Kubik Grounded Theory	245
Ralf Schieferdecker Dokumentarische Methode	265
Torsten Pflugmacher Objektive Hermeneutik	283
Jochen Heins Qualitative Inhaltsanalyse	303
Benjamin Uhl Quantitative Inhaltsanalyse	323
Iris Kleinbub Kriteriengeleitetes Rating	341
Miriam Langlotz Grammatische Textanalyse	359
Anke Schmitz Interaktionsanalysen	375
Lisa König, Lisa Steck & Jan Wawrzynek Kurz-Glossar	391

Vorwort

Empirisches Forschen ist ein spannendes Handwerk. Nicht nur, dass man den Blick von der reinen Theorie auf die Praxis ausweitet und erleben kann, ob sich die kunstfertig im Denkkammer ersonnenen Gedanken in der realen Welt bewähren, man erhält zudem Antworten auf Fragen, die man sich zuvor nur heuristisch beantworten konnte. Hierbei sind die Möglichkeiten nahezu unbegrenzt: Im Laufe eines Akademikerlebens kommt man mit vielen Forschungsprojekten, -methodiken und -stilen in Kontakt und bildet sich im Laufe der Zeit eine mehr oder weniger stark ausgeprägte Landkarte der Möglichkeiten.

Nachwuchswissenschaftlerinnen und Nachwuchswissenschaftlern, die noch nicht über einen so großen Erfahrungsschatz verfügen, sollte die durch das BMBF geförderte Tagung ‚Empirische Bildungsforschung in der Deutschdidaktik – Erhebungs- und Auswertungsverfahren und ihre Praxis‘ einen ersten Zugang in Form eines Überblicks in diese vielfältige Disziplin eröffnen.

Das grundlegende Anliegen bestand allerdings nicht so sehr in der Durchführung der Tagung an der PH Ludwigsburg, deren Aktualität durch über 90 Teilnehmerinnen und Teilnehmern aus ganz Deutschland, Spanien, Ägypten und Malta im September 2015 eindrücklich belegt wurde, sondern vielmehr in der Schaffung eines grundlegenden Einführungsbandes in die Erhebungs- und Auswertungsverfahren im Rahmen deutschdidaktischer Forschung, der Studierenden, Doktoranden und auch anderen Empirie-Interessierten eine erste Orientierung im Dickicht der empirischen Verfahren eröffnen sollte. Aus diesem Selbstverständnis heraus ist dieser Band auch nicht als umfassendes Handbuch konzipiert, sondern stellt einen Erstzugang dar, indem die einzelnen Verfahren und Methoden kurz und bündig mit konkreten Beispielen mitsamt Potenzialen und Fallstricken vorgestellt werden. Er soll Nachwuchsforschenden helfen, bei der Wahl ihres empirischen Instrumentariums einen breiten Überblick über aktuell verwendete Verfahren zu gewinnen, bevor in einer zweiten Phase eine vertiefende Auseinandersetzung mit den gewählten Methoden separat erfolgt, bei der allerdings die kommentierten Literaturempfehlungen der einzelnen Beiträge hilfreich genutzt werden können.

Ich danke den Vorträgerinnen und Vorträgern der Tagung und Beiträgerinnen und Beiträgern dieses Bandes, die sich trotz engen Zeitplans, straffer Vorgaben in Bezug auf die Strukturierung ihrer Beiträge und die eingeforderte anfängergerechte Sprache auf dieses Vorhaben eingelassen und es mitgetragen haben. Besonders danke ich Frederike Schmidt, die mir schon im Vorfeld der Antragstellung eine wichtige Ratgeberin und Ansprechpartnerin war, und deren Anregungen die Tagung und den Band auf vielfältige Weise bereichert haben.

Für die Hilfe während der Tagung danke ich meinen Hilfskräften Lisa König, Kalinka Neudeck, Peter Schurz, Lisa Steck und Jan Wawrzynek. Ebenso danke ich der Forschungsförderungsstelle der Pädagogischen Hochschule Ludwigsburg, insbesondere Alexandra Findeis, für die Unterstützung in Rat und Tat in allen

Phasen des Projekts. Besonders hervorgehoben sei zudem Kathrin Oedingen, die mir als Leiterin des Tagungsbüros und als Erstlektorin dieses Bandes in allen Phasen des Projekts eine sehr große Stütze war.

Großer Dank geht zudem an das Bundesministerium für Bildung und Forschung für die finanzielle Förderung der Tagung und dieses erweiterten Tagungsbandes.

Ludwigsburg im Dezember 2015

Jan M. Boelmann

Vorwort zur Wiederauflage in der Reihe ***Empirische Forschung in der Deutschdidaktik***

Wenn ein Band nach etwas über zwei Jahren zum dritten Mal wiederaufgelegt wird, scheint er eine Marktlücke getroffen zu haben. Im Fall des vorliegenden Bandes betrifft diese Lücke das empirische Arbeiten in der Deutschdidaktik, genauer gesagt, das empirische Arbeiten in der Deutschdidaktik für Nachwuchswissenschaftlerinnen und Nachwuchswissenschaftler. Die zahlreichen Rückmeldungen zu Band und Tagung zeigten uns schnell, dass dieser Band viele, aber bei weitem nicht alle Leerstellen füllen konnte. Deshalb beschlossen wir gemeinsam mit dem Schneider-Verlag, die Reihe *Empirische Forschung in der Deutschdidaktik* zu gründen.

Zwar erschien der Erhebungs- und Auswertungsverfahren-Band chronologisch vor allen anderen Bänden der Reihe, doch wird er in der Reihe als zweiter Band ausgewiesen. Ihm voran stellen wir einen Band über die *Grundlagen* empirischen Forschens, in denen die in diesem Band zu gering ausgeleuchteten theoretischen und praktischen Fragen auf dem Weg zu einem Forschungsprojekt grundgelegt werden. Auch zeigt dieser Band mit den konkreten Projekten im zweiten Teil jedes Beitrags ein Spektrum aktueller Forschungsprojekte auf, doch sollte dies noch systematischer erfolgen, wodurch sich die Existenz des dritten Bands der Reihe *Forschungsfelder* erklärt.

Ich danke allen Kolleginnen und Kollegen, seien es Arrivierte oder Nachwuchsforschende, für die Rückmeldungen, Anregungen, Bestärkungen ebenso wie für gerechtfertigte Kritik zu den Bänden. Auch den Autorinnen und Autoren dieses Bandes danke ich für die Durchsicht und teils Überarbeitung ihrer noch recht jungen Beiträge. Mein größter Dank gilt meiner Doktorandin Lisa König, die den Prozess der Reihengründung, wie auch die Planung und Herausgabe der weiteren Bände und die Wiederauflage dieses Bandes konstruktiv, kritisch und entlastend begleitet hat.

Ludwigsburg im August 2018

Jan M. Boelmann

Zur Konzeption des Bandes

Die durch die PISA-Studie maßgeblich vorangetriebene empirische Wende in der Bildungsforschung hat auch in der Deutschdidaktik einen Paradigmenwechsel nach sich gezogen. Konnte der Literaturdidaktik als wichtiger Teildisziplin der Deutschdidaktik noch bis in die 2000er-Jahre hinein vorgeworfen werden, lange an der „hermeneutischen Grundstruktur des Faches festgehalten“ (Groeben/Hurlmann 2006, 13) zu haben und empirische Fragestellungen nur randständig zu behandeln, so ist ihr Selbstverständnis mittlerweile eng mit der empirischen Bildungsforschung verwoben. Dies zeigt sich neben einer stetig steigenden Zahl von empirischen Forschungsprojekten arrivierter Forscherinnen und Forscher insbesondere in der Ausrichtung von wissenschaftlichen Qualifikationsarbeiten. Noch vor einem Jahrzehnt stellten insbesondere Staatsexamens- und Masterarbeiten mit empirischem Vorgehen eine Besonderheit dar; mittlerweile gehört es zu den grundlegenden Anforderungen an diese Arbeiten wie auch an Promotions- und Habilitationsschriften, dass in ihnen nicht nur die heuristische Entwicklung von Modellen und Theorien betrieben wird, sondern dass diese zusätzlich mit empirischer Beweisführung belegt werden.

Als ein hochrelevantes Problemfeld der Qualifikantinnen und Qualifikanten, aber auch von erfahrenen Forschenden, kristallisiert sich heraus, dass die Deutschdidaktik bislang nur partiell eine eigenständige Methodik ausgeprägt hat und somit Erhebungs- und Auswertungsverfahren im interdisziplinären Austausch mit den Nachbarwissenschaften entlehnt werden müssen. Auch wenn dies zu „soliden methodischen Settings“ (Bönnighausen/Brune 2014, 117) führt, besteht an dieser Stelle weiterer Vertiefungs- und Reflexionsbedarf. Daniel Scherf konstatierte in seiner Preisrede zur Verleihung des ‚Förderpreises des Symposium Deutschdidaktik 2014‘, dass der Mangel an ‚deutschdidaktischen‘, empirischen Verfahren eine große Herausforderung darstelle. Die in deutschdidaktischen Forschungsarbeiten verwendeten, den Nachbardisziplinen entlehnten Verfahren empirischen Erkenntnisgewinns würden im Rahmen dieser Forschungen schließlich zumindest partiell ihrem wissenschaftlich-philosophischen bzw. methodologischen Hintergrund entrissen und in ein neues Fachverständnis überführt. Damit stehe ihre Verwendung vor dem Problem, einerseits den methodologischen Prämissen, andererseits den Wissensproduktionsnormen ihres ‚neuen‘ Verwendungskontextes genügen zu müssen. Insbesondere Vertreter des deutschdidaktischen Nachwuchses, der als erster Jahrgang im Selbstverständnis der Deutschdidaktik als empirische Bildungsforschung ausgebildet wurde, nehmen sich dieser Schwierigkeit an und adaptieren verschiedene psychologische und sozialwissenschaftliche Verfahren problembewusst und auf kreative, gewinnbringende Weise.

Der Band zielt darauf ab, seinen Leserinnen und Lesern einen Überblick über verschiedene, in aktuellen deutschdidaktischen Projekten eingesetzte Erhebungs- und Auswertungsverfahren zu verschaffen und deren spezifische Leistungsfähigkeit zu reflektieren. Hierbei werden die einzelnen Methoden insbesondere für ihre Eignung, in Qualifikationsarbeiten eingesetzt zu werden, kritisch hinterfragt. Der vorliegende Band versteht sich nicht als umfassendes Kompendium, sondern vielmehr als Orientierung vermittelndes Einstiegswerk, das seinen Leserinnen und Lesern – um mit Jochen Heins zu sprechen – dabei hilft, ‚Schneisen in den Dschungel der Empirie‘ zu schlagen. Zugleich muss darauf hingewiesen werden, dass es neben den von den Autorinnen und Autoren dieses Bandes geschlagenen Schneisen auch sehr viele Methoden gibt, die ebenfalls produktiv in deutschdidaktischen Forschungsprojekten nutzbar sind. Dies zeigt sich auch daran, dass nahezu alle Beitragenden in ihren Ausführungen darauf hinweisen, dass neben den prototypisch vorgestellten Anwendungen auch zahlreiche weitere Ausformungen der jeweiligen Methoden denkbar, möglich und produktiv abzuwandeln sind. Die einzelnen Beiträge wollen somit das empirische Feld nicht verengen, sondern den Einstieg durch ausgewählte Beispiele vereinfachen und in einem ersten Schritt den nicht Empirie erfahrenden Leserinnen und Lesern ein Gefühl für die jeweilige Methode, ihre Einsatzmöglichkeiten und deren spezifische Fallstricke in der praktischen Nutzung vermitteln. Der zweite Schritt, namentlich die vertiefte Einarbeitung in eine Methode, kann anschließend über die methodenspezifischen Standardwerke erfolgen, wobei alle Autorinnen und Autoren im Rahmen eines kommentierten Literaturverzeichnisses entsprechende Titel empfehlen.

Um den Leserinnen und Lesern den Erstzugriff zu erleichtern, haben alle Autorinnen und Autoren einen weitestgehend standardisierten Aufbau der Beiträge gewählt: In einem ersten Teil, in dem die jeweilige Methode theoretisch vorgestellt wird, findet zuerst eine Vorstellung der Funktionsweise des Verfahrens statt, der ein Blick auf die benötigten Ausgangsdaten und die Ergebnisse ihrer Anwendung folgt. Anschließend werden Vor- und Nachteile des Verfahrens sowie dessen Gütekriterien benannt. Eine kurze kommentierte Literaturliste mit einschlägiger methodenspezifischer Forschungsliteratur für das vertiefende Selbststudium schließt den Theorieteil ab. Im zweiten Teil der Beiträge wird der Einsatz der Methode in einem konkreten Forschungsprojekt vorgestellt und vor dem Hintergrund ihrer Passung zur Forschungsfrage reflektiert. Darüber hinaus stehen hierbei insbesondere Fragen der Praktikabilität für den Einsatz in Qualifikationsarbeiten im Fokus.

Die Zuordnung der einzelnen Verfahren erfolgt in diesem Band forschungschronologisch: Unter Erhebungsverfahren fallen hier solche empirischen Techniken, deren Ziel es ist, in einem ersten Schritt Daten zu generieren, die in einem oder mehreren Folgeschritten mittels anderer empirischer Verfahren ausgewertet werden müssen. Als Auswertungsverfahren werden in diesem Band entsprechend Verfahren bezeichnet, die in einem zweiten, bzw. späteren Schritt eingesetzt werden, um die zuvor erhobenen bzw. ausgewählten Daten zu analysieren oder transformieren. Aus dieser pragmatischen Zuordnung folgen in gewissen Punkten aber spezifische Unschärfen: So verstehen sich zahlreiche quantifizierende Verfahren,

die hier als Verfahren des zweiten Schrittes ‚Auswertungsverfahren‘ genannt werden, wie etwa ‚kriteriengeleitete Ratings‘ oder die ‚quantitative Inhaltsanalyse‘ im Selbstverständnis als Erhebungsverfahren, da sie Daten generieren, die in der Folge mit statistischen Mitteln weiterverarbeitet werden müssen, um aussagekräftige Ergebnisse zu erzielen.

Zu den einzelnen Beiträgen

In dem einleitenden Beitrag betrachtet *Jan M. Boelmann* die einzelnen Phasen des Forschungsprozesses, in den die in diesem Band fokussierten Erhebungs- und Auswertungsverfahren eingebettet sind. Hierbei zeigt er insbesondere für studentische Forschende die Verzahnung der einzelnen Phasen miteinander auf.

Der erste Teil des Bandes widmet sich den Erhebungsverfahren und wird mit einem Überblicksartikel zu Interviewverfahren von *Frederike Schmidt* eröffnet, in dem sie allgemeine Prinzipien für deren Einsatz diskutiert und in die Vorstellung verschiedener Interviewformen einleitet.

Eva Maus fokussiert in ihrem Beitrag problemzentrierte Interviews. Anhand ihres Promotionsprojektes ‚Wer (ver)führt zum Lesen?‘ zeigt sie, wie diese theoriegeleitete und zugleich offene Form des Interviews zur Befragung jugendlicher Probandinnen und Probanden genutzt werden kann, um deren Vorlieben und Motivationsgründe bzgl. des Leseverhaltens und der Lektürepräferenzen zu untersuchen. Dabei legt sie ein besonderes Augenmerk auf die konkrete Planung, Durchführung und Auswertung einer Interviewstudie.

Als zweite Interviewform stellt *Frederike Schmidt* das Potenzial des leitfadengestützten Interviews als Erhebungsinstrument anhand zweier deutschdidaktischer Projekte vor.

Danach nutzt *Ina Lindow* ihr Dissertationsprojekt ‚Literaturunterricht als Fall. Kasuistisches Wissen von Deutschlehrenden‘ um zu verdeutlichen, wie das narrative Interview für die deutschdidaktische Forschung genutzt werden kann. Sie benennt und erklärt dezidiert die einzelnen Phasen dieser Interviewvariante, geht dabei auf die konkrete Durchführung ein und gibt Empfehlungen zur Auswertung der erhobenen Daten. Abschließend diskutiert sie kritisch, welche Vor- und Nachteile die Verwendung des narrativen Interviews vor dem Hintergrund ihres Forschungssettings aufwiesen.

Es folgt der Beitrag von *Daniel Scherf*, der an einem Beispiel aus der Lehrerprofessionsforschung aufzeigt, wie Gruppendiskussionen genutzt werden können, um kollektive Wissensbestände bestimmter Personengruppen in den Blick zu nehmen. Dabei führt er in die Methode der Gruppendiskussionen ein und zeigt auf, wie sich diese auswerten lässt.

Anica Betz setzt sich mit dem Fragebogen als ein klassisches, quantitatives Erhebungsinstrument auseinander. Sehr ausführlich erörtert sie die Fragebogenkon-

struktion, gibt Hinweise zur grafischen Gestaltung und stellt eine sprachdidaktische Studie vor, die mittels dieser Methode die Bedingungen des Lernens in einem linguistisch ausgerichteten Schülerlaborprojekt untersucht.

Am Beispiel des „Salzburger Lesescreening 5-8 zur Erhebung von basalen Leseverständnisleistungen“ (SLS) zeigt *Juliane Dube* in dem folgenden Beitrag, wie sich standardisierte Testverfahren in deutschdidaktischen Forschungsarbeiten einbetten lassen. Anhand ihres Projektes „Meine LeseZeit“, das auf die Förderung von Lesekompetenz abzielt, diskutiert sie beispielhaft die Passung der Leistungsfähigkeit der von ihr eingesetzten standardisierten Tests und ihrer Forschungsfrage und liefert zusätzlich einen Vergleich verschiedener deutschdidaktischer Screeningverfahren.

Wiebke Dannecker widmet sich in ihrem Beitrag dem Lauten Denken, welches mittlerweile als anerkanntes Verfahren deutschdidaktischer Forschung zur Erhebung komplexer Kognitionen gilt. In ihrem Beispielprojekt „Lautes Denken zu einem leisen Text“ geht sie der Frage nach, wie Schülerinnen und Schüler der Sekundarstufe II die Lektüre, Reflexion und Bewertung eines literarischen Textes bewältigen. Mit einem klaren Bezug zur Grounded Theory behandelt sie ausführlich die Wahl der Methode, die Planung und Durchführung der Erhebung sowie die Auswertung der Daten. In einem abschließenden Fazit geht sie nicht nur auf die Passung der Methode ein, sondern spricht sich zusätzlich für eine Methodentransformation zu Gunsten der Gegenstandsangemessenheit aus.

Diana Maak gibt im folgenden Beitrag konkrete Hinweise für die Gewinnung von und den forschungspraktischen Umgang mit videografischem Datenmaterial. Sie stellt die Vor- und Nachbereitung sowie die Erhebung von videografischen Daten in den Vordergrund und gibt einen Überblick über mögliche Programme zur Aufbereitung und Analyse dieses Materials. Anhand mehrerer Projekte klärt sie über die Vielfalt der Möglichkeiten bei der videografischen Datenerhebung auf.

Es folgt der Beitrag von *Ylva Schwinghammer*, in dem sie zunächst die teilnehmende Beobachtung in das Feld der wissenschaftlichen Beobachtungen einordnet. Anschließend umreist sie die vier Phasen der teilnehmenden Beobachtung und betont, dass deren Daten sowohl qualitativ als auch quantitativ ausgewertet werden können. In ihrem zweiten Teil präsentiert sie mit dem „Arbeitskoffer zu den Steirischen Literaturpfaden des Mittelalters“ ein mediävistisches Sparkling Science Projekt, in dem die teilnehmende Beobachtung zur Fundierung einer anschließenden schriftlichen Erhebung und zugleich zur Generierung neuer Erkenntnisse und Erfahrungen genutzt wurde. Abschließend stellt sie die Chancen und Probleme dieser Methode in Bezug zu ihrem Projekt dar.

Anja Binanzer stellt mit den Bildimpulsen ein Elizitationsverfahren vor, das die Probandinnen und Probanden durch die Kombination von Sprachproduktionsaufgaben und Bildstimuli gezielt zur Datenproduktion auffordert. Zur Verdeutlichung beschreibt sie neben ihrer eigenen Studie knapp mehrere deutschdidaktische Forschungsprojekte, die Bildimpulse bzw. Bildfolgen in Verbindung mit Sprachproduktionsaufgaben einsetzen. Zum Abschluss zieht sie ein methodisches

Fazit, in dem sie Vor- und Nachteile sowie konkrete Konstruktionsfehler reflektiert.

Das letzte Erhebungsinstrument in diesem Band behandelt *Jasmin Benz*. Sie thematisiert Vignetten, die in der didaktischen Forschung genutzt werden, um Lehr- und Lernsituationen abzubilden und Anschlussfähigkeit an verschiedene andere Erhebungsverfahren besitzen. Sie zeigt auf, dass Vignetten sowohl im Rahmen qualitativer als auch quantitativer Forschungen eingesetzt werden können. Im ersten Abschnitt geht die Autorin auf die unterschiedlichen Formen von Vignetten ein und erörtert anhand von vorliegenden Studien unterschiedliche Bearbeitungsformate. Mithilfe ihres Dissertationsprojektes „ProfiL_13: Vignetten als Erhebungsinstrument in Weiterbildungen für Deutschlehrkräfte“ veranschaulicht sie im zweiten Teil den Einsatz von schriftlichen Vignetten. Dazu bespricht sie nicht nur das gewählte Forschungssetting, sondern ebenfalls die Konzeption der Vignetten.

Der zweite Teil des Bandes wird mit dem Beitrag von *Vivien Heller* und *Miriam Morek* zur Gesprächsanalyse eröffnet, welche sich als qualitatives, mikroanalytisches Verfahren zur Beschreibung sozialer Interaktion definieren lässt. Neben methodischen Prinzipien werden der ethnomethodologische Entstehungshintergrund sowie die konstitutiven Eigenschaften von Gesprächen dargestellt. Eine Besonderheit in diesem Beitrag stellt der Exkurs zur Transkription dar. Im zweiten Teil nutzen die Autorinnen das interdisziplinäre Projekt „InterPass“ um aufzuzeigen, wie der gesprächsanalytische Zugang für die empirische Bearbeitung sprachdidaktischer Fragestellungen nutzbar gemacht werden kann.

Silke Kubik beschäftigt sich im folgenden Beitrag mit der Grounded Theory, einem hypothesengenerierenden und -erweiternden Verfahren, das den gesamten Forschungsprozess – von der Datenerhebung bis zur Datenauswertung mit bestimmten Kodierverfahren – abdecken kann. Sie stellt detailliert den Prozess des Kodierens dar und benennt die spezifischen prozessorientierten Merkmale dieser Methode. Ihr Projekt zu literarischen Reflexionsprozessen von Jugendlichen nutzt sie zur Veranschaulichung des forschungspraktischen Arbeitens mit der Grounded Theory.

Ralf Schieferdecker widmet sich der dokumentarischen Methode und stellt übersichtlich und ausführlich die einzelnen methodischen Schritte dieses Auswertungsverfahrens vor. Im zweiten Teil skizziert er die Forschungsfrage, das Vorgehen und die Ergebnisse seines Forschungsprojektes zum Thema „Orientierung von Lehrenden im Themenfeld Heterogenität“. Darin nutzt er die dokumentarische Methode, um konjunktive Erfahrungsräume von Lehrkräften zu rekonstruieren und zu Typen zu verdichten.

Es folgt der Beitrag von *Torsten Pflugmacher*, der sich der Objektiven Hermeneutik, einer Methode zur sequenzanalytischen Rekonstruktion von Unterrichtsprozessen, zuwendet. Er bespricht ausführlich deren fünf Maximen und zeigt anhand seines Beispielprojektes zur realistischen Literaturvermittlung in knapper

Form, wie eine mögliche Fallanalyse mithilfe der Objektiven Hermeneutik aussehen könnte.

Die nächsten beiden Beiträge widmen sich der Inhaltsanalyse. *Jochen Heins* bespricht die qualitative Inhaltsanalyse und konzentriert sich zunächst auf die Merkmale dieses Verfahrens und das methodische Vorgehen. Diese Ausführungen illustriert er anhand eines Projektes der Lernaufgabenwirkungsforschung im Literaturunterricht im zweiten Teil des Beitrages.

Benjamin J. Uhl betrachtet hingegen die quantitative Inhaltsanalyse und thematisiert detailliert deren Herzstücke: das Kategoriensystem und die darauf basierende Kodierung. Diese Überlegungen veranschaulicht er unter der Zuhilfenahme seines Dissertationsprojektes „Tempus, Narration und Medialität. Eine Studie über die Entwicklung schriftlicher Erzählfähigkeit an der Schnittstelle zwischen Grammatik und Schreiben“. Dabei gelingt es ihm, die Ergebnisse der Studie auszugeweiht vorzustellen und zugleich eine Reflexion des Einsatzes der gewählten Methode im Kontext des Forschungsprojektes vorzunehmen.

Anschließend beschäftigt sich *Iris Kleinbub* mit kriteriengeleiteten Ratings, die komplexe verhaltensmäßige, soziale und emotionale Vorgänge messen können. Dabei geht sie u.a. auf die spezifischen Charakteristika des Verfahrens ein. Anhand ihrer Beispielstudie „VERA – Gute Unterrichtspraxis“ stellt sie anschaulich dar, wie ein lesedidaktisch spezifiziertes Ratingverfahren konstruiert werden kann.

Mit der grammatischen Textanalyse stellt *Miriam Langlotz* im folgenden Beitrag ein korpusbasiertes, quantitatives Untersuchungsverfahren der Schreibentwicklungsforschung vor, dessen konkrete Umsetzung anhand ihres Dissertationsprojektes „Junktion und Schreibentwicklung“ erläutert wird.

Anke Schmitz behandelt das Verfahren der Interaktionsanalyse. Auf Grund ihrer statistischen Anlage unterscheidet sich dieses Verfahren von vielen in diesem Band vorgestellten Methoden, birgt aber für die deutschdidaktische Forschung spannende Mehrwerte, wie Schmitz durch die Vorstellung ihrer Studie zur Wirkung der globalen Textkohäsion auf das Textverständnis nachweist.

Abgeschlossen wird der Band mit einem Kurz-Glossar von *Lisa König, Lisa Steck* und *Jan Wawrzynek*, die zentrale Begriffe der empirischen Bildungsforschung für unerfahrene Leserinnen und Leser kurz mit dem Ziel skizzieren, Verständnisschwierigkeiten vorzubeugen.

Literatur

Bönnighausen, Marion; Brune, Carlo: Literarästhetische Kompetenzen – Zwei Sammelbände. In: Didaktik Deutsch (2014), 37, 43-48.

Groeben, Norbert/Hurrelmann, Bettina: Empirische Unterrichtsforschung in der Literatur- und Lesedidaktik. Ein Weiterbildungsprogramm. Weinheim/München: Juventa 2006.

Von der Idee zum Forschungsprojekt

Ein Überblick für Unerfahrene

1. Phasen des Forschungsprozesses

Wenn in diesem Band verschiedene empirische Verfahren zur Erhebung und Auswertung von Daten vorgestellt werden, darf dies nicht verschleiern, dass deren Einsatz nur einen spezifischen Ausschnitt des gesamten Forschungsprozesses darstellt. Der Einsatz von Erhebungs- und Auswertungsverfahren findet an zentralen Gelenkstellen eines Forschungsprojekts statt, ist aber in einen größeren Prozess eingebunden: So findet eine Phase vor der Datenerhebung statt, in der die wissenschaftliche Fundierung des Projekts mit der Entwicklung der Fragestellung und der Konzeptualisierung des empirischen Vorgehens durchgeführt wird, und eine Phase nach der Auswertung, die zumeist aus der statistischen Weiterverarbeitung der Analyseergebnisse und immer aus der Interpretation der Daten besteht. In den meisten Fällen gibt es auch eine Phase zwischen der Erhebung und der Auswertung der Daten, in denen die erhobenen Daten transformiert, weiterverarbeitet und gesichert werden. Ebenso ist es möglich, mehrere Erhebungs- und/oder Auswertungsrunden durchzuführen.

Wie der Forschungsprozess konkret abläuft, unterscheidet sich zwar im Einzelfall sehr stark und muss immer in Hinblick auf das Untersuchungsziel und die Forschungsfrage spezifiziert werden, dennoch lässt sich der grundlegende Ablauf eines Forschungsprojekts mit folgendem Schema¹ beispielhaft skizzieren:

1	Entwicklung einer Fragestellung (auf Grundlage eines umfassenden Literaturstudiums)
2	Konzeptualisierung der Studie (Auswahl des Gegenstands, der Probandengruppe, der empirischen Methodik, Pilotierung)
3	Datenerhebung mittels Erhebungsverfahren
4	Datenaufbereitung (Transkription, Dateneingabe, -modifikation), Datensicherung
5	Datenanalyse mittels Auswertungsverfahren
6	Statistische, computergestützte Weiterverarbeitung der Analyseergebnisse
7	Interpretation der Ergebnisse, Rückbezug auf Fragestellung, Theorie und Hypothesen

¹ Ähnliche Ablaufschemata finden sich bei Friedrichs 1990; Abel/Möller/Treumann 1998; Atteslander 2003; Diekmann 2005 sowie Raithel 2006.

Insbesondere für unerfahrenere, studentische Forschende sollen diese Fixpunkte kurz skizziert werden, um die Verschränkung der in diesem Band vorgestellten Verfahren mit den anderen Phasen der empirischen Erhebung zu verdeutlichen. Hierbei werden die einzelnen Phasen getrennt voneinander dargestellt, wenngleich sie häufig ineinander übergehen oder zirkulär ablaufen. Insbesondere die ersten beiden Phasen sind stark ineinander verschränkt, da sich die Konzeptualisierung aus der Forschungsfrage ableitet, diese aber häufig im Rahmen der Konzeptualisierung Anpassungen erfährt.

1.1 Entwicklung einer Fragestellung

Zu Beginn eines Forschungsvorhabens steht eine bislang unbeantwortete Frage, für deren Beantwortung der Einsatz empirischer Mittel genutzt wird. Hierbei unterscheidet sich die Entwicklung wissenschaftlicher Fragestellungen von denen des Alltags grundlegend: Der sogenannte gesunde Menschenverstand sieht ein Problem und versucht, dieses in Gänze zu lösen. In der Wissenschaft hingegen werden in der Regel komplexe Probleme betrachtet, die nicht als Ganzes zu erforschen sind. Vielmehr ist es die Aufgabe der Forschenden, das Problem in einzelne Aspekte zu zerlegen und sich diesen ausführlich zu widmen. Dieser Aspekt oder auch Teilaspekt des Problems wird im Rahmen eines Forschungsprojekts umfassend bearbeitet, in seiner Bedeutung für das gesamte Problem reflektiert und anschließend auf dieses zurückbezogen. Zwar kann prinzipiell jede Frage als Forschungsfrage genutzt werden, allerdings stellt jede Frage in der Folge spezifische Herausforderungen an den anschließenden Forschungsprozess.

Die vorschnell gewählte Frage ‚Lesen Kinder mit Migrationshintergrund schlechter als Kinder ohne Zuwanderungsgeschichte?‘ wäre beispielsweise aus mehreren Gründen nicht als Forschungsfrage für eine Qualifikationsarbeit geeignet: Zum einen kann die Frage mit Blick auf die nationalen und internationalen Vergleichsstudien wie PISA als bereits beantwortet gelten; eine erneute Untersuchung dieser Frage wäre nur dann sinnvoll, wenn die Arbeit neue, aussagekräftigere Erkenntnisse als die bereits durchgeführten Studien generiert. Aber auch wenn es die großen Studien nicht gegeben hätte, wäre die Frage für eine Qualifikationsarbeit ungeeignet, da sie zweitens das gesamte Spektrum eines komplexen fachlichen Gegenstands in den Blick nimmt: Für die Erfassung des Faktors ‚Lesen‘ müsste zuerst der Gegenstand ‚Lesen‘ präzisiert werden: Welches begriffliche Verständnis von Lesen liegt der Arbeit zu Grunde? Welche Prozesse und Begleitfaktoren genau werden betrachtet? Anschließend müsste neben einem validen Lesekompetenzmodell eine entsprechend erprobte und das gesamte Modell abbildende Testbatterie entwickelt werden, die dann an einer repräsentativen Stichprobengröße zur Anwendung gebracht würde, was als dritter großer Grund gegen eine Behandlung der Fragestellung in einer Qualifikationsarbeit spricht. Die in der Fragestellung benannte Zielgruppe ‚Kinder mit Migrationshintergrund‘ schließt gleichermaßen alle Altersstufen und Herkunftsländer ein und beachtet auch nicht, dass wahrscheinlich nur solche Kinder mit Migrationshintergrund gemeint sind, die in

Deutschland leben bzw. in Deutschland aufgewachsen sind. Liest man die Fragestellung ganz genau, könnte man eine globale Komponente in sie hineinlesen: ‚Lesen in jedem Land der Welt Kinder mit Migrationshintergrund schlechter als Kinder ohne Zuwanderungsgeschichte?‘

Nun soll nicht der Eindruck entstehen, dass der Forschungsbereich ‚Lesen im Kontext von Zuwanderung‘ ausgeforscht und es nicht ratsam sei, hier eine Forschungsarbeit zu schreiben. Vielmehr sollte bei der Entwicklung einer Fragestellung darauf geachtet werden, dass die Frage so gestellt wird, dass sie im Rahmen einer Master- oder Dissertationsarbeit zu beantworten ist – dies geschieht durch die bereits angesprochene Zerlegung des Problems in kleinere bearbeitbare Aspekte. Um das oben genannte Beispiel aufzugreifen, sollte in jedem Fall die Probandengruppe verkleinert werden, etwa durch die Festlegung auf eine Jahrgangsstufe und/oder Schulform – auch die Eingrenzung auf ein Herkunftsland ist denkbar. Zudem ließe sich auch der Untersuchungsgegenstand fokussieren, indem beispielsweise nicht der gesamte Leseprozess und die Lesekompetenzen der Probandinnen und Probanden in den Blick genommen werden, sondern nur einzelne Aspekte, wie die Leseflüssigkeit oder die lokale Kohärenzbildung. Ebenso sollte die Erhebung nicht als Selbstzweck betrachtet werden, sondern in einen Kontext eingebunden werden, der das Interesse am Gegenstand legitimiert.

Sowohl die Auswahl des Forschungsthemas, wie auch die Entwicklung der Forschungsfrage erfolgt großteilig theorie- und somit literaturgestützt: Im Rahmen eines Literaturstudiums wird das Forschungsfeld eingegrenzt, bestehende Forschungslücken identifiziert und Forschungsbedarfe herausgearbeitet. Insbesondere in studentischen Forschungsarbeiten muss kein gänzlich neues Forschungskonzept geschaffen werden. Es ist hier durchaus vertretbar, sich an bestehenden Forschungsansätzen zu orientieren und eine überschaubare eigene Fragestellung zu verfolgen, die sich ggf. nur gering von bereits beforschten Gegenständen unterscheidet. Hierbei kann diese Referenzforschung als Vergleichsfolie genutzt werden, um die eigenen Ergebnisse zu verorten. In Promotionsarbeiten hingegen muss ein eigenes Forschungsfeld eröffnet werden, das in dieser Form noch nicht erforscht wurde. Sowohl die Auswahl des Forschungsfeldes, wie auch alle Schwerpunktsetzungen, die in der Fragestellung vorgenommen werden, sollten sich logisch durch Sekundärliteratur stützen lassen.

Im Rahmen des vertiefenden Literaturstudiums sollte nach bereits vorhandenen Arbeiten, die sich mit dem Problemfeld befassen, gesucht und deren Nutzen für die eigene Arbeit hinterfragt werden: Gibt es bereits Antworten auf die eigene Frage? Sind diese Antworten ausreichend, um die eigene Fragestellung zu beantworten? Wenn dem nicht so ist, kann das Vorgehen der bestehenden Studien hinterfragt werden: Was kann ich aus den bestehenden Studien inhaltlich und methodisch lernen? Bringen alternative Ansätze neue Erkenntnisse?

Insbesondere diese grundsätzlichen Fragen spielen bereits in die zweite Phase, die Konzeptualisierung der eigenen Studie, hinein. Die Konzeptualisierung des Projekts und Anpassung der Forschungsfrage können als sich abwechselnde Phasen verstanden werden, die so lange durchlaufen werden, bis eine im Rahmen der dem

Projekt möglichen Mittel beantwortbare Forschungsfrage mit einer entsprechenden methodischen Anlage entwickelt wurde.

Insbesondere in studentischen Arbeiten, die unter größerem Zeitdruck verfasst werden, stellt sich die Entwicklung der Forschungsfrage häufig als neuralgischer Punkt heraus, da durch eine vorschnell gewählte Fragestellung das gesamte Projekt gefährdet werden kann: Gibt die Fragestellung eine unrealistische Probandengruppengröße vor? Werden Untersuchungsgegenstände in den Fokus gerückt, die nur schlecht oder gar nicht messbar bzw. mit empirischen Mitteln nur schwierig fassbar sind? Benötigt man zur validen Beantwortung der Forschungsfrage mehr Zeit als im Rahmen der Arbeit zur Verfügung steht? Die Wahl der Fragestellung ist somit der zentrale Schritt im gesamten Forschungsprozess, da sie alle folgenden Schritte bestimmt.

1.2 Konzeptualisierung der Studie

Für die Entwicklung einer gelungenen Forschungskonzeptualisierung ist es wichtig, einen Blick auf das Ende des Forschungsprojekts zu werfen: Die Konzeptualisierung legt fest, welche Ergebnisse die Studie generieren wird und welche Tragweite diese letztendlich für sich beanspruchen kann. Damit die Ergebnisse der durchgeführten Untersuchung wissenschaftlichen Ansprüchen genügen, sollten sie objektiv, reliabel, also zuverlässig, und valide, das heißt gültig, sein. Am Ende des Projekts sollte die Forschungsfrage somit erstens beantwortet sein – und zwar so beantwortet, dass die Ergebnisse zweitens weder zufällig noch beliebig ausfallen. Auch sollten sie drittens gültig beantwortet werden, was bedeutet, dass die der Antwort zugesprochene Tragweite auch eingelöst werden kann.

Um die zu Eingang aufgeworfene Forschungsfrage: ‚Lesen Kinder mit Migrationshintergrund schlechter als Kinder ohne Zuwanderungsgeschichte?‘ objektiv, reliabel und valide zu beantworten, müsste eine Studie im Umfang der PISA-Studie entworfen werden. Dass dies ein für eine Qualifikationsarbeit unleistbares Arbeitspensum bedeutet, muss nicht erneut betont werden. Zugleich können arbeitsökonomischere Vorhaben nur weniger generalisierende Aussagen valide belegen. So würde es für die Beantwortung der Fragestellung nicht ausreichen, die Leseflüssigkeit von Schülerinnen und Schülern in drei Parallelklassen einer Schule in Kombination mit biographischen Daten zu erheben, um eine generalisierende Aussage über ein so großes Forschungsfeld (Lesen) und eine so große Population (alle Kinder mit Migrationshintergrund) zu treffen.

In der Forschungspraxis ist somit die Konzeptualisierung der Studie immer mit einer Revision der Fragestellung verbunden: Erst wenn die Fragestellung so gewählt ist, dass sie auch realistisch mit Mitteln der empirischen Sozialforschung beantwortet werden kann – und dies bei Qualifikationsarbeiten von einer Person in einem begrenzten Zeitrahmen –, wird die Konzeptionsphase abgeschlossen. Hierbei ist zu beachten, dass sich die begrenzten finanziellen, zeitlichen und personalen Ressourcen nicht darin widerspiegeln sollten, dass die Erhebung beliebig wird oder keinen Erkenntnisgewinn jenseits des Offensichtlichen generiert. Es gilt, aus den begrenzten Ressourcen das Bestmögliche herauszuholen und sich auf

spezielle Aspekte zu fokussieren und somit der Spezialisierung den Vorzug vor der Bagatellisierung zu geben.

Bezogen auf das Beispiel könnte dies bedeuten: Wenn Sie sich aus forschungspraktischen Aspekten dazu in der Lage sehen, eine Erhebung in drei Parallelklassen einer Schule durchzuführen, lohnt es sich nicht, nur die Leseflüssigkeit der Schülerinnen und Schüler in Verbindung mit ihren biographischen Daten zu erheben, da das Studium einschlägiger Fachliteratur vermutlich aussagekräftigere und auf eine größere Gesamtpopulation übertragbare Ergebnisse hervorbringen würde. Die Erhebung würde somit nur bereits mehrfach belegte Ergebnisse erneut reproduzieren und Ihre Studie hätte keinen besonderen Mehrwert. Eine Ausnahme ergäbe sich allerdings, wenn zu erwarten wäre, dass die Schülerinnen und Schüler mit Migrationshintergrund nicht die empirisch mehrfach nachgewiesene schwächere Leistung im Vergleich zu den Schülerinnen und Schülern ohne Migrationshintergrund zeigten. Die Gründe hierfür könnten etwa in einem besonderen Fördersystem der Schule oder in speziellen anderen Rahmenbedingungen liegen, denen Sie sich in der Folge der Erhebung ausführlich widmen müssten. Ein solches Vorgehen hätte für die wissenschaftliche Gemeinschaft einen Mehrwert, da Wege aus dem bestehenden Dilemma aufgezeigt würden. Dieses Vorgehen zöge eine Anpassung der Fragestellung nach sich: ‚Können durch das FöSys-Programm der Albert-Einstein-Schule migrationshintergrundbedingte Ungleichheiten in Bezug auf die Leseflüssigkeit verringert werden?‘. In der Folge müsste durch diese Fragestellung wiederum die Konzeptionalisierung der Studie angepasst werden, da für den Nachweis der Wirksamkeit des Programms sowohl Daten vor dem Eintritt in das Programm sowie nach dem Durchlaufen des Programms erhoben werden müssten. Erst so könnte nachgewiesen werden, dass sich die Ungleichheit im Verlauf der Programmteilnahme verringert hat.

Eine andere Form der Spezialisierung kann in der Wahl des gewählten Ausschnitts des Forschungsfelds bestehen. Anstatt ein komplexes Feld wie ‚Lesen‘ in Gänze in den Blick zu nehmen – was wie bereits gezeigt wurde, im Rahmen einer Qualifikationsarbeit nur schwer leistbar ist –, können auch Teilaspekte, etwa Lesestrategieanwendung, besondere Ausformungen, wie das Verstehen von Instruktionstexten, oder spezifische Kombinationen, wie die Herstellung von Kohärenz in diskontinuierlichen Texten, sein. Auch die Einbeziehung spezifischer Kontextvariablen der Probandinnen und Probanden wie Geschlecht, Herkunftsländer, Vorlesezeiten in der Kindheit, etc. kann einen Mehrwert generieren.

Sobald mit der Fragestellung der konkrete Gegenstand eingegrenzt wurde, müssen grundlegende Entscheidungen für das empirische Vorgehen vorgenommen werden: Wird im Projekt quantitativ gearbeitet, also mit einer großen Probandengruppe zum Zweck möglichst generalisierbare Antworten zu erhalten, oder soll qualitativ vorgegangen werden, was die intensivere Arbeit mit kleineren Probandengruppen nach sich zieht. Hieran schließt sich die Wahl einer entsprechenden Erhebungs- und Auswertungsmethodik an. Sobald diese Entscheidungen getroffen wurden, gehört es zur guten wissenschaftlichen Praxis, das gesamte empiri-

sche Set vor der eigentlichen Datenerhebung an einer kleinen Gruppe von Probandinnen und Probanden zu testen. Ziel dieser sogenannten Pilotierung ist es, das eigene Material auf seine Praxistauglichkeit zu überprüfen, mögliche Probleme im Vorfeld zu identifizieren, unerwartete Fallstricke auszumerzen und Routine im Einsatz der Verfahren aufzubauen. Manche Verfahren, wie etwa spezifische Interviewformen oder andere Erhebungsinstrumente, in denen der Forschende aktiv in die Datenerhebung eingreift, erfordern sorgfältiges Training und Vorbereitung. Insbesondere bei studentischen Projekten wird die Pilotierung häufig halbherzig durchgeführt oder gar ganz ausgespart, wovon ausdrücklich zu warnen ist. Werden schwerwiegende Fehler erst im Rahmen der Datenerhebung deutlich, besteht die Gefahr, dass diese ungültig wird oder in Gänze erneut durchgeführt werden muss.

1.3 Datenerhebung mittels Erhebungsverfahren

Nachdem die ersten beiden Phasen – abgesehen von der Pilotierung des Instrumentariums – theoriebasiert abliefen, beginnt mit der Datenerhebung der praktische Teil des Forschungsprojekts: Es werden in Interviews Probandinnen und Probanden befragt, über Fragebögen Einstellungen und Wissen überprüft, standardisierte Tests durchgeführt oder auf andere Arten und Weisen Daten erhoben, die zu einem späteren Zeitpunkt ausgewertet werden. Da ein Schwerpunkt dieses Buches auf der Vorstellung verschiedener Erhebungsverfahren liegt, soll dem hier nur wenig Raum geboten werden.

Hingegen soll die mit der Auswahl des Erhebungsverfahrens häufig einhergehende Frage nach der Größe der Untersuchungsstichprobe kurz thematisiert werden: Grundsätzlich gilt, dass eine größere Stichprobe weitreichendere Schlussfolgerungen zulässt als eine kleine. Zugleich ist es stark von der Forschungsfrage und den gewählten Methoden abhängig, wie viele Probandinnen und Probanden in die Untersuchung eingebunden werden müssen, um belastbare Ergebnisse zu gewinnen – die Autoren geben in ihren Beiträgen hierzu gute Hinweise. Aus forschungspraktischer Sicht macht es einen großen Unterschied, ob das Ziel der Untersuchung die Entwicklung einer oder mehrerer subjektiven Theorie(n) sein soll oder repräsentative Aussagen über eine Gesamtpopulation am Ende des Projekts erwartet werden. In ersterem Fall ließen sich mit Mitteln der qualitativen Forschung auch anhand einer kleineren Probandengruppe Ergebnisse generieren, wohingegen für die Erhebung repräsentativer Daten ungleich größere Fallzahlen im Rahmen quantitativer Forschung benötigt werden. Hinter diesem Punkt steht zudem die Frage, welche Tragweite die Ergebnisse des Projekts haben können bzw. müssen. Ohne eine abschließende Antwort hierauf zu liefern, sollte bei der Wahl einer Forschungsfrage der Rahmen der Möglichkeiten mitbedacht werden: So ist es etwa im Rahmen einer Masterarbeit, die innerhalb von drei Monaten verfasst werden muss, nicht möglich, eine Langzeitstudie durchzuführen. Auch Interventionsstudien, die beispielsweise die Wirksamkeit von spezifischen Unterrichtsmethoden oder -konzepten erheben wollen, sind in solch einem kurzen Zeitraum in

der Regel nicht realisierbar – ganz davon abgesehen, dass hierbei eine angemessene Dokumentation des Vorgehens bei Qualifikationsschriften mit fest begrenztem Seitenumfang häufig nicht leistbar ist. Grundsätzlich gilt: Eine sauber durchgeführte und dokumentierte Erhebung mit kleinen Fallzahlen, die belastbare Rückschlüsse auf eine kleine Population zulässt, ist besser als eine mit hohem Aufwand, aber unzulänglichen Mitteln durchgeführte Erhebung, bei der zwar mit vielen Probandinnen und Probanden gearbeitet wurde, deren Ergebnisse auf Grund von fehlender wissenschaftlicher Fundierung und Transparenz nicht belastbar sind.

1.4 Datenaufbereitung und -sicherung

Die Dauer und Ausformung dieser Phase des Forschungsprozesses unterscheidet sich je nach gewählter Erhebungsmethodik stark. Während die Dateneingabe kleinerer Fragebogenerhebungen vergleichsweise schnell realisiert werden kann – etwa in eine Statistiksoftware –, benötigt die Verschriftung von Interviewmitschnitten, die sogenannte Transkription (s. hierzu den Exkurs bei Heller/Morek in diesem Band), häufig mehrere Wochen intensiver Arbeit am Material.

In jedem Fall sollte die Phase nach der Erhebung genutzt werden, Sicherheitskopien des Datenmaterials anzufertigen und die Originale sicher zu verwahren. Auch sollten die Daten – sofern dies nicht bereits im Rahmen der Erhebung geschehen ist – anonymisiert werden. Üblicherweise bekommen die Probandinnen und Probanden einen Kode zugeteilt, der ihre Äußerungen mit ebenfalls erhobenen biographischen Daten verbindet, so wird eine Identifizierung der realen Person ausgeschlossen, die für den Forschungsprozess wichtigen Informationen bleiben aber verbunden. Beides geschieht einerseits um Ihrer selbst willen, da ein Verlust der Originaldaten (etwa durch einen Diebstahl eines Rucksacks oder Laptops) das gesamte Forschungsprojekt verzögern oder zum Abbruch bringen könnte und die Anonymisierung dabei hilft, Störfaktoren wie etwa Halo-Effekte zu vermeiden. Andererseits ist es aus forschungsethischer Sicht wichtig, die Persönlichkeitsrechte der Probandinnen und Probanden zu schützen, die im Fall eines Verlusts der Daten durch Diebstahl nicht gewährleistet werden könnten.

1.5 Datenanalyse mittels Auswertungsverfahren

Nachdem die Daten gesichert und aufbereitet wurden, kann ihre Auswertung beginnen. Hierbei muss eine Passung von Erhebungsinstrument und Auswertungsinstrument bestehen, da sich die Auswertungsverfahren nicht auf beliebige Datenkorpora anwenden lassen. Über die geschickte Kombination von Auswertungs- und Erhebungsverfahren lassen sich aber aus einem Datensatz höchst unterschiedliche Ergebnisse generieren. So ließen sich etwa die Transkripte einer Gruppendiskussion je nach Fragestellung theoretisch mit allen in diesem Band vorgestellten Auswertungsverfahren, ausgenommen der Interaktionsanalyse, bearbeiten. Die Ergebnisse wären – den unterschiedlichen Fragestellungen angemessen – jedoch höchst unterschiedlich.

1.6 Statistische, computergestützte Weiterverarbeitung der Analyseergebnisse

Während die Ergebnisse der Auswertung von qualitativen Verfahren zumeist schon aussagekräftige Ergebnisse generieren, stellt sich insbesondere bei quantitativen Forschungsvorhaben die Frage nach der weiteren Verarbeitung der erhobenen Daten. Um der Datenberge Herr zu werden, die etwa eine Fragebogenerhebung nach sich zieht, müssen statistische Verfahren herangezogen werden, mit denen sich die Daten aufbereiten und Verbindungen zwischen einzelnen Werten herstellen lassen, aber auch bei qualitativen Erhebungen kann die statistische Auswertung der erhobenen Daten Erkenntnispotenziale bergen. Während einzelne Berechnungen, wie die von Mittelwerten, noch mit schulischen Mathematikkenntnissen zu meistern sind, erfordert die statistische Ermittlung komplexerer Zusammenhänge, wie etwa die Ausprägung der Korrelationen zwischen zwei Merkmalen, höhere Mathematik. Üblicherweise wird in dieser Phase des Forschungsprojekts auf Computerunterstützung in Form von Statistiksoftware zurückgegriffen. Im wissenschaftlichen Kontext ist die Software SPSS der Firma IBM am weitesten verbreitet. Wenig statistikaffine Nutzer kann der Erstkontakt mit solcher Software verunsichern, da diese in der Regel nicht allzu intuitiv aufgebaut sind. Dennoch gelingt die Einarbeitung zumeist schnell und die Ergebnisse lohnen den Aufwand, da empirische Beweisführungen häufig nur über den Einsatz komplexer statistischer Verfahren erfolgen kann.

1.7 Interpretation der Daten, Rückbezug auf Fragestellung, Theorie und Hypothesen

Die Auswertung der Daten allein stellt nur den ersten Schritt in der Beantwortung der Forschungsfrage dar, da alle Ergebnisse des Forschungsprozesses in der Regel einer Interpretation bedürfen. Ein in studentischen Arbeiten häufig unterlaufender Fehler besteht darin, die Ergebnisse der empirischen Erhebung als selbsterklärend misszuverstehen. Als Forschende(r) sollten Sie auch im Sinne Ihres Projekts darauf achten, dass die Schlüsse, die Sie aus den Ergebnissen Ihrer Forschung ziehen, für alle Rezipienten Ihrer Arbeit ersichtlich und logisch herleitbar sind. Zudem schlagen Sie durch die Interpretation Ihrer Daten mit einem Rückbezug auf die Fragestellung einen Bogen zum Anfang Ihrer Arbeit und legitimieren somit auch rückblickend Ihr Vorgehen.

2. Sechs Praxistipps

Im Rahmen der Tagung wurden die Vortragenden nach ihren Ratschlägen für Nachwuchsforscherinnen und Nachwuchsforscher gefragt, die am Anfang ihres Vorhabens stehen. Die folgenden sechs Tipps fassen die Antworten zusammen.

1. Machen Sie es richtig!

Empirische Forschung kann man nicht „mal eben“ oder „ein bisschen“ betreiben, auch lassen sich Daten nicht nebenbei erheben oder auswerten. Um saubere Ergebnisse zu erzielen, muss man sich seinem Projekt mit voller Aufmerksamkeit widmen. Der Versuch, halbherzig an seinem Projekt zu arbeiten, führt häufig dazu, dass Potenziale vergehen oder Ergebnisse verfälscht werden. Auch kann es passieren, dass ganze Erhebungen oder Auswertungen erneut durchgeführt werden müssen. Insgesamt gilt: Wer vorher mehr nachdenkt, muss hinterher weniger ausbessern.

2. Formulieren Sie eine klare Fragestellung!

Die Formulierung einer klaren Fragestellung ist der Grundstein eines gelingenden Forschungsprojekts. Erst wenn Sie Ihr Forschungsinteresse klar benennen können, empfiehlt es sich, im Prozess fortzufahren. Ohne eine Fragestellung fehlt Ihrem Vorhaben die Richtung und es besteht die Gefahr, dass die auf dieser Grundlage erhobenen Ergebnisse beliebig oder ohne Anknüpfung an das anvisierte Forschungsfeld verbleiben. Natürlich können Sie sich über die Formulierung von Hypothesen, also Aussagen, die negierbar oder verifizierbar sind, zunehmend an die finale Form annähern, aber am Ende führt kein Weg an einer klaren Fragestellung vorbei.

3. Kennen Sie Ihr Feld!

Diesen Tipp kann man in verschiedene Richtungen deuten. Zentral ist die fachliche Perspektive: Wenn Sie sich mit empirischen Verfahren einer Fragestellung nähern, sollten Sie bereits Experte in Ihrem Feld sein und versuchen, einen wissenschaftlichen Mehrwert zu generieren. Das verhindert, dass bereits erforschte Aspekte erneut – ggf. sogar mit weniger aussagekräftiger Methodik – erhoben werden, ohne dass ein Erkenntnisgewinn generiert würde. Auch bei der Verortung der Tragweite Ihrer Ergebnisse hilft fundiertes Wissen im Forschungsfeld weiter, da die Anschlussfähigkeit an andere Forschungsarbeiten und Fachdiskurse gewährleistet werden kann.

Aber auch die lokale Lesweise ist für die Praxis hilfreich: Im Rahmen von Feldforschung sollten Sie wissen, worauf Sie sich einlassen. So ist es etwa bei Beobachtungen unerlässlich, dass Sie die Abläufe, die Sie beobachten wollen, so gut kennen, dass Sie nicht zentrale Aspekte versäumen, da Sie zufällig Ihren Fokus auf andere weniger wichtige Aspekte gelegt haben. Auch bei der Arbeit mit Videographie hilft die vorherige Kenntnis der räumlichen Begebenheiten, um die Kameraperspektiven sorgfältig zu bestimmen.

Gleiches gilt in Bezug auf technische Voraussetzungen: Machen Sie sich im Vorfeld mit der von Ihnen verwendeten Technik vertraut. Eine mit viel Aufwand und Organisationstalent vorbereitete Erhebung kann durch eine volle Speicherkarte oder einen übereifrigen Stand-by-Modus ruiniert werden – was im schlimmsten Fall erst nach einem Erhebungstag auffällt.

4. Tauschen Sie sich aus!

Das Sprichwort: „Vier Augen sehen mehr als zwei.“ bewahrheitet sich auch in der empirischen Forschung. Insbesondere dann, wenn man sehr mit seiner Literatur und seinen Planungen vertraut ist, entwickelt man blinde Flecken, schätzt mögliche Probleme geringer ein, als sie sich in der Praxis ergeben könnten oder verliert aus anderen Gründen den Durchblick. Suchen Sie sich also aktiv Rückmeldungen zu Ihrem Projekt. In jedem Fall und in jeder Phase des Projekts hilft der Diskurs mit anderen Forschenden. Besonders in der Frühphase der Projektplanung kann der Austausch mit anderen Forschenden dabei helfen, Desiderate der Forschung zu entdecken, eigene Forschungsfelder einzugrenzen und Forschungsfragen zu spezifizieren. Hierzu bieten Hochschulen und Verbände, wie etwa das Nachwuchsnetzwerk des Symposium Deutsdidaktik, regelmäßig Kolloquien und Workshops für Nachwuchswissenschaftlerinnen und Nachwuchswissenschaftler an, in denen Sie sich beraten lassen können und leicht mit anderen Forschenden in Kontakt treten.

5. Seien Sie kreativ!

Empirische Verfahren bieten eine Vielzahl von Möglichkeiten, auf unterschiedliche Weise wissenschaftliche Fragen zu beantworten. Es kann bei der Entwicklung eines eigenen empirischen Settings große Freude bereiten, eigene Wege zu gehen und bestehende Verfahren abzuwandeln oder mit Blick in die Nachbarwissenschaften dort verbreitete Methoden für deutschdidaktische Fragestellungen zu adaptieren. Gleichzeitig versteht sich dieser Tipp nicht als Legitimation für hinter Kreativität verschleierte Faulheit. Auch in Methodenfragen gilt Tipp 3: ‚Kennen Sie Ihr Feld!‘.

6. Achten Sie auf den Arbeitsaufwand!

Ein großes Problem in Qualifikationsarbeiten ist das Gleichgewicht von Ambition und Leistbarkeit. Arbeiten von Nachwuchswissenschaftlerinnen und Nachwuchswissenschaftlern haben meist geringere zeitliche, finanzielle und personelle Ressourcen als drittmittelgeförderte Großprojekte und müssen diesen in der Planung und Durchführung Rechnung tragen. Manche Verfahren erscheinen auf den ersten Blick verlockend leicht durchzuführen oder spezielle Nebenwege zur eigentlichen Forschungsfrage bei geringem Aufwand sehr ertragreich – all dies muss man im Einzelfall entscheiden. Wichtig ist aber, das zentrale Ziel nicht aus den Augen zu verlieren und sich nicht zu verzetteln. Auch hier hilft Ihnen der Austausch mit Fachkolleginnen und -kollegen.

Literatur

Abel, Jürgen/Möller, Renate/Treumann, Klaus Peter: Einführung in die empirische Pädagogik. Stuttgart: Kohlhammer 1998.

Atteslander, Peter: Methoden der empirischen Sozialforschung. Berlin: Walter de Gruyter 2003.

Diekmann, Andreas: Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen. Reinbek bei Hamburg: Rowohlt-Taschenbuch-Verl. 2005.

Friedrichs, Jürgen: Methoden empirischer Sozialforschung. Opladen: Westdt. Verl. 1990.

Raithel, Jürgen: Quantitative Forschung. Ein Praxiskurs. Wiesbaden: VS, Verl. für Sozialwiss. 2008.

Interviewverfahren

Ein Überblick

1. Interviews und ihr Potenzial in Forschungsprojekten

In der Praxis der empirisch forschenden Deutschdidaktik gehören Interviews mittlerweile zum festen Repertoire und finden breite Anwendung im Rahmen verschiedener Fragestellungen und Forschungskontexte. *Das* Interview zur Erhebung von inhaltlichen Daten gibt es allerdings nicht – unter dem Begriff „Interview“ wird in der Methodenliteratur eine Fülle an Erhebungsmethoden summiert, die jeweils mit unterschiedlichen inhaltlichen Forschungsinteressen verknüpft sind. Bevor sich die nachfolgenden Beiträge dieses Sammelbandes konkret mit einzelnen Interviewformen befassen, sollen in diesem Kapitel zunächst allgemeine Prinzipien und forschungspraktische Schritte diskutiert werden, die *übergreifend* für den Einsatz von Interviewverfahren im Rahmen deutschdidaktischer Forschungsprojekte gelten.

1.1 Funktion und Grundprinzipien von Interviews

Im Gegensatz zur videographischen Datenerhebung (s. Maak in diesem Band) oder dem Einsatz von Fragebögen (s. Betz in diesem Band) werden Interviews eingesetzt, um einen vorrangig deutenden und verstehenden Zugang zu *subjektiven Sichtweisen* von Personen zu erhalten.¹ Ziel von Interviews ist die Erhebung von verbalen Daten zur *Rekonstruktion* von Wissen, Erfahrungen und Ereignissen aus der individuellen Perspektive der Befragten.² Im Bereich der Deutschdidaktik bieten sich Interviews beispielsweise zur Erfassung forschungsrelevanter Rezeptions- und Interpretationsprozesse von Schülerinnen und Schülern (z.B. Pieper et al. 2004) oder zur Erhebung der Perspektiven von (angehenden) Lehrkräften zu bestimmten Themen oder Methoden des Deutschunterrichts an (u.a. Wieser 2008). Neben dieser inhaltsbezogenen Ebene, die in diesem Beitrag im Vorder-

¹ Einzige Ausnahme ist das standardisierte Interview (vgl. Schnell 2012), bei dem (externe) Wissensbestände durch vorgefertigte Fragen und Antwortmöglichkeiten bereits vorgegeben sind.

² Interessieren hingegen vor allem die *kollektiv geteilten* Wissensbestände, so bietet sich das Erhebungsverfahren der Gruppendiskussion (s. Scherf in diesem Band) an.

grund steht, können Interviews alternativ dazu verwendet werden, die Sprachproduktion von Personen, z.B. im Bereich Deutsch als Fremdsprache/Deutsch als Zweitsprache, zu analysieren (vgl. ausführlich Mezger/Schroeder/Şimşek 2014).

Im Anschluss an Helfferich (2014, 561ff.) können vier grundlegende Prinzipien formuliert werden, die bei der Verwendung von Interviewverfahren zu bedenken sind:

- *Kommunikationssituation*: Der Zugang zu den Sichtweisen der Erforschten ergibt sich in einer Kommunikationssituation, in der interaktiv zwischen Interviewerin bzw. Interviewer und Befragten ein ‚Text‘ erzeugt wird, der nachfolgend Basis für die Auswertung ist (vgl. Abschnitt 1.3). Es gilt daher „stets zu reflektieren, unter welchen Bedingungen die spezifische [Text-]Version entstanden ist“ (ebd., 562).
- *Größtmögliche Offenheit*: Die Perspektiven von Fachdidaktikern und Lernenden bzw. Lehrenden können sich unterscheiden – Aufgabe des Forschenden ist es daher, das Interviewsetting und die Kommunikation so zu gestalten, dass den Erforschten ein offener Äußerungsraum angeboten wird, in dem sie ihre Sichtweisen – in der ihnen *eigenen Logik* und *mit ihren eigenen Begriffen* – frei entfalten können.
- *Strukturierung*: Als Interviewerin bzw. Interviewer ist man *Ko-Konstrukteur* der eigenen Daten. Anhand der ausgewählten Erzählaufforderungen und der weiteren Nachfragen im Interview (vgl. Abschnitt 1.2.2) werden bereits bestimmte Setzungen zum Untersuchungsgegenstand vorgenommen, die die Kommunikation strukturieren.
- *Rollenausgestaltung im Interview*: Auch wenn idealtypisch eine Begegnung ‚auf Augenhöhe‘ anzustreben ist, bildet die Interviewsituation immer eine asymmetrische Kommunikationssituation. So verhält sich die interviewende Person notwendigerweise strategisch, da sie von einem bestimmten Forschungsinteresse geleitet ist. Diesen Aspekt gilt es sowohl in der Gesprächssituation als auch für die Analyse zu reflektieren.

Bereits die Auswahl eines bestimmten Interviewverfahrens hat zentrale Auswirkungen auf den weiteren Untersuchungsprozess: Die Interviewform bestimmt (mit), *welche* Daten und *welche* Forschungsergebnisse in einer Studie tatsächlich generiert werden (vgl. Friebertshäuser/Langer 2010, 438f.). Wie einleitend angeführt, ist in der Forschung eine Vielzahl an sehr unterschiedlichen Formen unter der Bezeichnung „Interview“ etabliert, die häufig sogar noch in vielfältigen Varianten konkret umgesetzt werden können.³ Das heißt auch: Im Rahmen einer Studie muss man sich nicht nur für eine spezifische Interviewform entscheiden, sondern auch dokumentieren, *wie* genau diese im Rahmen der eigenen Untersuchung

³ Die Systematisierung der einzelnen Varianten ist in der Methodenliteratur nicht immer einheitlich: „Mal zielt die Bezeichnung auf einen spezifischen Forschungsgegenstand oder Anwendungsbereich, mal wird eine Unterscheidung nach Besonderheiten der Erhebungs- oder auch Auswertungsstrategien getroffen“ (Helfferich 2011, 36).

methodisch umgesetzt wurde (vgl. Kruse 2015, 149f.). Tabelle 1 gibt einen (selektiven) Einblick in Interviewverfahren, mit denen deutschdidaktische Forschungsfragen bearbeitet werden können. In dem tabellarischen Überblick habe ich die Grundzüge der ausgewählten einzelnen Verfahren auf fach- bzw. deutschdidaktische Fragestellungen hin konkretisiert und um mögliche ‚Anwendungsbereiche‘ ergänzt; die mit einem * gekennzeichneten Verfahren werden ausführlich in diesem Sammelband diskutiert.

Interviewverfahren	Methodik	z.B. Erhebung von ...
Dilemma-Interview (Colby et al. 1987)	Zu Beginn des Interviews wird den Befragten ein Dilemma (z.B. Schlüsselprobleme in Unterrichtssituationen) mit der Aufforderung präsentiert, mögliche Lösungsansätze zu erläutern	Bewertungen und Urteilsprozessen; Argumentationsgängen
Experteninterview (Meuser/Nagel 1991)	Befragung von Experten zu einem bestimmten Thema oder Bereich des Deutschunterrichts	Vorstellungen von Ausbildern, Fachdidaktikern, Fachwissenschaftlern etc.
Fokussiertes Interview (Merton/Kendall 1979)	Vorgabe von Reizen zur Fokussierung auf einen spezifischen deutschdidaktischen Gegenstand (Filmausschnitt, Unterrichtsmaterial, Schreibprodukte von Lernenden etc.)	subjektivem Erleben und Sinnverstehen zu einem Gegenstand oder einer Unterrichtssituation
Leitfadeninterview* (Hopf 1978)	Strukturierung des Gesprächs durch einen vorab erstellten Interviewleitfaden, Fragefolge und -formulierung ist aber flexibel	Wissensbeständen und Erfahrungen zu einem bestimmten deutschdidaktischen Thema
Narratives Interview* (Schütze 1976)	Stegreif- oder Spontanerzählung mittels einer erzählgenerierenden Frage/Erzählaufforderung	autobiographischen Erfahrungsausschnitten und Handlungsprozessen; Subjektiven Theorien
Problemzentriertes Interview* (Witzel 1985)	Spezifisches Problem oder Gegenstandsbereich wird im Interview aus Sicht der Befragten dargestellt und erörtert	problem- und kontextorientiertem Sinnverstehen, Vorstellungen
Standardisiertes Interview/ Survey-Befragung (Schnell 2012)	Quantitative Interviewvariante, bei der Fragen, Reihenfolge und Antwortmöglichkeiten weitgehend vorgegeben sind	quantitativen Aussagen, z.B. zum Gebrauch von bestimmten Medien oder Unterrichtsmaterialien

Tab. 1: Auswahl Interviewverfahren (eigene Darstellung, modifiziert nach Börtz/Döring 2015, 315)

Entscheidend für die Auswahl einer bestimmten Interviewform ist – wie generell bei der Bestimmung der Erhebungs- und Auswertungsmethoden im Forschungsprozess – die *Gegenstandsangemessenheit* der Methode (vgl. u.a. Przyborski/Wohlrab-Sahr 2014, 36ff.). Mit anderen Worten: Welche Interviewform

ist das richtige Erhebungsinstrument für das Ziel meiner Untersuchung bzw. mit welchem Verfahren können wesentliche Aspekte meiner Forschungsfrage erfasst werden? Die einzelnen angeführten Interviewverfahren schließen sich dabei wechselseitig nicht grundsätzlich aus: In der Forschungspraxis werden zunehmend verschiedene Verfahren miteinander kombiniert, z.B. leitfadengestützte Interviews mit narrativen Elementen oder Experteninterviews mit problemzentrierten Anteilen – entscheidend ist auch hier das Erkenntnisinteresse der eigenen Untersuchung. Eine Orientierungshilfe für die Interviewauswahl bieten die Fragen bei Helfferich (2011, 45).

1.2 Interviews planen und durchführen

Die bisherigen Ausführungen sollten bereits implizit einen Einblick gegeben haben, dass die Anwendung von Interviewverfahren einen durchaus hohen Anspruch aufweist. Nachfolgend werden daher einige zentrale Aspekte der Planung und Durchführung von Interviews diskutiert, die – unabhängig von der gewählten Interviewform – für den Untersuchungsprozess zu reflektieren sind.

1.2.1 Inhaltliche und organisatorische Vorbereitung

Interviews beginnen mit dem Zugang zum Feld, also der *Kontaktaufnahme* zu potentiellen Studienteilnehmerinnen und -teilnehmern (zu geeigneten Strategien vgl. ebd., 171ff.). Wichtig ist es, die Teilnehmenden über Inhalte und Ziele des Projekts zu informieren. Um beispielsweise Effekte der sozialen Erwünschtheit im Gespräch zu vermeiden, dürfen in der Regel keine forschungstheoretischen Annahmen (i.S.v. „Ich gehe davon aus, dass ...“) gegenüber den potentiellen Interviewpartnern mitgeteilt werden (vgl. Kruse 2015, 254ff.).

Neben der Kontaktaufnahme und Terminabsprache sollte vorab die Wahl des *Erhebungsortes* reflektiert werden, um eine möglichst angenehme und damit offene Gesprächsatmosphäre zu schaffen. Hier empfiehlt es sich, dass ein Ort ausgewählt wird, der für die Befragten vertraut ist bzw. an dem sie sich am wohlsten fühlen (z.B. in deren Wohnung oder in einem öffentlichen, ruhigen Raum). Es bietet sich deshalb an, die Interviewpartner zuvor zu fragen, welcher Treffpunkt für sie angenehm und geeignet erscheint (vgl. dazu Przyborski/Wohlrab-Sahr 2014, 63ff.). Natürlich sollte auch das Interviewmaterial (Aufnahmegerät, Ersatzbatterien, Interviewleitfaden, etc.) vor Gesprächsbeginn sorgfältig geprüft werden, um Unterbrechungen, technische Probleme u.Ä. in der Erhebungssituation zu vermeiden.

Kommt es schließlich zum vereinbarten Interviewtermin, sollte man nicht direkt mit dem Interview beginnen. Um eine geeignete Gesprächsatmosphäre zu schaffen, ist zunächst eine kurze *Einführungsphase* sinnvoll, in der man sich kennenlernt und den Rahmen des Interviews umreißt (*Joining*, vgl. ebd., 67f.). Gegebenenfalls kann das Aufnahmegerät bereits während dieser Small-Talk-Phase eingeschaltet sein – vorausgesetzt der bzw. die Befragte ist einverstanden. So können sich alle Beteiligten an die Aufnahmesituation gewöhnen und inhaltlich Zentrales, das bereits vor dem eigentlichen Interview besprochen wird, wird zugleich digital gesichert.

1.2.2 Durchführung und Aufzeichnung

In der Rolle als Interviewender ist man dafür verantwortlich, dass die Kommunikation gelingt, schließlich ist die Mitwirkung der Erforschten eine maßgebliche Voraussetzung für die Qualität der Daten. Ein zentrales Interviewprinzip ist deshalb das (aktive) *Zuhören*: Die Erzählerperson und seine bzw. ihre Darstellungen stehen im Mittelpunkt, weshalb man als Interviewer eigene Deutungen und Bewertungen im Gesprächsverlauf zurücknimmt.⁴ Während des Interviews sollte der Interviewende vielmehr eine ‚interessierte Naivität‘ durch Nachfragen zeigen, um die Perspektive(n) der Erforschten zu verstehen. Und auch wenn dies manchmal schwer fällt: Aktives Zuhören umfasst die Fähigkeit, Gesprächs- und Denkpausen in der Kommunikation auszuhalten und sich mit Fragen oder Kommentaren zurückzunehmen bis sog. Abschlussmarkierer von Seiten des Erforschten erfolgen (z.B. „Das war’s.“, „Mehr fällt mir dazu nicht ein.“). Nur so kann sichergestellt werden, dass die Interviewten Gelegenheit haben, *ihren individuellen Sinn*, d.h. was für sie selbst zum Thema wichtig ist, in der ihnen eigenen Art und Weise zu entwickeln (vgl. Kruse 2015, 148).

Die konkreten *Fragen und Stimuli* in einem Interview fokussieren und strukturieren das Gespräch, um aussagekräftiges Material zu gewinnen (vgl. Friebertshäuser/Langer 2010, 439). Dazu vorneweg: *Forschungsfragen sind keine Interviewfragen*. Die zentralen Fragen der Untersuchung können deshalb nicht einfach ungefiltert für die Kommunikation übernommen werden. Zum einen muss also für das Interview geklärt werden, welche Themenkomplexe für den gewählten Gegenstand relevant sind. Zum anderen müssen die zentralen Fragen der eigenen Untersuchung *hinreichend konkret* und *angemessen anschaulich* in Interviewfragen überführt werden. Dazu ein Beispiel aus einem schreibdidaktischen Seminar, in dem Interviewfragen zum Thema ‚Schreibprozesse im Beruf‘ entwickelt werden sollten:

Entwurf

(1) Welche kommunikativen Prozesse gibt es in Ihrem Alltag?

Überarbeitung

(2) Wenn du deinen heutigen Tag Revue passieren lässt, wann und in welchen Situationen hast du heute geschrieben?

Bsp. 1: Entwurf und Überarbeitung einer Interviewfrage
(eigene Darstellung, entnommen aus Lehnen/Schindler 2010, 246f.)

Was fällt auf? Erkennbar ist vor allem, dass der Fragenentwurf, neben der eher ungünstigen Passiv-Formulierung, doch relativ abstrakt bleibt – auf welche Inhalte zielt die Formulierung „kommunikative Prozesse“ bzw. was soll hier erfragt werden? Im Gesprächskontext ist zudem die *Orientierung am Adressaten* zentral (s.o.). So können zu abstrakte Begriffe ggf. nicht verstanden werden oder die Ver-

⁴ So wird nicht selten unterschätzt, dass das kompetente (Durch-)Führen von Interviews einen Erwerbsprozess darstellt und daher vor der Haupterhebung geübt werden sollte.

wendung fachdidaktisch einschlägiger Begriffe auf die Befragten eher abschreckend wirken. Dies kann im schlechtesten Fall bedeuten, dass die Gewinnung aussagekräftiger empirischen Daten eingeschränkt wird.

Die *Einstiegsfrage* ist bei jeder Interviewform das Einfallstor für das Gespräch und (mit) entscheidend für den weiteren Interviewverlauf und entstehende Forschungsergebnisse. Am Anfang der Kommunikation sollte daher eine erzählgenerierende und so spezifisch wie nötig gehaltene Eingangsfrage stehen, die im Sinne der Vergleichbarkeit möglichst bei allen Interviews gleich gehalten wird (vgl. Kruse 2015, 68ff.). Ein Interviewausschnitt aus einer Studie, in der die Erhebung der Sichtweisen von Deutschlehrkräften zu einem lesediagnostischen Diagnoseinstrument ein zentrales Element darstellte (vgl. Schmidt 2013), soll dies verdeutlichen:

I: Ja, dann würde ich dich erst einmal bitten, dein, äh, den allgemeinen Eindruck, den du gewonnen hast von dem Tool, zu beschreiben.

Ans: Ja, ich habe es eigentlich gerne gemacht. Äh, ich muss dazu sagen, zeitlich war es ein bisschen, ähm, äh, problematisch (I: Mmh.). Weil wir eben auch die mündlichen Prüfungen hatten, die schriftlichen Zentralprüfungen (I: Mmh.) und so weiter. Und auch das gezielte, äh, Üben dafür. Aber ich habe mir dann doch zwei-, dreimal die Zeit genommen (I: Mmh.) und einen Schüler fast fertig auch bekommen. (I: Mmh.) Und das, äh, lief also meiner Meinung nach auch wirklich gut. Und ich habe das nachher also auch wirklich gerne gemacht. Also nachher be/das, was ich wirklich gut fand: es war immer das gleiche System. (I: Mmh.) Und wenn man das erst einmal/wenn man erst einmal so reingekommen ist, allein, ich glaube, immer in diese fünf Punkte (I: Mmh.), äh, äh, und nachher das System drauf hatte, äh, sage ich aber nachher noch was zu, dann hat das richtig Spaß gemacht. (I: Ja.) Und das ging eigentlich auch ratzfatz muss ich sagen. Zu einigen, äh, äh, Unterpunkten zu den, äh, Peergroups zum Beispiel oder Elternhaus. (I: Mmh.) Das war ein bisschen problematisch für mich. Äh, also ich denke, da muss man noch so ein bisschen, oder hätte ich die Schüler ein bisschen besser kennenlernen müssen. (I: Mmh.) Also (.) aber ansonsten muss ich sagen, äh, zum Formulierungen war teilweise ein bisschen, äh, lang. Also, die Sätze (I: Mmh.) ein bisschen lang. Und, und teilweise auch ein bisschen kompliziert. Manchmal musste man umdenken. (I: Mmh.) Äh, man hat, die Frage davor war eigentlich positiv formuliert (I: Mmh.) oder, oder die Beobachtung. Die nächste war dann negativ oder so, wenn ich das noch richtig in Erinnerung habe. (I: Und dass man dann) Und das hat so ein bisschen, äh, Kuddelmuddel gemacht. Aber, oder sagen wir so, man musste sich dann sehr konzentrieren. (I: Mmh.) Ne. Aber ansonsten, muss ich sagen, hat das schon Spaß gemacht. Und, äh, vor allen Dingen, wenn man überlegt, äh, als ich das nachher farbig ausgedruckt gesehen habe. (I: Mmh.) Ja, da ist mir dann auch ((leicht amüsiert)) einiges klar geworden.

Bsp. 2: Einstiegsfrage für ein Interviewgespräch

Der Interviewausschnitt zeigt, dass der *erzählgenerierende Stimulus*, hier die Frage nach dem individuellen Eindruck über ein lesediagnostisches Beobachtungsinstrument, den Erfahrungsbereich sowie subjektive Schwerpunktsetzungen von Seiten der Deutschlehrkraft entfaltet. Im Anschluss an diese erste längere Passage des Erzählens erfolgen in der Regel *immanente Nachfragen*, in denen Äußerungen der Befragten aufgegriffen und verständnisgenerierende Fragen ge-

stellt werden („Kannst du mir das noch einmal genauer beschreiben?“, „Was genau meinst du mit XY?“ usw.). Im Sinne des „Fremdverstehens“ (Helfferich 2011, 84) eröffnet dieses Vorgehen die Möglichkeit, dass die Befragten ihre Sichtweisen breiter bzw. vertiefend entfalten können. Es ist wichtig, den Gesprächseinstieg – wie auch die weiteren entwickelten Interviewfragen – im Rahmen einer *Vorerhebung* (Pilotstudie) zu überprüfen. So können etwa Frageänderungen, die während der Haupterhebung erfolgen, zur Verzerrung der Validität der Forschungsergebnisse führen. Eine Pilotstudie bietet darüber hinaus Gelegenheit, das eigene Interviewverhalten zu üben und erste Erkenntnisse zum Untersuchungsgegenstand zu sammeln, ohne dass bereits die ‚eigentliche‘ Erhebungssituation begonnen hat (s. ausführlicher Schmidt in diesem Band).

Wenn die interviewende Person den Eindruck gewinnt, dass sämtliche immanenten Nachfragen umfassend besprochen wurden, können weitere *exmanente Nachfragen* erfolgen, die sich auf das spezifische Erkenntnisinteresse der Untersuchung beziehen. Auch Bildimpulse (s. Binanzer in diesem Band) oder Vignetten (s. Benz in diesem Band) können hier herangezogen werden, um die Sichtweisen der Erforschten im Interview vertiefend zu erfassen. Insgesamt ist das dargestellte Vorgehen davon abhängig, welche der verschiedenen Interviewformen (Tab. 1) für die Untersuchung ausgewählt wurde.

Eine besondere Herausforderung für alle Varianten von Interviews ist die *Gesprächsführung* bzw. genauer: Die Fähigkeit, sich im Interview *führen zu lassen*. Die Strukturierung der Kommunikation sollte nämlich weitestgehend durch die Befragten erfolgen. Ein wiederkehrendes Problem in Interviews sind vor diesem Hintergrund geschlossene Fragen, die vom Prinzip her mit „ja“ oder „nein“ beantwortet werden können. Diese Interviewfragen eröffnen nur einen geringen Spielraum für die Erforschten und bieten somit keinen Erkenntnisgewinn in Bezug auf ihr individuelles Sinnverstehen. Ebenso gilt es Suggestivfragen oder Fragen mit negativem Unterton in Form von „Sind Sie nicht auch der Meinung, dass ...“ zu vermeiden, da mit diesen Fragetechniken wiederum vielmehr die eigenen Ansichten als die der Befragten erhoben werden. Eine Übersicht zur Zulässigkeit bestimmter Frageformen bei einzelnen Interviewvarianten findet sich bei Helfferich (2011, 107).

Auch der *Abschluss des Interviews* darf nicht unterschätzt werden. Im letzten Gesprächsteil sollte den Befragten nochmals Freiraum für die Darstellung individueller Schwerpunkte geschaffen werden. Sehr häufig wird deshalb die folgende Abschlussfrage verwendet: „Ist dir/Ihnen während des Interviews noch etwas durch den Kopf gegangen, das wir noch nicht besprochen haben?“. Nach meiner Erfahrung nutzen die Interviewten oftmals den mit dieser Frage verbundenen Gestaltungsspielraum, um Aspekte anzusprechen, die im Interviewverlauf noch nicht thematisiert wurden.

Es bietet sich an, erst nach dem Interview den ergänzenden *Kurzfragebogen* für die Erfassung der biographischen Daten einzubringen (wenn dies für den Interviewverlauf nicht zentral ist), damit das Interview nicht von diesem eher formalen

Vorgang beeinflusst wird. Unerlässlich ist es weiterhin, die Befragten eine *Einverständniserklärung*⁵ unterzeichnen zu lassen, damit die erhobenen Daten für Forschungszwecke verwendet werden können (vgl. ebd., 191f.). Nach dem offiziellen Ende des Interviews schließt sich eine *informelle Gesprächsphase* an, die nochmals Aufschlussreiches für die Untersuchung eröffnen kann (vgl. Przyborski/Wohlrab-Sahr 2014, 73). Mit geringem zeitlichen Abstand sollte anschließend ein *Interviewprotokoll* erstellt werden, in dem vor allem Interviewsituation und Interviewverlauf – wie Rahmenbedingungen des Interviews, Auftreten des Interviewten, Verfassung des Interviewers etc. – erfasst werden (für Beispiele vgl. Kruse 2015, 278ff.). Die *systematische Dokumentation* von Eindrücken, Beobachtungen usw. bietet im Rahmen der Auswertung eine wichtige Ergänzung für das vorliegende Datenmaterial.

Interviews sollten als Audiodaten aufgenommen und digital gespeichert werden. Dazu gehört es auch, unmittelbar im Anschluss an das Interview Sicherheitskopien zu erstellen. Im Rahmen der Erhebung verbaler Daten muss das Datenmaterial vor der Analyse transkribiert, also zu schriftlichen Texten verarbeitet werden (s. Abschnitt 1.3).

1.3 *Quo vadis?* Datenaufbereitung und Analyse von Interviews

Die Daten, mit denen man im Rahmen von Interviewerhebungen arbeitet, sind verbale Daten, die in Form von Audio- oder Videodaten aufgezeichnet werden. Für die weitergehende Analyse müssen diese verbalen Daten in eine Schriftform überführt, d.h. transkribiert werden. Dabei handelt es sich jedoch nicht nur um einen simplen Transferschritt: Die Transkription ist eine „Schlüsselstelle“ (Przyborski/Wohlrab-Sahr 2014, 162) im Forschungsprozess (s. hierzu Exkurs im Beitrag von Heller/Morek in diesem Band). Sie ist bereits der *erste interpretative Zugang* zum Datenmaterial sowie das *Fundament* für alle weiteren Analyse-schritte, da nachfolgend die Ursprungsdaten in der Regel nur noch in Zweifelsfällen hinzugezogen werden (vgl. Langer 2010, 517). Das Sichten der verbalen Daten zur Verschriftlichung bildet insofern einen wichtigen Erkenntnisprozess, in dem ferner erste Interpretationshypothesen für die Untersuchung gebildet werden. Darüber hinaus ist die Anfertigung von Transkripten eine Arbeit, die mit einem gewissen Zeitaufwand verbunden ist – selbst mit einem recht einfachen Transkriptionssystem muss man als geübte Person mit etwa 10 Minuten Transkriptionaufwand pro Interviewminute rechnen. Folglich ist sowohl aus inhaltlicher als auch forschungspraktischer Perspektive zu entscheiden, *welches* Transkriptionssystem für die eigene Untersuchung zielführend ist. Mempel und Mehlhorn (2014, 148) nennen zentrale Fragen, die beim Entscheidungsprozess helfen können (s. Abb. 1).⁶

⁵ Ein Beispiel findet sich bei Helfferich (2011, 203).

⁶ Weiterführende Hinweise zu den Themen *Interviewaufnahmetechnik* und *Transkription von Audio- und Videodateien* finden sich bei Kruse (2014, 264ff.). Umfassende Informationen zur wohl gängigsten Transkriptionssoftware *f4* bietet die Internetseite <http://www.audiotranskription.de>.

- An welche Adressaten richtet sich das Transkript?
 - Welches Forschungsziel wird verfolgt?
 - Welche Informationen soll das Transkript enthalten?
- Wie genau/detailliert soll transkribiert werden?
- Welche Aufgabe hat das Transkript (Arbeits-, Feintranskript)?
 - Wofür möchte ich die Transkription verwenden?
 - Ist es die ausschließliche Auswertungsgrundlage?
- Wie hoch kann der Arbeitsaufwand für die Erstellung sein?

Abb. 1: Fragen zur Transkriptausswahl (Mempel/Mehlhorn 2014, 148)

Im Anschluss an die Anfertigung der Transkripte erfolgt der ‚eigentliche‘ Analyseprozess. Zwar bestehen in der methodischen Literatur mitunter einige Empfehlungen, dennoch sind die einzelnen Interviewverfahren *nicht* an eine bestimmte Auswertungsmethodik gebunden. Zur Analyse von Interviews werden in Forschungsprojekten häufig das Verfahren der qualitativen Inhaltsanalyse (s. Heins in diesem Band), die Grounded Theory (s. Kubik in diesem Band) oder die dokumentarische Methode (s. Schieferdecker in diesem Band) genutzt (für einen Überblick vgl. Kruse 2015, 390ff.). Entscheidend dafür, welches Verfahren zur Analyse des empirischen Datenmaterials gewählt werden sollte, ist das Forschungsinteresse und der Forschungsgegenstand der Untersuchung.

1.4 Potenzial von Interviewverfahren

Welche „Qualität qualitativer Daten“ (Helfferich 2011) bieten Interviews? Interviews eignen sich im Besonderen, um individuell vorliegendes, biographisch erworbenes Wissen zu erheben, da sie einen offenen Äußerungsraum bieten, in dem die Erforschten ihre Wissensbestände, ihre Einstellungen oder auch Ereignisse präsentieren bzw. in der ihnen eigenen Logik und in eigener Sprache erzählen können. Dies ermöglicht eine dichtere Beschreibung der subjektiven Perspektiven. Zentral ist allerdings, dass sich die expliziten Auskünfte der Beforschten nicht von selbst erschließen, sondern diese interpretativ vom Forschenden erschlossen werden müssen (vgl. Abschnitt 1.3). Möchte man neben individuell vorliegenden Perspektiven auch Gruppenphänomene erheben bzw. rekonstruieren, ist es fruchtbar, Interviews und Gruppendiskussionen (s. Scherf in diesem Band) miteinander zu kombinieren (vgl. Kruse 2015, 147).

Der Einsatz von Interviews mag für die Gewinnung von empirischen Daten zunächst relativ einfach und forschungsökonomisch gut durchführbar erscheinen. Allerdings gehören gut geführte Interviews zu den anspruchsvollen empirischen Methoden, die man in der Deutschdidaktik heranziehen kann. Dies wird gelegentlich unterschätzt, sodass das Datenmaterial im Extremfall weniger die individuellen Relevanzsetzungen der Beforschten widerspiegelt, sondern vielmehr die eigenen Sichtweisen auf einen bestimmten Gegenstand (vgl. Abschnitt 1.3).

1.5 Literaturempfehlungen

Daase/Hinrichs/Settinieri (2014): Befragung.

Der Beitrag bietet einen Überblick zu Interviews und weiteren Formen der Befragung aus Perspektive der DaF-/DaZ-Forschung mit Anwendungs- und Übungsaufgaben am Ende des Kapitels. Vor allem auf Grund der fachdidaktischen Perspektive auf die Erhebungsmethode „Interview“ bietet der Beitrag eine sinnvolle Ergänzung zu den Standardwerken der empirischen Sozialforschung.

Helfferrich (2011): Qualität qualitativer Daten.

Das vielbeachtete Lehrbuch bietet Leserinnen und Lesern eine komprimierte und sehr gut verständliche Einführung zur Planung und Durchführung von Interviews, ergänzt um konkrete Beispiele, viele forschungspraktische Hinweise und Vorlagen für die Organisation, beispielsweise für Erhebungsprotokolle oder Vereinbarungen zum Datenschutz, die in den meisten Werken zum Thema fehlen. Sowohl zum Einarbeiten in die Thematik als auch zum Vertiefen einzelner Aspekte ist das Manual eine geeignete Lektüre.

Kruse (2015): Qualitative Interviewforschung: Ein integrativer Ansatz.

Das Lehrbuch bietet einen breiten Überblick und – ähnlich zu Helfferrich (2011) – zugleich eine praxisnahe Orientierung für die Planung und Durchführung von Forschungsprojekten, die mit Interviews als Erhebungsmethode arbeiten. Vor allem zum Thema Fragestil und Gesprächsführung enthält das Lehrbuch viele zentrale Hinweise.

Przyborski/Wohlrab-Sahr (2014): Qualitative Sozialforschung.

Przyborski und Wohlrab-Sahr erörtern allgemeine Prinzipien von Interviews und diskutieren ausführlich das Potenzial einzelner Interviewverfahren. Besonders die Erläuterungen hinsichtlich forschungspraktischer Schritte zur Erhebungsmethode „Interview“ und zahlreiche praktische Hilfestellungen, beispielsweise in Bezug auf Interviewfragen, bieten eine gute Orientierung.

Reinders (2012): Qualitative Interviews mit Jugendlichen führen.

Interviews mit Schülerinnen und Schülern stellen Forschende vor spezifische Herausforderungen. Auch wenn das Lehrbuch eher übergreifend das Thema Interviews mit Jugendlichen fokussiert, bietet es eine fundierte und anschaulich aufbereitete Orientierung zur Thematik. Merksätze und knappe Zusammenfassungen nach jedem Kapitel bündeln zudem jeweils die zentralen Aspekte.

Literatur

- Bortz, Jürgen/Döring, Nicola: Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Kartonierte Sonderausgabe. Heidelberg: Springer Verlag 2015.
- Colby, Anne et al. (Hrsg.): Measurement of Moral Judgement. Vol. 1. Theoretical Foundations and Research Validitation. New York: Cambridge University Press 1987.
- Daase, Andrea/Hinrichs, Beatrix/Settinieri, Julia: Befragung. In: Settinieri, Julia et al. (Hrsg.): Empirische Forschungsmethoden für Deutsch als Fremd- und Zweitsprache. Eine Einführung. Paderborn: Schöningh 2014, 103-122.
- Friebertshäuser, Barbara/Langer, Antje: Interviewformen und Interviewpraxis. In: Friebertshäuser, Barbara/Langer, Antje/Prenzel, Annedore (Hrsg.): Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft. Weinheim/München: Juventa 2010, 437-455.
- Helfferrich, Cornelia: Leitfaden- und Experteninterviews. In: Baur, Nina/Blasius, Jörg (Hrsg.): Handbuch Methoden der empirischen Sozialforschung. Wiesbaden: Springer VS 2014, 559-574.
- Helfferrich, Cornelia: Qualität qualitativer Daten. Manual für die Durchführung qualitativer Sozialforschung. Wiesbaden: VS Springer 2011.
- Hopf, Christel: Die Pseudo-Exploration – Überlegungen zur Technik qualitativer Interviews in der Sozialforschung. In: Zeitschrift für Soziologie (1978), 7, 2, 97-115.
- Kruse, Jan: Qualitative Interviewforschung: Ein integrativer Ansatz. Weinheim/Basel: Beltz Juventa 2015.
- Langer, Antje: Transkribieren – Grundlagen und Regeln. In: Friebertshäuser, Barbara/Langer, Antje/Prenzel, Annedore (Hrsg.): Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft. Weinheim/München: Juventa 2010, 515-526.
- Lehnen, Katrin/Schindler, Kirsten: Berufliches Schreiben als Lernmedium und -gegenstand. Überlegungen zu einer berufsbezogenen Schreibdidaktik in der Hochschullehre. In: Pohl, Thorsten/Steinhoff, Torsten (Hrsg.): Textformen als Lernformen. Duisburg: Gilles & Francke 2010, 233-256.
- Merton, Robert K./Kendall, Patricia L.: Das fokussierte Interview. In: Hopf, Christel/Weingarten, Elmar (Hrsg.): Qualitative Sozialforschung. Stuttgart: Klett-Cotta 1979 [zuerst 1945/1946], 171-204.
- Meuser, Michael/Nagel, Ulrike: ExpertInneninterviews – vielfach erprobt, wenig bedacht. Ein Beitrag zur qualitativen Methodendiskussion. In: Garz, Detlef/Kraimer, Klaus (Hrsg.): Qualitativ-empirische Sozialforschung. Konzepte, Methoden, Analysen. Opladen: Westdeutscher Verlag 1991, 441-471.
- Mezger, Verena/Schroeder, Christoph/Şimşek, Yazgül: Elizitierung von Lernersprache. In: Settinieri, Julia et al. (Hrsg.): Empirische Forschungsmethoden für Deutsch als Fremd- und Zweitsprache. Eine Einführung. Paderborn: Schöningh 2014, 73-86.
- Pieper, Irene et al.: Lesesozialisation in schriftfernen Lebenswelten. Lektüre und Mediengebrauch von HauptschülerInnen. Weinheim/München: Juventa 2004.
- Przyborski, Aglaja/Wohlrab-Sahr, Monika: Qualitative Sozialforschung. Ein Arbeitsbuch. München: Oldenbourg Wissenschaftsverlag 2014.

- Reinders, Heinz: Qualitative Interviews mit Jugendlichen führen. Ein Leitfaden. München: Oldenbourg Wissenschaftsverlag 2012.
- Schmidt, Frederike: Den Schüler im Blick – Zur Entwicklung und Erprobung eines Diagnostiktools für Lesekompetenz in der Sekundarstufe I. In: Komorek, Michael/Prediger, Susanne (Hrsg.): Der lange Weg zum Unterrichtsdesign – Zur theoretischen Begründung und Umsetzung eines genuin fachdidaktischen Forschungs- und Entwicklungsprogramms. Münster: Waxmann 2013, 257-272.
- Schnell, Rainer: Survey-Interviews. Methoden standardisierter Befragungen. Wiesbaden: VS Verlag 2012.
- Schütze, Fritz: Zur Hervorlockung und Analyse von Erzählungen thematisch relevanter Geschichten im Rahmen soziologischer Feldforschung. In: Arbeitsgruppe Bielefelder Soziologen (Hrsg.): Kommunikative Sozialforschung. München: Fink 1976, 159-260.
- Wieser, Dorothee: Literaturunterricht aus Sicht der Lehrenden. Eine qualitative Interviewstudie. Wiesbaden: Springer VS 2008.
- Witzel, Andreas: Das problemzentrierte Interview. In: Jüttemann, Gerd (Hrsg.): Qualitative Forschung in der Psychologie. Grundfragen, Verfahrensweisen, Anwendungsfelder. Weinheim/Basel: Beltz 1989, 227-256.

Problemzentrierte Interviews

Zwischen Offenheit und Theorie

1. Vorstellung des problemzentrierten Interviews

Das problemzentrierte Interview (im Folgenden PZI) ist eine häufig genutzte hypothesengenerierende Interviewmethode, die Theoriegeleitetheit und Offenheit miteinander verbindet. Im Interview sollen die Befragten zum freien Erzählen animiert werden. Gleichzeitig hat der Forschende die Möglichkeit, Nachfragen zu stellen und nach einem offenen Leitfaden weitere Aspekte des spezifischen Problems, das erörtert werden soll, aufzugreifen (vgl. Witzel 1982, 66). Gewonnene Erkenntnisse können auf diese Weise stetig überprüft bzw. vertieft werden, sodass sich ein deduktiv-induktives Wechselspiel ergibt (vgl. Witzel 2000). Da Andreas Witzel (1982) das problemzentrierte Interview als Methodenmix einführte, ist auch das Einbeziehen weiterer Erhebungsinstrumente möglich.

1.1 Funktionsweise des problemzentrierten Interviews

Das PZI kombiniert narrative und strukturierte Interviewelemente. Der Umgang mit dem theoretischen Vorwissen hat bei dieser Interviewmethode eine besondere Bedeutung. Es ist eine wichtige Voraussetzung zur Planung, Durchführung und Auswertung der Interviews, muss aber konsequent offengelegt werden und darf weder bei der Datenerhebung noch bei ihrer Auswertung den Blick auf Neues, Widersprechendes oder Überraschendes verstellen. Begründet wird dieses Vorgehen damit, dass Vorwissen oder bereits bestehende – evtl. unbewusste – Konzepte zur Fragestellung der Erhebung immer existieren (vgl. Lamnek 2010, 333). Daher sollten diese besser wissenschaftlich fundiert, bewusst reflektiert und für Datenerhebung genutzt werden, anstatt ungefiltert, unbewusst und unreflektiert in die Forschung einzugehen.

Nachdem der Forschende ein gesellschaftlich relevantes Problem identifiziert hat, ist seine Aufgabe daher, sich sein Wissen zu dem Thema bewusst zu machen und dieses zu erweitern. Zu diesem Vorwissen gehört Alltagswissen genauso wie Kenntnis über relevante Forschungsergebnisse und Theorien sowie eventuell praktische Gegebenheiten, Abläufe oder Ähnliches.

At this stage answers need to be found to guiding questions like these: What do we already know about the problem in question (from previous research, media, discussions, everyday experience, etc.)? What are the key dimensions of the problem? Which concepts and theories could be used to frame the issue; which are typically used; and which could provide alternative perspectives? What do I want to know in addition to all this; and what do I want to find out in the course of my research? (Witzel/Reiter 2012, 40)

Auf der Basis dieses Wissens wird anschließend ein Gesprächsleitfaden für die Interviews entwickelt, der vor allem dazu dient, die wichtigsten Themenbereiche des Problems festzuhalten. Denn dieser Leitfaden soll keinesfalls in einem Frage-Antwort-Spiel abgearbeitet werden, sondern dient als Gedächtnisstütze und thematische Gliederung für das geplante Interview. Da Witzel (1994) für das PZI ursprünglich die Möglichkeit eines Methodenmixes vorgesehen hat, können problemlos weitere Instrumente für die Datenerhebung wie Gruppendiskussionen (s. Scherf in diesem Band) oder die Erhebung biographischer Daten einbezogen werden (vgl. Witzel/Reiter 2012, 7). Von dieser Möglichkeit sieht die Mehrzahl der Forschenden, die mit dem PZI arbeiten, ab. Lediglich Kurzfragebögen werden häufig genutzt, um vorab (oder im Anschluss an das Interview) von den Probandinnen und Probanden gezielt allgemeine Angaben zu erheben, die keines Gesprächs bedürfen wie beispielsweise das Alter, die Staatszugehörigkeit, den Beruf oder die Muttersprache des Befragten. Durch die Ausgliederung dieser Angaben kann ein Beitrag dazu geleistet werden, dass das Interview eher einem Gespräch und weniger einem ‚Ausfragen‘ der Probandinnen und Probanden ähnelt (vgl. Schmidt-Grunert 1999, 42).

Neben diesen inhaltlichen Vorbereitungen ist auch zu erwägen, welche Zielgruppe interviewt werden soll, wie sinnvoll Kontakt zu den Probandinnen und Probanden aufgenommen werden kann, wie viele Probeinterviews notwendig sind, um den Gesprächsleitfaden zu testen und zu überarbeiten und schließlich wo die Interviews stattfinden sollen. Zudem gilt es, die Handhabung eines geeigneten Audio-Aufnahmegerätes sicher zu beherrschen (vgl. ebd., 47ff.). Als Verfahren zur Auswahl der Probandinnen und Probanden empfiehlt Witzel (1994) das *Theoretische Sampling* nach Glaser und Strauss, da es sehr gut mit dem induktiv-deduktiven Vorgehen der Methode zusammenwirkt. So kann auch die Probandenwahl als Prozess begriffen werden, der auf erste Erkenntnisse oder neu aufkommende Fragen reagieren kann. Denkbar sind aber auch andere Methoden, um eine sinnvolle Auswahl von Probandinnen und Probanden zu treffen (vgl. Witzel 1994, 80 und Witzel/Reiter 2012, 61ff.).

Bei der Durchführung der Interviews ist es bedeutsam, dass die Probandinnen und Probanden Vertrauen zu den Interviewenden fassen und möglichst authentisch eigene Standpunkte und Einsichten teilen. Dazu gilt es unter anderem den richtigen Ton und den richtigen Einstieg ins Interview zu finden. Witzel und Reiter (2012) beschreiben diesen Vorgang als eine Art „sound-check“, um mittels „fine-tuning“ den richtigen Umgang mit dem jeweiligen Befragten zu finden (vgl. Witzel/Reiter 2012, 67f.). Dazu eignet sich insbesondere eine Frage, die einen narrativen Erzählfluss der Probandin oder des Probanden anregt. Dabei ist es weniger

entscheidend, ob die Erzählung des Befragten im Sinne der Fragestellung von Interesse ist. Wichtig ist vielmehr, dass zunächst eine lockere Erzählatmosphäre initiiert wird und auch die folgende Neugier des Forschenden kann durch diesen ersten Erzählfluss gerechtfertigt werden (vgl. ebd., 70f.). Denn im weiteren Interviewverlauf ist von größtem Interesse, dass die Probandin oder der Proband motiviert ist, möglichst unabhängig vom Gewünschten oder Vermuteten – auch im Widerspruch zum Vorwissen des Forschenden – ehrlich zu erzählen und zu erklären. Daher ist es wichtig, dass der Forschende sich selbst als Lernenden versteht, die Haltung annimmt und auch entsprechend wahrgenommen wird. (vgl. Witzel 1994, 73).

Das Interview läßt den Befragten möglichst frei zu Wort kommen, um einem offenen Gespräch nahezukommen. Es ist aber zentriert auf eine bestimmte Problemstellung, die der Interviewer einführt, auf die er immer wieder zurückkommt. (Schmidt-Grunert 1999, 40f.)

Im Folgenden stellt der Forschende offene Fragen, sodass der Proband oder die Probandin möglichst narrative Antworten gibt. Der Unterschied zum narrativen Interview liegt darin, dass auch Nachfragen gestellt werden können. Keinesfalls darf es dabei allerdings zu einem ‚verhörartigen‘ Ausfragen kommen. Der Forschende sollte daher immer voller Interesse das Problem mit der Probandin oder dem Probanden erörtern und sich vorurteilsfrei und unvoreingenommen zeigen (vgl. Witzel 1994, 76 und Schmidt-Grunert 1999, 45). Das Vorwissen des Forschenden sollte der Probandin oder dem Probanden zudem nicht deutlich werden. Erst im Nachhinein kann überprüft werden, ob bereits erarbeitete Konzepte mit den Aussagen der Probandin oder des Probanden übereinstimmen oder angepasst werden müssen (vgl. Lamnek 2010, 333).

Statt den Gesprächspartner auf das gesellschaftlich Wünschbare zu orientieren, versucht der Forscher durch die Art des Nachfragens, die originäre Problemsicht des Befragten im Laufe des Gesprächs zu entlocken und hinter oft kürzelhaften, stereotypen Darstellungen und Andeutungen auch auf widersprüchliche Problemsichten zu stoßen. (Witzel 1994, 70)

Umso besser sich der Forschende in die Aussagen der Probandin oder des Probanden hineindenken kann und umso umfassender er sich im Vorfeld informiert hat, desto leichter wird es ihm fallen, nur diejenigen Fragen zu stellen, die ein solches Interesse ausdrücken, das Vertrauen der Probandin oder des Probanden nicht verletzen und seinen Gesprächsfluss nicht unterbrechen. Da dies nur einem gut geschulten und für diese Feinheiten sensibilisierten Interviewer gelingen kann, ist es sinnvoll, wenn der Forschende selbst das Interview führt und diese Aufgabe nicht delegiert (vgl. Witzel 1994, 77 und 117). Konkret sollte sich möglichst ein Gespräch zum relevanten Problem entspinnen. Bei Verständnisfragen kann der Forschende das von der Probandin oder dem Probanden Gesagte zurückspiegeln und somit genauere Erläuterungen provozieren. Durch das Wiederholen der Aussage des Interviewten kann dies geschehen, ohne dass die Formulierungen einer Nachfrage eine Wertung suggeriert oder unbeabsichtigt eine bestimmte Antwort nahelegt. Aspekte des Problems, die der Gesprächsleitfaden vorsieht, die

sich jedoch nicht natürlich im Gespräch ergeben, können anschließend durch sogenannte Ad-Hoc-Fragen eingeführt werden, die wiederum als Gesprächsanlass dienen (vgl. Witzel 1994, 117). Der Gesprächsleitfaden gibt somit keine konkreten Frageformulierungen vor oder gliedert das Gespräch zwangsläufig, sondern bildet einen ‚doppelten Boden‘ auf den zurückgegriffen werden kann, um wichtige Fragestellungen nicht zu vergessen.

Zusätzlich erstellt der Interviewende nach jedem Interview ein Postskriptum, in dem möglichst ohne größere, zeitliche Distanz zum Gespräch alles festgehalten wird, das auf der Audio-Aufnahme nicht zu hören ist und für die korrekte Auswertung des Interviews von Interesse sein könnte, wie zum Beispiel die Mimik, Gestik, das Erscheinungsbild der Probandin oder des Probanden oder der erste Eindruck, den der Forschende von ihm hatte. Zudem können beispielsweise die Gesprächsatmosphäre, der „flow“ des Interviews oder die Voraussetzungen am Ort des Interviews von Interesse sein (vgl. Witzel/Reiter 2012, 95ff.).

Das Interview, das üblicherweise mit einem Audio-Aufnahmegerät aufgenommen wird, muss anschließend vollständig, in begründeten Fällen nur teilweise, transkribiert werden. Dabei können bereits erste Memos und Stichpunkte notiert werden (vgl. ebd., 98).

1.2 Ausgangsdaten und Ergebnisse

Nach dem Interview liegen dem Forschenden somit ein Transkript der Befragung, ggf. ein ausgefüllter Kurzfragebogen mit schnell abfragbaren (sozio-biographischen) Daten, ein Postskriptum des Interviews sowie der Gesprächsleitfaden vor. Letzterer kann bei der Auswertung als Orientierung dienen, indem die für den Leitfaden erarbeiteten thematischen Bereiche als ‚Etappen‘ der Interpretation dienen (vgl. Schmidt-Grunert 1999, 42f. und 51).

Lamnek (2010) formuliert im Folgenden drei Stufen der Auswertung (vgl. Lamnek 2010, 335): Zuerst findet eine „Methodologische Kommentierung“ statt, in der mögliche Missverständnisse oder Fehler des Interviewenden, wie zum Beispiel Suggestivfragen, markiert und kommentiert werden. In einem zweiten Schritt werden die einzelnen Interviews separat voneinander analysiert, was idealerweise durch eine Forschergruppe in Form von Kontrollen und Diskussionen begleitet werden kann. Wichtig ist dabei, die individuelle Sicht der Probandin oder des Probanden nachzuvollziehen.

Die Kunst eines derartigen ‚Sinnverstehens‘ besteht darin, den Bedeutungsgehalt der Textaussagen in der Perspektive des Befragten zu rekonstruieren und in deren ‚Sinn‘ zu interpretieren. (Schmidt-Grunert 1999, 50)

Bei der nun folgenden „vergleichenden Systematik“ gilt es, durch das Vergleichen der erhobenen und analysierten Einzelfälle Typisches und Kollektives zu finden. So können zum Beispiel sich wiederholende Formulierungen oder stets wiederkehrende Argumente zu einer interviewübergreifenden Interpretation führen (vgl. Witzel 1994, 110ff.).

Generell gibt das PZI keine konkrete Auswertungsmethode vor. Die gewonnenen Daten lassen sich mit allen Verfahren zur Auswertung qualitativer Daten analysieren (vgl. Witzel 2000). Naheliegender ist es jedoch die Auswertungsmethode der *Grounded Theory* nach Corbin und Strauss zu nutzen (vgl. Witzel/Reiter 2012, 100ff.), da sie ähnlich wie das PZI auf die Generierung neuer Theorien – im Gegensatz zum Überprüfen bestehender Theorien – ausgelegt ist und ohne vorab formulierte Forschungshypothesen auskommt.

1.3 Vorteile und Nachteile des problemzentrierten Interviews

Wie bei vielen qualitativen Verfahren ist ein Nachteil des PZIs aus zeitökonomischer Sicht, dass die (zumeist vollständige) Transkription der Daten zeitaufwendig ist. Zudem ist diese Interviewtechnik – vielleicht mehr noch als andere – anspruchsvoll, da nicht nur zeitgleich aktiv zugehört, flexibel mit dem Leitfaden umgegangen und das Interviewgespräch geführt und sich auf die jeweilige Probandin oder den jeweiligen Probanden eingestellt werden muss, sondern man idealerweise auch erste Interpretationshypothesen und einen Abgleich mit evtl. bereits erhobenen Daten leisten muss. Nur so können neue, aber relevante Aspekte vertieft oder die richtigen Informationen erfragt werden. Schließlich stellt die Tatsache, dass der Forschende zwar eine theoretische Analyse vorgenommen hat, dieses Wissen bei den Interviews aber nicht anwenden – nicht einmal erkennen lassen – darf, eine weitere große Herausforderung dar.

Eng mit diesen Nachteilen verwoben sind die Vorteile der Methode, denn die ‚andere Seite der Medaille‘ ist eine Kombination der Vorteile narrativer und strukturierter Interviewmethoden. So haben die Probandinnen und Probanden ausdrücklich die Möglichkeit, neue Aspekte des Problems aufzubringen und auch (relevante) Nachbarthemenbereiche zu erörtern. Ein sehr offener Umgang mit dem Leitfaden kann zudem dabei helfen, Vertrauen auf- und Unsicherheiten abzubauen. Auf diese Weise können Erkenntnisse, die nicht durch die vorangegangene theoretische Analyse nahegelegt wurden, entdeckt werden. Eine Vertiefung solcher neuer Ansätze, kann dann durch die Prozesshaftigkeit der Methode in späteren Interviews weiter verfolgt werden. Auch der bewusste Umgang von theoretischem Vorwissen und das gegebenenfalls problemlose Einbetten anderer Methoden (s.o.) beugen Fallstricken vor. Zusammenfassend lässt sich also sagen, dass die Gleichzeitigkeit von Fokussiertheit auf ein bestimmtes Problem einerseits und Offenheit gegenüber den je individuellen Sichtweisen und Motiven andererseits zu den größten Schwächen wie auch Stärken der Methode zählt.

1.4 Gütekriterien für den Einsatz des problemzentrierten Interviews

Für das PZI als qualitatives Erhebungsverfahren gelten die üblichen Gütekriterien für qualitative Forschung. In Abgrenzung von den Gütekriterien quantitativer Verfahren betont Schmidt-Grunert (1999) zum Beispiel, dass nicht die Anzahl der Befragungen entscheidend sei, sondern eine authentische Darstellung individueller Sinnzusammenhänge. Neben der ‚Repräsentativität‘ sei auch die ‚Reliabilität‘ in der qualitativen Forschung nicht als grundsätzliches Gütekriterium zu verstehen. Wichtig sei es hingegen im Sinne der ‚Explizierung‘ Unklarheiten immer durch Nachfragen zu klären, sodass alle relevanten Aspekte ausreichend genau ausgeführt werden. ‚Valide‘ werden die Ergebnisse beispielsweise durch den Austausch in Forschergruppen zu den jeweiligen Interpretationen und/oder durch wiederholte Gespräche mit Probandinnen und Probanden. Wichtig sei zudem ‚Reflexivität‘, das permanente Reflektieren aller Daten, die aufeinander bezogen und durch neue Erkenntnisse ergänzt und verändert werden müssen. Eine gründliche und nachvollziehbare Dokumentation aller Arbeitsprozesse stelle zudem sicher, dass weder gefälscht noch fehlinterpretiert wurde. Dies ist nur durch die Dokumentation in allen Arbeitsschritten möglich, durch die weitestgehende Transparenz erreicht wird (Schmidt-Grunert 1999, 55f.).

Neben diesen allgemeinen Gütekriterien der qualitativen Forschung lassen sich dem PZI drei Grundpositionen zuordnen, die in allen Schritten des Forschungsprozesses berücksichtigt werden müssen: Problemzentrierung, Gegenstandsorientierung und Prozessorientierung.

Ausgangspunkt der Forschung ist ein vom Forschenden identifiziertes Problem, das zunächst theoretisch erkundet werden muss. Das Problem steckt dabei den thematischen Rahmen der Befragung ab. Es geht „[...] um individuelle und kollektive Handlungsstrukturen und Verarbeitungsmuster gesellschaftlicher Realität.“ (Witzel 1994, 67f.) Dieses zu bearbeitende Problem soll vor allem an der Sicht der Befragten orientiert erforscht werden (vgl. ebd., 116).

Auf der Basis dieser Überlegungen bekommt das Kriterium der Problemzentrierung eine doppelte Bedeutung: einmal bezieht es sich auf eine relevante gesellschaftliche Problemstellung und ihre theoretische Ausformulierung als elastisch und handhabendes Vorwissen des Forschers. Zum anderen zielt es auf Strategien, die in der Lage sind, die Explikationsmöglichkeiten der Befragten so zu optimieren, daß sie ihre Problemsicht auch gegen die Forscherinterpretation und in den Fragen implizit enthaltene Unterstellungen zur Geltung bringen können. (ebd., 69)

Mit der Gegenstandsorientierung ist eine Fokussierung auf die Sicht der Probandinnen und Probanden und auf das originäre Handlungsfeld angemahnt. Der Forschende muss unvoreingenommen an den Forschungsgegenstand herantreten, um die Forschungsfrage adäquat im Sinne des PZIs bearbeiten zu können. Ziel ist es, den Untersuchungsgegenstand authentisch zu erfassen (vgl. Schmidt-Grunert 1999, 53). Theoretische und methodische Überlegungen sollten diesem Zugang zur Empirie nicht im Wege stehen (vgl. Schmidt-Grunert 1999, 42).

Die Prozessorientierung bezieht sich schließlich auf den Anspruch, während der Forschung flexibel und im gesamten Forschungsprozess offen zu bleiben für die Überarbeitung oder Revidierung bereits ermittelter Ergebnisse, für Anpassungen der Forschungsinstrumente, der Wahl der Probandinnen und Probanden etc. Diese Position kann zudem auf die Interviews selbst bezogen werden, da auch das Gespräch einem Prozess gleicht, bei dem „im Idealfall durch Vertrauenszuwachs [...] bestimmte Themen wiederholt, korrigiert und detailliert expliziert werden.“ (Witzel 1994, 116)

1.5 Literaturempfehlungen

Witzel/Reiter (2012): The Problem-Centred Interview.

Das Buch schildert ausführlich sowohl theoretische Grundlagen und Entstehungskontexte als auch Hinweise, wie problemzentrierte Interviews vorbereitet und durchgeführt werden. Die Methode wird anschaulich und detailliert erläutert, ergänzt durch konkrete Beispiele.

Witzel (1994): Verfahren der qualitativen Sozialforschung.

Witzel erläutert in diesem Buch Probleme gängiger, qualitativer Methoden der Sozialforschung. Er entwickelt aus diesen kritischen Darstellungen die Methode des problemzentrierten Interviews und stellt damit sowohl die Methode selbst als auch ihre Vorteile im Vergleich mit anderen Interviewformen dar.

Lamnek (2010): Qualitative Sozialforschung.

Lamnek stellt im Überblick verschiedene Methoden dar. Leicht verständlich und strukturiert wird auch das problemzentrierte Interview vorgestellt. Das Werk eignet sich wunderbar als Einstieg, der bereits die konkrete Umsetzung der Methode im Blick hat.

2. Beispielprojekt: Wer (ver)führt zum Lesen?

In dem Promotionsprojekt „Wer (ver)führt zum Lesen?“ (Maus 2014) wurde das problemzentrierte Interview als eine von mehreren Methoden genutzt. Ziel des Projekts war es, den Einfluss von Geschlechtermustern auf die Lesemotivation von Jugendlichen zu analysieren, wobei sowohl der Einfluss der Geschlechtermuster der Jugendlichen selbst als auch die Geschlechtermuster der Figuren in ihren Lektüren untersucht wurden.

Diese Forschungsfrage ergab sich aus der empirisch gesicherten Tatsache, dass Jungen statistisch weniger und vor allem weniger gern und (damit) auch weniger gut lesen als Mädchen. Für diesen international nachweisbaren und in Deutschland besonders ausgeprägten Zusammenhang zwischen Lesemotivation und -kompetenz sowie dem Geschlecht (vgl. z.B. Stanat/Kunter 2001, 51) gibt es verschiedene Erklärungsansätze, die allerdings bislang kaum empirisch untersucht wurden. Eine gängige These lautet, Literatur werde von Beginn an vor allem von weiblichen Personen vermittelt, sodass sich Jungen, die sich in der Pubertät von allem Weiblichen distanzieren wollen, auch das Lesen ablehnen. Zudem wählten Mütter, Erzieherinnen und Lehrerinnen oft auch die Lektüren für ihre Söhne, Schüler etc. aus, sodass die männlichen Bedürfnisse an ihre Lektüre zu kurz kämen (vgl. Garbe 2007, 77). In didaktischen Fachbüchern und populärwissenschaftlichen Titeln zu dem Thema wird immer wieder die Meinung vertreten, dass männliche Schüler spätestens in der Pubertät gängige Schullektüren ablehnen müssten, da sie ihrem Männlichkeitskonzept widersprächen. Jungen bräuchten nicht literarische Figuren, die sozial und „verweiblicht“ seien und vor allem in den Augen von Pädagogen als Vorbild dienten (vgl. Kliewer 2004, 30). Ganz im Gegenteil wollten die Jungen echte Helden, die für eine stereotype Männlichkeit stehen (vgl. Weißenburger 2009, 187).

Die hier vorgestellte Studie zielte darauf ab, diese These zu hinterfragen und die Berücksichtigung auch der weiblichen Perspektive in die Diskussion einzubringen, da Unterricht in der Regel koedukativ geschieht und damit auch die Bedürfnisse der Mädchen etwa bei der Lektürewahl für den Literaturunterricht berücksichtigt werden sollten.

2.1 Die Wahl der Methode(n)

Bei der Planung des Dissertationsprojekts stellte sich früh heraus, dass ein quantitatives Vorgehen zur Überprüfung dieser dargestellten These zum Zusammenhang von Lesemotivation und Geschlecht auf zu vielen unbewiesenen Vorannahmen aufbauen würde, sodass nur ein qualitatives, theoriegenerierendes Herantreten an die Forschungsfrage wissenschaftlichen Standards und dem Gegenstand der Untersuchung gerecht werden konnte. Aus der Forschungsfrage ergaben sich als zentrale Arbeitsschritte die Ermittlung von relevanten, typischen Texteigenschaften populärer Jugendlektüre und die Ermittlung von Lesereigenschaften und deren Interaktion mit den untersuchten (und weiteren) Texten.

Für die Analyse der Texteigenschaften wurde die Methode der *Qualitativen Inhaltsanalyse* (vgl. den Beitrag von Heins in diesem Band) genutzt. Die Protagonisten vierer populärer Jugendbuchreihen wurden auf deduktiv festgelegte Kategorien stereotyper Männlichkeit und stereotyper Weiblichkeit hin untersucht. In den Interviews wurden anschließend die Geschlechterrolleneinstellung und die Geschlechterrollenidentität¹ der Probandinnen und Probanden ermittelt und ihre Einstellung zu den Protagonisten der Beispielliteratur und anderer Lektüren abgefragt. Allerdings konnten valide Aussagen auf die Forschungsfrage „Wer (ver)führt zum Lesen?“ nur dann erwartet werden, wenn die Jugendlichen auch die Möglichkeit bekamen, eigene Vorlieben und Motivationsgründe in Bezug auf ihr Leseverhalten und ihre Lektürepräferenzen zu nennen.

Für diese Befragung bot sich daher die Methode des PZIs an. Sie trägt dem Anspruch gleichzeitiger Strukturiertheit und Offenheit der Befragungen Rechnung, wie im ersten Teil dargestellt, und eignet sich daher in besonderem Maße für die Bearbeitung der Forschungsfrage. Da die Methode darauf abzielt, unvoreingenommen den individuellen (oder kollektiven) Umgang und die je eigene Deutung von gesellschaftlicher Realität zu ermitteln (vgl. Witzel 2000), ist ihr Einsatz immer dann besonders sinnvoll, wenn es eine gesellschaftliche, soziale Erwünschtheit für das Handeln oder Denken einzelner gibt, die die Ergebnisse der Datenerhebung zu verzerren droht. Gerade bei der Thematisierung der Kategorie *Geschlecht* in der hier vorgestellten Studie war dies der Fall, da Pubertierende bei der Positionierung in dieser Kategorie hohem sozialen Druck und dem Einfluss starker Normen ausgesetzt sind. Durch die Offenheit und die Möglichkeit narrativer Elemente erlaubte das PZI den Probandinnen und Probanden, relativ frei möglichst viele Einflüsse und Beweggründe zu schildern, die für sie wichtig waren – unabhängig davon, ob sie erwartet wurden. Das offene Design erlaubte zudem auch die Rekonstruktion dessen, was die Probandinnen und Probanden eher nicht preisgeben wollten beziehungsweise konnten. Durch vorsichtiges Nachfragen an diesen Stellen erlaubt das PZI weiteren Motiven und Einstellungen auf die Spur zu kommen. Mithilfe eines Leitfadens konnte sichergestellt werden, dass trotzdem alle wichtigen Fragen gestellt wurden, die die vorab gesichtete Theorie als bedeutsam nahelegte.

Ein weiterer, wichtiger Aspekt, der für das PZI als Interviewmethode sprach, war, dass gerade bei der Befragung von Jugendlichen eine wichtige Rolle spielt, dass sie als Experten anerkannt und nicht ‚abgefragt‘ werden. Eine vertrauensvolle At-

¹ Die Studie legte bewusst nicht das biologische, sondern die in diesem Kontext sehr viel aussagekräftigere Kategorie des sozialen Geschlechts zugrunde. Die untersuchten Dimensionen dieses sozialen Geschlechts sind Geschlechterrolleneinstellung und Geschlechterrollenidentität. Ersteres bezieht sich auf die persönliche Annahme, wie Männer bzw. Frauen sich üblicherweise verhalten bzw. verhalten sollten (vgl. Alfermann 1996, 47). Letzteres bezieht sich auf die Selbstwahrnehmung und -inszenierung in diesem Kontext, die ganz persönliche Kombination verschiedener Geschlechterrollen zu einer individuellen Identität. (vgl. Spence/Buckner 1996, 57).

mosphäre auf Augenhöhe kann besonders in diesem Kontext die Wahrscheinlichkeit für ein in weiten Teilen offenes Gespräch deutlich erhöhen. So war es möglich, die Angst, ähnlich wie in der Schule falsche Antworten geben zu können, genauso abzubauen wie anfängliche Schüchternheit. Andererseits wäre beispielsweise die Ungeleitetheit eines rein narrativen Interviews eine Überforderung für die Jugendlichen gewesen, wie sich spätestens im Testinterview herausstellte. Gerade zu Beginn des Interviews brauchte es ein ‚Gerüst‘, an dem die jungen Probandinnen und Probanden sich orientierten bis sie die Situation so gut einschätzen konnten, dass sie auch von sich aus ins Erzählen kamen.

2.2 Die Planung der Interviews

Bei der Erstellung des Leitfadens gab es bereits einige Fragen und Themen, die sich logisch durch die Fragestellung ergaben wie etwa „Liest du gern?“ oder „Was liest du gern?“. Andere Aspekte erwiesen sich nach Sichtung der theoretischen Fachliteratur als relevant wie zum Beispiel Fragen nach einem gelungenen Perspektivwechsel (vgl. Bartsch et al., 39) oder dem Einfluss der Herkunftsfamilie (vgl. Dijkstra 1994, 28).

Um den Druck der sozialen Erwünschtheit möglichst zu minimieren, wurden alle Fragestellungen, die den Themenbereich *Gender* eindeutig betrafen, relativ spät im Interview gestellt. Im Pretest wurde nämlich deutlich, dass die Jugendlichen nach der expliziten Erwähnung der Kategorie *Geschlecht* über den vollständigen Zeitraum des Interviews hinweg diese Kategorie ‚mitdachten‘ und dieser persönliche Filter die Antworten beeinflusste. Die Probandinnen und Probanden wussten im Vorfeld der Interviews daher zwar, zu welchem Zweck diese geführt wurden, zur Fragestellung der Arbeit war ihnen jedoch nur bekannt, dass die Studie sich für ihre Lesevorlieben interessierte. So wurde der Fokus auf das Thema *Figuren* ebenfalls nicht thematisiert, um möglichst authentische Antworten zu erhalten.

Aus diesen und ähnlichen Überlegungen und den im ersten Teil dieses Beitrags ausgeführten Hinweisen zur Durchführung des PZIs (s.o.) heraus ergaben sich die folgenden, groben Themenbereiche für den Interviewleitfaden:

1. Einführung in das Interview
2. Einstiegskommunikation
3. Arbeit mit Auszügen der analysierten Literatur als Beispieltexte
4. Erhebung der Rezeptionsmuster (Lesestrategien/Involvement)
5. Erhebung des Leseverhaltens (Lesestrategien/Lesemilieu)
6. Erhebung der Geschlechterrolleinstellung und der persönlichen Geschlechterrollenidentität
7. Gelegenheit für weitere Anmerkungen und/oder Fragen
8. Erhebung persönlicher Daten (Alter, besuchte Schule etc.)

Bei der Formulierung der konkreten Fragen galt es, nah an der Sprache der Jugendlichen zu bleiben. Vor allem der Gebrauch von Fremdwörtern schied damit aus. Im Pretest zeigte sich auch, dass der auffordernde Charakter der Fragen besonders wichtig war, da die Jugendlichen in der Regel zunächst zurückhaltend

waren und nur schwer ins Erzählen kamen. Mit einfachen, leicht fassbaren Frageformulierungen gelang es hingegen gut, gehaltvolle Daten im Sinne der Fragestellung zu gewinnen: „Kannst du dir Figur XY vorstellen? Was ist das für eine? Beschreib mal!“ oder „Wenn du in ein Buch hinein könntest, in welches würdest du dann gern schlüpfen und warum ausgerechnet in dieses Buch?“. Ebenfalls aus der Hoffnung heraus, relevante Aussagen zu provozieren und um Aussagen zu gleichen Texten zu erhalten, wurden den Probandinnen und Probanden zudem kurze Auszüge aus zwei der analysierten Beispieltexen vorgelegt, die sie in der Interviewsituation lesen und anschließend kommentieren sollten. Obwohl im Vorfeld viele konkrete Fragen erarbeitet wurden, war es sinnvoll, während der Interviews den Leitfaden nur in der oben aufgeführten, groben Gliederung vorliegen zu haben. Ergänzend wurden fünf Fragen ausgewählt, die so zentral für die Erhebung waren, dass sie wortwörtlich gestellt und daher ebenfalls auf die Leitfaden-Version für den unmittelbaren Gebrauch bei der Befragung gestellt wurden.

Eine Ausnahme bei diesem Vorgehen bildete die Erhebung des sozialen Geschlechts durch die Ermittlung der Geschlechterrollenidentität und Geschlechterrolleneinstellung, die strukturiert fast am Ende des Interviews vorgenommen wurde. Den Jugendlichen wurden beispielsweise mögliche Regeln des gesellschaftlichen und familiären Lebens vorgelesen, die die Verteilung von Geschlechterrollen betrafen. Diese sollten anschließend kommentiert und bewertet werden. Hier wich das Vorgehen zwar deutlich von der Methode des PZIs ab, aber als Ergänzung im Sinne der Mixed-Methods konnte ein solches Vorgehen begründet in das Forschungsdesign integriert werden.

Die Wahl der Probandinnen und Probanden erfolgte durch das *Theoretische Sampling* (s.o.), sodass die Probandenwahl sich nach prozesshaft ermittelten Kriterien richtete. Beispielsweise fiel bei den ersten Auswertungsschritten der bereits geführten Interviews auf, dass sich die einzige Viel-Leserin im Sample im Sinne der Forschungsfrage sehr von den anderen Probandinnen und Probanden unterschied. Somit musste mindestens ein weiterer – möglichst männlicher – Viel-Leser befragt werden, um festzustellen, ob diese gravierenden Unterschiede ein zweites Mal und auch geschlechtsunabhängig auftraten. Da die Präferenzen von Jungen und Mädchen untersucht werden sollten, ergab sich auch insgesamt die Notwendigkeit, Jugendliche beiderlei Geschlechts zu interviewen. Um einen möglichst hohen Grad an Vergleichbarkeit zu gewährleisten, wurden weitere Kriterien vorab festgelegt, die die Probandinnen und Probanden erfüllen mussten: So sollten sie die achte Klasse eines Gymnasiums besuchen und zwischen 13 und 15 Jahren alt sein. Eine theoretische Sättigung ergab sich nach sieben Interviews.

Die Rekrutierung von Probandinnen und Probanden geschah teilweise über Kontakte, die im Rahmen einer vorab durchgeführten, schriftlichen Befragung zur Ermittlung geeigneter Beispielliteratur in Schulklassen entstanden waren. Zum Teil konnten auch private Beziehungen genutzt werden, um Probandinnen und Probanden zu gewinnen, wobei dabei nur solche Jugendliche in Frage kamen, die bisher keinen oder kaum Kontakt zur Forscherin hatten und nur über Dritte vermittelt wurden. Die Probandinnen und Probanden und ihre Eltern mussten eine

Einverständniserklärung zur Datenerhebung und späteren (anonymisierten) Veröffentlichung der Daten unterschreiben. Den Ort des Interviews durften sich die Jugendlichen selbst aussuchen, um eine für die Schülerinnen und Schüler möglichst angenehme Atmosphäre entstehen zu lassen. So fanden Interviews teilweise in der Schule der Jugendlichen statt, aber auch an der Universität und bei Schülerinnen und Schülern zu Hause.

2.3 Durchführung der Interviews

Zunächst galt es, die Interviewsituation aufzulockern, sich betont aufgeschlossen und interessiert zu zeigen und das Vertrauen der Jugendlichen zu gewinnen. Als Einstiegskommunikation hat es sich bewährt, die Probandinnen und Probanden vorab zu bitten, ihr aktuelles Lieblingsbuch mitzubringen. Den Intervieweinstieg konnten sie damit leicht selbst bestreiten, indem sie dieses Buch vorstellten und erklärten, was sie daran mochten. Auf diese Weise konnten die Jugendlichen auf ‚sicherem Terrain‘ ins Interview starten.

Ein Problem stellte die Tatsache dar, dass einige Jugendliche annahmen, dass häufiges Lesen anspruchsvoller Literatur erwünscht wäre und deswegen triviale Lesevorlieben beziehungsweise Leseunlust eher verbargen. Einen positiven Effekt auf die Authentizität der Aussagen hätte es daher gehabt, sich als Liebhaber von (auch trivialer) Jugendliteratur zu offenbaren und die Wichtigkeit des Freizeitlesens zu relativieren. Andererseits mussten suggestive Fragen oder Bemerkungen ausgeklammert werden. Der Balanceakt zwischen neutraler Haltung und Signalisierung von bedingungsloser Anerkennung blieb daher eine große, stets zu reflektierende Herausforderung in jedem Interview.

Es stellte sich heraus, dass die meisten Fragen des Leitfadens bereits durch ein offenes Gespräch beantwortet wurden oder zumindest Ansatzpunkte für ein entsprechendes Nachfragen lieferten, sodass der Leitfaden vor allem genutzt wurde, um am Ende der Interviews noch einmal zu kontrollieren, ob unverzichtbare Themenaspekte ausgelassen wurden. Als besonders ergiebig erwiesen sich die Postskripte, da häufig relevante Aussagen nach Ausschalten des Tonbandgeräts (s.o.) gemacht wurden.

2.4 Auswertung der Daten

Da es sich um eine qualitative, theoriegenerierende Studie handelt, lag zur Auswertung die Nutzung der *Grounded Theory* nahe, die auch Witzel (1994) als besonders geeignet für die Verwendung bei der Auswertung von PZIs ansieht (s.o.). Die gewonnenen Audio-Daten wurden zunächst vollständig transkribiert und anschließend kodiert. Bereits bei den Auswertungen der ersten Interviews zeigte sich dabei, dass es, wie bereits erwähnt, große Unterschiede zwischen dem Einfluss der Geschlechtermuster der Figuren auf Viel- und auf Nicht- bzw. Wenig-Leser herrschten. Ein weiteres, im Forschungszusammenhang unerwartetes Ergebnis, war zudem, dass fast alle Jugendliche viele Worte darauf verwendeten,

immer wieder ihr eigenes Selbstkonzept darzustellen und zu bestätigen. Lars² zum Beispiel betonte wiederkehrend seine Technikaffinität und seine Ablehnung des Bücherlesens aus pragmatischen Gründen heraus. Das Lesen von fiktiven Geschichten sei keine bedeutsame Tätigkeit im Hinblick auf im späteren Beruf geforderte Kompetenzen etc. Gleichzeitig war er besonders beeindruckt von Artemis Fowl, dem Titelhelden einer Science Fiction-Reihe, der sich durch ein besonders ausgeprägtes technisches Geschick auszeichnet und trotz seiner Jugend auch geschäftlich ausgesprochen erfolgreich ist. Auch bei der Auswertung anderer Interviews drängte sich die These auf, dass das jeweils dargestellte Selbstkonzept eng mit den präferierten Figuren zusammenhing. Ein solcher Zusammenhang wurde zum Teil sogar explizit von den Befragten formuliert:

Diana: ‚[...] und ich konnte genau sehen: Petra is so ungefähr so wie ich. Und dann äh dacht ich immer, hab ich mich gefreut, wenn Petra dran gekommen is, um zu lesen und dann so zu sagen, ah das hätt ich auch gemacht. [...]‘ (Maus 2014, 92)

Dieses unerwartete Ergebnis wurde in die Planung der weiteren Interviews einbezogen und durch die Sichtung entsprechender Literatur theoretisch begleitet, sodass die Themenschwerpunkte der Datenerhebung im Sinne der Gegenstandsbezogenheit und der Prozesshaftigkeit um den Aspekt des *Selbstkonzeptes* erweitert wurden.

Ergebnis der Studie war eine neue Theorie zum Zusammenhang von Geschlechtermustern auf die Lesemotivation von männlichen und weiblichen Jugendlichen. Es lässt sich, grob zusammengefasst, festhalten, dass alle Jugendlichen Lesen als eine sozial sehr erwünschte Tätigkeit empfinden und besonders Mädchen diesem sozialen Druck auch nachgeben wollen, es aber nicht immer schaffen. Dass Jungen vor allem über männliche Figuren lesen, war bereits bekannt, aber die These von den erwünschten Helden ließ sich nicht bestätigen. Vielmehr scheint eine Ähnlichkeit des individuellen Selbstkonzeptes der Jugendlichen zu der Figur entscheidend zu sein, womit die Bedürfnisse – auch der interviewten Jungen – sich nur schwer verallgemeinern lassen. Dies gilt vor allem für Jugendliche, die nicht viel lesen und gern über für sie ‚normale‘ Figuren, die ihnen auch in ihrem Geschlechtermuster gleichen, lesen wollen, weil sie ihre Perspektive leichter übernehmen können. Dies hat häufig eine Ablehnung von fantastischen Büchern zur Folge. Viel-Leser hingegen genießen (auch) das Fremde, sind teilweise in der Lage sich in schwierige und ihnen gänzlich unähnliche Figuren hineinzudenken und Lesemotivationen aus diesem Prozess zu generieren. Durch die Identifikation mit der Tätigkeit Lesen selbst ergibt sich somit eine größere Unabhängigkeit der Lesemotivation von den in der Lektüre handelnden Figuren und ihren Geschlech-

² Die Probandinnen und Probanden durften sich selbst einen Vornamen wählen, unter dem sie bei der Darstellung der Forschungsergebnisse auftauchen wollten. Auf diese Weise können sie selbst sich wiedererkennen, bleiben insgesamt aber anonym. Es handelt sich also nicht um die realen Namen der Jugendlichen.

termustern. Aus der Studie konnten sich so einige Implikationen für die Förderung insbesondere von weniger lesemotivierten Schüler unter anderem durch Auswahl geeigneter Lektüren ableiten.

2.5 Fazit

Das PZI eignet sich besonders für Forschungsfragen, die den individuellen (oder kollektiven) gedanklichen oder tatsächlichen Umgang mit gesellschaftlicher Realität in den Blick nehmen. Die Kombination von Offenheit und Strukturiertheit kann im Umgang mit sozialen Erwünschtheiten, aber auch mit Schüchternheit und unsicheren Probandinnen und Probanden helfen, Hilfestellungen zu leisten ohne für die Erhebung gewinnbringende, narrative Prozesse zu verhindern. Durch die Integrationsmöglichkeit weiterer Methoden, lässt sich das Forschungsdesign sehr gut auf verschiedene Forschungsfragen zuschneiden und damit auch an Fragestellungen der Deutschdidaktik anpassen. Die Grundannahmen der Problemzentrierung, der Gegenstandsorientierung und der Prozessorientierung erleichtern diesen Prozess der Anpassung deutlich, da auch während der Phase der Datenerhebung, entsprechende Modifikationen geschehen können, sofern sie reflektiert, dokumentiert und begründet werden.

Literatur

- Alfermann, Dorothee: Geschlechterrollen und geschlechtstypisches Verhalten. Stuttgart: Kohlhammer 1996.
- Bartsch, Anne/Mangold, Roland/Viehoff, Reinhold/Vorderer, Peter: Emotional gratification during media use. An integrative approach. In: Communications. The European Journal of Communication Research (2006) 31, 3, 261-278.
- Corbin, Juliet/Strauss, Anselm L.: Grounded Theory. Grundlagen qualitativer Sozialforschung. Weinheim: Beltz 1996.
- Flick, Uwe: Qualitative Sozialforschung. Eine Einführung. Hamburg: Rowohlt 2007.
- Garbe, Christine: Lesen – Sozialisation – Geschlecht. Geschlechterdifferenzierende Leseforschung und -förderung. In: Bertschi-Kaufmann, Andrea (Hrsg.): Lesekompetenz – Leseleistung – Leseförderung. Grundlagen, Modelle und Materialien. Zug: Klett 2007, 66-82.
- Kliwer, Annette: Jungenbücher – nur für Mädchen? Jungen als Helden und Leser der aktuellen Adoleszenzliteratur. In: Dies./Schilcher, Anita (Hrsg.): Neue Leser braucht das Land! Zum geschlechterdifferenzierenden Unterricht mit Kinder- und Jugendliteratur. Baltmannsweiler: Schneider 2004, 23-34.
- Lamnek, Siegfried: Qualitative Sozialforschung. Basel: Beltz 2010.
- Maus, Eva: Wer (ver)führt zum Lesen? Der Einfluss von Geschlechtermustern auf die Lesemotivation von Jungen und Mädchen. Baltmannsweiler: Schneider 2014.
- Schmidt-Grunert, Marianne: Grundlagen. In: dies. (Hrsg.): Sozialarbeitsforschung. Problemzentrierte Interviews als qualitative Erhebungsmethode. Freiburg: Lambertus 1999, 11-68.

- Spence, Janet/Buckner, Camille: Masculinity and Femininity. Defining the Undefinable. In: Kalbfleisch, Pamela J./Cody, Michael J. (Hrsg.): Gender, power, and communication in human relationships, Hillsdale, NJ: Erlbaum 1995, 105-138.
- Stanat, Petra/Kunter, Mareike: Geschlechterunterschiede in Basiskompetenzen, in: Baumert, Jürgen/Neubrand, Michael (Hrsg.): PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. Opladen: Leske+Budrich 2001, 251-270.
- Weißburger, Christian: Helden lesen! Die Chancen des Heldenmotivs bei der Leseförderung von Jungen. Baltmannsweiler: Schneider 2009.
- Witzel, Andreas/Reiter, Herwig: The Problem-Centred Interview. Principles and Practice. London u.a.: Sage 2012.
- Witzel, Andreas: Das problemzentrierte Interview. In: FQS. Forum: Qualitative Social Research (2000), <http://www.qualitative-research.net/index.php/fqs/article/viewArticle/1132/2519> (letzter Zugriff: 01.08.2018).
- Witzel, Andreas: Verfahren der qualitativen Sozialforschung. Überblick und Alternativen. Frankfurt a.M.: Campus 1994.

Leitfadeninterviews

Ein Verfahren zur Erhebung subjektiver Sichtweisen

1. Vorstellung des Leitfadeninterviews

Zu den zentralen und wahrscheinlich auch meistgenutzten Erhebungsverfahren in der Forschungspraxis gehört das leitfadengestützte Interview, das in der Methodoliteratur auch als semi- bzw. halbstrukturiertes Interview bezeichnet wird. *DAS* Leitfadeninterview gibt es genau genommen aber nicht. An sich ist die Bezeichnung ein „Oberbegriff“ (Kruse 2015, 203) für eine bestimmte *Vorgehensweise der Interviewführung*, unter der in der empirischen Sozialforschung verschiedene Interviewarten – so z.B. das fokussierte Interview (Merton/Kendall 1976) und das problemzentrierte Interview (s. Maus in diesem Band) – zusammengeführt werden. Vor diesem Hintergrund sollen nachfolgend spezifische Prinzipien und Forschungsschritte diskutiert werden, die grundlegend zu beachten sind, wenn ein Gesprächsleitfaden in der Interviewkommunikation eingesetzt wird. In einem zweiten Schritt werden Ablauf und Einsatz von leitfadengestützten Interviews anhand von zwei deutschdidaktischen Forschungsprojekten konkretisiert.¹

1.1 Funktionsweise des Leitfadeninterviews

Wie bereits ausgeführt, umfasst der Begriff des Leitfadeninterviews eine bestimmte *Art und Weise der Gesprächsführung*: Die Interviewkommunikation wird mit Hilfe eines vorbereiteten Leitfadens strukturiert, der durch offene, erzählgenerierende Fragen gegliedert ist. Auf dieser Grundlage folgt der Interviewverlauf einem bestimmten „Themenweg“ (Kruse 2015, 209), innerhalb dessen man etwas von den Befragten in Erfahrung bringen möchte. Ganz allgemein können Leitfadeninterviews unterschiedliche *Strukturierungsniveaus* aufweisen, die sich auf einem Kontinuum von stark bis weniger strukturierend bewegen – und entsprechend ihre Vor- und Nachteile haben (s. Abschnitt 2). Der Interviewleitfaden ist aber *kein Selbstzweck* für die Kommunikation: Er ist vielmehr ein Gerüst zur Strukturierung des Interviewverlaufs, mit dem man fokussiert die Sichtweisen der

¹ Aspekte, die grundlegend für die Planung und (Durch-)Führung von Interviews zu beachten sind, wurden bereits an anderer Stelle erörtert (s. Schmidt in diesem Band) und werden in diesem Beitrag nur vereinzelt aufgegriffen.

Interviewpartner erhebt (vgl. Niebert/Gropengießer 2014, 127). Leitfadeninterviews stehen insofern vor der Herausforderung, zwischen *Offenheit* und *Strukturierung* in der Gesprächssituation zu vermitteln: Ein Leitfaden soll das Interview in gewissem Maße lenken, darf aber den Gesprächsfluss nicht einengen.

Auf die Gestaltung eines Interviewleitfadens ist große Sorgfalt zu verwenden, da davon auch die Qualität der generierten Daten abhängt – oftmals wird diese Komplexität in der Forschungspraxis unterschätzt (vgl. Kruse 2015, 226). Die Leitfadenkonstruktion setzt zunächst die vertiefende Auseinandersetzung mit dem theoretischen Hintergrund zum Gegenstand voraus: Was möchte ich vom Erzählenden erfahren? Welche Themenbereiche sollen angesprochen werden? Welche Fragen müssen wann und wie gestellt werden? (vgl. Helfferich 2011, 141). Dabei ist zu beachten, dass *Forschungsfragen keine Interviewfragen* sind (s. Schmidt in diesem Band).

Für die konkrete Erstellung eines Leitfadens wird von Helfferich (2011, 178ff.) das sogenannte *SPSS-Prinzip* (Sammeln, Prüfen, Sortieren, Subsumieren)² vorgeschlagen:

SAMMELN	Zunächst werden in einem Brainstorming <i>sämtliche Fragen</i> , die von Interesse für die Untersuchung sein könnten, in einer Liste zusammengetragen – Bedenken bzgl. der Eignung der Fragen, der konkreten Formulierung usw. werden in diesem Schritt hinten angestellt. Eine Empfehlung ist, in diese Entwicklungsphase auch Fachkolleginnen und -kollegen einzubeziehen. Im Sinne des „Fremdverstehens“ (ebd., 84) eröffnet dieses Vorgehen die Möglichkeit, Verstehensprozesse in der Gruppe zu reflektieren und mitunter neue Verstehenskonstruktionen zu entwickeln (vgl. Daase/Hinrichs/Settinieri 2014, 111). Ferner können Fragen oder Leitfäden aus thematisch verwandten Studien eine weitere Orientierung für mögliche Interviewfragen bieten. Im Anschluss an Kruse (2015, 230f.) sollte diese Phase im Übrigen dazu genutzt werden, nochmals die Fragestellung(en) der eigenen Untersuchung zu reflektieren und diese ggf. weiterzuentwickeln.
PRÜFEN	Die zusammengestellte Frageliste ist <i>kein Leitfaden</i> . Die Liste wird deshalb in einem nächsten Schritt reduziert und geordnet. Dazu wird kritisch reflektiert, inwiefern die einzelnen Fragen lediglich vorhandenes Vorwissen erheben oder nur Faktenfragen darstellen – diese Fragen werden gestrichen. Ebenso werden zu detaillierte Fragen (z.B. „Wie viele Stunden in der Woche verwenden Sie in der Woche genau für die Leseförderung in Ihrer Klasse?“) aus der Liste entfernt oder ggf. durch offenere Formulierungen („Welchen Stellenwert hat das Thema Leseförderung in Ihrem Unterricht?“) ersetzt. Eine Orientierung für diesen Schritt bilden die Prüfaspekte bei Kruse (2015, 228ff.).
SORTIEREN	Anschließend werden die verbleibenden Fragen nach Bereichen, d.h. nach <i>inhaltlicher</i> oder – bei Interesse an einer zeitlichen Dimension – nach <i>chronologischer Zusammengehörigkeit</i> , geordnet. Insgesamt sollen in dieser Strukturierungsphase zwischen einem bis vier Fragenbündel entstehen. Es kann sein, dass einige Fragen als Einzelfragen stehen bleiben – sie bekommen einen gesonderten Platz im Leitfaden.

² Eine Weiterentwicklung des Verfahrens, das S²PS²-Verfahren, findet sich bei Kruse (2015, 230-236).

SUBSUMIEREN	In einem letzten Schritt wird für die einzelnen Themenkomplexe eine weitgehend offene (Erzähl-)Aufforderung gesucht, unter die andere Einzelaspekte untergeordnet werden können (z.B. „Beschreibe mir doch bitte, was deiner Meinung nach eine gute Möglichkeit ist, das Schreiben im Deutschunterricht zu fördern.“). ³ Dieser offene Einstieg in die jeweiligen Themenfelder soll den Erforschten ermöglichen, dass sie die Bereiche <i>aus ihrer Perspektive</i> und <i>mit ihren Schwerpunktsetzungen</i> erläutern können. Zudem werden konkrete Nachfragen formuliert, mit denen ggf. diejenigen Aspekte erfragt werden können, die nicht bereits auf Grundlage der jeweiligen Erzählaufforderung beantwortet wurden (s. Abb. 1).
-------------	--

Das SPSS-Prinzip sollte man allerdings nicht als starren Ablauf ansehen – die einzelnen Phasen sind vielmehr miteinander verwoben. Grundsätzlich folgt der Leitfaden einem systematischen Aufbau ‚*vom Allgemeinen zum Spezifischen*‘; auf diese Weise werden die Interviewprinzipien der Offenheit und Strukturierung (s.o.) ausbalanciert.

Darüber hinaus gibt es noch die Möglichkeit, den Leitfaden um *Gesprächsstimuli* zu ergänzen, die im Sinne des Forschungsinteresses *gezielt* zu ausführlichen bzw. vertiefenden Schilderungen anregen (vgl. Helfferich 2014, 565). Im Bereich der Lehr- und Lernforschung bieten sich in diesem Zusammenhang u.a. folgende Gesprächsimpulse an (modifiziert nach Niebert/Gropengießer 2014, 128ff.):

- Aufgabenstellungen mit und ohne Material⁴
- Interpretation von Texten, Tabellen, Karikaturen usw.
- Vignetten (s. Benz in diesem Band)
- Bildimpulse (s. Binanzer in diesem Band)
- Filmausschnitte

1.2 Auf die Leitfragen kommt es an – Gütekriterien für den Einsatz des Verfahrens

1.2.1 Allgemeine Prinzipien von leitfadengestützten Interviews

Der eingesetzte Interviewleitfaden bestimmt maßgeblich die Qualität der möglichen Forschungsergebnisse. Vor diesem Hintergrund gibt es verschiedene Anforderungen, die für die Konstruktion beachtet werden sollten, wenn man sich für die Durchführung von Leitfadeninterviews entschieden hat.

Gestaltung: Ein Interviewleitfaden muss *übersichtlich* gestaltet sein, damit er im Interviewverlauf flexibel und gut handhabbar ist – schließlich sollen die Fragen im Interview nicht abgelesen werden. Es bietet sich daher an, den Leitfaden tabellarisch aufzubauen, um in der Interviewkommunikation ggf. schnell einen Überblick gewinnen zu können. Ob wiederum die jeweiligen Fragen im Leitfaden ausformuliert werden oder der Leitfaden lediglich Stichworte enthält, kann man

³ Vgl. auch die Beispiele in Abschnitt 2.

⁴ Etwa „Schau dir das folgende Aufgabenset zum Thema *Inhaltsangabe* an. Würdest du es in deinem Unterricht einsetzen?“, oder „Bitte zeichne doch einmal, wie du dir die Szene vorstellst“.

nach eigener Vorliebe entscheiden (vgl. Daase/Hinrichs/Settinieri 2014, 111). Eine Empfehlung aus eigenen Untersuchungen ist, den Leitfaden noch um die Spalte „Notizen“ zu erweitern (s. Abb. 1): Dies bietet dem Interviewenden die Möglichkeit, Stichworte u.Ä. im Gespräch aufschreiben – z.B. um sich Anhaltspunkte für Nachfragen im Anschluss an einzelne Äußerungen zu notieren oder einzelne Begriffe der Beforschten noch genauer zu erfragen. Allerdings sollte man darauf achten, dass für die Interviewten dabei nicht der Charakter eines Verhörprotokolls entsteht.

Leitfrage/Stimulus/ Erzählaufforderung	Inhaltliche Aspekte Stichworte – nur erfragen, wenn nicht von allein thematisiert	(Nach-)Fragen mit obligatorischer Formulierung	Notizen
Erzählaufforderung			
Erzählaufforderung			
Erzählaufforderung			
Bilanzierungsfragen Einstellungsfragen			
Abschlussfrage: Fehlt etwas?			
Nach dem Interview: Ergänzender Fragebogen für Faktenfragen			

Abb. 1: Aufbauprinzip eines Leitfadens (modifiziert nach Helfferich 2014, 568)

Struktur: Der Leitfaden muss eine in sich *logische Struktur* haben. Dabei gilt: So offen wie möglich, so strukturiert wie nötig (vgl. Helfferich 2011, 181). Da die Schwerpunktsetzungen und Denkstrukturen von den Erforschten zu einem bestimmten Thema anders strukturiert sein können als die eigenen Sichtweisen, ist der Leitfaden bzw. die dortige Fragefolge aber nur als eine Gedankenstütze für das Interview anzusehen (vgl. Abschnitt 1.1).

Offenheit: Der Leitfaden ist eine flexible Checkliste und *kein Drehbuch*. Abgesehen von der Einstiegsfrage sollte sich die Fragefolge aus der Gesprächssituation ergeben. Eine zu starke Orientierung am Leitfaden birgt die Gefahr einer „Leitfadenbürokratie“ (Hopf 1978, 101ff.) – Fragen könnten im Gesprächsverlauf einfach nur ‚abgearbeitet‘ (s. Abb. 2) oder Ausführungen der Befragten im falschen Moment unterbrochen werden. Mit einem solchen Vorgehen werden nur oberflächliche Äußerungen erhoben, die im weiteren Forschungsprozess keine tiefergehenden Erkenntnisse über die subjektiven Sichtweisen der Personen ermöglichen.

I: kennst du die adressaten?

B: ja!

I: weisst du nach welchen kriterien sie die texte zur kenntnis nehmen/nutzen?

B: ich denke ja

I: und was ihnen wichtig ist?

B: ja!

Abb. 2: „Leitfadenbürokratie“ (Beispiel entnommen aus Lehnen/Schindler 2010, 249)

Sprachebene: Fragen und Impulse in einem leitfadengestützten Interview müssen *adressatenorientiert* sein, da sich Forschersprache und die Sprache von Lernenden oder Lehrenden voneinander unterscheiden können. Ein Gütekriterium für Interviewleitfäden ist daher, dass die Leitfragen *widerspruchsfrei, anschaulich* und *explizit* formuliert sind.

Qualität vor Quantität: Ein Interviewleitfaden darf *nicht überladen* sein – Priorität haben schließlich die im Gespräch produzierten Ausführungen der Erforschten. Zu viele angesprochene Einzelaspekte im Interviewverlauf können zu oberflächlichen Aussage der Befragten führen und/oder die interviewende Person zur Leitfadenbürokratie verleiten (s.o.).

1.2.2 Fehler erlaubt: Die Pilotierung des Interviewleitfadens

Einen Interviewleitfaden zu erstellen ist eine besondere Herausforderung; das sollte bereits deutlich geworden sein. In der Regel erfolgen daher mehrere Entwürfe, bevor ein Leitfaden letztlich zur Datenerhebung eingesetzt wird (vgl. auch Niebert/Gropengießer 2014, 130):

Erste Fassung und Vorstudie: Nach der Entwicklung eines ersten Entwurfs sollte der Leitfaden in einer Vorstudie (Pilotierung) auf seine Validität und Praktikabilität getestet werden. Dies stellt ein wichtiges *Gütekriterium* für die Qualität der generierten Daten in Leitfadeninterviews dar. Neben einer ersten gezielten Übung der Interviewsituation, in der die eigene Interviewerrolle reflektiert werden kann, können durch die Vorerhebung *inhaltliche und sprachliche Probleme* bzgl. der erstellten Leitfragen aufgedeckt werden: Welche (Fach-)Begriffe sind für Schülerinnen und Schüler oder Lehrkräfte zu abstrakt gehalten oder für sie zu vage formuliert? Welche Fragen sind zu komplex oder zu allgemein gehalten? Wie passen die angedachten Themenblöcke im Leitfaden zueinander? Dies sind nur einige Fragen, zu denen Probeinterviews wichtige Antworten liefern. Darüber hinaus können durch die Analyse der Vorstudien Daten noch neue Erkenntnisse zum Untersuchungsgegenstand abgeleitet werden, die in der Hauptstudie Berücksichtigung finden.

Zweite Fassung auf Grundlage erster Forschungserkenntnisse: Im Anschluss an die durchgeführten Probeinterviews erfolgt eine Re-Analyse der Leitfragen, so dass der Leitfaden *begründet* weiterentwickelt werden kann. Dieser Schritt muss zur Nachvollziehbarkeit und Transparenz des Forschungsprozesses dokumentiert werden.

1.3 Generierte Daten und die weitere Analyse

Leitfadengestützte Interviews werden mit dem Ziel durchgeführt, strukturiert die subjektiven Sichtweisen von Personen zu bestimmten Themen zu erheben. Im Interview wird immer sprachliches Datenmaterial generiert, das in Form von Audio- oder Videodaten digital aufgezeichnet wird.⁵ Die Aufbereitung und Analyse der verbalen Daten ist ein *methodisch kontrolliertes Fremdverstehen*, mit dem die Äußerungen Anderer erfasst werden und man als forschende Person den Äußerungen der Interviewten Sinn verleiht, d.h. sie interpretiert. Mit anderen Worten: Von der erfassten Sprache der Befragten wird auf ihr Denken geschlossen.

Zu diesem Zweck erfolgt zunächst die Anfertigung eines Transkripts, das den *ersten interpretativen Zugang* zum Datenmaterial darstellt und Gegenstand der ‚eigentlichen‘ Analyse ist (s. Schmidt in diesem Band). Soweit man von einer gängigen Analysepraxis sprechen kann, werden leitfadengestützte Interviews überwiegend inhaltsanalytisch (s. Heins in diesem Band) ausgewertet, da mit diesem Verfahren eine regelgeleitete Auswertung von Interviewdaten möglich ist. Für deutschdidaktische Forschungskontexte ist in diesem Zusammenhang der Ansatz des Biologiedidaktikers Harald Gropengießer (2005) interessant, der das Verfahren der qualitativen Inhaltsanalyse für fachdidaktische Zwecke adaptiert hat. Als weitere Analyseverfahren für leitfadengestützte Interviews bieten sich das Kodierparadigma der Grounded Theory (s. Kubik in diesem Band) an; weitaus seltener werden sequenzanalytisch-sinnverstehende Verfahren, wie die dokumentarische Methode (s. Schieferdecker in diesem Band), für die Auswertung herangezogen. Insgesamt ist bei der Entscheidung für ein spezifisches Analyseverfahren zentral, *welches Erkenntnisinteresse* im Rahmen der eigenen Untersuchung besteht.⁶

1.4 Stärken und Schwächen leitfadengestützter Interviews

Leitfadeninterviews sind geeignet, wenn individuelle Wissensbestände zu bestimmten Themen möglichst *offen* erhoben, aber gleichzeitig *relevante Themen und Fragerichtungen* für das Forschungsinteresse im Interview zur Sprache kommen sollen. Ein maßgeblicher Vorteil dieser Interviewform ist darüber hinaus, dass der Leitfaden im Erhebungsprozess absichert, dass zentrale Aspekte für das Forschungsinteresse in *allen Interviews* angesprochen werden (und trotzdem spezifisch auf die jeweilige Situation eingegangen werden kann). Insofern sind die einzelnen Interviews gut miteinander vergleichbar, da in den jeweiligen Erhebungen gleiche bzw. ähnliche Fragen die Basis der Kommunikation bilden (vgl. Helfrich 2014, 565).

⁵ Abhängig von den erstellten Leitfragen und verwendeten Impulsen können auch beschriftete Zeichnungen u.Ä. zum Datenmaterial gehören (vgl. Abschnitt 1.1), die entsprechend mit in die Analyse einbezogen werden.

⁶ Wichtig ist weiterhin, dass die genannten Aufbereitungs- und Analyseschritte nachvollziehbar und ausführlich dokumentiert werden, damit die Anwendung im jeweiligen Forschungskontext deutlich wird.

Der Einsatz des Leitfadens setzt weiterhin voraus, dass die interviewende Person die Fragestellungen, den theoretischen Ansatz und den Forschungsstand zum Projekt gut kennt, um *flexibel* in der Gesprächssituation zu agieren – die Interviews können insofern nur von Personen durchgeführt werden, die mit dem Projekt vertraut sind (vgl. Niebert/Gropengießer 2014, 126). Nur dann ist man auch in der Lage einzuschätzen, ob eine Interviewfrage bereits in einem anderen Zusammenhang diskutiert wurde (und somit ausgelassen werden kann), wann man vom Leitfaden abweichen sollte oder an welchen Punkten man im Gespräch vertiefende Nachfragen zu Äußerungen der Erforschten stellt.

Ganz allgemein stellen die gewählten Leitfragen in gewisser Form eine Erkenntnisgrenze für die Untersuchung dar, da sich die eingebrachten Fragen für das Interview aus dem Vorverständnis des Gegenstands von Seiten des Forschenden ergeben. Insofern ‚beschränken‘ die ausgewählten Leitfragen die Gestaltungsspielräume für die Schwerpunktsetzungen der Befragten, wodurch bestimmte Bereiche des Untersuchungsgegenstands ausgeblendet werden (vgl. Friebertshäuser/Langer 2010, 439). Steht die weitestgehend weitläufige Ausgestaltung der Kommunikation und der einzelne Fall im Vordergrund des Forschungsinteresses, so ist beispielsweise das narrative Interview (s. Lindow in diesem Band) dem Leitfadeninterview vorzuziehen.

1.5 Literaturempfehlungen⁷

Helfferrich (2010): Leitfaden- und Experteninterviews.

Der Beitrag bietet einen Überblick über allgemeine Prinzipien und forschungspraktische Schritte zum Leitfadeninterview.

Kruse (2015): Qualitative Interviewforschung.

Zur Entwicklung von Interviewleitfäden bietet das Lehrbuch eine praxisnahe Orientierung. Insbesondere die Themen *Fragestil* und *Gesprächsimpulse* werden anhand von etlichen Beispielen umfassend diskutiert sowie konkrete Anregungen für die Planung und Durchführung leitfadengestützter Interviews präsentiert.

Niebert/Gropengießer (2014): Leitfadengestützte Interviews.

Niebert und Gropengießer erörtern aus naturwissenschaftsdidaktischer Perspektive das Potenzial von leitfadengestützten Interviews für die Lehr- und Lernforschung. Gerade da einzelne Forschungsschritte anhand von eigenen (biologiedidaktischen) Forschungsarbeiten konkretisiert werden, ist der Lehrbuchartikel ebenso für Deutschdidaktiker sehr gut geeignet, um sich in das Thema einzuarbeiten.

⁷ Neben den hier diskutierten Literaturempfehlungen, die sich spezifisch mit leitfadengestützten Interviews auseinandersetzen, empfehle ich, sich vertiefend mit den Lehrbüchern und Beiträgen zu beschäftigen, die bereits in dem Überblicksbeitrag zum Thema Interviews (s. Schmidt in diesem Band) diskutiert wurden.

2. Von der Forschungsfrage zum Interviewleitfaden – Zwei deutschdidaktische Forschungsprojekte

Wie sieht nun ausdrücklich fachdidaktische Forschung aus, in der konkret leitfadengestützte Interviews angewendet werden? Nachfolgend soll das Potenzial der Erhebungsmethode anhand von zwei deutschdidaktischen Projekten veranschaulicht werden, in denen recht unterschiedliche Schwerpunkte im Fokus standen. Die gewählten Beispiele verdeutlichen darüber hinaus, wie das jeweilige Strukturierungsniveau des Gesprächsleitfadens vom Erkenntnisinteresse der Untersuchung abhängt.

2.1 Lesediagnostisches Wissen von Deutschlehrkräften

2.1.1 Ziel und Fragestellung der Studie

In einem Forschungsprojekt zur Entwicklung eines *praxisnahen* lesediagnostischen Instruments für die Sekundarstufe I bildete die Erhebung der subjektiven Sichtweisen von Lehrenden des Faches Deutsch – konkret ihrer Vorstellungen und Orientierungen⁸ zum Thema Diagnose von Lesekompetenz – einen zentralen Untersuchungsschritt (vgl. Schmidt 2013; 2015). Die weiteren Ausführungen konzentrieren sich auf diesen Erhebungsschritt der Untersuchung. Ziel war es zu erfassen, über welche handlungsrelevanten lesediagnostischen Wissensbestände Deutschlehrkräfte verfügen und wie Lehrkräfte ein auf fachdidaktischen Erkenntnissen entwickeltes Diagnoseverfahren reflektieren (vgl. vertiefend ebd., 262).

2.1.2 Studienanlage und Vorgehen

Zentraler Ansatzpunkt für die methodische Anlage der Untersuchung war, dass die beteiligten Deutschlehrkräfte im Rahmen der Erhebung *ihren Sinn* zum gewählten Untersuchungsgegenstand *entfalten*, sie also – in der ihnen eigenen Sprache und Logik – Zusammenhänge und Bedeutungen zum Gegenstand verbalisieren können. Da im Untersuchungszusammenhang die konkrete Perspektive der Erforschten auf die Diagnose von Lesekompetenz bzw. das entwickelte Diagnoseverfahren interessierte, wurde ein leitfadengestütztes Interview als zielführend angesehen: So konnten wesentliche Themen für das Erkenntnisinteresse der Studie angesprochen und zugleich den Befragten die Möglichkeit gegeben werden, ihre Sichtweisen zum Gegenstand zu entfalten (vgl. Abschnitt 1.1).⁹

Zur Ermittlung der Sichtweisen der Deutschlehrkräfte wurden zu zwei Messzeitpunkten leitfadengestützte Interviews durchgeführt. Während in der ersten Erhebung die handlungsrelevanten Wissensbestände der Deutschlehrkräfte zur Diagnose von Lesekompetenz in der Sekundarstufe I im Mittelpunkt standen, wurde

⁸ Zur genauen Begriffsbestimmung vgl. Schmidt (2015, 91f.).

⁹ In Abschnitt 1.1 wurde bereits erörtert, dass der Begriff des Leitfadeninterviews lediglich einen Oberbegriff darstellt. Im Rahmen der vorliegenden Untersuchung habe ich konkret problemzentrierte Interviews (s. Maus in diesem Band) durchgeführt.

in der zweiten Erhebung die Reflexion des entwickelten Beobachtungsinstruments fokussiert (vgl. vertiefend zum Forschungsdesign ebd., 262f.). Anliegen war es, die Erkenntnisse beider Interviewerhebungen für die (Weiter-)Entwicklung des lesediagnostischen Beobachtungsverfahrens fruchtbar zu machen (vgl. ebd., 261f.). Um konkret zu zeigen, wie das Forschungsinteresse der Untersuchung in einen Leitfaden überführt wurde, ist nachfolgend ein Auszug aus dem Leitfaden der ersten Interviewerhebung abgebildet:

Leitfrage/Stimulus/Erzählaufforderung	(Nach-)Fragen mit obligatorischer Formulierung
VIGNETTE	
ZIELE LESEUNTERRICHT Inwiefern ist Ihres Erachtens die Förderung von Lesen wichtig? Was gehört für Sie persönlich zur Lesefähigkeit der Schülerinnen und Schüler? Haben Sie Beispiele für Schülerinnen und Schüler, die Sie als gute bzw. schwache Leserinnen und Leser bezeichnen würden?	Können Sie mir beschreiben warum? Welche Teilaspekte? Welche Bereiche muss man fördern? Können Sie die Beispiele genauer erläutern?
(...)	
Was wünschen Sie sich bzgl. des Themas für die Zukunft?	In der Schule? In Fortbildungen?
Jetzt haben wir einiges besprochen. Ist Ihnen während des Interviews noch etwas durch den Kopf gegangen, das bisher noch nicht zur Sprache gekommen ist, was Ihnen aber wichtig ist?	

Bsp. 1: Leitfadenauszug zur Erhebung des lesediagnostischen Wissens von Lehrkräften (Schmidt 2013; 2015)

Zur Erläuterung: Als Gesprächseinstieg wurde in der ersten Interviewerhebung eine Vignette (s. Benz in diesem Band) gewählt, mit der die Lehrkräfte aufgefordert wurden, einem Studierenden zu erläutern, inwiefern diagnostische Aspekte im Bereich Lesekompetenz bei der Planung einer Deutscheinheit berücksichtigt werden sollten (vgl. Schmidt 2015, 98). Dieser Impuls zielte darauf zu erfahren, welches handlungsrelevante Wissen die befragten Lehrkräfte selbst zum Thema ‚Diagnose von Lesekompetenz‘ argumentativ entfalten. Nachfolgend war unter anderem von besonderem Interesse, welches Lesekonzept und welche Diagnosekriterien die Deutschlehrkräfte für ihren Unterricht anlegen (vgl. Themenblock „Ziele Leseunterricht“). Der dargestellte Auszug verdeutlicht exemplarisch, wie man sich im Rahmen eines thematischen Blocks von offenen zu spezifischen Fragen bewegt (vgl. Abschnitt 1.1).¹⁰ Durch die Nutzung des Leitfadens in allen Interviews nahmen die einzelnen Gespräche einen ähnlichen Themenverlauf und waren somit für die Analyse gut miteinander vergleichbar (vgl. Abschnitt 1.4).

¹⁰ Allerdings bleibt festzuhalten, dass im Interview die Erforschten und ihr Ordnungsmuster im Vordergrund stehen (vgl. Abschnitt 1.2.1).

2.1.3 Analyse und Ergebnisse

Insgesamt sind in die Auswertung die Interviewdaten von zehn Hauptschullehrerinnen und Hauptschullehrern eingeflossen, die zum Erhebungszeitpunkt in mindestens einer Klasse das Fach Deutsch unterrichteten (vgl. Schmidt 2013, 262). Das Ziel der Auswertung der Interviews war eine möglichst umfassende Rekonstruktion der Schwerpunktsetzungen der Deutschlehrkräfte zum Untersuchungsgegenstand, bei der sowohl die individuellen Wissensbestände als auch die kollektiven Orientierungen¹¹ relevant waren, um diese für die Entwicklung eines Diagnoseinstruments produktiv zu machen. Vor diesem Hintergrund wurde das Datenmaterial mithilfe einer modifizierten Form der dokumentarischen Methode (s. Schieferdecker in diesem Band) ausgewertet (vgl. ebd., 263).

Mit Blick auf das Erkenntnisinteresse der Untersuchung ist es von einem nicht unerheblichen Interesse, welchen Lesebegriff die befragten Lehrkräfte für ihren Unterricht heranziehen und auf welche fachdidaktischen Indikatoren sie zur Diagnose von Lesekompetenz bei ihren Schülerinnen und Schülern zurückgreifen (vgl. Abschnitt 2.1.2). Ein Ergebnis der Untersuchung ist, dass das „flüssige Lesen“ und das Textverstehen – konkret: die Inhaltssicherung – in den Vorstellungen der Deutschlehrkräfte eine zentrale Rolle einnehmen und von ihnen als bedeutsam für Diagnose von Lesekompetenz bewertet werden.¹² Hingegen spielen diejenigen Bereiche von Lesekompetenz, die nur indirekt oder durch gezieltes Beobachten erfasst werden können, wie beispielsweise das Reflektieren über Texte, in den Aussagen der befragten Lehrkräfte eine eher untergeordnete Rolle (vgl. Schmidt 2015, 99f., 104). Vertiefende Kenntnisse aktueller lesedidaktischer Positionen (vgl. exemplarisch Rosebrock/Nix 2014) zeigen sich in den Äußerungen der Lehrkräfte indes nicht als Quelle des handlungsrelevanten Lehrerwissens.

Dieser exemplarische Befund verdeutlicht, wie anhand von leitfadengestützten Interviews Erkenntnisse über Konstrukte oder Begriffe aus der Logik der Beforschten erhoben werden können. Hier zeigt sich aber nicht nur ein wichtiger Ansatzpunkt für die Weiterentwicklung des lesediagnostischen Beobachtungsverfahrens. Zugleich können derartige Erkenntnisse wichtige Anhaltspunkte dafür bieten, welche Perspektiven von Deutschlehrkräften im Rahmen von Maßnahmen zur Lehrprofessionalisierung berücksichtigt werden müssen (vgl. Schmidt 2015, 105).

¹¹ Zur Erhebung kollektiver Orientierungen im Rahmen von Einzelinterviews vgl. vertiefend Nohl (2009).

¹² Vgl. ausführlicher zu ersten Erkenntnissen der Untersuchung Schmidt (2013, 267f.) und Schmidt (2015, 99ff.).

2.2 Spielfilmverstehen von Schülerinnen und Schülern

2.2.1 Ziel und Fragestellung der Studie

In den letzten Jahren ist wiederholt von fachdidaktischer als auch bildungspolitischer Seite betont worden, dass Lernende in der Lage sein sollen, Filme zu analysieren und kritisch zu reflektieren (vgl. etwa Abraham 2009; BS 2012, 24). Ausgangspunkt der folgenden Untersuchung (Winkler/Schmidt 2012; Schmidt/Winkler 2015) war, dass es hingegen nur wenig empirisch gesichertes Wissen darüber gibt, *welches filmbezogene Vorwissen* Schülerinnen und Schüler bereits in den Deutschunterricht mitbringen (vgl. Abraham 2009, 43) – dieses Desiderat ist von besonderer Brisanz, da die Perspektiven der Lernenden ein wesentliches Element für die Strukturierung des Unterrichts darstellen (vgl. Kattmann et al. 1997). Anliegen der Studie war es vor diesem Hintergrund, die *Verstehensprozesse* von Lernenden nach der Erstrezeption des Films „Krabat“ (Kreuzpaintner 2008)¹³ zu untersuchen, um zu Erkenntnissen über die „Spielfilmkompetenz“ (Abraham 2009) von Schülerinnen und Schülern zu gelangen.

2.2.2 Studienanlage und Vorgehen

Zentrale Erhebungsmethode, um einen Zugang zu den subjektiven Rezeptions- und Interpretationsprozessen der Lernenden zu erhalten, waren sog. „Filmgespräche“ (Abraham 2009, 78ff.), die mit den Schülerinnen und Schülern durchgeführt wurden. Ursprünglich handelt es sich beim Filmgespräch um eine Unterrichtsmethode, die im Rahmen unserer Untersuchung allerdings als Erhebungsmethode eingesetzt wurde (vgl. Schmidt/Winkler 2015, 84f.). Zentral für die Durchführung war, dass für die Schülerinnen und Schüler möglichst viel Raum zum narrativen Austausch über die Filmwahrnehmung besteht – wir als Gesprächsleiterinnen gaben lediglich Diskussionsimpulse und formulierten nur ggf. Nachfragen zur Konkretisierung einzelner Ausführungen. Das Vorgehen weist eine Nähe zur Methode des *Gruppeninterviews*¹⁴ auf: Hier wird ebenfalls eine gemeinsame Befragung von Personen vorgenommen, die im Vergleich eher gering durch einen Interviewleitfaden strukturiert ist (vgl. Niebert/Gropengießer 2014, 124).¹⁵ Der bzw. die Interviewende greift weitestgehend nicht in die Kommunikation ein – erst wenn die Diskussion in der Gruppe ‚abebbt‘, erfolgen immanente und exmanente Nachfragen an die Beforschten.

Orientiert am Forschungsinteresse der Studie wurde von uns ein spezifisch angepasster *Gesprächsleitfaden* (Bsp. 2) entwickelt, in dem die Fragen weitgehend offen und erzählgenerierend formuliert waren (vgl. Abschnitt 1.2.1). Dieses Vorgehen sollte eine möglichst selbstläufige Kommunikation bei den Schülerinnen

¹³ Zur Begründung der Filmauswahl vgl. Winkler/Schmidt (2012, 232ff.).

¹⁴ Die Durchführung von leitfadengestützten Interviews kann als Einzel- oder Gruppeninterview erfolgen (vgl. u.a. Niebert/Gropengießer 2014, 124).

¹⁵ Hier bestehen fließende Übergänge zum Verfahren der Gruppendiskussion (s. Scherf in diesem Band), bei dem allerdings die *kollektiven Orientierungen einer Gruppe* im Vordergrund stehen.

und Schülern anregen, in der aber zugleich forschungsrelevante Aspekte – zur Filmwirkung, Reflexion filmischer Darstellungsmittel, Figurenwahrnehmung etc. – im Gespräch fokussiert werden sollten. Zentral für den Einsatz von Gesprächsimpulsen im Rahmen leitfadengestützter Interviews ist es, dass durch sie gezielt vertiefende Äußerungen der Erforschten generiert werden können (vgl. Abschnitt 1.1). In unserer Untersuchung haben wir in diesem Zusammenhang zwei Filmausschnitte verwendet, die von uns als besonders affektbesetzt eingestuft wurden und zugleich als Schlüsselstellen innerhalb der Handlungsstruktur des Filmes anzusehen sind (vgl. Winkler/Schmidt 2012, 237). Die ausgewählten Sequenzen wurden den Schülerinnen und Schülern zum Ende der Filmgespräche wiederholt vorgespielt, um gezielt vertiefende Äußerungen in Bezug auf Wirkung, Deutung und Analyse filmischer Mittel zu erheben.

Leitfrage/Stimulus/Erzählaufforderung	(Nach-)Fragen mit obligatorischer Formulierung
Wenn ihr an den Film zurückdenkt, was kommt euch als erstes in den Sinn?	
(...)	
Würdet ihr anderen empfehlen, den Film anzuschauen?	Warum?
(...)	
FILMAUSSCHNITT: Neumondnacht Ich zeige euch noch einmal die Szene ... Seht sie euch bitte noch einmal an. Wie wirkt die Szene auf euch, und warum ist das so?	Vielleicht fällt euch noch etwas ein, was bisher noch nicht gesagt wurde?
(...)	
Habt ihr Lust, nachdem ihr den Film geschaut habt, das Buch (nochmal) zu lesen?	
Gibt es noch etwas, dass euch wichtig ist, und das wir noch nicht besprochen haben?	

Bsp. 2: Leitfadenauszug aus einer Studie zum Spielfilmverstehen von Schülerinnen und Schülern (Winkler/Schmidt 2012; Schmidt/Winkler 2015)

Die Filmgespräche, die auf diesem Gesprächsleitfaden basierten, wurden mit insgesamt 18 Lernenden des 6. und 7. Gymnasialjahrgangs durchgeführt. Die Schülerinnen und Schüler wurden in der Studie in verschiedene Gruppen eingeteilt – in Lernende, die das Buch *Krabat* bereits im Vorfeld gelesen hatten, und solche, denen das Buch vorab nicht bekannt war (zum Forschungsdesign vgl. Winkler/Schmidt 2012, 234ff.). Somit war davon auszugehen, dass im Vorfeld unterschiedliche Wissensvoraussetzungen in den jeweiligen Gruppen bestanden, wodurch unterschiedliche mentale Repräsentationen des Films erwartbar waren. Zugleich waren die erhobenen Daten gut vergleichbar, da durch die Leitfragen zentrale Aspekte für das Forschungsinteresse jeweils in beiden Gruppen angesprochen wurden (vgl. Abschnitt 1.4). Alle Schülerinnen und Schüler hatten den Film „Krabat“ vor unserer Studie nicht gesehen hatten – somit konnten die Erst-

eindrücke der Filmrezeption erfasst werden. In Filmsitzungen wurde den Lernenden zunächst der Film als Ganzes vorgeführt und unmittelbar im Anschluss die leitfadengestützten Filmgespräche in den einzelnen Gruppen durchgeführt.¹⁶

2.2.3 Analyse und Ergebnisse

Um empirisch fundierte Hypothesen über das filmbezogene Vorwissen der Lernenden zu formulieren, wurde das erhobene Datenmaterial anhand von deduktiv gewonnenen Auswertungskategorien analysiert, die sich aus dem Forschungsinteresse der Studie ergaben (vgl. ausführlich Winkler/Schmidt 2012, 238f.; Schmidt/Winkler 2015, 86). Die verbalen Daten in unserer Untersuchung sind als Produkte der gemeinschaftlichen Auseinandersetzung der Schülerinnen und Schüler mit dem Film zu betrachten. Die Ergebnisse bieten unter anderem Anhaltspunkte dafür, dass Schülerinnen und Schüler bereits zu einem frühen Zeitpunkt der Sekundarstufe über ein beachtliches und recht nuanciertes Filmwissen verfügen (vgl. ebd., 250; ebd., 93).¹⁷ Überdies wird auf Grundlage des vorliegenden Datenmaterials deutlich, dass die Lernenden nicht unmittelbar mit spezifischer Fachterminologie argumentieren, sondern informelles, d.h. außerschulisch erworbenes Wissen explizieren (vgl. Schmidt/Winkler 2015, 93). Der letztgenannte Befund zeigt nochmals einen wesentlichen Vorteil des Einsatzes von leitfadengestützten Interviewverfahren für deutschdidaktische Fragestellungen: Die induktiv und eher unbewusst in außerschulischen Rezeptionssituationen erworbenen Wissensbestände der Schülerinnen und Schüler wurden in den Filmgesprächen explizit gemacht und in *der Logik und Sprache der Beforschten* erhoben (vgl. Abschnitt 1.2.1).

2.3 Bilanz

In diesem Abschnitt wurde die Anwendung leitfadengestützter Interviewverfahren in deutschdidaktischen Forschungskontexten dargestellt und tiefergehend diskutiert. Im Fokus stand dabei jeweils die ‚Überführung‘ der Forschungsfrage(n) in einen konkreten Gesprächsleitfaden. Zusammenfassend sollte anhand der hier exemplarisch diskutierten Studien deutlich geworden sein, dass die Erhebungsmethode facettenreiche und tiefergehende Erkenntnisse für die Deutschdidaktik eröffnet – sofern die in Abschnitt 1.1 und 1.2 diskutierten Prinzipien berücksichtigt werden. In den angeführten Beispielen dürfte ferner deutlich geworden sein, dass leitfadengestützte Interviews im fachdidaktischen Kontext eine probate Methode darstellen, um einen Zugang zum *Verstehen* von Lernenden und Lehrenden zu bestimmten Themen bzw. Themengebieten der Deutschdidaktik zu erhalten. Die gewählten Forschungsprojekte unterscheiden sich neben der Zielgruppe auch

¹⁶ Gerade für Erhebungen mit Schülerinnen und Schülern ist es wichtig, darauf hinzuweisen, dass es im Interview nicht um eine Leistungsmessung geht, sondern ihre Ansichten zu bestimmten Themen interessieren.

¹⁷ Für eine umfassende Darstellung der Ergebnisse vgl. Winkler/Schmidt (2012, 239ff.) und Schmidt/Winkler (2015, 86ff.).

hinsichtlich der Strukturierung des Gesprächsablaufs. Mit dieser bewussten Kontrastierung in der Darstellung wird in Ansätzen deutlich, welches Spektrum an Ausgestaltungen eines Leitfadens für Untersuchungen besteht (vgl. Abschnitt 1.1); dies zeigt, wie stark die Leitfadenkonstruktion mit dem Erkenntnisinteresse der Untersuchung zusammenhängt. Und nicht zuletzt: Empirische Befunde, wie die in Abschnitt 2 diskutierten, bieten wichtige Erkenntnisse, um aus deutschdidaktischer Perspektive *begründet* Konzepte für den Unterricht zu entwickeln (vgl. Kattmann et al. 1997) und somit zur Verbesserung des Lehrens und Lernens beizutragen.

Literatur

- Abraham, Ulf: Filme im Deutschunterricht. Seelze: Klett/Kallmeyer 2009.
- BS 2012 = Bildungsstandards im Fach Deutsch für die Allgemeine Hochschulreife. Beschluss der Kultusministerkonferenz vom 18.10.2012, http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2012/2012_10_18-Bildungsstandards-Deutsch-Abi.pdf, (letzter Zugriff: 01.08.2018).
- Daase, Andrea/Hinrichs, Beatrix/Settinieri, Julia: Befragung. In: Settinieri, Julia et al. (Hrsg.): Empirische Forschungsmethoden für Deutsch als Fremd- und Zweitsprache. Eine Einführung. Paderborn: Schöningh 2014, 103-122.
- Friebertshäuser, Barbara/Langer, Antje: Interviewformen und Interviewpraxis. In: Friebertshäuser, Barbara/Langer, Antje/Prenzel, Annedore (Hrsg.): Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft. Weinheim/München: Juventa 2010, 437-455.
- Gropengießer, Harald: Qualitative Inhaltsanalyse in der fachdidaktischen Lehr-Lernforschung. In: Mayring, Philipp/Gläser-Zikuda, Michaela (Hrsg.): Die Praxis der qualitativen Inhaltsanalyse. Weinheim/Basel: Beltz 2005, 172-189.
- Helferich, Cornelia: Leitfaden- und Experteninterviews. In: Baur, Nina/Blasius, Jörg (Hrsg.): Handbuch Methoden der empirischen Sozialforschung. Wiesbaden: Springer VS 2014, 559-580.
- Helferich, Cornelia: Qualität qualitativer Daten. Manual für die Durchführung qualitativer Sozialforschung. Wiesbaden: VS Springer 2011.
- Hopf, Christel: Die Pseudo-Exploration – Überlegungen zur Technik qualitativer Interviews in der Sozialforschung. In: Zeitschrift für Soziologie 1978 (7), 2, 97-115.
- Kattmann, Ulrich: Das Modell der Didaktischen Rekonstruktion – Ein Rahmen für naturwissenschaftsdidaktische Forschung und Entwicklung. In: Zeitschrift für Didaktik der Naturwissenschaften (1997), 3, 3, 3-18.
- Kreuzpaintner, Marco: Krabat. DVD, 115 Min., Frankfurt a.M.: Twentieth Century Fox 2008.
- Kruse, Jan: Qualitative Interviewforschung: Ein integrativer Ansatz. Weinheim/Basel: Beltz Juventa 2015.

- Lehnen, Katrin/Schindler, Kirsten: Berufliches Schreiben als Lernmedium und -gegenstand. Überlegungen zu einer berufsbezogenen Schreibdidaktik in der Hochschullehre. In: Pohl, Thorsten/Steinhoff, Torsten (Hrsg.): Textformen als Lernformen. Duisburg: Gilles & Francke 2010, 233-256.
- Merton, Robert K./Kendall, Patricia L.: Das fokussierte Interview. In: Hopf, Christel/Weingarten, Elmar (Hrsg.): Qualitative Sozialforschung. Stuttgart: Klett-Cotta 1979 [zuerst 1945/1946], 171-204.
- Niebert, Kai/Gropengießer, Harald: Leitfadengestützte Interviews. In: Krüger, Dirk/Parchmann, Ilka/Schecker, Horst (Hrsg.): Methoden in der naturwissenschafts-didaktischen Forschung. Berlin/Heidelberg: Springer 2014, 121-132.
- Nohl, Arnd-Michael: Interview und dokumentarische Methode. Anleitungen für die Forschungspraxis. Wiesbaden: VS Verlag 2009.
- Rosebrock, Cornelia/Nix, Daniel: Grundlagen der Lesedidaktik und der systematischen schulischen Leseförderung. Baltmannsweiler: Schneider Hohengehren 2014.
- Schmidt, Frederike/Winkler, Iris: An informelles Filmwissen anknüpfen! Empirische Befunde zum Spielfilmverstehen von Schülerinnen und Schülern. In: Didaktik Deutsch 2015 (20), 38, 80-96.
- Schmidt, Frederike: Den diagnostischen Blick schärfen: Vorstellungen und Orientierungen von Deutschlehrerinnen und Deutschlehrern zur Diagnose von Lesekompetenz. In: Bräuer, Christoph/Wieser, Dorothee (Hrsg.): Lehrende im Blick. Empirische Lehrerforschung in der Deutschdidaktik. Wiesbaden: Springer VS 2015, 89-109.
- Schmidt, Frederike: Den Schüler im Blick – Zur Entwicklung und Erprobung eines Diagnosetools für Lesekompetenz in der Sekundarstufe I. In: Komorek, Michael/Prediger, Susanne (Hrsg.): Der lange Weg zum Unterrichtsdesign – Zur theoretischen Begründung und Umsetzung eines genuin fachdidaktischen Forschungs- und Entwicklungsprogramms. Münster: Waxmann 2013, 257-272.
- Schütze, Fritz: Zur Hervorlockung und Analyse von Erzählungen thematisch relevanter Geschichten im Rahmen soziologischer Feldforschung. In: Arbeitsgruppe Bielefelder Soziologen (Hrsg.): Kommunikative Sozialforschung. München: Fink 1976, 159-260.
- Winkler, Iris/Schmidt, Frederike: Das Unheimliche im Film – Zur Wirkung des Films Krabat auf Schülerinnen und Schüler. In: Dettmar, Ute/Oetken, Mareile/Schwagmeier, Uwe (Hrsg.): SchWellengänge. Zur Poetik, Topik und Optik des Fantastischen in Kinder- und Jugendliteratur und -medien. Frankfurt a.M.: Peter Lang Verlag 2012, 231-254.

Narrative Interviews

Stegreiferzählungen als Datengrundlage empirischer Forschung

1. Vorstellung des narrativen Interviews

Das narrative Interview wurde von Fritz Schütze in den 1970er Jahren im Rahmen soziologischer Feldforschung entwickelt; es hat sich gegenwärtig besonders in der Biographieforschung etabliert (Loch/Rosenthal 2002, 221; Flick 2009, 228).

Das narrative Interview stellt die Sichtweise der interviewten Person in spezifischer Weise in den Mittelpunkt: Die thematische Ausgestaltung wird zunächst weitgehend der interviewten Person überlassen (Schütze 1976, 163f.), um eine (ausführliche) Erzählung eigenerlebter Geschichte bzw. eigenerlebter Ereignisse (Hermanns 1981, 18ff.) zu generieren. Die Datenerhebung erfolgt somit nicht hypothesengeleitet. Sie setzt vielmehr auf den Nachvollzug von Handlungsabläufen sowie auf empirisch fundierte handlungstheoretische Konzeptionen (Loch/Rosenthal 2002, 221).

1.1 Funktionsweise des narrativen Interviews

Das narrative Interview umfasst spezifische Phasen, die aufeinander folgend die besondere Funktionsweise der Interviewform konstituieren. Im Einzelnen besteht das narrative Interview aus einem von der interviewenden Person gegebenen Erzählstimulus, der darauf einsetzenden Stegreiferzählung der interviewten Person sowie einem narrativen Nachfrageteil und einer Bilanzierungsphase. Anhand des *Erzählstimulus* wird die interviewte Person angehalten, die Geschichte eines Gegenstandsbereichs, an der sie teilgenommen hat, bzw. die Geschichte eines Ereigniszusammenhangs darzulegen. Je nach Erkenntnisinteresse kann der Erzählstimulus unterschiedlich stark – offen, leicht vorstrukturiert oder geschlossen – strukturiert (vgl. ebd., 226f.) sein: Ein offener Erzählstimulus enthält keinerlei thematische Einschränkung, er gewährt somit der interviewten Person einen größtmöglichen Handlungsspielraum; ein leicht vorstrukturierter Erzählstimulus benennt dagegen den Forschungskontext und gibt damit der Erzählung der interviewten Person einen thematischen Schwerpunkt; ein geschlossener Erzählstimulus bindet die Erzählung an eine bestimmte Themenstellung und setzt so der Erzählung Rahmen und Fokus.

In einem narrativen Interview soll die interviewte Person dazu bewegt werden, aus dem Stand heraus zu erzählen. Derartige *Stegreiferzählungen* sind nach Schütze dadurch charakterisiert, dass der Kommunikationspartner „keine systematische Ausarbeitung der Erzähltechnik vorzunehmen vermochte, die Formulierungen kalkulieren oder gar schriftlich abzirkeln und dann für die Präsentation einüben konnte“ (Schütze 1987, 237, zitiert nach Küsters 2006, 25). Stegreiferzählungen werden also anlassbezogen und spontan erzählt. Sogenannte Zugzwänge des Erzählens (vgl. Kallmeyer/Schütze 1977) beeinflussen dabei, ohne dass die oder der Erzählende sich dessen bewusst ist, die Auswahl und Darbietung des Erzählinhalts; sie fungieren somit als formgebende Mechanismen. So ist die oder der Erzählende durch den *Detaillierungszwang* beispielsweise angehalten, sich an die tatsächliche Abfolge von erlebten Ereignissen zu halten und gleichzeitig die wesentlichen kausalen bzw. motivationalen Übergänge zwischen den Ereignissen darzulegen (ebd., 188). Der Gestalterschließungszwang bewirkt wiederum, dass der Gesamtzusammenhang des Erlebten offengelegt und eine begonnene Geschichte auch zu Ende erzählt wird (ebd.). Auf diese Weise wird sichergestellt, dass die Erzählung auch von Außenstehenden verstanden werden kann. Zugleich kann der Gestalterschließungszwang dazu führen, dass die interviewte Person Informationen preisgibt, die sie unter anderen Umständen verschwiegen hätte (Küsters 2006, 28). Der Relevanzfestlegungs- und Kondensierungszwang veranlasst die Erzählerin bzw. den Erzähler, nur tatsächlich Relevantes zu erzählen und somit eine fortlaufende Gewichtung und Bewertung von Einzelereignissen unter dem Gesichtspunkt der Gesamtaussage vorzunehmen (Kallmeyer/Schütze 1977, 188). Entsprechend lassen sich aus dem Erzählten Rückschlüsse auf das Relevanzsystem der interviewten Person ziehen (vgl. Loch/Rosenthal 2002, 224).

Während die interviewte Person *erzählt*, ist es an der interviewenden Person, die Perspektive der interviewten Person nachzuvollziehen, dabei von Werturteilen abzusehen und stattdessen bei der Verbalisierung insbesondere von Gefühlen und schweren Lebenserfahrungen zu unterstützen (Herrmanns 1995, 184; Loch/Rosenthal 2002, 229). Das Ende der Haupterzählung lässt sich daran erkennen, dass die interviewte Person eine Koda im Sinne von „Tja, das war’s so im großen und ganzen. Ich hoffe, Sie konnten was damit anfangen.“ (Herrmanns 1995, 184) gebraucht.

Im Rahmen des *narrativen Nachfrageteils*, der sich an die Haupterzählung anschließt, hat die interviewende Person die Möglichkeit, die Interviewpartnerin bzw. den Interviewpartner durch narrative Fragen dazu zu bewegen, fragmentarisch gebliebene Erzählansätze zu vervollständigen (ebd.; Flick 2009, 229). Hierzu gibt sie der interviewten Person erneute Erzählstimuli in Bezug auf das bereits Thematisierte. Derartige erzählgenerierende Nachfragen sind besonders in Bezug auf jene Passagen der Haupterzählung wichtig, in denen die interviewte Person nicht erzählt, sondern beschrieben oder argumentiert hat. Loch und Rosenthal (2002, 230) unterscheiden die folgenden sechs narrativen Nachfragtypen:

1. Ansteuern einer Lebensphase: ‚Können Sie mir über die Zeit (ihrer Kindheit, etc.) noch etwas mehr erzählen?‘
2. Eröffnung eines temporalen Rahmens bei scheinbar statischen Themen: ‚Sie erwähnten Ihre Mutter, können Sie einmal von Ihren frühesten Erinnerungen erzählen und was Sie mit Ihrer Mutter im Laufe Ihres Lebens erlebt haben?‘
3. Ansteuern einer benannten Situation: ‚Sie erwähnten vorhin die Situation x, können Sie mir diese noch einmal genauer erzählen?‘
4. Ansteuern einer Erzählung zu einem Argument: ‚Können Sie sich noch an eine Situation erinnern, in der Sie sich hilflos beim Anblick eines Sterbenden gefühlt haben?‘
5. Ansteuern von Tradiertem bzw. Fremderlebtem: ‚Können Sie sich noch an eine Situation erinnern, in der Ihnen davon erzählt wurde, wie Ihr Vater gestorben ist?‘
6. Ansteuern von Phantasien oder von Zukunftsvorstellungen: ‚Können Sie sich noch an eine Situation erinnern, in der Sie sich vorgestellt haben, wie Ihr Großvater im Krieg verwundet wurde (oder: dass Sie aus dem Pflagedienst ausscheiden möchten)?‘

Generiert die interviewte Person keine neuen Erzählungen mehr, geht das Interview in die *Bilanzierungsphase* über. Während dieser Phase wird die interviewte Person als Expertin bzw. Experte und Theoretikerin bzw. Theoretiker angesehen (vgl. Schütze 1983a). Ihr können und sollen deshalb nun Fragen gestellt werden, die sie dazu anhalten, zu beschreiben und zu argumentieren, das Erfahrene bzw. Erzählte also theoretisch zu erklären und zu bilanzieren (Hermanns 1995, 184; Flick 2009, 230).

Nach dieser letzten Phase des narrativen Interviews empfiehlt es sich – mündlich oder schriftlich –, solche soziodemographischen Merkmale der Interviewpartnerin bzw. des Interviewpartners zu erfassen, die bei der Auswertung eine zusätzliche Perspektive auf das Datenmaterial eröffnen können.

1.2 Ausgangsdaten und Ergebnisse

Die dominante Darstellungsform im narrativen Interview ist die *Erzählung* (Schütze 1984, 91; Hermanns 1995, 183): Aus der Perspektive gegenwärtiger Erinnerung werden in ihr vergangene Erfahrungen rekonstruiert und in einen Zusammenhang gebracht; eine Ereignis- und/oder eine Zustandsänderung bildet hierbei den Ausgangspunkt (Schütze 1984, 91). Üblicherweise enthält eine Erzählung nicht nur narrative Passagen, sondern auch Passagen, in denen die Darstellungsformen der *Beschreibung* und der *Argumentation* überwiegen. Argumentative Passagen erlauben einen Einblick in die orientierungs-, erklärungs- und/oder biographie- bzw. identitätstheoretischen Vorstellungen der interviewten Person; in beschreibenden Passagen werden zum Beispiel soziale Bedingungen und Merkmale von Aktanten oder situative Umstände geschildert (ebd.) bzw. Zustände, Situationen und typische Verfahrensweisen dargestellt, denen die interviewte Person eine Relevanz für das Verständnis des dargestellten Ereignisses beimisst (Hermanns 1995, 183). Werden die von der narrativen Kommunikationsstruktur abweichenden Darstellungsformen nicht funktional in Bezug auf das Erzählte gebraucht, können sie ein Indiz dafür darstellen, dass die interviewte

Person versucht, etwas zu verschleiern bzw. von etwas abzulenken. Weitere Indizien für ein derartiges Verhalten sind Verzögerungspausen sowie der Versuch, einen Wechsel der Sprecherrollen herbeizuführen (vgl. Schütze 1976, 198). Dass die Erzählung nicht (mehr) den Zugzwängen des Erzählens (s. Kap. 1.1) folgt, kann schließlich auch daran deutlich werden, dass die interviewte Person von dialektal gefärbter, spontaner Umgangssprache in eine förmlich reflektierte Standardsprache übergeht (vgl. Lindow 2013, 129).

Vor diesem Hintergrund empfiehlt es sich zum einen, die Interviews aufzuzeichnen, anschließend den konkreten Wortlaut niederzuschreiben und dabei auch Pausen, Dehnungen, Verzögerungssignale wie „äh“ und Rezeptionssignale wie „hm“ sowie vorgenommene Akzentuierungen zu transkribieren (s. hierzu den Exkurs von Heller/Morek in diesem Band). Zum anderen sollten die Interviewdaten mit einer Methode ausgewertet werden, die die unterschiedlichen Darstellungsformen bzw. Textsorten explizit berücksichtigt und zugleich unterscheidet zwischen dem, was gesagt wird, und dem, wie es gesagt wird. Eine Auswertung nach der *dokumentarischen Methode* (vgl. Bohnsack 2007; s. Schieferdecker in diesem Band) ergänzt um die Elemente der *Narrationsstrukturanalyse* (Schütze 1983a, b), wie sie Nohl (2009) erstmalig vorschlägt, bietet sich deshalb besonders an. Hierbei wird davon ausgegangen, dass sich das konjunktive Wissen der interviewten Personen, das auf einer gemeinsam gelebten, sozialen Handlungs- und Erfahrungspraxis beruht (Mannheim 1964, 98) und in impliziter Form vorliegt (vgl. Nohl 2009, 10), in den narrativen und beschreibenden Passagen widerspiegelt, und dass das kommunikative Wissen als ein öffentliches und reflexiv zugängliches Wissen mit den Textsorten der Argumentation und Bewertung korrespondiert (vgl. ebd., 49).

Mittels einer *vergleichenden Sequenzanalyse* über alle Interviews hinweg lassen sich schließlich in mehrfacher Hinsicht Typen bilden: in Bezug auf die Semantik der Erzählungen (sinngenetische Typenbildung) sowie hinsichtlich spezifischer Erfahrungshintergründe der interviewten Personen (soziogenetische Typenbildung). Die Mehrdimensionalität der erhobenen Daten erfährt auf diese Weise in besonderer Weise Würdigung.

1.3 Vorteile und Nachteile des narrativen Interviews

Narrative Interviews eignen sich für autobiographische Forschungsfragen ebenso wie für die Erhebung und Untersuchung von thematisch gebundenen Erzählungen (vgl. ebd., 23).

Die Besonderheit narrativer Interviews – die Zurückhaltung der interviewenden Person sowie der lange, zusammenhängende und komplexe Redefluss der interviewten Personen – bringen für die Datengenerierung spezifische Vorteile. Dazu gehören insbesondere die relative Offenheit der Erhebung, die es ermöglicht, Forschungsfragen nachzugehen, die auf komplexe Zusammenhänge abzielen und sich nicht standardisiert untersuchen lassen, sowie die Generierung von Daten, die relativ wenig durch die interviewende Person verzerrt sind.

Die Verteilung des Rederechts im narrativen Interview aber stellt bei der Durchführung des Interviews für die interviewte Person wie für die interviewende Person auch eine besondere Herausforderung dar. Diese Herausforderung ist umso größer, je mehr auf beiden Seiten die allgemeine Vorstellung von einem Interview besteht, wonach die maßgebliche Gesprächssteuerung von der interviewenden Person ausgeht. Die Interviewsituation kann unter diesen Umständen ungewohnt bzw. sogar künstlich erscheinen (vgl. Küsters 2006, 22; Hermanns 1995, 184f.), die eigene Rolle zu akzeptieren und sich darauf einzulassen, als vergleichsweise schwierig wahrgenommen werden. Für die interviewende Person ist es deshalb umso wichtiger, sich immer wieder dazu anzuhalten, tatsächlich nach der Erzählung einer Geschichte von Ereignissen zu fragen und nicht nach Motiven, Zuständen oder Routinen (vgl. Hermanns 1995, 185).

1.4 Gütekriterien für den Einsatz des narrativen Interviews

Spezifische Gütekriterien für das narrative Interview ergeben sich zunächst einmal aus dem formalen Anspruch, Stegreiferzählungen zu generieren. So muss sich der Forschungsgegenstand tatsächlich auf der Grundlage von Erzählungen erheben lassen und so muss der Erzählimpuls tatsächlich auch Erzählungen hervorbringen können.

Die Gütekriterien für das narrative Interview berühren aber auch die Annahme, dass sich über Erzählungen ein Zugang zu tatsächlichen Erfahrungen eröffnet (vgl. Schütze 1984, 78f.). Diese Annahme einer Analogie von retrospektiver Erzählung und Erfahrung wird mittlerweile jedoch nicht mehr aufrecht erhalten. Stattdessen wird das ursprünglich von Schütze ausgewiesene Ziel, aus Erzählungen vergangene Ereignisse oder vergangenes Handeln zu erschließen, in der Zielstellung aufgehoben, die Erfahrungsaufschichtung des Handelnden zu durchdringen (vgl. Küsters 2006, 32f.). Die Frage nach dem Wahrheitsgehalt des Erzählten rückt somit in den Hintergrund. Angesichts dessen ist die Forscherin bzw. der Forscher dazu gezwungen, präzise zu reflektieren, worin genau das Forschungsinteresse ihrer bzw. seiner Arbeit besteht – mit anderen Worten – was genau überhaupt erhoben werden soll.

Schließlich ist auch zu beachten, dass in einem narrativen Interview, die zu erforschenden Gegenstände bzw. Sachverhalte nicht nur erhoben, sondern zum Teil auch konstituiert werden können (vgl. Küsters 2006, 38). Bei der Analyse des Datenmaterials ist deshalb zu untersuchen, ob bzw. inwieweit die argumentierenden Passagen erkennen lassen, wann der (wesentliche) Reflexionsprozess stattgefunden hat bzw. wodurch er initiiert wurde. Der Anteil der Interviewsituation bzw. der interviewenden Person an der Erfahrungsaufschichtung der interviewten Person sollte möglichst gering sein.

1.5 Literaturempfehlungen

Schütze (1976): Zur Hervorlockung und Analyse von Erzählungen thematisch relevanter Geschichten im Rahmen soziologischer Feldforschung.

In dem Beitrag vermittelt Fritz Schütze anhand eines soziologischen Forschungsprojektes einen Einblick in die theoretische Begründung und Anwendung der narrativen Interviewform.

Loch/Rosenthal (2002): Das narrative Interview.

Der Beitrag stellt eine zentrale Aufarbeitung der narrativen Interviewform dar. Die Autorinnen unterlegen ihre Ausführungen mit zahlreichen Beispielen – für die (erste) konkrete Planung einer eigenen Interviewstudie eine sehr hilfreiche Unterstützung.

Nohl (2009): Interview und dokumentarische Methode.

Nohl bietet eine umfassende theoretische Begründung zum Einsatz der dokumentarischen Methode für die Auswertung, auch und insbesondere von narrativen Interviews. Er zeigt auf, wie sich die dokumentarische Methode mit Elementen der Narrationsstrukturanalyse bei der Auswertung narrativer Interviews verknüpfen lässt.

Flick (1996): Psychologie des technisierten Alltags.

In Hinblick auf das narrative Interview empfehlen sich die Ausführungen von Flick besonders hinsichtlich seiner Abgrenzung vom episodischen Interview, das mit seinem Fokus auf Situationen und Episoden eine große Nähe zum narrativen Interview aufweist.

2. Literaturunterricht als Fall – die narrative Erhebung kasuistischen Wissens von Deutschlehrenden

Wie das narrative Interview für die deutschdidaktische Forschung genutzt werden kann, wird z.B. im Dissertationsprojekt „Literaturunterricht als Fall. Kasuistisches Wissen von Deutschlehrenden“ (Lindow 2013) deutlich.

2.1 Fragestellung und Zielsetzung der Interviewstudie

Die Dissertation greift die fachdidaktische Forschungslinie zu Kognitionen und Überzeugungen von Lehrkräften in Bezug auf die kasuistische Wissensform auf. Kasuistisches Wissen ist ein episodisches Wissen (Strube/Janetzko 1990), das durch narratives Denken (Bruner 1985) entsteht bzw. in narrativer Form (Shulman 2004) vorliegt sowie auf der theoriebasierten Analyse und Interpretation spezifischer Ereignisse und deren Kontextmerkmalen beruht (Shulman 1986). Dabei gilt: Nicht jede Begebenheit im Unterricht ist zwangsläufig ein Fall, kann durch Analyse und Interpretation aber immer zu einem Fall gemacht werden (vgl. ebd.). Kasuistisches Wissen kann demnach auch, solange nicht verallgemeinert, als ein subjektives Wissen (vgl. Neuweg 2011) verstanden werden, das (a) Aufschluss über die kognitiven Strukturen von Lehrenden gibt, (b) eine Parallelisierung mit explizitem Professionswissen im Sinne eines objektiven Ausbildungswissens (ebd.) erlaubt und (c) eine Rekonstruktion von Handlungsepisoden ermöglicht, in denen sich das Können bzw. Handeln der betreffenden Lehrperson spiegelt. Lehrende können ihr kasuistisches Wissen zumindest rekonstruieren und damit auch explizieren (vgl. Shulman 1986).

Die Studie untersucht, worin für Lehrende des Faches Deutsch problematische Fälle des gymnasialen Literaturunterrichts bestehen. Im Einzelnen werden folgende Fragen gestellt:

1. Welche Orientierungsmuster haben Lehrpersonen aus ihren Unterrichtserfahrungen als Handlungs- und Reflexionsgrundlage ihrer gegenwärtigen Lehrtätigkeit generiert?
2. Innerhalb welcher Problemfelder des Literaturunterrichts sind diese Orientierungsmuster organisiert?
3. Welche Bewältigungsstrategien haben Lehrende im Rahmen dieser Problemfelder angewendet bzw. entwickelt?

Zu Beantwortung der Fragen wurde die Generierung von Typologien in Bezug auf Problemfelder des Literaturunterrichts einerseits und Bewältigungsstrategien von Lehrenden andererseits angestrebt.

2.2 Methodisches Vorgehen bei der Datenerhebung und Auswertung

Fragestellung und Zielsetzung der Arbeit ließen die Wahl auf das narrative Interview als Erhebungsinstrument und auf die dokumentarische Methode (s. Schieferdecker in diesem Band) als Auswertungsverfahren fallen.

2.2.1 Entscheidung für die narrative Interviewform

Die Entscheidung für das narrative Interview als Erhebungsinstrument der Studie ging zum einen aus den *Fragestellungen* der Untersuchung hervor und begründete sich zum anderen aus den Merkmalen des zu erhebenden *Konstrukts*, der kasuistischen Wissensform. Die narrative Form der kasuistischen Wissensform und ihre Explizierbarkeit stellen schließlich die Merkmale dar, die für die Erhebung im Rahmen der Dissertation zentral sind. Sie bilden die Voraussetzung dafür, die kasuistische Wissensform über *Erzählungen* zu erfassen.

Eine derartige Erhebung von Erzählungen wäre grundsätzlich auch mit dem episodischen Interview möglich gewesen. Das episodische Interview ist eine von Uwe Flick (1996) entwickelte Interviewform, die zwischen narrativ-episodischem Wissen einerseits und semantischem Wissen in Bezug auf einen Gegenstandsbereich andererseits unterscheidet und beide Wissensformen auch getrennt erhebt (vgl. Flick 1996, 238ff.). Die Entscheidung für das narrative Interview und gegen das episodische Interview folgte schließlich im Wesentlichen zwei Überlegungen: Erstens, nur das narrative Interview erlaubt es, die kasuistische Wissensform als Ganzes, d.h. ohne künstliche Trennung in narratives und reflexiv-theoretisches Wissen zu erheben. Zweitens, nur mit dem narrativen Interview lässt sich eine Ausrichtung der Erhebung an Erzählungen über Begebenheiten, die Deutschlehrende vor fachspezifische Probleme gestellt haben, durchführen. Das episodische Interview würde dagegen der Beantwortung einer Frage nach dem Wissen von Deutschlehrenden in einem bestimmten Bereich bzw. in Bezug auf einen bestimmten Unterrichtsgegenstand dienen, unterlegt mit konkreten Beispielen aus dem Unterricht.

Die Entscheidung für das narrative Interview im Rahmen des Dissertationsprojekts erfolgte schließlich auch unter Berücksichtigung weiterführender Gütekriterien der Interviewform (s. Kap. 1.4). Zur *Homologie-Annahme* und zur Frage des *Wahrheitsgehalts* sei an dieser Stelle ausgeführt: Da die Interviewstudie nicht darauf abzielt, objektive Unterrichtsereignisse zu erheben, sondern den Fokus auf das Wissen, das Deutschlehrenden hinsichtlich einer bestimmten Begebenheit und deren Analyse und Reflexion im Gedächtnis geblieben ist, setzt und dabei auch die Frage, ob dieses Wissen handlungsleitendes Wissen darstellt, ausklammert, ist die Debatte um die Annahme einer Analogie von retrospektiver Erzählung und Erfahrung für das Projekt von untergeordneter Bedeutung.

2.2.2 Ausgestaltung des narrativen Interviews

Für die Durchführung des narrativen Interviews wurde eine leicht vorstrukturierte Eingangsfrage bzw. *Narrationsaufforderung* gewählt (vgl. Loch/Rosenthal 2002, 227). Sie lautet:

Mit meiner Arbeit möchte ich herausfinden, worin im gymnasialen Literaturunterricht fachliche bzw. fachdidaktische Probleme und Problemsituationen bestehen und welche Möglichkeiten es gibt, mit diesen umzugehen. Dazu bin ich an problematischen Fällen interessiert, die die unterrichtende Lehrperson in ihrem Handeln als Deutschlehrerin bzw. Deutschlehrer nachhaltig beeinflusst haben, Fälle, die zum Beispiel bewirkt haben, dass die Lehrperson eine bestimmte Methode oder eine Schwerpunktsetzung geändert hat oder zumindest reflektiert hat. Ich möchte Sie bitten, mir einen solchen Fall aus Ihrem Literaturunterricht (in Bezug auf Lyrik) zu erzählen, möglichst so, dass ich mir ein genaues Bild von der Begebenheit, der Situation machen kann. (Lindow 2013, 121)

In dem Vertrauen, dass die interviewten Lehrpersonen im Rahmen ihrer Erzählung auch Informationen preisgeben, die sie in einem anderen Interviewkontext verschwiegen hätten (vgl. Kallmeyer/Schütze 1977, 188), benennt diese Narrationsaufforderung den hinsichtlich seiner Ausrichtung auf problematische Situationen durchaus sensiblen Forschungskontext. Darüber hinaus macht die Narrationsaufforderung Vorgaben in Bezug auf den thematischen Schwerpunkt der Erzählung (unerwartete bzw. problematische Situation, gymnasialer Literaturunterricht, Reflexivität und subjektive Bedeutsamkeit). Die Eingrenzungen, die auf diese Weise vorgenommen wurden, dienten dazu, ein Mindestmaß an Vergleichbarkeit der generierten Fallschilderungen zu sichern.

Die Narrationsaufforderung lässt aber auch Spielraum für Varianz unter den Fallschilderungen. So blieb es den interviewten Deutschlehrenden überlassen, sowohl die zeitliche Extension (z.B. Unterrichtsstunde oder eine mehrere Stunden umfassende Unterrichtseinheit) als auch den Fokus (eigene Person, Klasse als Ganzes, einzelne Schülerinnen und Schüler, Lehrer-Schüler-Interaktion usw.) des Falles sowie die thematische bzw. methodische Verankerung im Fach frei zu wählen. Nur für die Hälfte der interviewten Lehrenden, und hier auch keinesfalls bindend, wurde der inhaltliche Fokus noch einmal auf den Lyrikunterricht eingeschränkt. Ziel dieser zum Teil zusätzlichen Einschränkung war es, für einen Bereich aus dem Literaturunterricht besonders differenzierte Aussagen treffen zu können.

Für die Nachfragephase wurden *narrative Nachfragen* zur Ansteuerung und Präzisierung einer benannten Situation, zur Eröffnung eines temporalen Rahmens für statisch thematisierte Themen sowie zur Ansteuerung einer Erzählung zu einem Argument (vgl. Loch/Rosenthal 2002, 230) wie folgt konzipiert:

- Ansteuern einer benannten Situation: ‚Sie erwähnten die Situation ... Können Sie mir diese noch einmal genauer erzählen?‘
- Eröffnung eines temporalen Rahmens in Bezug auf statisch präsentierte Themen: ‚Sie erwähnten ... Können Sie die Begebenheit/Ausschnitte der Begebenheit noch einmal in Bezug auf diesen Gesichtspunkt erzählen?‘
- Ansteuern einer Erzählung zu einem Argument: ‚Sie haben argumentiert, dass ... Können Sie die Begebenheit/Ausschnitte der Begebenheit noch einmal in Bezug auf dieses Argument erzählen?‘ (Lindow 2013, 286)

Die *Bilanzierungsphase* des Interviews wurde schließlich dazu genutzt, einen genaueren Einblick in die reflexive Komponente des kasuistischen Wissens der interviewten Lehrenden zu erhalten. Entsprechende Fragen zielten deshalb auf eine Präzisierung der Rahmenbedingungen der erzählten Begebenheit sowie auf eine Reflexion des eingetretenen Problems.

Nach Beendigung des Interviews wurden die interviewten Lehrpersonen gebeten, einen soziodemographischen Fragebogen auszufüllen. Diese Erhebung der soziodemographischen Daten erfolgte entgegen der Empfehlung von Küsters (2006, 48) schriftlich, um das eigentliche Interview nach der Bilanzierungsphase tatsächlich beenden zu können.

2.2.3 Auswertung der narrativen Interviewdaten

Die erhobenen Daten wurden mit der dokumentarischen Methode nach Nohl (2009; s. Schieferdecker in diesem Band) analysiert. Dazu wurden die narrativen Textsequenzen – unter besonderer Berücksichtigung des die jeweilige Begebenheit konstituierenden Problems – einer komparativen Sequenzanalyse unterzogen (vgl. Nohl 2009, 51f.). Passagen, in denen die interviewten Lehrenden argumentierten und bewerteten, wurden vor allem auf Schlüssigkeit im Hinblick auf die Erzählung untersucht.

Ziel der komparativen Sequenzanalyse war die Bildung von sinn- und soziogenetischen Typen (vgl. Bohnsack 2008; Nohl 2009) und damit zugleich die Generalisierung der gewonnenen Erkenntnisse. Im Rahmen der *sinn-genetischen Typenbildung* wurden die jeweiligen Orientierungsrahmen in den Interviews rekonstruiert und im Fallvergleich abstrahiert. Es wurde also untersucht, welches Problem die Lehrpersonen jeweils vordergründig wahrnahmen und welche grundlegenden Probleme des Literaturunterrichts sich in diesen Problemen widerspiegelten. Im Sinne einer *soziogenetischen Typenbildung* wurden die einzelnen Orientierungsrahmen außerdem auf soziale Zusammenhänge analysiert und die Prüfung einer Gruppierung der Fälle unter soziographischen Aspekten geprüft.

2.3 Ergebnisse

Mithilfe der *sinngenetischen Typenbildung* ließen sich folgende für den Literaturunterricht typische Problemfelder herausarbeiten:

- ‚Lernende nehmen eine ablehnende Haltung gegenüber einem literarischen Text ein‘,
- ‚Lernenden bereitet es Schwierigkeiten, einen sinnhaften Zugang zu einem literarischen Text zu finden‘ sowie
- ‚Lernende haben Probleme, sprachliche Mittel inhaltlich zu deuten‘ und ‚Lernende vertreten eine von der Lesart der Lehrperson abweichende Interpretation‘.

Innerhalb dieser Problemfelder dokumentieren sich Handlungsdilemmata und somit systematisch bedingte Unsicherheiten für Lehrende im Literaturunterricht (vgl. auch Lindow/Wieser 2013; Lindow 2015).

Im Zuge der *soziogenetischen Typenbildung* konnten die von den Lehrenden eingesetzten Bewältigungsstrategien ‚affektiv-statisch‘ und ‚pragmatisch-flexibel‘ identifiziert werden. Diese Formen von Unterrichtshandeln weisen eine Abhängigkeit vom Alter bzw. von der Berufserfahrung der interviewten Lehrenden zum Zeitpunkt der erfahrenen Begebenheit auf: Ältere bzw. erfahrene Lehrpersonen erscheinen in ihrem Unterrichtshandeln eher emotional gelassen zu sein und fähig, methodisch und inhaltlich flexibel auf Herausforderungen zu reagieren. In der Analyse und Bewertung von problematischen Unterrichtssituationen reflektieren sie auch ihre eigene Rolle. Jüngere bzw. weniger erfahrene Lehrpersonen neigen dagegen dazu, emotional auf Unterrichtsprobleme zu reagieren. Bei der Problembewältigung versuchen sie, an ihrer ursprünglichen Unterrichtsplanung festzuhalten; bei der Problemreflexion tendieren sie dazu, ausschließlich Merkmale der Schülerinnen und Schüler sowie des fachlichen Gegenstands in den Blick zu nehmen.

2.4 Methodendiskussion

Für die Erhebung kasuistischen Lehrerwissens erwies sich das narrative Interview als geeignetes Erhebungsinstrument. Im Einzelnen lassen sich folgende Erfahrungen in Bezug auf den Einsatz der Interviewform im Rahmen des Dissertationsprojekts zusammenfassen.

Die *Narrationsaufforderung* wurde von den interviewten Deutschlehrenden insgesamt positiv aufgenommen: Die Lehrpersonen zeigten sich mehrheitlich aufgeschlossen gegenüber der Interviewform, und sie schätzten das Interesse an ihren konkreten Unterrichtserfahrungen sowie die ihnen ermöglichte Freiheit bei der Ausrichtung und Darbietung ihrer Fallerzählung. Einige Lehrpersonen hinterfragten die Interviewform jedoch auch offensiv bzw. entzogen sich der Erzählung eines konkreten Unterrichtsfalls, indem sie sich bzw. das eigene Tun immer wieder zu rechtfertigen bzw. zu legitimieren versuchten. In der Regel war es während des Interviews jedoch möglich, das Vertrauen der Lehrpersonen zu gewinnen und

nach und nach auch über sensiblere Aspekte ihrer Unterrichtspraxis zu sprechen und diese in konkrete Erzählungen zu überführen.

Auf der Grundlage der (ursprünglichen) Narrationsaufforderung, die den Literaturunterricht im Allgemeinen für Fall Erzählungen öffnete, wurde eine Vielzahl sehr unterschiedlicher Fall Erzählungen generiert. Diese hohe thematische Varianz unter den Fall Erzählungen erwies sich in Bezug auf das Ziel der Arbeit, Typologien und somit auch verallgemeinerbare Ergebnisse zu generieren, als problematisch. Aus diesem Grund wurde der Fokus der Narrationsaufforderung im Zuge der Erhebung auf den Lyrikunterricht eingegrenzt.

Der eigentliche Hauptteil des Interviews, die *Stegreiferzählung* der Deutschlehrenden, gestaltete sich in der zusammenhängenden Form oftmals sehr kurz. Einige Lehrpersonen gaben sogar an, sich nicht an Details erinnern zu können. Die relative Kürze bzw. Knappheit einiger Fallschilderungen lässt sich mitunter auf die teilweise beobachtete anfängliche Unsicherheit von interviewten Lehrpersonen zurückführen bzw. mit der für ein Interview eher ungewohnten Redeverteilung erklären (vgl. Kap. 1.3). Sie kann aber auch als Indiz dafür angesehen werden, dass kasuistisches Wissen nicht (grundsätzlich) in der angenommenen Detailliertheit vorliegt oder von Lehrenden nicht immer aus dem Stegreif in Gänze expliziert werden kann. Es wäre dann die Frage zu stellen, ob und inwieweit das Konstrukt kasuistisches Wissen – mithilfe der narrativen Interviewform – tatsächlich operationalisiert werden kann. Da jedoch die erhobenen Fall Erzählungen mehrheitlich einen hohen, für die Fragestellung der Arbeit hinreichenden Grad an Detailliertheit aufweisen, erscheint es legitim, von validen Daten auszugehen. Die Lehrpersonen, deren Erinnerungen an ein Ereignis zunächst zu abstrakt ausfielen, sind in der Regel dazu übergegangen, von einer zeitnahen Begebenheit zu erzählen.

Im Rahmen des *Nachfrageteils* der Interviewstudie wurde von der strengen Abfolge und damit von der Trennung narrativer und bilanzierender Nachfragen, wie sie in der Literatur konzipiert ist (vgl. Kap. 1.1), abgewichen. Diese Abweichung entstand einerseits, weil die interviewten Lehrpersonen während der Bilanzierungsphase häufig neue Aspekte einbrachten, die für die erzählte Begebenheit wichtig und bereichernd erschienen und deshalb von mir als Interviewerin noch einmal narrativ nachgefragt wurden. Andererseits wurde die konsequente Abgrenzung von narrativer Nachfragephase und Bilanzierungsphase aufgebrochen, um auch Fragen stellen zu können, die zwar zeitlich verzögert entstanden, deren Beantwortung mir aber bedeutsam erschien. Der konzeptionierte lineare Aufbau des Interviews gestaltete sich so in der Durchführung der Interviewstudie teilweise spiralförmig.

In der Nachfragephase wurde darüber hinaus deutlich, dass die den Lehrpersonen gestellten Fragen vereinzelt zur Reflexion im Rahmen der Interviews geführt haben. Der Forschungsgegenstand, das kasuistische Wissen der Deutschlehrenden, wurde in einigen Fällen somit teilweise erst im Rahmen des Interviews konstituiert.

Trotz dieser kritischen Momente in der Durchführung der narrativen Interviews ist abschließend hervorzuheben, dass mittels der narrativen Interviewform Fallergählungen erhoben werden konnten, die in ihrer Komplexität einen tiefen Einblick in die Problem- und Spannungsfelder von Literaturunterricht erlauben, und die es darüber hinaus ermöglichen, Antinomien zu rekonstruieren, die den Schwierigkeiten der Lehrpersonen zugrunde liegen (vgl. auch Lindow/Wieser 2013) und letztendlich auch eine systematische Unsicherheit von Literaturunterricht bedingen (vgl. Lindow 2015).

Literatur

- Bohnsack, Ralf: Typenbildung, Generalisierung und komparative Analyse: Grundprinzipien der dokumentarischen Methode. In: Bohnsack, Ralf/Nentwig-Gesemann, Iris/Nohl, Arnd-Michael (Hrsg.): Die dokumentarische Methode und ihre Forschungspraxis. Grundlagen qualitativer Sozialforschung. Wiesbaden: VS Verlag für Sozialwissenschaften 2007, 225-253.
- Bruner, Jerome S.: Narrative and Paradigmatic Modes of Thought. In: Eisner, Elliot (Hrsg.): Learning and Teaching the Ways of Knowing. Chicago, Illinois: The University of Chicago Press 1985, 97-115.
- Flick, Uwe: Psychologie des technisierten Alltags. Opladen: Westdeutscher Verlag 1996.
- Flick, Uwe: Qualitative Sozialforschung. Eine Einführung. Reinbek: Rowohlt 2009.
- Hermanns, Harry: Das narrative Interview in berufsbiographisch orientierten Untersuchungen. Arbeitspapiere des Wissenschaftlichen Zentrums für Berufs- und Hochschulforschung an der Gesamthochschule Kassel (1981), 9.
- Hermanns, Harry: Narratives Interview. In: Flick, Uwe et al. (Hrsg.): Handbuch Qualitative Sozialforschung. Weinheim: Beltz 1995, 182-185.
- Kallmeyer, Werner/Schütze, Fritz: Zur Konstitution von Kommunikationsschemata der Sachverhaltsdarstellung. In: Wegner, Dirk (Hrsg.): Gesprächsanalysen. Vorträge, gehalten anlässlich des 5. Kolloquiums des Instituts für Kommunikationsforschung und Phonetik, Bonn, 14.-16. Oktober 1976. Hamburg: Helmut Buske Verlag 1977, 159-274.
- Küsters, Ivonne: Narrative Interviews. Grundlagen und Anwendungen. Wiesbaden: VS Verlag für Sozialwissenschaften 2006.
- Lindow, Ina/Wieser, Dorothee: Literaturunterricht als Fall philologischer Praxis: Antinomien und Lösungsversuche. In: Zeitschrift für Germanistik (2013), 2, 390-404.
- Lindow, Ina: Literaturunterricht als Fall. Kasuistisches Wissen von Deutschlehrenden. Wiesbaden: VS Verlag für Sozialwissenschaften 2013.
- Lindow, Ina: Systematische Schwierigkeiten in der schulischen Förderung literarischer Rezeptionskompetenz. In: Bräuer, Christoph/Wieser, Dorothee (Hrsg.): Lehrende im Blick. Empirische Lehrerforschung in der Deutschdidaktik. Wiesbaden: Springer VS Verlag für Sozialwissenschaften 2015, 35-48.
- Loch, Ulrike/Rosenthal, Gabriele: Das narrative Interview. In: Schaeffer, Doris/Müller-Mundt/Gabriele (Hrsg.): Qualitative Gesundheits- und Pflegeforschung. Bernet et al.: Hans Huber 2002, 221-232.

- Mannheim, Karl: Beiträge zur Theorie der Weltanschauungsinterpretation. In: Wolff, Kurt H. (Hrsg.): Karl Mannheim. Wissenssoziologie. Aus dem Werk. Neuwied/Berlin: Luchterhand 1964, 91-154.
- Neuweg, Georg Hans: Das Wissen der Wissensvermittler. Problemstellungen, Befunde und Perspektiven der Forschung zum Lehrerwissen. In: Terhart, Ewald/Bennewitz, Hedda/Rothland, Martin (Hrsg.): Handbuch der Forschung zum Lehrerberuf. Münster: Waxmann 2011, 451-477.
- Nohl, Arnd-Michael: Interview und dokumentarische Methode. Anleitungen für die Forschungspraxis. Wiesbaden: VS Verlag für Sozialwissenschaften 2009.
- Schütze, Fritz: Biographieforschung und narratives Interview. In: Neue Praxis (1983a), 3, 283-293.
- Schütze, Fritz: Das narrative Interview in Interaktionsfeldstudien I. Studienbrief der Fernuniversität Hagen 1987.
- Schütze, Fritz: Kognitive Figuren des autobiographischen Stegreiferzählens. In: Kohli, Martin/Robert, Günther (Hrsg.): Biographie und soziale Wirklichkeit. Neue Beiträge und Forschungsperspektiven. Stuttgart: J.B. Metzlersche Verlagsbuchhandlung 1984, 78-117.
- Schütze, Fritz: Prozeßstrukturen des Lebenslaufs. In: Matthes, Joachim/Pfeifenberger, Arno/Stosberg, Manfred (Hrsg.): Biographie in handlungswissenschaftlicher Perspektive. Nürnberg: Verlag der Nürnberger Forschungsgemeinschaft 1983b, 67-156.
- Schütze, Fritz: Zur Hervorlockung und Analyse von Erzählungen thematisch relevanter Geschichten im Rahmen soziologischer Feldforschung. In: Arbeitsgruppe Bielefelder Soziologen (Hrsg.): Kommunikative Sozialforschung. München: Fink 1976, 159-260.
- Shulman, Lee S.: Those Who Understand: Knowledge Growth in Teaching. In: Educational Researcher (1986), 15, 2, 4-14.
- Shulman, Lee S.: Just in Case: Reflections on Learning from Experience. In: Wilson, Suzanne (Hrsg.): Lee S. Shulman. The Wisdom of Practice. Essays on Teaching, Learning, and Learning to Teach. San Francisco.: Jossey-Bass 2004, 461-482.

Gruppendiskussionen

Ein Verfahren zur Erhebung kollektiver Orientierungen¹

1. Vorstellung der Gruppendiskussion

Als Gruppendiskussion wird eine Sonderform der Gruppenbefragung durch einen (Sozial-)Forschenden verstanden, die darauf zielt, das gültige Relevanzsystem einer Gruppe auszubreiten (vgl. Schäffer 2006, 76).² Entsprechend soll sich in einer Gruppendiskussion (zumindest partiell) ein sogenannter selbstläufiger Diskurs der Teilnehmenden entwickeln, „in dem die Gruppe sich (...) ihrer kollektiven Erfahrungen in Erzählungen und Beschreibungen versichert“ (ebd.). Aus diesen könne wiederum die kollektive Wahrnehmungsperspektive von Gegebenheiten und somit für alle Teilnehmerinnen und Teilnehmer der Gruppe geltende Relevanzsetzungen rekonstruiert werden. Der „Gegenstand“ von Gruppendiskussionen sind folglich „kollektive Orientierungen und Wissensbestände“ (Przyborski/Wohlrab-Sahar 2014, 93). Das Verfahren findet vornehmlich Verwendung in der Jugend-, Milieu- und Geschlechterforschung sowie in der Wissens-, Professions- und Evaluationsforschung (vgl. ebd., 93ff.).

Die Entstehungskontexte der Erhebungsform sind einerseits im Bereich der Marktforschung, andererseits in der Sozialforschung zu sehen (vgl. Bohnsack

¹ Mein Beitrag widmet sich dem Erhebungsinstrument Gruppendiskussion somit in einem engen Rahmen – er bezieht sich nur auf Gespräche, die zum genannten Zweck generiert werden und deren Erhebung und Auswertung im Rahmen dokumentarischer Methodologie stattfindet (s. Schieferdecker in diesem Band). Diese Schwerpunktsetzung erschien mir nötig, um dem Anspruch nach Übersichtlichkeit zu genügen, die der vorliegende Band in sich trägt. Zu Zielsetzungen, methodologischen Grundlegungen und praktischen Durchführung von Gruppendiskussionen in weiteren/anderen methodologischen Kontexten vgl. z.B. Loos/Schäffer (2001); Lamnek (2005) sowie Kühn/Koschel (2011).

² Begründung dieses Ansinnens ist die Überzeugung, dass eine soziale Gruppe gemeinhin über eine spezifische sowie verbindende ‚Weltanschauungsweise‘ verfügt, die sie von anderen unterscheidet (vgl. Scherf 2013, 46ff.; Nutzung des Begriffs ‚Weltanschauung‘ nach Mannheim [1964]). So verfügten Angehörige einer Gruppe über gemeinsame Erfahrungen, deren Wahrnehmungsweise man im Blick auf diesbezüglich getätigte gemeinsame Erzählungen rekonstruieren könne (vgl. Schäffer 2006, 76).

2014, 107). Letztere beschäftigt sich seit den 80er Jahren verstärkt mit Gruppendiskussionen. Sie beruft sich hierbei auf grundlagentheoretische Konzepte Karl Mannheims (vgl. z.B. Bohnsack 1989). Die Konzeption der *dokumentarischen Methode* erfolgte in Auseinandersetzung mit dem Gruppendiskussionsverfahren, weshalb sie als methodologische Verortung für die Erstellung und weitergehende Bearbeitung von Gruppengesprächsdaten als herausragend geeignet gilt (s. Schieferdecker in diesem Band).

1.1 Ausgangsdaten und Ergebnisse

Der Wissenssoziologe Karl Mannheim gelang zur Erkenntnis, dass soziale Gemeinschaften über handlungspraktisch sehr bedeutendes Wissen verfügen, das den Mitgliedern der Gruppe aber verborgen bleibt – die gemeinsam Handelnden „wissen im Grunde gar nicht, was sie alles wissen“ (ebd., 281f.). So seien bestimmte Wahrnehmungsweisen, Konventionen und Muster des Deutens von Situationen den Akteuren nicht präsent, sie stellten vielmehr die *selbstverständlichen Orientierungen des miteinander Handelns* (und Kommunizierens) dar. Solcherart geteilt vorliegendes Wissen wird nicht zum Gegenstand von Gesprächen, es formt vielmehr den Rahmen, in dem thematische Diskurse stattfinden (und kann aus solchen rekonstruiert werden). Mannheim prägte hierfür den Begriff des ‚atheoretischen Wissens‘, welches als inkorporiertes Wissen „das Handeln relativ unabhängig vom subjektiv gemeinten Sinn orientiert“ (Bohnsack et al. 2007, 11/vgl. Mannheim 1964).³

Gruppendiskussionsdaten werden erhoben und ausgewertet, um die atheoretischen Wissensbestände einer Gruppe – inkludiert sind hierin auch (eher) unbewusst vorliegende Werthaltungen – zu extrapolieren, die ihren konkreten Handlungen zugrunde liegen. Diese werden wiederum gemeinhin mit *kollektiven Orientierungen* des Handelns benannt. In der Analyse von Gruppendiskussionsdaten findet somit eine Akzentverschiebung vom „Was“ hin zum „Wie“ statt: Der Inhalt („Was wird gesagt?“) tritt in den Hintergrund, wodurch sogenannte Orientierungsmuster („Wie werden in einer Gruppe Themen verhandelt? Welcher Sinngehalt kann die Grundlage jener kollektiv hervorgebrachten Äußerungen sein?“) in den Blick genommen werden können (vgl. Przyborski/Wohrlab-Sahr 2014, 277ff.).

Um die Rekonstruktion atheoretischen Wissens vornehmen zu können, müssen allerdings zuerst einmal Daten geschaffen werden, die eine solche Tätigkeit zulassen – in ihnen müssen sich Anhaltspunkte dafür ausmachen lassen, dass sich die Mitglieder des Handlungszusammenhangs zumindest teilweise *unmittelbar*

³ Ich möchte diesen Umstand mit einem Beispiel aus meiner Interpretationsarbeit illustrieren: So wurde die Haltung einer von mir beforschten Lehrergruppe, als professionelle Lehrer für den eigenen Unterricht alleinverantwortlich zu sein und Inhalte und Vorgehensweisen folglich nicht mit den Kollegen abstimmen zu müssen, im Diskurs nicht expliziert. Sie ließ sich vielmehr anhand der getätigten Handlungserzählungen rekonstruieren. Allen Erzählungen liegt genannte Haltung schließlich selbstverständlich zugrunde (vgl. Scherf 2013, 51).

verstehen, d.h., in „erlebnismäßigen Gemeinsamkeiten“ (ebd., 92) wiederfinden und auf Grundlage kollektiver Sichtweisen ihr Gespräch begehen. Solche Daten sollen mit der Erhebung von Gruppendiskussionen geschaffen werden.

1.2 Funktionsweise der Gruppendiskussion

Der methodologischen Bedeutung des Gruppendiskussionsverfahrens wurde von Seiten der Marktforschung keine Rechnung getragen; primär galten Gruppendiskussionen in dieser Disziplin als durchführens-wert, da sie Interviews (zeit-)ökonomisch überlegen sind: „Mehrere Interviews soll[t]en zugleich erreicht werden“ (Bohnsack 2014, 107). Aus sozialforscherischer Perspektive kann von einer Gruppendiskussion aber nur gesprochen werden, wenn „die methodologische Bedeutung von Interaktions-, Diskurs- und Gruppenprozessen für die Konstitution von Meinungen, Orientierungs- und Bedeutungsmustern in einem zugrunde liegenden theoretischen Modell [...] verankert sind“ (ebd.) und somit die Entstehungsform der Daten bei ihrer weitergehenden Prozessierung und Analyse Beachtung findet.

Wird der Gegenstand der Erhebung konsequent kollektiv konzipiert – was beim Untersuchungsgegenstand *kollektive Orientierungen* der Fall ist – muss bei dessen Erhebung sichergestellt werden, dass dieser auch gemeinschaftlich hervorgebracht wird. Dies ist nach Überzeugung der dokumentarischen Methodologie im Besonderen dann der Fall, wenn die Untersuchten einen sog. *selbstläufigen*, also einen durch sie selbst strukturierten Diskurs vollziehen. Die bedeutendste Aufgabe eines als Initiator einer Gruppendiskussion fungierenden Forschenden ist es somit, ein engagiertes Gespräch *zwischen* den Teilnehmenden in Gang zu bringen (vgl. Przyborski/Wohlrab-Sahr 2014, 96).

Ein selbstläufiger Diskurs kann nun recht einfach in einer realen Gruppe erzeugt werden. Für solche Gemeinschaften kann schließlich davon ausgegangen werden, dass „sie durch existentielle Gemeinsamkeiten zusammengehalten werden“ (ebd., 95), was sich in der Lebendigkeit der Diskussion niederschlägt. Zu erwarten ist somit ergiebige Datenmaterial auf Grund selbst strukturierter, interaktiv dichter Diskurse.

Welche Berührungspunkte die beforschte Gruppe formt, muss vor der Diskussion nicht zwingend bekannt sein – auf Grundlage der herzustellenden Daten kann schließlich ausgemacht werden, welche Gemeinsamkeiten ihr zugrunde liegen. Welche kollektiven Orientierungen für eine beforschte Gruppe als typisch gelten können bzw. welche sie im Besonderen aufweist, kann folgend insbesondere durch komparative Analyse herausgearbeitet werden. Durch den Vergleich „mit angrenzenden Gruppen, Zusammenhängen oder Phänomenen lassen sich die unterschiedlichen Erfahrungsräume, die konstitutiv für soziale Zusammenhänge [...] sind, voneinander lösen und man kann zu der je spezifischen Überlagerung von Erfahrungsräumen vordringen, die bestimmte Gruppen konstituieren“ (ebd., 95f.).

Diskussionsgruppen können aber auch durch den Forschenden und damit vor dem Hintergrund seines spezifischen Interesses zusammengestellt werden. Um aussagekräftiges Material gewinnen zu können, ist allerdings eine *gemeinsame Erfahrungsbasis* der Diskussionsteilnehmenden Voraussetzung – wobei „eine gemeinsame Erfahrungsbasis nicht voraussetzt, dass diese Erfahrungen von den anwesenden Personen auch tatsächlich gemeinsam gemacht wurden [...]: Ausschlaggebend ist die Strukturidentität der Erfahrungen, ihre Homologie, nicht aber, dass diejenigen, die durch derartige Erfahrungen miteinander verbunden sind, einander tatsächlich kennen oder diese Erfahrungen an einem identischen Ort zu einem identischen Zeitpunkt gemacht haben.“ (ebd., 96)

Gemeinsamkeiten können z.B. in Form einer gleichartigen Sozialisationsgeschichte, ähnlichen erlittenen biographischen Erfahrungen (schwere Erkrankungen, Flucht) oder einer vergleichbaren Berufsausübung vorliegen. Gelingt die Auswahl von Teilnehmenden dahingehend, dass in dem vom Forschenden erwarteten Bereich Gemeinsamkeiten vorliegen, dann sollte die Gruppe auch zahlreiche Passagen des Diskurses diesen widmen (vgl. ebd.).

Für Gruppendiskussionen wurden „in dichter Auseinandersetzung mit [...] der Rekonstruktion der Durchführung von Gruppendiskussionen“ (ebd., 97) insgesamt acht Prinzipien der „Initiierung und Leitung von Gruppendiskussionen“ (Bohnsack 2014, 225) entwickelt, von denen im Folgenden die m. E. für erfolgreiche Datengenerierung wichtigsten vorgestellt werden.

Bedeutendstes Gesprächsprinzip erscheint zunächst, forscherseitige *Interventionen immer an die ganze Gruppe zu richten* (vgl. ebd., 226). Statt also einem einzelnen Sprecher beispielsweise mit der Frage „Was meinst du dazu?“ (Przyborski/Wohlrab-Sahr 2014, 97) ein gesondertes Rederecht einzuräumen, sind alle Sprechenden gleichermaßen zu Wortäußerungen zu animieren (Beispiel: „Sie alle haben ja nun mit dem Thema ... zu tun. Wie sehen Sie denn...?“). Durch ein solches Vorgehen wird einerseits sichergestellt, dass das *Rederecht* durch die Gruppe und somit wie in der Gruppe üblich und nicht durch den Diskussionsleiter verteilt wird, woraus bereits auf bestimmte kollektive Orientierungen geschlossen werden kann (vgl. hierzu z.B. Scherf 2013, 342ff.). Andererseits erleichtern solche Impulse der Gruppe das Finden existentieller Gemeinsamkeiten (vgl. Przyborski/Wohlrab-Sahr 2014, 97). Impulse des Diskussionsleiters müssen nicht sprachlicher Natur sein – gerade zu Beginn der Diskussion können nonverbale Äußerungen wie „ermunternde Blicke“ (ebd., 100) hilfreich sein, um die Kommunikation der Teilnehmenden in Gang zu bringen.

Ein zentraler Impuls stellt der sog. „Eingangsstimulus einer Gruppendiskussion“ (ebd.) dar. Mit diesem wird gemeinhin das Gespräch initiiert und das (erste) Thema der Diskussion forscherseitig gesetzt. Vermieden werden sollte, mit der Themensetzung sog. „Propositionen“ (Bohnsack 2014, 226) aufzurufen, also Vorgaben dahingehend, innerhalb welcher Orientierungen das Thema zu bearbeiten ist. Generell sollten folglich *Themenvorschläge* des Diskussionsleiters *ohne themenbezogene Orientierungsrahmen* geäußert werden – durch die Themensetzung

sollte sich „lediglich das Interesse an der Entfaltung des Themas durch die Teilnehmer dokumentieren“ (Przyborski/Wohlrab-Sah 2014, 99), nicht die eigenen, an Erfahrungen rückgebundenen und zwangsläufig wertenden Orientierungen: „Man fragt also z.B. nicht: ‚Familie ist doch eine schöne Sache, oder?‘, sondern eher: ‚Welche Erfahrungen habt ihr denn mit eurer Familie gemacht, wie ist es denn so mit Familie?‘“ (ebd.).

Fragestellungen und Impulse seitens des Diskussionsleiters sind im Weiteren *betont vage zu halten*. Mit der von Forscherseite somit gewissermaßen „demonstrierten Vagheit“ (Bohnsack 2014, 226) soll den Teilnehmerinnen und Teilnehmern signalisiert werden, dass der Forschende eben nicht über präzises (Vor-) Wissen hinsichtlich der „milieuspezifischen Orientierungen der Gruppe“ (ebd.) verfügt. Damit wird nun einerseits „der Respekt gegenüber dem Relevanzsystem, der Erfahrungswelt der Beforschten“ (ebd., 227) bekundet, andererseits wird den Teilnehmerinnen und Teilnehmern die Rolle des „Experten für ihre Belange“ (Przyborski/Wohlrab-Sah 2014, 99) zugewiesen, die sie ausführlich vor dem milieufremden Forschenden ausbreiten können – *detaillierte Darstellungen* sollten die Folge sein.

Während der gesamten Diskussion sollte der Diskussionsleiter folglich äußerst *zurückhaltend agieren* – was durch die vorgenannten Prinzipien bereits deutlich geworden sein dürfte. Nur, wenn der gruppenfremde Forschende nicht bzw. kaum zum Gespräch beiträgt, können sich schließlich selbstläufige Diskurse einstellen, können die Diskussionsteilnehmerinnen und -teilnehmer Themen inklusive der den thematischen Diskursen zugrunde liegenden Orientierungen aufwerfen, selbst gewählt wieder beenden, durch ein weiteres Thema ergänzen oder ablösen. Um sicherzustellen, dass Gruppen in Diskussionen ‚bei sich‘ bleiben, verlassen manche Forschende nach dem Eingangsstimulus sogar die Diskussion und überlassen die Diskutanten sich selbst (vgl. hierzu z.B. Schmelz 2009).

Gruppendiskussionen werden zumeist in Form von Audio- oder Videodaten aufgezeichnet. Aufgrund der in der Wissenschaftswelt geltenden „Primordialität des Textuellen und Schriftlichen“ (Langer 2010, 515) werden für die weitergehende Analyse der Diskurse allerdings häufig in Schriftsprache verfügbare Daten verlangt, weshalb die Gespräche in Transkripte überführt werden. Das Transkribieren von Gruppengesprächen ist eine langwierige Tätigkeit, die m. E. allerdings als ein wichtiger Erkenntnisprozess bezüglich der Daten angesehen werden muss (vgl. auch Langer 2010, 517) – das intensive, wiederholte Hören bzw. Sichten der Daten zum Zwecke der Verschriftlichung sensibilisiert schließlich für das Gehörte bzw. Gesehene und kann zu ersten „Assoziationen, Hypothesen, Irritationen, Fragen“ (ebd.) führen, die als Interpretationsanregungen festzuhalten sind. Zudem kann eine Transkription nie eine mündliche Kommunikation vollständig in schriftliche Symbole fassen (vgl. Fuhs 2007, 84), weshalb im Prozess des Verschriftlichens beständig interpretative Entscheidungen gefällt werden.

Schon bevor mit der eigentlichen Transkription begonnen wird, gilt es, sich bzgl. des Transkriptionsumfangs sowie der zu befolgenden Transkriptionsregeln Klar-

heit zu verschaffen. Die dokumentarische Methodologie verlangt nicht, die vorliegenden Diskussionsdaten in Gänze zu transkribieren; vielmehr können auf Grundlage eines vom Forschenden erstellten Kurzprotokolls, des sog. „thematische[n] Verlaufs“ einer Diskussion (Przyborski/Wohlrab-Sahr 2014, 292), bestimmte „Passagen“ (ebd.) zur Transkription ausgewählt werden. Als Auswahlkriterien gelten z.B. deren inhaltliche Relevanz sowie die interaktive oder metaphorische Dichte des Gesprächs (vgl. ebd.; Bohnsack 2014, 125). In intensiven, bildreichen Gesprächspassagen mit raschen Sprecherwechseln kämen schließlich „jene kollektiven Orientierungen, welche im *Fokus* der Gruppe bzw. des Milieus stehen, in besonders prägnanter und/oder elaborierter Weise zum Ausdruck“ (ebd., Hervorhebung im Original).

Entscheidungen bezüglich des Transkriptionsumfangs, aber auch der Art und Weise der Transkription, sind mit Bedacht zu treffen. Die Transkripte sind letzten Endes die Produkte, mit denen in allen auf die Transkription folgenden Forschungsschritten weitergearbeitet wird, die ursprünglichen Audio-/Videodaten werden nur selten zur Kontextualisierung in den Interpretationsprozess einbezogen (vgl. Fuhs 2007, 84). So kann beispielsweise ein (zu) detailliertes, sehr ‚feinkörniges‘ Transkript die Analyse des Diskurses sogar erschweren, wenn die Menge der (ggf. überflüssigen) Information die Sicht auf Wesentliches behindert. Wird andererseits etwas für die Auswertung Entscheidendes nicht im Transkript vermerkt, ist es für den Auswertungsprozess „gewissermaßen nicht mehr existent“ (Langer 2010, 516) und mindert seine Qualität. Liegen Videodaten vor, können Transkripte auch mit Fotogrammen versehen werden – Bewegungen der Diskurspartner im Raum sowie deren Gestik, Mimik und ggf. genutzte Gegenstände sind auf diese Weise gut ins Transkript zu integrieren (vgl. hierzu z.B. Asbrand et al. 2013).

Auf die Transkription von Gruppendiskursen folgt gemeinhin deren dokumentarische Interpretation, die auf die Rekonstruktion kollektiver Orientierungen zielt.⁴ Diese erfolgt sequenziell und in (mindestens) drei Schritten: Zunächst wird in der *formulierenden Interpretation* der betrachtete Diskursabschnitt durch den Forschenden paraphrasiert, um seinen kommunikativen Sinngehalt zu erschließen. Die Frage, die der Forschende hierbei „zu beantworten sucht, lautet: ‚Was wird gesagt?‘“ (Przyborski/Wohlrab-Sahr 2014, 293). In der anschließenden *reflektierenden Interpretation* widmet sich der Forschende der Rekonstruktion der kollektiven Wissensbestände, die dem Diskurs zugrunde liegen. Der Forschende geht hierbei der Frage nach: „[W]elcher Sinngehalt kann die Grundlage der konkreten Äußerung sein?“ (ebd., 295). Im Ver- und Abgleich mit weiteren Diskurssequenzen der gleichen oder anderen Gruppen widmet sich der Forschende dann nach und nach der *Typenbildung*. Diese zielt – verkürzt – darauf, bestimmte Orientierungsgehalte als für bestimmte Gruppen spezifische auszumachen (s. Schieferdecker in diesem Band).

⁴ Zum Umstand, dass Gruppendiskussionsdaten auch in ganz anderen methodologischen Kontexten Verwendung finden können, vgl. Fußnote 1.

1.3 Vorteile und Nachteile der Gruppendiskussion

Gruppendiskussionen eignen sich im Besonderen, um kollektive Orientierungen einer Gruppe und damit deren implizite, sozial geteilt vorliegenden Wissensbestände zu explorieren. Beruft man sich auf das grundlagentheoretische Konzept Karl Mannheims, kann man davon ausgehen, dass diese nicht erst im Diskurs *entstehen*, sondern durch diesen lediglich *repräsentiert* werden – in diesem Fall bewegt sich der Forschende im Auswertungsprozess von Gruppendiskussionen tatsächlich auf für die untersuchte Gruppe geltende Weltanschauungsweisen zu. Gerade für Gruppen, die vom Forschenden zusammengestellt werden oder sich in ihrer üblichen verbindenden Lebenswelt wenig (verbal) miteinander verständigen, gilt es allerdings, den Konstrukt-Charakter der untersuchten Kommunikation bei der Datenauswertung zu berücksichtigen.⁵

Das Verfahren eignet sich prinzipiell nicht zur Bearbeitung von Fragen, „bei denen Individuen die zu untersuchende Einheit darstellen“ (Przyborski/Wohlrab-Sahr 2014, 93). Individuell vorliegendes, biographisch erworbenes Wissen, das nur einzelne Gruppenmitglieder aufweisen, kommt in Gruppendiskussionen z.B. häufig gar nicht erst zur Sprache; wird auf individuelle Handlungen verwiesen, geschieht dies vor dem Hintergrund des Diskussionsgeschehens.

Hat man im Blick, neben kollektiven Orientierungen individuell vorliegende Wissensstrukturen zu erheben bzw. zu rekonstruieren, bietet sich folglich die *triangulierende* Betrachtung verschiedener Verbaldaten an: In Interviews kommt es gegenüber dem „milieufremden Interviewer“ (ebd., 279) z.B. zu detaillierten Schilderungen von Handlungspraktiken. Jene spielen in Gruppendiskussionen naturgemäß eine viel kleinere Rolle – da die Diskursteilnehmerinnen und -teilnehmer davon ausgehen, dass ihre Mitstreiterinnen und Mitstreiter die Praktiken, auf die rekurriert wird, teilen (vgl. Scherf 2013, 75). Auf zusätzliche Daten aus teilnehmenden Beobachtungen zurückzugreifen, kann sicherstellen, sowohl den kollektiven Habitus bestimmter Gruppen als auch den individuellen Habitus *und* die Handlungspraktiken der Mitglieder umfassend rekonstruieren zu können und somit valide wissenschaftlich abzubilden (vgl. Bohnsack et al. 1995; Przyborski/Wohlrab-Sahr 2014, 22ff., 278ff.). Ein solches Design, das die einander ergänzende Betrachtung all jener Daten vorsieht, weisen allerdings bislang recht wenige Studien auf (vgl. ebd.) – vermutlich auf Grund der hierfür erforderlichen Mittel.

⁵ Für eine Lehrerguppe, die sich zum Zweck der Gruppendiskussion eine Dreiviertelstunde miteinander unterhält, deren Mitglieder in ihrem sonstigen professionellen Alltag aber kaum miteinander sprechen, können die entstandenen Diskursdaten z.B. keineswegs als ‚natürliche‘ betrachtet werden (vgl. hierzu z.B. Scherf 2013, 159ff.).

1.4 Gütekriterien für den Einsatz der Gruppendiskussion

Bereits deutlich wurde, dass Gruppendiskussionen genutzt werden können, um kollektive Wissensbestände bestimmter Personengruppen in den Blick zu nehmen. Grundsätzlich gilt es folglich zu prüfen (und ist für die Geltung der Erkenntnisse entscheidend), ob mit ihrem Einsatz dieses Ziel auch verfolgt wird.

Weiterhin gilt die Zusammensetzung einer Diskussionsgruppe als wichtiger Parameter, um die Güte der entstehenden Daten zu beeinflussen: „Dort, wo die Teilnehmer gemeinsame Erfahrungen haben, wo sich Homologien in der Erlebnisaufschichtung finden, werden sich Zentren der Diskussion ergeben. Für diese Bereiche lassen sich kollektive Orientierungen herausarbeiten“ (ebd., 94). Umgekehrt lässt sich formulieren, dass nur dann, wenn eine Gruppe über gemeinsame oder strukturgleiche Erfahrungen verfügt, ein lebendiges Gespräch entstehen kann, aus dem auf kollektiv gültige Rahmungen bzw. „ein ähnliches Weltbild“ (ebd., 95) der Teilnehmerinnen und Teilnehmer geschlossen werden kann.

Auch das Agieren des Diskussionsleitenden kann die entstehenden Diskursdaten in ihrer Qualität maßgeblich beeinflussen. Zur Rekonstruktion kollektiver Orientierungen einer beforschten Gruppe können streng genommen schließlich nur Diskurssequenzen herangezogen werden, die diese komplett selbst verantworten – was erfordert, dass die gruppenfremde Initiatorin und Moderatorin bzw. der gruppenfremde Initiator und Moderator des Gesprächs ausgesprochen zurückhaltend agiert sowie auf eine Teilnehmerrolle weitgehend verzichtet. Zu prüfen, ob dies gelang, ist insofern auch der erste, m. E. unverzichtbare Schritt der Datenauswertung.

Laut Przyborski und Wahlrab-Sahr (2014) entscheidet nicht zuletzt die Länge des Diskurses über die Nutzbarkeit der Daten: „Sobald man zwischen 30 und 60 Minuten überwiegend selbstläufiges Gespräch ‚im Kasten‘ hat, kann man (fast) nichts mehr falsch machen“ (ebd., 100). Begründet wird diese Haltung damit, dass es gewisse Zeit brauche, bis die (einander ggf. unbekannt) Teilnehmerinnen und Teilnehmer einander und die ungewohnte Erhebungssituation ‚abgetastet‘ hätten. Erst im Anschluss daran würden sich Diskurse ergeben, die die gewünschte Intensität und Interaktivität aufwiesen.

1.5 Literaturempfehlungen

Um sich bezüglich der Erhebung und adäquaten Nutzung von Gruppendiskussionsdaten weitergehend zu informieren, empfehle ich – was nahe liegt – die Werke, auf die sich die von mir vorgelegte Beschreibung des Verfahrens stützt.

Przyborski/Wahlrab-Sahr (2014): Qualitative Sozialforschung.

Das Arbeitsbuch enthält eine umfangreiche Darstellung der Erhebungsform Gruppendiskussion (vgl. ebd., 88ff.). Die Autorinnen geben in dieser etliche konkrete Anregungen für die praktische Durchführung. Darüber hinaus wird die zur Auswertung von Gruppendiskursen präferierte Dokumentarische Methode bezüglich ihrer Entstehungskontexte und methodologischen Grundlagen, aber auch in ihrer praktischen Anwendung vorgestellt (vgl. ebd., 277ff.).

Bohnsack (2014): Rekonstruktive Sozialforschung.

Das Werk widmet sich umfassend der – von ihm geprägten – Dokumentarischen Methode (vgl. ebd., 33ff.). Die *Erhebung* von Gruppendiskussionen begründet er mit den gleichen methodologischen Überlegungen wie deren (dokumentarische) Auswertung (vgl. ebd., 107ff., v.a. 115ff.). Besonders geeignet halte ich das Werk entsprechend, um sich über die sozialwissenschaftlichen Grundannahmen der Erhebung von Gruppendaten sowie deren deutenden Nutzung, sprich der Rekonstruktion von kollektiven Wissensbeständen, zu informieren.

Weitere Empfehlungen

Die Erstellung und Verwendung von Gruppendiskussionsdaten in deutschdidaktischen Lehrerforschungen erläutern Schmelz (2009) sowie Scherf (2013). Inwiefern Gruppendiskussionen überhaupt Verwendung in deutschdidaktischen Forschungen fanden bzw. finden können, erfassen schließlich Ballis et al. (2014) – allerdings eher am Rande: zentral widmen sie sich der Verwendung dokumentarischer Methodologie.

2. Beispiel für den Einsatz des Verfahrens: Leseförderung aus Lehrersicht (Scherf 2013)

2.1 Fragestellung, Studienanlage und Vorgehen

Auf welchen Wissensgrundlagen betreiben Lehrende Leseförderung? Der Beantwortung dieser Frage widmete sich meine Dissertationsstudie (vgl. Scherf 2013). Ausgehend von meinem Standpunkt als Lesedidaktiker war für mich (u.a.) von besonderem Interesse, inwiefern sich das Wissen zu Leseförderung im Kontext innovativer lesedidaktischer Wissensangebote verändert. Berechtigung erhält diese spezielle Perspektive auf Lehrerwissen durch das Selbstverständnis meiner wissenschaftlichen Disziplin: Die Lesedidaktik zielt mit ihren Konzeptionen auf die Veränderung von Lehrerwissen ebenso wie auf die Umgestaltung der schulischen Leseförderpraxis; sie hat den Anspruch, „praktische Wissenschaft“ (Ossner 1999, 37) zu sein, indem sie auf die Praxis – positiv verändernd – einwirkt.

Wie sich Lehrerwissen zu Leseförderung gestaltet und welche Wissensangebote auf welche Weise in Lehrerwissen zu Leseförderung eingehen, war bei Studienanlage allerdings kaum erforscht. Folgerichtig hatte man sich dem Forschungsgegenstand explorativ zu nähern (vgl. Flick et al. 2005, 17). Die auf Hypothesen (weitgehend) verzichtende rekonstruktive Ausrichtung der Studie ist folglich dem bisherigen Erkenntnisstand sowie dem Erkenntnisinteresse der Studie geschuldet; ihr Ziel war es zu erfassen, welches umfassende Wissen Lehrende zu Leseförderung (tatsächlich) aufweisen, welche Wissensbestände im Handeln einen Niederschlag finden und wie sich das Wissen (und Handeln) durch jüngst entstandene lesedidaktische Wissensangebote verändert.

Zunächst war geplant, zur Bearbeitung genannter Forschungsfrage eine Erhebung durchzuführen, in der Lehrerwissen als *individuelles* (didaktisches) *Wissen* gefasst wird (vgl. z.B. Kunze 2004, 78ff.); die Planung lehnte sich an kognitionspsychologisch orientierte Erhebungen von (Experten-)Wissen an (vgl. z.B. Groeben et al. 1988; Bromme 1992). Ein besonderes Augenmerk wurde schon bei der Voruntersuchung auf die mögliche (berufs-)biographische Genese von Leseförderwissen gelegt, da sich Hinweise auf die biographische Bedingtheit von handlungsrelevantem Wissen aus bisherigen Forschungsarbeiten ableiten ließen (vgl. z.B. Wieser 2008, 231).

Bei der Interpretation der Vorstudienaten wurde dann deutlich, dass Lehrerwissen zu Leseförderung tatsächlich in Abhängigkeit zur gewählten Profession und zur eigenen (Berufs-)Biographie konstruiert wird. Es wurde aber auch deutlich, dass es bezüglich des Leseförderhandelns offenbar eine umfassende *kollektive* Wissensbasis innerhalb einer *Lehrergruppe* gibt – Verweise auf Förderbedingungen an der eigenen Schule, auf bestimmte Sichtweisen, die unter *allen* Kollegen der eigenen Schule herrschten, häufige Begründungen, die ein kollektivierendes ‚wir‘ oder ‚man‘ enthielten etc., ließen dies vermuten. Eine alleinige Betrachtung von Lehrerwissen als ‚individuell‘ würde jenen Umstand missachten und erschien folglich unangemessen. Wollte man Lehrerwissen zu Leseförderung umfassend

wissenschaftlich abbilden, galt es, auch die kollektiv vorliegenden Wissensbestände in den Blick zu nehmen.

Um *kollektive Orientierungen* bezüglich Leseförderung zu rekonstruieren, mussten nun Diskurse von (realen) Leseförder-Gruppen untersucht werden. Da jedoch in der Voruntersuchung auch deutlich wurde, dass biographisch induzierte Wissensbestände der einzelnen Informantinnen und Informanten⁶ – die sich innerhalb einer Gruppe von Leseförderlehrerinnen und -lehrern voneinander unterscheiden müssen – ebenso für das Leseförderhandeln bedeutsam sind, wurde infolge der Voruntersuchung ein fallbezogenes Erhebungsverfahren entwickelt, das zwei verschiedene Erhebungsformen berücksichtigt: Die Erhebung von Gruppendiskussionen mit allen Informantinnen und Informanten einer Schulgruppe sollte auf die Emergenz von kollektiven Orientierungen zielen, teilnarrative Einzelinterviews zum Thema Leseförderung mit jeder Lehrerin bzw. jedem Lehrer alleine sollten dagegen die Darstellung von individuell vorliegenden Wissensbeständen evozieren. Erst in der Triangulation von Daten und Perspektiven sollten fallbezogene Erkenntnisse generiert werden (vgl. Scherf 2013, 62ff.)⁷.

Ergebnis der Studie ist u.a., dass sich Lehrerwissen zu Leseförderung als sehr komplexes, an Varietäten reiches Wissen unterschiedlichster Quellgebiete und „Resonanzböden“ (Nittel 1997, 31) darstellt, welches tatsächlich zu Teilen biographisch erworben individuell und zu Teilen in der sozialen Arena eines Kollegiums geteilt vorliegt. Zudem zeigen sich gleichermaßen implizite wie verbalisierbare Wissensbestände der Beforschten als für ihr unterrichtliches Handeln relevant. *Kollektive Orientierungen* zum professionellen Selbstverständnis erscheinen z.B. als die Wissensbestände, die besonders in den Umgang mit lesedidaktischen Innovationen hineinspielen.⁸

Bezüglich des methodischen Vorgehens zeigten sich sowohl die generierten Daten als auch deren Auswertungsweise als geeignet, um sich den disparaten sowie formal vielschichtigen Lehrerwissensbeständen zu Leseförderung zu nähern. Die Vermutung, dass sich das Leseförderhandeln eines Lehrenden zwischen kollektiven Orientierungen einer sozialen Arena sowie individuellen Wissensbeständen aufspannt, bestätigte sich während der Hauptuntersuchung – sowohl Gruppendiskussionen als auch Interviews zu führen und folgend sozial geteilte *und* individuelle Wissensbestände zu rekonstruieren, erwies sich insofern als erkenntnisstiftendes Vorgehen für mein Forschungsanliegen.

⁶ Als Informantinnen und Informanten bezeichne ich in meiner Studie die Lehrerinnen und Lehrer, die bei der Datenaufnahme mitwirkten – und mir hierdurch ermöglichten, bzgl. ihrer Einstellungen und Wissensbestände zu Leseförderung Informationen zu erlangen.

⁷ Zum Problem, *individuell vorliegende Wissensbestände* im Rahmen des (kollektivierenden) Wissensbegriffs Mannheims zu verhandeln, vgl. Scherf 2013, 66.

⁸ Zu weiteren Erkenntnissen der Studie vgl. Scherf 2013, 411ff.; zu letztgenanntem Punkt vgl. Scherf (2016).

Exemplarisch konnte folglich für fachdidaktische Wissensforschungen aufgezeigt werden, was auf Grund der wenigen in einem solchen Setting durchgeführten sozialwissenschaftlichen Studien vermutet wurde: dass nämlich die Triangulation von Gruppendiskussions- und narrativen Interviewdaten eine sinnvolle Perspektiverweiterung darstellt, sich die erhobenen Daten auf geeignete Weise „ergänzen bzw. kompensieren“ (Przyborski/Wohlrab-Sahr 2014, 278). So wurde in den Gruppendiskussionen z.B. selten ausführlich thematisiert, wie ganz praktisch im Unterricht gehandelt wird. Dies schien entweder den Lehrerinnen und Lehrern nicht der Rede wert, weil man davon ausging, dies sei ohnehin allen bekannt, oder aber das eigene Agieren im Unterricht galt kollektiv nicht als angemessenes Gesprächsthema. Gegenüber dem Interviewenden, den die Informantinnen und Informanten als (mehr oder weniger) milieufremd auffassten, wurde das Unterrichtshandeln dagegen ausführlich expliziert. Persönliche, biographische Bedingungen des Leseförderhandelns brachen sich eher im Zwiegespräch mit dem Interviewenden Bahn; institutionelle und gesellschaftliche Bedingungen, denen man sich unter Leseförderlehrerinnen und -lehrern gemeinschaftlich ausgesetzt sah, bestimmten dagegen die Diskurse innerhalb der Teilkollegien. Letztlich konnten also mit den gewählten Diskursformen wie geplant eher individuell bzw. eher kollektiv vorliegende Wissensbestände erhoben werden; und letztlich zeigten sich die jeweils erhobenen Wissensbestände gleichermaßen als bedeutend für das Lehrerhandeln. (Vgl. Scherf 2013, 427f.)

2.2 Erhebung und Auswertung von Gruppendiskussionen: Beispielsequenzen

Im Rahmen meiner Studie wurden insgesamt fünf Fälle betrachtet, wobei als ‚Fall‘ jeweils das Wissen zu Leseförderung eines Teilkollegiums gefasst wurde. Die Teilkollegien der Albert-Einstein-Schule (AES) sowie der Christian-Morgenstern-Schule (CMS) wurden als sog. ‚Eckfälle‘ ausgewählt und erfuhren sowohl während der Auswertung als auch im Ergebnisbericht gesonderte Beachtung.⁹

Mit den Informanten aller fünf Teilkollegien wurden Gruppendiskussionen durchgeführt, die durch die Nutzung gleichartiger thematischer Impulse einen vergleichbaren Verlauf nehmen sollten (vgl. Scherf 2013, 466). Der erste Impuls zielte jeweils darauf zu erfahren, was die beforschte Gruppe ausmacht – erwünscht wurden Erzählungen zur Gruppenzusammensetzung, aber auch zur Schule und deren Besonderheiten. Erst nachfolgend wurde ein Impuls gesetzt, der sich auf Leseförderung bezog („Leseförderung an der ...-Schule“: Was gibt es dazu zu erzählen?“). Mit allen Beforschten wurden zudem Interviews geführt.

Um zu zeigen, wie die von mir generierten Diskussionsdaten sich gestalten und welche Schlüsse aus ihnen im Rahmen der dokumentarischen Auswertung gezogen wurden, möchte ich im Folgenden jeweils eine Sequenz der Fälle AES und CMS sowie skizzenhaft deren Deutung präsentieren. Beide Sequenzen finden sich

⁹ Zu den Falldarstellungen vgl. Scherf 2013, 133ff. und 233ff. Die Schul- wie Klarnamen der Beforschten wurden anonymisiert.

zu Beginn der jeweiligen Diskussion; sie folgen auf den erstgenannten Impuls, etwas zu sich bzw. der eigenen Schule zu erzählen.

In der Informantengruppe AES entspannt sich kurz nach der Frage des Interviewers (I), wie die Informanten ihre Schule sähen, folgender Diskurs:

107. Aiw (...) Ähm, die Schüler sind in
 108. den letzten Jahren insgesamt einfach schwieriger]
 109. Tim [Mh.]
 110. Aiw [geworden. Äh, das]
 111. Naw [Wir haben
 112. eine Negativauswahl.]
 113. Aiw [So empfinde ich das. Eine Negativauswahl.]
 114. Tim [Ja, das stimmt.]
 115. Aiw [ähm, (.) weil die
 116. Gesellschaft das Prinzip Integrierte Gesamtschule (.) zwar von Finnland und den
 117. nordischen Ländern in den Himmel hebt, aber hier keiner drauf springt und auf
 118. diesem Pferd reitet und das ist das Problem.
 119. I Mh. Was heißt denn Negativauswahl? Also, ich meine, es gibt ja eine
 120. Negativauswahl und eine Negativ-Negativ-Negativauswahl. Wie seht ihr das?
 121. Naw Also wir, ähm, leiden darunter, dass wir in direkter Nachbarschaft, äh, Gymnasien
 122. haben, also zwei eigentlich. Ne, im R. gibt es zwei (.) Gesamtschulen, die
 123. Gymnasialklassen ab Fünf anbieten]
 124. I [Mh.]
 125. Naw [und, ähm, die graben uns
 126. natürlich das Wasser ab. Und das, was du sagst hier von wegen, äh, Integrierte
 127. Gesamtschule und Ganztagschule und alles wird in den Himmel gehoben. Ähm, ja,
 128. alle schwärmen davon]
 129. Aiw [Ja.]
 130. Naw [aber keiner]
 131. Aiw [keiner macht es]
 132. Naw [schickt seine Kinder hin, also vor allem
 133. Mh.]
 134. Naw [nicht die, ähm, die etwas, ähm, potenteren Kinder.]
 135. I [Mh.]
 136. Naw [Also wenn die
 137. Kinder in der Grundschule gute Noten haben, dann werden die auf Deibel komm
 138. raus ins Gymnasium geschickt (.) und nicht zu uns. Viel zu wenige, wir leiden
 139. darunter, dass bei uns die Spitze wirklich abgesägt ist.
 140. Aiw Mh.
 141. Tim Mh.
 142. Naw Ne, wir haben SEHR, sehr wenige Kinder, die, ähm, so, ähm, (.) aus]
 143. Aiw [richtiges
 144. Gymnasialniveau]
 145. Naw [richtiges Gymnasialniveau haben.]
 146. Tim [Mh.]

Zunächst zum Verlauf: In dieser Gesprächsphase sind erstmals alle Informantinnen und Informanten beteiligt. Das Thema wird von den Informantinnen und Informanten gesetzt (der Interviewer fragt immanent nach, vgl. Z. 120); der Diskurs entspinnt sich weitgehend innerhalb der Informantengruppe. Auch auf Grund der interaktiven und metaphorischen Dichte kann die Sequenz m. E. zur Auswertung herangezogen werden – in ihr sollten gültige Orientierungen der Gruppe zum Ausdruck kommen.

Welche Deutungen die Sequenz ermöglichte, kann hier nur skizziert werden (s. Schieferdecker in diesem Band): Als bedeutende wie problematische Rahmenbedingung von Lehrerarbeit an der AES wird hier die Schülerklientel benannt. Zwar werde von allen das Prinzip der Gesamtschule „in den Himmel gehoben“ (AES, 127), aber man schicke seine Kinder, sofern sie gute Noten hätten, auf das Gymnasium. Zu ihnen käme eine „Negativauswahl“ (AES, 112). Das Fehlen leistungs- ‚potenter‘ Schüler wird von allen als Erschwernis der Unterrichtsarbeit gesehen.

Die Leistungsstärke und Verhaltensunauffälligkeit früherer Schülergenerationen bildet den positiven Gegenhorizont innerhalb der Orientierungsfigur ‚Schüler‘. (Dass frühere Schülergenerationen leistungsstärker gewesen seien, wird in anderen Sequenzen ausgeführt.) Obschon im Folgenden einige Male auf die Heterogenität der Schülerschaft verwiesen wird, erscheint die Klientel den Informantinnen und Informanten bezogen auf ihre ‚Beschulbarkeit‘ *homogen* – die ‚Schüler heute‘ sind leistungsschwach und im Verhalten problematisch. Die gesamte Orientierungsfigur besitzt den Charakter einer zweckrationalen ‚Entschuldigung‘ des – im weiteren Verlauf der Diskussion so bewerteten – erfolglosen Arbeitens: Erfolgreiches Belehren der ‚Schüler heute‘ erscheint unter diesen Umständen kaum möglich.

Dass paradoxerweise das Prinzip der gemeinschaftlichen Beschulung im gesamtgesellschaftlichen Diskurs eine positive Wertung erfährt, lässt die empfundene Lehrerarbeits-Wirklichkeit der eigenen Gesamtschule umso frustrierender erscheinen: Der generellen Hochschätzung der Schulform zum Trotz sei man nicht Gesamtklientel-, sondern ‚Negativauswahl-Lehrer‘ und könne vor diesem Hintergrund nicht das leisten, was geleistet werden soll: Man habe schlicht „keine Chance“ (AES, Z. 502) und leide hierunter (vgl. Scherf 2013, 159ff.).

Ein gänzlich anderer Diskurs führt die Informantengruppe der CMS. Auf einen ähnlichen Impuls folgt diese Sequenz:

147. I (...) (2) Ähm, (3) generell, also die, die Schule, an der Sie alle sind, diese ARS,
 148. ähm (.), vielleicht auch die Deutschkollegen, was kann man denn über die so
 149. erzählen, was kann man über die (.) CMS sagen, wie empfinden Sie die?
 150. (6)
 151. Kam Also, ich denke, im Deutschbereich zumindest in den Jahrgangsstufen Fünf,
 152. Sechs, (.) glaube ich schon, dass das eine relativ gute Zusammenarbeit ist, das
 153. kann man sagen. (.) Also auch ein guter Austausch. Also, in den höheren
 154. Jahrgangsstufen (.) nimmt das immer mehr ab. Aber (.), ähm, wir (.) bereiten ja
 155. auch gemeinsam Klassenarbeiten vor, sechs, sieben, ja. (.) Also, (.), ähm, in den
 156. Unterstufen ist es wirklich, ähm, gesamtschulmäßig, sage ich mal, in der
 157. Oberstufe geht, geht das dann über so allmählich in das Gymnasiale und da ist das
 158. Einzelkämpfertum (.) etwas stärker verbreitet, zumindest im Fach Deutsch.
 159. How Aber ich erlebe generell einen wirklich guten Austausch. Also, wenn ich das auch
 160. damit vergleiche, woher ich komme,]
 161. Kam [Mh.]
 162. How [war es nicht (.) unbedingt so. Ja, also
 163. die Kommunikation ist schon eine andere.
 164. Kam [Mh.]
 165. Kem [Mh.]
 166. How Und ich bin ja nun ein Frischling.]
 167. Kam [((lacht))]

168. How [Also ich wurde hier sehr gut aufgenommen,
 169. Kem und habe, war also gleich drin (.) sozusagen.]
 170. [JA DIE MÖGEN DICH NOCH. ((ironisch))]
 171. How ((alle lachen laut))
 172. Ja, ne? Was habe ich nicht alles getan! ((lacht)) Genau, ja, also das ist jetzt, das ist eindeutig positiv zu sagen. (...)

Zum Verlauf: Bei der Auswertung dieser Sequenz sollte beachtet werden, wie sie begann – es ist die vom Interviewer angebotene Verengung der Ausgangsfrage, welche die Informantinnen und Informanten auf die Fachschaft Deutsch Bezug nehmen lässt. Zunächst wird auf die breite Kooperation in der Fachschaft verwiesen, welche hoch geschätzt wird. Sie wird der Schulform Gesamtschule sowie eher den unteren Jahrgängen zugeordnet und damit exemplifiziert, dass man Klassenarbeiten gemeinsam vorbereite und sich austausche. Die Ausführung einer Informantin, dass der Austausch an der CMS generell besser funktioniere als an ihrer alten Wirkungsstätte, wird für den Informanten Kem zur ‚Vorlage‘ für einen alle erheiternden Witz.

Zur Auswertung: Deutlich wird ausgeführt, dass man die innerschulische Kommunikation für besonders gut hält – so sehr, dass man über die Elaboration von How, sie sei schnell gut aufgenommen worden, gleich humoristisch deren noch nicht abgelaufene Phase der ‚Bewährung‘ thematisieren kann. Die kollektiv verstandene Ironie zeigt nun an, dass man sich innerhalb der Gruppe tatsächlich unmittelbar versteht – auch, was die Hierarchisierung der Gruppe angeht, in der How mit dem Bild des Frischlings (vgl. CMS, 113) ihren Lerner-Status unterstreicht.

Kollegiale Kommunikation wird wiederum unter zwei Blickwinkeln in die Sequenz eingebracht: Zum einen sieht man sie fachlich bzw. unterrichtsbezogen, als Form der Abläufe homogenisierenden fachbezogenen Zusammenarbeit, die man schätzt, z.B. im gemeinsamen Vorbereiten von Klassenarbeiten. Ihr wird das mit der Schulform Gymnasium verbundene fachliche Einzelkämpfertum, was an der CMS vor allem in der Oberstufe stärker vertreten sei, als negativer Horizont gegenübergestellt. Zusammenarbeit wird weiterhin eingebettet in den darüber hinausgehenden guten Austausch, der im Kontext quasi-privaten „sich Mögens“ weitergeführt wird.

Es kristallisiert sich in der Sequenz m. E. erstmals ein kollektiv gültiges Bild der eigenen *guten* und *erfolgreichen* Arbeit heraus, in das die Schulform Gesamtschule *und* die Einzelschule CMS ungebrochen positiv hineinspielen, während ein Verweis auf die problematische Klientel bzw. eine sich hierdurch ergebende schlechte Ausgangslage der Unterrichtsarbeit unterbleibt. Während die Informanten der AES den gymnasialen Schüler als ‚erwünschte‘ Vergleichsdimension heranzogen, wählen die der CMS das Gymnasium als zu vergleichende Schulform – um allerdings darauf zu verweisen, dass diese im Kontrast zu ihrer defizitär sei. Man sieht sich somit als kooperierende und damit professionelle Gesamtschullehrer – erfolgreich agierend in der ‚richtigen‘ Schulform.

Dass sich auf die Implementation von lesedidaktischen Innovationen auswirkt, wie die Teilkollegien ihre Arbeitsbedingungen und -erfolge einschätzen, ist nun

wenig verwunderlich, kann hier aber nicht mehr ausgeführt werden (vgl. Scherf 2016).

2.3 Einordnung des Beispiels, Ausblick

Im vorliegenden Beitrag sollte sowohl die Erhebung von Gruppendiskussionen als auch die Arbeit an in jenem Setting entstehenden Daten vorgestellt werden. Dezidiert sollte hierbei Bezug auf die Verwendung des Verfahrens in deutschdidaktischen Forschungskontexten genommen werden.

Deutlich geworden sein dürfte anhand der angeführten Beispiele, dass Gruppendiskussionen zu aufschlussreichen und reichhaltigen Erkenntnissen auch in dieser Disziplin beitragen können, sofern bei ihrer Erstellung und Interpretation deren methodologische Einbettung sowie Prinzipien der Durchführung beachtet werden. So unterscheiden sich die gewählten Beispiele z.B. sowohl bezüglich des kommunikativen Handelns, das der Diskussionsleiter an den Tag legt, als auch bezüglich der interaktiven Dichte – woraus forschenseitig Schlüsse dahingehend gezogen werden müssen, wie weit man im Vorgang der Interpretation zu gehen vermag.

Symptomatisch ist das gewählte Beispiel insofern, dass Gruppendiskussionen in der Deutschdidaktik bisher fast ausschließlich im Kontext der Lehrerforschung Verwendung fanden (vgl. Schmelz 2009, Scherf 2013). Auch die anderen an Deutschunterricht beteiligten Akteure – Schülerinnen und Schüler und deren Eltern, zu beschulende Erwachsene etc. – verfügen allerdings über kollektive Wissensbestände, denen im genannten Setting nachgegangen werden kann (vgl. zu Gruppendiskussionen mit Kindern z.B. Przyborski/Wohlrab-Sahr 2014, 102ff.)¹⁰. Diesbezüglich methodologische ‚Anpassungsarbeit‘ zu leisten erscheint wichtig – genauso wichtig, wie in jenen Feldern Erfahrungen zu sammeln und fachbezogene Expertise zu ermöglichen.

Literatur

- Asbrand, Barbara/Martens, Matthias/Petersen, Dorthe: Die Rolle der Dinge in schulischen Lehr-Lernprozessen. In: Zeitschrift für Erziehungswissenschaft (2013), 16, 1, 171-188.
- Ballis, Anja/Penzold, Michael/Scherf, Daniel/Schieferdecker, Ralf: Die dokumentarische Methode und ihr Potenzial für Forschungen (nicht nur) in der Fachdidaktik Deutsch. In: Didaktik Deutsch (2014), 19, 37, 92-104.
- Bohnsack, Ralf/Loos, Peter/Schäffer, Burkhard/Städtler, Klaus/Wild, Bodo: Die Suche nach Gemeinsamkeit und die Gewalt der Gruppe. Hooligans, Musikgruppen und andere Jugenddeliquen. Opladen: Leske+Budrich 1995.

¹⁰ Heizmann (2011) beginnt hiermit. In seiner Dissertationsstudie untersucht er Gesprächsdaten, die mit Grundschülergruppen generiert wurden, in Bezug auf deren Orientierungen beim Interpretieren von Literatur.

- Bohnsack, Ralf/Nentwig-Gesemann, Iris/Nohl, Arnd-Michael: Einleitung. Die dokumentarische Methode und ihre Forschungspraxis. In: Dies. (Hrsg.): Die dokumentarische Methode und ihre Forschungspraxis. Grundlagen qualitativer Sozialforschung. Wiesbaden: VS Verlag für Sozialwissenschaften 2007, 9-27.
- Bohnsack, Ralf: Generation, Milieu und Geschlecht – Ergebnisse aus Gruppendiskussionen mit Jugendlichen. Opladen: Leske+Budrich 1989.
- Bohnsack, Ralf: Rekonstruktive Sozialforschung. Einführung in qualitative Methoden. Opladen/Toronto: Verlag Barbara Budrich 2014.
- Bromme, Rainer: Der Lehrer als Experte. Zur Psychologie des professionellen Wissens. Bern, Göttingen, Toronto: Hans Huber 1992.
- Flick, Uwe/Kardoff, Ernst von/Steinke, Ines: Was ist Qualitative Forschung? Einleitung und Überblick. In: Dies. (Hrsg.): Qualitative Forschung. Ein Handbuch. Reinbek bei Hamburg: Rowohlt 2005, 13-29.
- Fuhs, Burkhard: Qualitative Methoden in der Erziehungswissenschaft. Darmstadt: Wissenschaftliche Buchgesellschaft 2007.
- Groeben, Norbert/Wahl, Diethelm/Schlee, Jörg/Scheele, Brigitte: Das Forschungsprogramm Subjektive Theorien. Eine Einführung in die Psychologie des reflexiven Subjekts. Tübingen: Francke 1988.
- Heizmann, Felix: Ästhetische Erfahrungen in Literarischen Gesprächen mit Grundschulkindern. Ergebnisse einer explorativen Studie. In: Kirschenmann, Johannes/Richter, Christoph/Spinner, Kaspar H. (Hrsg.): Reden über Kunst. München: kopaed 2011, 93-119.
- Kühn, Thomas/Koschel, Kay-Volker: Gruppendiskussionen. Ein Praxis-Handbuch. Wiesbaden: VS Verlag für Sozialwissenschaften 2011.
- Kunze, Ingrid: Konzepte von Deutschunterricht. Eine Studie zu individuellen didaktischen Theorien von Lehrerinnen und Lehrern. Wiesbaden: VS Verlag für Sozialwissenschaften 2004.
- Lamnek, Siegfried: Gruppendiskussionen. Theorie und Praxis. Weinheim et al.: Beltz 2005.
- Langer, Antje: Transkribieren – Grundlagen und Regeln. In: Friebertshäuser, Barbara/Langer, Antje/Prenzel, Annedore (Hrsg.): Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft. Weinheim und München: Juventa 2010, 515-526.
- Loos, Peter/Schäffer, Burkhard: Das Gruppendiskussionsverfahren. Theoretische Grundlagen und empirische Anwendung. Opladen: Leske+Budrich 2001.
- Mannheim, Karl: Beiträge zur Theorie der Weltanschauungs-Interpretation. In: Ders.: Wissenssoziologie. Auswahl aus dem Werk. Eingeleitet und herausgegeben von Kurt H. Wolff. Neuwied a. R.: Luchterhand 1964, 91-154 [EA: 1921-22, Jahrbuch für Kunstgeschichte XV, 4].
- Nittel, Dieter: Von der Kreissäge zur Weiterbildung eines Großkonzerns – das berufsbiographische Portrait eines (un)gewöhnlichen Erwachsenenbildners. In: Nittel, Dieter/Marotzki, Winfried (Hrsg.): Berufslaufbahn und biographische Lernstrategien. Eine Fallstudie über Pädagogen in der Privatwirtschaft. Baltmannsweiler: Schneider 1997, 12-35.

- Ossner, Jakob: Das Profil der Fachdidaktik. Grundzüge einer praktischen Disziplin. In: Radtke, Frank-Olaf (Hrsg.): *Lehrerbildung an der Universität. Zur Wissensbasis pädagogischer Professionalität*. Frankfurter Beiträge zur Erziehungswissenschaft. Frankfurt: Johann Wolfgang Goethe-Universität 1999, 24-45.
- Przyborski, Aglaja/Wohlrab-Sahr, Monika: *Qualitative Sozialforschung*. Ein Arbeitsbuch. München: Oldenbourg 2014.
- Schäffer, Burkhard: Gruppendiskussion. In: Bohnsack, Ralf/Marotzki, Winfried/Meuser, Michael (Hrsg.): *Hauptbegriffe Qualitativer Sozialforschung*. Opladen/Farmington Hills: Verlag Barbara Budrich 2006, 75-80.
- Scherf, Daniel: Implementationskulturen: Ermöglichungs- und Begrenzungsraum individuellen innovativen Handelns im Leseunterricht. In: Philipp, Maik/Souvignier, Elmar (Hrsg.): *Implementation von Lesefördermaßnahmen. Perspektiven auf Gelingensbedingungen und Hindernisse*. Münster et al.: Waxmann. 2016, 23-51.
- Scherf, Daniel: *Leseförderung aus Lehrersicht. Eine qualitativ-empirische Untersuchung professionellen Wissens*. Wiesbaden: Springer VS 2013.
- Schmelz, Markus: *Texte überarbeiten im Deutschunterricht der Hauptschule. Eine empirische Untersuchung zur Rezeption schreibdidaktischer Neuerungen*. Baltmannsweiler: Schneider 2009.
- Wieser, Dorothee: *Literaturunterricht aus Sicht der Lehrenden. Eine qualitative Interviewstudie*. Wiesbaden: VS Verlag für Sozialwissenschaften 2008.

Fragebögen

Am Beispiel einer Studie zu linguistischer Wissenschaftsvermittlung im Schülerlabor¹

1. Vorstellung von Fragebögen

Der Fragebogen als klassisches empirisches Erhebungsinstrument, das in vielen Fachdisziplinen (u.a. der Psychologie und den Sozialwissenschaften) eingesetzt wird, kommt v.a. in quantitativ angelegten Studien zur Anwendung. Es handelt sich dabei um eine schriftliche strukturierte Befragung (vgl. Eid/Gollwitzer/Schmitt 2013, 28), eine Methode, die als Gruppenverfahren beschrieben werden kann, da eine gleichzeitige Durchführung mit mehreren Befragten möglich ist und die gleichen Informationen bei allen Personen erhoben werden. Außerdem werden Fragebögen als reaktives Verfahren klassifiziert, indem es die Probandinnen und Probanden auffordert, auf Fragen zu antworten (vgl. ebd., 20f.). Zusammengefasst können Fragebögen wie folgt definiert werden:

Ein Fragebogen ist eine mehr oder weniger standardisierte Zusammenstellung von Fragen, die Personen zur Beantwortung vorgelegt werden mit dem Ziel, deren Antworten zur Überprüfung der den Fragen zugrundeliegenden theoretischen Konzepte und Zusammenhänge zu verwenden. Somit stellt ein Fragebogen das zentrale Verbindungsstück zwischen Theorie und Analyse dar. (Porst 1996, 738)

Hierbei wird deutlich, dass der Fragebogen theoretisch fundiert sein muss und auf zuvor definierten Konstrukten basiert, die anschließend durch die Gewinnung von Daten systematisch empirisch überprüft werden.

Doch welche Art von Konstrukten wird in Fragebogenuntersuchungen abgefragt? Fragebögen gelten als „wissenschaftliches Routineverfahren“ (Lienert 1969, 7) zur Erhebung persönlicher Merkmale von Probandinnen und Probanden, d.h. es werden deren spezifische Eigenschaften, Emotionen, Einstellungen, Interessen und Selbstkonzepte gemessen und durch mathematisch-statistische Verfahren abgesichert, sodass Aussagen über die getesteten Konstrukte möglich sind (vgl. Eid/Gollwitzer/Schmitt 2013, 30). Die genannten Untersuchungsgegenstände

¹ Diese Forschungsarbeit entsteht im Rahmen des Promotionskollegs „Wissenschaftsvermittlung im Schülerlabor“ der Professional School of Education der Ruhr-Universität Bochum.

stammen zwar aus der Psychologie, sind jedoch nicht auf diese Disziplin beschränkt, sondern wirken in viele Didaktiken hinein und sind beispielsweise für die Untersuchung von Lehr-Lernkontexten unterschiedlicher Fächer, d.h. auch in der Deutschdidaktik, relevant. Und auch genuin deutschdidaktische Fragestellungen lassen sich durch Fragebogenstudien untersuchen wie beispielsweise die Lehrerinnen- und Lehrerbefragung zu Grammatikunterricht und Grammatikterminologie in der Schule von Topalović und Dünschede (2014) oder die Bielefelder Erhebung zur Selbsteinschätzung der Schreibkompetenzen von Studierenden von Sennewald und Mandalka (2012) zeigen. Generell kann in jüngster Zeit von einer zunehmenden Empirisierung der Deutschdidaktik gesprochen werden, die Fragebogenstudien mit einschließt und besonders die Forschung in der Sprachdidaktik bereichert.²

Das Erhebungsinstrument Fragebogen kann dabei je nach Studienausrichtung und Forschungsinteresse angepasst werden und in unterschiedlicher Weise strukturiert sein, was Gegenstand der folgenden Unterkapitel ist. Dabei wird zunächst das Erhebungsinstrument unabhängig von seiner Ausrichtung und der Disziplin, die den Hintergrund der Studie bildet, beschrieben und im Anschluss beispielhaft anhand eines deutschdidaktischen empirischen Forschungsprojektes gezeigt, wie die Fragebogenentwicklung ablief und warum und in welcher Form ein Fragebogen als Erhebungsinstrument verwendet wurde.

Das Vorgehen beim Einsatz von Fragebögen bei empirischen Erhebungen besteht aus verschiedenen Schritten, die exemplarisch folgendermaßen ablaufen: Die Entwicklung des Fragebogens beginnt mit der theoretischen Fundierung, die in die Aufgaben- bzw. Itemerstellung, -anordnung und -revision und nach einer Pilotierung in die Analyse und Prüfung der Gütekriterien mündet. Darauf folgen die Datenerhebung und schließlich die Auswertung. Der Schwerpunkt soll hier auf die ersten Schritte gelegt werden, die Teil der Fragebogenkonstruktion sind, wobei vor allem Überlegungen zur Fragebogenerstellung in den Mittelpunkt rücken. Dabei kann nur ein grober Überblick über die Optionen gegeben werden, für eine vertiefte Auseinandersetzung mit dem Instrument Fragebogen wird auf die Literaturempfehlungen verwiesen (vgl. Kap. 1.5). Die Datenerhebung und -auswertung sind demgegenüber stark von der inhaltlichen Ausrichtung der Studie und dem Aufbau des Fragebogens abhängig, sodass eine Vielzahl an individuellen Zugängen und Vorgehensmöglichkeiten besteht, die an dieser Stelle nicht vollständig dargestellt, sondern nur umrissen werden kann.

² Eine speziell auf quantitative Studien in der Linguistik und Sprachlehrforschung ausgerichtete Anleitung für die empirische Studienplanung bieten Albert und Marx (2010).

1.1 Funktionsweise von Fragebögen

Bei der Fragebogenkonstruktion ist prinzipiell „auf die qualitative und quantitative Übereinstimmung des Instrumentariums mit dem Fragebogenziel zu achten“ (Porst 2011, 15), womit die vollständige und inhaltlich angemessene Operationalisierung des theoretischen Konstruktes gemeint ist. Um dies zu gewährleisten ist zunächst eine Literaturrecherche notwendig, die auf eine Eingrenzung des zu überprüfenden Merkmals, die Definition des Konstruktes sowie die Abgrenzung von anderen Konstrukten ausgerichtet ist. Nach der theoretischen Vorarbeit ist für die Entwicklung eines diesen theoretischen Überlegungen entsprechenden Fragebogens eine intensive Auseinandersetzung mit verschiedenen Aspekten, die einen Fragebogen bestimmen, wie der Fragebogenform, den Arten von verwendeten Items, der Formulierung der Items etc. notwendig.

Zentral für die Erstellung des Fragebogens ist die Bestimmung des Itemformats. Items werden als „kleinste Beobachtungseinheit“ (Rost 2004, 55) definiert, die sich aus zwei Komponenten zusammensetzt: dem Itemstamm, der aus einer Frage oder Aussage besteht, und dem Antwortformat, das beispielsweise anzukreuzende Antwortalternativen, eine leere Zeile o.ä. enthält, in die durch die Befragten Daten eingespeist werden. Items als Einheit versteht Rost (2004) auch im Sinne einer Einheitlichkeit, die nicht nach zwei Dingen gleichzeitig fragen sollte, denn „das im Antwortformat registrierte Verhalten muss eindeutig auf die im Itemstamm vorgegebene Situation (Frage) zurückzuführen sein, wenn das Testverhalten Rückschlüsse auf die Personeigenschaft erlauben soll“ (ebd., 55).

Als mögliche Items stehen verschiedene Frageformate zur Auswahl, die hinsichtlich ihres Standardisierungsgrades Unterschiede aufweisen: Bei einer quantitativen Ausrichtung des Fragebogens mit hohem Standardisierungsgrad werden überwiegend geschlossene Items verwendet und bei stärker qualitativem Fokus offene Fragen, die frei formulierte Antworten verlangen; es kann jedoch auch eine Kombination aus beidem zum Einsatz kommen. Als Beispiele für offene Aufgabenformate werden Ergänzungsaufgaben (z.B. Ergänze eine Definition: Unter Rechtschreibkompetenz versteht man ...) und der Kurzaufsatz genannt (vgl. Bühner 2006, 64f.). Vorteile dieses Frage-Formates bestehen nach Raab-Steiner und Benesch (2010) in der Tatsache, dass keine Kategorien o.ä. vorgegeben werden und dadurch eine weniger starke Lenkung stattfindet. Nachteilig kann sich hingegen auswirken, dass die Offenheit möglicherweise Verbalisierungshemmungen auslöst und eine Auswertung zeitaufwendig ist (vgl. Raab-Steiner/Benesch 2010, 48). Aufgrund des größeren Aufwands ziehen daher Probandinnen und Probanden häufig geschlossene Itemformate vor (vgl. ebd.). Demgegenüber zeichnen sich halboffene Aufgabenformate durch vorgegebene Antwortkategorien aus, die jedoch von der befragten Person um weitere ergänzt werden können (z.B. Welches ist Ihr Lieblingsfach? Deutsch; Mathe; Englisch; Ein anderes, und zwar: ...). Der Vorteil liegt hier in der leichteren Auswertbarkeit und der Offenheit gegenüber zusätzlichen Kategorien, die vorher nicht bedacht wurden. Geschlossene Items geben stattdessen nur begrenzte und definierte Antwortkategorien vor, es wird

von den Probandinnen und Probanden folglich eine Auswahl aus einem eingeschränkten Verhaltensbereich getroffen (vgl. Rost 2004, 59), sodass die Auswertungsökonomie hoch ist, aber keine individuellen Anpassungen möglich sind. Dazu gehören Items wie beispielsweise, Welche der folgenden Begriffe sind Lernbereiche des Deutschunterrichts? Sprachreflexion; Orthographie; Präsentieren; Sprechen und Zuhören; ...‘ Bei sogenannten Single-Choice-Aufgaben, die zu den geschlossenen Aufgaben gehören, ist nur eine Antwort zulässig, bei Multiple-Choice-Aufgaben sind Mehrfachnennungen möglich; Fragebögen sollten daher Hinweise zur geforderten Anzahl der Antworten enthalten.

Weiterhin wird in Bezug auf diesen Aufgabentyp unterschieden zwischen Zuordnungsaufgaben (bei denen Inhalte anderen Inhalten zugeordnet werden), Umordnungsaufgaben (die dazu auffordern, Fragmente in eine bestimmte Reihenfolge zu bringen), Richtig-Falsch- bzw. Ja-Nein-Aufgaben (mit dichotomen Antwortformaten) und Ratingskalen (bei denen die Befragten die Wahl haben zwischen mehr als zwei abgestuften Antwortkategorien (vgl. Bühner 2006, 53ff.; Raab-Steiner/Benesch 2010, 47ff.) und diese „Antwortkategorien *item-unspezifisch* formuliert“ (Rost 2004, 64), d.h. für mehrere Items gleich benannt und gültig sind). Ratingskalen, die in vielen Fragebögen eingesetzt werden, sind ‚informationsreicher‘ als andere geschlossene Antwortformate, indem sie eine differenzierte Äußerung beispielsweise von Zustimmung oder Ablehnung in verschiedenen Abstufungen anhand einer Skala, d.h. eines Bezugssystems ermöglichen (vgl. ebd., 65), z.B. ‚Stimmt gar nicht‘, ‚Stimmt wenig‘, ‚Stimmt ziemlich‘, ‚Stimmt völlig‘. Man unterscheidet verschiedene Skalenniveaus, die für Ratingskalen infrage kommen. Die Wichtigsten für die Testkonstruktion sind erstens Nominal-Skalen, die eine Zuordnung in wertfreie Kategorien vorsehen (z.B. männlich/weiblich) und Aussagen über Gleichheit bzw. Ungleichheit zulassen, zweitens Ordinal-Skalen, die vergleichende Aussagen (z.B. größer/kleiner) generieren und sich ordnen lassen, und drittens Intervall-Skalen, die Relationen und Unterschiede abbilden (Maßeinheiten) und deren Abstand gleich ist (vgl. Bühner 2006, 73ff.). Diese wiederum können Unterschiede dahingehend aufweisen, ob sie unipolar oder bipolar verlaufen, d.h. in eine Richtung oder zwischen einem negativen und einem positiven Pol verlaufen, welche Anzahl an Abstufungen sie enthalten und ob eine numerische oder eine verbale Skalenbezeichnung gewählt wird, wobei letztere verschiedene Urteilsdimensionen wie Häufigkeit, Intensität, Wahrscheinlichkeit und Bewertung zulässt (vgl. Raab-Steiner/Benesch 2010, 54ff.). Die Verwendung von Ratingskalen ist verbunden mit Vorteilen wie einer ökonomischen Durch- und Ausführung der Erhebung und Auswertung sowie differenzierten Antworten, aber auch mit dem Nachteil, dass Abstufungen von den Teilnehmerinnen und Teilnehmern unterschiedlich aufgefasst werden (vgl. Bühner 2006, 55f.) und Urteilstendenzen auftreten können wie z.B. die Tendenz zur Mitte (bei der Extremurteile vermieden werden), der Primacy-Effekt (wobei die Beurteilung der ersten Items die Urteile über die weiteren beeinflusst) und der Halo-Effekt (wenn das Urteil über ein Merkmal auf andere Merkmale ausstrahlt) (vgl. Hussy/Schreier/Echterhoff 2010, 76ff.; s. Kleinbusch in diesem Band). Generell sollte bei der Konstruk-

tion von Fragebögen beachtet werden, dass Antworttendenzen vorkommen können, die die Ergebnisse verfälschen. Dazu gehören neben den genannten die absichtliche Verstellung, Antworten gemäß der sozialen Erwünschtheit, d.h. nach Meinung der Testperson der sozialen Norm entsprechend, die Akquieszenz, bei der unabhängig vom Inhalt eher die Antwort ‚ja‘ gewählt wird als ‚nein‘, die Bevorzugung von extremen, unbestimmten oder besonders platzierten Antwortkategorien usw. (vgl. Raab-Steiner/Benesch 2010, 59ff.).³

Nach Festlegung der Itemformate ist der nächste Schritt die inhaltliche Itemerstellung bzw. -sammlung. Dabei können verschiedene Konstruktionsprinzipien verfolgt werden: Beim ‚Rationalen Konstruktionsprinzip‘ beispielsweise werden die Items aus der Theorie abgeleitet, beim ‚Induktiven Konstruktionsprinzip‘ werden Items aus Faktorenanalysen oder korrelationsanalytischen Verfahren gewonnen, weiterhin können vorhandene und bereits erprobte Fragebögen abgewandelt werden, qualitative Voruntersuchungen und Expertenbefragungen zur Itemerstellung genutzt werden (vgl. Mummendey/Grau 2008, 63ff.). Die Itemsammlung sollte zunächst „möglichst mindestens das Dreifache der letztlich erforderlichen Itemanzahl“ (ebd., 65) umfassen, damit nach der Itemrevision, die v.a. hinsichtlich der Verständlichkeit, Kürze und Prägnanz der Formulierungen, Art der Frage- richtung erfolgt, noch genügend Items für den Fragebogen zur Verfügung stehen (vgl. ebd., 66).⁴ Entsprechend der Zielgruppe des Fragebogens, die sich aus dem Forschungsinteresse ergibt, sollten die Formulierungen verständlich, präzise und auf die Zielgruppe abgestimmt sein (z.B. die Anrede ‚Du‘ oder ‚Sie‘ verwenden). Bei der Itemrevision ist auch die Reihenfolge der Items zu beachten.

Für eine sinnvolle Gestaltung der Struktur des Fragebogens gilt: Unabhängig davon, ob eine Papier-Bleistift-Version gewählt wird oder der Fragebogen online ausgefüllt werden soll, beginnt der Fragebogen in der Regel mit einer kurzen Instruktion, die das Interesse der Befragten am Thema hervorrufen und einige Hintergrundinformationen liefern soll. So enthält sie eine kurze Darstellung der Einrichtung bzw. Person, die die Erhebung durchführt, eine Erklärung über die Weiterverwendung der gewonnenen Daten, die wichtigsten Hinweise zum Ausfüllen des Fragebogens, eine Zusicherung der Anonymität sowie eine Danksagung für die Bearbeitung (vgl. Raab-Steiner/Benesch 2010, 49f.). Insgesamt sollte die Logik des Befragungsablaufs transparent sein, d.h. Fragen zu einem Thema sind als ein Fragenblock zusammengefasst und mit einer Überschrift versehen. Nach einer leichten ‚Aufwärmfrage‘ können die für die Forscherinnen und Forscher besonders interessanten und relevanten Items aufgeführt werden, während heikle Fragen am Ende des Fragebogens platziert werden sollten, sodass bei einer Nichtbeantwortung und schwindenden Motivation zu den wichtigsten Fragen bereits Daten vorliegen (vgl. Porst 2011, 142f.). Die Testlänge (Anzahl der Testaufgaben)

³ Für Vermeidungsmöglichkeiten einiger Antworttendenzen bei der Fragebogenkonstruktion bzw. für weiterführende Informationen s. Mummendey/Grau 2008, 154ff.

⁴ Tipps für die Itemformulierung sind beispielsweise nachzulesen in Porst (2011, 95ff.) oder Hussy/Schreier/Echterhoff (2010, 73f.).

sowie die eventuell vorgegebene Testzeit zur Beantwortung des Fragebogens sollten – genau wie das Layout – der Zielgruppe entsprechend gestaltet und für die Befragten transparent sein. Um für die Beachtung der Zeit sowie weiterer Regeln beim Ablauf der Erhebung zu sorgen, kann der Einsatz von Testleiterinnen und Testleitern (z.B. Lehrerinnen und Lehrer oder Hilfskräfte) helfen, die zuvor geschult werden und einen Testleitfaden mit den wichtigsten beim Ausfüllen zu beachtenden Regeln und Hinweisen zur Durchführung an die Hand bekommen. Die Testleiterinnen und Testleiter führen die Erhebung durch, kontrollieren die Einhaltung der Regeln und sammeln die ausgefüllten Fragebögen ein. Eine weitere Option stellt das Verschicken der Fragebögen per Post bzw. Mail mit der Bitte um Rücksendung dar, allerdings birgt dieses Verfahren die Gefahr, dass nicht alle Fragebögen wieder zurückkommen.

Die grafische Gestaltung von Fragebögen erleichtern kostenlose Tools wie ‚Umfrage-Online‘ (<https://www.umfrageonline.com/>) oder auch ‚Häkchen‘ (<http://www.haekchen.at/haekchen/home.asp>), ein Tool, das den Ausdruck von Word- oder PDF-Dateien zusätzlich zur Möglichkeit von Onlineumfragen erlaubt. Mit der Software ‚EvaSys‘ (<https://www.evasys.de/startseite.html>) lässt sich darüber hinaus nicht nur die Fragebogenerstellung umsetzen, sondern das Programm ist auch in der Lage, diese auszuwerten, allerdings ist dessen Nutzung nicht kostenlos.

Bevor ein Fragebogen bei einer empirischen Studie verwendet wird, ist es üblich, ihn zu pilotieren, d.h. ihn von Probandinnen und Probanden ausfüllen zu lassen und dann einer Analyse und ggf. einer Überarbeitung zu unterziehen, bevor er in der Hauptstudie eingesetzt wird (vgl. Mummendey/Grau 2008, 90f.). Dabei wird erstens überprüft, ob der Fragebogen bestimmten Gütekriterien entspricht, und versucht, diese Standards zu erfüllen, sowie zweitens eine Itemanalyse durchgeführt, um herauszufinden, welche Items beibehalten werden.

1.2 Ausgangsdaten und Ergebnisse

Aus der Erhebung, in der der erstellte Fragebogen zum Einsatz kommt, gehen die Daten hervor, die anschließend so aufbereitet werden müssen, dass eine Auswertung möglich ist. Bei halboffenen und geschlossenen Antwortformaten „wird lediglich ausgewertet, welche der vorgegebenen Antwortmöglichkeiten genutzt wurde“ (Eid/Gollwitzer/Schmitt 2013, 31), bei offenen Fragen sind Inhaltsanalysen sowie die Generierung eines Systems von Antwortkategorien notwendig, denen die Antworten zugeordnet werden können (vgl. ebd., 32; Bühner 2006, 65). Eine geeignete Methode dafür ist z.B. die ‚Qualitative Inhaltsanalyse‘ nach Mayring (vgl. Mayring 2008 und Heins in diesem Band), die hilfreich bei der Strukturierung der ausformulierten Antworten der Befragten und der Kategorienbildung sein kann.

Für die Datenaufbereitung quantitativer Fragebögen werden – wenn kein Fragebogentool verwendet wurde, welches die Daten automatisch einliest und in Zahlenwerten ausgibt – die Daten in Zahlen überführt, indem die vorgegebenen Antwortkategorien Zahlenwerten zugeordnet werden. Dabei werden die Daten von

den Befragungsbögen in ein Format gebracht, das mit Statistikprogrammen kompatibel ist, und anschließend einzelne Variablen zu Gesamtwerten zusammengefasst, fehlende Werte kodiert etc. (vgl. Hussy/Schreier/Echterhoff 2010, 160ff.). Für die Auswertung dieser aufbereiteten Daten werden Statistikkennnisse vorausgesetzt, insofern sollte vorher eine Einarbeitung in die verwendete Statistiksoftware, häufig ‚SPSS‘, stattfinden. Die individuelle Fragestellung der Studie bedingt die statistischen Verfahren, die für die Auswertung der Daten infrage kommen. Beispielsweise können verschiedene deskriptiv-statistische Methoden verwendet werden, um Häufigkeitsverteilungen, Korrelationen, Regressionen zu berechnen und/oder inferenz-statistische Methoden für Hypothesentests wie Signifikanztests oder Varianzanalysen eingesetzt werden (vgl. ebd., 163ff.). Letztlich zielen diese Verfahren darauf ab, aus den Erhebungsdaten Ergebnisse zu generieren und Aussagen über Persönlichkeitsmerkmale, Interessen o.Ä. ableiten zu können und damit schließlich die Forschungsfrage zu beantworten.

1.3 Vorteile und Nachteile von Fragebögen

Mit der Wahl des Erhebungsinstruments Fragebogen für eine empirische Studie gehen Vor- und Nachteile einher, über die sich die Studienleiterin oder der Studienleiter bewusst sein sollte.

Als Vorteile nennen Eid, Gollwitzer und Schmitt (2013, 29) die Ökonomie hinsichtlich der Durchführung ohne zeitintensiven Schulungsaufwand und als Gruppenverfahren, den geringen Materialaufwand, die hohe Durchführungs- und Auswertungsobjektivität sowie die hohe Vergleichbarkeit der erhobenen Informationen über mehrere Personen. Weiterhin ist zu erwähnen, dass es sich bei Fragebögen um ein klassisches und häufig eingesetztes Verfahren handelt, sodass für viele Konstrukte bereits Items und Skalen bestehen und diese empirisch erprobt sind, sodass eine Orientierung daran möglich ist.

Allerdings gehen mit der Entscheidung für dieses Instrument auch Nachteile einher: Fragebögen sind unflexibel, d.h. die Daten beschränken sich auf die gestellten Aufgaben und es können Verfälschungen der Ergebnisse durch Urteilstendenzen wie die ‚Tendenz zur Mitte‘ und die ‚Soziale Erwünschtheit‘ auftreten, die schwerer aufzudecken sind als bei mündlichen Befragungen (vgl. Eid/Gollwitzer/Schmitt 2013, 29).

1.4 Gütekriterien für den Einsatz von Fragebögen

Die wichtigsten Kriterien in der klassischen Testtheorie zur Eignungsüberprüfung des eingesetzten Fragebogens als Forschungsinstrument sind die Objektivität, die Reliabilität und die Validität (vgl. Rost 2004, 33). Diese Gütekriterien, die jeder Test in einem Mindestmaß zu erfüllen hat, werden wie folgt definiert: Die Objektivität gibt an, „inwiefern das Testergebnis unabhängig ist von jeglichen Einflüssen außerhalb der getesteten Person, also vom Versuchsleiter, der Art der Auswertung, den situativen Bedingungen“ (ebd., 33) etc., die Reliabilität bestimmt die Zuverlässigkeit eines Tests, d.h. „wie genau der Test das misst, was er misst (egal, was er misst)“ (ebd., 33), d.h. bei Messwiederholung liefert das Instrument

bei gleichbleibenden Eigenschaften die gleichen Messwerte, und mit Validität als Indikator für den Grad der Gültigkeit der Messung ist gemeint, „inwieweit der Test das misst, was er messen soll“ (ebd., 33).

Weitere Untergliederungen innerhalb der drei Hauptgütekriterien definieren die Standards, an denen Fragebögen gemessen werden, genauer. Hinsichtlich der Objektivität werden die ‚Durchführungs-‘, ‚Auswertungs-‘ und ‚Interpretationsobjektivität‘ als Teilbereiche unterschieden. Die angestrebte Objektivität betrifft bei der Durchführung die Bedingungen der Fragebogenerhebung wie z.B. die Instruktion, den Umgang mit Rückfragen, die zeitlichen Vorgaben, die bei allen Messungen gleich sein müssen, bei der Auswertung eine gleiche Punktvergabe bei gleichen Antworten und genaue Auswertungsvorschriften, die bei offenen Aufgaben schwieriger zu definieren sind als bei geschlossenen, und bei der Interpretation eine gleiche Beurteilung und Deutung der Ergebnisse unabhängig von der auswertenden Person (vgl. Bühner 2006, 34f.).

In Bezug auf die Reliabilität wird differenziert zwischen erstens der ‚Inneren Konsistenz‘, bei der der Fragebogen halbiert und die Teile miteinander korreliert werden, um dessen Äquivalenz zu testen, und die standardmäßig durch den Cronbach- α -Koeffizienten bestimmt wird (vgl. ebd., 35, 131f.), zweitens der ‚Retest-Reliabilität‘, welche die Korrelation von Testergebnissen an zwei Erhebungszeitpunkten angibt, und drittens der ‚Paralleltestreliabilität‘, wobei die Korrelationen von zwei Tests, die dieselben Merkmale messen sollen, berechnet werden (vgl. ebd., 35f.).⁵

Weiterhin wurden drei Validitätsarten klassifiziert: die ‚Inhaltsvalidität‘, die schwer zu bestimmen ist und angibt, ob ein Item das zu erfassende Konstrukt auch wirklich misst, die ‚Kriteriumsvalidität‘, die auf Korrelationen zwischen Testkennwerten und Kriterien (z.B. Schulnoten) beruht, und die ‚Konstruktvalidität‘, aus der abgeleitet wird, dass der Fragebogen die Merkmale misst, auf die er abzielt (vgl. ebd., 36ff.).

Nach der Pilotierung des erstellten Fragebogens wird außerdem – nicht ausschließlich, aber vor allem bei Items zur Leistungsmessung – eine Itemanalyse durchgeführt, die in der statistischen Überprüfung jedes Items besteht und mithilfe derer entschieden wird, welche Items in den endgültigen Fragebogen aufgenommen werden (vgl. Mummendey/Grau 2008, 97). Dazu gehört die Überprüfung des Schwierigkeitsindexes, der zur Tilgung solcher Items dienen soll,

die entweder von fast niemandem oder von fast allen Personen im Sinne des zu messenden Merkmals bejaht werden, die also entweder viel zu *schwierig* oder viel zu *leicht* sind, so dass sie sehr konforme Reaktionen hervorrufen und letztlich zur Unterscheidung zwischen verschiedenen Individuen nicht gut brauchbar sind. (ebd., 97)

⁵ Bühner (2006, 124ff.) widmet der Erklärung der Reliabilität und ihrer Berechnung ein vollständiges Kapitel, das auch das Vorgehen im Statistikprogramm SPSS einschließt.

Ein mittlerer Schwierigkeitsgrad von $p=50\%$ wird als optimal angesehen; aussortiert werden meist zu leichte bzw. zu schwere Items mit einem Schwierigkeitsgrad von unter 20% und über 80% (vgl. ebd., 97).

Als weiterer Indikator für die Überprüfung der Items gilt die Trennschärfe, die angibt, „wie stark jedes Item den endgültigen Fragebogen-Wert bereits vorher sagt, wie gut es das gesamte im Fragebogen gemessene Konstrukt repräsentiert“ (ebd., 98). Eine hohe Trennschärfe von ca. über .4 (wobei eine Akzeptanz von niedrigeren Werten aus verschiedenen Gründen auch möglich ist) bescheinigt den Items eines Konstruktes eine Homogenität, d.h. jedes Item trägt viel zum Fragebogen-Gesamtwert bei (vgl. ebd., 98f.).⁶

1.5 Literaturempfehlungen

Bühner (2006): Einführung in die Test- und Fragebogenkonstruktion.

Diese Einführung liefert anschauliche Erklärungen in Bezug auf die Testkonstruktion, außerdem enthält sie eine ausführliche Darstellung verschiedener Auswertungsmöglichkeiten (auch in SPSS).

Eid et al. (2013): Statistik und Forschungsmethoden.

In diesem Band werden verschiedene Forschungsmethoden vorgestellt, er ist klar gegliedert und leicht verständlich geschrieben.

Hussy et al. (2010): Forschungsmethoden in Psychologie und Sozialwissenschaften.

Ein disziplinenübergreifendes Lehrbuch, das Grundbegriffe erklärt und quantitative wie qualitative Erhebungs-, Forschungs- und Auswertungsmethoden sowie Mixed Methods vorstellt.

Raab-Steiner/Benesch (2010): Der Fragebogen.

Dieser Band ist auf die Methode ‚Fragebogen‘ ausgerichtet und mit einem ausführlichen Teil zu statistischen Analysen (u.a. in SPSS) und hilfreichen Hinweisen zum statistischen Auswertungsbericht versehen.

Rost (2004): Testtheorie – Testkonstruktion.

Rost thematisiert Aspekte der Fragebogenerstellung wie Testtheorie, -konstruktion, -modelle und -optimierung, besonders hervorhebenswert sind die griffigen Definitionen und vielen veranschaulichenden Beispiele.

⁶ Hinweise zur Berechnung von verschiedenen Schwierigkeitsindizes (u.a. in SPSS) finden sich bei Bühner 2006, 83ff., weiterführende Informationen zur Trennschärfe ebd., 95ff.

2. Eine Fragebogenstudie zur linguistischen Wissenschaftsvermittlung im Schülerlabor

Dass Fragebögen auf Grund ihrer Vorteile und verschiedenen Adaptionmöglichkeiten an individuelle Forschungsfragen auch in empirischen deutschdidaktischen Studien einsetzbar sind, zeigt nicht nur der erste Teil dieses Beitrags, sondern auch ihre Verwendung in verschiedenen Studien (z.B. Topalović/Dünschede 2014; Sennewald/Mandalka 2012). Im Folgenden soll nun anhand einer aktuellen sprachdidaktischen Studie, welche mittels Fragebögen die Bedingungen des Lernens in einem linguistisch ausgerichteten Schülerlaborprojekt erforscht, beispielhaft dargelegt werden, welche Gründe zu der Wahl dieses Forschungsinstruments führten und wie dessen konkrete Umsetzung aussah.

2.1 Zielsetzung und Konzeption der Studie

Schülerlabore sind außerschulische Lernangebote für Schülerinnen und Schüler, die seit den 90er Jahren vor dem Hintergrund eines mangelnden Interesses der Jugendlichen an naturwissenschaftlichen Fächern und Berufen sowie unzureichender Leistungen im MINT-Bereich entstanden sind und sich großer Beliebtheit erfreuen (vgl. Pawek 2009, 2f.). Neben Schülerlaboren, die Projekte zu naturwissenschaftlichen oder technischen Themen anbieten, gibt es in einigen Institutionen seit 2006 auch Lernangebote aus dem geisteswissenschaftlichen Bereich. Um das Lernen und dessen Bedingungen in Schülerlaboren empirisch zu erforschen, sind im letzten Jahrzehnt einige empirische Wirkungsstudien durchgeführt worden, die neben weiteren Variablen v.a. das aus dem Laborbesuch resultierende Interesse an den naturwissenschaftlichen Fächern in den Blick nehmen (z.B. Engeln 2004; Scharfenberg 2005; Pawek 2009). Geisteswissenschaftliche Schülerlaborforschung sucht man allerdings bisher vergeblich. Ein Forschungsdesiderat besteht darin, zu überprüfen, ob sich ein geisteswissenschaftliches Schülerlaborprojekt ähnlich auswirkt wie ein naturwissenschaftliches und beispielsweise in der Lage ist, das Interesse an dem entsprechenden Unterrichtsfach zu fördern.

Auf diese Forschungslücke abzielend ist im Kontext des Promotionskollegs ‚Wissenschaftsvermittlung im Schülerlabor‘ der Ruhr-Universität Bochum eine Studie konzipiert worden, deren Erkenntnisinteresse und Methodik an die bisher erfolgten quantitativ ausgerichteten Schülerlabor-Fragebogenstudien anknüpft, aber kein naturwissenschaftliches, sondern ein deutschdidaktisches, in diesem Falle linguistisches, Projekt auf seine Wirkung hin untersucht. Dahinter verbirgt sich das Studienziel, mögliche Implikationen für einen Grammatikunterricht abzuleiten, der das Interesse an linguistischen Themen fördert. Thema des Laborprojektes ist das Ruhrdeutsche, das als Gegenstand linguistischer Forschung (an der Ruhr-Universität) den Schülerinnen und Schülern nähergebracht und dessen Merkmale anhand eines von den Lernenden selbst erstellten Transkriptes sprachlich erforscht werden.

Dabei wird neben verschiedenen Hintergrundvariablen v.a. die Wirkung eines Laborbesuchs hinsichtlich der psychologischen Konstrukte ‚Interesse‘ und ‚Fähigkeitsselbstkonzept‘ erforscht und eine Bedingung des Lernens in Schülerlaboren, die in der Literatur häufig als charakteristisches Merkmal dieser genannt wird, in den Mittelpunkt der Studie gerückt: die ‚Authentizität‘. Darunter wird in einigen der Schülerlaborstudien der Einblick der Lernenden in die Wissenschaftsdisziplin und das Lernen unter der ‚realen‘ Forschungswelt nahekommenden Bedingungen verstanden, die sich positiv auf die Motivation, das Interesse und die Kognition auswirken sollen (vgl. Pawek 2009, 21; Scharfenberg 2005, 24). Die theoretische Auseinandersetzung mit Authentizität in Lehr-Lernkontexten legt ein mehrdimensionales Authentizitätskonzept nahe, das als Basis für die Operationalisierung im Fragebogen dient. Der verwendete Fragebogen zielt darauf ab, zu klären, ob das für das Schülerlaborprojekt gewählte Lernsetting von den Schülerinnen und Schülern als authentische Wissenschaftsvermittlung aufgefasst wird und welche Rolle der Authentizität in Hinblick auf die Interessensentwicklung an der Beschäftigung mit der Grammatik des Deutschen zukommt.

Das dazu gewählte Studiendesign ist ebenfalls an die bisherige bundesweite Schülerlaborforschung angelehnt, um die Vergleichbarkeit mit dort generierten Ergebnissen zu gewährleisten. Die hier vorgestellte deutschdidaktische Studie basiert auf quantitativ ausgerichteten Fragebögen zu drei Messzeitpunkten, d.h. in einem Pre-, Post-, Follow-Up-Design, die eine Intervention zu den sprachlichen Besonderheiten des Ruhrdeutschen mit Schülerinnen und Schülern aus sechs Gymnasialklassen der Jahrgangsstufe acht begleiten. Ein Kontrollgruppendesign sorgt dafür, dass es zwei unterschiedlich authentische Lernumgebungen gibt: die eine Hälfte der Klasse nimmt an dem Projekt im Alfred-Krupp-Schülerlabor der Ruhr-Universität Bochum teil und tritt dabei in Kontakt mit einem Forscher, die andere Hälfte erlebt das Projekt an der eigenen Schule ohne Forscher, wobei die Inhalte, Methoden und das Material in beiden Projekten gleich sind.

Die Fragebögen, die als Forschungsinstrument bei der Studie dienen, sollen nun vor dem Hintergrund der die Fragebogenkonstruktion begleitenden Überlegungen dargestellt werden.

2.2 Die verwendeten Fragebögen

Neben der Konzeption der bisherigen Schülerlaborstudien, die ebenfalls größtenteils mit quantitativ ausgerichteten Fragebögen operieren und als Orientierung für die hier vorgestellte Studie zur linguistischen Wissenschaftsvermittlung im Schülerlabor dienen, gibt es weitere Gründe für die Entscheidung für das Erhebungsinstrument Fragebogen: Das Forschungsinteresse richtet sich besonders auf Konstrukte der pädagogischen Psychologie, die bereits vielfach in Studien durch Fragebögen erfasst wurden, was sowohl eine theoretische Fundierung als auch eine empirische Umsetzung erleichtert. Außerdem spielen die Erhebungs- und Auswertungsökonomie des Verfahrens eine wichtige Rolle. Die Durchführung in Projektgruppen bedeutet einen nur geringen zeitlichen Aufwand und die Wahl vorrangig geschlossener Items eine recht unkomplizierte und schnelle Auswertung.

Die Fragebogenstruktur ergibt sich einerseits aus den drei Messzeitpunkten, an denen teils unterschiedliche Konstrukte gemessen werden und mit denen verschiedene Zielsetzungen verknüpft sind, um die Wirkung der Schülerlaborintervention zu erfassen und statistisch nachzuweisen. Bezogen auf die in der Zielsetzung genannten Konstrukte, zu denen anhand der Fragebögen Daten gesammelt werden, bedeutet dies, dass im Pre- und Follow-Up-Test das ‚individuelle Interesse‘, ‚Fähigkeitsselbstkonzept‘ und die ‚wahrgenommene Authentizität‘ erhoben werden, während der Posttest das ‚situationale Interesse‘ und ebenfalls die ‚wahrgenommene Authentizität‘, die in unmittelbarem Zusammenhang mit der Intervention stehen, misst. Zudem werden im Pretest (ungefähr eine Woche vor der Projektteilnahme) einige Hintergrundinformationen der Schülerinnen und Schüler (wie z.B. der sozioökonomische Status) und Einstellungen erfasst. Im Posttest werden v.a. die Erfahrung direkt nach der Intervention erfragt und im Follow-Up-Test nach zehn bis 14 Tagen noch einmal einige der Konstrukte der ersten Erhebung erhoben, deren Beurteilung sich möglicherweise durch die Intervention geändert hat. Dies heißt konkret, dass nicht alle drei Fragebögen identisch sind, sondern sich je nach Messzeitpunkt aus unterschiedlichen Konstrukten zusammensetzen, die teils mehrfach, teils aber auch nur einmal im Rahmen der Studie erfragt werden. Die Skalen zur Erfassung der jeweiligen Konstrukte sind jedoch über die Erhebungszeitpunkte identisch.

Diese mehrdimensional angelegten Konstrukte als zentrale Fragebogenbestandteile wurden zuvor aus der Theorie heraus definiert, sodass das ‚Rationale Konstruktionsprinzip‘ für die Skalenkonzeption genutzt wurde. Am Beispiel der Interessenskala als einem zentralen Teil der Fragebögen wird im Folgenden beschrieben, wie aus der Theorie Fragebögen entstehen und die zuvor definierten Merkmale des Konstruktes systematisch umgesetzt werden.

Die Grundlage für die Interessenskala bildet die Person-Gegenstands-Konzeption des Interesses der ‚Münchener Schule‘ um Schiefele, Prenzel und Krapp (z.B. Krapp/Prenzel 1992), die einerseits eine Unterscheidung in ein stabiles, längerfristig wirkendes ‚individuelles Interesse‘ und ein an eine Gegebenheit gekoppeltes ‚situationales Interesse‘ vorsieht (vgl. Krapp 1992, 11ff.) und andererseits die Operationalisierung aller drei Interessenskomponenten, der emotionalen, der epistemischen und der wertbezogenen, nahelegt (ebd., 312). Weiterhin kann das Interesse auf verschiedene Gegenstände, d.h. beispielsweise Themen, ausgerichtet sein, sodass in diesem Fall das ‚individuelle Interesse‘ über die drei Bereiche ‚Individuelles Fachinteresse Deutschunterricht‘, ‚Individuelles Sachinteresse am Nachdenken über Sprache‘ und ‚Individuelles Fachinteresse Grammatikunterricht‘ modelliert wird, die für die Projektteilnahme als relevant erachtet und über mehrere Items (zwischen vier und sieben Items je Bereich) in den Fragebogen aufgenommen wurden. Bei der Itemformulierung wird außerdem darauf geachtet, dass pro Bereich alle Interessenskomponenten abgefragt werden. Damit besteht die Skala zum ‚individuellen Interesse‘ aus insgesamt 16 Items, die als Aussagen formuliert sind und auf Grundlage der Zustimmung (vier Stufen: ‚Stimmt gar nicht‘, ‚Stimmt wenig‘, ‚Stimmt ziemlich‘, ‚Stimmt völlig‘) oder mit ‚Weiß nicht‘

von den Schülerinnen und Schülern bewertet werden. Das gewählte Antwortformat ist also eine vierstufige Ratingskala, die daher keine Mittelkategorie umfasst, ordinalskaliert ist und auf verbalen Skalenbezeichnungen basiert. Als Orientierung für die sprachliche Formulierung der zu beurteilenden Aussagen dienten Items von Pawek (2009) und Willems (2011).

Das ‚situationale Interesse‘ wurde ebenfalls über die drei Komponenten des Interesses operationalisiert und die Itemerstellung der geschlossenen Itemformate erfolgte auf Grundlage einiger von Willems (2011), Pawek (2009) und Zinn (2008) entworfenen Interessensitems, die für die Studie angepasst wurden. Auch die so zusammengestellte Interessensskala zum ‚situationalen Interesse‘ enthält 16 Items, die auf einer Ratingskala mit den gleichen vier Abstufungen und der Möglichkeit, ‚Weiß nicht‘ anzukreuzen, beurteilt werden. Damit entspricht der Interessensfragebogen der Definition von Rost, der darunter „eine Selbstauskunft über innere Zustände oder Vorgänge“ (Rost 2004, 51) versteht und Ratingskalen als geeignetes Antwortformat dafür sieht, denn „Interessen sind wie Einstellungen objektbezogen, d.h. man hat ein Interesse an etwas oder für etwas, und man drückt sein Interesse wie seine Meinung gerne graduell abgestuft aus“ (Rost 2004, 51). Für andere Skalen wie z.B. das Fragebogenkonstrukt ‚Authentizität‘ standen stattdessen kaum Items aus anderen Studien zur Adaption zur Verfügung (lediglich vier Items aus der Studie von Pawek (2009) eigneten sich), sodass auf der Theorie basierend selbstständig Items entwickelt werden mussten.

Für alle drei Fragebogenkonstrukte, ‚individuelles‘ und ‚situationales Interesse‘ sowie die ‚wahrgenommene Authentizität‘, ergab die Reliabilitätsprüfung mit SPSS Cronbachs- α -Werte von über .8, sodass sich die Operationalisierung durch eine gute innere Konsistenz auszeichnet und diesbezüglich keine Veränderungen der Skalen notwendig waren. Weiterhin wurde die Trennschärfe zwischen den Items der Konstrukte ermittelt und nur Items mit einer Trennschärfe von über .3 in den Skalen behalten.

Neben den genannten Konstrukten umfasst der Fragebogen für die Studie über linguistische Wissenschaftsvermittlung im Schülerlabor einige Multiple- bzw. Single-Choice-Fragen, die sich nicht auf Einschätzungen und Einstellungen, sondern auf exakte Angaben (z.B. Geschlecht) und Daten beispielsweise zum sozioökonomischen Status der teilnehmenden Schülerinnen und Schüler beziehen, und vereinzelte offene Items, die auf das Vorwissen, die Schülervorstellungen oder eine Projektevaluation abzielen. Während die Wahl vorrangig geschlossener Items die Auswertung erleichtert und weniger zeitintensiv ist, zielen diese Aufgaben bewusst auf frei formulierte Antworten, ohne dass den Befragten Kategorien vorgegeben und damit die Antworttexte beispielsweise zu ihren eigenen Vorstellungen beeinflusst werden. Im Fragebogen sind diese offenen Fragen mit einem Vermerk versehen, der eine Antwort in Sätzen oder Stichworten freistellt, um Schreibhemmnisse zu vermeiden.

Die auf diese Weise konzipierten Fragebögen für die drei Erhebungszeitpunkte sind dem Studienziel entsprechend auf Schülerinnen und Schüler der Mittelstufe ausgerichtet, für diese verständlich formuliert und verwenden die Anrede ‚du‘.

Außerdem umfassen sie nicht mehr als 60 Items pro Fragebogen, um die Schülerinnen und Schüler nach der Intervention bzw. im Unterricht maximal 20 Minuten aufzuhalten und keine allzu großen Konzentrationsabfälle zu provozieren. Damit ein ortsunabhängiges Ausfüllen der Fragebogen möglich ist, wurde ein Papier-Bleistift-Test gewählt, der den Schülerinnen und Schülern von einer Testleiterin oder einem Testleiter ausgeteilt und nach der Beantwortung wieder eingesammelt wird. Für eine größtmögliche Durchführungsobjektivität sorgt dabei ein Testleitfaden, der den Ablauf der Erhebung regelt; die Auswertungsobjektivität wird durch die Erstellung des Fragebogens mithilfe des Programms EvaSys erfüllt, da auf diese Weise die ausgefüllten Fragebögen eingescannt und von einem Programm ausgelesen werden, sodass die Auswertung der geschlossenen Items automatisch erfolgt bzw. direkt die Rohdaten in Zahlenform vorliegen, wodurch die Fehleranfälligkeit im Vergleich zu einer manuellen Dateneingabe verringert wird. Diese Rohdaten werden für statistische Auswertungen mit der Software ‚SPSS‘ genutzt und dadurch Aussagen zu Häufigkeiten, Konstruktwerten, Entwicklungen über die Messzeitpunkte in Bezug auf die Konstrukte, Korrelationen zwischen diesen etc. möglich, sodass Antworten auf die Forschungsfragen gegeben werden können.⁷ Die offenen Fragen der Fragebögen werden mit der ‚Zusammenfassenden qualitativen Inhaltsanalyse‘ nach Mayring analysiert und daraus induktiv Kategorien zur Einordnung der Schülerantworten abgeleitet. Weiterhin können die Anzahl der Nennung dieser Antwortkategorien, d.h. deren Häufigkeiten, ausgewertet und z.B. in Diagrammform dargestellt werden.

In Bezug auf die Fragestellung der Wirkungsüberprüfung eines linguistischen Schülerlaborprojektes stellen Fragebögen ein empirisch erprobtes, durchführungs- und auswertungsökonomisches Erhebungsinstrument dar, das auf Grund seiner Adaptionmöglichkeiten auf die beispielhaft skizzierte deutschdidaktische Fragestellung angepasst werden kann und Aussagen über die erhobenen Konstrukte zulässt, sodass daraus Schlüsse für eine aktivierende Gestaltung von Grammatikunterricht möglich sind. Dabei überwiegen die Vorzüge des Verfahrens gegenüber möglichen Einschränkungen hinsichtlich der erhobenen Daten, die beispielsweise weniger flexibel auf einzelne Schülerinnen und Schüler eingehen und Items vorgeben, ohne dass die Testteilnehmerinnen und -teilnehmer Themen vorschlagen bzw. selbst ansprechen könnten. Vor allem vor dem Hintergrund bisheriger Wirksamkeitsstudien zu Schülerlaboren, die ebenfalls vorrangig quantitativ ausgerichtete Fragebogenskalen zur Erhebung von Selbstauskünften der Schülerinnen und Schüler einsetzen, ist die Wahl des Forschungsinstruments vorteilhaft, um an deren Forschungsergebnissen anzuknüpfen bzw. diese auszu-differenzieren.

⁷ Aufgrund der Tatsache, dass zum Zeitpunkt des Verfassens dieses Beitrags die Datenerhebung noch nicht vollständig abgeschlossen war, können hier noch keine Studienergebnisse vorgestellt werden.

Literatur

- Albert, Ruth/Marx, Nicole: Empirisches Arbeiten in Linguistik und Sprachlehrforschung. Anleitung zu quantitativen Studien von der Planungsphase bis zum Forschungsbericht. Tübingen: Narr Francke Attempto 2010.
- Bühner, Markus: Einführung in die Test- und Fragebogenkonstruktion. München: Pearson Studium 2006.
- Eid, Michael/Gollwitzer, Mario/Schmitt, Manfred: Statistik und Forschungsmethoden. Lehrbuch. Weinheim/Basel: Beltz 2013.
- Engeln, Katrin: Schülerlabors: authentische, aktivierende Lernumgebungen als Möglichkeit, Interesse an Naturwissenschaften und Technik zu wecken. Berlin: Logos 2004.
- EvaSys: <https://www.evsys.de/startseite.html> (letzter Zugriff: 01.08.2018).
- Häkchen: <http://www.haekchen.at/haekchen/home.asp> (letzter Zugriff: 01.08.2018).
- Hussy, Walter/Schreier, Margit/Echterhoff, Gerald: Forschungsmethoden in Psychologie und Sozialwissenschaften. Für Bachelor. Berlin/Heidelberg: Springer 2010.
- Krapp, Andreas/Prenzel, Manfred (Hrsg.): Interesse, Lernen, Leistung. Neuere Ansätze der pädagogisch-psychologischen Interessenforschung. Münster: Achendorffsche Verlagsbuchhandlung 1992.
- Krapp, Andreas: Interesse. In: Rost, Detlef H. (Hrsg.): Handwörterbuch Pädagogische Psychologie. Weinheim/Basel: Beltz 2010, 311-323.
- Krapp, Andreas: Konzepte und Forschungsansätze zur Analyse des Zusammenhangs von Interesse, Lernen und Leistung. In: Krapp, Andreas/Prenzel, Manfred (Hrsg.): Interesse, Lernen, Leistung. Neuere Ansätze der pädagogisch-psychologischen Interessenforschung. Münster: Achendorffsche Verlagsbuchhandlung 1992, 9-52.
- Lienert, Gustav A.: Testaufbau und Testanalyse. Weinheim: Beltz 1969.
- Mayring, Philipp: Qualitative Inhaltsanalyse. Grundlagen und Techniken. Weinheim/Basel: Beltz 2008.
- Mummendey, Hans Dieter/Grau, Ina: Die Fragebogen-Methode. Göttingen u.a.: Hogrefe 2008.
- Pawek, Christoph: Schülerlabore als interesselördernde außerschulische Lernumgebungen für Schülerinnen und Schüler aus der Mittel- und Oberstufe. Universität Kiel 2009.
- Porst, Rolf: Fragebogen. Ein Arbeitsbuch. Wiesbaden: VS 2011.
- Porst, Rolf: Fragebogenerstellung. In: Goebel, Hans/Nelde, Peter H./Starý, Zdenek/Wölck, Wolfgang (Hrsg.): Kontaktlinguistik. Contact linguistics. Linguistique de contact. Ein Internationales Handbuch zeitgenössischer Forschung. Berlin/New York: de Gruyter 1996, 737-744.
- Raab-Steiner, Elisabeth/Benesch, Michael: Der Fragebogen. Von der Forschungsidee zur SPSS/PASW-Auswertung. Wien: Facultas 2010.
- Rost, Jürgen: Testtheorie – Testkonstruktion. Bern: Hans Huber 2004.
- Scharfenberg, Franz-Josef: Experimenteller Biologieunterricht zu Aspekten der Gentechnik im Lernort Labor: empirische Untersuchung zu Akzeptanz, Wissenserwerb und Interesse. Universität Bayreuth 2005.

- Sennewald, Nadja/Mandalka, Nicole: Akademisches Schreiben von Studierenden. Die Bielefelder Erhebung zur Selbsteinschätzung der Schreibkompetenzen. In: Preußner, Ulrike/Sennewald, Nadja (Hrsg.): Literale Kompetenzentwicklung an der Hochschule. Frankfurt a.M.: Peter Lang 2012, 143-166.
- Topalović, Elvira/Dünschede, Susanne: Weil Grammatik im Lehrplan steht? Bundesweite Umfrage zum Grammatikunterricht in der Schule. In: Schlobinski, Peter/Oliver Stenschke (Hrsg.): Schulgrammatik – Grammatik in der Schule. Der Deutschunterricht (4) 2014, 76-81.
- Umfrage Online: <https://www.umfrageonline.com/>(letzter Zugriff: 01.08.2018).
- Willems, Ariane S.: Bedingungen des situationalen Interesses im Mathematikunterricht – eine mehrbenenanalytische Perspektive. Münster: Waxmann 2011.
- Zinn, Bernd: Physik lernen, um Physik zu lehren. Eine Möglichkeit für interessanteren Physikunterricht. Kassel: Logos 2008.

Standardisierte Testverfahren

Messung basaler Lesefähigkeiten am Beispiel des Salzburger Lesescreenings (SLS)

1. Vorstellung standardisierter Testverfahren

Unter standardisierten Testverfahren werden all jene Erhebungsverfahren zusammengefasst, deren Instruktionen, Items (insbesondere deren Menge, Reihenfolge, Gestaltung), Antwortmöglichkeiten (zur Berechnung von Rohwerten und Normwerten) sowie Interpretationen nach einem festgelegten, standardisierten Schema erfolgen und wissenschaftlichen Gütekriterien entsprechen.¹ Nur so kann eine objektive Analyse des zu untersuchenden Merkmals sowohl zur Aufklärung eines zeitlichen Verlaufes (*intraindividuelle* Unterschiede) oder eines Vergleiches zwischen Personen (*interindividuelle* Unterschiede) erfolgen.

Die Zahl der allein im deutschen Sprachraum vorhandenen psychologischen Tests² kann auf mehrere Tausend geschätzt werden, sodass im Laufe seines Lebens vermutlich ein jeder schon mit standardisierten Tests konfrontiert war. Zu erkennen sind sie daran, dass die Instruktion wortgetreu wiedergegeben werden muss, sodass der Vortrag oftmals etwas künstlich wirkt. Fragen dürfen nur an vorgegebenen Stellen und Antworten nur nach vorgegebenen Mustern erfolgen. Wengleich dieses Vorgehen manchmal etwas befremdlich wirkt, wird nur so mit

¹ Sind einige Elemente variabel, spricht man von teil- oder halbstandardisierten Verfahren wie z.B. beim qualitativen halbstandardisierten Interview. Hier kann der Interviewer neben den vorgegebenen Fragen in Abhängigkeit vom Gesprächsverlauf und dem Gesprächsverhalten des Teilnehmers oder der Teilnehmerin weitere Fragen ergänzen bzw. umformulieren.

² Die Einteilung der Testverfahren erfolgt nach ganz unterschiedlichen Klassifikationsmerkmalen. Ohne die Vielzahl dieser wiedergeben zu wollen, kann danach unterschieden werden, *welche Merkmale* gemessen (u.a. Intelligenz, Gedächtnis, Aufmerksamkeit, Motivation, u.a.), *auf welche Art* die Merkmale erhoben (Urteilstests, Leistungstests, Deutungstests) und für welche *Fragstellungen* die Tests einsetzbar sind (Eignungstests, Schultests, u.a.). Eine weitere Einteilung bietet die Differenzierung in Hard Skills/Fachkompetenz (vorwiegend Leistungsmerkmale) oder Soft Skills (soziale Kompetenzen, Neigungen, Interessen, Persönlichkeitsmerkmale im engeren Sinne).

Sicherheit gewährleistet, dass alle Testteilnehmerinnen und Testteilnehmer (Probandinnen und Probanden) auch dieselben Informationen erhalten und somit eines der Hauptgütekriterien, die Objektivität, erfüllt wird. Weitere Hauptgütekriterien bzw. Qualitätskriterien für standardisierte Testverfahren sind neben der Objektivität auch die Reliabilität und die Validität. Will das Kriterium der Reliabilität erfüllt sein, muss der Test, sofern sich die Fähigkeiten der Probandin oder des Probanden nicht verändert haben, auch bei mehrmaliger Durchführung immer wieder zum gleichen Ergebnis kommen. Als valide wird ein Test hingegen bezeichnet, wenn er tatsächlich nur das misst, was er messen soll.

Für die Qualitätsmessung von standardisierten Verfahren können zusätzlich eine Reihe von Nebengütekriterien berücksichtigt werden. Hierzu zählen u.a. Testfairness, Utilität, Ökonomie, Transparenz, Unverfälschbarkeit, Zumutbarkeit, und Verfügbarkeit von Normen.³

Viele standardisierte Testverfahren zeichnen sich zudem dadurch aus, dass mit ihnen Punkte (Rohwertpunkte⁴) ermittelt werden können, die einen Vergleich der ermittelten Werte mit der Eichstichprobe⁵ erlauben sowie die Zuordnung zu Prozenträngen und T-Werten.

Prozentrang-Skala

Der Prozentrang (PR) gibt Auskunft über die relative Stellung, die eine Probandin oder ein Proband in einer bestimmten Stichprobe einnimmt bzw. wie viel Prozent der Getesteten aus der Normierungsstichprobe eine gleich gute, bessere oder schlechtere Leistung erzielt haben. So bedeutet ein PR von 50, dass jemand mit seinem Testwert 49% der Stichprobe übertrifft und selbst von 50% übertroffen wird (vgl. u.a. Moosbrugger/Kelava 2008, 168).

T-Wert-Skala

T-Werte stellen eine Umwandlung der Prozentränge in normal verteilte Standardwerte dar (vgl. Modell der Gauß'schen Normalverteilung), wodurch eine Normierung auch nicht-normal verteilter Eichstichproben möglich wird. Obwohl gegenüber den Prozentrang-Angaben weniger anschaulich, können sie Unterschiede im gesamten Leistungsspektrum damit statistisch korrekt abbilden.⁶

Die Spitze der T-Wert-Kurve kennzeichnet den T-Wert 50. Schlussfolgernd gelten Leistungen mit einem T-Wert zwischen 60 und 70 (entspricht +1 Standardabweichung) als überdurchschnittlich und mit einem T-Wert zwischen 30 und 40

³ Erläuterungen zu den Nebengütekriterien finden sich in jedem guten Handbuch zur Testpsychologie u.a. in Brickenkamp, Rolf: Handbuch psychologischer und pädagogischer Tests. Göttingen et al.: Hogrefe 2007.

⁴ Als Rohwerte werden die noch unnormierten Werte des gemessenen Merkmals bezeichnet.

⁵ Als Eichstichprobe wird die Stichprobe bezeichnet, die bei der Gewinnung von Normwerten für die Testinterpretation herangezogen wurde.

⁶ Aus T-Werten können gleichfalls Durchschnittswerte berechnet werden, was für Prozentränge nicht erlaubt ist, da das Datenniveau jener diese Operationen nicht zulässt.

(entspricht -1 Standardabweichung) als unterdurchschnittlich (vgl. Übersicht in Lenhard 2013, 75). Da die Qualität und der Stand der Entwicklung von Test zu Test sehr unterschiedlich sein können, ist es obligatorisch eine eingehende Analyse existierender Testverfahren im Vorfeld des Einsatzes durchzuführen. Was hierbei berücksichtigt werden muss, insbesondere wenn Testverfahren innerhalb eines Forschungsprojektes eingesetzt werden, soll nun am Beispiel von für die Deutschdidaktik besonders relevanten Lesetests erläutert werden.

1.1 Standardisierte Testverfahren zur Messung basaler Lesefähigkeiten

Mit dem Ziel, objektive und zuverlässige Auswertungsverfahren zur Beobachtung der Mikro- und Makroprozesse, zur Überprüfung von Lernergebnissen, zur Erfolgskontrolle von Förderprogrammen sowie zur Ermittlung und Beschreibung der beim Kind beobachteten Schwierigkeiten zu erfassen, wurden Mitte des 20. Jahrhunderts – analog zu den Überlegungen zum Leseverständnis – die ersten standardisierten Lesetests entwickelt. Erkenntnisse aus diesen sollten nicht nur die Unterrichtsplanung (vgl. Baumann 2003, 869), sondern auch Selektions- und Förderentscheidungen unterstützen. Letztgenannte basierten zuvor lediglich auf subjektiven Urteilen der Lehrperson, die nun durch eine wissenschaftlich fundierte Analyse (Ausweisung der Ergebnisse in Prozenträngen und T-Werten), die u.a. den Vergleich individueller Leistung mit einer normierten Bezugsgruppe umfasste, abgesichert werden konnten.⁷

Aufgrund ihrer starken Orientierung am Legastheniekonzept, das später scharf kritisiert wurde, stagnierte die Testentwicklung jedoch bis in die 90er Jahre (vgl. Ingenkamp 1985). Erst die wichtigen Erkenntnisse über den Leseprozess, den Leseerwerb und die dabei auftretenden Schwierigkeiten durch Arbeiten der Schriftspracherwerbsforschung (vgl. Baumann 2003, 873) sowie die Veröffentlichung der PISA-Ergebnisse rehabilitierte sowohl die Testentwicklung als auch die Testanwendung (vgl. Rackwitz 2009, 42).

Zum heutigen Zeitpunkt können über die Testzentrale des Hogrefe-Verlags 27 Lesetests für die unterschiedlichsten Jahrgangsstufen abgerufen werden.⁸ Für eine

⁷ Wie stark Lehrerurteile und Ergebnisse von standardisierten Tests zuweilen auseinanderliegen können, belegen die Fragebogenauswertungen bei PISA (vgl. Artelt/Sta-
nat/Schneider/Schiefele 2001) und IGLU (vgl. Valtin/Hornberg et al. 2010). So wurden in der PISA-Studie lediglich 11,4% der schwächsten Schülerinnen und Schüler auch als diese von den Lehrkräften identifiziert. Bei fast 90% wurde der hohe Förderbedarf dementsprechend nicht erkannt. Ein ähnliches Ergebnis findet sich auch in den IGLU-Daten wieder. Die Leistungen von mehr als einem Drittel der Schülerinnen und Schüler mit dem stärksten Förderbedarf wurden von den Lehrkräften mit „befriedigend“ und besser bewertet.

⁸ <http://www.testzentrale.de>

wesentliche Kategorisierung dieser, wenngleich auf Grund der Mehrdimensionalität des Leseprozesses nicht immer trennscharf umzusetzen⁹, schlägt Baumann (2003) die Unterscheidung sowohl in Lesefertigkeitstests, welche die Erfassung der basalen Lesefähigkeit anstreben, als auch Leseverständnistests, welche die Fähigkeiten des sinnentnehmenden Lesens oberhalb der Wortebene zu erheben versuchen, vor.

1.2 Screeningverfahren vs. Individualtests

Im Vorfeld des Einsatzes von Lesetests gilt es dementsprechend, sich mit den unterschiedlichen Zielsetzungen der Verfahren, die im Wesentlichen die Vorgehensweise des Erhebungsinstruments bestimmen, vertraut zu machen. Unterschieden werden hierbei Screeningverfahren von Individualtests.

Screeningverfahren wie z.B. das *Salzburger Lesescreening (SLS)* für die Klassenstufen 5-8 von Auer/Gruber/Mayringer/Wimmer eignen sich auf Grund ihrer leicht zu handhabenden Durchführung und Auswertung besonders für das Erstellen von allgemeinen Aussagen zu größeren Probandenzahlen. Hier bieten sie einen schnellen Überblick über die Leistungsfähigkeit einer ganzen Klasse und identifizieren dabei lesestarke und -schwache Schülerinnen und Schüler. Screeningverfahren stellen damit einen ersten Schritt in den diagnostischen Prozess dar.

Eine individuelle, qualitative Analyse der Schülerantworten bieten Individualtests wie z.B. der *Zürcher Lesetest II (ZLT II)* für die 2. bis 6. Klasse¹⁰ von Petermann und Daseking (2012). Im Rahmen von Einzelsitzungen mit der Schülerin oder dem Schüler wird die Lesefertigkeit mithilfe von sieben Untertests sowohl über Lesegeschwindigkeit, Lesegenauigkeit, Leseflüssigkeit als auch über auditive Merkfähigkeit und phonologische Bewusstheit im weiteren Sinne erhoben.¹¹ Die Auswertung der Ergebnisse dieser teilweise sehr aufwendigen Sitzungen (ca. 25-35 Min. pro Schülerin oder Schüler) bietet im Rahmen von prozessorientierter Lernforschung die Grundlage für Analysen zu den angewendeten Strategien und zur Erstellung von individuellen Fehlerprofilen.

⁹ Lesefertigkeitstests z.B. zur Lesegeschwindigkeit versuchen, Rateprozesse durch den Rückgriff auf Verständnisfragen zu vermeiden.

¹⁰ Normtabellen liegen bis zur 8. Jahrgangsstufe vor.

¹¹ Phonologische Bewusstheit im weiteren Sinne umfasst die Analyse- und Synthesefähigkeit größerer Einheiten, d.h. Wörter und Silben, während sich die phonologische Bewusstheit im engen Sinne lediglich auf einzelne Laute bezieht (vgl. Rackwitz 2009, 2).

1.2.1 Das Salzburger Lesescreening 5-8: Testmaterial und Testdurchführung

Das *Salzburger Lese-Screening für die Klassenstufen 5 bis 8 (SLS 5-8)* zur Erhebung von basalen Leseverständnisleistungen wurde bereits mehrfach in Bildungstudien eingesetzt.¹²

In einer Liste mit inhaltlich einfachen Sätzen, die auf das Wissen der Schülerinnen und Schüler abgestimmt wurden (vgl. SLS Handbuch 2005, 9), soll innerhalb einer dreiminütigen Bearbeitungszeit der Wahrheitsgehalt dieser Sätze bewertet werden. Damit wird sowohl die Lesegenauigkeit als auch der Automatisierungsgrad des Lesens über die zu messende Lesezeit expliziert.

Während der Sommerferien ist schulfrei.	✓	✗
Auf Kirschbäumen wachsen Kichererbsen.	✓	✗
Hunde und Katzen gehören zur Familie der Fische.	✓	✗
Mit Turnschuhen kann man besser laufen als mit Gummistiefeln.	✓	✗

Abb. 1: Beispiel aus dem Testheft des Salzburger Lesescreening 5-8, Form A1

Die Abfolge richtiger und falscher Sätze ist auf den Testbögen in einem ausgeglichenen Verhältnis zufällig gewählt. Dabei steigt ihre Länge schrittweise, ohne jedoch die Schülerinnen und Schüler mit schwachen Leseleistungen und/oder Migrationshintergrund zu überfordern.

1.2.2 Ausgangsdaten und Ergebnisse

Nach der Durchführung des SLS werden mithilfe einer Auswertungsfolie die korrekt beurteilten Sätze der jeweiligen Schülerin oder des jeweiligen Schülers gezählt. Die Auswertungsfolie ist dabei so beschaffen, dass das Testheft über die Folie hinausragt. So können am Rand falsch beurteilte oder ausgelassene Sätze markiert und der letzte bearbeitete Satz gekennzeichnet werden. Für die Auswertung und den anschließenden Vergleich der Ergebnisse ist die Ermittlung eines Rohwertes (RW) nötig. Hierfür werden die falsch bearbeiteten und die ausgelassenen Sätze von der Gesamtzahl der bearbeiteten abgezogen (vgl. Auer et al. 2005, 18).

Gelingt es beispielsweise einer Schülerin einer fünften Klasse an der Hauptschule nach drei Minuten Bearbeitungszeit 20 Sätze auf ihre Richtigkeit zu prüfen, würde sie theoretisch 20 Rohwertpunkte erhalten. In der Kontrolle zeigt sich jedoch, dass drei der Sätze falsch sind und einer ausgelassen wurde. Schlussfolgernd reduziert sich der Rohwert des Testergebnisses auf 16 und wird ins Testheft eingetragen.

¹² U.a. im Projekt STIKK_4 (Stundenplanverankerung, Individualisierung, Kompetenzstufenorientierung, Klassenübergreifende Organisation, 4 Förderstunden in der Woche für alle), vgl. dazu die Ausführungen in Sigel (2007). Zudem hat das Nationale Bildungspanel (NEPS) bei eigenen Untersuchungen die Testkonstruktionsprinzipien des SLS als Grundlage genommen.

Unter Vorbehalt sollten Ergebnisse bewertet werden, wenn Schülerinnen und Schüler weniger als zehn Sätze geprüft oder mehrere Sätze nicht bearbeitet bzw. fünf Sätze falsch beantwortet wurden. In allen Fällen ist die Aussagekraft der Ergebnisse beeinträchtigt. Ganz abzusehen ist von einer Auswertung, wenn eine komplette Seite überblättert oder mehr als zehn Fehler im Test gewertet wurden (vgl. Mayringer/Wimmer o.J., 14).

Ist der Rohwert richtig ermittelt, kann dieser mit der im Testmanual aufgeführten Normwertskala für den Lesequotienten (LQ-Skala) abgeglichen werden, welche die Ergebnisse einer einzelnen Person zu den Werten einer Vergleichsgruppe derselben Altersgruppe und Schulform in Bezug setzen.

Lesequotient (LQ)-Skala

Die Umrechnung des Rohwerts in einen für die Vergleichsgruppe angemessenen Normwert erfolgt über die dem Test mitgelieferten Normwertskalen. Diese Tabellen mit Vergleichswerten für die Klassen 5-8 der Haupt- und Realschule einerseits und des Gymnasiums andererseits ermöglichen dem Tester das Ablesen des Lesequotienten, welcher Auskunft über die Leseleistung gibt.

Rohwert	Lesequotient (LQ)		
	Gesamt	Hauptschule Realschule	Gymnasium
< 14	< 59	< 57	< 45
14	59	57	45
15	61	59	48
16	62	62	50

Tab. 1: Ausschnitt der Auswertungstabelle vgl. im Manual, 26

Neben den allgemeinen Aussagen zur Entwicklung der Lesegeschwindigkeit erlaubt die Umrechnung der im Lesescreening erbrachten Leistung in die LQ-Werte zudem eine Zuordnung in folgende Kategorien, die den Tester bei der Interpretation der Ergebnisse unterstützt. Die Hauptschülerin aus dem oben genannten Beispiel erreicht demnach mit einem Rohwert von 16 einen Lesequotienten von 62, was sich als außerordentlich schwache Leistung einordnen lässt und weitere Diagnoseschritte einleiten würde

Lesequotient (LQ)	Leistung
> 125	ausgezeichnet
111-125	überdurchschnittlich
90-110	durchschnittlich
75-89	unterdurchschnittlich
< 75	schwach

Tab. 2: Zuordnung des Lesequotienten zur Leseleistung (vgl. Handbuch SLS 2005, 18)

Es empfiehlt sich jedoch, die erhobenen Daten auf Grund der unzureichenden Vergleichbarkeit mit der Eichstichprobe des SLS, die sich ausschließlich aus Schülerinnen und Schülern aus Bayern sowie den österreichischen Bundesländern Salzburg und Oberösterreich zusammensetzt, primär mit dem Leistungsstand der Klasse und der Parallelklassen zu vergleichen.

1.3 Vorteile und Nachteile des Salzburger Lesescreenings

Dem Verfahrensmodus des Screeningverfahrens, wie beim *Salzburger Lesescreening*, intendiert, ist eine globale und möglichst ökonomische Ermittlung der Leseleistung, die sowohl einen allgemeinen Überblick zum Leistungsstand der Klasse bietet, aber auch leseschwache Schülerinnen und Schüler identifiziert. Das SLS bietet sich daher insbesondere für die Erfassung der allgemeinen Ausgangslage z.B. im Vorfeld von (fachbezogenen) Implementationsstudien mit großen Stichproben an.¹³

Durch das Vorliegen zweier Testversionen mit inhaltlich unterschiedlichen, den gestellten Leseanforderungen nach aber vergleichbaren Sätzen («Satzversionen« Form A und Form B) kann der Test bereits nach kurzer Zeit wiederholt werden. Darüber hinaus liegen auch für beide Satzversionen zwei Varianten mit geringfügig veränderter Satzabfolge vor, sodass das Abschreiben von Testergebnissen wesentlich erschwert wird.

Nachteilig beim SLS erweist sich hingegen der unzureichende Vergleich mit der Eichstichprobe. Wie bereits erwähnt, setzt sich diese in wesentlichen Teilen aus Schülerinnen und Schülern aus den Kantonen Oberösterreich und Salzburg zusammen. Nur ein kleiner Teil der Eichstichprobe stammt hingegen aus Bayern, sodass die Repräsentativität der Stichprobe für den deutschen Bildungsraum als eingeschränkt bewertet werden muss.

Darüber hinaus fehlt es, wie in vielen anderen Tests auch, an separaten Normwerten für mehrsprachige Kinder. Diese müssen demnach, ungeachtet ihrer Sprachbiographie, ihre Leistungen mit jenen der deutschen Muttersprachler vergleichen lassen. Vor dem Hintergrund der Ergebnisse aus der Leseforschung wären ebenso geschlechtsspezifische Prozentrangnormen, wie sie im SLS 1-4 ausgewiesen sind, wünschenswert.

Der kritische Blick auf die Validität des Testverfahrens offenbart zudem, dass diese allein auf Übereinstimmungen mit anderen Tests beruht. Wenngleich dies ein gängiges Verfahren zur Ermittlung der Testgütekriterien ist, bietet dieses Vorgehen, so Rackwitz (2009, 43), selbst im Zusammenhang mit dem Abgleich des Lehrerurteils und der Deutschnote keine ausreichende Aussagekraft für die Testgültigkeit.

¹³ Sprachliche Kompetenz als Grundlage für alle Fächer gilt es auch außerhalb des Deutschunterrichts in empirischen Studien zu erfassen. Dementsprechend empfiehlt es sich auch in anderen Fächern, Testverfahren zur allgemeinen Lesefähigkeit einzusetzen, um sprachliche Schwierigkeiten als Grundlage für schlechtere Testergebnisse auszuschließen.

Im Rahmen von Qualifizierungsarbeiten empfiehlt es sich, zuletzt die hohen Kosten des SLS zu berücksichtigen bzw. mit denen von anderen Testverfahren zu vergleichen.

In toto gilt für den SLS, dass er allgemeine Fähigkeiten abfragt, die jedoch insbesondere im Bereich der Förderdiagnostik nur oberflächlich differenzieren. Gezielte Beobachtungen und Dokumentationen von Lernverläufen gelten daher als sinnvolle Ergänzungen zum vorgestellten Testverfahren (vgl. Rackwitz 2009).

1.4 Gütekriterien für den Einsatz des Verfahrens

Die Testgütekriterien sind ausführlich im Handbuch aufgeführt und sollen an dieser Stelle nur gekürzt wiedergegeben werden.

Die Testbewertung anhand der zentralen Testgütekriterien: Reliabilität (Zuverlässigkeit), Validität (Gültigkeit) und Objektivität spricht zunächst für die Empfehlung des Testverfahrens. So ist die Reliabilität des Lesescreenings mit $r_{tt}=.89$ in allen Klassenstufen sehr hoch, die Validität liegt mit $r=.78$ auch im akzeptablen Bereich und mit dem beigelegten Handbuch, welches umfassende Erläuterungen zu Durchführung, Auswertung und Interpretation der Ergebnisse liefert, ist auch gewährleistet, dass der Test immer wieder im gleichen Muster durchgeführt und ausgewertet wird. Ein weiteres wichtiges Kriterium für den Einsatz von Lesetests in Forschungsarbeiten ist das Ökonomie-Kriterium. Hier zeigt sich der SLS mit seinem geringem Bearbeitungs- und Auswertungsaufwand (7 Min. Testinstruktion, 3 Min. Bearbeitungszeit, ca. 5 Min. Auswertungszeit) anderen Testverfahren deutlich überlegen. Die Prüfung des Repräsentativitäts-Kriteriums ergibt jedoch, dass die Eichstichprobe zwar ausreichend groß (zwischen 714 und 850 Schülerinnen und Schüler pro Jahrgangsstufe) hingegen lediglich aus Bayern sowie den österreichischen Bundesländern Salzburg und Oberösterreich stammt. Als Referenzwerte für den Vergleich mit den Ergebnissen von deutschen Schülerinnen und Schülern scheint sie damit eher ungeeignet. Zudem fehlen jegliche Angaben zum Zeitpunkt der Erhebung.

1.5 Literaturempfehlungen

Ingenkamp (1985): Lehrbuch der pädagogischen Diagnostik.

Diese Monographie gibt einen ausführlichen und sehr interessanten Überblick u.a. zur historischen Entwicklung der pädagogischen Diagnostik.

Baumann (2003): Lesetests.

Im Artikel werden die Unterschiede von älteren und neueren Testverfahren anhand von Beispielen gegenübergestellt und bewertet.

Lenhard (2013): Leseverständnis und Lesekompetenz.

Lenhard bietet einen umfassenden und leserfreundlichen Einblick in die Testdiagnostik u.a. mit einer Gegenüberstellung von standardisierten Testverfahren und informeller Diagnostik sowie der Vorstellung und Bewertung zahlreicher Beispiele.

Rackwitz (2009): Diagnose und Förderung.

Kleiner Überblick zu Papier- und Bleistifttests sowie computer- und internetgestützte Verfahren. Darüber hinaus grundlegende Informationen, die bei der Testauswahl berücksichtigt werden sollten.

Rackwitz (2010): Tests unter der Lupe: Können sie halten, was sie versprechen?

In diesem Artikel wird die Zusammenstellung der Gütekriterien von standardisierten Testverfahren kritisch beleuchtet, deren Ausführungen einmal mehr auf die notwendige qualitativ differenzierte Bewertung der quantitativen Ergebnisse z.B. durch die Lehrperson oder den Testleiter verweisen.

2. „Ich kann jetzt besser lesen“ – Konzeptionierung, Transfer und Evaluation eines Recreational Reading Programs in der Sekundarstufe I¹⁴

2.1 Zielstellung der Forschungsarbeit

Die in den letzten Jahren publizierten Ergebnisse der Schulleistungsstudie PISA attestieren den deutschen Schülerinnen und Schülern für den Bereich der Lesekompetenz einen anhaltenden Förderbedarf. Wenngleich der signifikante Leistungszuwachs im Bereich der Lesekompetenz der aktuellen schulischen Leseförderung erste Erfolge attestiert, besitzen durchschnittlich immer noch 14,5% der befragten 15-Jährigen eine unzureichende Lesekompetenz.

Der differenzierte Blick auf die nationalen Ergebnisse zeigt, dass der Handlungsdruck insbesondere bei männlichen Lernern¹⁵ an der Hauptschule¹⁶, der bereits auch in den vorangegangenen Studien diagnostiziert wurde, fortbesteht. Damit zählen die Unterstützung von Erwerbsprozessen und die Förderung von Lesekompetenz jener Schülerschaft zu den zentralen Bildungszielen der schulischen Aus- und Weiterbildung.

Sich dieser Zielsetzung verpflichtet, konzentrierte sich die Forschungsarbeit auf die Förderung von Lesekompetenz. Die geplante Fördermaßnahme sollte sich dabei jedoch nicht allein auf die Vermittlung technischer Lesefertigkeiten konzentrieren, sondern gleichfalls einen engagierten Leseförderunterricht gestalten, der auch eine Steigerung der Lesemotivation, des lesebezogenen Selbstkonzepts und des freizeithlichen Leseverhaltens fokussiert.

Für die Umsetzung dieser Vorstellungen wurden auf das aus dem angloamerikanischen Raum entlehnte *Guided Independent Reading* zurückgegriffen. Da internationale Förderprogramme jedoch an die hiesige Unterrichtspraxis mit ihren vorherrschenden curricularen, personalen und materiellen Rahmenbedingungen angepasst werden müssen, entstand die Idee zum Projekt *Meine LeseZeit*.

Mit dem Projekt *Meine LeseZeit* wurde eine ganzheitliche Förderung angestrebt, die neben einer zweiwöchigen Einführungsphase drei zentrale Projektbausteine umfasste.

Der erste Projektbaustein zur literarischen Kontextualisierung (*Around Literature*) beinhaltete die Vorbereitung und Begleitung der Viellesephasen. Speziell bedeutete dies einerseits die Unterstützung bei der Auswahl der Buchlektüre so-

¹⁴ Weitere Ausführungen zur Forschung beziehen sich auf das durchgeführte Dissertationsprojekt, welches 2014 im Schneider Verlag unter dem angegebenen Titel veröffentlicht wurde.

¹⁵ Während der Anteil der Mädchen unter Stufe zwei aktuell nur noch 9% beträgt, ist die Gruppe der Jungen mit 20% nach wie vor besonders groß (OECD 2014, 215).

¹⁶ Immer noch lesen 43,8% der Hauptschülerinnen und -schüler auf der untersten Kompetenzstufe (vgl. Hohn et al. 2013, 237).

wie während des Leseprozesses und andererseits die reflexive Auseinandersetzung mit den unterschiedlichen historischen und gesellschaftlichen Bezügen von Literatur.

Der zweite Baustein (*Book Talks*) schaffte den Rahmen für unterschiedliche didaktische Settings, in denen die Schülerinnen und Schüler Möglichkeiten erhielten, sich über das Gelesene auszutauschen.

Zentrum des Leseförderprojektes war jedoch der dritte Baustein. Dieser umfasste wöchentlich stattfindende zwanzigminütige Lesezeiten (*Uninterrupted Silent Reading*), in deren Mittelpunkt die selbstbestimmte Auswahl einer Lektüre und ein frei von weiterführenden Leseaufgaben im eigenen Lesetempo erfolgender Leseprozess stand (vgl. Dube 2014b).

2.2 Forschungsdesign

Dem Ziel folgend, die Wirksamkeit des auf dem Konzept des *Guided Independent Reading* beruhenden Projektes *Meine LeseZeit* auf die unterschiedlichen Dimensionen von Lesekompetenz zu untersuchen, wurde ein Forschungsdesign für eine Langzeitstudie (Longitudinalstudie) mit drei Erhebungszeitpunkten entworfen. Die Implementationsstudie mit Projekt-Kontrollgruppen-Design umfasste dabei einen Erhebungszeitraum von 12 Monaten (Sept. 2010 - Sept. 2011). In jener Zeit nahmen die sechs Projektklassen (PG) neben vier wöchentlichen Deutschstunden auch zweimal in der Woche am Projekt *Meine LeseZeit* teil, während sieben Kontrollklassen (KG) ausschließlich sechs Stunden den Deutschunterricht besuchten.

Aufgrund des Bestrebens den ganzheitlichen Ansatz des Leseförderprojektes auf seine Wirksamkeit zu überprüfen, mussten eine Vielzahl von Variablen sowohl auf kognitiver Ebene (Leseverstehen, Dekodier-/Lesegeschwindigkeit, Wortschatz, allgemeine Sprachkompetenz), auf motivationaler Ebene (Lesemotivation und lesebezogenes Selbstkonzept) als auch auf habitueller Ebene (Lesesozialisation, Leseverhalten in der Freizeit) im Forschungsdesign berücksichtigt werden. Das Tableau zur Datenerhebung ist dementsprechend breit und umfasst neben quantitativen (Fragebogen, standardisierte Lesetests) auch qualitative Erhebungsinstrumente (halbstandardisierte Interviews).

Für jede der Zielvariablen galt es im Vorfeld abzuwägen, welche der Instrumente und Methoden sich am besten für die Erfassung dieser eignen. Die Entscheidungen für den SLS zur Erfassung der Variable Lesegeschwindigkeit möchte ich nachfolgend beispielhaft skizzieren.

2.3 Warum passt das Salzburger Lesescreening zu meiner Forschungsfrage?

Ein Indikator, der Auskunft über den Grad der Beherrschung basaler Lesefähigkeiten gibt, ist die Lesegeschwindigkeit. Als Maß für einen sicheren und schnellen Zugriff auf das lexikalische Gedächtnis während der schriftlichen Sprachrezeption (vgl. Ritz-Fröhlich 1973) bildet sie neben der Lesegenauigkeit das Kernelement der Leseflüssigkeit, welche als bedeutungstragende Differenzierungsvariable die guten von den schlechten Lesern trennt (vgl. Landerl/Wimmer 2008). Schwierigkeiten im lexikalischen Zugriff werden dementsprechend neben der Häufung von Lesefehlern auch in einer verminderten Lesegeschwindigkeit explizit (vgl. Tischler et al. 2013, 215), welche das Speichern von Sequenzen oder Zusammenhängen im Arbeitsgedächtnis¹⁷ erschwert.

Aufgrund der Stichprobengröße von N=353 Schülerinnen und Schülern aus der fünften Jahrgangsstufe aus einer Oberhausener und zwei Essener Gesamtschulen wurden Individualtests bereits zu Beginn ausgeschlossen. Zudem sollten die gewählten Verfahren keinem förderdiagnostischen Zweck dienen, sondern einen Überblick zum Leistungsstand der Lernenden bieten. Die Gruppe der in Frage kommenden Screeningverfahren für die weiterführende Schule umfasste anschließend neben dem *Salzburger Lesescreening* für die Klassen 5-8 (SLS 5-8), noch den *Salzburger Lese- und Rechtschreibtest SRLT 2*¹⁸ und *Ein Leseverständnis für Erst- bis Sechstklässler* (ELFE 1-6).

Den Ausführungen zur Qualität von standardisierten Testverfahren folgend, wurden die in Frage kommenden Lesetests im Hinblick auf die Erfüllung ihrer Haupt- und Nebentestgütekriterien miteinander verglichen.

Dieser strukturelle Vergleich ergab, dass der SLRT konstruktionsbedingt nicht als Gruppentest durchgeführt werden kann und die Eichstichprobe bereits veraltet ist.¹⁹ Der direkte Vergleich mit ELFE 1-6 brachte die Erkenntnis, dass sowohl bessere Werte der Reliabilität als auch der Validität zu Gunsten des SLS

¹⁷ Dieses hat nach LaBerge und Samuels (1974) schnell die Grenzen seiner Kapazität erreicht, wenn die Informationsverarbeitung im Übersetzen von visueller Information in semantischen Gehalt nicht simultan und automatisiert abläuft. Sinkende Konzentration und die Gefahr, gedanklich von einem Textinhalt abzuschweifen, welche nur durch ein wiederholtes Lesen aufgefangen werden kann, sind die Folge eines verlangsamten Rezeptionsvorgangs (vgl. Scheerer-Neumann et al. 1978; Hunziker 2006, 106).

¹⁸ Der *SLRT* stellt einen klassischen Lesetest dar, bei dem Listen von Wörtern und Pseudowörtern sowie Texte gelesen werden müssen. Das Hauptkriterium ist dabei die Leszeit.

¹⁹ Analog zum SLS fehlen auch beim SLRT die Angaben zur Erhebungszeit der Eichstichprobe. Da das Manual aber bereits aus dem Jahr 1997 stammt, kann davon ausgegangen werden, dass die Eichstichprobe sich nur noch eingeschränkt für Vergleiche eignet.

vorliegen.²⁰ Neben besseren Haupttestgütekriterien überzeugte der SLS zudem durch seine ökonomische Handhabung in Durchführung und Auswertung, dem Vorliegen von zweifachen A- und B-Versionen, die insbesondere in Longitudinalstudien mit Projekt-Kontrollgruppen-Design relevant sind sowie dem gut verständlichen und umfangreichen Handbuch. Zudem unterstützten Berichte aus vorangegangenen Forschungsarbeiten die Entscheidung, den SLS auch im geplanten Vorhaben in die Testbatterie mit aufzunehmen.

2.3.1 Einsatz des SLS im Forschungsprojekt „Meine LeseZeit“

Bereits beim Austeilen der Testhefte zeigten sich viele Schülerinnen und Schüler durch die Länge und die kurze Bearbeitungszeit demotiviert. Erklärungen, dass es um das konzentrierte Arbeiten unter Zeitdruck gehen würde und auch die Lehrpersonen Schwierigkeiten hätten, in der vorgegebenen Zeit alle Sätze auf ihren Wahrheitsgehalt zu überprüfen, trugen jedoch dazu bei, dass sich alle Heranwachsenden nach kurzer Verwirrung offen und interessiert am Test zeigten. Geschult in der Testinstruktion konnten anschließend weitere Fragen beantwortet und gemeinsam mit den Schülerinnen und Schülern einzelne Beispiele besprochen werden. Die Probandinnen und Probanden zeigten keine Schwierigkeiten die Instruktionen nachzuvollziehen und die Beispiele zu lösen. Erwartungsgemäß konnte der Test somit bereits nach kurzer Bearbeitungszeit wieder eingesammelt werden.

2.3.2 Auswertung des Forschungsvorhabens „Meine LeseZeit“ mit dem SLS

Es wird erwartet, dass die Implementierung des Projekts *Meine LeseZeit* in das schulische Curriculum zur Reduzierung der Zahl disfluen-ter LeserInnen führt: Das Projekt soll zur Verbesserung basaler Lesefähigkeiten im Bereich Vorwissen (Sach-, Text- und Weltwissen), im Bereich grundlegendes Textverstehen und im Bereich Lesegeschwindigkeit beitragen. (Dube 2014a, 121)

Vor dem Hintergrund dieser Hypothesenprüfung wurden die Testhefte entsprechend des bereits dargestellten Prozederes im Anschluss an allen drei Erhebungszeitpunkten (Pre-, Post- und Follow-Up-Test) klassenweise ausgewertet und anschließend für den Projekt-Kontrollgruppen-Vergleich zusammengeführt.

²⁰ Während die die Retestrelia- bilität beim SLS $r_{tt}=.89$ beträgt, liegt sie bei ELFE mit $r_{tt}=.85$ darunter. Zudem erfolgte die Validitätsprüfung für ELFE für die Klassen fünf und sechs ausschließlich auf dem Lehrerurteil. Wenngleich die Validität mit $r=.84$ in Klasse 5 noch sehr hoch ist, kann sie in Klasse 6 mit $r=.45$ nicht mehr vertreten werden. Es muss daher stark gezweifelt werden, dass ELFE in der Klasse sechs noch die Lesegeschwindigkeit misst.

	Projektgruppe RW-Mittelwert (SD)	Kontrollgruppe RW-Mittelwert (SD)
Pretest (t1)	21,75 (5,527)	25,27 (5,555)
Posttest (t2)	27,55 (6,048)	28,83 (5,693)
Follow-Up-Test (t3)	30,31 (6,273)	30,33 (5,860)

Tab. 3: Projekt-Kontrollgruppen-Vergleich

Bei der Auswertung aller Daten zeigte sich, dass sich die Projektgruppe im Mittel vom Pre- zum Posttest um $M=5,80$ und vom Pretest zum Follow-Up-Test sogar um deutliche $M=8,56$ Rohwertpunkte verbessern konnte, während die Kontrollgruppe bis zur Follow-Up-Testung ihre Lesegeschwindigkeit nur um $M=5,09$ Rohwertpunkte steigerte.

Die Umrechnung dieser Rohwerte in LQ-Werte veranschaulichte anschließend, dass die Projektgruppe ihre zunächst weit unterdurchschnittliche ($LQ=74$) auf eine durchschnittliche Lesegeschwindigkeit ($LQ=89$) verbessern konnte.

Auf den ersten Blick erscheint die deutlich positive Veränderung des Rohwerts der Projektgruppe die Effektivität der Intervention zu bestätigen, dennoch lassen Mittelwertdifferenzen alleine keine validen Aussagen darüber zu, inwieweit die Ergebnisse in der Projektgruppe und in der Kontrollgruppe durch den Unterricht beeinflusst wurden. Zwar liefern die durch standardisierte Tests generierten Mittelwert-Ergebnisse einen Hinweis auf die Wirksamkeit der Intervention, müssen aber durch weitere statistische Verfahren abgesichert werden. Deshalb gilt es zum Beispiel mittels der zweifaktoriellen Varianzanalyse mit Messwiederholung Interaktionseffekte bzw. Wechselwirkungen zwischen den Variablen aufzudecken und hinsichtlich ihrer Signifikanz zu prüfen. Da jedoch auch Signifikanzwerte noch nichts über die Wirksamkeit einer Methode aussagen, sondern lediglich prüfen, ob die Ergebnisse nicht zufällig zustande gekommen sind, sollten sich Berechnungen zur Effektstärke anschließen.²¹

Diese weiterführenden Untersuchungen ergaben, dass die Steigerung der Lesegeschwindigkeit in Folge der Projektteilnahme nicht nur hoch signifikant ($p < .001$), sondern der Einsatz des Vielleseprojekts mit einer Effektstärke von $d=1,02$, insbesondere vor dem Hintergrund der schlechteren Ausgangsbedingungen, als hoch bedeutsam einzuschätzen ist. Dies bedeutet, dass das Bild zur basalen Lesekompetenz der zu untersuchenden Stichprobe mit den Ergebnissen des SLS zur Lesegeschwindigkeit komplementiert und Aussagen über die Wirksamkeit des Förderprogramms im Hinblick auf jene Zielvariable generiert werden konnten.

Zusammenfassend kann dementsprechend festgehalten werden, dass der Einsatz standardisierter Testverfahren im Projekt *Meine LeseZeit* dazu beigetragen hat, die Wirksamkeit der Intervention insbesondere für leseschwache Jugendliche zu belegen.

²¹ All jenes kann an dieser Stelle nicht erläutert werden. Ausführliche Beschreibungen befinden sich jedoch in Dube (2014a).

Literatur

- Artelt, Cordula et al.: PISA 2000. Zusammenfassung zentraler Befunde. Berlin: Max-Planck-Institut für Bildungsforschung 2001.
- Auer, Michaela/Gruber, Gabriele/Mayringer, Heinz/Wimmer, Heinz: SLS 5-8 – Salzburger Lese-Screening für die Klassenstufen 5-8. Manual. Bern: Verlag Hans Huber 2005.
- Baumann, Monika: Lesetests. In: Bredel, U./Günther, H./Klotz, P./Ossner, J./Siebert-Ott, G. (Hrsg.): Didaktik der deutschen Sprache. Ein Handbuch. 2. Teilband. Paderborn et al.: Ferdinand Schöningh 2003, 869-882.
- Bos, Wilfried et al. (Hrsg.): IGLU 2006. Die Grundschule auf dem Prüfstand. Vertiefende Analysen zu Rahmenbedingungen schulischen Lernens. Münster: Waxmann 2010.
- Dijk, Teun van/Kintsch, Walter: Strategies of discourse comprehension. New York: Academic Press 1983.
- Dube, Juliane: „Ich kann jetzt besser lesen“ – Konzeption, Transfer und Evaluation eines Recreational Reading Programs in der Sekundarstufe I. Hohengehren: Schneider 2014a.
- Dube, Juliane: Erfolgreiche Leseförderung für ZweitsprachenlerInnen mit einem Recreational Reading Program. In: German as a foreign language (2014b), 2/2014, 48-65.
- Hohn, Katharina/Schiepe-Tiska, Anja/Sälzer, Christine/Artelt, Cordula: Lesekompetenz in PISA 2012: Veränderungen und Perspektiven. Online unter: http://www.ebf.edu.tum.de/fileadmin/tueds06/www/Mitarbeiter_PISA/Hohn_et_al__2013__Lesekompetenz_in_PISA_2012.pdf (letzter Zugriff: 01.08.2018).
- Ingenkamp, Karlheinz: Lehrbuch der pädagogischen Diagnostik. Weinheim: Beltz 1985.
- LaBerge, David/Samuels, Jay S.: Toward a theory of automatic information processing in reading. *Cognitive Psychology* (1974), 6, 293-323.
- Landern, Karin/Wimmer, Heinz: Development of Word Reading Fluency and Spelling in a Consistent Orthography. An 8-Year Follow-Up. In: *Journal of Education Psychology* (2008), 100.2008, 1, 150-161.
- Lenhard, Wolfgang/Schneider, Wolfgang: ELFE 1-6. Ein Leseverständnistest für Erst- bis Sechstklässler. Göttingen: Hogrefe 2006.
- Lenhard, Wolfgang: Leseverständnis und Lesekompetenz. Grundlagen – Diagnostik – Förderung. Stuttgart: Kohlhammer 2013.
- Linder, Maria/Grissemann, Hans: Zürcher Lesetest (ZLT). Bern: Verlag Hans Huber 2000.
- Mayringer, Heinz/Wimmer, Heinz: Salzburger Lese-Screening 1-4. Handbuch. Online unter: http://www.eduhi.at/dl/Salzburger_Lesescreeing_Handbuch.pdf, (letzter Zugriff: 01.08.2018).
- Moosbrugger, Helfried/Kelava, Augustin: Testtheorie und Fragebogenkonstruktion: mit 43 Tabellen. Heidelberg: Springer 2008.
- Perfetti, Charles A.: Levels of Language and Levels of Process. In: Cermak, L.S. & Craik, F.J.M. (Hrsg.): Levels of processing in human memory. Hillsdale: Erlbaum 1979, 159-181.

- Perfetti, Charles A.: Reading ability. New York: Oxford University Press 1985.
- Petermann, Franz/Daseking, Monika: ZLT-II Zürcher Lesetest II. Bern: Verlag Hans Huber 2012.
- Rackwitz, Rüdiger-Philipp: Diagnose und Förderung: Möglichkeiten und Grenzen von Testverfahren. In: Deutsch 5 – 10 (2009), 1/2009, 18, 42-44.
- Rackwitz, Rüdiger-Philipp: Dialogische Lernbeobachtung statt standardisierter Tests. Wie finde ich heraus, was meinen Schülern fehlt? In: PÄDAGOGIK (2011). Heft 5/2011, 34 – 37.
- Rackwitz, Rüdiger-Philipp: Ist die phonologische Bewusstheit wirklich Voraussetzung für einen erfolgreichen Schriftspracherwerb? Schwäbisch Gmünd: Päd. Hochschule 2009.
- Rackwitz, Rüdiger-Philipp: Tests unter der Lupe: Können sie halten, was sie versprechen? In: Die Grundschulzeitschrift (2010), 24. Jg., 234, 4-9.
- Ritz-Fröhlich, Gertrud: Verbale Interaktionsstrategien im Unterricht: Impuls, Denkanstoß, Frage. Ravensburg: Maier 1973.
- Sigel, Richard/Feneberg, Sabine. Leseförderung und Entwicklung von Lesefreude mit STIKK_4 in den Jahrgangsstufen fünf und sechs. Leitfaden für die lesekompetente Hauptschule. München: Maß 2007.

Lautes Denken

Leise lesen und laut denken. Eine Erhebungsmethode zur Rekonstruktion von ‚Lesespuren‘

1. Vorstellung des Lauten Denkens

Die Erhebungsmethode Lautes Denken ist insbesondere für die Rekonstruktion von Prozessen gut geeignet, die höhere Verstehensebenen betreffen, wie etwa das Lesen von (literarischen) Texten oder die Organisation von Schreibprozessen. Das Laute Denken ermöglicht nach Heine (2005, 165) vor allem Einblicke in Zwischenprodukte, die den Lösungsweg deutlich machen. Gerade diese ‚Lesespuren‘ (Bertschi-Kaufmann 2004, 48) sind es, die für die Rekonstruktion von Verstehensprozessen wichtige Anhaltspunkte liefern. Darüber hinaus bietet das Laute Denken gegenüber der nachträglichen Auswertung von Interviews oder Schülerprodukten, wie zum Beispiel Aufsätzen und Klassenarbeiten, den entscheidenden Vorteil, dass eine zeitnahe Erfassung der bei der Lektüre von (literarischen) Texten verbalisierten Äußerungen der Probandinnen und Probanden möglich ist.

Ursprünglich stammt die Methode aus der Psychologie und wurde dort als Hilfsmittel der Gesprächstherapie eingesetzt. Erstmals wurde das Laute Denken in den 1980er Jahren auch in der Leseverstehensforschung angewendet und gilt heute als anerkanntes Verfahren deutschdidaktischer Forschung zur Erhebung komplexer Kognitionen (vgl. Schreier 2006, 410). In der Schreibforschung gilt das Laute Denken seit den 1980er Jahren als „eine der wichtigsten Datenquellen sowohl für modelltheoretische Konzeptionen als auch für die Untersuchung von Teilprozessen“ (Weinzierl/Wrobel 2017, 223). Inzwischen wird das Laute Denken in der deutschdidaktischen Forschung zur Erfassung von domänenspezifischem Vorwissen sowie zur Erhebung von Schülervorstellungen (vgl. Winkler 2007; Stark 2010a; Meissner 2010) und zur Rekonstruktion von literarischen Verstehensprozessen (vgl. Dannecker 2012) genutzt.

1.1 Funktionsweise des Lauten Denkens

Das Laute Denken ermöglicht einen Zugang zu mentalen Prozessen beim Lesen und wird definiert durch die Instruktion, alles laut auszusprechen, was den Probandinnen und Probanden durch den Kopf geht (vgl. Stark 2010b, 58). Damit ist das Laute Denken dem Obergriff der Introspektion zuzuordnen und bezeichnet sämtliche Aussagen einer Person „durch die Einblicke in ihre eigenen mentalen Aktivitäten gewährt werden“ (Heine 2005, 167). Allerdings sind die Lautdenkprotokolle von metakognitiven (fremdadressierten) Verbalprotokollen begrifflich abzugrenzen: Beide Formen sind den Verbalprotokollen zuzuordnen, unterscheiden sich allerdings hinsichtlich des Grades der Reflexion der eigenen Handlungen bzw. Metakognitivität der Äußerungen. Da Probandinnen und Probanden als Teilnehmende in quasi-experimentellen Settings oftmals aufgefordert werden, ganz bestimmte Aspekte zu erklären, zu interpretieren oder zu hinterfragen und diese Aussagen Ericsson und Simon zufolge grundsätzlich auf dem obersten Verbalisierungslevel anzusiedeln seien (vgl. Ericsson/Simon 1993, 17), gilt es genau zu prüfen, welche Art von Daten man generieren möchte.

Grundsätzlich darf man sich zunächst nicht davon irritieren lassen, dass sich in verschiedenen Forschungsbereichen eine Fülle an Begriffen, wie Denke-Laut-Methode, Gedankenprotokoll, ‚Thinking Aloud Protocol‘, ‚Talk Aloud Interview‘, ‚Think Aloud‘ oder ‚Verbal Protocol‘ finden lässt. Schließlich ist allen Bezeichnungen gemein, dass sie sich auf Produkte lauten Denkens, also auf Verbalisierungen von Probandinnen und Probanden beziehen, die als verbale Daten aufgenommen, transkribiert, ausgewertet und interpretiert werden müssen (vgl. Konrad 2010, 476f.). Dabei gilt es zu bedenken, dass diese Daten keine direkte Kopie der Gedanken der Probandinnen und Probanden darstellen, denn es werden lediglich solche Kognitionen und mentalen Operationen aufgezeichnet, die den Befragten bewusst sind (vgl. Weidle/Wagner 1982, 82). Überdies legt die Bezeichnung Lautes Denken zwar nahe, dass lediglich Gedanken der Probandinnen und Probanden erfasst würden, doch tatsächlich lassen sich auf diese Weise auch Beschreibungen von Gefühlen und Empfindungen festhalten. Bertschi-Kaufmann bemerkt hinsichtlich literaturdidaktischer Fragestellungen treffend:

Den Prozess der Lektüre selber sowie seine Folgen für das weitere Leseverhalten zu beobachten ist selbstverständlich nur mit Einschränkung möglich, denn am Vorgang des Lesens sind nur der Text und das lesende Subjekt beteiligt, er entzieht sich der Beobachtung weitgehend. Wenn Leserinnen und Leser unmittelbar nach ihren Lektüren aber Erinnerertes in eigenen – mündlichen oder schriftlichen – Texten rekapitulieren, geben sie immerhin Einblick in den Vorgang der Adaption und der Sinnkonstruktion und zeigen damit Spuren, die der eben vollzogene Leseprozess hinterlassen hat. (Bertschi-Kaufmann 2004, 48)

Das Laute Denken ist insofern als anspruchsvolles Erhebungsverfahren einzuschätzen, da es eine reflektierte Aufmerksamkeit hinsichtlich seiner Anwendung verlangt. Allerdings eröffnet dieses Verfahren bedeutende Chancen für die Rekonstruktion von Lektüre- und Verstehensprozessen, da es bezüglich des Standardisierungsgrads als offene Verbalisationsform einzuschätzen ist. Im Gegensatz zu

standardisierten Fragebögen mit geschlossenen Antwortformaten (s. Betz in diesem Band), wird den Probandinnen und Probanden beim Lauten Denken die Chance zur Äußerung eigener Gedanken und Interpretationen gegeben (vgl. Haußer 1994, 66). Überdies ermöglichen verschiedene Varianten des Lauten Denkens, die Verbalisierungen der Probandinnen und Probanden zu unterschiedlichen Zeitpunkten im Verlauf der Lektüre zu erfassen. Entweder nach jedem Satz (*sentence-by-sentence-talking*), nach bedeutenden Textstellen (*selective-talking*) oder nach dem Lesen des gesamten Textes (*after-the-fact-talking*). Für die Auswahl des Erhebungszeitpunktes kommt es wiederum auf das Forschungsinteresse an. Die fokussierte Variante des *sentence-by-sentence-talking* wäre zum Beispiel für die Überprüfung einer bestimmten vorab theoretisch hergeleiteten Forschungsfrage, wie etwa die Frage nach der Bedeutung von Vorhersagen über das weitere Geschehen im Text, geeignet (vgl. Olson/Duffy/Mack 1984, 259). Damit beweist das Laute Denken auch seine Einsatzmöglichkeit als Erhebungsverfahren für quantitativ angelegte Studien. Die Unterbrechung nach bedeutenden Textstellen (*selective-talking*) wäre indes geeignet, um etwa Rezeptionsweisen bestimmter Sinnabschnitte eines literarischen Textes genauer untersuchen zu können. Die Variante, die sich auf die Lektüre des gesamten Textes bezieht (*after-the-fact-talking*), ist sinnvoll, wenn es beispielsweise um die Rekonstruktion von Prozessen ästhetischer Erfahrung geht, die ein Verständnis des ganzen Textes voraussetzt. Nicht nur im letztgenannten Fall kann es erforderlich sein, die Probandinnen und Probanden mit standardisierten Aufforderungen, wie „Was denkst Du?“, an das Laute Denken zu erinnern.

Bei der Durchführung des Lauten Denkens ist unbedingt darauf zu achten, dass die Methode mit einem standardisierten Instruktionstext eingeleitet wird. Dies ist erforderlich, um Verzerrungen zu vermeiden und die Vergleichbarkeit der Fälle zu gewährleisten. Beispielsweise könnte man dies sicherstellen, indem man einen Anleitungstext vorab aufnimmt und den Probandinnen und Probanden jeweils vorspielt. Zudem hat es sich als sinnvoll erwiesen, die Methode mit einer Aufgabe aus einem anderen Gegenstandsbereich, zum Beispiel mit einer Rechenaufgabe, vorzuführen und die Probandinnen und Probanden auf diese Weise die Methode vorab üben zu lassen (vgl. Dannecker 2012, 126). Damit wird sichergestellt, dass die Probandinnen und Probanden die Aufforderung, während des Lesens laut zu denken, verstanden haben. Die Wahl eines anderen fachlichen Gegenstandes ist angeraten, da die Probandinnen und Probanden andernfalls dazu neigen, die Lese- und Verbalisierungsweise des Versuchsleitenden zu imitieren.

1.2 Ausgangsdaten und Ergebnisse

Nach Abschluss der Datenerhebung liegen den Forschenden zunächst Audio-Daten vor, die es zu transkribieren gilt. Je nach Fragestellung kann auf eine vollständige Transkription und die Berücksichtigung prosodischer und phonetischer Merkmale zugunsten einer Transkription ausgewählter Passagen im Wortlaut verzichtet werden. Allerdings kann es sinnvoll sein, die Transkription um Angaben

zu Gesprächspausen, Betonungen und Sprechweisen zu ergänzen, weil diese Aspekte für die spätere Auswertung bedeutsam sein könnten (zur Transkription s. Exkurs von Heller/Morek in diesem Band).

Es hat sich als hilfreich erwiesen, bereits im Rahmen der Datenerhebung sogenannte Postskripte zu erstellen. Dabei handelt es sich um Forschungsnotizen, die sich der Versuchsleitende im Anschluss an das Laute Denken zur Erhebungssituation (äußere Einflüsse, Ort und Dauer der Erhebung, Beobachtungen etc.) macht. Das Führen eines Forschungstagebuchs, ein Notizheft, das begleitend zu Datenerhebung und -auswertung geführt wird, kann ebenfalls dabei behilflich sein, den Forschungsprozess nachvollziehen und reflektieren zu können.

Auch die Wahl der Auswertungsmethode hängt von der jeweiligen Fragestellung des Forschungsvorhabens ab. Denkbar wäre etwa ein Vorgehen in Anlehnung an die Qualitative Inhaltsanalyse (s. Heins in diesem Band). Hierzu gibt es seitens der Biologiedidaktik einen forschungspragmatischen Vorschlag, der das Datenmaterial sinnvoll strukturieren hilft (vgl. Gropengießer 2005, 175). Gegenüber einem solchen Auswertungsverfahren, das sich dem Material eher hypothesenprüfend nähert, ließe sich aus einer Reihe hypothesengenerierender Forschungsmethoden wählen, die sich dem Gegenstand mit einer explorativ-entdeckenden Haltung nähern. Für die Auswahl ist entscheidend, „dass die Methode ihrem Gegenstand, nämlich dem sozialen Handeln, der Kommunikation derjenigen adäquat ist, die Gegenstand der Forschung sind“ (Bohnsack 2003, 17). Einer qualitativen Forschungslogik folgend wird dabei von der Entwicklung von Hypothesen aus dem Datenmaterial heraus ausgegangen. Die Forschenden verfügen zwar über Vorwissen über den Gegenstandsbereich, doch geht es bei der Auswertung nicht um die Bestätigung vorab formulierter Thesen, sondern um das Entwickeln neuer Hypothesen. Dirks beschreibt demzufolge den Erkenntnisprozess, der sich während der Auswertung ereignet, als explorativen Prozess, der meist durch einen abduktiven Blitz, also die „Zufälligkeit des guten Einfalls“ (Dirks 2007, 45), gekennzeichnet sei. Diese sich überraschend ereignende und den Gegenstand erhellende Erklärung für eine neue Regel bleibt solange provisorisch, bis sie mithilfe des Datenmaterials verifiziert werden kann. Durch das abduktive Schließen soll vermieden werden, dass dem Datenmaterial mit der am wahrscheinlichsten erscheinenden Erklärung begegnet wird.

Die Darstellung der Ergebnisse richtet sich nach der Wahl der Auswertungsmethode. Fällt die Entscheidung auf die Präsentation von Fallrekonstruktionen, so dient es der Anschaulichkeit, jedem Fall ein zusammenfassendes Zitat voranzustellen, das als charakterisierende Gesamtdeutung fungiert. Grundsätzlich ist es ratsam, die Ergebnisse zu visualisieren, so etwa in Form von Tabellen oder Schaubildern. Beim Zitieren aus den Datenprotokollen bietet sich ein Verweis auf die jeweiligen Zeilennummern an. Um den Lesefluss zu erleichtern, können für die Darstellung der Ergebnisse Füllwörter wie ‚ähm‘ oder sprachliche Verschleifungen wie ‚so’n‘ für ‚so ein‘ geglättet werden.

Im Anschluss an die Auswertung der einzelnen Verbalprotokolle kann es sinnvoll sein, fallvergleichende Analysen anzuschließen. Auch wäre zu überlegen, ob man

die Probandinnen und Probanden im Sinne eines ‚stimulated recall‘ mit ihren eigenen Äußerungen konfrontiert, um so zu tieferen Einsichten zu gelangen. Damit würde man das Laute Denken in der Form des Nachträglichen Lauten Denkens verwenden, das in der Unterrichtsforschung häufig zur Kommentierung von Unterrichtsvideos und somit zu einer retrospektiven Selbstbeobachtung und Reflexion des eigenen pädagogischen Handelns genutzt wird.

1.3 Vorteile und Nachteile des Lauten Denkens

Entscheidet man sich für ein qualitativ-empirisches Vorgehen, so muss man sich bewusst sein, dass es durch die Explikation und Interpretation der Datenprotokolle zunächst zu einer Vermehrung des Datenmaterials kommt. Erscheint ein qualitatives Vorgehen zunächst attraktiv, weil nur mit vergleichsweise wenigen Fällen gearbeitet wird, kann sich diese Entscheidung schnell als trügerisch herausstellen. Die Auswertung der Verbalprotokolle, abgesehen von der notwendigen (zumeist vollständigen) Transkription der Daten, nimmt viel Zeit in Anspruch. So sollte man etwa mit einem Zeitraum von drei bis vier Monaten pro Fallrekonstruktion rechnen. Allerdings ist dieser Aufwand hinsichtlich der Tiefe der Einsichten gerechtfertigt.

Zudem bedarf es einer großen Sensibilität und Disziplin seitens der Forschenden bei der Anwendung dieser anspruchsvollen Methode. So ist eine gewissenhafte Durchführung unabdingbar, um qualitativ hochwertige Protokolldaten zu generieren: „Die Datenerhebung muss gut vorbereitet, mehrfach erprobt und gewissenhaft durchgeführt werden“ (Stark 2010b, 78). Dieser Aufwand scheint allerdings ebenfalls gerechtfertigt, da sich so im Sinne der Gegenstandsangemessenheit qualitativ hochwertige Daten generieren lassen, die neue Einsichten ermöglichen. So lässt sich sagen, dass sich die Schwächen der Methode durch eine gewissenhafte und reflektierte Vorgehensweise schließlich in Stärken umwandeln lassen.

1.4 Gütekriterien für den Einsatz des Lauten Denkens

Während bei quantitativen Erhebungsverfahren die Gütekriterien Validität, Reliabilität und Objektivität unbedingt zu berücksichtigen sind, lassen sich diese Kriterien nur eingeschränkt auf qualitativ-empirisch angelegte Studien übertragen. Trotzdem ist es auch bei qualitativer Forschung von großer Bedeutung, forschungsethisch korrekt und für Dritte nachvollziehbar vorzugehen. Grundsätzlich sind das Vorverständnis sowie das methodische Vorgehen offen zu legen. Dazu gehört sich das eigene Vorverständnis bewusst zu machen und zu reflektieren sowie alle Schritte der Erhebung und Auswertung zu dokumentieren. Zudem gilt es, Interpretationen der erhobenen Daten argumentativ abzusichern und ggf. durch den Austausch mit anderen Forschenden (kommunikative Validierung) und/oder durch eine Kombination mit anderen Erhebungs- bzw. Auswertungsmethoden (Triangulation) zu validieren. Dies bezieht die Gegenstandsangemessenheit, die Angemessenheit der Erhebungs- und Auswertungsmethoden sowie die Konsistenz- und Kohärenzprüfung der entwickelten Theorie ein. Zusammenfassend lässt

sich sagen, dass die intersubjektive Nachvollziehbarkeit, also eine Verpflichtung zu Reflexion und Transparenz des Vorgehens, und die Gegenstandsangemessenheit stets zu berücksichtigende Gütekriterien qualitativer Forschung sind (vgl. Dannecker 2009, 122). Zudem liegt es in der Verantwortlichkeit der Forschenden, verschiedene Probleme, die sich im Verlauf der Datenerhebung ergeben könnten, zu antizipieren. So sind Störungen des Versuchsablaufs durch äußere Einflüsse und technische Defekte grundsätzlich zu vermeiden. Zudem spielt das Vertrauen der Probandinnen und Probanden zum Versuchsleitenden bei der Erhebung eine große Rolle. Darum sollte die Erhebungssituation eine vertrauensvolle Atmosphäre bieten. Das Einfühlungsvermögen sowie die Wahrung einer professionellen Distanz bei gleichzeitiger Anzeige von Sympathie und Verständnis tragen dabei zum Vertrauen seitens der Probandinnen und Probanden bei.

Ein häufig berichtetes Problem bei der Erhebung von Daten mit dem Lauten Denken stellt die Leistungsmotivation und -bereitschaft der Probandinnen und Probanden hinsichtlich der Verbalisierungen dar, die sich allerdings nur bedingt beeinflussen lassen. Bei der Durchführung der Erhebung ist daher insbesondere darauf zu achten, dass die Forschenden sich auf die situativen Bedingungen und Befindlichkeiten der Probandinnen und Probanden einzustellen vermögen (vgl. Flick 2002, 97). Dementsprechend ist es von Vorteil, den Probandinnen und Probanden mit einer zugewandten und offen-anererkennenden Haltung zu begegnen. Die Vorstellung der eigenen Person und des Anliegens sowie die Einholung des Einverständnisses zur Aufzeichnung sowie der Hinweis auf Anonymisierung schaffen eine gegenüber Dritten geschützte Atmosphäre. Dies ist vor allem bei einer Datenerhebung im Schulkontext zu berücksichtigen.

Grundsätzlich sollten sich Forschende mit der Kommentierung und Bewertung von Antworten zurückhalten, um die Datenerhebung nicht zu beeinflussen. Außerdem müssen Forschende im Verlauf des Forschungsprozesses ihre eigene Rolle fortwährend reflektieren. Diese so genannte ‚reflektierte Subjektivität‘ kann durch die begleitende Selbstbeobachtung, die Reflexion persönlicher Voraussetzungen, die Gewährleistung einer Vertrauensbeziehung zwischen Forschenden und Probandinnen und Probanden oder etwa auch die Reflexion des Feldeinstiegs realisiert werden (vgl. ausführlich dazu Steinke 2005, 330f.).

1.5 Literaturempfehlungen

Dannecker (2009): „Eigentlich ist es schön zu lesen, nur es ist etwas irritierend“ – Zur Frage der Operationalisierbarkeit literarischen Verstehens.

Dieser Beitrag geht vor allem auf die Herausforderung der Gegenstandsangemessenheit empirischer Forschungszugänge in der Literaturdidaktik ein und stellt die Methode des Lauten Denkens am Beispiel eines Verbalprotokolls einer Oberstufenschülerin zur Lektüre eines Romananfangs vor.

Konrad (2010): Lautes Denken.

In diesem Artikel gibt Konrad einen guten Überblick über die Entstehungsgeschichte des Lauten Denkens sowie deren Anwendung in der psychologischen Forschung.

Stark (2010): Lautes Denken in der Leseprozessforschung.

Dieser Beitrag bietet einen umfassenden und weiterführenden Einblick in die Methode des Lauten Denkens und diskutiert zugleich Grenzen und Chancen des Erhebungsverfahrens kritisch.

Dannecker (2013): Literaturvermittlung und Binnendifferenzierung – empirische Zugänge.

Dieser Band versammelt Beiträge zu empirischen Forschungsprojekten von Absolventen der Deutschdidaktik, die vom Umfang her gut im Rahmen von Bachelor- und Masterarbeiten bewältigt werden können.

Weinzierl/Wrobel (2017): Schreibprozesse untersuchen.

Dieser Beitrag gibt einen guten Überblick über den Einsatz des Lauten Denkens in der Schreibforschung.

2. Beispielprojekt: Lautes Denken zu einem leisen Text. Literarische Texte reflektieren und bewerten – eine empirische Erhebung mit Schülerinnen und Schülern der Sekundarstufe II

Im Folgenden soll nun als Beispiel eine im qualitativen Forschungsparadigma verortete Studie herangezogen werden, die sich der Frage widmete, wie Schülerinnen und Schüler der Sekundarstufe II die Lektüre, Reflexion und Bewertung eines literarischen Textes bewältigen. Ausgehend von den PISA-Ergebnissen, die gezeigt haben, dass deutsche Schülerinnen und Schüler beim Kompetenzbereich ‚Reflektieren und Bewerten von Texten‘ besonders schlecht abschneiden, zielte die Erhebung darauf, zu erforschen, welche Rezeptionsprozesse Schülerinnen und Schüler leisten müssen, um zu einer Reflexion und Bewertung eines ausgewählten literarischen Textes zu kommen.

Dass das Lesen und Verstehen literarischer Texte andere Anforderungen an die Rezipienten stellt als pragmatische Texte, ist seitens der Literaturdidaktik häufig postuliert worden (siehe dazu: Eggert 2002; Kammler 2006; Kämper-van den Boogaart 2005; Zabka 2006; Spinner 2007; Pieper/Kämper-van den Boogaart 2008; Abraham/Masaneck/Winkler 2010). Hinsichtlich einer konkreten Bestimmung dessen, was sich beim Lesen literarischer Texte ereignet, besteht in der deutschdidaktischen Diskussion allerdings keine Einigkeit. Auch der Eintrag im Lexikon Deutschdidaktik zur literarischen Kompetenz verweist darauf, dass der Begriff bisher noch nicht angemessen modelliert worden sei. Zwar gäbe es einige Versuche, ein Modell literarischer Bildung oder auch literarischer Kompetenz auf theoretischer Ebene zu formulieren, doch die Zusammenführung zu einer theoretisch und empirisch fundierten Modellierung stehe noch aus (Pieper 2006, 441). Diesem Desiderat folgen einige quantitativ-empirische Projekte, die sich etwa der Erforschung literarisch kodierter, intendierter bzw. evozierter Emotionen widmen und damit einer kompetenztheoretisch gerahmten Erfassung emotionaler Anteile des literarischen Verstehens beabsichtigen (vgl. Frederking/Brüggemann 2012, 21). Daran anknüpfend untersuchte Steinhauer Dimensionen der Involviertheit beim Lesen literarischer Texte (vgl. Steinhauer 2010, 168). Den Nachvollzug sprachlicher und literarischer Bildung nehmen einige qualitativ-empirisch vorgehende Studien in den Blick (vgl. Rosebrock 2008, Bertschi-Kaufmann 2012; 2014). Ergänzend dazu versteht sich die hier vorgestellte Studie als ein weiterer Beitrag zur Grundlagenforschung in diesem Bereich und nimmt dabei die Lektüreprozesse von Schülerinnen und Schülern der Sekundarstufe II in Bezug auf einen ausgewählten Kurzprosatext in den Blick.

Die Auswertung der im Rahmen dieser Studie erhobenen, verbalen Daten und die fallvergleichenden Analysen konnten zeigen, dass sich die Aneignung eines literarischen Textes höchst individualisiert vollzieht. Da jeder Leseakt eine individuelle, aktuelle mentale Konstruktion darstellt, ergeben sich unterschiedliche Leseerfahrungen und Textlektüren, welche die Rezipienten in Abhängigkeit von ihrem Wissen und ihren Erfahrungen zu unterschiedlichen Reflexionen anregen. So

verfolgen die Lesenden in Abhängigkeit vom literarischen Text unterschiedliche ‚Aneignungspfade‘. Dabei zeigte sich, dass das Wissen „um die spezifischen Spielregeln literarischer Kommunikation“ (Kämper-van den Boogaart/Pieper 2008, 59) ebenso bedeutsam ist wie das Zusammenspiel von Imagination und Metakognition.

Demzufolge lässt sich kein ‚typischer‘ Verlauf von Textbegegnung, -aneignung und -deutung beschreiben. Vielmehr beschreiten die Rezipienten unterschiedliche und individuell geprägte ‚Textaneignungspfade‘, die sich je nach persönlich-emotionaler Disposition und sozial-kultureller Situierung unterschiedlich gestalten. Dabei hat sich ein Wechselspiel zwischen kognitiv-analytischen und emotional-nachvollziehenden Vorgängen als bedeutsam erwiesen. Diese sind zudem abhängig vom jeweiligen Text. Die mentale Vorstellungsbildung stellt dafür ein zentrales Element dar, da sie zu einer gedanklichen Repräsentation des Textes führt. Diese ist wiederum durch die persönlichen Erfahrungen, Erlebnisse und Erinnerungen der Lesenden geprägt. Sie fungiert damit gleichsam als Kontinuum zwischen kognitiv-analytischer und emotionaler Textaneignung.

Ein weiteres zentrales Ergebnis ist, dass die Probandinnen und Probanden dieser Untersuchung die Ebene des Reflektierens und Bewertens nur mithilfe didaktisch-methodischer Unterstützungsangebote erreichen (vgl. Dannecker 2012, 185). Dieses Ergebnis ist insofern als bedeutsam einzuschätzen, da es die herkömmliche Forschungsmeinung, die von einer sich quasi automatisch ereignenden Reflexion des literarischen Texts ausging (vgl. Leubner/Saupe 2008, 20), ergänzen kann. Zugleich liefert dieses Ergebnis wichtige Hinweise für die didaktisch-methodische Gestaltung von Unterrichtsszenarien. So ist nicht nur die motivational-emotionale Bereitschaft und die subjektive Beteiligung, sondern auch die Kommunikation und Bedeutungsaushandlung mit ‚kompetenten Anderen‘ maßgeblich für das Verstehen von und Reflektieren über literarischer Texte. In einer sich anschließenden Studie ließe sich nun überprüfen, ob sich beispielsweise kooperative Lernformen als förderlich hinsichtlich der Anregung von Reflexions- und Wertungsprozessen erweisen.

2.1 Die Wahl der Methode(n)

Während in der Vergangenheit häufig retrospektive Untersuchungen zum Leseprozess durchgeführt wurden und in den meisten Fällen daher der Lektüreprozess nicht mit in die Untersuchung einbezogen wurde (vgl. Odag 2007, 19), bietet das Laute Denken die Möglichkeit, Spuren der Verstehenskonstruktion und der Auseinandersetzung mit einem (literarischen) Text in größtmöglicher zeitlicher Nähe zur Lektüre des Textes zu erfassen. Obschon sich Lektüre, Reflexion und Bewertung eines literarischen Textes auch anhand von schriftlichen Äußerungen von Probandinnen und Probanden nachvollziehen lassen, kam die Aufgabenstellung eines Interpretationsaufsatzes bei der vorliegenden Untersuchung nicht in Frage, da durch dieses Vorgehen einerseits zugleich die Schreibkompetenz der Schülerinnen und Schüler Gegenstand der Untersuchung gewesen und andererseits die Nähe zur Aneignungssituation des Textes nicht mehr gegeben gewesen wäre. Um

die überaus flüchtigen und höchst individualisiert ablaufenden Prozesse bei der Lektüre erfassen zu können, wurde für diese Untersuchung daher das Laute Denken als Erhebungsmethode gewählt.

2.2 Die Planung der Erhebung

Für die hier vorgestellte Untersuchung wurde das Laute Denken um ein halbstandardisiertes Gespräch über den Text und weitere Aufgaben ergänzt, da die Pilotierung zu dieser Studie gezeigt hatte, dass das Untersuchungsdesign hinsichtlich der Einführung der Erhebungsmethode, der Textauswahl und der didaktischen Strukturierung modifiziert werden musste. Dies verweist noch einmal darauf, dass die Erprobung des Erhebungsdesigns von Bedeutung ist (vgl. Dannecker 2009, 126f.).

So hat sich beispielsweise im Vergleich zur Pilotierung, bei der mit einem Romananfang gearbeitet wurde, gezeigt, dass ein abgeschlossener Text sinnvoll ist, um höhere Leseprozesse erfassen zu können. So wurde für die Hauptuntersuchung ein Kurzprosatext gewählt, da dessen ‚strukturelle Einfachheit‘ ermöglicht, dass sich die Rezipienten nur mit wenigen Figuren auseinandersetzen und nur einfache Handlungsstränge verfolgen müssen. Rosebrock zufolge bietet kurze Prosa den Vorteil, dass „mit den kleinen Textsorten [...] die verschiedenen formalen Ausdrucksformen des Epischen sozusagen *en miniature* kennen gelernt und für die Konzentration auf die wunderbare Vielfalt literarischen Ausdrucks genutzt werden“ (Rosebrock 2007, 11) können. Zudem bot die Kürze des Textes die Chance, den Probandinnen und Probanden eine ganzheitliche Erfahrung mit einem abgeschlossenen literarischen Text zu ermöglichen.

Da es den Probandinnen und Probanden der Pilotierung schwerfiel, ohne die Anregungen der Versuchsleiterin zu einer Interpretation des Textes zu gelangen, führte dies zur Konzeption eines dreistufigen Aufgabendesigns, das zunächst die Erfassung spontaner Reaktionen und Vorstellungen zum Text mittels des Lauten Denkens, dann eine textanalytische Auseinandersetzung und schließlich die Stellungnahme zu einer Experten-Interpretation vorsah. Dadurch ergab sich eine didaktische Strukturierung der Erhebungssituation: Im Anschluss an die Textlektüre wurden den Probandinnen und Probanden Fragen zu Textrezeption und Aufgaben zur Analyse des Textes gestellt. Dazu wurde ein Leitfaden konstruiert, ähnlich wie bei einem halbstandardisierten Interview, der eine grobe Strukturierung und ein Mindestmaß an Vergleichbarkeit des literarischen Gesprächs ermöglichte. Zudem bot dieser offene Leitfaden der Versuchsleiterin die Möglichkeit, auf Aspekte, die von den Schülerinnen und Schülern angesprochen wurden, einzugehen. Zugleich konnte so das Kommunikationsbedürfnis der Probandinnen und Probanden berücksichtigt werden. Die Aufgaben und die Experteninterpretation wurden den Probandinnen und Probanden in Form eines Aufgabenblatts vorgelegt. Vor allem die Auseinandersetzung mit einer Experteninterpretation hat sich dabei hinsichtlich der Forschungsfrage als zielführend erwiesen, da die Probandinnen und Probanden dadurch zu einer kritischen Einschätzung ihrer eigenen Auseinandersetzung mit dem Text und ihrer Interpretation aufgefordert wurden.

Eine solche Modifikation sozialwissenschaftlicher Methoden ist nicht nur angesichts des Primats der Gegenstandsangemessenheit empirischer Forschung gerechtfertigt, sondern verweist auch auf die Entwicklung und Emanzipation deutschdidaktischer empirischer Forschungsmethodik. Zwar kann die empirisch forschende Deutschdidaktik auf die Methodendiskussion der empirischen sozialwissenschaftlichen und psychologischen Forschung zurückgreifen, doch ist sie angehalten, dabei eine überlegte und sorgfältige Auswahl von Erhebungs- und Auswertungsmethoden hinsichtlich ihrer genuinen Forschungsfragen anzustreben und gegebenenfalls Methoden im Sinne der Gegenstandsangemessenheit zu modifizieren.

2.3 Durchführung der Erhebung

Die Voruntersuchung wurde mit insgesamt sechs Probandinnen und Probanden durchgeführt. Dabei handelte es sich um drei Schülerinnen und drei Schüler einer Integrierten Gesamtschule im Stadtgebiet Hannovers. Die Datenerhebung erfolgte im Rahmen einer didaktisch strukturierten Erhebungssituation außerhalb des Unterrichts, um die Zahl der Einflussvariablen kontrollieren zu können. Da die Untersuchung im Schulkontext durchgeführt werden sollte, mussten die Gegebenheiten des Feldes, wie beispielsweise die Zugehörigkeit zu einer Schulklasse oder die Bereitschaft zur Teilnahme an der Untersuchung, als unabdingbar angesehen werden. Die Auswahl der Probandinnen und Probanden erfolgte daher unter forschungspraktischen Gesichtspunkten kriteriengeleitet und nicht im strengen Sinne nach der von Glaser und Strauss vorgeschlagenen Strategie des ‚Theoretical Sampling‘ (vgl. Flick 2002, 102). Festgelegt wurde vor der Untersuchung das Schichtungsmerkmal ‚Zugehörigkeit der Schüler zum 13. Jahrgang, die das Fach Deutsch als Schwerpunkt gewählt hatten‘. Es handelt sich also bei der Stichprobe um eine geschichtete Zufallsstichprobe. Das heißt, die Vorauswahl erfolgte nach Merkmalen, die dann eine Zufallsstichprobe mit spezifischer Repräsentativität bezüglich dieser Merkmale ergab (vgl. Beller 2008, 87).

Im Rahmen der Hauptuntersuchung wurden insgesamt neun Schülerinnen und Schüler des 13. Jahrgangs untersucht. Darunter waren fünf männliche und vier weibliche Probandinnen und Probanden (vgl. Dannecker 2012, 132). Die Auswertung konnte auf neun Fälle beschränkt werden, da sich nach der Auswertung des neunten Falls eine so genannte ‚theoretische Sättigung‘ ergab, d.h. sich der Gehalt an Neuem für die zu entwickelnde Theorie auf Grund der Auswertung weiterer Fälle nicht ergab (vgl. Flick 2002, 102).

2.4 Auswertung der Daten

Als Auswertungsmethode wurde ein rekonstruktiv-sequentielles Verfahren gewählt, da damit ein sensibler Umgang mit dem Untersuchungsgegenstand der Rekonstruktion von Lese-, Verstehens-, Bewertungs- und Reflexionsprozessen ermöglicht wird. Zudem ermöglicht diese Methode im Gegensatz zu subsumptionslogischen Auswertungsverfahren, die dem Text mit einem festen Variablensystem

begegnen, den unterschiedlichen Textzugängen der Probandinnen und Probanden, auch in ihrer sequentiellen Abfolge, Rechnung zu tragen. Aus der forschungstheoretischen Überzeugung heraus, dass sich die Gültigkeit von Aussagen über die Wirklichkeit nur durch die empirische Erforschung der sozialen Wirklichkeit begründen lässt, wurden die Daten in Anlehnung an die ‚Grounded Theory‘ ausgewertet (s. Kubik in diesem Band). Die ‚Grounded Theory‘ ist insofern nicht nur ein Auswertungsverfahren, sondern die Entscheidung für diesen Forschungsansatz beeinflusst auch das Vorgehen bei der Probandenwahl und somit auch die Datenerhebung.

Zentraler Bestandteil bei der Datenauswertung in Anlehnung an die ‚Grounded Theory‘ ist das Kodieren. Damit ist gemeint, dass „Daten aufgebrochen, konzeptualisiert und auf neue Art zusammengesetzt werden“ (vgl. Strauss/Corbin 1996, 39). Ziel ist dabei, Themen aus den Daten zu extrahieren oder ein deskriptives, theoretisches Netz von locker verwobenen Konzepten zu entwickeln. Das Verfahren des Kodierens umfasst drei Haupttypen: offenes Kodieren, axiales Kodieren und selektives Kodieren (vgl. ebd., 40). In die Auswertung der Daten fließen zwar die vorher dokumentierten theoretischen Vorannahmen ein, doch es wird stets vom vorliegenden Datenmaterial ausgegangen. Dazu wurde mit ‚In-Vivo-Kodes‘ gearbeitet, die eine Orientierung an der Wortwahl der Probandinnen und Probanden ermöglichten. So lässt sich die ‚Grounded Theory‘ als deutend-rekonstruierendes Verfahren charakterisieren, das eine Generierung von Deutungstexten aus dem Material heraus anstrebt, bevor mit der Kategorienbildung begonnen wird. Ziel ist dabei die Generierung „einer Theorie, die dem untersuchten Gegenstandsbereich gerecht wird und ihn erhellt“ (ebd., 9). Im Vergleich zu anderen Verfahren ermöglicht die ‚Grounded Theory‘ eine größtmögliche Offenheit bei der Auswertung. Dies entspricht der explorativen Forschungshaltung und dem Erkenntnisinteresse der hier vorgestellten Studie.

2.5 Fazit

Auch wenn die Methode des Lauten Denkens wesentlich von der Verbalisierungsfähigkeit und -willigkeit der Probandinnen und Probanden abhängt, so bietet diese Erhebungsmethode bei allen Schwächen, wie beispielsweise die notwendige Auswahl von Kognitionen und Bildern seitens der Probandinnen und Probanden oder die Beeinflussung der Daten durch die Erwartungen der Versuchsleitenden – am ehesten und vollständigsten die Möglichkeit, die im Individuum ablaufenden Kognitionen zu erfassen, die – in bestimmtem Umfang – Rückschlüsse auf Gefühle und Empfindungen sowie lektürebegleitende und -strukturierende Operationen während des Lesens zulassen.

Da die empirisch forschende Deutschdidaktik noch nicht auf ein bewährtes Repertoire empirischer Forschungsmethoden zurückgreifen kann, ist sie darauf angewiesen, aus dem Methodenangebot verschiedener Disziplinen, wie etwa der Soziologie, Sozialpsychologie oder der pädagogischen Psychologie, eine bewusste Auswahl in Bezug auf ihre eigenen Fragestellungen und Gegenstände zu

treffen. Die Adaption ist dabei oftmals geboten, um der Gegenstandsangemessenheit fachdidaktischer Forschungsvorhaben gerecht zu werden. Mit der Entwicklung fachspezifischer Forschungsinstrumente beschreitet die Fachdidaktik Deutsch den Weg zu einer empirisch forschenden Wissenschaft, die sich der Aufgabe stellt, „wissenschaftlich begründete Kompetenzbeschreibungen, adäquate Erhebungsinstrumente für die je erreichten Lernstände und nicht zuletzt auch methodische Konzepte für die Verbesserung des Unterrichts zur Verfügung zu stellen, die auf ihre Wirksamkeit hin überprüft sind“ (Groebe/Hurrelmann 2006, 11).

Literatur

- Abraham, Ulf/Masanek, Nicole/Winkler, Iris: Zur Einführung. Poetisches Verstehen in Zeiten der Kompetenzorientierung. In: Dies. (Hrsg.): Poetisches Verstehen. Literaturdidaktische Positionen – empirische Forschung – Projekte aus dem Deutschunterricht. Baltmannsweiler: Schneider Verlag Hohengehren 2010, 5-8.
- Bertschi-Kaufmann, Andrea: „Also während dem Lesen bin ich eigentlich wie weg.“ Wenn Heranwachsende ihre Literalität entwickeln. In: Leseräume. Zeitschrift für Literalität in Schule und Forschung. Siehe: <http://leseraeume.de/wp-content/uploads/2015/10/lr-2014-1-bertschi-kaufmann.pdf> (letzter Zugriff: 01.08.2018).
- Bertschi-Kaufmann, Andrea: Literarisches Lernen und seine Spuren in den Selbstaussagen von heranwachsenden Leserinnen und Lesern. In: Frickel, Daniela A./Kammler, Clemens/Rupp, Gerhard (Hrsg.): Literaturdidaktik im Zeichen von Kompetenzorientierung und Empirie. Freiburg: Fillibach 2012, 165-186.
- Bohnsack, Ralf: Rekonstruktive Sozialforschung. Einführung in qualitative Methoden. Opladen: UTB 2003.
- Brüggemann, Jörn/Frederking, Volker: Literarisch kodierte, intendierte bzw. evozierte Emotionen und literarästhetische Verstehenskompetenz. Theoretische Grundlagen einer empirischen Forschung. In: Frickel, Daniela A./Kammler, Clemens/Rupp, Gerhard (Hrsg.): Literaturdidaktik im Zeichen von Kompetenzorientierung und Empirie. Freiburg: Fillibach 2012, 15-40.
- Dannecker, Wiebke (Hrsg.): Literaturvermittlung und Binnendifferenzierung – empirische Zugänge. Bielefeld: Aisthesis 2013. (= Hannoversche Beiträge zur Kulturvermittlung und Didaktik, Bd. 3).
- Dannecker, Wiebke: Literarische Texte reflektieren und bewerten – zwischen theoretischer Modellierung und empirischer Rekonstruktion. Trier: WVT 2012.
- Dannecker, Wiebke: „Eigentlich ist es schön zu lesen, nur es ist etwas irritierend“ – Zur Frage der Operationalisierbarkeit literarischen Verstehens. In: Lütge, Christiane et al. (Hrsg.): Empirische Forschung – Konzepte und Perspektiven. Frankfurt: Lang 2009 (= Fremdsprachendidaktik inhalts- und lernerorientiert, Bd. 16), 121-133.
- Dirks, Una: Fremdsprachenforschung als ‚Entdeckungsreise‘. Im Spannungsfeld von Abduktion, Deduktion und Induktion. In: Vollmer, Helmut Johannes (Hrsg.): Synergieeffekte in der Fremdsprachenforschung. Empirische Zugänge, Probleme, Ergebnisse. Frankfurt a.M.: Lang 2007. (= Kolloquium Fremdsprachenunterricht, Bd. 27), 43-58.

- Eggert, Hartmut: Literarische Texte und ihre Anforderungen an die Lesekompetenz. In: Groeben, Norbert/Hurrelmann, Bettina (Hrsg.): Lesekompetenz: Bedingungen, Dimensionen, Funktionen. Weinheim/München: Juventa 2002, 186-194.
- Ericsson, Anders K./Simon, Herbert A.: Protocol Analysis: Verbal Reports as Data. Cambridge, Ma: MIT Press 1993.
- Flick, Uwe: Qualitative Sozialforschung. Eine Einführung, Reinbek: Rowohlt 2002.
- Groeben, Norbert/Hurrelmann, Bettina: Einleitung. Die Grundkonzeption des Weiterbildungsprogramms. In: Dies. (Hrsg.): Empirische Unterrichtsforschung in der Literatur- und Lesedidaktik. Ein Weiterbildungsprogramm. Weinheim/München: Juventa 2006, 11-30.
- Gropengießer, Harald: Qualitative Inhaltsanalyse in der fachdidaktischen Lehr-Lernforschung. In: Mayring, Philipp/Gläser-Zikuda, Michaela (Hrsg.): Die Praxis der Qualitativen Inhaltsanalyse. Weinheim: Beltz 2005, 172-189.
- Hauber, Karl: Forschungsinteraktion und Forschungskonzeption. In: Huber, Günter L./Mandl, Heinz (Hrsg.): Verbale Daten. Eine Einführung in die Grundlagen und Methoden der Erhebung und Auswertung. Weinheim: Beltz 1994, 61-77.
- Heine, Lena: Lautes Denken als Forschungsinstrument in der Fremdsprachenforschung. In: Zeitschrift für Fremdsprachenforschung (2005), 16, 2, 163-185.
- Hurrelmann, Bettina: Leseleistung – Lesekompetenz. Folgerungen aus PISA, mit einem Plädoyer für ein didaktisches Konzept des Lesens als kultureller Praxis. In: Praxis Deutsch (2002), 176, 6-18.
- Kammler, Clemens: Literarische Kompetenzen – Standards im Literaturunterricht. Anmerkungen zum Diskussionsstand. In: Ders. (Hrsg.): Literarische Kompetenzen – Standards im Literaturunterricht. Modelle für die Primar- und Sekundarstufe. Seelze: Kallmeyer 2006, 7-23.
- Kämper-van den Boogaart, Michael/Pieper, Irene: Literarisches Lesen. In: Didaktik Deutsch Sonderheft 2 (2008), 46-65.
- Kämper-van den Boogaart, Michael: Lässt sich normieren, was als literarische Bildung gelten soll? Eine Problemskizze am Beispiel von Brechts Erzählung Der hilflose Knabe. In: Rösch, Heidi (Hrsg.): Kompetenzen im Deutschunterricht. Frankfurt a.M.: Lang 2005. (= Beiträge zur Literatur-, Sprach- und Mediendidaktik, Bd. 9), 27-50.
- Konrad, Klaus: Lautes Denken. In: Mey, Günter/Mruck, Katja (Hrsg.): Handbuch Qualitative Forschung in der Psychologie. Wiesbaden: Verlag für Sozialwissenschaften 2010, 476-490.
- Lamnek, Siegfried: Qualitative Sozialforschung. Weinheim: Beltz 2010.
- Leubner, Martin/Saupe, Anja: Textverstehen im Literaturunterricht und Aufgaben. Baltmannsweiler: Schneider Hohengehren 2008.
- Meissner, Ruth: Domänenspezifisches Vorwissen und literarisches Verstehen. Erste Ergebnisse einer qualitativ-empirischen Studie zur Interpretation von Lyrik. In: Winkler, Iris/Masanek, Nicole/Abraham, Ulf (Hrsg.): Poetisches Verstehen. Literaturdidaktische Positionen – empirische Forschung – Projekte aus dem Deutschunterricht. Baltmannsweiler: Schneider Verlag Hohengehren 2010, 133-147.

- Odag, Özen: Wenn Männer von der Liebe lesen und Frauen von Abenteuern... Eine empirische Rezeptionsstudie zur emotionalen Beteiligung von Männern und Frauen beim Lesen narrativer Texte. Lengerich: Pabst 2007.
- Olson, Gary M./Duffy, Susan A./Mack, Robert L.: Thinking-Out Loud as a Method for Studying Real-Time Comprehension Processes. In: New Methods in Reading Comprehension Research. Just. Hillsdale 1984, 253-286.
- Pieper, Irene: Literarische Kompetenz. In: Kliewer, Heinz-Jürgen/Pohl, Inge (Hrsg.): Lexikon Deutschdidaktik. Bd. 1. Baltmannsweiler: Schneider Verlag Hohengehren 2006, 441-443.
- Rosebrock, Cornelia: Lesesozialisation und Leseförderung – literarisches Lernen in der Schule. In: Kämper-van den Boogaart, Michael (Hrsg.): Deutschdidaktik. Leitfaden für die Sekundarstufe I und II. Berlin: Cornelsen Scriptor 2003, 153-174.
- Rosebrock, Cornelia: Literarische Erfahrung mit dem Erlkönig: Ein Blick auf Prozessebenen des Lesens. In: Härle, Gerhard/Rank, Bernhard (Hrsg.): „Sich bilden ist nichts anders, als frei werden“. Sprachliche und literarische Bildung als Herausforderung für den Deutschunterricht. Baltmannsweiler 2008, 89-110.
- Rosebrock, Cornelia: Neue kurze Prosa. In: Praxis Deutsch (2007b), 206, 6-16.
- Rosenthal, Gabriele: Interpretative Sozialforschung. Eine Einführung. Weinheim/München: Juventa 2008.
- Schreier, Margrit: Qualitative Verfahren der Datenerhebung. In: Groeben, Norbert/Hurrelmann, Bettina (Hrsg.): Empirische Unterrichtsforschung in der Literatur- und Lesedidaktik. München: Juventa 2006, 399-420.
- Spinner, Kaspar H.: Lesen als ästhetische Bildung. In: Bertschi-Kaufmann, Andrea (Hrsg.): Lesekompetenz – Leseleistung – Leseförderung. Seelze: Kallmeyer/Klett 2007, 83-95.
- Stark, Tobias: Lautes Denken in der Leseprozessforschung. Kritischer Bericht über eine Erhebungsmethode. In: Didaktik Deutsch (2010b), 29, 58-78.
- Stark, Tobias: Zur Interaktion von Wissensaktivierung, Textverstehens- und Bewertungsprozessen beim literarischen Lesen. Erste Ergebnisse einer qualitativen empirischen Untersuchung. In: Winkler, Iris/Masaneck, Nicole/Abraham, Ulf (Hrsg.): Poetisches Verstehen. Literaturdidaktische Positionen – empirische Forschung – Projekte aus dem Deutschunterricht. Baltmannsweiler: Schneider Verlag Hohengehren 2010a, 114-132.
- Steinhauer, Lydia: Involviertes Lesen: eine empirische Studie zum Begriff und seiner Wechselwirkung mit literarästhetischer Urteilskompetenz. Freiburg: Fillibach 2010.
- Steinke, Ines: Gütekriterien qualitativer Forschung. In: Uwe Flick/Ernst Kardorff/Steinke, Ines (Hrsg.): Qualitative Forschung. Ein Handbuch. Reinbeck: Rowohlt 2005, 319-331.
- Strauss, Anselm/Corbin, Juliet: Grounded Theory: Grundlagen Qualitativer Sozialforschung. Weinheim: Beltz 1996.
- Weidle, Renate/Wagner, Angelika C.: Die Methode des Lauten Denkens. In: Huber, Günter L./Mandl, Heinz (Hrsg.): Verbale Daten. Eine Einführung in die Grundlagen und Methoden der Erhebung und Auswertung. Weinheim: Beltz 1994, 81-103.

- Weinzierl, Christian/Wrobel, Arne: Schreibprozesse untersuchen. In: Becker-Mrotzek, Michael/Grabowski, Joachim/Steinhoff, Torsten (Hrsg.): Forschungshandbuch empirische Schreibdidaktik. Münster/New York: Waxmann 2017. 221-237.
- Winkler, Iris: Welches Wissen fördert das Verstehen literarischer Texte? Zur Frage der Modellierung literarischen Wissens für den Deutschunterricht. In: Didaktik Deutsch (2007), 22, 71-94.

Videographische Datenerhebung

Lehr-/Lernkontexte unter dem Mikroskop?

1. Vorstellung der videographischen Datenerhebung

Videographisch orientierte Forschung verwendet Filmmaterial als Datenbasis im Rahmen wissenschaftlicher Untersuchungen, wobei im Folgenden ausschließlich zur wissenschaftlichen Dokumentation erhobene bzw. erstellte videographische Daten im Vordergrund stehen¹, die in der Regel von Forschenden selbst generiert werden (Corsten 2010, 8). Videographische Datenerhebung stellt eine Form wissenschaftlicher Online-Beobachtung dar. Wissenschaftlich ist diese, da sie zielgerichtet, systematisch und methodisch kontrolliert bzw. reflektiert erfolgt (vgl. Ricart Brede 2014, 138; Kromrey 2009, 325ff.). Von Online-Beobachtungen spricht man immer dann, wenn digitale Beobachtungsdaten generiert werden (vgl. Ricart Brede 2014, 139). Insbesondere in den letzten zehn Jahren hat die videographische Datenerhebung, u.a. auf Grund technischer Entwicklungen, im Kontext der Erforschung von Lehr- und Lernprozessen an Bedeutung gewonnen. Im Folgenden stehen die Vor- und Nachbereitung sowie die eigentliche Erhebung von videographischen Daten im Vordergrund und es werden konkrete Hinweise für die Forschungspraxis gegeben.

1.1 *Roll it!* – Funktionsweise des Verfahrens

Im Fokus dieses Teilkapitels steht die eigentliche Erhebung von videographischen Daten. Die Ausführungen beziehen sich vornehmlich auf Erhebungen, die in Lehr-/Lernkontexten, z.B. in Schulen, stattfinden. Idealtypisch lassen sich für die videographische Datenerhebung vier Schritte bzw. Phasen ausmachen:

1. Felderschließung und Einholung aller notwendigen Einverständnisse
2. Planung und Organisation der Erhebung
3. Erhebung der videographischen Daten sowie ggf. zusätzlicher Daten
4. Datensicherung und -archivierung

¹ Alternativ können zweitens Aufnahmen zur Erinnerung an besondere Anlässe und Begebenheiten im Alltag von Filmamateuren oder drittens mit künstlerischem oder journalistischem Anspruch von berufsmäßigen Filmemachern erstellt werden. Ein Grenzfall stellt die Variante dar, bei der Forschende Menschen bitten, Szenen ihrer Lebenspraxis selbst per Video festzuhalten. (Vgl. Corsten 2010, 8).

Videographische Datenerhebung beginnt in der Regel mit der Erschließung des Forschungsfeldes, also der Kontaktaufnahme zu den Studienteilnehmerinnen und -teilnehmern. Diese müssen einerseits über Erkenntnisinteresse und Anlage der Studie informiert werden und andererseits auf Basis dieser Informationen ein Einverständnis zur Teilnahme an der Studie geben. Als geeignet haben sich Informationsschreiben mit angehängter Einverständniserklärung erwiesen.² Bei Minderjährigen ist zu berücksichtigen, dass das Einverständnis von Eltern bzw. Erziehungsberechtigten einzuholen ist. Es bietet sich bei videographischen Datenerhebungen an, die Einverständniserklärungen mehrstufig zu gestalten, sodass z.B. eine Erlaubnis zur Aufnahme von a) Audio- oder b) Audio- und Videoaufnahmen gegeben werden kann. Soll z.B. Unterricht videographiert werden, hat das den Vorteil, dass Schülerinnen und Schüler, für die ein Einverständnis für a), nicht aber für b) vorliegt, zumindest außerhalb des von der Kamera erfassten Sichtfeldes platziert werden können, ohne dass auf Aufnahmen ganz verzichtet werden muss. Sollen videographische Daten an Schulen erhoben werden, dann ist auch ein Einverständnis des jeweiligen Kultusministeriums bzw. des zuständigen Schulamtes erforderlich. Je nach Bundesland variieren die Anforderungen und Vorgaben – z.B. für entsprechende Anträge – sehr stark. Da es sich bei videographischen Daten um sensible Daten handelt, ist ein ethisch und rechtlich³ vertretbarer Umgang wesentlich. Diese erste Phase kann im Extremfall sechs bis zwölf Monate dauern und es ist nicht ausgeschlossen, dass die Studie nicht oder nicht überall bzw. nicht wie ursprünglich geplant stattfinden kann.

Sind aber einmal alle notwendigen Einverständnisse eingeholt, dann gilt es in einer zweiten Phase die Erhebung zu planen. Hierbei bietet es sich an, im Rahmen von ersten umfassenden Felddbesuchen, das Forschungsfeld besser kennenzulernen. In Gesprächen mit Akteuren des Feldes können wichtige Informationen zu Strukturen und Gegebenheiten vor Ort gesammelt werden. Ferner können erste Offline-Beobachtungen, d.h. Beobachtungen, die auf das eigene „Sehen“ zurückgreifen, ohne dass bereits digitale Daten erhoben werden, weitere Einblicke eröffnen und ein Vertrautmachen mit dem Feld und dessen räumlichen Gegebenheiten ermöglichen. Hierbei kann bereits die Vorgehensweise bei der videographischen Datenerhebung, z.B. die Position von Kamera(-Personen), geplant und ggf. mit Akteuren abgesprochen werden. Auch können eventuell bereits

² Ein Beispiel für ein Informationsschreiben für Erhebungen an Schulen findet sich in Krelle (2014, 84).

³ Insbesondere relevant sind das APR (allgemeines Persönlichkeitsrecht) im Grundgesetz: Art. 2 Abs. 1 (Recht auf Freiheit der Person) und Art. 1 Abs. 2 (Unantastbarkeit der Würde des Menschen). Darüber hinaus gilt: „Bildnisse dürfen nur mit Einwilligung des Abgebildeten verbreitet oder öffentlich zur Schau gestellt werden. Die Einwilligung gilt im Zweifel als erteilt, wenn der Abgebildete dafür, daß er sich abbilden ließ, eine Entlohnung erhielt“ (§ 22 KunstUrhG). Ausnahmen stellen z.B. Bilder von öffentlichen Versammlungen, Aufzügen und ähnlichen Vorgängen, an denen die dargestellten Personen teilgenommen haben, dar.

(sprach-)biographische Daten aller beobachteten Akteure erhoben werden.⁴ Ein solches Vorgehen gestattet es auch, Vertrauen und eine angenehme Erhebungssituation zu schaffen.

Schließlich findet in der dritten Phase die eigentliche Datenerhebung statt. Zu den vereinbarten Terminen sind Kamera und Kamerapersonen wie abgesprochen platziert bzw. werden entsprechend eingesetzt, wobei die Kamera im Idealfall vor Betreten des Raums durch die Akteure läuft. Bereits vor der Erhebung sollte eine Raumskizze angefertigt worden sein, in die vor, während bzw. im Anschluss an die Datenerhebung Sitzplan-Notizen eingetragen werden, sodass z.B. die eindeutige Zuordnung (sprach-)biographischer Daten möglich ist. Auch das Vermerken von Sitzplatzwechseln ist empfehlenswert, da dies maßgeblich die Aufbereitung und Analyse der Daten erleichtert. Ferner sollte jeweils zeitnah im Anschluss an die Erhebung ein Kurzprotokoll angefertigt werden, das Kernaspekte/-inhalte der videographierten Situation wiedergibt sowie auch Auffälliges bzw. für die Forschenden Interessantes dokumentiert. Ebenso bietet sich die zumindest stichpunktartige Dokumentation von Kommentaren seitens der Akteure im Anschluss an die Aufnahmen an. All diese Maßnahmen verhindern auch, dass der Überblick über die Daten verloren geht. Denn insbesondere während der dritten Phase werden teilweise sehr schnell sehr viele Daten erhoben. Um ein einheitliches Vorgehen zu sichern und Transparenz zu gewährleisten, bietet es sich an, einen Leitfaden bzw. ein Kameraskript für die Erhebung zu erstellen.⁵

Da die Teilnahme an einer Studie immer auf freiwilliger Basis erfolgt und jederzeit abgebrochen werden kann und da videographische Daten verhältnismäßig invasive und sensible Daten darstellen, ist m.E. eine „good-practice“-Empfehlung, im Anschluss an die Erhebung anzubieten, die Daten zu löschen. Hat man z.B. Unterricht videographiert und die gefilmte Lehrkraft teilt im Anschluss mit, dass sie sich während der Aufnahme sehr unwohl gefühlt und nicht den Eindruck hat, dass es sich um eine für sie ‚typische‘ Stunde handelt, kann das Angebot das gegenseitige Vertrauen stärken und eine Bereitschaft zur Kooperation damit aufrecht erhalten. Gleichzeitig stellt dies ein Recht der Untersuchungsteilnehmerinnen und -teilnehmer dar. Erfahrungsgemäß wird dieses Angebot nicht angenommen – zumal auch Teilnehmerinnen und Teilnehmern der Aufwand der Datenerhebung durchaus bewusst ist – aber als Erleichterung empfunden, dass dies prinzipiell möglich ist.

Phase vier, die Überlappungen mit der dritten Phase aufweist, wird häufig vernachlässigt. Aber dass die Daten möglichst zeitnah gesichert, pseudonymisiert und archiviert werden, stellt eine wichtige Aufgabe dar. Die Sicherung besteht in

⁴ Diese sind häufig unabdingbar für eine weitere Aufbereitung und Analyse der Daten. Im Fall von Studien zum Fremd- oder Zweitspracherwerb z.B. betrifft dies Informationen dazu, wie lange die Fremd- oder Zweitsprache bereits gelernt wird, welche anderen Fremd- oder Zweitsprachen beherrscht werden, was die Erstsprache ist etc.

⁵ Ein konkretes Beispiel findet sich in Ricart Brede (2011).

der Regel im Überspielen der Dateien (ggf. in ein anderes Format) auf einen sicheren (mindestens passwortgeschützten) Datenträger, im Anfertigen einer Kopie und ggf. der Löschung der Aufnahmedaten, z.B. von Audioaufnahmegegeräten. Bei der Pseudonymisierung werden Namen bzw. andere Identifikationsmerkmale durch Pseudonyme bzw. Codes ersetzt, sodass Personen (auch Schulen etc.) nicht identifiziert werden können, aber Bezüge verschiedener Datensätze zueinander weiterhin erhalten bleiben (Virtuelles Datenschutzbüro 2013).⁶ Werden z.B. Daten an zwei Schulen erhoben, dann können diesen die Zahlen 1 und 2 zugeordnet werden. Datensätze, die an Schule 1 erhoben wurden, haben diese Zahl dann immer als erste im Benennungscode. Tatsächlich ist die Erstellung und Zuordnung eines eindeutigen Code-Systems eine Herausforderung und muss projektweise angepasst werden. Die Archivierung der Daten schließlich erfolgt durch die ‚Ab-lage‘ an verschiedenen Orten, z.B. in abschließbaren und für Unbefugte unzu-gänglichen Schränken. Ferner sollte der Datenbestand in einer Metadatei fortlau-fend dokumentiert werden, sodass ein schneller Rückgriff auf die Daten möglich ist.

1.2 *To be continued...* Generierte Daten und ihre weitere Nutzung

‚Ergebnis‘ einer videographischen Datenerhebung sind zunächst Videodaten, die visuelle und auditive Daten beinhalten. Diese Videodaten können in unterschiedlicher Art und Weise weitergenutzt werden. Ganz grob lassen sich zwei Arten der Verwendung unterscheiden:

1. Verwendung der Filmdaten für die Analyse von (Hör- und) Sichtbarem direkt am (audio-)visuellen Datenmaterial
2. Nutzung der audio-visuellen Daten zur Anfertigung von Transkripten und deren anschließende Feinanalyse

Im ersten Fall spricht Ricart Brede (2011, 92f.) z.B. von Sichtstrukturen. Hierbei stehen also sichtbare, direkt beobachtbare Merkmale einer (Lehr-Lern-)Situation im Vordergrund. Solche Analysen erfordern in der Regel die Erstellung eines Beobachtungssystems.⁷ Auch Collagen und das (Zusammen-)Schneiden von video-graphierten Daten, wie es im Rahmen von Kamera-Ethnographie⁸ erfolgt, ist dieser Herangehensweise zuzuordnen. Weitere Analyse-möglichkeiten, deren Gegenstand die Videodaten selbst sind, stellt z.B. auch die von Dinkelaker und Herrle

⁶ Im Gegensatz dazu können anonymisierte Daten nicht mehr rückbezogen werden.

⁷ Weiterführende Hinweise dazu finden sich z.B. in Ricart Brede et al. (2010).

⁸ „Eine ethnographische Untersuchung zielt in der Regel darauf ab, Menschen über einen längeren Zeitraum in ihrem alltäglichen Leben zu beforschen. Das heißt, der/die EthnographIn bzw. FeldforscherIn nimmt physisch mit der Gesamtheit seiner/ihrer Person über einen **längeren Zeitraum an ausgewählten Lebenswelten** teil, mit dem Ziel, **Daten zu erheben** und **Beschreibungen** anzufertigen, die als Grundlage für spätere **Analysen** dienen. [...] Ethnographische Feldforschung zeichnet sich durch eine besonders intensive und langfristige, über die reine Datenerhebung hinausgehende In-Beziehung-Setzung zum Untersuchungsfeld aus.“ (Halbmayer/Salat o.J.; Hervorhebungen im Original, Anm. der Verfasserin)

(2009, 64ff.) vorgestellte Konfigurationsanalyse dar. Diese ermöglicht es, einen Überblick über die Gesamtordnung des Interaktionsgeschehens zu erhalten und z.B. Rückschlüsse daraus zu ziehen, wie Raum genutzt wird bzw. auf welche Raumbereiche Geschehen fokussiert wird. Fuchs et al. (2014) können mit Hilfe dieser Analyse zeigen, inwiefern die Raumnutzung Rückschlüsse auf Möglichkeiten der Erstsprachen-Nutzung im Rahmen von zweitsprachlicher Sprachförderung zulässt und inwiefern dies durch die Lehrkraft gesteuert wird.

Maßgeblicher Vorteil von Untersuchungen, die direkt am Datenmaterial arbeiten, besteht darin, dass eine aufwändige Transkription nicht notwendig und der unmittelbare Bezug zum Original-Datenmaterial jederzeit gewährleistet ist. Insbesondere die Fokussierung von non-verbalen Aspekten bietet sich für eine solche Vorgehensweise an.

Ferner dienen Videodaten insbesondere im Kontext von Unterrichtsforschung häufig als Grundlage für die Anfertigung von Transkripten,⁹ welche dann in Feinanalysen weiter untersucht werden, sodass diese ‚Sekundärdaten‘ Gegenstand der eigentlichen Analyse sind und die Originaldaten lediglich in Zweifelsfällen hinzugezogen werden. Erst Transkriptionen erlauben eine detaillierte Untersuchung sprachlicher Handlungen oder Formulierungen (Brünner 2012, 55). So lässt sich komplexes kommunikatives Unterrichtsgeschehen überhaupt erst im Transkript rekonstruieren:

Die detaillierte Dokumentation des kommunikativen Geschehens, das heißt die möglichst umfassende und genaue Verschriftlichung der verbalen und non-verbalen Anteile, ist die notwendige Voraussetzung für die Analyse und Erklärung der beobachteten Lehr-Lern-Prozesse. (Becker-Mrotzek/Vogt 2009, 62)

Jedoch sind mit der Transkription immer bereits Selektions- und Interpretationsprozesse verbunden (Schramm/Aguado 2010, 194).

Vorteil dieser Vorgehensweise ist, dass videographische Daten, insbesondere von Unterrichtssituationen, eine bessere Qualität von Transkripten ermöglichen. Steht z.B. lediglich eine Audioaufnahme zur Verfügung, um ein Transkript einer Unterrichtsstunde mit ca. 30 Schülerinnen und Schülern anzufertigen, dann ist eine durchgängige eindeutige Sprecherzuordnung so gut wie unmöglich. Darüber ermöglichen videographische Daten eine umfassendere Berücksichtigung des Kontextes bzw. der Situation (Was wird gesagt? vs. Was wird getan?).

Es handelt sich hierbei um eine sehr grobe Unterscheidung bzgl. der Art der Weiterverwendung videographierter Daten. In der Forschungspraxis finden sich zahlreiche Kombinationen bzw. fließende Übergänge. Dennoch besteht eine große Herausforderung videographischer Forschung darin, zu entscheiden, welche Daten für welche Art von Aufbereitung und Analyse notwendig sind. Nicht selten liegen nach der Datenerhebung zwei bis drei Videoaufnahmen des gleichen Geschehens aus unterschiedlichen Perspektiven plus ein bis zwei Audioaufnahmen vor, ohne dass vorab hinterfragt wurde, wie diese dann für eine zweckgemäße

⁹ Ein Beispiel findet sich in Krelle (2014, 82f.).

Aufbereitung und Analyse zusammenzubringen bzw. nutzbar zu machen sind, ohne den Aufwand bis ins Unpraktikable hin zu steigern. Daher bietet es sich an, im Rahmen einer Pilotstudie den vollständigen Forschungsprozess (Datenerhebung, -aufbereitung und -analyse) zu erproben.

Tabelle 1 gibt einen ersten (unvollständigen) Einblick in die Programme, mit denen videographische Daten aufbereitet bzw. analysiert werden können. Aufgeführt ist, ob bis zu zwei Videospuren eingespielt werden können und ob eine Analyse bzw. Annotation direkt im Programm möglich ist. Bei F4 Video z.B. handelt es sich um eine Transkriptionssoftware, die keine weitere Analyse ermöglicht.

Programm	Transkription	Eine Videospur	Zwei Videospuren	Analyse/Annotation direkt im Programm
ELAN	✓ Partitur	✓	✓	✓
EXMA-RaLDA	✓ Partitur	✓	--	✓ (umfassende linguistische Annotationen und Analysen mgl.)
F4 Video	✓ Textnotation	✓	--	--
MAXQDA	--	✓ Ab Vers. 13	--	✓
TRANSANA	✓ Textnotation	✓	--	--
Videograph	✓ Textnotation	✓	✓	✓ (insbesondere Sichtstrukturen)

Tab. 2: Programme zur Aufbereitung und Analyse von Videodaten

Es bietet sich an, die vorhandenen Programme in Abhängigkeit des eigenen Erkenntnisinteresses zu prüfen und zu erproben. Unter Nutzung solcher Programme können häufig mehr Daten bearbeitet werden und dies kann verhältnismäßig einfach und effektiv geschehen. Ferner können diese Programme zu einer besseren Transparenz in der Darstellung der Vorgehensweise bei Datenaufbereitung und -analyse führen.

1.3 *Film reviews...* Gütekriterien für den Einsatz videographischer Datenerhebung

Für videographische Daten wird vornehmlich die Frage der Validität, also der Frage danach, ob tatsächlich erfasst wird, was erfasst werden soll, kritisch diskutiert. Eng damit verbunden ist die Auseinandersetzung mit dem so genannten Beobachterparadoxon (Labov 1971), das darauf verweist, dass durch die Gegenwart eines Beobachters – dies gilt auch für die Gegenwart einer Kamera – das beo-

bachtete Geschehen bereits verändert wird und sich die Gegenwart auf das Beobachtete auswirkt. Die Anwesenheit der Kamera sowie der Forschenden ist daher für jeden Erhebungskontext kritisch zu reflektieren.

Im Zusammenhang mit dieser grundlegenden Herausforderung videographischer Beobachtung wird ferner der Aspekt der Invasivität, also der durch die Filmsituation und Kamera(-personen-)präsenz verursachte Einfluss auf die Situation, diskutiert. Für die Erhebung videographischer Daten in Unterrichtskontexten fassen Maak und Ricart Brede deren Auswirkungsmöglichkeiten wie folgt zusammen:

Möglich ist dabei eine Beeinflussung sowohl der Lehrenden als auch der Lernenden, die sich darin äußern kann, dass diese verbal, paraverbal und/oder nonverbal sowie bezogen auf fachliche, didaktische und/oder soziale Inhalte auf eine für sie ungewöhnliche Art und Weise handeln bzw. Handlungen unterlassen. Diese Veränderungen müssen nicht per se beobachtbar sein und können sich auch allein im Denken der Lehrenden und Lernenden manifestieren. (Maak/Ricart Brede 2014, 151f.)

Bezogen auf Unterricht vertritt Helmke die Auffassung (2009, 301; ähnlich Petko et al. 2003, 270), dass jeder Unterricht durch routinehafte Abläufe geprägt ist und sich wünschbare Verhältnisse nur selten kurzfristig herstellen lassen. Petko et al. (2003, 270) räumen aber ein, dass Lehrende einen aus ihrer Sicht idealen Unterricht zu halten versuchen. Stigler spricht von einer idealisierten Version dessen, was der Lehrende normalerweise im Klassenraum täte (Stigler 1998, 141). Umfassende Untersuchungen in diesem Bereich stehen noch aus. Befragungen von Lehrenden sowie Schülerinnen und Schülern zur Typikalität von Unterricht weisen darauf hin, dass dieser als ‚normal‘ wahrgenommen wird (vgl. Petko et al. 2003, 270). Als Selbstaussagen sind diese Angaben entsprechend kritisch zu reflektieren. Maak und Ricart Brede (2014) untersuchen durch die Kamera verursachte beobachtbare Invasivität und kommen zu dem Schluss, dass nicht notwendigerweise anderes, aber häufig weniger Unterrichtsgeschehen erhoben wird; immerhin 10% der analysierten Aufnahmezeit dokumentierten invasive Effekte. Frankhauser (2013, Abs. 1) spricht davon, dass insbesondere innerhalb des quantitativen Paradigmas das Ideal einer Null-Invasivität angestrebt wird, ein Unsichtbarmachen der Technik, wobei sie diesen Anspruch kritisch hinterfragt und diskutiert.

Andererseits folgern etwa Tuma et al. (2013, 33f.) aus der Tatsache, dass Aussagen über Daten am Originaldatum validiert werden können, dass es sich bei Videodaten um außergewöhnlich valide, vermutlich sogar die valideste Datensorte in der sozialwissenschaftlichen Forschung handelt. Auch Corsten (2010, 9) verweist darauf, dass Videodaten „[...] hinsichtlich der Gütekriterien der Exaktheit, Lückenlosigkeit und Zuverlässigkeit als Beobachtungsprotokoll kaum zu übertreffen [...]“ sind. Dementsprechend wird mehrheitlich davon ausgegangen, dass Videodaten den gängigen Gütekriterien gerecht zu werden vermögen.

Ein weiterer Aspekt der Diskussion um die Validität von videographischen Daten ist der Anspruch, das Geschehen möglichst umfassend und vollständig zu doku-

mentieren. Dem steht die Vorgehensweise bei der Kamera-Ethnographie entgegen. Hierbei erfolgen ethnographische Verstehensprozesse und Beschreibungsversuche durch den Kameragebrauch und Videoschnitt (Mohn 2008, 61):

Da Situationen zu komplex und unüberschaubar sind, als dass man sie bei laufender Kamera einfach aufzeichnen könnte – ein gerade beim sozialwissenschaftlichen Kameragebrauch verbreiteter Irrtum – sind Entscheidungen zu treffen, von wo aus und wohin blickend denn sinnvolle Blickspuren mit der Kamera angelegt werden können. Die Blickschneise durch den Dschungel des komplexen Geschehens wird bei der Kamera-Ethnographie zum Dreh- und Angelpunkt eines selektiven, Bilder produzierenden Hinschauens. (Ebd., 62)

Die Kamera als Werkzeug der Formulierung audiovisueller Feldnotizen soll also nicht das ganze Geschehen erfassen und Objektivität stellt hier kein angestrebtes Gütekriterium dar, vielmehr geht es um reflektierte Subjektivität. Dies bedeutet, dass abhängig von Forschungsfrage, Erkenntnisinteresse und Herangehensweise unterschiedliche Gütekriterien in unterschiedlicher Gewichtung und Tragweite Berücksichtigung finden müssen. Wesentlich – für jede Form des Zugriffs, sei es qualitativ oder quantitativ orientiert – ist jedoch das Gütekriterium der Transparenz bzw. der intersubjektiven Nachvollziehbarkeit (Schmelter 2014, 43f.; Steinke 2007, 325ff.) – eine nachvollziehbare Darstellung von Erhebung, Aufbereitung, Analyse und Präsentation der Ergebnisse ist immer anzustreben.

1.4 *See the film or not?* Stärken und Schwächen videographischer Datenerhebung

Mit der Darstellung der vier Phasen der Erhebung videographischer Daten und der Diskussion der Gütekriterien ging bereits die – zumeist implizite – Darstellung von Stärken und Schwächen der Methode einher. Diese sind daher nachfolgend verkürzt in einem tabellarischen Überblick zusammengefasst. Dem ist hinzuzufügen, dass je nach Erkenntnisinteresse und Herangehensweise bei der Erhebung die Grenze zwischen Stärke und Schwäche fließend ist bzw. sich dieses Verhältnis auch umkehren kann (vgl. z.B. Hinweise zur Kamera-Ethnographie im vorangegangenen Teilkapitel). Der Großteil der in Tabelle 2 aufgeführten Aspekte ist ferner selbsterklärend. Daher wird nachfolgend lediglich auf ausgewählte Aspekte näher eingegangen.

STÄRKEN	SCHWÄCHEN
<ul style="list-style-type: none"> • Iteration (Petko et al. 2003, 265; Schramm/Aguado 2010, 1866ff.; Tuma et al. 2013, 33f.) <p><i>u.a. wiederholtes Ansehen, unterschiedliche Analysen, gleiche Analysen zu verschiedenen Zeitpunkten, Möglichkeiten der besseren Validierung, größere Intersubjektivität</i></p> <ul style="list-style-type: none"> • Auditive + visuelle Daten (Dinkelaker/Herrle 2009, 14) → Dichte der Daten, die eine „Mikroskopie“, eine viel detailliertere Betrachtung als mit dem ‚bloßem Auge‘, ermöglicht (Tuma et al. 2013, 34) • Beobachtung in Echtzeit (Dinkelaker/Herrle 2009, 14) • weniger subjekt- und theoriegebundene Qualität der Daten → analytische Fragestellungen müssen nicht vor der Erhebung festgelegt werden (Petko et al. 2003, 265) • Festhalten von visuell erfassbaren Vorgängen in mimetischer Form innerhalb ihres natürlichen Kontexts (Tuma et al. 2013, 31) • Nutzung technischer Entwicklungen für die Analyse; u.a. rasches Spulen, Zeitlupen, Zoomen, Standbildanfertigung (Tuma et al. 2013, 32) • ‚Alltagsnähe‘ der wissenschaftlichen Untersuchung (praxisnahe Unterrichtsforschung) • Offenheit bzw. Möglichkeit, Unerwartetes zu finden • Nutzung für Lehrer(aus)- und -fortbildung (z.B. Helmke et al. 2007) → Brücke zwischen Theorie und Praxis (Petko et al. 2003, 265; Krammer/Reusser 2005; Knapp/Ricart Brede 2012, 220ff.) 	<ul style="list-style-type: none"> • Vorlaufzeit (Einverständniseinholung etc.) – zeit- und arbeitsaufwändige Erschließung des Feldes • systematische Stichprobenverzerrung (geringere Teilnahmbereitschaft, wenn Video-/Bildaufnahmen angefertigt werden) • transparente und überprüfbare Darstellung von Forschungsprozessen und -ergebnissen erschwert, da die Daten i.d.R. nicht veröffentlicht werden können (Dinkelaker/Herrle 2009, 18) • Invasivität (vgl. z.B. Maak/Ricart Brede 2014) bzw. Beobachterparadoxon (Labov 1971) • Veränderung des Beobachteten durch Veränderung des Feldes (z.B. durch das Umsetzen von Schülerinnen und Schülern) • Transformation lebensweltlicher Situationen (Tuma et al. 2013, 34f.) • Aufwand für Datenaufbereitung bei Transkription • Transkription stellt bereits Interpretation der Daten dar (Schramm/Aguado 2010, 194) • Selektivität; Zeigen immer auch Nicht-Zeigen (Frankhauser 2013)

Tab. 3: Stärken und Schwächen videographischer Datenerhebung

Wenn Tuma et al. (2013, 31ff.) von visuell erfassbaren Vorgängen in mimetischer Form innerhalb ihres natürlichen Kontexts sprechen, dann gehen sie davon aus, dass Videos insofern mimetisch sind als sie etwas abbilden, das tatsächlich stattgefunden hat. Da auf Grund der Daten viel detailliertere Betrachtungen als mit dem ‚bloßem Auge‘, ferner Zeitlupe, beschleunigtes Vor- und Zurückspulen sowie Zoomen möglich ist, verwenden sie die Metapher von der „soziologischen Mikroskopie“ (ebd., 34). Gleichzeitig bedeutet dieser Realitätscharakter und diese

Möglichkeiten der Mikroskopie jedoch nicht, dass es sich um eine Abbildung der Wirklichkeit handelt, sondern vielmehr um eine Transformation lebensweltlicher Situationen:

So reduziert das Video den dreidimensionalen Raum auf eine zweidimensionale Bildschirmfläche, schränkt den natürlichen Sehwinkel ein und verkleinert das Sichtfeld auf einen viereckig begrenzten Ausschnitt. Aufzeichnungen vermindern die Wahrnehmungsfülle sozialer Situationen auch in anderer Hinsicht. Video ist nicht in der Lage, nicht-akustische und nicht-visuelle Erfahrungsqualitäten mit einzufangen. (ebd.)

Etwas, das also der Beobachtung durch die Kamera entzieht, ist z.B. die Atmosphäre; Frankhauser bezeichnet dies als „Problem der Entsinnlichung“ (Frankhauser 2013, 27f.). Zusammenfassend gilt für die videographische Datenerhebung wie für jede andere Datenerhebungsmethode, dass deren Einsatz gut durchdacht und begründet sein sollte. Ferner darf nicht aus dem Blick verloren werden, dass – auch videographische – Beobachtung immer selektiv ist (ebd., 26), denn: „Die Kunst des Sehens und Zeigens ist immer auch eine des Weglassens und des Nichtzeigens. Wer immer alles gleichzeitig im Bildausschnitt der Kamera haben möchte, der hat gewissermaßen alles und doch nichts.“ (Mohn 2008, 63)

1.5 Literaturempfehlungen

Ricart Brede (2014): Beobachtung.

Sollten Sie noch nicht mit der Methode ‚Beobachtung‘ vertraut sein, können Sie mit dem Beitrag von Ricart Brede einen ersten Überblick zum Thema zu erhalten.

Dinkelaker/Herrle (2009): Erziehungswissenschaftliche Videographie.

Diese Einführung in die Videographie ist besonders gut geeignet für ein erstes Einarbeiten in die Thematik – nicht zuletzt auch, weil hier für die deutschdidaktische Unterrichtsforschung eher wenig gängige Analyseverfahren vorgestellt werden, die dennoch von Interesse sein können.

Tuma/Schneider/Knoblach (2013): Videographie.

Auch diese Einführung bietet eine umfassende Auseinandersetzung mit dem Thema – wobei die Rezeption der Perspektive interpretativer Sozialforschung für deutschdidaktische Fragestellungen eine anregende Bereicherung darstellen kann.

Neumann/Mahler (2014): Empirische Methoden der Deutschdidaktik: audio- und videografierte Unterrichtsforschung.

Dieser Band umfasst zahlreiche Beiträge, die sich mit videographischer Unterrichtsforschung in der Deutschdidaktik auseinandersetzen und jeweils ganz konkrete Anwendungsmöglichkeiten und Vorgehensweisen nachzeichnen. Er bietet Möglichkeiten zur vertieften – spezifisch deutschdidaktischen – Auseinandersetzung.

2. Videographische Datenerhebung konkret: Beispielprojekte

Nachfolgend werden konkrete Erhebungssettings vorgestellt, die aus unterschiedlichen Forschungsprojekten stammen. Damit soll verdeutlicht werden, wie stark die konkrete Erhebung vom Erkenntnisinteresse abhängt. Es wird jeweils auf das Erkenntnisinteresse der Studie, die Anzahl von Kameras und Kamerapersonen sowie Foki bzw. Methoden der Analyse eingegangen, um dies zu illustrieren.

Geht es um videographische Datenerhebung von Unterricht, dann wird generell empfohlen, mit zwei Kameras aufzunehmen, wobei eine Kamera die Lehrperson(en) bzw. den vorderen Bereich des Klassenraums, in welchem sich Tafel und ‚Lehrerpult‘ befinden, fokussiert. Eine weitere Kamera erfasst nach Möglichkeit in der Totale alle Schülerinnen und Schüler im Klassenzimmer (vgl. z.B. Irion/Knecht 2010, 1; Dinkelaker/Herrle 2009, 25; Seidel et al. 2005, 31f.).

Auch im Fach-DaZ-Projekt wurde diesen Empfehlungen gefolgt, wie Abbildung 1 zeigt (vgl. Ahrenholz 2013 für weiterführende Informationen zum Projekt). Schülerinnen und Schüler, welche keine Erlaubnis zur Aufnahme von Videodaten gegeben haben, werden außerhalb des von der Kamera erfassten Sichtfeldes platziert (Seidel et al. 2005). In Abbildung 1 betraf dies die Schülerinnen und Schüler S03, S05 und S07 (gekennzeichnet durch Unterstreichung).

Erkenntnisinteresse: Erfassung und Beschreibung der sprachlichen Mittel der Wissensvermittlung in der Schule auf einer empirischen Basis für verschiedene Fächer und Altersgruppen.

Anzahl von Kameras, Kamerapersonen und Audioaufnahmegeräten: I.d.R. wurden zwei Kameras, eine Lehrer- und eine Schülerkamera eingesetzt, wobei immer eine Kameraperson an der schwenkbaren Lehrerkamera positioniert war, z.T. eine weitere Kameraperson an der Schülerkamera. Ein Audioaufnahmegerät lag i.d.R. am Lehrerpult und bei Bedarf, z.B. bei Gruppenarbeit, wurden weitere Audioaufnahmegeräte verwendet, die auf den Tischen der Schülerinnen und Schüler positioniert waren. Die Kamerapersonen waren angehalten, sich möglichst zurückhaltend zu verhalten.

Fokus bzw. Methoden der Analyse: Transkription und Feinanalyse des sprachlichen, fachlichen Inputs, um Charakteristika dessen herauszuarbeiten (Maak 2014b).

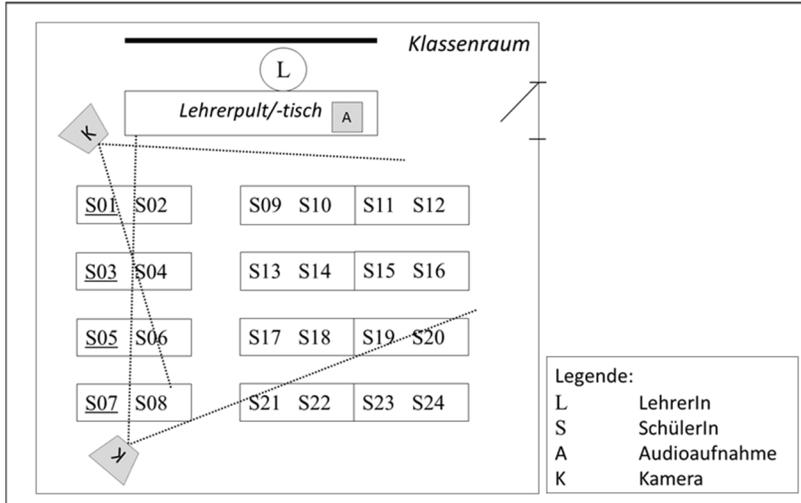


Abb. 1: Kameraeinstellung bei Unterrichtsaufnahmen, Frontalunterricht (Abbildung stammt aus Maak 2014b)

Dieses Setting ist also ein gängiges. Allerdings verdeutlichen Kamerastandpunkte auch die Perspektive von Forschenden auf das Feld – die Blickrichtung der Unterrichtsforschung (Frankhauser 2013, 29). Frankhauser (2013, 31) argumentiert dafür, dass unterrichtsvideographische Datenerhebung häufig die Rollenformigkeit eines institutionell gerahmten Unterrichts, und damit auch normative Setzungen und Interessen, wiederholt. Für die obige Konstellation bedeutet dies:

Unschwer ist zu erkennen, dass hier der Blick der Lehrperson auf ihre Klasse und derjenige des Schülers/der Schülerin auf ihre Lehrperson auf die beiden Kameraaugen übertragen worden sind. Die beiden Perspektiven verdeutlichen ein Strukturmodell von Unterrichtsinszenierung, wie sie vielen Unterrichtsbeobachtungen zugrunde liegt: hier die Lehrperson, auf der Hauptbühne neben der Tafel, die das Geschehen in der Klasse verfolgt, orchestriert und steuert – dort der oder die einzelne Lernende, dessen/deren Anpassungsfähigkeit sich nicht zuletzt darin zeigt, wie gut er/sie der Lehrperson ‚folgen‘ kann. Der Statik der Standkamera, die gewissermaßen das Lehrer/innenaue repräsentiert, welches das Ganze überblicken soll, steht dabei die Flexibilität der Handkamera, die das Schüler/innenaue verbildlicht, komplementär gegenüber. (ebd., 30)

Eigene Erhebungssettings sind entsprechend kritisch zu reflektieren. Nachfolgend werden weitere Formen videographischer Datenerhebungen in verschiedenen Projekt- und Forschungskontexten vorgestellt, um die Vielfalt der Möglichkeiten bei der videographischen Datenerhebung zu veranschaulichen.

2.1 Das MaTS-Projekt: Interviews mit DaZ-Seiteneinsteigerinnen und -einsteigern

Erkenntnisinteresse: Perspektive von DaZ-Seiteneinsteigerinnen und -einsteigern auf die Eingliederung in das deutsche Schulsystem

Anzahl von Kameras, Kamerapersonen und Audioaufnahmege­räten: Es wurde jeweils eine Kamera eingesetzt sowie ein Audioaufnahmege­rät. Beides wurde vorab von der Interviewenden positioniert. Eine Kameraperson war nicht anwesend (AJ01 = Personencode für Schüler, Int = Interviewerin, LT = Lehrertisch, K = Kamera).

Fokus bzw. Methoden der Analyse: Transkription und Auswertung mittels qualitativer Inhaltsanalyse (Ahrenholz/Maak 2013; Maak 2014a)

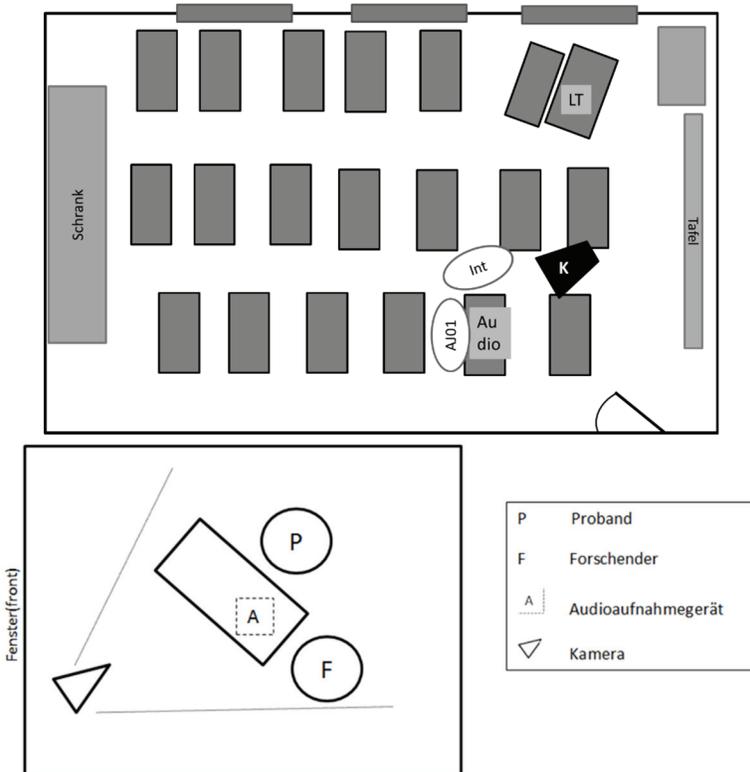


Abb. 2: Beispiel für die Aufnahme von Interviews mit ein bis zwei Schülerinnen und Schülern

Die Kamera ist so positioniert, dass der Schüler bzw. die Schülerin frontal fokussiert wird, der Interviewende ‚auf Eck‘ sitzt und so vornehmlich von der Seite sichtbar ist. Da sich die befragten Schülerinnen und Schüler während des Gesprächs dem Interviewenden zuwenden, wird vermieden, dass die Kamera ständig in den Blick genommen wird. Auf dem Tisch liegt zusätzlich ein Audioaufnahmege­rät.

2.2 Ricart Brede (2011): Sprachförderung in Kleingruppen vorschulischer Sprachförderung:¹⁰

Erkenntnisinteresse: u.a. „Welche Aktivitäten werden im Rahmen der Sprachförderheiten in welcher Häufigkeit und in welchem zeitlichen Umfang ausgeführt?“ (Ricart Brede 2011, 100)

Anzahl von Kameras, Kamerapersonen und Audioaufnahmegeräten: Eingesetzt wurden zwei Kameras, eine Erzieherinnenkamera (E-K) und eine Gruppenkamera (G-K). Die Erhebung wurde von einer Kameraperson durchgeführt, die ggf. die Veränderung der Kamera (vgl. Abb. 3) vornahm und die sich möglichst zurückhaltend verhielt. Eine zusätzliche Audioaufnahme wurde nicht gemacht.

Fokus bzw. Methoden der Analyse: vornehmlich Sichtstrukturen mittels niedrig- bis mittelinferentem¹¹ Beobachtungssystem (vgl. ebd.)

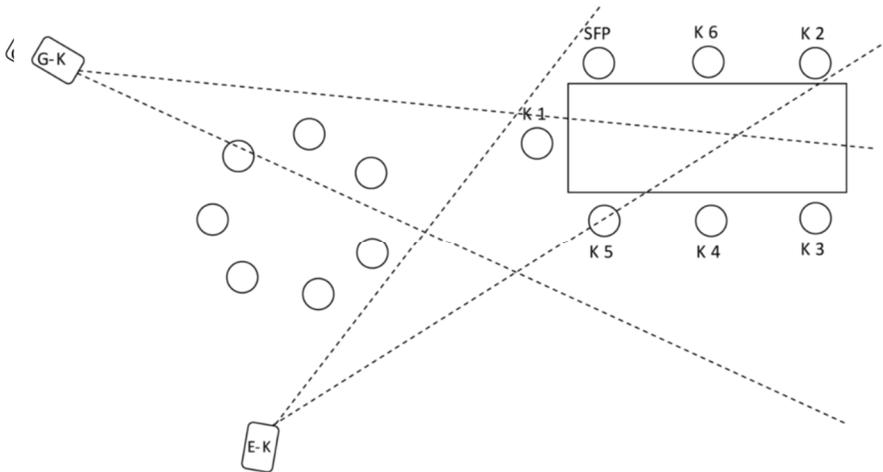


Abb. 3: Beispiel für eine Kamerapositionierung mit Ortswechsel während der Sprachförderung (entnommen mit freundlicher Genehmigung aus Ricart Brede 2011, 100)

¹⁰ Die Publikation ist im Buchhandel nicht mehr erhältlich. Bei Interesse wenden Sie sich aber gerne an Julia Ricart Brede (julia.ricart.brede@uni-flensburg.de).

¹¹ *Niedrig-inferente Kodierungen* beziehen sich auf die Erfassung von *direkt beobachtbaren* Aspekten der Sicht- oder Oberflächenstruktur des Unterrichts, wie z.B. die Unterrichtsorganisation hinsichtlich der Interaktionsstruktur und der Sozialformen, die Formulierung des Lektionsziels durch die Lehrperson beim Stundeneinstieg oder die Verwendung von technischen Hilfsmitteln, wie Taschenrechner und Computer im Verlauf der Lektion, usw. (Petko et al. 2003, 275). Hoch-inferente Kodierungen bedeuten i.d.R. eine höhere Interpretationsleistung (ebd., 275f.).

Ricart Brede begründet die Kamerapositionierung wie folgt:

Die beiden Kameras sind darin so positioniert, dass sie den zuvor erfragten Wechsel der Gruppe vom Sitzkreis an den Tisch ausschließlich über kleine Schwenks und Zooms der Kamera, d.h. ohne Veränderung der Stativposition mitvollziehen können. (ebd., 99)

Dieses Beispiel verdeutlicht, wie entscheidend die gründliche Vorbereitung von videographischer Datenerhebung sein kann. Ohne im Vorfeld erfolgte Absprachen und entsprechende Abstimmung der Kamerapositionen hätte die Sprachförderperson z.B. plötzlich mit dem Rücken zu E-Kamera sitzen oder die Kameras hätten beim Positionswechsel im Weg stehen können. Ferner wäre eine Neupositionierung höchst invasiv, insbesondere da es sich hier um die Aufnahme in einer Kleingruppe handelt.

2.3 Kamera-Ethnographie

Die in den vorherigen Beispielen angestrebte ‚Standardisierung‘ der Erhebung dient u.a. der besseren Vergleichbarkeit von Daten bzw. Ergebnissen. Eine solche Standardisierung findet im Rahmen der Kamera-Ethnographie nicht statt. Vielmehr dient eine flexible Handhabung der Kamera, häufig von der Kameraperson getragen und im Raum bewegt, dazu, bewusst bestimmte Aspekte zu fokussieren. Diese Fokussierung kann sich spontan im Laufe der Erhebung ergeben und muss nicht vorab geplant sein. Kamera und Kameraperson sind somit beweglich und verändern ihre Position wie auch Fokussierung ggf. im Laufe der Aufnahme.¹²

Zusammenfassend ist festzuhalten, dass die videographische Datenerhebung eine Fülle von Möglichkeiten bzgl. Aufbereitung und Auswertung eröffnet. Ferner wollte der Beitrag aber auch verdeutlichen, dass die Erhebung selbst wohlüberlegt und gut vorbereitet werden sollte – daher entfällt auch bei dieser Methode nicht die Notwendigkeit der Konkretisierung von Forschungsfragen und Erkenntnisinteresse, bevor die eigentliche Erhebung stattfindet. Auch die Möglichkeit des Rückgriffs auf bereits vorhandene Videodaten sollte daher stets überprüft werden (vgl. z.B. Hinweise in Krelle 2014, 83).

Literatur

- Ahrenholz, Bernt: Sprache im Fachunterricht untersuchen. In: Röhner, Charlotte/Hövelbrinks, Britta (Hrsg.): Fachbezogene Sprachförderung in Deutsch als Zweitsprache. Theoretische Konzepte und empirische Befunde zum Erwerb bildungssprachlicher Kompetenzen. Weinheim: Beltz Juventa 2013, 87-98.
- Ahrenholz, Bernt/Maak, Diana: Zur Situation von SchülerInnen nicht-deutscher Herkunftssprache in Thüringen unter besonderer Berücksichtigung von Seiteneinsteigern. Abschlussbericht zum Projekt "Mehrsprachigkeit an Thüringer Schulen (MaTS)", durchgeführt im Auftrag des TMBWK. (2013) Unter Mitarbeit von Isabel Fuchs,

¹² Für weitere Hinweise vgl. auch Mohn (2007) und Mohn/Amann (1998).

- Britta Hövelbrinks, Julia Ricart Brede und Wolfgang Zippel. (Berichte und Materialien <http://www.daz-portal.de>, 1). Online unter: http://www.daz-portal.de/images/Berichte/bm_band_01_mats_bericht_20130618_final.pdf. (letzter Zugriff: 01.08.2018).
- Becker-Mrotzek, Michael/Vogt, Rüdiger: Unterrichtskommunikation. Linguistische Analysemethoden und Forschungsergebnisse. Tübingen: Niemeyer (Germanistische Arbeitshefte, Band 38) 2009.
- Brünner, Gisela: Analyse mündlicher Kommunikation. In: Becker-Mrotzek, Michael (Hrsg.): Mündliche Kommunikation und Gesprächsdidaktik. Baltmannsweiler: Schneider Verlag Hohengehren (Deutschunterricht in Theorie und Praxis, 3) 2012, 52-65.
- Corsten, Michael: Videographie praktizieren – Ansprüche und Folgen. Ein methodischer Streifzug durch die Beiträge des Bandes. In: Corsten, Michael/Krug, Melanie/Moritz, Christine (Hrsg.): Videographie Praktizieren. Herangehensweisen, Möglichkeiten und Grenzen. Wiesbaden: VS Verlag für Sozialwissenschaften 2010, 7-22.
- Dinkelaker, Jörg/Herrle, Matthias: Erziehungswissenschaftliche Videographie. Eine Einführung. Wiesbaden: VS Verlag für Sozialwissenschaften 2009.
- Frankhauser, Regula: Videobasierte Unterrichtsbeobachtung: die Quadratur des Zirkels? In: Forum: Qualitative Sozialforschung (2013), 14, 1, Artikel 24.
- Fuchs, Isabel/Maak, Diana/Ahrenholz, Bernt: Die Erstsprache(n) als Ressource beim Spracherwerb von SeiteneinsteigerInnen. In: Lütke, Beate/Petersen, Inger (Hrsg.): Deutsch als Zweitsprache – erwerben, lernen und lehren. Beiträge aus dem 9. Workshop "Kinder mit Migrationshintergrund, 2013. Stuttgart: Fillibach bei Klett 2014, 71-91.
- Halbmayer, Ernst/Salat, Jana: Ethnographie als Prozess der Datenerhebung (o.J.), <http://www.univie.ac.at/ksa/elearning/cp/qualitative/qualitative-51.html> (letzter Zugriff: 01.08.2018).
- Helmke, Andreas/Göbel, Kerstin/Hosendfeld, Ingmar/Schrader, Friedrich-Wilhelm/Helmke, Tuyet/Wagner, Wolfgang: Die Videostudie im DESI-Projekt: Anlage, Ziele, Kameramanual. Universität Koblenz-Landau: Campus Landau, 2007.
- Helmke, Andreas: Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts; Franz Emanuel Weinert gewidmet. Stuttgart: Klett 2009.
- Irion, Thomas/Knecht, Andreas: Empfehlungen zur Aufzeichnung und Aufbereitung von Unterrichtsaufnahmen für die Datenanalyse. Ergänzung auf beigefügter CD-Rom zum Artikel von Thomas Irion: Hypercoding in der empirischen Lehr-Lern-Forschung. In: Corsten, Michael/Krug, Melanie/Moritz, Christine (Hrsg.): Videographie Praktizieren. Herangehensweisen, Möglichkeiten und Grenzen. Wiesbaden: VS Verlag für Sozialwissenschaften 2010.
- Knapp, Werner/Ricart Brede, Julia: Videographie als Methode zur Aufzeichnung und Analyse sprachlicher Lehr- und Lernsituationen. Vorschläge zur Systematisierung am Beispiel (vor-)schulischer Sprachförderung. In: Ahrenholz, Bernt (Hrsg.): Einblicke in die Zweitspracherwerbsforschung und ihre methodischen Verfahren. Berlin: De Gruyter Mouton (DaZ-Forschung. Deutsch als Zweitsprache, Mehrsprachigkeit und Migration, 1) 2012, 219-236.

- Krammer, Kathrin/Reusser, Kurt: Unterrichtsvideos als Medium der Lehrerinnen- und Lehrerbildung. In: Beiträge zur Lehrerbildung (2005), 23, 35-50.
- Krelle, Michael: Videographie von Deutschunterricht – Methodische Anmerkungen und Hinweise. In: Neumann, Astrid/Mahler, Isabelle (Hrsg.): Empirische Methoden der Deutschdidaktik: audio- und videografierte Unterrichtsforschung. Baltmannsweiler: Schneider Verlag Hohengehren 2014, 78-97.
- Kromrey, Helmut: Empirische Sozialforschung. Stuttgart: UTB 2009.
- Labov, William: Das Studium der Sprache im sozialen Kontext. In: Klein, Wolfgang (Hrsg.): Aspekte der Soziolinguistik. Frankfurt a.M.: Athenäum 1971, 111-194.
- Maak, Diana/Ricart Brede, Julia: Empirische Erfassung von Invasivität in videografierten Lehr-Lernsituationen: Entwicklung und Erprobung eines Beobachtungssystems. In: Neumann, Astrid/Mahler, Isabelle (Hrsg.): Empirische Methoden der Deutschdidaktik. Audio- und videografierte Unterrichtsforschung. Baltmannsweiler: Schneider Verlag Hohengehren 2014, 151-173.
- Maak, Diana: „es WÄre SCHÖN, wenn es nich (.) OFT so diese RÜCKSchläge gäbe“ – Eingliederung von SeiteneinsteigerInnen mit Deutsch als Zweitsprache in Thüringen. In: Ahrenholz, Bernt/Grommes, Patrick (Hrsg.): Zweitspracherwerb im Jugendalter. Berlin: de Gruyter Mouton 2014a, 319-340.
- Maak, Diana: „Wo kommt das blut HER?“ Sprachliche Merkmale des fachlichen Inputs im Fachunterricht Biologie der Sekundarstufe I – eine Konzept-orientierte Analyse der Enkodierung von Bewegungsereignissen (Dissertation zur Erlangung des akademischen Grades eines Doctor philosophiae, eingereicht an der Friedrich-Schiller-Universität Jena im September 2014b).
- Mohn, Bina Elisabeth: Die Kunst des dichten Zeigens. Aus der Praxis kamera-ethnographischer Blickentwürfe. In: Binder, Beate/Neuland-Kitzerow, Dagmar/Noack, Karoline (Hrsg.): Kunst und Ethnographie: Zum Verhältnis von visueller Kultur und ethnographischem Arbeiten. Berliner Blätter 46/2008, LIT Verlag 2008, 61-72.
- Mohn, Bina Elisabeth: Kamera-Ethnografie: Vom Blickentwurf zur Denkbewegung. In: Brandstetter, Gabriele/Klein, Gabriele (Hrsg.): Methoden der Tanzwissenschaft. Modellanalysen zu Pina Bauschs „Sacre du Printemps“. Tanz Scripte, Band 4, Bielefeld: transcript-Verlag 2007, 173-194 (erweiterte Neuauflage 2015: 209-230).
- Mohn, Elisabeth/Amann, Klaus: Forschung mit der Kamera. In: Anthropolitan: Visuelle Anthropologie. Mitteilungsblatt der GeFKA 1998, 6, 4-20.
- Neumann, Astrid/Mahler, Isabelle: Empirische Methoden der Deutschdidaktik: audio- und videografierte Unterrichtsforschung. Baltmannsweiler: Schneider 2014.
- Petko, Dominik/Waldis, Monika/Pauli, Christine/Reusser, Kurt: Methodologische Überlegungen zur videogestützten Forschung in der Mathematikdidaktik. In: ZDM (2003), 35, 6, 265-280.
- Ricart Brede, Julia/Knapp, Werner/Gasteiger-Klicpera/Kurcharz, Diemut: Die Entwicklung von Beobachtungssystemen in der videobasierten Forschung am Beispiel von Qualitätsanalysen vorschulischer Sprachfördereinheiten. In: Aguado, Karub/Schramm, Karen/Vollmer, Helmut Johannes (Hrsg.): Fremdsprachliches Handeln beobachten, messen, evaluieren. Reihe: KFU, Band 37. Frankfurt a.M.: Peter Lang 2010, 257-275.

- Ricart Brede, Julia: Beobachten. In: Settinieri, Julia/Demirkaya, Sevilen/Feldmeier, Alexis/Gültekin-Karakoç, Nazan/Riemer, Claudia (Hrsg.): Einführung in empirische Forschungsmethoden für Deutsch als Fremd- und Zweitsprache. Stuttgart.: UTB 2014, 137-146.
- Ricart Brede, Julia: Videobasierte Qualitätsanalyse vorschulischer Sprachfördersituationen. Freiburg: Fillibach 2011.
- Schmelter, Lars: Gütekriterien. In: Settinieri, Julia/Demirkaya, Sevilen/Feldmeier, Alexis/Gültekin-Karakoç, Nazan/Riemer, Claudia (Hrsg.): Einführung in empirische Forschungsmethoden für Deutsch als Fremd- und Zweitsprache. Stuttgart: UTB 2014, 33-56.
- Schramm, Karen/Aguado, Karin: Videographie in den Fremdsprachendidaktiken. Ein Überblick. In: Aguado, Karin/Schramm, Karen/Vollmer, Helmut Johannes (Hrsg.): Fremdsprachliches Handeln beobachten, messen, evaluieren. Neue methodische Ansätze der Kompetenzforschung und der Videographie. Frankfurt a.M./New York: Peter Lang (Kolloquium Fremdsprachenunterricht, 37) 2010, 185-214.
- Seidel, Tina/Dalehefte, Inger Marie/Meyer, Lena: Standardized guidelines – How to collect videotapes. In: Seidel, Tina/Prenzel, Manfred/Kobarg, Mareike (Hrsg.): How to run a video study. Technical report of the IPN video study. Münster u.a.: Waxmann 2005. 29-53.
- Steinke, Ines: Gütekriterien qualitativer Forschung. In: Flick, Uwe/von Kardorff, Ernst/Steinke, Ines (Hrsg.): Qualitative Forschung. Ein Handbuch. Reinbek bei Hamburg: Rowohlt Taschenbuch-Verlag 2007, 319-331.
- Stigler, James W.: Video Surveys: New Data for the Improvement of Classroom Instruction. In: Paris, Scott G./Wellman, Henry M. (Hrsg.): Global prospects for education. Development, culture, and schooling. Washington, D.C. American Psychological Association 1998, 129-168.
- Tuma, René/Schnettler, Bernt/Knoblauch, Hubert: Videographie. Einführung in die interpretative Videoanalyse sozialer Situationen. Wiesbaden: Springer VS 2013.
- Virtuelles Datenschutzbüro: Anonymisierung und Pseudonymisierung i. S. d. BDSG. <https://www.datenschutz.de/feature/detail/?featid=101>, Stand 11.07.2013 (letzter Zugriff: 06.09.2015).

Teilnehmende Beobachtung

Am Beispiel der Textverstehenserhebung des Projektes *Arbeitskoffer zu den Steirischen Literaturpfaden des Mittelalters*

1. Vorstellung der teilnehmenden Beobachtung

Die teilnehmende Beobachtung hat ihre Wurzeln in der ethnologischen Feldforschung. Es handelt sich dabei um eine Sonderform wissenschaftlichen Beobachtens, bei der der Forschende aktiv am zu beobachtenden Geschehen partizipiert. Je nach Anwendung und Fragestellung können unterschiedlich stark strukturierte Daten gewonnen werden und mittels verschiedenen qualitativen wie auch quantitativen Verfahren zur Auswertung gelangen. Teilnehmende Beobachtung kommt häufig in Zusammenhang mit der Grounded Theory (s. Kubik in diesem Band) zum Einsatz.

1.1 Funktionsweise der teilnehmenden Beobachtung

Wissenschaftliche Beobachtung unterscheidet sich von Alltagsbeobachtung und -wahrnehmung vor allem dadurch, dass sie gezielt eingesetzt, kontrolliert und protokolliert wird (vgl. Beer 2008, 167). Je nach Beobachtungsfeld, Beobachterstatus und Beobachtungsinstrument können verschiedene Arten bzw. Varianten der Beobachtung unterschieden werden, wobei vielfach auch Zwischenstufen möglich sind:

- *Feld- vs. Laborbeobachtung* (auch: natürliche vs. künstliche Beobachtung): Von einer Feldbeobachtung spricht man, wenn Personen in ihrem natürlichen sozialen Umfeld beobachtet werden (z.B. Schülerinnen und Schüler im Klassenzimmer), während im Labor eine künstliche Beobachtungssituation geschaffen wird.
- *teilnehmende vs. nicht-teilnehmende Beobachtung* (auch: aktive und passive Beobachtung): Der nicht-teilnehmende oder passive Beobachtende (auch ‚complete observer‘ oder ‚reiner Beobachter‘) partizipiert nicht an der Situation und interagiert nicht mit den Beobachteten, wobei unterschiedliche Grade der Distanz möglich sind: Der Forschende kann dabei etwa im Raum ständig sichtbar sein oder sich versteckt hinter einer Einwegscheibe befinden. Der teilnehmende Beobachtende zeichnet sich durch seine Interaktion mit

dem sozialen Feld aus, wobei auch hier der Grad seiner Partizipation stark variieren und von einer Randposition als teilnehmend Anwesender bis hin zur totalen Integration in das soziale Geschehen reichen kann.

- *offene vs. verdeckte Beobachtung*: Offenheit bezieht sich auf die Transparenz der Beobachtungssituation für die Beobachteten, also ob sie wissen, dass sie beobachtet werden oder nicht. Auch Mischformen sind möglich, wenn etwa den Probandinnen und Probanden zwar bekannt ist, dass sie beobachtet werden, ihnen der Grad der Beobachtung oder das tatsächliche Ziel aber nicht bewusst sind.
- *strukturierte vs. unstrukturierte Beobachtung* (auch: systematisch/standardisiert vs. unsystematisch/unstandardisiert): Strukturierter Beobachtung liegt ein vorab festgelegtes Beobachtungsschema zu Grunde, das u.a. Beobachtungseinheiten und Kategorien festlegt, während Wahrnehmung und Aufzeichnung bei der unstrukturierten Beobachtung zunächst noch nicht eingeschränkt werden, also möglichst das ganze Geschehen beobachtet werden soll. Strukturierte Beobachtung findet vor allem dann Anwendung, wenn das Feld relativ gut bekannt ist und bereits Forschungshypothesen formuliert wurden. Umgekehrt kann unstrukturierte Beobachtung dazu dienen, das Feld überhaupt erst kennen zu lernen und in weiterer Folge Forschungsfragen zu entwickeln.
- *technisch vermittelte (auch: apparative) vs. unvermittelte Beobachtung*: Alternativ bzw. ergänzend zu menschlichen Beobachtenden können heute auch technische Hilfsmittel wie Kameras als Informationsüberträger zum Einsatz kommen, man spricht hierbei von technisch vermittelter Beobachtung (s. Maak in diesem Band), während bei unvermittelter Beobachtung ausschließlich die Wahrnehmung des menschlichen Beobachtenden als Forschungsgrundlage dient (vgl. hierzu etwa Lötscher 2014, 193ff.; Atteslander 2010, 86ff.; Greve/Wentura 1997, 26ff.; Beer 2008, 169ff.).

Wissenschaftliche Beobachtung kann sowohl zur Entwicklung als auch zur Beantwortung von Forschungsfragen genutzt werden (vgl. Fromm 2012, 9). Speziell in ersterem Fall kommt oft die Sonderform der teilnehmenden Beobachtung zum Einsatz, die ihren Ursprung in der ethnologischen Feldforschung hat und anfänglich vor allem zur näheren Erforschung von fremden Volksgruppen, Kulturen und Milieus eingesetzt wurde. Heute findet die teilnehmende Beobachtung u.a. auch in der Pädagogik und Didaktik Anwendung, wobei ‚Teilnahme‘ hier sehr unterschiedlich ausgelegt werden und ein ganzes Spektrum verschiedenartigen Engagiert-Seins des Forschenden umfassen kann (vgl. Hauser-Schäublin 2008, 38). Das reicht von einer Minimaldefinition der ‚physischen Anwesenheit‘ des Beobachtenden bis hin zur völligen Integration in die Gruppe bzw. das Feld.

Die Vorgehensweise bei teilnehmenden Beobachtungen unterscheidet sich – abhängig u.a. von der Art der Teilnahme, dem Feld und den Zielen der Erhebung – im Einzelnen mitunter beträchtlich, weshalb sie im Folgenden auch nur ganz grob

umrissen werden soll. Grundsätzlich lassen sich 4 Phasen unterscheiden: Planung, Beobachtung, Protokollierung und Auswertung.

1. Bei der Planung einer Beobachtung wird zunächst eine Phase der Vorbereitung und Informationssammlung erfolgen: Hierunter fallen etwa das grundsätzliche Vertrautmachen mit dem Feld sowie die Klärung rechtlicher und organisatorischer Details und gegebenenfalls das Einholen von Einverständniserklärungen. Außerdem gilt es, eine Entscheidung zu treffen über den Beobachtungsmodus, die Definition der Beobachterrolle (Teilnahme und Offenheit) und den Grad der Strukturiertheit der Beobachtung. Je nach Fragestellung kann es auch sinnvoll sein, Mischformen anzuwenden, etwa zwischen offenen Beobachtungsprotokollen und hochstrukturierten Beobachtungsinstrumenten. (vgl. Löttscher 2014, 197ff.) Agiert man das erste Mal als (teilnehmender) Beobachtender, empfiehlt es sich, genaue Wahrnehmung und Aufzeichnung vorab zu trainieren.
2. Es folgt die Phase der eigentlichen Beobachtung (wobei die Dauer der Beobachtungseinheit stark variieren kann) – also der Datensammlung, in der Eindrücke, Erfahrungen und Wahrnehmungen gesammelt werden. Je nach Beobachterrolle (insbesondere bei offener Beobachtung) ist es unter Umständen möglich und sinnvoll, bereits während der Beobachtung erste Notizen anzufertigen, die später als Gedächtnisstütze dienen können.
3. Das Verfassen des eigentlichen Beobachtungsprotokolls findet jedoch in der Regel im Anschluss an die teilnehmende Beobachtung statt und sollte möglichst unmittelbar geschehen, damit die Eindrücke noch präsent sind.
4. Anschließend erfolgt die (schrittweise) Auswertung: Ausgangspunkt zur Generierung von Forschungsergebnissen wird in der Regel das zunächst eventuell noch sehr umfangreiche und unstrukturierte Beobachtungsprotokoll sein. Je nach Anwendung und Fragestellung können unterschiedlich stark strukturierte Daten gewonnen werden und mittels verschiedenen qualitativen wie auch quantitativen Verfahren zur Auswertung gelangen.

1.2 Ausgangsdaten und Ergebnisse

Die teilnehmende Beobachtung erfolgt auf Grund ihrer Offenheit und Gegenstandsorientierung in den meisten Fällen unstrukturiert, d.h. ohne ein vorher festgelegtes Beobachtungsinstrument. Die Basis zur weiteren Auswertung stellt also das schriftliche Protokoll bzw. bei längerfristigen Beobachtungen eventuell zusätzlich ein Forschungstagebuch dar (vgl. hierzu Girtler 2001, 133ff.). Je nach Untersuchungssituation ist es auch dem teilnehmenden Beobachtenden unter Umständen möglich, bereits während der Beobachtungssituation erste Notizen anzufertigen, die später als Grundlage für die Protokollierung dienen können. Die gesammelten Eindrücke, Erfahrungen und Wahrnehmungen werden zunächst niedergeschrieben. Dabei gilt es zu beachten, dass das Protokoll so vollständig und ausführlich wie möglich den genauen Verlauf der Situation wiedergibt und dabei unmissverständlich formuliert wird. Verbale Kommunikation wird – soweit von

Forschungsinteresse – möglichst wörtlich notiert (vgl. hierzu etwa Lötscher 2014, 199ff.; Atteslander 2010, 99f.). Der Soziologe Roland Girtler empfiehlt, sich auch bei unstrukturierter teilnehmender Beobachtung an einem Rahmenschema zu orientieren und das Augenmerk auf folgende Punkte zu richten, die möglichst exakt beschrieben werden sollen:

1. die Teilnehmenden an den sozialen Situationen
2. die Durchführung der sozialen Situation
3. die Schaffung der sozialen Situation
4. die die Teilnehmenden an der Situation determinierenden Normen
5. die Regelmäßigkeit der sozialen Situation
6. die Reaktion, wenn die Teilnehmenden an der Situation den an sie gerichteten Erwartungen nicht entsprechen
7. den Unterschied zwischen Behauptetem und Getanem (vgl. Girtler 2001, 134ff.).

Zur Auswertung bieten sich Verfahren wie die objektive Hermeneutik (s. Pflugmacher in diesem Band) oder die qualitative Inhaltsanalyse (s. Heins in diesem Band) an. Ausgehend von ersten vorläufigen Themenkomplexen werden im laufenden Forschungsprozess schrittweise Daten und Hypothesen verknüpft und verallgemeinert (vgl. Atteslander 2010, 99). Wurde vorab ein teil- oder hochstrukturiertes Beobachtungsinstrument, etwa ein Kodierhandbuch, erstellt oder im Zuge der Auswertung ein Kategoriensystem entwickelt, können im Anschluss (zusätzlich) auch quantitative Verfahren zum Einsatz kommen.

1.3 Vorteile und Nachteile der teilnehmenden Beobachtung

Sorgfältig geplant und situationsadäquat eingesetzt, bietet die teilnehmende Beobachtung zahlreiche Vorzüge gegenüber anderen Methoden und die Chance auf die Entdeckung neuer Perspektiven. Der Blick in die einschlägige Forschungsliteratur offenbart jedoch auch mögliche Schwierigkeiten und Schwächen des Verfahrens, die allerdings nicht zwingend bei jeder Fragestellung in gleichem Maße relevant sind und sich teilweise bei entsprechend sorgfältiger Methodenauswahl, Planung und Durchführung vermeiden bzw. minimieren lassen.

Beobachtungen können dort ansetzen, wo Befragungen an ihre Grenzen stoßen, weil sie unnötig kompliziert wären, die Befragten nichts über das zu untersuchende Phänomen wissen, sprachliche Kommunikation nicht möglich ist – etwa bei Kleinkindern (vgl. Fromm 2012, 8) – oder wesentliche Aspekte ausgeklammert blieben. Natürliche Situationen können in ihrem natürlichen Umfeld beobachtet werden. Konkrete Handlungs- und Vorgehensweisen, Kommunikation und Interaktion in sozialen Situationen können erfasst werden, ohne auf die Einschätzungen und Darstellungen der Betroffenen angewiesen (vgl. Lötscher 2014, 192) oder – wie etwa bei Fragebögen und standardisierten Tests – von ihrer Verσιertheit im Umgang mit dem Erhebungsinstrument abhängig zu sein.

Die teilnehmende Beobachtung als Sonderform erlaubt es dem Forschenden, direkt mit seinem Feld zu interagieren, was durch Selbsterfahrung und

-wahrnehmung zu einem vertieften Verständnis führt (vgl. ebd., 193). Durch Interaktion mit den Beteiligten kann der Beobachtende gezielt in die Situation eingreifen und Vorgänge steuern, um so zum Erkenntnisgewinn beizutragen.

Teilnehmende Beobachtung bietet sich insbesondere dann an, wenn über ein Forschungsfeld noch wenig bekannt ist. Sie ermöglicht einen explorativen Zugang, der auf ähnlich vielgestaltige und informationsreiche Weise kaum anders herstellbar wäre (vgl. Fromm 2012, 14) und auch dazu genutzt werden kann, um überhaupt erst Forschungsfragen zu generieren.

Beobachtungsverfahren sind allerdings in der Regel relativ aufwändig in der Durchführung, sodass es schwierig ist, größere Stichproben zu untersuchen und statistisch auszuwerten (vgl. ebd., 9). Wie bei jeder Form der offenen Beobachtung besteht auch bei (offener) teilnehmender Beobachtung die Möglichkeit, dass sich das Verhalten der Beobachteten durch die Beobachtung verändert. Inwieweit durch die Methode der teilnehmenden im Gegensatz zur nicht-teilnehmenden Beobachtung derartige Reaktivitätseffekte vermindert werden oder sogar stärker auftreten (können), gilt als umstritten (vgl. Greve/Wentura 1997, 29). Die Konzentration auf einzelne Situationen oder Personen, kann – wie die eigene Interaktion in der Gruppe – dazu führen, dass das Geschehen nur selektiv wahrgenommen wird und so zeitgleich ablaufende Vorgänge verborgen bleiben. (Teil-)strukturierte Beobachtungsprotokolle können zwar zur Fokussierung auf forschungsrelevante Aspekte innerhalb der Beobachtung hilfreich sein, allerdings bergen sie auch die Gefahr, den Blick schon zu Beginn allzu sehr einzuschränken. Durch die Partizipation am Geschehen ist eine umfassende begleitende Datensammlung bzw. Protokollierung kaum möglich. Die Protokollierung muss also in der Regel im Nachhinein geschehen und ist daher u.a. stark von der Gedächtnisleistung des Beobachtenden abhängig. In diesem Zusammenhang kann es neben subjektiven Verzerrungen auch zum sogenannten ‚primacy-recency‘-Effekt kommen. (vgl. ebd., 67) Zudem steigt durch die zeitliche Distanz unter Umständen die Gefahr von Fehlinterpretationen. Die unter Umständen auftretende emotionale Involviertheit des teilnehmenden Beobachtenden durch seine Interaktion im sozialen Geschehen kann vor allem bei längerfristigen Beobachtungen Selektions- und Verzerrungseffekte begünstigen.

1.4 Gütekriterien für den Einsatz der teilnehmenden Beobachtung

Teilnehmende Beobachtung basiert in jedem Fall auf sozialen Beziehungen zwischen ForscherIn und den zu untersuchenden Menschen. Sie ist – als Methode – das Gegenteil von Untersuchungen im Labor, bei denen möglichst Dinge und Sachverhalte untersucht werden, auf die der Beobachter gar nicht oder nur gering einwirkt. Laborexperimente lassen sich – unabhängig von der Person des Forschers – nahezu beliebig wiederholen; die Resultate gelten deshalb als intersubjektiv überprüfbar und objektiv. Teilnehmende Beobachtung ist von jeder Situation vor Ort, von den Menschen, die gerade daran teilhaben und wie sie miteinander agieren, abhängig; sie lässt sich deshalb nie genau gleich wiederholen. Die mittels Teilnehmender Beobachtung gewonnenen Daten sind immer von den Interaktionen des Forschers mit seinem Untersuchungsfeld geprägt. (Hauser-Schäublin 2008, 38)

Hier wird deutlich, dass teilnehmende Beobachtung nicht immer den Kriterien für wissenschaftliche Beobachtungen im Allgemeinen entsprechen kann, für die in der Regel Replizierbarkeit und Objektivität neben Validität und Generalisierbarkeit als maßgebliche Gütekriterien gelten (vgl. Greve/Wentura 1997, 13, 50ff.). Abhängig von der Ausgestaltung der Beobachterrolle bzw. des Grads der Teilnahme werden die Ergebnisse einer teilnehmenden Beobachtung eher selten genau replizierbar und damit intersubjektiv überprüfbar sein, was nicht bedeuten muss, dass hier Validität und Generalisierbarkeit nicht angestrebt werden können. Um Fehlinterpretationen und subjektiven Fehleinschätzungen vorzubeugen, können etwa im Interpretations- bzw. Auswertungsprozess auch mehrere Forschende das vorhandene Datenmaterial einer Beobachtung sichten. Girtler schlägt in diesem Zusammenhang in seinen Überlegungen zur freien teilnehmenden Beobachtung zudem vor, das erarbeitete Material gemeinsam mit einzelnen Mitgliedern der beobachteten Gruppe kritisch zu sichten (vgl. Girtler 2001, 128f.), was allerdings gerade im didaktischen Kontext sicherlich nicht bei jeder Untersuchung gleich gut anwendbar oder zielführend sein dürfte.

Der Schlüssel für eine erfolgreiche Datensammlung und -auswertung liegt in der Person des Beobachtenden selbst, die als Wahrnehmungsinstrument auf unterschiedlichen Ebenen fehleranfällig ist. So kann es im Beobachtungsprozess neben Wahrnehmungsfehlern auch zu Deutungsfehlern, Erinnerungsfehlern und Wiedergabefehlern kommen, denen man durch Planung, Sorgfalt und Training Herr werden sollte (zur Vermeidung von Beobachtungsfehlern sowie zu Auswahl, Training und Kontrolle von Beobachtern vgl. Greve/Wentura 1997, 74ff.).

Das Qualitätskriterium für die Verschriftlichung von Beobachtungen liegt einerseits in der Ausführlichkeit und Detaillierung des Schreibens und andererseits in der Reflexion und zunehmenden Fokussierung und analytischen “Verdichtung” des Materials. (Breidenstein 2012, 33)

Für die Protokollierung und damit den Grundstein zur weiteren Aufbereitung wie Verwertung der gesammelten Daten sollte also ein hoher Grad an Detaillierung und hinreichender Intensität angestrebt werden (vgl. ebd., 42), was ausschließlich zeitnah zur Beobachtungseinheit geschehen kann. Da es auf Grund der Offenheit und Gegenstandsorientierung keine standardisierten Auswertungsverfahren für

die teilnehmende Beobachtung gibt, liegt es letztlich alleine in der Verantwortung des Forschenden, zu einer sinnvollen und die Wirklichkeit abbildenden Interpretation, theoretischen Einordnung und Darstellung der gewonnenen Daten zu gelangen (vgl. Girtler 2011, 141ff.; Atteslander 2010, 99f.).

1.5 Literaturempfehlungen

Atteslander (2010): Methoden der empirischen Sozialforschung.

Das Einführungswerk von Peter Atteslander bietet einen umfassenden Überblick über Theorie und Praxis der empirischen Sozialforschung. Ein eigenes Kapitel befasst sich auf rund 30 Seiten mit der Beobachtung als Methode zur Erhebung sozialer Daten, auch auf die teilnehmende Beobachtung wird in diesem Zusammenhang näher eingegangen.

de Boer/Reh (2010): Beobachtung in der Schule – Beobachten lernen.

Der von Heike de Boer und Sabine Reh herausgegebene Sammelband widmet sich ausführlich den unterschiedlichsten Aspekten von Beobachtungen in der Schule und geht dabei auch immer wieder näher auf die Methode der teilnehmenden Beobachtung und ihren Einsatz im schulischen Kontext ein.

Fromm (2012): Beobachtung. Anleitung und Übung.

Der schmale Band setzt laut eigener Angabe dort an, „wo die meisten Lehrbücher zur Forschungsmethodik aufhören“ (Fromm 2012, 7). Auf 52 Seiten werden Anregungen und Hilfestellungen zur praktischen Umsetzung von Beobachtungen gegeben; im letzten Teil finden sich zwei Übungen zu den Beobachtungssystemen von Melbin sowie Jecker-Maccoby-Breitrose.

Girtler (2001): Methoden der Feldforschung.

Der Soziologe und Kulturanthropologe Roland Girtler gilt als Pionier der qualitativen Sozialforschung und hat die Entwicklung der teilnehmenden Beobachtung als Methode maßgeblich mitgeprägt. Auch wenn sich methodisch aus dem vorliegenden Handbuch zur ethnologischen Feldforschung vieles nicht unverändert in den Bereich der Deutschdidaktik übernehmen lassen wird, bietet es doch einen guten Einstieg in die Arbeit mit der teilnehmenden Beobachtung.

Greve/Wentura (1997): Wissenschaftliche Beobachtung.

Das Einführungswerk für Studierende der Psychologie von Werner Greve und Dirk Wentura soll mit den wichtigsten Problemen und Lösungsansätzen im Umgang mit Beobachtungsverfahren vertraut machen und bietet somit vor allem für Einsteigerinnen und Einsteiger einen guten ersten Überblick.

2. Die Textverstehenserhebung im Rahmen des Sparkling Science Projektes *Arbeitskoffer zu den Steirischen Literaturpfaden des Mittelalters*

Das Sparkling Science Projekt *Arbeitskoffer zu den Steirischen Literaturpfaden des Mittelalters*¹ lehnt sich bereits im Titel eng an sein ‚Mutterprojekt‘, die *Steirischen Literaturpfade des Mittelalters*,² an, dessen bildungstouristisches Angebot rund um die regionale ältere deutsche Literatur als Ausgangspunkt für ein innovatives didaktisches Vermittlungskonzept dient, das die vielfältige wie reizvolle (literarische) Welt des steirischen Mittelalters für Lehrerinnen und Lehrer, Schülerinnen und Schüler und Studierende als (außer-)schulischen und digitalen Lernort erschließen soll. Ausgehend von zwölf Basistexten aus dem Umfeld der acht *Steirischen Literaturpfade* waren im Rahmen des Arbeitskoffer-Projekts zunächst die Voraussetzungen zu klären, die für eine auf die Bedürfnisse von Lehrenden und Lernenden ausgerichtete Erschließung dieser Textzeugnisse im und für den Unterricht nötig sind. Die anschließende bedarfsgerechte Aufbereitung der mittel- und frühneuhochdeutschen Literatur schloss auch die Entwicklung von qualitativ hochwertigen und niederschwellig zugänglichen Unterrichtsmaterialien mit ein, die den Anforderungen des gegenwärtigen Bildungssystems Rechnung tragen und den aktuellen fachdidaktischen Forschungsstand berücksichtigen.

Im Zuge dieses Projektes (als dessen wissenschaftliche Koordinatorin die Verfasserin tätig ist) wird seit 2013 auch eine mehrteilige empirische Untersuchung durchgeführt, die sich damit befasst, wie Lernende mittelalterliche Texte verstehen. Auf die im Folgenden näher dargestellte erste Phase, in der die teilnehmende Beobachtung zum Einsatz kam, folgte eine groß angelegte schriftliche Erhebung, die bis dato noch nicht abgeschlossen ist. Lernende aus Schule und Hochschule werden in dieser zweiten Phase mit mittel- und frühneuhochdeutschen Textauschnitten konfrontiert. Dies geschieht mittels spezieller Textblätter, die drei Ebenen mit Angaben zur Ermittlung des Textverstehens enthalten und später zur statistischen Auswertung kodiert in SPSS-Datenbanken übertragen werden. Die teilnehmende Beobachtung stellte in diesem Zusammenhang eine Art von Vorerhebung dar, auf deren Basis die weitere Vorgehensweise – die genauen Forschungsfragen sowie das endgültige Design bis hin zur Textstellenauswahl der schriftlichen Erhebung – entwickelt wurde. Gleichzeitig dienten die in der Beobachtungsphase gewonnenen Erkenntnisse später bei der Auswertung der Studie als Hilfestellung zur Interpretation der Ergebnisse. Es gelang so u.a., ein Gefühl dafür zu bekommen, welche Hilfestellungen gegeben werden können und müssen, um die erfolgreiche Entschlüsselung eines mittelalterlichen Textes in Originalsprache zu

¹ Nähere Informationen finden sich über das Textportal des Projektes <http://gams.uni-graz.at/literaturpfade-arbeitskoffer> sowie unter [http://www.sparklingscience.at/de/projects/show.html?--typo3_neos_nodetypes-page\[id\]=599](http://www.sparklingscience.at/de/projects/show.html?--typo3_neos_nodetypes-page[id]=599) (1. Laufzeit 2012-2015) und [http://www.sparklingscience.at/de/projects/show.html?--typo3_neos_nodetypes-page\[id\]=742](http://www.sparklingscience.at/de/projects/show.html?--typo3_neos_nodetypes-page[id]=742) (2. Laufzeit 2014-2016).

² <http://literaturpfade.uni-graz.at/de/>

ermöglichen. Darüber hinaus lieferten die Beobachtungsprotokolle und Erfahrungen auch weiterführende Erkenntnisse, die im Rahmen zukünftiger Projekte untersucht werden sollen.

2.1 Zentrale Fragestellungen

Ausgehend von der Leitfrage, wie Schülerinnen und Schüler mit mittel- und frühneuhochdeutschen Texten umgehen, erschienen im Zuge der Beobachtung unter anderem folgende Aspekte von Interesse:

- Welche Textstellen sind für die Zielgruppe besonders schwierig oder auch spannend?
- Welche Strategien wenden Schülerinnen und Schüler an, um Texte der älteren Sprachstufen zu entschlüsseln?
- Welche Hilfestellungen erleichtern den Zugang zum Text (Hinweise sprachlicher und inhaltlicher Natur, Vorlesen, Hörverstehen ...)
- Wie kommen Schülerinnen und Schüler mit dem mittelhochdeutschen Wörterbuch zurecht?
- Wie verläuft die Anschlusskommunikation und welche Schlüsse lassen sich daraus ziehen?
- Welche Texte und Teilaspekte eignen sich für welche Einsatzzwecke im Unterricht?
- Was kann besonders gut mit welchen Methoden unterrichtet werden?

2.2 Erhebung

Zwölf Gruppen aus je vier bis acht Schülerinnen und Schülern der Sekundarstufe II arbeiteten im Rahmen des Sparkling Science Projektes etwas über ein Semester lang mit jeweils einem steirischen Text der mittel- oder frühneuhochdeutschen Sprachstufe. Zu ihren Aufgaben zählte dabei neben der Entschlüsselung und Übersetzung des Textes auch seine Aufbereitung für das Textportal des Projektes (inklusive Kommentierung und Verfassen von begleitenden Wiki-Artikeln), die gemeinsame Konzeption und Durchführung einer Unterrichtseinheit sowie Entwicklung und Umsetzung eines frei gewählten Kreativprojektes zum Text (z.B. Video, Bildband, Frauenmagazin). Als teilnehmende Beobachterinnen und Beobachter fungierten zwölf Lehramtsstudierende der Universität Graz, die im Vorfeld entsprechend geschult und instruiert wurden. Sie begleiteten jeweils eine Gruppe über fünf Monate hinweg in der Arbeit mit einem mittelalterlichen Text.

In vorliegendem Fall kam eine halb-offene, teil-standardisierte teilnehmende Beobachtung unter ‚kontrolliert-natürlichen‘ Versuchsbedingungen zum Einsatz.

- *halb-offen*: Die Jugendlichen wussten, dass sie Teil eines wissenschaftlichen Projektes sind und wir anhand ihrer Übersetzungen herausfinden wollten, wie Schülerinnen und Schüler allgemein mit diesen Texten umgehen können und wie Texte aufbereitet werden sollen, damit sie bestmöglich für den Einsatz im Unterricht geeignet sind. Sie konnten während der Übersetzungsarbeit

auch sehen, dass die Beobachterinnen und Beobachter Markierungen und Notizen anfertigten. Der Grad der Beobachtung bzw. alle Forschungsinteressen waren ihnen aber nicht bekannt.

- *teil-standardisiert*: Neben dem freien Beobachtungsprotokoll stand den teilnehmenden Beobachterinnen und Beobachtern mit den gemeinsam mit den Schülerinnen und Schülern ausgefüllten Textblättern und einem Kodierungsschema ein zusätzliches standardisiertes Instrument zur Verfügung, das allerdings nur auf einzelne Fragestellungen anwendbar war. Die teilnehmenden Beobachterinnen und Beobachter, die im Rahmen des Projektes ‚Textcoaches‘ genannt wurden, hatten in der Übersetzungsphase vor sich eine Version des mittelalterlichen Textes liegen, auf der sie zunächst jedes Wort mit einem farbigen Marker (grün, gelb, rot) kennzeichneten, je nachdem, ob die Schülerinnen und Schüler es auf Anhieb verstanden hatten, es mit Hilfestellung entschlüsseln konnten oder aber gar nicht in der Lage waren, die Textstelle zu verstehen. Zusätzlich gab es ein System aus Buchstabencodes, mit dem weitere Informationen festgehalten werden konnten (z.B. WB – Wörterbuch wurde zu Hilfe genommen, V – Textstelle wurde zusätzlich laut gelesen/vorgelesen, S – Hinweis auf Sprachentwicklung, D – Hinweis auf dialektale Varianten etc.). Darüber hinaus hatten sie die Möglichkeit, sich während der Beobachtungseinheiten, die jeweils 4-6 Schulstunden betrug, bei Bedarf mittels eigener Kürzel zusätzliche kleine Notizen zu machen. Das eigentliche Beobachtungsprotokoll wurde auf Basis der Notizen und einiger grundlegender Richtlinien und Anweisungen, die sie vorab erhalten hatten, jeweils am Tag selbst verfasst (inkl. einer kritischen Reflexion bzw. Einschätzung der eigenen Beobachterrolle) und gemeinsam mit den Textblättern zeitnah an die Untersuchungsleiterin übermittelt.
- *teilnehmend*: Die Textcoaches waren während der gesamten Projektzeit von Februar bis Juli 2013 Teil der Gruppe, sie begleiteten ihre Schülerinnen und Schüler als Gruppenleiterinnen und -leiter sowie als Vertrauenspersonen durch alle Phasen der gemeinsamen Projektarbeit. Auch wenn sie als Studierende des Unterrichtsfaches Deutsch am Ende des ersten Studienabschnittes gegenüber den Jugendlichen natürlich einen Wissensvorsprung hatten, was die Übersetzung mittelalterlicher Texte anbelangt und sie sich im Vorfeld mit ihrem jeweiligen Text vertraut machen konnten, waren ihre eigenen Erfahrungen und Wahrnehmungen während der Textarbeit durchaus ebenso von Interesse für die Untersuchung. Durch ihre Rolle in der Gruppe hatten sie zudem jederzeit die Möglichkeit, gezielt steuernd einzugreifen und mussten dies sogar tun, indem sie etwa an entsprechender Stelle Hinweise zur Entschlüsselung des Textes gaben.
- *kontrolliert-natürlich*: Die Erhebung fand in der natürlichen Umgebung der Schülerinnen und Schüler statt, nämlich an ihrer Schule, in ihrer jeweiligen Klasse in Form einer Gruppenarbeit. Die Arbeitssituation an sich (das Übersetzen eines mittelalterlichen Textes und seine weitere Bearbeitung) war aber

vom Projekt vorgegeben, also gewissermaßen eine künstliche bzw. kontrollierte.

2.3 Ergebnisse und Fazit

Als Daten lagen am Ende der Beobachtungsphase die farbig markierten Textblätter inklusive Kodierungen sowie umfangreiche Protokolle und Reflexionen der einzelnen Gruppenleiterinnen und -leiter vor, die seitens der Untersuchungsleiterin noch durch eigene Beobachtungen ergänzt wurden. In der anschließenden Auswertungsphase offenbarten sich zwei große Mankos der Methode: der Aufwand und die Beobachterabhängigkeit.

Zur personell wie zeitlich ressourcenintensiven Erhebungsphase, die nur durch die Mitwirkung studentischer Praktikanten verwirklicht werden konnte, kam der nicht zu unterschätzende Aufwand in der Auswertung: Neben den jeweiligen farbig markierten Textblättern inklusive Kodierungen und Notizen im Umfang von zehn bis zwölf Seiten lagen aus jeder der zwölf Gruppen pro Workshop drei- bis siebenseitige Beobachtungsprotokolle als Materialgrundlage vor. Die Zusammenführung und Auswertung der trotz Leitfadens doch recht heterogenen Aufzeichnungen und Beobachtungsprotokolle konnte durch die vorab festgelegten Kodierungen und die Möglichkeit, offene Fragen im Zuge von Einzelnachbesprechungen mit den Textcoaches zu klären sowie durch Ergänzungen der eigenen Beobachtungen der Studienleiterin ein wenig erleichtert werden. Als praktikabel erwies sich die Arbeit in Kleingruppen anstelle des Klassenverbandes, die sicherte, dass die Beobachterinnen und Beobachter auch wirklich alle Schülerinnen und Schüler gut im Blick hatten, allerdings auch zum großen Umfang der Beobachtungsprotokolle beitrugen.

Die Abhängigkeit von der Person des einzelnen Beobachtenden machte sich gerade in der Teamarbeit bzw. im Vergleich der einzelnen Gruppen bemerkbar und führte trotz gemeinsamer Vorbereitung und Richtlinien mitunter zu sehr heterogenen Ergebnissen. Zwei von zwölf Beobachtungen waren auf Grund massiver Mängel in der Durchführung und Protokollierung gar nicht zu verwerten, die beiden Beobachterinnen mussten im Laufe des Projektes ausgetauscht werden.

Trotz dieser Einschränkungen erwies sich die Vorstudie mittels teilnehmender Beobachtung als richtungsweisend für die spätere schriftliche Erhebung und führte zusätzlich zu den erwarteten Ergebnissen auch zu Erkenntnissen, die ohne den explorativen Charakter der teilnehmenden Beobachtung wohl kaum ans Tageslicht gekommen wären: u.a. etwa, dass die Übersetzungsarbeit und die damit verbundenen herausfordernden Dekodierungsleistungen einen unmittelbaren Einfluss auf prozedurale Ebenen der Lesekompetenz haben dürften. Das zeigte sich vor allem deutlich in jenen Gruppen, die zu Beginn kein oder kaum Interesse an den Gegenständen Literatur und Mittelalter hatten, zum Teil auch zu den schwächeren Schülerinnen und Schülern im Unterrichtsfach Deutsch zählten und zu Semesteranfang auch kaum Vorwissensbestände abrufen und benennen konnten. Diese Jugendlichen entwickelten im Laufe der Entschlüsselung des Textes zunehmend ‚need for cognition‘, das Bedürfnis nach einer kognitiven Durchdringung

des Textes, was sich unmittelbar auf ihre Lesemotivation und in weiterer Folge auf das Abrufen und Verknüpfen mit vorhandenen Wissensbeständen auswirkte. Im Vergleich zu anderen Schulen, mit denen im Rahmen des Projektes gearbeitet wurde, verlief in allen Gruppen, die mittel- und frühneuhochdeutsche Texte selbst übersetzt hatten, die Anschlusskommunikation (sowohl in der kreativen Auseinandersetzung mit den Texten und Themen als auch in den Gesprächen der Schülerinnen und Schüler untereinander) auf deutlich höherem Niveau. Diese Beobachtungen legen die Annahme nahe, dass gerade die Übersetzungsarbeit an älteren deutschen Texten, die (nicht nur) im schulischen Kontext meist bestenfalls als notwendiges Übel erachtet wird, viel häufiger jedoch als Ausschlusskriterium für diese Art von Literatur gilt, eine Schlüsselrolle im Zusammenhang mit genuinen Bildungspotentialen mittelalterlicher Texte spielen könnte (vgl. Schwinghammer 2015, 158ff.).

Auch wenn die teilnehmende Beobachtung „aufwändig und nicht von einer unmittelbaren ‚Verwertbarkeit‘ der Ergebnisse gekennzeichnet [ist]“ (Breidenstein 2012, 42f.) und letztlich vorweg niemand sagen kann, „ob und inwiefern die Analysen ‚hilfreich‘ für die Praxis sein werden“ (ebd., 43): Für das gegenständliche Projekt hat sich der Einsatz der teilnehmenden Beobachtung in jedem Fall gelohnt, da er es nicht nur erlaubte, die große schriftliche Erhebung auf ein sicheres Fundament zu stellen, sondern auch neue Erkenntnisse und Erfahrungswerte offenbarte, die in zwei neuen Projekten münden konnten und im Rahmen des Fortsetzungsprojekts ‚Arbeitskoffer 2.0‘ sowie im Zuge eines noch im Planungsstadium befindlichen Habilitationsprojektes (LLÄDT – Literarisches Lernen anhand älterer deutscher Texte) näher untersucht werden sollen.

Literatur

- Aeppli, Jürg/Gasser, Luciano/Gutzwiller, Eveline/Tettenborn, Annette: Empirisches wissenschaftliches Arbeiten. Ein Studienbuch für die Bildungswissenschaften. Bad Heilbrunn: Klinkhardt 2014.
- Atteslander, Peter: Methoden der empirischen Sozialforschung. Berlin: ESV 2010.
- Beer, Bettina: Systematische Beobachtung. In: Beer, Bettina (Hrsg.): Methoden ethnologischer Feldforschung. Berlin: Reimer 2008, 167-189.
- Breidenstein, Georg: Ethnographisches Beobachten. In: de Boer, Heike/Reh, Sabine (Hrsg.): Beobachtung in der Schule – Beobachten lernen. Wiesbaden: Springer 2012, 27-43.
- de Boer, Heike/Reh, Sabine: Beobachtung in der Schule – Beobachten lernen. Wiesbaden: Springer 2012.
- Fromm, Martin: Beobachtung. Anleitung und Übung. Norderstedt: Books on Demand 2012.
- Girtler, Roland: Methoden der Feldforschung. Wien/Köln/Weimar: Böhlau 2001.
- Greve, Werner/Wentura, Dirk: Wissenschaftliche Beobachtung. Eine Einführung. Weinheim: Beltz 1997.

- Hauser-Schäublin, Brigitta: Teilnehmende Beobachtung. In: Beer, Bettina (Hrsg.): Methoden ethnologischer Feldforschung. Berlin: Reimer 2008, 37-58.
- Lötscher, Hanni: Beobachtung. Gastbeitrag in: Aeppli, Jürg/Gasser, Luciano/Gutzwiller, Eveline/Tettenborn, Annette: Empirisches wissenschaftliches Arbeiten. Ein Studienbuch für die Bildungswissenschaften. Bad Heilbrunn: Klinkhardt 2014, 192-203.
- Mayring, Philipp: Qualitative Inhaltsanalyse. Grundlagen und Techniken. Weinheim/Basel: Beltz 2015.

Literatur zum Beispielprojekt:

- Schwinghammer, Ylva: Alte Sprache – schwere Sprache? Empirische Erhebungen und praktische Erfahrungen zum Einsatz mittel- und frühneuhochdeutscher Texte im Unterricht. In: Hofmeister, Wernfried/Schwinghammer, Ylva (Hrsg.): Literatur-Erlebnisse zwischen Mittelalter und Gegenwart. Aktuelle didaktische Konzepte und Reflexionen zur Vermittlung deutschsprachiger Texte. Wien: Peter Lang 2015, 147-173.
- Schwinghammer, Ylva et al.: Das Sparkling Science Projekt Arbeitskoffer zu den Steirischen Literaturpfaden des Mittelalters: Die Welt des Mittelalters als Ausgangspunkt für regionale und digitale Literaturerlebnisse. In: Hofmeister, Wernfried/Schwinghammer, Ylva (Hrsg.): Literatur-Erlebnisse zwischen Mittelalter und Gegenwart. Aktuelle didaktische Konzepte und Reflexionen zur Vermittlung deutschsprachiger Texte. Wien: Peter Lang 2015, 9-72.

Bildimpulse und Bildfolgen

Elizitation von Sprachproduktionsdaten anhand visueller Stimuli¹

1. Vorstellung von Bildimpulsen und Bildfolgen

In diesem Beitrag stelle ich vor, wie Sprachproduktionsaufgaben mit visuellen Stimuli (Bildimpulse und Bildfolgen²) kombiniert werden können, um mündliche oder schriftliche Sprachproduktionsdaten zu elizitieren³. Solche Daten ermöglichen die Analyse von produktiven sprachlichen Fähigkeiten⁴. Dabei konzentriere ich mich auf den Kontext der Spracherwerbs- bzw. Sprachentwicklungsforschung (Erstsprache (L1) und Zweitsprache (L2)), in dem dieses Erhebungsverfahren häufig verwendet wird und zur Beantwortung sehr unterschiedlicher Forschungsfragen zum Einsatz kommen kann (vgl. z.B. für L1: Berman/Slobin 1994; Boueke et al. 1995; Augst et al. 2007; Becker-Mrotzek et al. 2014; Uhl 2015; L2 z.B. Knapp 1997; Griefßhaber 2005; Reich/Roth 2007; Jeuk 2012; Wieler 2013; Mezger/Schroeder/Şimşek 2014).

¹ Ich bedanke mich bei Andreas Bittner, Miriam Langlotz und Verena Wecker für wertvolle Anregungen zum Manuskript des vorliegenden Beitrags.

² Als weiteres Datenerhebungsfahren, das sich bildlicher Stimuli bedient und häufig eingesetzt wird, sind außerdem Videoclips zu nennen (vgl. z.B. Dimroth 2005; 2012; Mehlem 2010; Blaschitz 2013; Redder/Guckelsberger/Graßer 2013), auf die in diesem Beitrag allerdings nicht weiter eingegangen wird.

³ Bei einer Elizitation (auch: Elizitierung, aus lat.: *elicere*: entlocken) fordern die Forschenden Probandinnen/Probanden gezielt zur Datenproduktion auf, vgl. ausführlicher Abschnitt 2.1).

⁴ Vgl. zu psycholinguistischen Erhebungs- und Analyseverfahren von Sprachverarbeitungsdaten den Überblicksartikel von Roberts (2012) am Beispiel der Zweitspracherwerbsforschung. Einen deutschsprachigen, einführenden Sammelband mit verschiedenen Anwendungsbeispielen aus der Zweitspracherwerbsforschung haben jüngst Schimke/Hopp (2018) herausgegeben.

1.1 Funktionsweise des Einsatzes von Bildimpulsen und Bildfolgen in Verbindung mit Sprachproduktionsaufgaben

Wie bei jeder empirischen Untersuchung steht am Anfang die Entwicklung einer Forschungsfrage, die wiederum zur Entscheidung für ein Erhebungsverfahren führt, durch dessen Anwendung der/die Forschende Antworten auf seine/ihre Forschungsfragen erwartet (vgl. Sunderland 2010). Bei dem hier zu besprechenden Verfahren muss dazu eine aufeinander abgestimmte Kombination von Bildstimulus und Sprachproduktionsaufgabe entwickelt werden, sodass die Probandin bzw. der Proband bei der Datenerhebung die im Forschungsinteresse stehenden sprachlichen Strukturen am wahrscheinlichsten produziert. Bei der Datenerhebung wird diese Kombination von Bildstimulus und Sprachproduktionsaufgabe den Probandinnen und Probanden präsentiert, die zur mündlichen oder schriftlichen Lösung der Aufgabe aufgefordert werden. Dabei müssen nonverbale statische visuelle Zeichen holistisch verarbeitet und bei der Versprachlichung in eine sukzessive lineare Struktur transformiert werden (vgl. Nöth 2000, 490). Sofern nicht interaktionale bzw. dialogische produktive Kompetenzen untersucht werden, intervenieren die Forschenden i.d.R. während der Sprachproduktion nicht weiter, d.h. die Sprachproduktion vollzieht sich dann monologisch und erfolgt gemäß der vom Probanden/der Probandin frei gewählten bzw. in Abhängigkeit von seiner/ihrer sprachlichen Entwicklung zur Verfügung stehenden sprachlichen Mittel.

Bei einem Bildimpuls besteht der visuelle Stimulus aus einem einzelnen Bild. Die dazu gestellte Sprachproduktionsaufgabe kann in Abhängigkeit vom Forschungsinteresse auf die Elizitation von kleineren sprachlichen Einheiten unterhalb der Satzebene oder auf die Elizitation einer satzübergreifenden zusammenhängenden Sprachhandlung abzielen. Beispielsweise kann ein Bildimpuls dazu eingesetzt werden, um den produktiven Wortschatz zu einem spezifischen lexikalischen Feld oder um bestimmten Flexionsformen zu elizitieren. Wird etwa ein Wimmelbild mit der Sprachproduktionsaufgabe kombiniert, die Tätigkeiten der darauf abgebildeten Personen zu benennen (z.B.: *Was tun die Menschen auf dem Bild?*), kann der produktive Wortschatz zum verbalen Lexikon elizitiert werden (vgl. z.B. Jeuk 2012). Ein Bildimpuls kann aber auch als Stimulus für satzübergreifende zusammenhängende Sprachhandlungen eingesetzt werden, etwa für Beschreibungen (*Beschreibe den auf dem Bild dargestellten Wasserkreislauf*, vgl. z.B. Oleschko 2014) oder Erzählungen (z.B. *Denk dir zu dem Bild eine Geschichte aus*, vgl. z.B. Augst et al. 2007). Der Bildimpuls muss so gestaltet bzw. gewählt sein, dass er dem Forschungsinteresse entsprechenden Anlass zur Sprachproduktion gibt.

Beim Einsatz einer Bildfolge muss die Probandin bzw. der Proband zwischen einer Abfolge von einzelnen, aber zusammenhängenden Bildern eine Beziehung herstellen und diese verbalisieren (vgl. z.B. Becker-Mrotzek et al. 2014 zu Instruktionen zu einer bildlich dargestellten Kochanleitung; vgl. z.B. Boueke et al. 1995 oder Becker 2001 zu (Nach-)Erzählungen einer Bildergeschichte). Entspre-

chend eliziert man mit dieser Kombination von visuellem Stimulus und Sprachproduktionsaufgabe immer satzübergreifende zusammenhängende Dateneinheiten. Dabei können die Bildfolge und die Sprachproduktionsaufgabe nichtsdestotrotz so konstruiert sein, dass die Probandin bzw. der Proband darüber hinaus dazu angeregt wird, auch spezifische kleinere sprachliche Einheiten zu produzieren. Sollen beispielsweise anhand einer bildlich dargestellten Kochanleitung Daten gewonnen werden, die die Analyse der satzübergreifenden Sprachhandlung ‚Instruktion‘ ermöglichen, kann zeitgleich außerdem der produktive Wortschatz in diesem Feld erfasst werden, wenn auf den Bildern viele verschiedene Kochutensilien und Zutaten abgebildet sind, die zur erfolgreichen Bewältigung der Sprachproduktionsaufgabe benannt werden müssen.

1.2 Ausgangsdaten

Die bisherigen Ausführungen lassen bereits erahnen, dass die Kombination von Sprachproduktionsaufgaben und Bildstimuli je nach Forschungsfrage sehr variabel gestaltet werden kann. Genauso variabel sind die dadurch zu gewinnenden Ausgangsdaten. Gemein ist solchen Verfahren, wie auch immer sie gestaltet sein mögen, dass Probandinnen und Probanden gezielt zur Sprachproduktion aufgefordert werden. Solche Erhebungsverfahren werden unter dem Terminus Elizitation subsummiert und in vielen Darstellungen den Datenerhebungsverfahren Beobachtung und Experiment gegenübergestellt (vgl. z.B. Chaudron 2005, 764; Mezger/Schroeder/Şimşek 2014, 74f.). Je nach Erhebungsverfahren müssen die dadurch generierten Ausgangsdaten unterschiedlich typisiert werden. Bei z.B. teilnehmender Beobachtung erhält man sogenannte natürlichsprachliche Daten (auch: spontansprachliche Daten), weil die Daten in authentischen Kommunikationssituationen ohne Lenkung bzw. ohne Kontrolle durch die Forschenden gewonnen werden. Unter experimentellen Daten versteht man dagegen solche Daten, die dekontextualisiert von einer authentischen Kommunikationssituation und unter durch den Forschenden exakt kontrollierten Sprachproduktionsbedingungen generiert werden (vgl. Schlobinski 1996, 32 oder Meindl 2011).

Bei einer Elizitation legen die Forschenden durch ihre Sprachproduktionsaufgabe die Parameter für die Sprachproduktion fest und die Probandin bzw. der Proband ist sich darüber bewusst, dass sie/er im Rahmen eines Forschungsprozesses sprachlich handelt. Dadurch kann die Natürlichkeit der Sprachproduktionsdaten maßgeblich verzerrt werden (vgl. Schu 2000 und zum sog. ‚Beobachter-Paradoxon‘ Labov 1971). Deshalb stellen durch Elizitation gewonnene Daten keine natürlichsprachlichen Daten dar. Solange die Datenelizitation aber auch nicht im Sprachlabor unter exakter Kontrolle aller Sprachproduktionsbedingungen erfolgt, liegen auch keine experimentellen Daten vor. Je nachdem, wie stark die Forschenden die Sprachproduktion durch ihre Aufgabenstellung lenken, können elizierte Daten also mehr oder weniger natürlich bzw. experimentell sein. Stellt man sich die Erhebungsverfahren Beobachtung und Experiment als zwei sich gegenüberliegende Endpunkte eines Kontinuums vor, sind Elizitationsverfahren auf diesem

Kontinuum deshalb in Abhängigkeit vom Grad der Steuerung durch die Forschenden entweder näher am Pol der kaum kontrollierten beobachtenden Verfahren oder näher am Pol der stark kontrollierten experimentellen Verfahren anzusiedeln. Dabei sind elizitierte Daten aufgrund der in jedem Fall gegebenen Steuerung durch die Forschenden von spontansprachlichen Daten abzugrenzen, schließen jedoch experimentelle Daten ein (vgl. Abb. 1).

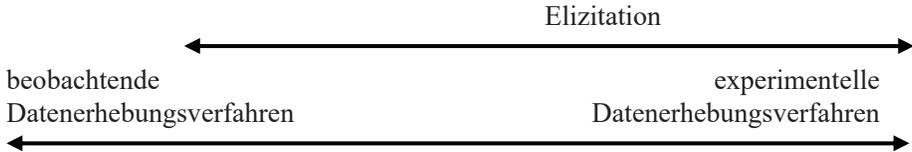


Abb. 1: Elizitation auf dem Kontinuum beobachtender und experimenteller Datenerhebungsverfahren

Inwieweit für das hier zu besprechende Elizitationsverfahren, bei dem Sprachproduktionsaufgaben mit Bildstimuli kombiniert werden, eine mehr oder weniger starke Lenkung vorliegt, entscheidet sich durch die Forschungsfrage und das daraus resultierende konkrete Design. Bei Forschungsfragen, bei denen es den Forschenden auf eine möglichst natürliche Sprachproduktion ankommt, wählen sie Bildstimuli und Sprachproduktionsaufgaben, die der Probandin bzw. dem Probanden großen Freiraum bei der Sprachproduktion geben. Dies ist dann sinnvoll und notwendig, wenn beispielsweise eine satzübergreifende zusammenhängende Sprachhandlung elizitiert werden soll. Kommt es den Forschenden darauf an, möglichst systematisch ein spezifisches Einzelphänomen zu untersuchen, von dem sie annehmen bzw. aus anderen Studien bereits wissen, dass es in der natürlichen, nicht kontrollierten Sprachproduktion selten vorkommt, gestalten sie ihr Forschungsdesign eher experimentell, indem sie durch die Kombination ihrer Sprachproduktionsaufgabe und ihres Bildstimulus die Produktion der gewünschten sprachlichen Strukturen forcieren.

Dabei kann der Kontrollgrad an zwei unterschiedlichen Stellen reguliert werden, nämlich a) durch den visuellen Stimulus und b) durch die konkrete Aufgabenstellung für die Sprachproduktion. Bereits die Entscheidung, entweder einen Bildimpuls oder eine Bildfolge einzusetzen, führt auf dem Kontinuum zwischen natürlichsprachlichen und experimentellen Daten zu einer unterschiedlichen Lokalisierung der zu gewinnenden Daten. Soll beispielsweise eine Erzählung elizitiert werden und die Forschenden setzen dazu eine Bildergeschichte ein, geben sie damit einen klar konturierten narrativen Rahmen vor, entlang dessen sich die Sprachproduktion vollziehen soll. Zwar intervenieren die Forschenden während der Sprachproduktion nicht weiter, dennoch lenken sie die Sprachproduktion nonverbal stetig durch die vorgegebene Bildfolge (vgl. Bredel 2001; Becker 2001). Entscheiden sie sich dagegen für einen einzelnen Bildimpuls als visuellen Stimulus, besteht ihre Lenkung lediglich durch die initiale Aufforderung, zu dem vorgelegten Bild eine Geschichte zu erzählen und lassen damit für die Sprachproduktion größeren Freiraum. So muss die Probandin bzw. der Proband bei einem Bildim-

puls die Etablierung eines narrativen Settings, den Planbruch und das Schlussgeschehen beispielsweise selbst entwickeln, ohne dass sie/er sich an einer Bildfolge orientieren kann bzw. muss (vgl. Knapp 1997; Uhl 2015). Bei diesem Beispiel ist der Kontrollgrad durch den Einsatz einer Bildfolge also höher als durch den Einsatz eines einzelnen Bildimpulses.

Bezogen auf den Kontrollgrad bei der Sprachproduktionsaufgabe kommt es auf die Spezifik der Aufgabenstellung an, mit der ebenfalls Verschiebungen der zu gewinnenden Daten auf dem Kontinuum zwischen natürlichsprachlichen und experimentellen Daten verbunden sind. Je spezifischer die Aufgabenstellung ausfällt, umso mehr nähern sich die gewonnenen Daten auf dem Kontinuum dem experimentellen Pol an. Gleichzeitig entscheidet sich mit der Spezifik der Aufgabenstellung, wie wahrscheinlich die Probandin bzw. der Proband das zu untersuchende sprachliche Phänomen bei der Bewältigung der kommunikativen Aufgabe verwenden wird. Am Beispiel von Aufgabenstellungen zur Untersuchung grammatischer Fähigkeiten unterscheidet Purpura (2005, 110) in Anlehnung an Loschky/Bley-Vroman (1993) drei Typen von Aufgabenstellungen, nämlich nach ihrer a) *task-naturalness*, b) *task-utility* und c) *task-essentialness*. Diese Typologie der Aufgabenstellung lässt sich auf das Kontinuum von natürlichsprachlichen und experimentellen Daten übertragen, wobei von a) bis c) der Natürlichkeitsgrad der zu gewinnenden Daten abnimmt. Bei einer Aufgabenstellung unter der Bedingung der *task-naturalness* kann eine bestimmte sprachliche Struktur zwar durch die Probandin bzw. den Probanden geäußert werden; die Verwendung der Struktur ist aber nicht zwingend notwendig. Unter der Bedingung der *task-utility* ist die Lösung der Aufgabe für die Probandin bzw. den Probanden am einfachsten, wenn er/sie die gewünschte sprachliche Struktur verwendet. Allerdings ist die Aufgabe auch ohne diese sprachliche Struktur lösbar. Lediglich unter der Bedingung der *task-essentialness* kann die Aufgabe nur durch die Verwendung der konkreten sprachlichen Struktur (zielsprachlich) bewältigt werden.

Die unterschiedlichen Konditionen seien an einem einfachen konstruierten Erhebungsdesign zur Untersuchung von produktiven grammatischen Fähigkeiten am Beispiel der Adjektivkomparation illustriert. Als Bildimpuls könnte Abbildung 2 eingesetzt werden. Durch die Abbildung von Figuren unterschiedlicher Körpergröße ist ein bildlicher Stimulus gegeben, der Anlass zur Verwendung der Adjektive *klein* bzw. *groß* in unterschiedlichen Komparationsstufen gibt.

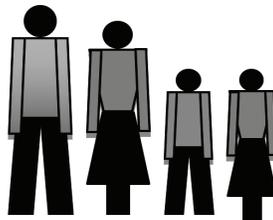


Abb. 2: Bildimpuls zur Generierung von Daten zur Adjektivkomparation

Für die unterschiedlichen Konditionen der Aufgabenstellung sei nachfolgend jeweils eine entsprechende Sprachproduktionsaufgabe angeführt:

- a. Beschreibe die auf dem Bild abgebildete Familie.
- b. Beschreibe die Familienmitglieder auf dem Bild im Hinblick auf ihre Körpergröße.
- c. Beschreibe zuerst die auf dem Bild abgebildeten Familienmitglieder im Hinblick auf ihre Körpergröße und beschreibe dann die Körpergrößen aller Familienmitglieder im Vergleich zueinander.

Durch die Aufgabenstellung a. besteht keine Notwendigkeit, komparierende Adjektive zu verwenden. Die Familie könnte beispielsweise dadurch beschrieben werden, dass die Anzahl der Personen oder die Bezeichnungen für die Familienmitglieder genannt werden. Durch Aufgabenstellungen wie b. und c. wird dagegen die Verwendung von komparierenden Adjektiven angeregt (in Aufgabenstellung b.) bzw. ist ihre Verwendung notwendig (in Aufgabenstellung c.). In gleichem Maße wie die Kontrolle der Sprachproduktion der Probandin bzw. des Probanden durch die zunehmende Spezifizierung der Aufgabenstellung von a. bis c. zunimmt, steigt die Wahrscheinlichkeit, dass die Probandin bzw. der Proband komparierende Adjektive verwendet und die Forschenden somit solche Ausgangsdaten erhalten, die ihrem Forschungsinteresse entsprechen.

Die bisher besprochenen Anwendungsbeispiele zeigen, dass durch den Einsatz von Bildstimuli in Kombination mit Sprachproduktionsaufgaben bezogen auf ihren natürlichen bzw. experimentellen Charakter sehr unterschiedliche Ausgangsdaten generiert werden können. Die Variabilität der Ausgangsdaten trifft auch auf die Analysemöglichkeiten zu, die sie eröffnen, da mit diesem Verfahren Ausgangsdaten generiert werden können, anhand derer alle produktiven sprachlichen Kompetenzen untersucht werden können. Legt man beispielsweise den Referenzrahmen zu sprachlichen Basisqualifikationen nach Ehlich et al. (2008) zugrunde, in dem sprachliche Fähigkeiten in phonische, pragmatische, lexikalisch-semantische, morphologisch-syntaktische, diskursive und literale Fähigkeiten unterschieden werden, sind für all diese Ebenen Erhebungsdesigns vorstellbar, die auf die Elizitation entsprechender Ausgangsdaten abzielen. Sollen phonische bzw. orthographische Fähigkeiten untersucht werden, sind die Forschenden natürlich an das Medium der Mündlichkeit bzw. Schriftlichkeit gebunden.

In der Tabelle 1 habe ich beispielhaft einige Studien bzw. Sprachstandsdiagnoseverfahren aus der Spracherwerbs- und Sprachentwicklungsforschung angeführt, in denen Sprachproduktionsaufgaben in Kombination mit Bildstimuli zum Einsatz kommen. Aus der dritten Spalte der Tabelle geht hervor, dass jeweils unterschiedliche der genannten produktiven sprachlichen Fähigkeiten im Fokus stehen. Im Feld der Sprachdiagnostik werden beispielsweise häufig mehrere Bereiche analysiert (vgl. z.B. Griebhaber 2005; 2009; 2012; Reich/Roth 2007; Reich/Roth/Döll 2009). In Studien, in denen die Sprachentwicklung hinsichtlich des Ausbaus literaler bzw. diskursiver Fähigkeiten untersucht wird, kommen Bil-

dimpulse bzw. Bildfolgen häufig zur Ermittlung von Diskurs- bzw. Textmusterwissen zum Einsatz (vgl. z.B. Berman/Slobin 1994; Boueke et al. 1995; Knapp 1997; Weinhold 2000; Becker 2001; Augst et al. 2007; Dannerer 2012; Becker-Mrotzek et al. 2014; Oleschko 2014; Uhl 2015).⁵ Den Erwerb pragmatischer bzw. diskursiv-interaktionaler Kompetenzen haben z.B. verschiedene Forschungsarbeiten zum Gegenstand, in denen Bilderbücher als Sprachproduktionsstimuli eingesetzt werden (vgl. z.B. Wieler 2013 oder den Forschungsbericht von Preußner 2015 aus didaktischer Perspektive).

Studie	eingesetzter Bildstimulus	untersuchte sprachliche Kompetenz	mündl./ schriftl.	L1/ L2
Berman/Slobin 1994	Bildfolge (Bildergesch.)	diskursive narrative Kompetenz	mündl.	L1
Boueke et al. 1995	Bildfolge (Bildergesch.)	diskursive narrative Kompetenz	mündl.	L1
Becker 2001	Bildfolge (Bildergesch.)	diskursive narrative Kompetenz	mündl.	L1
Reich/Roth 2007	Bildfolge (Bildergesch.)	phonische, lexikalisch-semantische, morphologisch-syntaktische, diskursive narrative, pragmatische Kompetenz	mündl.	L2
Grießhaber 2005	Bildimpuls	lexikalisch-semantische, morphologisch-syntaktische Kompetenz	mündl.	L2
Jeuk 2012	Bildimpuls (Wimmelbild)	lexikalisch-semantische Kompetenz	mündl.	L2
Dannerer 2012	Bildfolge (Bildergesch.)	diskursive narrative Kompetenz	mündl./ schriftl.	L1/ L2
Wieler 2013	Bildfolge (Bilderbuch)	narrative diskursive und literale sowie interaktionale Kompetenz	mündl./ schriftl.	L2
Weinhold 2000	Bildimpuls	literale Kompetenz von Schreibnovizen	schriftl.	L1
Augst et al. 2007	Bildimpuls	literale narrative Kompetenz	schriftl.	L1
Uhl 2015	Bildimpulse	morphologisch-syntaktische (Tempusgebrauch), literale narrative Kompetenz	schriftl.	L1
Becker-Mrotzek et al. 2014	Bildfolge (Instruktionsillustrationen)	literale instruktive Kompetenz (Kochanleitungen)	schriftl.	L1
Knapp 1997	Bildimpuls	literale narrative Kompetenz	schriftl.	L2

⁵ Diese Studien schließen aber auch die Untersuchung von sprachlichen Mitteln bzw. Strukturen unterhalb der Satzebene ein, die spezifisch für bestimmte Diskurs- bzw. Textsorten sind.

Ballis 2008	Bildimpuls	literale narrative Kompetenz (textmusterspezifische Wendungen in Märchen)	schriftl.	L2
Grißhaber 2009	Bildimpuls	lexikalisch-semantische, morphologisch-syntaktische, literale narrative Kompetenz	schriftl.	L2
Reich/Roth/ Döll 2009	Bildfolge (Instruktions- illustrationen)	lexikalisch-semantische, morphologisch-syntaktische, literale instruktive, pragmatische Kompetenz (Bauanleitung zu einem Bumerang)	schriftl.	L1/ L2
Oleschko 2014	Bildimpuls	deskriptive literale Kompetenz	schriftl.	L2

Tab. 1: Beispiele für untersuchte sprachliche produktive Kompetenzen beim Einsatz von Sprachproduktionsaufgaben in Kombination mit Bildstimuli

Bis auf Jeuk (2012) zielen alle der angeführten Studien darauf ab, Daten zu elizitieren, die die Analyse satzübergreifender zusammenhängender Sprachhandlungen erlauben ((Nach-)Erzählung, Instruktion, Beschreibung). Bezogen auf den Kontrollgrad durch das Forschungsdesign sind die darin vorzufindenden Sprachproduktionsaufgaben deshalb der Kondition der *task-naturalness* zuzuordnen. Sie geben Beispiele für gering lenkende Sprachproduktionsaufgaben, die zum Ziel haben, eher natürlichsprachliche Daten zu generieren (vgl. z.B. Augst et al. 2007, 46: „Denk dir zu diesem Bild eine interessante Geschichte aus und schreibe sie für das Geschichtenbuch auf“ oder Reich/Roth/Döll 2009, 232: „Sende uns [...] als Arbeitsprobe einen *Artikel*, in dem erklärt wird, wie der Bumerang *Triton IV* gebaut wird. Der Artikel muss ohne Abbildungen verständlich sein.“). Ein Beispiel für ein Erhebungsdesign, das den Konditionen der *task-utility* bzw. der *task-essentialness* entspricht, gebe ich im zweiten Teil dieses Beitrags, in dem ich im Rahmen einer Zweitspracherwerbsstudie von mir entwickelte Kombinationen von Sprachproduktionsaufgaben und Bildimpulsen im Detail vorstelle.

1.3 Vorteile und Nachteile des Einsatzes von Bildimpulsen und Bildfolgen in Verbindung mit Sprachproduktionsaufgaben

Vorteile

Variabilität des Erhebungsdesigns: Ein großer Vorteil dieses Verfahrens liegt darin, dass die Bildimpulse und Bildfolgen sowie die dazu gehörigen Sprachproduktionsaufgaben je nach Forschungsinteresse variabel gestaltet werden können und die mündlichen oder schriftlichen Ausgangsdaten dadurch die Analyse verschiedener produktiver sprachlicher Fähigkeiten ermöglichen.

Voraussetzungen der Probandinnen und Probanden: Zum Einsatz dieses Erhebungsverfahrens müssen bezogen auf die Probandinnen und Probanden wenige Voraussetzungen gegeben sein. Zu diesen zählen eine so weit fortgeschrittene kognitive Entwicklung, dass die Aufgabenstellung verstanden und gelöst werden

kann, sowie ein Minimum an Sprachkompetenz in der jeweiligen zu untersuchenden Teilkompetenz. Z.B. werden Grundschul Kinder der ersten Klasse mangels Schriftkompetenz kaum eine Erzählung selbst aufschreiben können oder Lernende des Deutschen als Fremdsprache im initialen Stadium ihres Fremdspracherwerbs noch nicht über ausreichende sprachliche Mittel verfügen, um eine Instruktion verbalisieren zu können.

Aufwand: Das Erhebungsverfahren ist organisatorisch, zeitlich, materiell bzw. technisch unaufwändig, sofern damit schriftsprachliche Daten erhoben werden. Eine Textelizitation lässt sich im schulischen Kontext beispielsweise im Klassenverband durchführen. Benötigt werden dafür lediglich die vervielfältigten Bildstimuli und Schreibproduktionsaufgaben sowie Schreibwerkzeug. Innerhalb eines kurzen Zeitraums können Texte von einer Vielzahl von Probandinnen und Probanden erhoben werden. Bei mündlichen Elizitationen steigt der Aufwand: Es wird ein Aufnahmegerät benötigt, für die Aufbereitung der Daten sind Transkriptionskenntnisse, ggf. eine Transkriptionssoftware, notwendig. Außerdem muss dafür mehr Zeit eingeplant werden, da jede Probandin bzw. jeder Proband bei der Sprachproduktion einzeln aufgezeichnet werden muss und die Daten durch die Transkription zeitaufwändig aufzubereiten sind.

Nachteile

Fehlende Evidenz: In Abhängigkeit von der Motivation der Probandinnen und Probanden oder der zur Verfügung stehenden Zeit, die Sprachproduktionsaufgabe zu lösen, kann sowohl die Datenqualität als auch die -quantität sehr unterschiedlich ausfallen. Für bestimmte, im Forschungsinteresse stehende sprachliche Strukturen sind möglicherweise keine Belege vorzufinden. Die Wahrscheinlichkeit, dass spezifische sprachliche Strukturen nicht enthalten sind, steigt außerdem mit abnehmendem Lenkungsgrad bei der Sprachproduktionsaufgabe, d.h. mit zunehmender Natürlichkeit der generierten Ausgangsdaten. Bei fehlender Evidenz darf aber nicht geschlossen werden, dass die Probandin bzw. der Proband diese Strukturen nicht beherrscht, weil es genauso gut sein kann, dass er/sie sie bei der Lösung der Aufgabe lediglich nicht verwendet hat.

Zur Beachtung beim Einsatz von Bildergeschichten zur Untersuchung narrativer Fähigkeiten: Fix (2006, 96) kritisiert zum Einsatz von Bildergeschichten beispielsweise, dass es sich bei deren Verschriftlichung um eine rein schulische Textsorte handele, durch die nur auf diese Sprachproduktionsaufgabe bezogene narrative Fähigkeiten ermittelt werden können. In Becker (2001), Bredel (2001) oder Knapp (2001) wird diesen Kritikpunkt betreffend aus methodischer Sicht ausführlich reflektiert, welchen Einfluss das Erhebungsdesign bzw. die Sprachproduktionsaufgabe auf die (kognitive) Erzählkonzeption der Probandin bzw. des Probanden nimmt. Hauptsächlich wird kritisiert, dass bei Bildergeschichten die einzelnen Bilder *beschrieben* werden und nicht eine Geschichte *erzählt* wird und die Probandinnen und Probanden somit eine andere Sprachhandlung als die zu untersuchende ausführen.

1.4 Literaturempfehlungen

Die in Tabelle 1 angeführten Studien seien der Lektüre empfohlen, um sich einen Überblick zu verschaffen, auf welche vielfältige Weise Sprachproduktionsaufgaben mit visuellen Stimuli kombiniert werden können. Hervorheben möchte ich zwei Titel

Berman/Slobin (1994): Relating Events In Narrative: A Crosslinguistic Developmental Study.

Zum Einsatz von Bildergeschichten sei besonders auf Berman und Slobin (1994) verwiesen. In diesem Band steht der Erwerb narrativer mündlicher Fähigkeiten aus kontrastiv-linguistischer Perspektive im Fokus. Darin enthalten sind Einzelstudien zum Erzählerwerb von Probandinnen und Probanden mit unterschiedlichen L1, wobei die Daten aller Einzelstudien anhand der gleichen Bildergeschichte elizitiert wurden. Dabei handelt es sich um die Bildergeschichte „Frog, where are you?“ (Mayer 1969), die in der Erzählforschung eine gewisse Berühmtheit erlangt hat.

Augst et al. (2007): Text-Sorten-Kompetenz.

Zur Elizitation von schriftlichen Sprachproduktionsdaten anhand eines Bildimpulses möchte ich die Studie von Augst et al. (2007) hervorheben, da es sich hier zum einen um eine echte Längsschnittstudie zum Ausbau narrativer Fähigkeiten im Grundschulalter handelt (L1 Deutsch). Zum anderen wird in dieser Studie der Zusammenhang zwischen der Entwicklung grammatischer und der Entwicklung textmusterspezifischer Fähigkeiten untersucht.

2. Elizitation von Sprachproduktionsdaten anhand von Bildimpulsen am Beispiel einer Zweitspracherwerbsstudie

In Binanzer (2017) habe ich den Erwerb des deutschen Genussystems durch insgesamt 195 Grundschul Kinder (2.-4. Klasse) untersucht, wovon 136 Kinder Deutsch als L2 und Türkisch oder Russisch als L1 sprachen. Zur Untersuchung des Genuserwerbs habe ich in zwei Teilstudien zwei unterschiedliche Datentypen elizitiert, wovon eine Studie auf zu Bildimpulsen schriftlich elizitierten Sprachproduktionsdaten basiert.

In Abschnitt 2.1 skizziere ich zunächst die Forschungsfrage meiner Untersuchung, in Abschnitt 2.2 die Vorannahmen, die ich zum Erwerb des deutschen Genussystems hatte, und in Abschnitt 2.3 die Erhebungsdesigns von Vorgängerstudien. Von diesen drei Aspekten ausgehend habe ich für meine eigene Untersuchung entschieden, mit zwei unterschiedlichen Elizitationsverfahren zu arbeiten, deren Daten ich trianguliert habe.⁶ In Abschnitt 2.4 stelle ich von den beiden von mir verwendeten Elizitationsverfahren das Textelizitationsverfahren vor, wofür ich vier verschiedene Bildimpulse mit unterschiedlichen Sprachproduktionsaufgaben (Erzählung, Beschreibung) miteinander kombiniert habe.⁷ Abschließend reflektiere ich die durch die Textelizitation erzielten Ergebnisse meiner Studie (2.5).

2.1 Forschungsfrage

Der Erwerb der Kategorie Genus durch Lernende des Deutschen als Zweit- und Fremdsprache wurde bisher vornehmlich unter der Perspektive untersucht, wie Genus als Klassifikationskategorie erworben wird, d.h. wie sich L2-Lernende die Genusklassenzugehörigkeit eines deutschen Nomens zu einer der drei Genusklassen Maskulinum, Femininum und Neutrum erschließen (z.B. *der, die* oder *das Wiese?*). Da die drei deutschen Genera Maskulinum, Femininum und Neutrum durch multiple, miteinander konkurrierende und dabei mehr oder minder valide formale (phonologische und morphologische) und semantische nominale Merkmale indiziert werden (vgl. Köpcke 1982; Köpcke/Zubin 2009), gilt der Erwerb der deutschen Genusklassifikation als komplexe Erwerbsaufgabe (vgl. z.B. Wegener 1995; Menzel 2004; Kostyuk 2005; Marouani 2006; Jeuk 2008; Spinner/Juffs 2008; Dieser 2009; Montanari 2010; Ruberg 2013). Im Zentrum der hier angeführten Genuserwerbsstudien stand entsprechend die Frage, in welcher Reihenfolge die o.g. formalen und semantischen Merkmale als Indikatoren für die Genusklassenzugehörigkeit erworben werden. Zur empirischen Untersuchung

⁶ Bei einer Datentriangulation werden zur Untersuchung eines Phänomens unterschiedliche Datenerhebungsmethoden eingesetzt, wodurch unterschiedliche Datentypen generiert und die Ergebnisse aufeinander bezogen werden können. Vorhandene Schwächen der unterschiedlichen Datentypen können dadurch ggf. kompensiert werden (vgl. ausführlicher zu Triangulation z.B. Flick (2011) oder im Überblick Aguado (2014)).

⁷ Neben den Textelizitationsverfahren kam außerdem ein Multiple Choice-Verfahren experimentellen Charakters zum Einsatz.

dieser Frage wurden in diesen Studien vornehmlich die Verwendungen von Definitartikeln analysiert, weil die Wahl des Definitartikels Aufschluss über die Genuszuweisungsstrategien von Lernenden gibt.

In meiner Untersuchung verlagerte ich entgegen dieser gängigen Herangehensweise die Perspektive auf den Genuserwerb, indem ich postulierte, dass Genuserwerb zunächst Kongruenzerwerb ist. Diese Verlagerung des Blickwinkels war durch die auf funktionalistischen Spracherwerbsmodellen basierende Prämisse motiviert, dass Lernende im Erwerbsprozess sprachliche Formen und Strukturen hinsichtlich ihrer kommunikativen Funktionen interpretieren. Daraus resultierte für meine Studie, die Funktion des Lerngegenstandes, also der Kategorie Genus, als begründenden Ausgangspunkt zu wählen. Diese Funktion manifestiert sich darin, als Kongruenzkategorie zwischen sprachlichen Einheiten eindeutige Referenzbezüge herzustellen („reference tracking“, Corbett 1991, 322). Die Grundannahme meiner Studie war deshalb, dass Genuserwerb v.a. Kongruenzerwerb ist, der den Klassifikationserwerb impliziert bzw. nach sich zieht.

Ich untersuchte deshalb nicht nur die Verwendung von Definitartikeln, sondern auch die Verwendung anderer sprachlicher Einheiten, die im Deutschen nach Genus flektiert werden. Diese sprachlichen Einheiten bezeichne ich nachfolgend in der Terminologie von Corbett (1979; 1991; 2006) als genussensitive ‚Targets‘ („Ziele“). Sie werden so bezeichnet, weil das Genusmerkmal eines Nomens auf sie übertragen werden muss⁸ und an ihnen die auf Genus basierende Kongruenz zwischen zwei sprachlichen Einheiten sichtbar wird. Zu den von mir untersuchten genussensitiven Targets zählen neben Definitartikeln auch attributive Adjektive, Relativ- und Personalpronomen (*der* – schön-*er* – *der* – *er*; *die* – schön-*e* – *die* – *sie*; *das* – schön-*es* – *das* – *es*).

2.2 Vorannahmen

Für den Erwerb von Genuskongruenz formulierte ich zwei zentrale Hypothesen:

- a) Semantische Kongruenz wird vor formal-grammatischer Genuskongruenz erworben.
- b) Targets mit enger morphosyntaktischer Bindung zum kongruenzauslösenden Nomen werden vor Targets mit geringerer morphosyntaktischer Bindung formal-grammatisch genuskongruent markiert (Artikel > attributives Adjektiv > Relativpronomen > Possessiv-, Personalpronomen).

Diese Vorannahmen seien nachfolgend erläutert. Verschiedene Vorgängerstudien zum Genuserwerb durch kindliche Lernende des Deutschen als Zweitsprache haben gezeigt, dass ab dem Grundschulalter semantische Merkmale als Indikatoren für die Genusklassenzugehörigkeit vor formalen Merkmalen erworben wer-

⁸ Eine Genusmarkierung kann auch durch „ein Bezugsnomen im Lexikon, ein mit einem lexikalischen Feld verbundenes Merkmal oder auch ein vom Lexikon unabhängiges pragmatisches Moment“ (Köpcke/Zubin 2009, 140) ausgelöst werden. Vgl. ausführlicher auch Abschnitt 2.2 oder Corbett (1991; 2006).

den (vgl. Wegener 1995; Bast 2003) und dass Personalpronomen die ersten Targets sind, die auf systematische Formverwendungen schließen lassen, die auf semantischen Merkmalen von Referenznomen basieren. So zeigen diese Studien, dass kindliche L2-Lernende die Personalpronomen *er* und *sie* dazu verwenden, um auf männliche bzw. weibliche Personenbezeichnungen zu referieren, während sie mit dem Personalpronomen *es* auf unbelebte Objekte referieren. In dieser Phase berücksichtigen die Lernenden das formal-grammatische Genus der Nomen noch nicht (z.B. *Mädchen_N* – *sie_F*; *Eisenbahn_F* – *es_N*), sondern richten sich in ihrer Formenwahl nach ihren semantischen Merkmalen (*Mädchen*_[+weiblich] – *sie*_[+weiblich]; *Eisenbahn*_[-belebt] – *es*_[-belebt]).

Von dieser Formverwendung im pronominalen Bereich ausgehend sollten die Lernenden, so meine Annahme, auch die anderen genussensitiven Targets semantisieren, die Formübereinstimmungen aufweisen, um ein konsistentes, semantisches Kongruenzsystem zu etablieren (für [+männlich]: *er* – *der* – schön-*er*; für [+weiblich]: *sie* – *die* – schön-*e*; für [-belebt]: *es* – *das* – schön-*es*). Mit zunehmender Zweitsprachkompetenz bzw. fortschreitendem Zweitspracherwerb sollte dieses semantisch basierte System in einer nächsten Erwerbsphase desemantisiert bzw. grammatikalisiert werden. In Anlehnung an die Agreement Hierarchy (‘Kongruenzhierarchie’) von Corbett (vgl. 1979; 1991; 2006) habe ich dabei angenommen, dass sich formal-grammatische Genuskongruenz vom Artikel ausgehend über die anderen Targets entfaltet: Zunächst werden morphosyntaktisch eng an das Bezugsnomen gebundene Targets (Artikel, attributives Adjektiv), danach weniger eng gebundene Targets (Relativ-, Personalpronomen) formal-grammatisch genuskongruent markiert. Demnach sollte sich beispielsweise zeigen, dass die Kinder in der ersten Grammatikalisierungsphase im Kontext eines Nomens wie *Mädchen* nominalgruppenintern damit beginnen, Artikel und Adjektive formal-grammatisch zu verwenden (*das Mädchen*, *ein schönes Mädchen*), zeitgleich nominalgruppenextern aber noch semantisch motivierte Formen wählen (*das Mädchen* – *sie*). Erst darauffolgend – wiederum mit zunehmender Zweitsprachkompetenz bzw. fortschreitendem Zweitspracherwerb – sollte das gesamte Kongruenzsystem desemantisiert und an allen Targets unabhängig von ihrer morphosyntaktischen Bindung an das Nomen formal-grammatische Genuskongruenz markiert werden (*das Mädchen* – *es*).

2.3 Erhebungsverfahren und Ausgangsdaten von Vorgängerstudien

Zur Untersuchung des Genuserwerbs wurden sowohl natürlichsprachliche als auch experimentelle Datenkorpora erhoben und ausgewertet. Natürlichsprachliche Korpora wurden z.B. in informellen Spiel-, Vorlese- oder Gesprächssituationen aufgezeichnet (vgl. Spinner/Juffs 2008; Montanari 2010; Ruberg 2013), die dann im Hinblick auf die darin vorkommenden genussensitiven Targets ausgewertet wurden. Systematisch erhobene Datenkorpora zum Genuserwerb fußen häufig auf experimentellen Genuszuweisungstests, bei denen die Probandinnen und Probanden gebeten werden, Kunstwörtern einen Definitartikel (*der*, *die* oder

das Troch?) zuzuweisen,⁹ der Aufschluss über Genuszuweisungsstrategien gibt (vgl. z.B. Mills 1986; Wegener 1995; Marouani 2006; Dieser 2009; Ruberg 2013). Mit diesem Verfahren konnte gezielt überprüft werden, welche Rolle den formalen und semantischen Merkmalen von Nomen bei der Genuszuweisung zukommt.

Mit den jeweiligen Datentypen sind Vor- und Nachteile verbunden. Als nachteilig ist für natürlichsprachliche Korpora die fehlende Evidenz zu bestimmten genus-sensitiven Targets (attributive Adjektive oder Relativpronomen) bzw. zu Nomen mit bestimmten semantischen oder formalen Merkmalen zu bewerten. Bei echten Nomen, insbesondere bei hochfrequenten, konnte außerdem auch nicht ausgeschlossen werden, dass die Artikelwörter lediglich holistisch mit den Nomen gespeichert wurden und somit keinen Aufschluss über etwaige Genuszuweisungsstrategien geben. Bei experimentellen Kunstwörterdaten wurde auf Grund der künstlichen Sprachproduktionsbedingungen dagegen die Generalisierbarkeit der Ergebnisse immer wieder in Frage gestellt. Zudem wurden dabei meistens nur Definitartikel und keine anderen genus-sensitiven Targets elizitiert.

Zum Ausgleich dieser mit den Datentypen einhergehenden Vor- und Nachteile habe ich mich zur Beantwortung meiner Forschungsfragen deshalb für eine Methoden- und Datentriangulation entschieden. Durch den Einsatz eines experimentellen Verfahrens konnte ich gezielt alle von mir angenommenen Einflussfaktoren überprüfen. In diesem Testverfahren waren die Kinder dazu aufgefordert, zwischen vorgegebenen Targets (Indefinit- und Definitartikel, attributive Adjektive, Relativ- und Personalpronomen) im Kontext verschiedener Nomen mit unterschiedlichen semantischen und formalen Merkmalen die zum Nomen kongruenten Formen (Maskulinum, Femininum, Neutrum) zu wählen. Zum anderen erhielt ich durch ein gemäßigt kontrolliertes Verfahren, die Textelizitation mittels Bildimpulsen, auch weitgehend natürlichsprachliche Sprachproduktionsdaten, die ich mit den Daten des experimentellen Designs vergleichen konnte. Nachfolgend stelle ich das letztgenannte Verfahren vor.

⁹ Die in solchen Testverfahren eingesetzten Kunstwörter entsprechen phonotaktisch der Struktur realer deutscher Nomen und werden mit den nominalen Merkmalen, die phonologische oder morphologische Hinweise auf die Genusklassenzugehörigkeit des Nomens geben, versehen. Beispiele aus Wegener (1995, 17): *Troch*: formales Merkmal: monosyllabisch; erwartete Genuszuweisung: Maskulinum; *Muhre*: formales Merkmal: Auslaut auf Schwa: erwartete Genuszuweisung: Femininum; vgl. auch schon Köpcke/Zubin (1983). Bei der Konstruktion der Kunstwörter muss darauf geachtet werden, dass die Artikelformen nicht in Analogie zu realen Nomen gewählt werden können.

2.4 Die Bildimpuls-Studie

Das weniger stark lenkende Datenerhebungsverfahren zur Elizitation von Sprachproduktionsdaten stellte die freie Textproduktion zu Bildimpulsen dar. Mit diesem Erhebungsverfahren habe ich Daten generiert, die Aufschluss darüber gaben, wie welche genussensitiven Targets in der freien, schriftlichen satzübergreifenden Sprachproduktion verwendet werden (im Gegensatz zur Verwendung der Targets im experimentellen Design).

Bei der Konzeption des Erhebungsdesigns bestand die erste Aufgabe darin, geeignete Nomen auszuwählen, in deren Kontext die zu untersuchenden genussensitiven Targets verwendet werden sollten. Zum einen mussten die Nomen nach den drei Genera Maskulinum, Femininum und Neutrum variiert werden, um über den Erwerb aller drei Genusparadigmen Aussagen treffen zu können. Zum anderen mussten die Nomen verschiedene semantische und formale Merkmale aufweisen, um herausfinden zu können, welche Rolle diesen Merkmale beim Kongruenzerwerb zukommt. Nur so konnte überprüft werden, ob Kongruenz zuerst bei solchen Nomen hergestellt wird, bei denen genussensitive Targets auf semantischer Basis miteinander verknüpft werden können, und ob formal-grammatische Genuskongruenz erst darauf folgend erworben wird. Diesen Kriterien entsprechend wurden aus den Bereichen Personen-, Tier-, Spielzeug- und Möbelbezeichnungen für die Bildimpuls-Studie insgesamt elf Nomen ausgewählt. Diese sind unter Angabe ihrer semantischen und formalen Merkmale der Tabelle 2 zu entnehmen.

semantische Merkmale	Testitem	formale Merkmale
	<i>Vater</i> _M	<i>X-er</i>
[+belebt]/	<i>Hund</i> _M	<i>X</i> _{monosyllabisch}
[+männlich]/	<i>Mutter</i> _F	<i>X-er</i>
[+weiblich]	<i>Katze</i> _F	<i>X-e</i>
	<i>Baby</i> _N	–
	<i>Drachen</i> _M	<i>X-en</i>
	<i>Schrank</i> _M	<i>X</i> _{monosyllabisch}
[–belebt]	<i>Tafel</i> _F	<i>X-el</i>
	<i>Eisenbahn</i> _F	<i>X</i> _{monosyllabisch}
	<i>Regal</i> _N	<i>X-al</i>
	<i>Schiff</i> _N	<i>X</i> _{monosyllabisch}

Tab. 2: Testitems in den Sprachproduktionsaufgaben

Die zweite konzeptionelle Anforderung bestand darin, die Sprachproduktionsaufgabe zu formulieren und dadurch die Textsorte bzw. die Sprachhandlung der Probandinnen und Probanden festzulegen. Dabei entschied ich mich für zwei unterschiedliche Textsorten, da ich einerseits narrative Texte (Erzählungen), andererseits deskriptive Texte (Beschreibungen) erhoben habe. Warum ich zwei Textsorten eliziert habe, erläutere ich nachfolgend bei der detaillierten Beschreibung

der einzelnen Bildimpulse und Sprachproduktionsaufgaben. Unabhängig von der Textsorte beinhalteten die Schreibaufgaben (z.T. auch die Bildüberschrift) die o.g. Nomen. Damit stellte ich sicher, dass die Kinder in ihren Texten tatsächlich diese Nomen und nicht etwa Synonyme verwendeten.

a) Narrative Texte

Auf dem ersten Bildimpuls mit dem Titel „Hund und Katze“ waren die beiden Tiere freundschaftlich aneinandergeschmiegt abgebildet. Die Kinder waren dazu aufgefordert, folgende Schreibaufgabe zu lösen: „Schau dir das Bild an. Was ist passiert? Warum verstehen sich Hund und Katze plötzlich so gut? Schreibe eine Geschichte!“

Der zweite Bildimpuls mit dem Titel „Eine schlaflose Nacht“ zeigte ein Elternpaar mit einem schreienden Baby. Der Schreibauftrag lautete, eine Geschichte zu erzählen, aus der hervorgeht, warum das Baby nicht schlafen kann und was Vater und Mutter dagegen unternehmen.

Beide Bildimpulse wiesen also kombiniert mit der Sprachproduktionsaufgabe Erzählwürdigkeit auf, weil durch sie eine Exposition, eine Komplikation und eine Auflösung eines narrativen Geschehens entwickelt werden konnte (vgl. Quasthoff 1980). Dadurch sollte gewährleistet werden, dass die Kinder satzübergreifende Sprachproduktionsdaten produzierten, in denen potentiell verschiedene genussensitive Targets verwendet werden konnten.

b) Deskriptive Texte

Die Texte, die zu den Nomen mit dem semantischen Merkmal [–belebt] (*Schrank, Tafel und Regal; Drachen, Eisenbahn und Schiff*) verfasst werden sollten, mussten deshalb deskriptiv sein, weil es zu vermeiden galt, dass die Kinder die Möbelstücke und Spielzeuge als belebte Handlungsträger interpretierten. Wären die Kinder dazu aufgefordert worden, zu diesen Gegenständen eine *Geschichte zu erzählen*, hätten sie die Gegenstände möglicherweise animiert und ihnen somit die gleichen semantischen Merkmale angedichtet, die die Nomen aufweisen, mit denen sie verglichen werden sollten ([+belebt], [+männlich], [+weiblich]).

Bei beiden Bildimpulsen, auf denen die Möbelstücke und Spielzeuge abgebildet waren (ein Klassenzimmer und ein Kinderzimmer), sollten diese deshalb hinsichtlich ihrer Eigenschaften bzw. Funktionen (Farbe, Größe, Position im Bild, Verwendung) beschrieben werden. Zur Beschreibung der Möbelstücke erhielten die Kinder folgenden Schreibauftrag: „Schau dir das Bild an. Beschreibe die drei rot eingekreisten Gegenstände (*Tafel, Regal, Schrank*) ganz genau. Wo sind sie im Klassenzimmer? Wie sehen sie aus? Welche Farben haben sie? Wozu brauchen wir diese Gegenstände im Klassenzimmer?“. Diese Reihe an Beschreibungsaufgaben sollte sicherstellen, dass auch bei diesen Bildimpulsen satzübergreifende sprachliche Einheiten elizitiert werden konnten. Zu den Spielzeugen war der Schreibauftrag identisch.

Unabhängig von der Textsorte kann für beide Textproduktionsaufgaben festgehalten werden, dass die Probandinnen und Probanden durch die Elizitation von

Texten, in denen vorgegebene Nomen verwendet werden mussten, dazu gezwungen waren, im Kontext dieser Nomen genussensitive Targets zu gebrauchen. Allerdings ergibt sich in der freien Textproduktion nicht für alle Targets die gleiche Verwendungsnotwendigkeit. Die einzigen Targets, für die die Kondition der *task-essentialness* gegeben war, stellten die Artikelwörter dar, da die Nomen funktional-pragmatisch als Textreferenten determiniert werden mussten. Dadurch, dass es sich um die Elizitation satzübergreifender Sprachproduktionsdaten handelte, war immerhin für Personalpronomen die Kondition der *task-utility* erfüllt, da durch Pronominalisierung satzübergreifende Referenzbezüge hergestellt werden können. Die Verwendung von Adjektiven, Relativ- oder Possessivpronomen war dagegen fakultativ bzw. war die Notwendigkeit bzw. Nützlichkeit, diese Targets zu verwenden, abhängig von den im Text zu äußernden Propositionen.

Sehr wohl abhängig von der Textsorte lässt sich aber konstatieren, dass die Gegenstandsbeschreibungen durch die spezifischere Sprachproduktionsaufgabe stärker gelenkt waren als die narrativen Geschichten. Bei dieser Textproduktionsaufgabe habe ich im Unterschied zu den narrativen Schreibaufträgen zudem gezielt Farbadjektive elizitiert (*task-essentialness*). Außerdem erhielten die Kinder für die Beschreibungen der unbelebten Gegenstände ein in drei Abschnitte gegliedertes Arbeitsblatt. In jedem Abschnitt sollte jeder Gegenstand einzeln beschrieben werden. Die Abschnitte wurden durch die jeweils zu beschreibenden Gegenstände (*Schrank, Tafel ...*) betitelt. Dadurch sollte forciert werden, dass die Kinder das zu beschreibende Nomen ggf. auch pronominalisieren (*task-utility*).¹⁰ Verortet man diese beiden unterschiedlichen Textproduktionsaufgaben auf dem in Abschnitt 1.2 vorgestellten Kontinuum von natürlichsprachlichen und experimentellen Daten, sind deshalb die narrativen Texte auf Grund ihrer freieren Aufgabenstellungen bezogen auf den zu schreibenden Text insgesamt näher am natürlichsprachlichen Pol zu verorten als die deskriptiven Texte.

2.5 Ergebnisse

Ohne an dieser Stelle die Ergebnisse meiner Untersuchung im Detail diskutieren zu können, kann zusammenfassend festgehalten werden, dass ich meine Annahmen zum Erwerb der deutschen Genuskongruenz in der Zweitsprache Deutsch durch die unterschiedlichen von mir erhobenen Datentypen verifizieren konnte. Die Daten beider Datenerhebungsverfahren offenbaren die gleichen Formverwendungsstrategien im Aufbau der morphologischen Genusparadigmen, wobei semantische Kongruenz vor formal-grammatischer Genuskongruenz erworben wird

¹⁰ In der Pilotierungsphase, in der zunächst keine vorstrukturierten Abschnitte verwendet worden waren, hatte sich gezeigt, dass die Kinder dazu neigten, ihre Texte nach den zu beschreibenden Eigenschaften zu gliedern. Dies hatte zur Folge, dass in den Textabschnitten zunächst beispielsweise für alle drei Nomen eine Eigenschaft genannt wurde (z.B. *Die Tafel hängt an der Wand. Das Regal steht unter dem Fenster. Der Schrank steht in der Ecke.*). Bei der Wiederaufnahme der zu beschreibenden Gegenstände wurden diese deshalb nicht pronominalisiert, sondern erneut benannt.

und sich formal-grammatische Genuskongruenz tatsachlich von der Nominalgruppe ausgehend entfaltet. Beispielhaft sei dies an drei Datenauszügen aus den Textkorpora illustriert:

- (1) **Das baby** hatte etwas gegesen es brach auch nichtz. Dann sagte die Mutter zu dem Vater ich glaube ich soll ihm etwas verbessern dannach ist **er** eingeschlafen. und alles war vorbei. (LOV, L1 Russisch)
- (2) **Der Drache** befindet sich an der Wand. **Er** sieht zweimal ein dreieck. **Es** hat grün, rot, gelb, weiÙe Farben. (CAR, L1 Russisch)
- (3) **Die Eisenbahn** im Zimer hat ganz viele Farben **es** hat rote, grüne, gelbe, oronge Farben. (LAS, L1 Russisch)

In allen drei Belegen sind die pranominalen Definitartikel formal-grammatisch genuskongruent zum nachfolgenden Nomen. Die Personalpronomen, die sich, anders als die Definitartikel, nicht in der gleichen syntaktischen Domane wie das kongruenzauflsende Nomen befinden, sind dagegen formal-grammatisch inkongruent. An den Personalpronomen zeigten sich aber systematische Interimssymbolisierungen der Lernenden, die sich an den semantischen Merkmalen der Nomen orientierten. Diese semantisch basierten Form-Funktionszuordnungen bestanden darin, Nomen mit dem Merkmal [+belebt] systematisch durch das maskuline Personalpronomen *er* zu ersetzen, sofern diese Nomen kein inharentes Sexusmerkmal aufwiesen (neben dem Nomen *Baby* wurde auch das Nomen *Kind* eingesetzt). Dadurch markierten die Kinder das semantische Merkmal [+belebt]. Personalpronomen im Kontext der Gegenstandsbezeichnungen wurden dagegen systematisch durch das neutrale Personalpronomen *es* wieder aufgenommen, um dadurch das Merkmal [-belebt] zu symbolisieren. Mit steigendem Sprachstand in der Zweitsprache Deutsch¹¹ war eine Reanalyse dieser Form-Funktionszuordnung beobachtbar, d.h. dass die Kinder mit bereits weiter fortgeschrittener Zweitsprachkompetenz die semantische Strategie zugunsten der dann entwickelten formal-grammatischen Strategie aufgaben und die Nomen tatsachlich ihrem formal-grammatischen Genus entsprechend pronominalisierten.

2.6 Methodisches Fazit

Bezogen auf das Datenerhebungsverfahren der Textelizitation anhand der Bildimpulse sei noch ein abschließendes Resümee gezogen, indem seine Vor- und Nachteile sowie konkrete Konstruktionsfehler bei meiner Untersuchung reflektiert werden.

Wie zu erwarten war, hat sich wie in anderen Vorgangerstudien gezeigt, dass die Texte, die die Kinder zu den Bildimpulsen verfassten, nicht fur alle zu untersuchenden genussensitiven Targets Evidenz boten. Nur zu Definitartikeln und Personalpronomen konnten anhand der freien Textproduktionsdaten verlassliche Aussagen getroffen werden, wahrend fur attributive Adjektive und Relativpronomen nur einzelne Belege zu finden waren.

¹¹ Der allgemeine Sprachstand in der Zweitsprache Deutsch wurde durch den C-Test ermittelt (vgl. zu diesem Sprachstandsdiagnoseverfahren z.B. Baur/Spettmann 2010).

Der Versuch, innerhalb der freien Textproduktion dennoch gezielt spezifische Targets zu elizitieren, führte bei den deskriptiven Texten zu einer Fehlkonstruktion der Sprachproduktionsaufgabe. Durch die darin enthaltene Frage „Welche Farben haben die Gegenstände?“ war für Farbadjektive zwar die Kondition der task-essentialness erfüllt, allerdings offenbaren die Belege (2) und (3) beispielsweise, dass die Kinder das Nomen *Farben* und nicht etwa die Spielzeugbezeichnungen *Drachen* bzw. *Eisenbahn* attribuierten (*es hat rote, grüne, gelbe, oronge Farben*). Noch häufiger verwendeten die Kinder die Farbadjektive prädikativ (z.B. *Der Drachen ist grün und rot*), wodurch sie bezogen auf Genusmarkierungen nicht auswertbar waren, da prädikativ verwendete Adjektive im Deutschen nicht nach Genus flektiert werden.

Als vorteilhaft erwies sich dieses Verfahren hingegen bezogen auf meine Untersuchungsgruppe, da ich damit auch Kinder jüngeren Alters in meine Untersuchung einschließen konnte. In der Pilotierungsphase zu den Datenerhebungsdesigns hatte sich gezeigt, dass nur Kinder der dritten und vierten Grundschulklassen über ausreichende Lesefähigkeiten verfügten, um die Multiple Choice-Tests bearbeiten zu können. Durch die Textelizitation konnte ich aber auch Sprachproduktionsdaten von Kindern des zweiten Jahrgangs erheben. So konnte ich die Daten von unterschiedlichen Untersuchungsgruppen (Alter bzw. Sprachstand in der Zweitsprache Deutsch) im Querschnitt miteinander vergleichen und dadurch Entwicklungstendenzen herausarbeiten, die sich mit dem ansteigenden Alter und der damit i.d.R. einhergehenden fortschreitenden zweitsprachlichen Entwicklung vollziehen.

Da die Ergebnisse zu den Definitartikeln und Personalpronomen außerdem die Befunde des Multiple Choice-Tests stützten, haben diese Daten auch den (methodischen) Zweck erfüllt, den ich durch sie erzielen wollte. Durch die Datentriangulation konnte ich ausschließen, dass die in den Multiple Choice-Tests für alle Targets beobachteten Formverwendungsstrategien nur ein Effekt des experimentellen Datenerhebungsdesigns waren. Die natürlichsprachlichen Daten verifizierten somit die Ergebnisse des Multiple Choice-Tests und unterstrichen sowohl deren Wert als auch den Wert des Textelizitationsverfahrens anhand der Bildimpulse.

Literatur

- Aguado, Karin: Triangulation. In: Settineri, Julia/Demirkaya, Sevilen/Feldmeier, Alexis/Gültekin-Karakoç, Nazan/Riemer, Claudia (Hrsg.): Empirische Forschungsmethoden für Deutsch als Fremd- und Zweitsprache. Eine Einführung. Paderborn: Schöningh 2014, 47-56.
- Augst, Gerhard/Disselhoff, Katrin/Henrich, Alexandra/Pohl, Thorsten/Völzing, Paul-Ludwig: Text-Sorten-Kompetenz. Eine echte Longitudinalstudie zur Entwicklung von Textsortenkompetenz. Frankfurt a.M.: Peter Lang 2007.
- Ballis, Anja: Formelhafter Sprachgebrauch in Schülertexten der Sekundarstufe I. Eine quantitative und qualitative Analyse von Märchen. In: Deutsch als Zweitsprache (2008), 1, 16-23.
- Bast, Cornelia: Der Altersfaktor im Zweitspracherwerb. Die Entwicklung der grammatischen Kategorien Numerus, Genus und Kasus in der Nominalphrase im ungesteuerten Zweitspracherwerb des Deutschen bei russischen Lernerinnen. Dissertation, Universität zu Köln 2003.
- Baur, Ruprecht/Spettmann, Melanie: Sprachstandmessung und Sprachförderung mit dem C-Test. In: Ahrenholz, Bernt/Oomen-Welke, Ingelore (Hrsg.): Deutsch als Zweitsprache. Baltmannsweiler: Schneider 2010, 430-441.
- Becker, Tabea: Kinder lernen erzählen. Zur Entwicklung der narrativen Fähigkeiten von Kindern unter Berücksichtigung der Erzählform. Hohengehren: Schneider 2001.
- Becker-Mrotzek, Michael/Grabowski, Joachim/Jost, Jörg/Knopf, Matthias/Linnemann, Markus: Adressatenorientierung und Kohärenzherstellung im Text. Zum Zusammenhang kognitiver und sprachlich realisierter Teilkomponenten von Schreibkompetenz. In: Didaktik Deutsch (2014), 37, 20-43.
- Berman, Ruth A./Slobin, Dan I.: Relating Events In Narrative: A Crosslinguistic Developmental Study. Hillsdale, NJ: Lawrence Erlbaum Associates 1994.
- Bınanzer, Anja: Genus – Kongruenz und Klassifikation. Evidenzen aus dem Zweitspracherwerb des Deutschen. Berlin, Boston: De Gruyter 2017.
- Blaschitz, Verena: Nacherzählen in der Zweitsprache – Narrativ-kreatives Potenzial des Nacherzählens eines Stummfilms. In: Becker, Tabea/Wieler, Petra (Hrsg.): Erzählforschung und Erzähldidaktik heute. Entwicklungslinien, Konzepte, Perspektiven. Tübingen: Stauffenburg 2013, 121-146.
- Boueke, Dietrich/Schüleın, Frieder/Büscher, Hartmut/Terhorst, Evamaria/Wolf, Dagmar: Wie Kinder erzählen. Untersuchungen zur Erzähltheorie und zur Entwicklung narrativer Fähigkeiten. München: Fink 1995.
- Bredel, Ursula: Ohne Worte – Zum Verhältnis von Grammatik und Textproduktion am Beispiel des Erzählens von Bildergeschichten. In: Didaktik Deutsch (2001), 11, 4-21.
- Chaudron, Craig: Data Collection in SLA Research. In: Doughty, Catherin J./Long, Michael H. (Hrsg.): The Handbook of Second Language Acquisition. Blackwell: Oxford 2005, 762-828.
- Corbett, Greville G.: Agreement. Cambridge: University Press 2006.
- Corbett, Greville G.: Gender. Cambridge: University Press 1991.

- Corbett, Greville G.: The Agreement Hierarchy. In: *Journal of Linguistics* (1979), 1, 203-224.
- Dannerer, Monika: Narrative Fähigkeiten und Individualität. Mündlicher und schriftlicher Erzählerwerb im Längsschnitt von der 5. bis zur 12. Schulstufe. Tübingen: Stauffenburg 2012.
- Dieser, Elena: Genuserwerb im Russischen und im Deutschen. Korpusgestützte Studie zu ein- und zweisprachigen Kindern und Erwachsenen. München: Verlag Otto Sagner 2009.
- Dimroth, Christine: The Finite Story (2005). Max-Planck-Institut für Psycholinguistik. <http://www.mpi.nl/departments/other-research/research-projects/information-structure-project/tools/the-finite-story> (letzter Zugriff 07.10.2015).
- Dimroth, Christine: Videoclips zur Elizitation von Erzählungen. Methodische Überlegungen und einige Ergebnisse am Beispiel der „Finite Story“. In: Ahrenholz, Bernt (Hrsg.): *Einblicke in die Zweitspracherwerbsforschung und ihre methodischen Verfahren*. Berlin: de Gruyter 2012, 77-97.
- Ehlich, Konrad/Bredel, Ursula/Reich, Hans H.: Referenzrahmen zur altersspezifischen Sprachaneignung. Berlin: Bundesministerium für Bildung und Forschung 2008.
- Fix, Martin: *Texte schreiben. Schreibprozesse im Deutschunterricht*. Paderborn: Schöningh 2006.
- Flick, Uwe: *Triangulation. Eine Einführung*. Wiesbaden: Springer 2011.
- Grieffhaber, Wilhelm: Die Profilanalyse. In: Ahrenholz, Bernt (Hrsg.): *Einblicke in die Zweitspracherwerbsforschung und ihre methodischen Verfahren*. Berlin: de Gruyter 2012, 173-193.
- Grieffhaber, Wilhelm: L2-Kenntnisse und Literalität in frühen Lernertexten. In: Ahrenholz, Bernt (Hrsg.): *Empirische Befunde zu DaZ-Erwerb und zur Sprachförderung. Beiträge aus dem 3. 'Workshop Kinder mit Migrationshintergrund'*. Freiburg: Fillibach 2009, 115-135.
- Grieffhaber, Wilhelm: Sprachstandsdiagnose im Zweitspracherwerb: Funktional-pragmatische Fundierung der Profilanalyse (2005), <https://spzwww.uni-muenster.de/griesha/pub/tprofilanalyse-azm-05.pdf> (letzter Zugriff: 01.08.2018).
- Jeuk, Stefan: „Der Katze jagt den Vogel.“ Aspekte von Genus und Kasus im Grundschulalter. In: Ahrenholz, Bernt (Hrsg.): *Zweitspracherwerb. Diagnosen, Verläufe, Voraussetzungen*. Freiburg: Fillibach, 2008, 135-150.
- Jeuk, Stefan: Sprachstandsbeobachtung in den Klassen 1 und 2. In: Oomen-Welke, Ingeborg/Decker-Ernst, Yvonne (Hrsg.): *Deutsch als Zweitsprache. Beiträge zur durchgängigen Sprachbildung. Beiträge aus dem 8. Workshop „Kinder mit Migrationshintergrund“*. Stuttgart: Fillibach 2012, 159-172.
- Knapp, Werner: Erzähltheorie und Erzählerwerb. Zur Diskussion neuerer Forschungsergebnisse. In: *Didaktik Deutsch* (2001), 10, 26-48.
- Knapp, Werner: *Schriftliches Erzählen in der Zweitsprache*. Tübingen: Niemeyer 1997.
- Köpcke, Klaus-Michael/Panther, Klaus-Uwe/Zubin, David: Motivating Grammatical and Conceptual Gender Agreement in German. In: Schmid, Hans-Jörg/Handl, Susanne (Hrsg.): *Cognitive Foundations of Linguistic Usage Patterns*. Berlin: de Gruyter 2010, 171-194.

- Köpcke, Klaus-Michael/Zubin, David: Die kognitive Organisation der Genuszuweisung zu den einsilbigen Nomen der deutschen Gegenwartssprache. *Zeitschrift für Germanistische Linguistik* (1983), 166-182.
- Köpcke, Klaus-Michael/Zubin, David: Genus. In: Hentschel, Elke/Vogel, Petra Maria (Hrsg.): *Deutsche Morphologie*. Berlin: de Gruyter 2009, 132-154.
- Köpcke, Klaus-Michael: Konkurrenz bei der Genuskongruenz. Überlegungen zum Grammatikunterricht in der Sekundarstufe II. In: *Der Deutschunterricht* (2012), 1, 36-46.
- Köpcke, Klaus-Michael: *Untersuchungen zum Genussystem der deutschen Gegenwartssprache*. Tübingen: Niemeyer 1982.
- Kostyuk, Natalia: *Der Zweitspracherwerb beim Kind. Eine Studie am Beispiel des Erwerbs des Deutschen durch drei russischsprachige Kinder*. Hamburg: Dr. Kovač 2005.
- Labov, William: Das Studium der Sprache im sozialen Kontext. In: Klein, Wolfgang/Wunderlich, Dieter (Hrsg.): *Aspekte der Soziolinguistik*. Frankfurt a.M.: Athenäum 1971, 111-194.
- Loschky, Lester/Bley-Vroman, Robert: Grammar and Task-based Methodology. In: Crookes, Graham/Gass, Susan M. (Hrsg.): *Tasks and Language Learning: Integrating Theory and Practice*. Clevedon Multilingual Matters 1995, 123-167.
- Marouani, Zahida: *Der Erwerb des Deutschen durch arabischsprachige Kinder. Eine Studie zur Nominalflexion*. Dissertation, Universität Heidelberg 2006.
- Mayer, Mercer: *Frog, where are you?* New York: Dial Press, 1969.
- Mehlem, Ulrich: Schreibbanlässe und Schreibprozesse in der Grundschule – Literalität durch Interaktion. In: Mehlem, Ulrich/Said, Sahel (Hrsg.): *Erwerb schriftsprachlicher Kompetenzen im DaZ-Kontext. Diagnose und Förderung*. Freiburg: Fillibach 2010, 133-160.
- Meindl, Claudia: *Methodik für Linguisten. Eine Einführung in Statistik und Versuchsplanung*. Tübingen: Narr 2011.
- Menzel, Barbara: *Genuszuweisung im DaF-Erwerb. Psycholinguistische Prozesse und didaktische Implikationen*. Berlin: Weißensee 2004.
- Mezger, Verena/Schroeder, Christoph/Şimşek, Yazgül: Elizitierung von Lernersprache. In: Settinieri, Julia/Demirkaya, Sevilen/Feldmeier, Alexis/Gültekin-Karakoç, Nazan/Riemer, Claudia (Hrsg.): *Empirische Forschungsmethoden für Deutsch als Fremd- und Zweitsprache. Eine Einführung*. Paderborn: Schöningh 2014, 73-86.
- Mills, Anne: *The Acquisition of Gender. A Study of English and German*. Berlin: Springer 1986.
- Montanari, Elke: *Kindliche Mehrsprachigkeit. Determination und Genus*. Münster: Waxmann 2010.
- Nöth, Winfried: Der Zusammenhang von Bild und Text. In: Brinker, Klaus/Antos, Gerd/Heinemann, Wolfgang/Sager Sven F. (Hrsg.): *Text- und Gesprächslinguistik. Band 1*. Berlin: de Gruyter 2000, 489-496.
- Oleschko, Sven: Sprachfähigkeit in der Domäne der Gesellschaftswissenschaften am Beispiel funktionaler Beschreibungen. In: Ahrenholz, Bernt/Grommes, Patrick (Hrsg.): *Zweitspracherwerb im Jugendalter*. Berlin: de Gruyter 2014, 193-210.

- Preußer, Ulrike: Das Bilderbuch aus didaktischer Perspektive. In: *Didaktik Deutsch* (2015), 39, 61-73.
- Purpura, James E.: *Assessing Grammar*. Cambridge: University Press 2005.
- Quasthoff, Uta: *Erzählen im Gespräch*. Tübingen: Narr 1980.
- Redder, Angelika/Guckelsberger, Susanne/Graßer, Barbara: *Mündliche Wissensprozessierung und Konnektierung. Sprachliche Handlungsfähigkeiten in der Primarstufe*. Münster: Waxmann 2013.
- Reich, Hans/Roth, Hans-Joachim/Döll, Marion: *Fast Catch Bumerang – Auswertungshinweise, Schreibimpuls und Auswertungsbogen* (2009). https://www.uni-due.de/imperia/md/content/prodaz/reich_roth_d_ll_fastcatchbumerang.pdf (letzter Zugriff: 01.08.2018).
- Reich, Hans/Roth, Hans-Joachim: *HAVAS 5 – das Hamburger Verfahren zur Analyse des Sprachstands bei Fünfjährigen*. In: Reich, Hans/Roth, Hans-Joachim/Neumann, Ursula (Hrsg.): *Sprachdiagnostik im Lernprozess. Verfahren zur Analyse von Sprachständen im Kontext von Zweisprachigkeit*. (= *FörMig Edition Band 3.*) Münster: Waxmann 2007, 71-94.
- Roberts, Leah: *Review Article: Psycholinguistic Techniques and Resources in Second Language Acquisition Research*. In: *Second Language Research* (2012), 28, I, 113-127.
- Ruberg, Tobias: *Der Genuserwerb ein- und mehrsprachiger Kinder*. Hamburg: Dr. Kovač 2013.
- Schimke, Sarah/Hopp, Holger (Hrsg.): *Sprachverarbeitung im Zweitspracherwerb*. Berlin: de Gruyter, 2018.
- Schlobinski, Peter: *Empirische Sprachwissenschaft*. Wiesbaden: Springer 1996.
- Schu, Josef: *Formen der Elizitation und das Problem der Natürlichkeit von Gesprächen*. In: Brinker, Klaus/Antos, Gerd/Heinemann, Wolfgang/Sager Sven F. (Hrsg.): *Text- und Gesprächslinguistik. Band 2*. Berlin: de Gruyter 2000, 1013-1021.
- Slobin, Dan I.: *Crosslinguistic Evidence for the Language-making Capacity*. In: Slobin, Dan I. (Hrsg.): *The Crosslinguistic Study of Language Acquisition*. Hillsdale: Erlbaum 1985, 1157-1249.
- Spinner, Patti/Juffs, Alan: *The Acquisition of Gender in a Complex Morphological System. The Case of German*. In: *International Review of Applied Linguistics in Language Teaching* (2008), 46 (4), 315-348.
- Steele, Susan: *Word Order and Variation: A Typological Study*. In: Greenberg, Joseph/Ferguson, Charles/Moravcsik, Edith (Hrsg.): *Universals of Human Language IV, Syntax*. Stanford: University Press 1978, 585-623.
- Sunderland, Jane: *Research Questions in Linguistics*. In: Litosseliti, Lia (Hrsg.): *Research Methods in Linguistics*. London: Continuum 2010, 9-28.
- Uhl, Benjamin: *Tempus – Narration – Medialität. Eine Studie über die Entwicklung schriftlicher Erzählfähigkeit an der Schnittstelle zwischen Grammatik und Schreiben*. Baltmannsweiler: Schneider Verlag Hohengehren 2015.
- Wegener, Heide: *Das Genus im DaZ-Erwerb. Beobachtungen an Kindern aus Polen, Russland und der Türkei*. In: Handwerker, Brigitte (Hrsg.): *Fremde Sprache Deutsch*. Tübingen: Narr 1995, 1-24.

- Weinhold, Swantje: Text als Herausforderung. Zur Textkompetenz am Schulanfang. Mit 296 Schülertexten aus Klasse 1. Freiburg: Fillibach 2000.
- Wieler, Petra: Die *Bildergeschichte als Lerngegenstand* vs. *Geschichtenerzählen zu Bildern* mit mehrsprachigen Grundschulkindern. In: Becker, Tabea/Wieler, Petra (Hrsg.): Erzählforschung und Erzähldidaktik heute. Entwicklungslinien, Konzepte, Perspektiven. Tübingen: Stauffenburg 2013, 255-278.

Vignetten

Konzeption und Auswertung von schriftlichen Vignetten zur Erhebung von Aspekten fachdidaktischer Analysekompetenz

1. Vorstellung von Vignetten

Bei Vignetten handelt es sich um ein Erhebungsinstrument, das sowohl im Rahmen quantitativer als auch qualitativer Forschung zum Einsatz kommen kann. Grundsätzlich bestehen Vignetten aus einer Situierung, wie beispielsweise der schriftlichen Beschreibung einer Lehr- und Lernsituation, und einem Bearbeitungsimpuls zu dieser Situation. Die gewonnenen Daten ermöglichen Rückschlüsse auf Einstellungen und Kompetenzen.

So setzte bereits Piaget (1932) bei seinen Untersuchungen zur Moralentwicklung von Kindern sogenannte ‚story situations‘ ein. In den letzten Jahren nutzten immer mehr Studien zum Lehren und Lernen den situativen Bezug von Vignetten, um Lehrerkompetenzen zu erheben und einzuschätzen (vgl. z.B. Brovelli et al. 2013; Oser et al. 2010).

1.1 Funktionsweise von Vignetten

Vignetten in der didaktischen Forschung stellen Lehr- und Lernsituationen (auch ausschnittsweise) dar. So beschreiben Rehm und Bölsterli (2014, 218) eine ihrer schriftlichen Vignetten zum naturwissenschaftlichen Unterricht wie folgt: Nach Kontextinformationen zu Klassenstufe und vorangegangenen Stundeninhalten wird auf ein Gespräch zwischen Lehrkraft und Klasse fokussiert, das die Besprechung eines Arbeitsblattes zum Aufbau von Tulpen wiedergibt. Diese etwa halbseitige Vignette schließt mit einem Bearbeitungsimpuls, der Lehramtsstudierende auffordert, eine Rückmeldung zu dieser Situation zu notieren.

Als Erhebungsinstrument¹ zielen Vignetten in der Lehr- und Lernforschung häufig darauf, Kompetenzen zu untersuchen. Das Beispiel zeigt, wie eine solche Vignette konzeptioniert sein kann – allerdings unterscheiden sich Vignetten in der Forschungspraxis in vielerlei Hinsicht: Erste Unterschiede ergeben sich daraus, für welche *Zielpopulation* (z.B. Lehrkräfte, die Deutsch fachfremd in Klasse 1

¹ Darüber hinaus gibt es auch den Ansatz, Vignetten als Lehr- und Lernmaterialien in der Lehrer(aus)bildung einzusetzen (vgl. z.B. Kleinbub 2014).

und 2 unterrichten) aus welchem *Anlass* (z.B. deren Lernentwicklungsverlauf in einer Weiterbildung) empirische Einblicke ermöglicht werden sollen. Da dies Fragen sind, die Erhebungsinstrumente im Allgemeinen betreffen, sei an dieser Stelle auf einen Überblicksartikel für den Bereich der Lehrerkompetenzmessung von Kunter und Klusmann (2010) verwiesen.

Vignetten im Besonderen können verschiedene *Inhalte zur Situierung* nutzen, wie beispielsweise Unterrichtssequenzen mit der ganzen Klasse, Dialoge zwischen einem Schüler und der Lehrkraft, Dialoge zwischen Lehrkräften (z.B. im Lehrerzimmer vgl. Rehm/Bölsterli 2014, 217ff.), Lehr- und Lernmaterialien oder auch Schülertexte. Für die inhaltliche Konzeption von Vignetten findet sich bei Atria et al. (2006, 237f.) eine Übersicht zentraler Aspekte, wie die Passung an die Zielgruppe und die Nähe zum Alltag. Eine Möglichkeit, dies zu erreichen, besteht darin, Lehr- und Lernsituationen aufzunehmen, um daraus anschließend Material für die Vignettenkonstruktion zu gewinnen.

Des Weiteren muss man sich als Forschender für ein *Medium*² entscheiden, wobei Kombinationsmöglichkeiten gegeben sind:

- *Schriftliche* Vignetten fassen Lehr- und Lernsituationen zusammen und/oder enthalten Transkripte. Brovelli et al. (2013, 311) begründen ihren Einsatz von schriftlichen Vignetten damit, dass im Vergleich zu Filmvignetten ‚größere Unterrichtszusammenhänge‘ wiedergegeben werden können.
- *Mündliche* Vignetten geben die Situierung entweder direkt (z.B. über den Forschungsleitenden) oder indirekt über eine Aufnahme wieder. Grundsätzlich ist dabei die Flüchtigkeit von mündlichen Äußerungen zu berücksichtigen.
- *Filmvignetten* enthalten echte oder nachgespielte Lehr- und Lernsituationen. Sie gelten als besonders authentisch und ermöglichen es, schulisches Lehren und Lernen mit einem realistischen Maß an Komplexität abzubilden (vgl. z.B. Itel 2015; Oser et al. 2010).
- Auch *statische Bilder*, wie beispielsweise Standaufnahmen, sind denkbar (vgl. Schwindt 2008, 45f.).

Der *Einsatz* von Vignetten kann im *Rahmen von Interviews und Fragebögen* erfolgen. Werden Vignetten unabhängig von einem anderen Erhebungsformat verwendet, dann ist, je nach Auswertungsverfahren, auch der Begriff *Vignettentest* gebräuchlich.

Darüber hinaus unterscheiden sich Vignetten in ihrem *Bearbeitungsimpuls bzw. Bearbeitungsformat*. Grundsätzlich gibt es offene und geschlossene Formate, die wiederum verschiedene Daten generieren. Mit *geschlossenen Formaten* arbeiten beispielsweise Seidel und Prenzel (2007, 207), indem sie verschiedene Aussagen zu Unterrichtssequenzen vorgeben, die über eine vierstufige Skala (‚trifft nicht

² Die Ausführungen beziehen sich auf das Medium, über welches die Vignette angeboten wird. Die Probandinnen und Probanden selbst antworten in der Regel entweder mündlich oder schriftlich.

zu‘ bis ‚trifft zu‘) eingeschätzt werden sollen. Geschlossene Formate sind oft aufwendig in ihrer Konzeption (vgl. zu einem möglichen Erarbeitungsprozess z.B. Kanert/Resch 2014), aber vergleichsweise zeitökonomisch in ihrer Auswertung – weitere Ausführungen dazu finden sich im nächsten Teilkapitel. *Offene Formate* nutzen Bearbeitungsimpulse, so formulieren Rehm und Bölsterli (2014, 218): „Geben Sie bitte stichwortartig Ihre Rückmeldung zu dieser Unterrichtssequenz“. Ein zentrales Argument für offene Impulse besteht darin, dass geschlossene Formate durch vorgegebene Aussagen möglicherweise erst die Wahrnehmung der Probandinnen und Probanden auf problematische Aspekte lenken (vgl. Brovelli et al. 2013, 309). Andererseits kann eine offene Aufgabenformulierung auch zu wenig konkret sein. Etwa wenn der Bearbeitungsimpuls im Verhältnis zu dem Kompetenzkonstrukt, das man erheben möchte, zu ungenau ist. Dies führt wiederum dazu, dass eventuell nicht erhoben wird, was erhoben werden soll – ein Problem hinsichtlich der Validität. Rehm und Bölsterli (2014, 216) weisen auf diese Schwierigkeit hin und schlagen vor, im Rahmen der Pilotierungsphase verschiedene Formulierungen zu testen.

Ein weiterer zentraler Aspekt bei der Konzeption von Vignetten ist die Entscheidung für einen *forschungsmethodischen Ansatz*: Beim *advokatorischen Ansatz* (vgl. Oser et al. 2010, 7f.) erfolgt die Kompetenzmessung indirekt, indem die teilnehmenden Lehrkräfte das in der Vignette beschriebene bzw. gezeigte Handeln einer anderen Lehrperson bewerten. Oser et al. (2010, 22f.) zeigen, dass Lehrende bei der Beurteilung von Gruppenunterricht im Vergleich zu Nicht-Lehrpersonen tendenziell besser abschneiden. Dies gibt einen Hinweis darauf, dass der advokatorische Ansatz Rückschlüsse auf die Kompetenz von Lehrkräften ermöglichen kann. Im Rahmen des vorliegenden Beitrags wird der zweite Ansatz als *eigenproduktiver Ansatz* bezeichnet. Damit ist gemeint, dass die Lehrpersonen durch die Vignette Informationen erhalten und dazu angeregt werden, die Situierung selbst weiterzuführen; damit liegt die Erstproduktion bei der jeweiligen Probandin bzw. dem jeweiligen Probanden. Eine solche Situierung kann beispielsweise einen Schülertext, den Kontext zu seiner Entstehung und den Impuls enthalten, nun eine fördernde schriftliche Rückmeldung an die Schülerin bzw. den Schüler zu verfassen (vgl. Beispielprojekt im zweiten Teil des vorliegenden Beitrags). Auch beim eigenproduktiven Ansatz sollte geprüft werden, inwieweit Rückschlüsse auf Kompetenzen im schulischen Alltag möglich sind. Für welche der hier dargestellten Konzeptionsmöglichkeiten man sich als Forschender entscheidet, hängt letztlich von Anlass und Fragestellung der jeweiligen Studie ab.

1.2 Ausgangsdaten und Ergebnisse

Vignetten haben als Erhebungsinstrument das Potenzial, unterschiedliche Arten von Daten zu generieren. Dies hängt vor allem mit dem gewählten Bearbeitungsformat zusammen. Geschlossene Formate, wie eine mehrstufige Zustimmungsskala von ‚trifft zu‘ bis ‚trifft nicht zu‘, erzeugen Daten, die im Rahmen quantitativer Forschung ausgewertet werden. Damit stehen zahlreiche statistische Verfah-

ren zur Verfügung, die je nach konkreter Datenlage (Skalenniveau, Stichprobenumfang, ...) und Forschungsinteresse zum Einsatz kommen können. Im Folgenden sollen anhand von bereits vorliegenden Studien zwei Optionen kurz vorgestellt werden:

- Oser et al. (2010) nutzen das geschlossene Bearbeitungsformat, um mit den daraus gewonnenen Daten u.a. ein theoretisch entwickeltes Kompetenzmodell statistisch zu überprüfen. Mit Hilfe von explorativer und konfirmatorischer Faktorenanalyse zeigen sie auf, inwieweit das theoretisch modellierte Kompetenzkonstrukt beibehalten werden kann. Bezugsnorm für eine spätere Auswertung sind Expertenurteile.
- Seidel und Prenzel (2007) setzen Varianzanalysen ein und weisen nach, dass sich Personen mit unterschiedlicher Expertise im Bereich Lehren und Lernen (hier: Lehramtsstudierende, Physiklehrende und Schulinspektoren) hinsichtlich der erhobenen Kompetenz (hier: Analysekompetenz bezogen auf Physikunterricht) entsprechend systematisch unterscheiden.

Da die Auswertung geschlossener Formate als vergleichsweise zeitökonomisch gilt, eignen sie sich in besondere Weise für die Erhebung größerer Stichproben. Andersherum empfehlen sich einige statistische Auswertungsverfahren in der Regel erst ab einem bestimmten Stichprobenumfang. In Abhängigkeit von der Forschungsfrage sind Daten, die aus geschlossenen Aufgaben folgen, nicht immer zielführend. Gerade bei explorativen Fragestellungen ist eine vergleichsweise enge Betrachtung durch geschlossene Formate zu vermeiden.

Im Hinblick auf die Forschung zu Lehrerkompetenzen wird etwa darauf verwiesen, dass offene Formate eher Einblick in Problemlösefähigkeiten ermöglichen und Antworten weniger lenken bzw. einschränken (vgl. Brovelli et al. 2013, 309f.; Kanert/Resch 2014, 26). Solche Daten können sowohl im Kontext quantitativer als auch qualitativer Forschung ausgewertet werden; darüber hinaus sind sie für Mixed-Methods-Verfahren zugänglich. Unabhängig vom forschungsmethodischen Paradigma empfiehlt es sich bei der Auswertung offener Vignettenimpulse, mit einem Kategoriensystem zu arbeiten. Die Arbeit mit Kategoriensystemen ist aufwendig und ihre Konzeption hat großen Einfluss darauf, inwieweit eine Untersuchung schließlich den Gütekriterien von Forschung entspricht. Kategorien können theoriegeleitet (deduktiv) oder aus dem Datenmaterial heraus (induktiv) abgeleitet werden. Atria et al. (2006, 239) empfehlen ein deduktiv-induktives Vorgehen in mehreren Durchgängen, um ein möglichst passgenaues Kategoriensystem zu erhalten. Vor diesem Hintergrund wird auch immer wieder auf die qualitative Inhaltsanalyse (in Anlehnung an Mayring 2015 – s. Heins in diesem Band) verwiesen (vgl. z.B. Itel 2015). Brovelli et al. (2013) ergänzen ihre inhaltsanalytischen Auswertungen durch den Einsatz statistischer Verfahren und bedienen sich (auch im Sinne von Mayring 2015, 8) dem Mixed-Methods-Ansatz. Vignetten eignen sich darüber hinaus ebenso für rein explorative Fragestellungen und können zur Hypothesenbildung beitragen.

1.3 Vorteile und Nachteile von Vignetten

Lehrer sind in ihrem schulischen Alltag häufig mit komplexen Lehr- und Lernsituationen konfrontiert, in denen sie situativ und kontextbezogen handeln (vgl. Oser et al. 2010, 5). Teilnehmende Beobachtungen (s. Schwinghammer in diesem Band) oder das Videographieren (s. Maak in diesem Band) von Unterricht sind aber nicht immer durchführbar und häufig ressourcenintensiv. Im Vergleich zu reinen Wissenstest bieten Vignetten vor diesem Hintergrund eine *ökonomischere Möglichkeit*, Lehrerkompetenzen mit einem gewissen Grad an *Authentizität* und *situativem Bezug* zu erheben (vgl. Rehm/Bölsterli 2014, 213). Dies gilt im Besonderen für die Lehrerforschung, ist aber ebenso für Studien zu Schülerkompetenzen nutzbar (vgl. z.B. Atria 2006).

Ein weiterer Vorteil von Vignetten besteht darin, dass sie Kompetenzen *objektiv* und *proximal* erfassen (vgl. z.B. Kanert/Resch 2014; Rehm/Bölsterli 2014). Diese begriffliche Klassifikation wird unter Rückgriff auf Kunter und Klusmann (2010) wie folgt verstanden:

	Distale Indikatoren	Proximale Indikatoren
Subjektiver Zugang	Indirekte Kompetenzerfassung durch Selbstauskunft, z.B. über Selbstauskunft zu Abschlüssen	Direkte Kompetenzerfassung durch Selbstauskunft, z.B. über subjektive Selbsteinschätzung im Interview
Objektiver Zugang	Indirekte Kompetenzerfassung durch Dritte, z.B. über eine Überprüfung der Abschlüsse	Direkte Kompetenzerfassung durch Dritte, z.B. über Wissenstest, Vignetten

Tab. 1: Ansätze der Kompetenzerfassung
(kombinierte Übersicht nach Kunter/Klusmann 2010 sowie Rehm/Bölsterli 2014)

Grundsätzlich liefert aber auch der Einsatz von Vignetten nur Indikatoren, die wiederum Rückschlüsse auf Aspekte von Kompetenzen ermöglichen. Eine *Grenze des Verfahrens* liegt darin, dass selbst besonders authentisch gestaltete Vignetten durch die Bearbeitungssituation immer ein Stück weit kontextentbunden sind (vgl. Atria 2006, 248). Gerade die schriftliche Beurteilung von Lehr- und Lernsituationen entspricht nur bedingt dem schulischen Alltag der Lehrkräfte. Rehm und Bölsterli (2014) greifen diese Problematik teilweise auf, indem sie die Bearbeitungszeit pro Vignette begrenzen; dies soll die Unmittelbarkeit von Lehrerhandeln im Unterricht abbilden.

Darüber hinaus stellt sich die Frage, inwieweit die Auswertung von Vignetten tatsächlich Rückschlüsse auf das (kompetente) Handeln in realen Situationen erlaubt (vgl. z.B. Atria et al. 2006, 247f.). Vignetten sind kein direktes Abbild von Kompetenzen oder realer Handlungsbereitschaft – allerdings gibt es Studien, die nahelegen, dass die Forschungsarbeit mit Vignetten durchaus solche Einblicke ermöglicht. Für das Fach Mathematik finden Kersting et al. (2012) Hinweise auf Zusammenhänge zwischen der Beurteilung von Filmvignetten, Lehrerhandlungen und Schülerleistungen.

Grundsätzlich empfiehlt es sich, die Grenzen eines Erhebungsinstrumentes durch seine sorgfältige Konzeption und die Einhaltung wissenschaftlicher Gütekriterien möglichst weit zu verschieben.

1.4 Gütekriterien für den Einsatz von Vignetten

Wie bereits dargestellt, können Vignetten sowohl im Rahmen quantitativer als auch qualitativer Forschung zum Einsatz kommen; darüber hinaus sind auch Mixed-Methods-Ansätze möglich. Entsprechend umfassend stellt sich die Diskussion um die Gütekriterien beim Einsatz von Vignetten dar. Der vorliegende Beitrag skizziert einen ersten Überblick und erhebt dabei keinen Anspruch auf Vollständigkeit. Die jeweils angegebene Literatur eignet sich aber zur Ergänzung und Vertiefung. Soweit möglich werden daraus exemplarisch ausgewählte Begriffe auf die Lehr- und Lernforschung mit Vignetten bezogen.

Im Kontext *quantitativer Forschung* empfiehlt sich beispielsweise das Lehrwerk von Bühner (2011) zur Test- und Fragebogenkonstruktion. Hier (ebd., 58ff.) wird grundsätzlich zwischen *Objektivität*, *Reliabilität* und *Validität* unterschieden. So ist beispielsweise auf *Durchführungsobjektivität* zu achten. Dies bedeutet für den Vignetteneinsatz u.a., dass die Bearbeitungszeit und die Instruktionen zur Bearbeitung vorgegeben und eingehalten werden sollten. Ebenso zentral ist die *Auswertungsobjektivität*. Sie zielt auf eine übereinstimmende Beurteilung durch mehrere Auswertende. Geläufig ist auch der Begriff der Interraterreliabilität, ein verbreitetes statistisches Maß dafür ist Cohens Kappa.

Hinsichtlich der Reliabilität gibt es u.a. die *Retestmethode*, bei der ein (Vignetten-)Test von denselben Personen zu zwei verschiedenen Messzeitpunkten bearbeitet wird. Dies bietet sich vor allem bei Querschnittsstudien an, da eine Intervention den Retest verzerren kann. Dagegen kommen bei der *Paralleltestmethode* zwei parallele Versionen eines Tests zum Einsatz. Dies ist auf Grund des situativen Bezugs für Vignetten nicht immer sinnvoll umzusetzen.

Die Validität fragt danach, ob man als Forscherer tatsächlich das misst, was man messen möchte. Dabei ist die *Kriteriumsvalidität* statistischen Berechnungen zugänglich, indem etwa die Korrelation (der Zusammenhang) zwischen den Ergebnissen der Vignettenbearbeitung und anderen Kriterien (wie z.B. der Berufserfahrung in Jahren) bestimmt wird. Um *Konstruktvalidität* nachzuweisen, werden Zusammenhänge zwischen dem eigenen und anderen Erhebungsinstrumenten geprüft. Dabei sollten Instrumente mit ähnlichen Gültigkeitsbereichen eine Korrelation aufweisen; allerdings macht man sich hier abhängig von der Güte andere Erhebungsinstrumente.

Steineke (2013, 319ff.) gibt einen Überblick *qualitativer Gütekriterien*, die dem qualitativen Forschungsparadigma eher entgegenkommen und unabhängig von statistischen Berechnungen sind. Ein wichtiger Aspekt ist die *intersubjektive Nachvollziehbarkeit* durch sorgfältige Dokumentation aller Phasen des Forschungsprozesses und Interpretationen in der Gruppe. Bei der *Indikation des Forschungsprozesses* geht es darum, seine Angemessenheit zu betrachten. Dies kann

beispielsweise damit beginnen, dass die Passung zwischen Forschungsfrage und qualitativem Zugang geprüft wird. Ist es das Ziel der Forschungsarbeit, einen standardisierten Vignettentest zu konzeptionieren, sollte eine quantitative Ausrichtung gewählt werden. Als ein weiteres Kriterium lenkt die *Limitation* den Blick darauf, inwieweit eine entwickelte Theorie verallgemeinerbar ist. Die Frage nach dem Gültigkeitsbereich von Ergebnissen kann gerade in der Lehrerforschung besonders relevant werden, wenn auf Grund des schulischen Alltags nur sehr kleine Stichproben zugänglich sind. Das Kriterium der *reflektierten Subjektivität* zielt auf den Forschenden als ein Subjekt, das eigene Vorannahmen und Interessen in den Forschungsprozess miteinbringt. Gerade bei Forschungsarbeiten im Bereich Lehren und Lernen ist dies ein zentrales Kriterium – schließlich hat hier bereits jeder (zumindest aus der Schülerperspektive) eigene Erfahrungen gesammelt.

Für die Auswertung von Vignetten mit offenen Bearbeitungsimpulsen kann auch auf die qualitative Inhaltsanalyse zurückgegriffen werden. Eine Diskussion und Übersicht zu *Gütekriterien der qualitativen Inhaltsanalyse* findet sich u.a. bei Mayring (2015, 123ff.; s. auch Heins in diesem Band). Neben anderen Kriterien wird zur Bestimmung der *Reproduzierbarkeit* auch auf die *Interkoderreliabilität* hingewiesen. Studien, die Vignetten mit offenen Antwortformaten nutzen, geben häufig den Grad an Übereinstimmung zwischen zwei oder mehr Auswertenden an. Im Kontext des Kriteriums *Konstruktvalidität* findet sich ein Verweis auf die Validierung durch Experten(-befragungen).

1.5 Literaturempfehlungen

Brovelli et al. (2013): Erfassen professioneller Kompetenzen für den naturwissenschaftlichen Unterricht.

Im Beitrag wird ein Forschungsprojekt vorgestellt, das schriftliche Vignetten mit offenem Bearbeitungsformat einsetzt. Ziel ist es, Kompetenzen von Lehramtsstudierenden der Naturwissenschaften zu erheben. Die Auswertung erfolgt in Anlehnung an Mayring inhaltsanalytisch.

Bühner/Ziegler (2009): Statistik für Psychologen und Sozialwissenschaftler.

Das Lehrwerk erklärt statistische Verfahren, die bei einer quantitativen Auswertung von Vignetten wichtig sind. Die Auswertung mit den Programmen SPSS und G*Power wird kleinschrittig beschrieben.

Itel (2015): Filmvignetten zur Einschätzung sprachförderrelevanten Wissens von fröhpädagogischen Fachpersonen.

Der Beitrag stellt eine Studie aus dem Bereich der Sprachdidaktik vor. Es werden Interviews mit Filmvignetten durchgeführt; die Auswertung erfolgt qualitativ inhaltsanalytisch.

Rehm/Bölsterli (2014): Entwicklung von Unterrichtsvignetten.

Der Beitrag bezieht sich auf die Entwicklung des LUZERNer-Vignettentests. U.a. wird auf die Entwicklung des Kodiermanuals eingegangen.

2. ProfiL_13: Vignetten als Erhebungsinstrument in Weiterbildungen für Deutschlehrkräfte

Das folgende Kapitel stellt den Einsatz von schriftlichen Vignetten im Teilprojekt 13 des Promotionskollegs *Professionalisierung im Lehrberuf* (ProfiL)³ vor. Zunächst wird ein Überblick zum theoretischen Hintergrund sowie zum Forschungssetting und zu den Forschungsfragen gegeben.

2.1 Theoretische Eckpfeiler

ProfiL_13 untersucht die Entwicklung von Aspekten fachdidaktischer Analysekompetenz in Weiterbildungen⁴ für Deutschlehrkräfte. *Fachdidaktische Analysekompetenz* wird in diesem Projekt als Teilaspekt professioneller Kompetenz (vgl. Baumert/Kunter 2011) verstanden. Das Konstrukt Analysekompetenz bezieht sich vor allem auf das Analysieren von Lehr- und Lernsituationen. Seidel und Prenzel (2007) unterscheiden die Facetten *Beschreiben*, *Erklären* sowie *Bewerten* bzw. *Vorhersagen* (letzte Facette vgl. Jahn et al. 2011). Für das Fach Physik zeigen Plöger und Scholl (2014), dass die Ausbildung von Analysekompetenz messbar fachspezifisch ist. Im Rahmen von ProfiL_13 wird das Konstrukt Analysekompetenz für die jeweiligen deutschdidaktischen Lernbereiche der untersuchten Lehrerweiterbildungen adaptiert und konkretisiert.

Die Forschung zu *Lehrerweiterbildungen* beruht auf drei theoretischen Eckpfeilern. So unterscheiden sich Weiterbildungen nach Lipowsky (2010, 52ff.) hinsichtlich ihrer *Reichweite*. Diese lässt sich anhand von vier Ebenen kategorisieren: Einschätzung (Akzeptanz) der Weiterbildung, Lehrerkognitionen, tatsächliches Lehrerhandeln im Unterricht und die Schülerenebene. ProfiL_13 fokussiert auf die ersten beiden Ebenen und untersucht mit Aspekten fachdidaktischer Analysekompetenz vor allem die Entwicklung von Lehrerkognitionen.

Neben der Reichweite einer Weiterbildung stellen *wirksame Merkmale* von Fortbildungen den zweiten theoretischen Eckpfeiler dar. Als eines dieser Merkmale gilt die Kopplung von Präsenz- und Zwischenphasen (vgl. Lipowsky/Rzejak 2012, 7f.). Das *Forschungsfeld von ProfiL_13* sind die Ludwigsburger Weiterbildungen⁵, die mit einer Kombination aus drei Präsenztagen und zwei Zwischenphasen arbeiten. In den etwa vierwöchigen Zwischenphasen sind die Lehrkräfte an ihren Schulen und haben die Möglichkeit, Inhalte sowie Materialien aus der Weiterbildung in ihrem schulischen Alltag zu erproben. Die Zwischenphasen

³ Weitere Informationen zum Gesamtprojekt ProfiL und zu den Teilprojekten unter <http://profil.ph-bw.de> (Gefördert vom Ministerium für Wissenschaft, Forschung und Kunst BW).

⁴ Die Begriffe *Fortbildung*, *Weiterbildung* und *Professionalisierungsmaßnahme* unterscheiden sich kontextabhängig. In Anlehnung an Lipowsky (2010, 51) werden sie im vorliegenden Beitrag allerdings synonym verwendet.

⁵ Weitere Informationen zu dem IQF- Projekt LuWe unter: <http://www.ph-ludwigsburg.de/luwe>.

werden durch eine Lernplattform begleitet; dies zielt auf einen Zuwachs an Reflexion und Austausch. In den Fächern Deutsch und Mathematik gibt es regelmäßig Angebote zu verschiedenen didaktischen Themenbereichen.

Der dritte theoretische Eckpfeiler bezieht sich auf die *individuellen Voraussetzungen* der teilnehmenden Lehrkräfte in den Bereichen der Kognition, Motivation, Volition und Persönlichkeit (vgl. Lipowsky 2010, 62ff.). So deutet eine Analyse verschiedener Studien (vgl. Rzejak et al. 2014, 141) darauf hin, dass individuelle Lehrervoraussetzungen Einfluss auf den Wirkungsertrag von Weiterbildungen haben. ProfiL_13 untersucht in diesem Kontext die Selbstregulation der teilnehmenden Lehrkräfte über das diagnostische Instrument AVEM (vgl. Schaar Schmidt 2004) – wie es im Rahmen von COACTIV bereits genutzt wurde (vgl. Klusmann 2011, 279).

2.2 Forschungsfragen und Forschungssetting

Die grundlegende Forschungsfrage im Rahmen von ProfiL_13 ist vor diesem Hintergrund: ‚Inwieweit und in Abhängigkeit wovon entwickelt sich fachdidaktische Analysekompetenz in den untersuchten Lehrerweiterbildungen für Deutschlehrkräfte?‘ Es ergeben sich die folgenden Teilfragen – jeweils konkretisiert für die deutschdidaktischen Lernbereiche einer Weiterbildung:

- Haben die untersuchten Fortbildungen einen empirisch nachweisbaren Effekt auf die Entwicklung von Aspekten fachdidaktischer Analysekompetenz?
- Welchen Einfluss hat die tatsächliche Nutzung der Zwischenphasen durch die teilnehmenden Lehrerinnen und Lehrer auf die Entwicklung von Aspekten fachdidaktischer Analysekompetenz?
- Welchen Einfluss haben Kontextfaktoren wie Selbstregulation und Arbeitsbelastung auf die Entwicklung von Aspekten fachdidaktischer Analysekompetenz?

ProfiL_13 basiert auf einer Längsschnittstudie und nutzt sowohl quantitative als auch qualitative Methoden. Um die Forschungsfrage in ihren Teilaspekten erfassen zu können, kommen an den drei Präsenztagen verschiedene Erhebungsinstrumente zum Einsatz⁶:

⁶ Das Forschungssetting wurde bis Mai 2015 zusammen mit Tanja Walz konzeptioniert. Dies schließt eine gemeinsame Arbeit an den Fragebögen bis Mai 2015 ein.



Abb. 2: Forschungssetting

Die erste Teilfrage zur Entwicklung von Aspekten fachdidaktischer Analysekompetenz stellt den Bezugspunkt für die anderen Teilfragen und gleichzeitig den Schwerpunkt des Forschungsprojekts dar. Sie wird in einem *Pretest-Treatment-Posttest-Design* über die Bewertung von Vignetten erhoben.

2.3 Konzeption und Einsatz der Vignetten

Gerade in Lehrerweiterbildungen ist der Bedarf an relevanten Handlungskompetenzen groß. Schließlich befinden sich die teilnehmenden Lehrkräfte mitten im schulischen Alltag. Entsprechend sind die Lernbereiche der untersuchten Lehrerweiterbildungen daraufhin konzeptioniert, die Aneignung bzw. Vertiefung wichtiger Teilkompetenzen für den Deutschunterricht zu unterstützen; dabei geht es auch um Aspekte fachdidaktischer Analysekompetenz. Vignetten kommen Lehr- und Lernsituationen in ihrer Komplexität und situativen Gebundenheit vergleichsweise nahe und eignen sich daher in besonderer Weise für dieses Forschungsprojekt (vgl. Oser et al. 2010, 5).

Da es um die Entwicklung von Kompetenzaspekten in einem spezifischen Fortbildungsangebot geht, ermöglichen inhaltlich passende Vignetten „[...] differenzierte Einblicke in die Wirkungsweise [...]“ (Atria et al. 2006, 247). Entsprechend orientiert sich die Konzeption der Vignetten in ProfiL_13 an den jeweiligen Lernbereichen einer Weiterbildung. Die folgenden Ausführungen beziehen sich exemplarisch auf eine Weiterbildung im Bereich der Schreibdidaktik für die Grundschule. In einem ersten Schritt wurden Vignetten zu den zentralen Lernbereichen theoriegeleitet erstellt: (1) Bewerten und Beurteilen von Schülertexten, (2) Überarbeiten von Schülertexten, (3) Übergang vom mündlichen zum schriftlichen Erzählen.

Da im Deutschunterricht der Umgang mit (teils umfangreichen) Aufgaben und Schülertexten immer wieder eine wichtige Rolle spielt, wurden schriftliche Vignetten als besonders zielführend erachtet. Die Anbindung an authentische Lehr- und Lernsituationen erfolgte zunächst durch Rücksprache mit dem jeweiligen Dozierenden der Weiterbildung. Alle Dozentinnen und Dozenten konzeptionieren ihre Weiterbildung selbst und verfügen sowohl über eigene Unterrichts- als auch

über mehrjährige Weiterbildungserfahrung. Die Pilotierungsphase⁷ bestand aus den folgenden Bausteinen:

- Qualitative Experteninterviews (n=3) zur ersten Validierung:
Inhaltlich ging es um die Authentizität und die (fachdidaktische) Relevanz der Situationen. Außerdem wurden mögliche Verständnisschwierigkeiten in Bezug auf die Kontextdarstellung thematisiert. Darüber hinaus befassten sich die Interviews mit der Frage, welche Aspekte bei der Auswertung eine Rolle spielen sollten.
- Pilotierung in einer Lehrerweiterbildung (Erhebungszeitpunkt t1 mit n=11 und t2 mit n=8):
Die Lehrkräfte bearbeiteten drei Vignetten vor und nach der Weiterbildung. Zusätzlich bewerteten sie jede Vignette hinsichtlich mehrerer Aspekte, wie beispielsweise deren Authentizität und (schulalltäglichen) Relevanz.
- Einsatz in einem sprachdidaktischen Seminar für Lehramtsstudierende: Da sich Einsatz und Intervention in diesem Seminar von der Lehrerweiterbildung unterschieden, werden diese Ergebnisse im Folgenden nur stellenweise herangezogen. So konnte die Interraterreliabilität für die Kodieranleitung über eine größere Stichprobe abgesichert werden (t1 mit n=22). Außerdem wurde zusätzlich eine Variation im Bearbeitungsimpuls getestet (t1 mit zusätzlich n=10).

Zum derzeitigen Stand des Forschungsprojekts liegen Daten und Ergebnisse aus der Pilotierung vor. Es handelt sich um erste Hinweise darauf, inwieweit der Einsatz von Vignetten hier geeignet ist, um Einblicke in die Wirksamkeit der Weiterbildung zu erhalten.

2.4 Kodieranleitung

Neben einer Vignette mit explorativem Charakter kamen zwei Vignetten zum Einsatz, die über Kodieranleitungen ausgewertet wurden. Die Entwicklung der Kategorien erfolgte in mehreren Schritten. Nach der ersten theoriegeleiteten (deduktiven) Konzeption erfolgte eine induktive Ausdifferenzierung am Datenmaterial; dabei fanden auch die Experteninterviews Berücksichtigung. Im Anschluss an die Kodierung von Teildatensätzen wurden die Anleitungen im Austausch mit dem Zweitrater geschärft. Zentrale Elemente einer *Kodieranleitung* sind der Kode (hier in Zahlen), eine prägnante Definition, die Regeln der Kodierung und Beispiele (vgl. z.B. Hammann/Jördens 2014, 175f.). Die Ankerbeispiele dienen dazu, typische Antworten für eine Kodierung darzustellen. Dagegen können Abgrenzungsbeispiele dabei helfen, die Kodierung in Grenzfällen eindeutiger zu gestalten.

Im Folgenden wird exemplarisch für die Vignette 3 „*Übergang vom mündlichen zum schriftlichen Erzählen*“ auf die Kodierung eingegangen. Die Vignette enthält

⁷ Die Haupterhebung umfasst den Zeitraum von Oktober 2015 bis Juli 2016 und strebt einen höheren Stichprobenumfang für Fortbildungen zu verschiedenen deutschdidaktischen Bereichen an.

einen Dialog zwischen der Lehrkraft und einer Schülerin. Die Schülerin arbeitet an einer schriftlichen Erzählung und kommt nicht weiter. Daraufhin fordert die Lehrkraft sie auf, mündlich zu erzählen. Nachdem die Schülerin das getan hat, schließt der Dialog mit dem Impuls der Lehrkraft an die Schülerin, das mündlich Erzählte nun aufzuschreiben. In Anlehnung an Rehm und Bölsterli (2014, 218) lautet der offene Bearbeitungsimpuls für die Weiterbildungsteilnehmer: „Bitte schreiben Sie Ihre Rückmeldung zu der dargestellten Situation – bezogen auf das Erzählen im Deutschunterricht der Grundschule – in das folgende freie Feld. Begründen Sie dabei auch.“ Die Kodierung zielt nun darauf, inwieweit die Lehrkräfte den Aspekt aufgreifen, dass zwar eine große Nähe zwischen mündlichen und schriftlichen Erzählungen besteht – diese aber wegen konzeptioneller Unterschiede nicht bedenkenlos genutzt werden kann (vgl. z.B. Fix 2006, 94). So gebrauchen mündliche Erzählungen beispielsweise das Perfekt und sind vergleichsweise dialogischer gestaltet.

Unter Bezugnahme auf das Konstrukt ‚Fachdidaktische Analysekompetenz‘ ist die Kodierung für diese Vignette wie folgt aufgebaut (unter Anlehnung an Brovelli et al. 2013; Rehm/Bölsterli 2014; Schwindt 2008; Seidel/Prenzel 2007):

- (0) Der Aspekt wird nicht benannt.
- (1) Der Aspekt wird benannt (im Sinne einer kurzen Beschreibung); die Verwendung von Fachbegriffen ist dabei nicht zwingend.
- (2) Der Aspekt wird benannt und bewertet. Dabei geht es nicht um eine Bewertung im Sinne von ‚richtig‘ oder ‚falsch‘. Als Bewertung gilt, wenn der Aspekt zur Diskussion gestellt wird.
- (3) Der Aspekt wird benannt, bewertet und die Antwort enthält einen alternativen Handlungsvorschlag. Dabei wird nur kodiert, wenn der Handlungsvorschlag funktional⁸ ist.
- (4) Der Aspekt wird benannt, bewertet und erklärt. Die Erklärung stellt einen Bezug zu fachdidaktischen Konzeptionen her.
- (5) Der Aspekt wird benannt, bewertet und erklärt. Zusätzlich enthält die Antwort einen alternativen und funktionalen Handlungsvorschlag.

Kode	Definition	Regeln zur Kodierung	Beispiele
0	Nicht- Benennung des „problematischen“ Übergangs vom mündlichen zum schriftlichen Erzählen <i>Hinweis:</i> Damit sind Unterschiede zwischen konzeptioneller Mündlichkeit und Schriftlichkeit gemeint. [...]	Dieser Kode wird vergeben, wenn die Antwort den „problematischen“ Übergang vom mündlichen zum schriftlichen Erzählen nicht aufgreift. <i>Hinweise zur Abgrenzung:</i> - Eine indirekte Thematisierung (z.B. durch den Verweis auf Strukturierungsmethoden) reicht nicht aus. [...]	„Ich denke zunächst würde ich mündlich Ideen gemeinsam mit der Klasse sammeln. Daraus entstehen oft neue Impulse für Kinder und auch eigene Schreibideen. Auch Stichworte → oder Bilder können Kindern helfen, die zunächst keine Ideen haben.“ (Ankerbeispiel) [...]

Abb. 3: Auszug aus der Kodieranleitung zu Vignette 3, Kode (0)

⁸ Der Begriff wird etwa von Ziegler und Dresel (2006) in Bezug auf Lernstrategien von Schülerinnen und Schülern verwendet und hier für den Bereich der Lehrerkompetenzen adaptiert.

2.5 (Erste) Ergebnisse

Aus der Pilotierungsphase liegen verschiedene Ergebnisse vor. Die Prüfung der *Interraterreliabilität* erfolgte für jede Kodieranleitung, jeden Messzeitpunkt (t1, t2) und jede Stichprobe (Lehrkräfte, Studierende). Neben der prozentualen Übereinstimmung wurde Cohens Kappa berechnet (vgl. z.B. Hamman/Jördens 2014, 177). Cohens Kappa berücksichtigt auch die zufälligen Übereinstimmungen zwischen den Auswertenden. Bortz und Döring (2006, 277) sprechen Werten zwischen 0,6 und 0,75 eine gute Übereinstimmung zu. Diese Vorgabe konnte bei allen Überprüfungen der Interraterreliabilität eingehalten und in der Mehrheit mit Werten $> 0,75$ übertroffen werden.

Die *Bewertung der Vignetten durch die Lehrkräfte* hinsichtlich Aspekten wie Authentizität und Relevanz diente dazu, eine erste Rückmeldung zur Akzeptanz der Vignetten einzuholen. Außerdem sind Lehrkräfte Experten für die Frage, inwieweit Lehr- und Lernsituationen als authentisch gelten können. Entsprechend gibt diese Rückmeldung zusätzliche Hinweise für den Validierungsprozess. In Anlehnung an Seidel et al. (2011, 264) sowie Seidel und Prenzel (2007, 208) erhielten die Lehrkräfte Aussagen zu mehreren Aspekten. Die fünfstufige Skala reichte von 1 (,trifft gar nicht zu‘) bis 5 (,trifft völlig zu‘); angegeben sind jeweils Mittelwert (M) und Standardabweichung (SD):

	Vignette 1 M (SD)	Vignette 2 M (SD)	Vignette 3 M (SD)
interessant	3,82 (0,98)	4,27 (0,65)	4,45 (0,82)
relevant	4,18 (1,25)	4,73 (0,65)	4,64 (0,67)
authentisch	4,64 (0,92)	4,73 (0,65)	4,73 (0,65)
leicht hineinzusetzen	3,82 (0,98)	4,18 (1,08)	4,45 (0,69)

Tab. 1: Bewertung der Vignetten (Lehrkräfte, t1 mit n=11)

Die Werte sprechen sowohl für die Akzeptanz der Vignetten als auch für deren Authentizität aus Sicht der Lehrkräfte.

Um einen ersten Hinweis auf die *Wirksamkeit der Weiterbildung* zu erhalten, wurde der t-Test eingesetzt. Der t-Test für abhängige Stichproben prüft, ob ein signifikanter Unterschied zwischen den Antworten der Lehrkräfte vor und nach der Weiterbildung besteht. Für Vignette 3 ‚Übergang vom mündlichen zum schriftlichen Erzählen‘ und der oben dargestellten Kodierung zeigte der t-Test mit $p < .001$ einen hoch signifikanten Unterschied zugunsten der Antworten nach der Weiterbildung (zum t-Test vgl. z.B. Bühner/Ziegler 2009, 232ff.). Im Anschluss daran wurde mit G*Power die Effektstärke als Maß dafür bestimmt, ob das Ergebnis des t-Tests aus statistischer Sicht praktisch bedeutsam ist. Mit $d_z=2.31$

liegt ein hoher Effekt vor; die Teststärke fällt mit $1 - \beta > .99$ ebenfalls hoch aus.⁹ Interessant ist außerdem die Verteilung der kodierten Lehrerantworten zu Messzeitpunkt t2. Über die Hälfte der Lehrkräfte erhält den Kode 3 und scheint damit auf den Handlungsvorschlag zu fokussieren. Dies könnte darauf hindeuten, dass Inhalte der Weiterbildung vor allem hinsichtlich ihres Handlungspotenzials abgespeichert werden. Aufgrund der Stichprobengröße ($n=8$) muss allerdings betont werden, dass die Aussagekraft hier begrenzt ist.

Auf diesem Weg lassen sich vor allem erste Hinweise auf die Wirksamkeit der Weiterbildung gewinnen. Eine darüber hinausgehende, qualitativ inhaltsanalytische Auswertung legt nahe, dass die Antworten mit der Kodierung 0 zwei Typen zuzuordnen sind. Typ A äußert sich zu der beschriebenen Vorgehensweise der Lehrkraft durchweg positiv. Typ B führt Handlungsalternativen an, die allerdings nicht auf die konzeptionellen Unterschiede zwischen mündlichem und schriftlichem Erzählen zielen. Des Weiteren ist noch qualitativ inhaltsanalytisch zu prüfen, inwieweit für Kodierung 3 und 5 ebenfalls Typen an funktionalen Handlungsvorschlägen unterschieden werden können.

Vignette 2 situiert einen fördernden Lehrerkommentar (vgl. z.B. Becker-Mrotzek/Böttcher 2011, 98). Dabei wurden mit verschiedenen Kodieranleitungen drei Aspekte ausgewertet: Positives Feedback, Überarbeitungshinweise und sprachliche Angemessenheit. Der t-Test zeigte für alle drei Aspekte einen signifikanten Unterschied ($p < .05$) zugunsten der Antworten nach der Weiterbildung. Auch die Effekt- und Teststärken fallen hoch aus; lediglich die Teststärke zur sprachlichen Angemessenheit liegt mit einem Wert von $1 - \beta = .77$ knapp unter der geforderten Vorgabe von $.80$. Darüber hinaus sollte noch qualitativ inhaltsanalytisch geprüft werden, inwieweit sich die Überarbeitungshinweise auf Revisionen unterschiedlicher Tiefe beziehen (vgl. Fix 2006, 166). Kodiert wurde lediglich, wie konkret die Überarbeitungshinweise gestaltet waren.

2.6 (Erstes) Fazit zur Arbeit mit den Vignetten

Die Rückmeldung der Lehrkräfte zu den einzelnen Vignetten legt nahe, dass sie die dargestellten Lehr- und Lernsituationen als interessant, relevant und authentisch einschätzten. Gerade ein gelungener situativer Bezug kann die Vignetten zu mehr als einem Erhebungsinstrument machen. So gab es nach dem zweiten Messzeitpunkt eine (freiwillige) Gesprächsrunde, in der Veränderungen im eigenen Antwortverhalten reflektiert wurden. Dadurch erhalten die Vignetten einen *zusätzlichen Mehrwert für die teilnehmenden Lehrkräfte* und werden potenziell zu einem Modulbaustein der Weiterbildung.

⁹ Zur weiteren Absicherung wurden die Differenzen auch mit einem non-parametrischen Verfahren geprüft. Die Ergebnisse des Wilcoxon-Tests bestätigen für diesen Aspekt (Vignette 3) sowie die Aspekte ‚positives Feedback‘ und ‚Überarbeitungshinweise‘ (Vignette 2, siehe weitere Ausführungen in diesem Teilkapitel) bei $p < .05$ signifikante Unterschiede mit starken Effekten.

Durch die Entscheidung für ein offenes Antwortformat waren die erzeugten Daten quantitativen und qualitativen Auswertungsverfahren zugänglich. Dadurch waren *verschiedene Perspektiven* auf die Entwicklung von Aspekten fachdidaktischer Analysekompetenz möglich. Schließlich bringen quantifizierende Verfahren im vorliegenden Fall eine abstufende Wertung der Lehrerantworten mit sich. Ein weiterer Blick auf strukturelle Veränderungen während der Weiterbildung sollte zunächst nicht zwingend von Wertigkeiten ausgehen. Dies öffnet die Forschungsperspektive hinsichtlich der Wirkungsweise der untersuchten Lehrerweiterbildung.

Die Vignetten gaben in der Pilotierungsphase – trotz des geringen Stichprobenumfangs – ernstzunehmende Hinweise auf die Wirksamkeit der untersuchten Fortbildung. Allerdings liegt die *Grenze des Verfahrens* darin, dass Vignetten zwar auf Wirklichkeit zielen, diese aber nicht abbilden. So bleiben die Gründe dafür, warum eine Probandin bzw. ein Proband das jeweilige Kreuz setzt oder eine bestimmte Antwort formuliert immer ein Stück weit verborgen. Im vorliegenden Fall würde es sich anbieten, einzelnen Lehrkräften außerhalb der Weiterbildung die Vignetten vorzulegen und mit der Methode des Lauten Denkens (s. Dannecker in diesem Band) bearbeiten zu lassen.

Selbst wenn man die Wirksamkeit der Weiterbildung für die untersuchten Teilbereiche als nachgewiesen annimmt, sind nur erste und keine direkten Rückschlüsse auf Veränderungen im schulischen Alltag leistbar. Es bleibt die Frage, wie die teilnehmenden Lehrkräfte nun in entsprechenden, realen Lehr- und Lernsituationen handeln. Um darauf erste Antworten zu finden, wären weitere Erhebungen notwendig, wie beispielsweise das Videographieren von Unterricht (vgl. z.B. Kleinbub 2010). Ist man sich dieser Grenzen bewusst, dann bieten Vignetten einen nutzbaren empirischen Zugang zu Aspekten von Kompetenzen. Vor diesem Hintergrund erscheinen die eingesetzten Vignetten im Rahmen von Profil_13 bisher als ein geeignetes Erhebungsinstrument, um die Entwicklung von Aspekten fachdidaktischer Analysekompetenz in einem ersten situativen Kontext zu erforschen.

Literatur

- Atria, Moira/Strohmeier, Dagmar/Spiel, Christiane: Der Einsatz von Vignetten in der Programmevaluation – Beispiele aus dem Anwendungsfeld «Gewalt in der Schule». In: Flick, Uwe (Hrsg.): *Qualitative Evaluationsforschung. Konzepte – Methoden – Umsetzung*. Reinbek bei Hamburg: Rowohlt Taschenbuch Verlag 2006, 233-249.
- Baumert, Jürgen/Kunter, Mareike: Das Kompetenzmodell von COACTIV. In: Kunter, Mareike et al. (Hrsg.): *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. Münster et al.: Waxmann 2011, 29-53.
- Becker-Mrotzek, Michael/Böttcher, Ingrid: *Schreibkompetenz entwickeln und beurteilen. Praxishandbuch für die Sekundarstufe I und II*. Berlin: Cornelsen Scriptor 2011.
- Bortz, Jürgen/Döring, Nicola: *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Heidelberg: Springer 2006.

- Brovelli, Dorothee/Bölsterli, Katrin/Rehm, Markus/Wilhelm, Markus: Erfassen professioneller Kompetenzen für den naturwissenschaftlichen Unterricht: Ein Vignettest mit authentisch komplexen Unterrichtssituationen und offenem Aufgabenformat. In: *Unterrichtswissenschaft* (2013), 41, 4, 306-329.
- Bühner, Markus/Ziegler, Matthias: *Statistik für Psychologen und Sozialwissenschaftler*. München et al.: Pearson Studium 2009.
- Bühner, Markus: *Einführung in die Test- und Fragebogenkonstruktion*. München et al.: Pearson Studium 2011.
- Fix, Martin: *Texte schreiben. Schreibprozesse im Deutschunterricht*. Paderborn: Schöningh 2006.
- Hammann, Markus/Jördens, Janina: Offene Aufgaben codieren. In: Krüger, Dirk et al. (Hrsg.): *Methoden in der naturwissenschaftsdidaktischen Forschung*. Berlin et al.: Springer Spektrum 2014, 169-178.
- Itel, Nadine: Filmvignetten zur Einschätzung sprachförderrelevanten Wissens von frühpädagogischen Fachpersonen. In: Bräuer, Christoph/Wieser, Dorothee (Hrsg.): *Lehrende im Blick. Empirische Lehrerforschung in der Deutschdidaktik*. Wiesbaden: Springer VS 2015, 301-319.
- Jahn, Gloria/Prenzel, Manfred/Stürmer, Kathleen/Seidel, Tina: Varianten einer computer-gestützten Erhebung von Lehrerkompetenzen: Untersuchungen zu Anwendungen des Tools Observer. In: *Unterrichtswissenschaft* (2011), 39, 2, 136-153.
- Kanert, Georg/Resch, Mario: Erfassung geschichtsdidaktischer Wissensstrukturen von Geschichtslehrkräften anhand eines vignettengestützten Testverfahrens. In: *Zeitschrift für Geschichtsdidaktik* (2014), 13, 15-31.
- Kersting, Nicole B./Givvin, Karen B./Thompson, Belinda J./Santagata, Rossella/Stigler, James W.: Measuring usable knowledge: Teachers' analyses of mathematics classroom videos predict teaching quality and student learning. In: *American Educational Research Journal* (2012), 49, 3, 568-589.
- Kleinbub, Iris: Grammatik unterrichten – Professionsorientierung in der Lehrerbildung. In: Bredel, Ursula/Schmellentin, Claudia (Hrsg.): *Welche Grammatik braucht der Grammatikunterricht?* Baltmannsweiler: Schneider Verlag Hohengehren 2014 (Thema Sprache – Wissenschaft für den Unterricht, Bd. 8), 135-159.
- Kleinbub, Iris: *Unterrichtsqualität im Leseunterricht. Eine videobasierte Analyse in vier-ten Klassen*. Trier: WVT 2010 (Koblenz-Landauer Studien zu Geistes-, Kultur- und Bildungswissenschaften, Bd. 9).
- Klusmann, Uta: Allgemeine berufliche Motivation und Selbstregulation. In: Kunter, Mareike et al. (Hrsg.): *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. Münster et al.: Waxmann 2011, 277-294.
- Kunter, Mareike/Klusmann, Uta: Kompetenzmessung bei Lehrkräften – Methodische Herausforderungen. In: *Unterrichtswissenschaft* (2010), 38, 1, 68-86.
- Lipowsky, Frank/Rzejak, Daniela: Lehrerinnen und Lehrer als Lerner – Wann gelingt der Rollentausch? Merkmale und Wirkungen wirksamer Lehrerfortbildungen. In: *Schulpädagogik heute* (2012), 3, 5, 1-17.
- Lipowsky, Frank: Empirische Befunde zur Wirksamkeit von Lehrerfortbildung. In: Müller, Florian H. et al. (Hrsg.): *Lehrerinnen und Lehrer lernen. Konzepte und Befunde zur Lehrerfortbildung*. Münster et al.: Waxmann 2010, 51-70.
- Mayring, Philipp: *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. Weinheim et al.: Beltz 2015.

- Oser, Fritz et al.: Die Messung der Qualität von professionellen Kompetenzprofilen von Lehrpersonen mit Hilfe der Einschätzung von Filmvignetten. In: *Unterrichtswissenschaft* (2010), 38, 1, 5-27.
- Piaget, Jean: *Le jugement moral chez l'enfant*. Paris: Alcan 1932.
- Plöger, Wilfried/Scholl, Daniel: Analysekompetenz von Lehrpersonen – Modellierung und Messung. In: *Zeitschrift für Erziehungswissenschaft* (2014), 17, 1, 85-112.
- Rehm, Markus/Bölsterli, Katrin: Entwicklung von Unterrichtsvignetten. In: Krüger, Dirk et al. (Hrsg.): *Methoden in der naturwissenschaftsdidaktischen Forschung*. Berlin et al.: Springer Spektrum 2014, 213-225.
- Rzejak, Daniela et al.: Facetten der Lehrerfortbildungsmotivation – eine faktorenanalytische Betrachtung. In: *Journal for Educational Research Online* (2014), 6, 1, 139-159. <http://www.j-f-b.de/index.php/jero/article/viewFile/412/197> (letzter Zugriff: 01.08.2018).
- Schaarschmidt, Uwe (Hrsg.): *Halbtagsjobber? Psychische Gesundheit im Lehrerberuf – Analyse eines veränderungsbedürftigen Zustandes*. Weinheim et al.: Beltz 2004.
- Schwindt, Katharina: Lehrpersonen betrachten Unterricht. Kriterien für die kompetente Unterrichtswahrnehmung. Münster et al.: Waxmann 2008 (*Empirische Erziehungswissenschaft*, Bd. 10).
- Seidel, Tina et al.: Teacher learning from analysis of videotaped classroom situations: Does it make a difference whether teachers observe their own teaching or that of others? In: *Teaching and Teacher Education* (2011), 27, 259-267.
- Seidel, Tina/Prenzel, Manfred: Wie Lehrpersonen Unterricht wahrnehmen und einschätzen. Erfassung pädagogisch-psychologischer Kompetenzen mit Videosequenzen. In: Prenzel, Manfred et al. (Hrsg.): *Kompetenzdiagnostik*. Wiesbaden: VS Verlag für Sozialwissenschaften 2007 (*Zeitschrift für Erziehungswissenschaft Sonderheft* 8), 201-216.
- Steineke, Ines: Gütekriterien qualitativer Forschung. In: Flick, Uwe (Hrsg.): *Qualitative Forschung. Ein Handbuch*. Reinbek bei Hamburg: Rowohlt Taschenbuch Verlag 2013, 319-331.
- Ziegler, Albert/Dresel, Markus: Lernstrategien: Die Genderproblematik. In: Mandl, Heinz/Friedrich, Helmut Felix (Hrsg.): *Handbuch Lernstrategien*. Göttingen et al.: Hogrefe 2006, 378-389.

Gesprächsanalyse

Mikroanalytische Beschreibung sprachlicher Interaktion in Bildungs- und Lernzusammenhängen

1. Vorstellung der Gesprächsanalyse

Der vorliegende Beitrag führt in die Gesprächsanalyse ethnomethodologischer und somit soziologischer Provenienz ein. Aus dieser ursprünglichen Form der Analyse von Interaktionen haben sich inzwischen weitere Spielarten der Gesprächsanalyse entwickelt. Insbesondere im deutschsprachigen Raum ist die ethnomethodologische Konversationsanalyse (KA) stark von der Linguistik rezipiert und im Hinblick auf linguistische Fragestellungen weiter entwickelt und genutzt worden. So machen u.a. die interaktionale Linguistik (Selting/Couper-Kuhlen 2000) und die Analyse kommunikativer Gattungen (Günthner/Knoblauch 1994) konversationsanalytische Vorgehensweisen fruchtbar. Forschungen im Bereich Sprach- und Diskurserwerb und Sprachdidaktik (Hausendorf/Quasthoff 1996; Heller 2012; Morek 2012; Morek/Heller/Quasthoff 2017; Harren 2015) nutzen die Konversationsanalyse verstärkt erst in jüngerer Zeit.

Die (ethnomethodologische) Gesprächsanalyse ist ein *qualitatives, mikroanalytisches* Verfahren der Beschreibung sozialer Interaktion. ‚Qualitativ‘ heißt: Sie ist – gemäß ihrer ethnomethodologischen Ausrichtung (vgl. 1.1) – primär an Fragen des ‚Wie‘ sozialer Interaktion interessiert und an der empirischen Aufdeckung von Prinzipien und Strukturen als solchen: Wie laufen bestimmte Arten von Gesprächen genau ab? Welche Art von Geordnetheit liegt dem sprachlichen (und nicht-sprachlichen) Handeln von Gesprächsteilnehmenden zugrunde? Auf welche regelhafte Weise bewerkstelligen Interagierende bestimmte kommunikative Aufgaben, welche sprachlichen und kommunikativen *Praktiken* (Deppermann 2008, 9) nutzen sie dabei? Fragen der Verbreitung oder Verteilung bestimmter Phänomene und statistische Zusammenhänge, wie sie in quantitativen Arbeiten erforscht werden, spielen für gesprächsanalytische Fragestellungen keine zentrale Rolle. Sie sind auch mit gesprächsanalytischen Methoden nicht oder nur mit ungeheurem zeitlichen und analytischen Aufwand zu bearbeiten. Denn ein ‚mikroanalytisches‘ Vorgehen bedeutet: Die Analyse berücksichtigt zentral die Ebene ‚kleinster‘ sprachlicher und interaktiver Phänomene. Sie betrachtet Gespräche

wie durch ein Mikroskop und zoomt im Analyseprozess ganz nah an Details des Miteinander-Sprechens heran.

Grundsätzlich verfährt sie dabei strikt *induktiv* (Silverman 1998, 59) und ist *rekonstruktiv* orientiert: Angesetzt wird nicht, wie bei deduktiven Ansätzen, bei theoretischen Konzepten, sondern beim konkreten Fall. Den Dreh- und Angelpunkt des Erkenntnisgewinns stellen die erhobenen, realen Gesprächsdaten dar, gemäß der methodischen Maxime, dass analytische Ergebnisse „from the data themselves“ (Schegloff/Sacks 1973, 201f.) gewonnen werden müssen. ‚Rekonstruktiven‘ Anspruch hat das gesprächsanalytische Vorgehen schließlich, weil es darauf zielt, nachzuvollziehen, wie die Gesprächsbeteiligten selbst wechselseitig ihr Handeln interpretieren und koordinieren. An die Stelle extern vom Forschenden an ein Gespräch herangetragenener Kategorisierungen und Deutungen tritt das Bestreben, aus der Innenperspektive eines Gesprächs heraus dessen geordnete Hervorbringung zu rekonstruieren. Die Konversationsanalyse arbeitet somit strikt empirisch und entwickelt Konzepte und Theorien – und i.d.R. sogar die genaue Formulierung ihrer Forschungsfragestellung – erst in der Analyse authentischer Daten sowie unter Beschränkung auf tatsächlich Beobachtbares.

Für die englische Bezeichnung „conversation analysis“ existieren im Deutschen unterschiedliche Bezeichnungen. Die wörtliche Übersetzung „Konversationsanalyse“ sollte nicht dahingehend missverstanden werden, sich nur auf die Analyse von Konversationen i.S. ‚gepflegter Gespräche‘ zu beziehen. Die Bezeichnungen „Gesprächsanalyse“ und „Gesprächsforschung“ tragen dem Interesse an jeglicher Art von Gesprächen deutlicher Rechnung.

Die ethnomethodologische Konversationsanalyse zeichnet sich dadurch aus, dass sie ihre Methodik an den konstitutiven Eigenschaften verbaler Interaktion, insbesondere ihrer

- Methodizität,
- Reflexivität,
- Interaktivität und
- Sequenzialität

ausrichtet (Deppermann 2000, 98). Bevor wir im folgenden Kapitel methodische Prinzipien und Vorgehensweisen erläutern, gilt es zunächst, den ethnomethodologischen Entstehungshintergrund sowie die konstitutiven Eigenschaften von Gesprächen näher zu betrachten.

1.1 Theoretischer Rahmen der Gesprächsanalyse

Der Begründer der Ethnomethodologie, Harold Garfinkel (1917-2011), beschäftigt sich mit einer Grundfrage der Soziologie, der Frage nämlich, wie angesichts der Tatsache, dass Menschen egoistische Ziele verfolgen, soziale Ordnung möglich ist. Garfinkel (1967) sucht eine Antwort auf diese Frage in den Prozessen der alltäglichen Sinnkonstitution. Er arbeitet heraus, wie sich Interaktanten den Sinn ihres Handelns wechselseitig anzeigen und verstehbar machen – und zwar auf

‚methodische‘ Art und Weise. Ziel der Ethnomethodologie ist es, diese *Methodizität* des Alltagshandelns offenzulegen, d.h.

die als selbstverständlich hingegenommenen Praktiken und Verfahren (Methoden) zu bestimmen, mittels derer die Mitglieder einer Gesellschaft (ethnos) in ihrem Handeln das eigene Tun wahrnehmbar und erkennbar machen und die Wirklichkeit um sich sinnhaft strukturieren und ordnen. (Bergmann 2008a, 51; vgl. auch Streeck 2005, 1417)

Diese alltäglich und unreflektiert eingesetzten Praktiken und Verfahren der Sinngebung bezeichnet Garfinkel als *Ethnomethoden*. Sie gewährleisten das Funktionieren unauffälliger, trivial erscheinender Routineabläufe wie dem Bilden einer Schlange vor der Käsetheke im Supermarkt. Für ethnomethodologische ebenso wie für konversationsanalytische Vorgehensweisen entscheidend ist, dass diese Praktiken bzw. Ethnomethoden von den Beteiligten zwar im Alltag genutzt und gesehen, jedoch unbemerkt bleiben („seen but unnoticed“, Garfinkel 1967, 36). Diese können daher nicht im Rahmen von Befragungen Auskunft über ihre Praktiken geben oder diese gar erklären.

Aus Sicht der Ethnomethodologie wird soziale Realität also durch das alltägliche, geordnete Handeln der Menschen selbst erzeugt. Garfinkel (1967, 10f.) bezeichnet sie deshalb als Vollzugswirklichkeit („achievement“ bzw. „ongoing accomplishment“). Wie aber kann man sich diesen Herstellungsprozess konkret vorstellen? Wirklichkeit wird erzeugt (und ist erfahrbar) nur im Rahmen sozial situierter Handlungs- und Gesprächskontexte. Sprache ist dafür ein fundamentales Mittel. Entscheidend ist nun, dass Interaktanten ihre an sich vagen (indexikalischen) Äußerungen mit Hinweisen darauf versehen, in welchem Kontext sie zu verstehen sind. Beispielsweise enthält die Äußerung „Wären Sie so freundlich, mir zu sagen, wo ich X finde?“ gleich mehrere solcher Hinweise. So gibt die der eigentlichen Frage vorgeschaltete Bitte dem Rezipienten zu verstehen, dass der Fragende ihm a) in Bezug auf den Gegenstand der Frage mehr Wissen zuschreibt als sich selbst; b) im Gewähren der Auskunft eher einen Akt der Hilfsbereitschaft (als dem Befolgen einer anderweitigen Verpflichtung) sieht. Die Wahl des Personalpronomens *Sie* zeigt an, dass der Fragende sich mit seinem Gegenüber in einem förmlichen und nicht vertrauten Verhältnis stehend sieht. Der Fragende dokumentiert also mit seinem Handeln, in welchem sozialen und kommunikativen Kontext seine Äußerung zu verstehen ist. Der Clou der Ethnomethodologie besteht nun darin, zu zeigen, dass eben jene Kontexthinweise den entsprechenden Kontext – die Bearbeitung einer Wissensasymmetrie in einem formellen Rahmen – überhaupt erst konstituieren. Das Verhältnis zwischen sprachlichen Handlungen und Kontext ist also ein *reflexives*: „the activities whereby members produce and manage settings of organized everyday affairs are identical with members’ procedures for making those settings ‚accountable‘ (Garfinkel 1967, 1).

Harvey Sacks (1935-1975) hat nun mit der Konversationsanalyse eine Untersuchungssystematik entwickelt, mit der sich rekonstruieren lässt, wie Menschen Zug um Zug Wirklichkeit im Gespräch hervorbringen. Prägend für deren Entwicklung war neben dem Austausch mit Harold Garfinkel auch der mit Erving

Goffman (Heritage/Clayman 2010, 8; Silverman 1998, 32). Zentrale konversationsanalytische Vorgehensweisen hat Sacks in der zweiten Hälfte der 1960er Jahre bis 1972 in Vorlesungsreihen entwickelt. Die meisten Vorlesungen sind posthum in den *Lectures on Conversation* (Sacks 1995) veröffentlicht worden.

Welches sind nun die konstitutiven Eigenschaften von Gesprächen und was folgt in methodischer Hinsicht aus ihnen? Erstens sind Gespräche im Ganzen wie auch einzelne Gesprächshandlungen – Gesprächseröffnungen, Erzählen, Beheben von Verstehensproblemen – grundlegend durch *Interaktivität* gekennzeichnet. Das Handeln der Gesprächsbeteiligten ist aufeinander bezogen. Dies gilt auch bei scheinbar monologischen Gesprächsformen wie einem Lehrervortrag; auch er ist nicht nur auf die Rezipienten (ihr Vorwissen, ihre Interessen) zugeschnitten (was die Konversationsanalyse mit dem Begriff *recipient design* fasst), sondern kann ohne die Anzeige der Aufmerksamkeit und des (Nicht-)Verstehens durch zumindest einige der Beteiligten nicht gelingen. In der Konsequenz heißt das: Die Konversationsanalyse untersucht Gespräche und einzelne Gesprächshandlungen als interaktiv hergestellte Phänomene.

Der zweite Punkt erscheint auf den ersten Blick trivial: Gespräche sind Prozesse, die in der Zeit stattfinden und eine gewisse Linearisierung erfordern. Diese *Sequenzialität* bedingt nun aber zwei weitere konstitutive Eigenschaften von Gesprächen. Erstes können Gesprächsbeiträge i.W. nur in einem zeitlichen Nacheinander erfolgen; zweitens erfolgen sie in geordneter Weise.

Ad 1: In Alltagsgesprächen steht weder von vornherein fest, welcher der Beteiligten den nächsten Turn übernimmt, noch wann dies geschieht; der *Sprecherwechsel* stellt somit ein organisatorisches Problem für die Beteiligten dar. Den ‚Regelapparat‘ für den Sprecherwechsel haben Sacks, Schegloff und Jefferson (1974) in einem ihrer bekanntesten Aufsätze „A Simplest Systematics for the Organization of Turn-Taking for Conversation“ beschrieben.

Ad 2: In sequenzieller Hinsicht erfolgen Gesprächsbeiträge nicht nur nacheinander; sie sind darüber hinaus auch in bestimmter Weise verkettet und verknüpfen sich erkennbar zu kleineren oder größeren Einheiten. Die *sequenzielle Organisation* solcher Einheiten kommt dabei über sog. konditionelle Relevanzen zustande. Damit sind normative Erwartbarkeiten zwischen bestimmten Typen von Äußerungen bezeichnet. So setzt der Äußerungstyp ‚Gruß‘ als nächste Handlung einen Gegengruß relevant. Solche konditionellen Relevanzen wurden zunächst für sog. Paarsequenzen beispielsweise im Kontext von Gesprächseröffnungen (Schegloff 1972) und -beendigungen (Schegloff/Sacks 1973) beschrieben (für einen Überblick vgl. Schegloff 2007). Insbesondere im deutschsprachigen Raum wurde die Untersuchung unterschiedlicher Sequenztypen für die Analyse kommunikativer Gattungen (Günthner/Knoblach 1994) wie Erzählen (Sacks 1972; Quasthoff 1980), Erklären (Keppler 1989; Morek 2012) und Argumentieren (Spranz-Fogasy 2003; Heller 2012) fruchtbar gemacht. Interaktanten orientieren sich beobachtbar an bestimmten Sequenztypen und reduzieren somit die koordinativen Anforderungen, die sich im Rahmen von Gesprächen ergeben.

Auch die Einsicht in die sequenzielle Organisiertheit von Gesprächen ist analytisch folgenreich: Die Bedeutung und Funktion einer einzelnen Äußerung ist nicht angemessen zu beschreiben, ohne den sequenziellen Kontext, in dem sie steht, zu beachten. Die grundlegende Analyseeinheit kann deshalb nicht eine Einzeläußerung sein, sondern die Handlungssequenz, deren Teil sie ist.

Die beschriebenen Organisationsmechanismen – der Regelapparat für die Organisation des Sprecherwechsels und die verfestigten Sequenztypen – unterstützen den reibungslosen Ablauf von Gesprächen. Dennoch bleiben Letztere störanfällig. Als dritter Organisationsmechanismus wurden in der Konversationsanalyse daher konversationelle Reparaturen untersucht (Schegloff/Jefferson/Sacks 1977, 381; für Unterricht: Macbeth 2004). Bei Formulierungs- und Verstehensproblemen greifen Interaktanten also auf geregelte Verfahren zu ihrer Behebung zurück.

1.2 Methodisches Vorgehen I: Datenerhebung und -aufbereitung

Ton- und/oder Videoaufzeichnungen (zu Videographie s. Maak in diesem Band) bilden die zentrale Datengrundlage der Gesprächsanalyse. Beobachtungsprotokolle, Selbstauskünfte in Befragungen oder Interviews sowie Kodierschemata sind dagegen nicht geeignet, um Einblicke in den tatsächlichen Zug-um-Zug-Aufbau realer Gespräche zu erlangen. Der Mitschnitt authentischer Interaktionssituationen stellt demgegenüber den Versuch dar, Gespräche in ihrer Zeitlichkeit, Interaktivität und Multimodalität möglichst „abbildgetreu“ zu konservieren (Deppermann 2008, 21; vgl. Bergmann 2008b, 531). ‚Authentizität‘ bedeutet hier vor allem, dass die Gespräche nicht eigens für Forschungszwecke arrangiert werden, sondern ‚sowieso‘, d.h. als Teil normaler Abläufe in einem bestimmten Kontext (z.B. Unterricht) stattfinden. Für spezielle Fragestellungen (z.B. bei Querschnittstudien zum Erwerb bestimmter Diskursfähigkeiten) können jedoch auch bestimmte Gesprächsaktivitäten elizitiert werden (z.B. Hausendorf/Quasthoff 1996); dabei ist jedoch die besondere Art der Interaktionssituation in der Analyse zu berücksichtigen.

Essenziell für das Audio-/Videographieren ist eine zuverlässige, gute technische Ausstattung, die eine möglichst hohe Aufnahmequalität garantiert.¹ Sie erleichtert die weiteren Schritte der Datenaufbereitung und -analyse. Wie viel Material aufgezeichnet wird (Anzahl und Dauer der einzelnen Gespräche) ist abhängig von der Forschungsfragestellung (vgl. beispielhaft: Abschnitt 2.2). Dabei sollte im Blick behalten werden, dass gesprächsanalytische Untersuchungen vor allem auf der feinkörnigen Analyse von Einzelfällen basieren und nicht auf großen Stichproben (vgl. Deppermann 2008, 28). Trotzdem ist eine gewisse Datenmenge nö-

¹ Sehr nützliche, ausführliche Hinweise u.a. zu technischen Fragen der Gesprächsdatenerhebung und -aufbereitung finden sich im Gesprächsanalytischen Informationssystem (GAIS) des Instituts für Deutsche Sprache: <http://prowiki.ids-mannheim.de/bin/view/GAIS/AufnahmeIndex>. Auch forschungspraktische und rechtliche Aspekte der Gesprächsaufzeichnung werden auf dieser Plattform eingehend behandelt. Vgl. außerdem Deppermann (2008, 23ff.).

tig: Für gesprächsanalytische Studien gilt i.d.R., dass wesentlich mehr aufgezeichnet wird, als letztlich mikroanalytisch untersucht wird. Erst eine breite(re) Datenbasis ermöglicht es, gezielt Fallvergleiche der interessierenden Phänomene und Praktiken anzustellen und Konstanz und Varianz in vorfindbaren Gesprächsstrukturen aufzuspüren.

Um jedoch den Blick auf mikrostrukturelle Details des sequenziellen, interaktiven Ablaufs von Gesprächen richten zu können, reichen Ton- oder Videoaufnahmen allein nicht aus – dafür sind Gespräche ein zu flüchtiges und komplexes Geschehen. Daher werden analyserelevante Gesprächspassagen *transkribiert*, d.h. nach bestimmten Konventionen verschriftet.

1.3 Exkurs: Transkription

Die gesprächsanalytische Transkription zielt darauf, Merkmale der mündlichen Kommunikation im Medium der Schrift zu bewahren. Es wird versucht, alles was gesagt wird – und *wie* es gesagt wird – möglichst exakt im Transkript abzubilden. Typisch Mündliches, wie z.B. Versprecher, dialektale Aussprache, Sprechpausen oder simultanes ‚Durcheinander‘-Reden werden somit nicht herausgefiltert, sondern mit verschriftet. Auch prosodische Faktoren (z.B. Lautstärke, Intonation) und nonverbale Aktivitäten (z.B. Gesten) finden – je nach Datenlage und Forschungsfragestellung (s. 2.1) – Eingang in die Transkription.

Diese Akribie in der Verschriftung von Gesprächen entspringt einer zentralen methodischen Prämisse der KA, der “order at all points”-Maxime (Sacks 1995, 484): Sie geht davon aus, dass jedes noch so kleine Detail in einem Gespräch – sei es ein kurzes Stocken, ein Räuspern oder eine besondere Betonung – nicht zufällig oder willkürlich ist, sondern einer Geordnetheit und Funktionalität unterliegt. Daher gilt es, solche Details nicht aus der Untersuchung auszuschließen, sondern sie als potenziell relevante Phänomene zu berücksichtigen. Transkripte erlauben es also, Gesprächsverläufe mit ihren klein(st)en Details gleichsam ‚in Zeitlupe‘ zu betrachten, mehrere Datensegmente unmittelbar miteinander zu vergleichen und analytische Beobachtungen ganz exakt an bestimmte Stellen bzw. Äußerungen oder Äußerungsteile im Gespräch zu binden. Trotzdem dürfen Transkripte nicht als 1:1-Abbilder realer Gespräche missverstanden werden (Ayass 2015; Deppermann 2008, 41): Der Schritt des Transkribierens stellt seinerseits eine Selektion und Abstraktion von der zugrunde liegenden Interaktion (bzw. deren wiederum u.U. selektiver Aufzeichnung) dar. Daher werden bei der Analyse zusätzlich zum Transkript stets die entsprechenden video-/audiographierten Daten hinzugezogen, um einen möglichst holistischen Zugang zum entsprechenden Gesprächsausschnitt zu erlangen.

Wie detailliert transkribiert wird, hängt in erster Linie vom jeweiligen Untersuchungsinteresse ab: Ist man beispielsweise an performativischen Aspekten schülerseitigen Vortragens interessiert, also etwa an Faktoren wie Lautstärke, Intonation und Akzentsetzung, so ist eine Transkription nötig, die die Ebene der Prosodie sehr feinkörnig berücksichtigt. Geht es dagegen aus stärker inhaltlicher Perspektive um die Wahl verschiedener Vertextungsverfahren beim schülerseitigen

Erklären (wie z.B. im Projekt InterPass, vgl. Kapitel 2), so ist eine Transkription (vorerst) ausreichend, die im Bereich der Prosodie nur minimale Angaben macht.

Im deutschsprachigen Raum haben sich in der linguistischen Gesprächsforschung vor allem zwei Transkriptionsverfahren etabliert: GAT (Gesprächsanalytisches Transkriptionssystem, Selting et al. 2009) und HIAT (Halbinterpretative Arbeitstranskription, Ehlich/Rehbein 1976). Im Folgenden konzentrieren wir uns auf die Transkription nach GAT, da sie in genuin gesprächsanalytischen Arbeiten bevorzugt wird, besonders leicht in Textverarbeitungsprogramme zu integrieren ist und den schrittweisen Ausbau des Detailliertheitsgrades von Transkripten ermöglicht (s.u.). Zudem ist GAT an die international gängigen Transkriptionskonventionen der Conversation Analysis angelehnt (Jefferson 2004) und liegt inzwischen auch in englischsprachiger Publikation vor (Selting et al. 2011). Ziele, Grundprinzipien und Notationsregeln von GAT sind in Selting et al. 2009 ausführlich und anhand verschiedener Beispieltranskripte beschrieben.² Die wichtigsten Konventionen der Basistranskription sind im Anhang aufgeführt.

Grundsätzlich erfolgt die Transkription bei GAT in literarischer Umschrift, d.h. die Wiedergabe des Gesprochenen orientiert sich an der Orthographie, allerdings in konsequent durchgängiger Kleinschreibung (Ausnahme: Akzente, s.u.). Abweichungen von der Standardausssprache werden entsprechend wiedergegeben (z.B. *hamwer* für 'haben wir'). Die Beiträge der einzelnen Gesprächsbeteiligten werden jeweils in Zeilenschreibweise untereinander notiert, sodass der Gesprächsverlauf Zeile für Zeile von oben nach unten geschrieben bzw. gelesen wird. Zeilen werden nummeriert, um eine exakte Bezugnahme auf einzelne Äußerungen zu ermöglichen. Längere Beiträge eines Sprechers werden gemäß ihrer Intonationsphrasen (Selting et al. 2009) auf mehrere Zeilen gesplittet. Folgender kurzer Transkriptausschnitt (vgl. ausführlich zu der entsprechenden Sequenz: Abschnitt 2.3) zeigt diese und einige andere der grundlegendsten Eigenschaften von GAT-Transkripten:

Beispiel-Transkript (Basistranskript), Leh: Lehrerin, Nah: Nahema

```
028 Nah    also ein brief (.) erklärt was (-) pasSIERT ist,
029        also was (die überschrift ist) zum BEIspiel- (.)
030        was ich gestern erLEEBT hab;
031 Leh    [hm_hm;      ]
032 Nah    [der erzählt] dann von GESTern;
```

Der Sprecherbeitrag („Turn“) des Schülers Nahema (Z. 28-32) besteht aus insgesamt vier Intonationsphrasen, die jeweils nacheinander geäußert werden. So lange Nahema seinen Turn fortsetzt, ohne dass sich ein anderer Interaktant einschaltet, braucht seine Sprechersigle (Nah) nicht ständig neu vor die entsprechende Äußerung geschrieben zu werden.

² Eine (leider nur in Teilen komplettiert vorliegende) digitale Einführung in die Transkription nach GAT (“GAT-TO”) mit einigen praktischen Übungen zur Verschriftung prosodischer Phänomene findet sich unter: <http://paul.igl.uni-freiburg.de/gat-to/>

Die eckigen Klammern in den Zeilen 31 und 32 signalisieren, dass die entsprechenden Äußerungen bzw. Äußerungsteile von Nahema und der Lehrperson zeitgleich gesprochen werden; die öffnende Klammer markiert den Beginn der Simultanpassage und die geschlossene deren Ende. Das *hm hm* der Lehrperson wird also unmittelbar *nach* Nahemas „gestern erLEBT hab;“ (Z. 30) gesprochen und parallel zu Nahemas „der erzählt“ (Z. 32). Sog. Rezeptionssignale wie *hm* (oder auch redupliziert: *hm hm*), *ja*, *nee* o.ä. gehören zu den typischerweise aus der Zuhörerrolle geäußerten Mitteln. Sie werden, so kurz und/oder unauffällig sie auch sein mögen, mitnotiert, da sie wichtige interaktive Funktionen erfüllen (z.B. Anzeigen von Aufmerksamkeit oder Verständnis; weiteres Zugeständnis des Rederechts an das Gegenüber). Zudem zeigt der Ausschnitt exemplarisch, dass auch Pausen und ihre Länge in GAT-Transkripten notiert werden, da sie ein wichtiger Parameter bei der zeitlichen Gestaltung von Interaktionen sind. Beispielsweise kann eine Pause nach dem Stellen einer Frage (z.B. „Kommst du heute Abend mit?“) ein Zeichen dafür sein, dass der Angesprochene eine strukturell „dispräferierte“ Antwort (Pomerantz/Heritage 2013; Levinson 1990, 305ff.) liefern wird (z.B. „Naja, also ich muss mal schauen.“). Pausen werden in runden Klammern verschriftet, z.B. (.) für Mikropausen, (-) für kurze Pausen von ca. 0.2 bis 0.5 Sekunden Dauer. Im obigen Ausschnitt z.B. markieren die drei Pausen in Nahemas Turn (Z. 28, 29) in Kombination mit unvervollständigten bzw. projizierten syntaktischen Konstruktionen den noch folgenden Ausbau und den Beibehalt des Rederechts.

Bei GAT sind grundsätzlich drei Detailliertheitsstufen der Transkription möglich: Minimaltranskript (nur als Arbeitstranskript zu verwenden), Basistranskript und Feintranskript. Das Basistranskript unterscheidet sich vom Minimaltranskript vor allem dadurch, dass Fokusakzente („Hauptbetonungen in einer Äußerung“) verschriftet werden sowie Tonhöhenbewegungen am Ende von Intonationsphrasen. In Nahemas Äußerung „was ich gestern erLEBT hab;“ (Z. 30) zum Beispiel ist *LEBT* die prosodisch am stärksten hervorgehobene Silbe, die informationsstrukturell in den Vordergrund tritt (sprich: Nahema betont, dass ein Brief wiedergibt, was man erLEBT hat.). ‚Unnatürlich‘ starke Akzentuierungen hingegen werden mit zusätzlichen Ausrufezeichen notiert: !ECHT! Auch alle anderen Interpunktionszeichen haben in GAT nicht ihre syntaktische Funktion, sondern werden anderweitig genutzt, nämlich zur Wiedergabe finaler Tonhöhenbewegungen. Zeile 28 endet beispielsweise mit leicht steigender Intonation, angezeigt durch ein Komma (weitere Intonationsmuster s. Anhang).

In der höchsten Detailliertheitsstufe, den Feintranskripten, können zusätzlich sehr feine sprecherische Markierungen verzeichnet werden wie z.B. Tonhöhen-sprünge, Veränderungen des Tonhöhenregisters, Lautstärke, Sprechgeschwindigkeit und Stimmqualität.

Grundlegend für GAT und einer seiner großen Vorteile ist das „Zwiebelprinzip“ (Selting et al. 2009, 356): Ein Transkript einer bestimmten Detailliertheitsstufe kann jederzeit, so es der Analyseprozess erfordert, zu einer differenzierteren Version ausgebaut bzw. verfeinert werden. Dies kann z.B. dann relevant werden,

wenn sich im Zuge der Analysen offenbart, dass bislang unberücksichtigte sprachlich-interaktive Details integraler Bestandteil der fokussierten Gesprächspraktiken bzw. -mechanismen sind. Insofern kommt GAT der Verwobenheit von Datenaufbereitung, Gegenstandskonstitution und Analyse entgegen (vgl. Abschnitt 1.4).

Aufgrund seines gängigen Zeicheninventars und seines einfach formatierten Layouts lässt sich GAT problemlos in gängigen Textverarbeitungsprogrammen nutzen. Komfortabler und unter forschungspraktischen Gesichtspunkten i.d.R. sinnvoller ist jedoch die Nutzung von Transkriptionseditoren. Der Hauptvorteil solcher Software besteht darin, dass sie Abspiel- und Verschriftungsfenster integriert, die Verknüpfung von Transkript und Aufnahme erlaubt, Zeitangaben (z.B. auch: Pausenmessung) einfügen lässt sowie über verschiedene Abspielmodi verfügt (z.B. langsamere Abspielgeschwindigkeit). Zu den gängigen Programmen, die auf diese Weise die Transkription unterstützen und vereinfachen, gehören: *FOLKER*, *f4*, *Elan*, *Transana*.³ Besonders nützlich ist dabei das Setzen sog. Zeitmarken im Transkript, die das sofortige Auffinden der entsprechenden Stelle in der zugehörigen Video-/Audiodatei erlauben. Auch vereinfachen Transkriptionseditoren die Arbeit mit multimodalen Daten und Transkripten. In diesem Zusammenhang sind auch spezifische theoretische und methodische Überlegungen angestellt worden zu deren Transkription (vgl. Stukenbrock 2009; Heath/Hindmarsh/Luff 2010).

1.4 Methodisches Vorgehen II: Analyse

Dafür, wie Konversationsanalytiker bei der Auseinandersetzung mit den Daten vorgehen, lässt sich kein ‚Rezept‘ in Form einer strikten Abarbeitung vorgegebener methodischer Schritte angeben. Kennzeichnend für konversationsanalytische Arbeiten ist vielmehr eine spezifische Analysehaltung:

[C]onversation analysts employ a wide range of essentially interpretative skills in their research. The conversation analytic mentality involves more a cast of mind, or a way of seeing, than a static and prescriptive set of instructions which analysts bring to bear on the data. (Hutchby/Wooffitt 2008, 89)

Nichtsdestotrotz lassen sich – basierend auf den oben skizzierten konstitutiven Eigenschaften von Gesprächen – grundlegende Prinzipien konversationsanalytischer Arbeitsweise formulieren. Sie ‚methodisieren‘ die interpretative und teils auf eigene kommunikative Intuitionen zurückgreifende Tätigkeit des Forschenden (Bergmann 2008b, 532).

Am Beginn konversationsanalytischer Arbeit steht die Identifikation eines analytisch interessanten Phänomens und dessen detaillierte Analyse anhand eines Einzelfalls, der dieses Phänomen bestmöglich repräsentiert. Das Phänomen kann dabei unterschiedlich ‚groß‘ sein (z.B. ein bestimmtes Rezeptionssignal oder eine

³ F4: <https://www.audiotranskription.de/f4.htm>; FOLKER: <http://agd.ids-mannheim.de/folker.shtml>; ELAN: <https://tla.mpi.nl/tools/tla-tools/elan/>; Transana: <http://www.transana.org/> (vgl. Schwab 2006).

größere kommunikative Aktivität wie z.B. Bearbeitung von Wissensasymmetrien). Der Analyseausschnitt ist dabei so zu wählen, dass er eine thematisch bzw. handlungslogisch abgeschlossene Einheit darstellt, deren Grenzen von den Interaktanten durch einen eindeutig erkennbaren Anfang und Abschluss markiert werden (Deppermann 2008). Erst ein solcher Radius ermöglicht es, den *sequenziellen Kontext*, d.h. die umgebenden Gesprächszüge sinnvoll zu berücksichtigen und das interessierende Phänomen in seinem konversationellen Zusammenhang zu betrachten. Hierhin besteht ein deutlicher Unterschied zu kodierenden Verfahren, die Einzeläußerungen kontextisoliert typisieren (vgl. Heller 2014, 130).

In der Konversationsanalyse wird also *sequenzanalytisch* vorgegangen. Das heißt: Die Gesprächssequenz wird Zug-um-Zug in ihrem interaktiven Zustandekommen nachvollzogen – mit Blick auf sequenzielle Folgerwartungen (vgl. Abschnitt 1.1) und deren Interpretation und Behandlung durch die Beteiligten. Der Blick richtet sich dabei auf sicht- und hörbare Formen der Interaktionsoberfläche und deren *Funktionen*. Äußerungen werden in ihrer spezifischen Beschaffenheit im Hinblick auf die Frage betrachtet, welches kommunikative Problem sie auf methodische Weise lösen. Eine Äußerung wie *Mir ist gestern was Krasses passiert!* etwa bearbeitet das ‚Problem‘, sich in einem Gespräch für einen längeren Turn, nämlich eine Erzählung, das Rederecht und Interesse der Gesprächspartner zu sichern. Es geht also nicht um mentalistische Deutungen von Sprecherabsichten o.ä., sondern um die Frage „Why that now?“ (Schegloff/Sacks 1973, 76):

Niemals wird gefragt: ‚Was will der Sprecher sagen?‘, sondern immer: ‚Welches sind die Funktionen und Leistungen, die dieses Phänomen in seiner konkreten sprachlichen Form für die aktuelle Interaktion erfüllt bzw. erbringt?‘ (Hausendorf/Quasthoff 1996, 113)

Was eine Äußerung tatsächlich „tut“, ist nicht intuitiv erschließbar, sondern wird nur an ihren sequenziellen Folgen, d.h. an der Folgeäußerung des Gegenübers, ersichtlich: Wie wird eine Äußerung vom Gegenüber verstanden, interpretiert und behandelt? Nicht nur die Beteiligten selbst, sondern auch die Analysierenden können über dieses Verfahren der „next turn proof procedure“ (Sacks/Schegloff/Jefferson 1974, 729) Aufschluss über die faktische Bedeutung einer Äußerung für die Gesprächsbeteiligten gewinnen.

Wenn die sequenzielle Organisation und deren zugrunde liegende Ordnung an einem Einzelfall rekonstruiert ist, geht es in einem zweiten Schritt darum, ähnlich gelagerte Fälle aus den Daten zu sammeln. Es werden *Kollektionen* (Hutchby/Wooffitt 2008, 88ff.) gebildet, um das in Frage stehende Phänomen fallvergleichend und auf größerer Datenbasis systematisch untersuchen zu können.

Um solche Vergleichsfälle auswählen zu können, ist eine intensive Kenntnis des erhobenen Gesprächsmaterials notwendig. Für einen ersten Überblick bietet sich die Erstellung grober Ablaufbeschreibungen an („Gesprächsinventare“, Deppermann 2008). Die zielgerichtete Auswahl analyserelevanter Sequenzen kann auf Basis dieser Inventarisierung und des wiederholten Sichtens ‚verdächtiger‘ Stellen in den Aufnahmen vorgenommen werden. Der Fallvergleich ermöglicht es,

sowohl kontextübergreifend gültige Strukturmuster zu identifizieren als auch kontextspezifische Realisierungen herauszuarbeiten. Besonders aufschlussreich für die Formulierung übergreifender sequenzieller Prinzipien können auch „abweichende Fälle“ („deviant cases“, Schegloff 1972) sein, d.h. Fälle, die zunächst von der bislang rekonstruierten Musterhaftigkeit abzuweichen scheinen. Lässt sich zeigen, dass die Gesprächsteilnehmenden diese selbst ebenfalls als Abweichungen von gesprächsbezogenen, normativen Erwartbarkeiten behandeln (z.B. durch einen *account* beim Nicht-Beantworten einer Frage wie z.B. *Es tut mir Leid, ich weiß es nicht.*), liefert dies besonders verlässliche Evidenz für die Geltung der rekonstruierten musterhaften Strukturen (z.B. ‚eine Frage macht eine Antwort sequenziell erwartbar‘).

Beim fallvergleichenden Vorgehen werden so lange neue Fälle hinzugezogen, bis eine ‚Sättigung‘ eintritt (Heller 2014, 132), d.h. bis die jeweils situierten Realisierungen eines Phänomens sich den herausgearbeiteten Methoden zur Gesprächsorganisation zuordnen lassen: Sie sind dann nur noch kontextspezifische Realisierungen *kontextfreier* Ressourcen (Sacks/Schegloff/Jefferson 1974) und als solche das angestrebte Untersuchungsziel konversationsanalytischer Arbeiten.

Das hier skizzierte methodische Vorgehen – von der detaillierten Analyse eines Einzelfalls hin zur Extraktion kontextübergreifender formaler Interaktionsmechanismen über den Fallvergleich – bringt folgende zwei Konsequenzen für den forschungspraktischen Prozess mit sich: Erstens sind Gegenstandskonstitution (z.B. Sequenzauswahl), Datenaufbereitung (Transkription) und Analyse keine trennbaren, einander nachgelagerten Etappen, sondern greifen im Forschungsverlauf immer wieder ineinander. Zweitens besteht letztlich, wenn es um die Präsentation von Untersuchungsergebnissen geht, eine Diskrepanz zwischen erfolgter Analyse und (aus)gewählter Darstellung. Das, was – etwa in einer Publikation – exemplarisch anhand ausgewählter Sequenzen vorgeführt wird, macht i.d.R. nur einen Bruchteil des gesamten Materials aus, für den es stellvertretend steht. Eine Ausnahme bilden Einzelfallanalysen, die ihren Wert aus der Identifikation eines Interaktionsphänomens und der diesbezüglichen Hypothesengenerierung allein beziehen (vgl. Deppermann 2008).

1.5 Gütekriterien des Verfahrens

Die klassischen Gütekriterien der quantitativen Forschung (Objektivität, Reliabilität, Validität) sind auf gesprächsanalytische Untersuchungen nicht anwendbar, da es sich bei diesen nicht um ein Berechnungs- und/oder Mess-, sondern ein interpretatives Verfahren handelt. Allerdings bringt es das strikt rekonstruktive Vorgehen der ethnomethodologischen Gesprächsanalyse mit sich, dass Verfahren der Validierung gleichsam *in* die Analyse integriert sind: Der Nachweis der Gültigkeit von Aussagen über bestimmte Gesprächspraktiken und -mechanismen geschieht somit durch die oben skizzierten methodischen Verfahren (Bergmann 2008b, 534): a) Fallvergleich zu funktional gleichartigen Phänomenen in den erhobenen Daten, b) Analyse ‚abweichender Fälle‘, c) *next-turn-proof-procedure*. Die Konversationsanalyse erlegt sich somit selbst strenge Ansprüche auf, was den

Nachweis betrifft, dass die in Frage stehenden Praktiken nicht forschenseitig ‚erfunden‘ sind, sondern tatsächlich für Gesellschaftsmitglieder eine Ressource ihres Handelns darstellen.

Schließlich sollte stets die Möglichkeit kollegialer Datensitzungen unbedingt genutzt werden (Schwarze 2014). In der gemeinsamen Auseinandersetzung mit dem Datenmaterial können (subjektive) Interpretationen abgeglichen, relativiert und korrigiert werden. Das Resultat sind intersubjektiv überprüfte und überprüfbare Analyseergebnisse.

1.6 Literaturempfehlungen

Deppermann (2008): Gespräche analysieren.

Konzise Einführung in gesprächsanalytische Theorie und ihr praktisches Vorgehen.

Hutchby/Wooffitt (2008): Conversation analysis.

Leitfaden zum konversationsanalytischen Arbeiten, demonstriert an Transkript-Beispielen.

Sidnell/Stivers (2013): The handbook of conversation analysis.

Breiter Überblick über klassische und aktuelle Untersuchungsfelder der Konversationsanalyse.

Hausendorf/Quasthoff (2005): Konversations-/Diskursanalyse: (Sprach-)Entwicklung durch Interaktion.

Diskussion der Nutzungsmöglichkeiten der KA für erwerbsbezogene Fragen.

Becker-Mrotzek/Vogt (2009): Unterrichtskommunikation.

Überblick über verschiedene Ansätze und Befunde der linguistischen Unterrichtsforschung.

2. Gesprächsanalytische Zugänge in der sprachdidaktischen Forschung: Das Beispiel InterPass

2.1 Forschungsfrage und Nutzung der GA

Im Folgenden werden wir gemäß der Ausrichtung dieses Bandes an einem konkreten Forschungsprojekt aufzeigen, wie der gesprächsanalytische Zugang für die empirische Bearbeitung sprachdidaktischer Fragestellungen nutzbar gemacht werden kann. Dazu wählen wir das interdisziplinäre Projekt InterPass.⁴ Die Darstellung konzentriert sich auf die gesprächsanalytischen Teile des Projekts und muss zwangsläufig an all denjenigen Stellen knapper ausfallen, an denen weitergehende theoretische und methodische Aspekte (z.B. zum Diskurserwerb) betroffen sind.

Ausgehend von Befunden internationaler Schulleistungsstudien zu der nach wie vor engen Kopplung von sozialer Herkunft und Bildungserfolg untersucht Interpass die subtile Rolle diskursiver und fachkultureller Praktiken im Unterricht. Denn: Bisherige Befunde der soziolinguistischen Erwerbsforschung belegen, dass Diskurspraktiken in den Familien von Schülerinnen und Schülern in sehr unterschiedlicher Weise realisiert und sozialisatorisch angeeignet werden. Gleichzeitig sind Diskurspraktiken (z.B. Erklären, Argumentieren) als zunehmend komplexe sprachlich-kommunikative Aktivitäts- und Anforderungsfelder eines jeden Fachunterrichts mit verschiedenen fachkulturellen Ausprägungen zu betrachten.

Im linguistischen Strang von Interpass stellt die Gesprächsanalyse den zentralen Zugang für die Bearbeitung zweier Fragekomplexe:

- *Klassenbezogene Partizipationsgelegenheiten*: Welche diskursiven (und fachlichen) Partizipationsgelegenheiten an explanativen und argumentativen Diskursaktivitäten schaffen Lehrkräfte (,Arbeitsteilung‘ zwischen Lehrenden und Lernenden beim Erklären und Argumentieren)? Lassen sich diesbezüglich Unterschiede zwischen Lehrkräften beobachten?
- *Individuelle Lerngelegenheiten* für ausgewählte Fokuskinder, die sich hinsichtlich ihrer sprachlichen und sozialen Ausgangsbedingungen unterscheiden: a) Orientieren sich die an das jeweilige Kind gestellten diskursiven (und fachlichen) Anforderungen an seiner gezeigten Kompetenz und zielen Angebote interaktiver Unterstützung auf den Ausbau der Kompetenz ab? b) Erweist sich die interaktive Unterstützung im Längsschnitt für die Partizipation und den Ausbau von Diskurskompetenz als wirksam?

⁴ Das Projekt „Interaktive Verfahren der Etablierung von Passungen und Divergenzen für sprachliche und fachkulturelle Praktiken im Deutsch- und Mathematikunterricht“ wird vom BMBF (01JC1112) gefördert (Leitung: Uta M. Quasthoff und Susanne Prediger). Als interdisziplinäres Projekt bearbeitet es sprachwissenschaftliche sowie mathematik- und sprachdidaktische Fragestellungen im Rahmen eines triangulativen Forschungsdesigns (Unterrichtsinteraktion und Gruppendiskussionen).

2.2 Datenerhebung und -aufbereitung

Zur Bearbeitung der genannten Fragestellungen arbeitet InterPass mit einem umfangreichen Korpus videographierter Unterrichtsinteraktionen, das in fünf Gesamtschul- und Gymnasialklassen aus privilegierten und benachteiligten Milieus aufgezeichnet wurde. Es wurden die ersten acht sowie nach einem halben Jahr vier bis sechs weitere Mathematik- und Deutschstunden des fünften Jahrgangs videographiert (insgesamt über 120 Stunden). Der Aufnahmezeitpunkt zu Beginn des fünften Schuljahres wurde gewählt, weil gerade beim Übergang an die neue Schulform die jeweilige Mikrokultur des Fachunterrichts konstituiert wird.

Um nicht nur den groben Verlauf von Unterricht, sondern Interaktionen der Lehrkraft mit unterschiedlichen Schülerinnen und Schülern zu dokumentieren, wurden vier Kameras eingesetzt. Eine fixe Kamera zeichnete das Geschehen im ganzen Raum aus der Totale auf. Eine zweite mobile Kamera dokumentierte die Lehrperson sowohl im Klassengespräch als auch in Interaktionen mit einzelnen Kindern. Um sprachliche Äußerungen der Lehrperson unabhängig von ihrer Bewegung im Raum zu erfassen, trug diese ein Funkmikrofon. Zwei weitere Kameras fokussierten auf ausgewählte Fokuskinder, wobei zur Ergänzung Audioaufnahmegeräte (auf den Tischen) eingesetzt wurden. Zusätzlich zur videographischen Dokumentation wurden der Stundenverlauf tabellarisch protokolliert und verwendete Lehrmaterialien und Schüleraufzeichnungen erfasst.

Die Datenaufbereitung erfolgte in *Transana* (s. Exkurs Abschnitt 1.3). Die Aufzeichnungen der Lehrerkamera wurden genutzt, um sämtliche Unterrichtsstunden im Hinblick auf Unterrichtsthemen, Sozialformen und Diskursaktivitäten zu annotieren. Dies ermöglichte einen ersten groben Überblick über die in den einzelnen Fächern präferierten diskursiven Gattungen. Auf dieser Grundlage wurden Sequenzen für die Transkription ausgewählt. Zur Erstellung von Kollektionen für ausgewählte Fokuskinder wurden anhand der Aufzeichnungen der Fokuskinder-Kameras zunächst alle für die Unterrichtspartizipation relevanten Aktivitäten dieser Kinder inventarisiert. In einem zweiten Schritt wurden ihre Diskursaktivitäten entsprechend der Konventionen des GAT 2 (s. Exkurs Abschnitt 1.3) transkribiert. Grundsätzlich wurden zunächst Basistranskripte erstellt, die nach Bedarf um ausgewählte Konventionen der Feintranskription erweitert wurden. Das Gesamtkorpus umfasst rund 420 Transkripte unterschiedlicher Länge.

2.3 Exemplarische Analyse

Die folgende Sequenz zeigt eine Erkläraktivität aus dem Deutschunterricht einer fünften Klasse. Erklärungen stellen größere Einheiten in Gesprächen dar, die ein- und ausleitend als Aktivität eines bestimmten Typs erkennbar gemacht werden (Morek 2012). Wie genau sich die Beteiligten dies wechselseitig anzeigen, wird im Folgenden exemplarisch vorgeführt. Wir machen die sequenzielle Organisation der Erkläraktivität sichtbar, indem wir *gesprächsstrukturelle Aufgaben* rekonstruieren, an denen sich die Beteiligten selbst beobachtbar orientieren. Mit der Analyse der – wie sich in vergleichenden Untersuchungen gezeigt hat (Hausen-

dorf/Quasthoff 1996; Morek 2012; Heller 2012) – kontextinsensitiven gesprächsstrukturellen Aufgaben gewinnen wir ein Instrument, mit dem sich Erkläraktivitäten auch kontextvergleichend und in erwerbsbezogener Perspektive untersuchen lassen.

T-01-1-DU-09-02400608-Geschäftsbrief2 (Leh: Lehrerin, Kev: Kevin, Nah: Nahema, Tha: Thasin)⁵

001 Leh zuERST;
 002 damit wir alle aber wieder DRIN sind; (.)
 003 WAS war das für ein brief (.) den ich euch gezeigt hab;
 004 SuS ((melden sich, auch Kev))
 005 KEvin; (.)
 006 Kev ein geSCHÄFTlichen brief, (.)
 007 Leh ein geschäftlicher brief war das; (=)
 008 geNAU; (.)
 009 SO; (.)
 010 was unterscheidet jetzt den geschäftlichen brief (.)
 011 von (.) dem persÖNlichen; (.)
 012 wer kann das nochmal Sagen; (.)
 013 SuS ((melden sich, inkl. Nah, (2.5)))
 014 Leh wer hat denn heute den SAÜbersten finger-
 015 ((lacht))
 016 ich mach nur SPASS; (.)
 017 nahEMA;
 018 (2.6)
 019 Leh <<☺> es war nur ein SCHERZ;> (1.5)
 020 Nah das war um !BÜ!cher (-) also BÜcher zu kaufen; (-)
 021 Leh JA:, (.)
 022 der handelte davon bücher zu KAUFen;=ja:, (-)
 023 das heißt kannst du den unterschied jetzt beNENnen, (--)
 024 zwischen einem |normalen |
 | |((führt Hände nach re)) |
 persönlichen brief und dem |bücher |
 | |((führt Hände nach li)) |
 |KAUF |(.).brief, (-)
 |((macht air quotes)) |
 025 |angebots[BRIEF,]
 026 Nah [ja:;]
 027 ein BRIEF ist (-) zum;
 028 also ein brief (.) erklärt was (-) passIERT ist,
 029 also was (die überschrift ist) zum beispiel- (.)
 030 was ich gestern erLEBT hab;
 031 Leh [hm_hm;]
 032 Nah [der erzählt] dann von GESTern;
 033 u:nd (.) ein geschäftlicher brief is:t (.) geht um
 KAUFen; (-)
 034 Leh hm:: nicht IMmer;=
 035 =da hätt ich jetzt n EINSpruch; (.)
 ((...))
 078 Tha bei einem geschäftlichen brief lässt man das persÖNliche
 sein;
 079 Leh DANKeschön; (-)
 080 geNAU. (-)

⁵ Die Transkriptionskonventionen sind im Anhang aufgeführt. Die vertikalen Striche (|) kennzeichnen simultane verbale und nonverbale Handlungen.

Die Lehrerin leitet mit dem Temporaladverb „zuERST“ den Übergang von der Klärung organisatorischer Fragen zum thematischen Unterrichtsgespräch ein. Mit ihrer Was-Frage (Z. 3) etabliert sie einen lokalen Zugzwang, der als nächste Handlung das Nennen des Gegenstands der vorigen Stunde relevant setzt. Nach schülerseitigen Turnbewerbungen erhält Kevin das Rederecht und bedient den Zugzwang mit der lokalen Antwort: „ein geSCHÄFTlicher brief,“. Die steigende Tonhöhenbewegung am Äußerungsende versteht die Aussage mit einer Versuchsmarkierung („try-marker“, Sacks/Schegloff 1979, 18) und macht somit eine Bestätigung erwartbar. Die Lehrerin initiiert zunächst eine Korrektur (Macbeth 2004) der Kasusmarkierung (Z. 7) und ratifiziert dann die Richtigkeit der Aussage. Über die *known answer question* (Mehan 1979), die das unterrichtstypische IRE-Schema (Initiation, Reply, Evaluation, vgl. ebd.) in Gang setzt, haben die Beteiligten gemeinsam einen Anschluss an die vorige Stunde hergestellt und einen bestimmten thematischen Gegenstand – „Geschäftsbriefe“ – relevant gesetzt. Aus gesprächsstruktureller Perspektive haben sie damit die für die Initiierung einer Erklärung notwendige, jedoch nicht hinreichende Aufgabe der *Etablierung von Inhaltsrelevanz* (Morek 2012) bearbeitet. Letztere bildet nicht zwangsläufig den Auftakt zum Erklären, sondern kann auch als Ausgangspunkt für andere Aktivitäten genutzt werden. Eine Erklärung wird erst erwartbar, wenn auch ein Erklärbedarf etabliert wird.

Den Einstieg in die Bearbeitung dieser Aufgabe projiziert die Lehrerin mit dem Diskursmarker „so“ (Schiffrin 1987). Mit ihrer Was-Frage⁶ (Z. 10) bewerkstelligt sie eine zweite für die Hervorbringung einer Erklärung wesentliche Aufgabe, das *Konstituieren eines Explanandums* (Morek 2012). Sie etabliert also einen Erklärgegenstand, der in diesem Fall die Unterscheidung von persönlichen und geschäftlichen Briefen betrifft. Die Frage setzt einen Zugzwang für einen bestimmten Typ nächster Handlung, nämlich eine *Erklärung*. Da die Bearbeitung eines Erklärbedarfs i.d.R. einen übersatzmäßigen Beitrag, d.h. eine explanative Diskurseinheit, verlangt, ist der Zugzwang globaler Natur. Dies impliziert auf gesprächsorganisatorischer Ebene, dass der nächste Sprecher für die Dauer der Diskurseinheit die Rolle des primären Sprechers übernimmt, während sich der Rezipient auf die Anzeige von Aufmerksamkeit und (Nicht-)Verstehen beschränkt.

Die Art, in der hier ein Erklärbedarf etabliert wird, ist für den institutionellen Kontext Unterricht spezifisch. Die Frage der Lehrerin „wer kann das nochmal SAgen;“ fordert nicht nur explizit zu Turnbewerbungen auf, sondern signalisiert durch das Adverb „nochmal“, dass die Erklärung nicht der Behebung eines Wissensdefizits der Fragenden, sondern vielmehr der Wissensdemonstration der Lernenden dient. Der Einschub in Zeile 12 schließlich („das müsste mir jetzt !JE!der sagen können::;“) kontextualisiert die Aktivität unmissverständlich als Wiederholung und verdeutlicht die Erwartung, dass das in Frage stehende Wissen verfügbar

⁶ Entgegen landläufiger Annahmen sind Fragen nicht immer mit einer steigenden Tonhöhenbewegung am Ende der Intonationsphrase versehen (Couper-Kuhlen 2012).

sein sollte. Die lange Pause (Z. 13) sowie die metadiskursive und spaßhafte Thematisierung der Sprecherauswahl⁷ (Z. 14) schaffen eine größere „offene Position“ (Mazeland 1983, 83) für Turnbewerbungen auch zögerlicher Schülerinnen und Schüler – und ermöglichen es wiederum der Lehrerin, das Verfügen von Wissen bereits anhand von Meldungen zu überblicken. Nahema wird auf seine Turnbewerbung hin das Rederecht erteilt. Seine ausbleibende Übernahme des Turns (Z. 18) interpretiert die Lehrerin als Verunsicherung bzw. Nichtverstehen ihrer scherzhaft gemeinten Frage: Mit *smile voice* (Selting et al. 2009, 306) gibt sie rückwirkend einen weiteren metadiskursiven Kommentar („es war nur ein SCHERZ;“), der Nahema zur Übernahme des Turns ermutigt.

Nahema bedient nun den globalen Zugzwang weder global, noch mit einer Erklärung. Zwar nennt er den Inhalt des in der vorigen Stunde von der Lehrerin gezeigten Geschäftsbriefes, doch expliziert er an keiner Stelle einen Unterschied. Nach einer Reformulierung (Z. 22) re-etabliert die Lehrerin daher den globalen Zugzwang (Z. 23). Dass die Erklärung einen Vergleich der Briefsorten erfordert, wird von ihr dabei nicht nur verbal expliziert, sondern auch gestisch veranschaulicht: Simultan zum verbalen Äußerungsteil „normalen“ führt sie ihre Hände nach rechts, um sie dann simultan zu „bücher“ links von ihrem Körper zu positionieren. Sie gibt damit eine visuelle Verstehenshilfe für die Implikationen des globalen Zugzwangs. Im Rahmen dieser Explikation greift sie das von Nahema ins Spiel gebrachte „kaufen“ mit dem Kompositum „bücher(.)kauf(.)brief“ auf; zugleich distanziert sie sich durch *air quotes* von dieser Bezeichnung und führt mit der Reformulierung „angebotsBRIEF“ in Zeile 25 einen angemessenen Begriff ein.

Überlappend mit der Lehrerin signalisiert Nahema zunächst sein Verstehen des Zugzwangs (Z. 26) und produziert dann eine explanative Diskurseinheit (Z. 27-33), in der er jeweils eine Eigenschaft für jede der beiden Briefsorten darstellt („erklärt was pasSIERT ist“ und „der erzählt dann von GESTern;“ vs. „geht um KAUFen;“). In gesprächsstruktureller Hinsicht arbeitet er also an der *Durchführung der eigentlichen Erklärung* (Morek 2012), d.h. an der sprachlichen Explizierung von Informationen und Zusammenhängen, die für die Verständigung bzw. die Wissensdemonstration als notwendig erachtet werden. Die Lehrerin produziert aus der Rolle der Rezipientin heraus lediglich Zuhörersignale (Z. 31: „hm_hm;“), mit denen sie das Verstehen von (nicht aber Einverständnis mit) Nahemas Erklärung signalisiert. Wie genau Nahema seine Erklärung strukturiert und gestaltet, verdient genauere Aufmerksamkeit. Wir fahren jedoch vorerst mit dem Herauspräparieren der gesprächsstrukturellen Aufgaben des Erklärens fort und kommen erst an späterer Stelle aus erwerbs- bzw. kompetenzbezogener Perspektive auf Nahemas Diskurseinheit zurück.

„KAUFen; (-)“ (Z. 33) markiert sowohl semantisch als auch prosodisch das Ende der explanativen Diskurseinheit. Die Lehrerin verlässt nun die Rolle der Zuhörerin und macht einen Einspruch geltend (Z. 34). Die Erklärung wird fortgesetzt,

⁷ Im Rahmen dieser Analyse gehen wir auf die Funktion der Äußerung nicht weiter ein.

wobei weitere Schülerinnen und Schüler Erklärversuche unternehmen (im Transkript nicht abgebildet). Die Lehrerin leitet erst dann die gesprächsstrukturelle Aufgabe des *Abschließens* (Morek 2012) der Erkläraktivität ein, nachdem Thasin eine aus ihrer Sicht zutreffende Erklärung geliefert hat (Z. 78).

Welche Form von Ergebnissen hat die sequenzielle Analyse erbracht? Die Rekonstruktion der sequenziellen Erwartungen und ihrer Bearbeitung hat es uns ermöglicht, gesprächsstrukturelle Aufgaben herauszupräparieren, die für die Hervorbringung einer Erklärung konstitutiv sind. Dass sich die Beteiligten tatsächlich selbst an diesen Aufgaben orientieren, wird besonders deutlich, wenn noch nicht voll kompetente Gesprächsbeteiligte sequenzielle Erwartungen nicht befolgen und diese re-etabliert werden, wie nach Nahemas zunächst lokaler Antwort der Fall. Darüber hinaus belegen auch vergleichende Analysen anhand einer breiten Datenbasis, dass sich Interaktanten unabhängig vom jeweiligen Kontext (Familie, Schule) an den genannten gesprächsstrukturellen Aufgaben orientieren (Morek 2012). Solche gesprächsstrukturellen Aufgaben wurden auch für das Erzählen (Hausendorf/Quasthoff 1996) und Argumentieren (Heller 2012) rekonstruiert (für einen Überblick vgl. Morek/Heller/Quasthoff 2017). Wir haben somit kontextinsensitive konversationelle Aufgaben herauspräpariert, die nun als *tertium comparationis* für weiterführende Forschungsfragen fruchtbar gemacht werden können.

Im Projekt InterPass dienten die gesprächsstrukturellen Aufgaben zur Untersuchung der beiden oben (2.1) erläuterten Fragestellungen:

- *Klassenbezogene Partizipationsgelegenheiten.* Um lehrerspezifische Muster des Stellens diskursiver (und fachlicher) Anforderungen zu rekonstruieren, wurden zu den insgesamt zehn Deutsch- und Mathematiklehrkräften Kollektionen erstellt und im Hinblick auf die jeweilige Arbeitsteilung beim Erklären untersucht. Dabei zeigte sich, dass Schülerinnen und Schüler in unterschiedlicher Weise an Erklärungen beteiligt werden: als „Kleinteillieferanten“ mit lediglich lokalen Anteilen oder als „primäre Beiträger“ mit globalen Anteilen (Quasthoff et al. i.V.; vgl. auch Morek 2012 für den Grundschulunterricht).
- *Individuelle Lerngelegenheiten.* Die Frage nach den individuellen Lerngelegenheiten wurde längsschnittlich anhand von 11 ausgewählten (s.o.) Fokuskindern, darunter auch Nahema, untersucht. Dazu wurden Kollektionen erstellt, die alle Diskursaktivitäten dieser Kinder über die zwei Erhebungszeiträume umfassen. Die Bearbeitung dieser zweiten Forschungsfrage werden wir im Folgenden noch weiter skizzieren.

Wir betrachten dazu Nahemas Erklärung aus *erwerbs- bzw. kompetenzbezogener Perspektive*. In Bezug auf die Wahrnehmung und Erfüllung sequenzieller Erwartbarkeiten (Fähigkeitsdimension: ‚*Kontextualisierung*‘, vgl. Quasthoff 2011) ist zu beobachten, dass Nahema den globalen Zugzwang erst nach einer Explizierung durch die Lehrerin als solchen erkennt. Im Fähigkeitsbereich der *Vertextung* (ebd.) zeigt sich, dass Nahema in seiner Diskurseinheit beide Briefsorten thematisiert, wobei er Funktionen des persönlichen Briefes an einem Beispiel veranschaulicht, Funktionen des geschäftlichen Briefes dagegen eher generalisierend (allerdings inhaltlich nicht zutreffend) darstellt. Ein zentraler Unterschied wird

jedoch nicht expliziert und muss vom Zuhörer erschlossen werden. Ein Blick auf den Fähigkeitsbereich der *Markierung* zeigt, dass Nahema sprachliche Mittel für eine eindeutige Referenz auf die Relata fehlen. So verwendet er in Zeile 27/28 das Hyperonym „Brief“ und kann auf diese Weise nicht eindeutig markieren, auf welches der beiden Relata sich seine Aussage bezieht. Darüber hinaus setzt er keine Konjunktionen ein, um den Vergleich auch sprachlich zu indizieren.

Für die Untersuchung unterrichtlicher Lerngelegenheiten für Erklärkompetenz betrachten wir das lehrerseitige Agieren im Hinblick auf Verfahren interaktiver Unterstützung (Hausendorf/Quasthoff 1996; Heller/Morek 2015). So unterstützt die Lehrerin Nahema beim Verstehen des Zugzwangs, indem sie diesen noch einmal expliziert und Nahema eine zweite Chance zu Bearbeitung des Erklärbedarfs gewährt. Im Hinblick auf die Vertextung und Markierung seiner Diskurseinheit erhält Nahema hier jedoch keine weitere Unterstützung.

Über den Längsschnitt kann nun untersucht werden, ob und inwiefern sich Nahemas Erklärfähigkeiten in Bezug auf die drei Kompetenzdimensionen verändern. Daran anschließend ist die Frage zu stellen, ob etwaige Veränderungen auf die lehrerseitige Unterstützung zurückzuführen sind. So einfach dieses analytische Vorgehen auf den ersten Blick erscheint, so komplex ist es doch. Es verlangt nicht nur eine ausreichende Datenbasis, sondern wird u.U. auch durch die mangelnde Vergleichbarkeit unterrichtlicher Erkläraktivitäten erschwert. An dieser Stelle zeigt sich, wie wichtig eine an formalen, funktionsbezogen verankerten Gesprächspraktiken und -prinzipien orientierte Analyse ist (vgl. Abschnitt 1.1), die über die – manchem vielleicht zunächst trivial erscheinende – Detailbeschreibung eines Einzelfalls weit hinausgeht.

3. Schlussbemerkung

Mit der exemplarischen Analyse wird deutlich, dass die Gesprächsanalyse in der deutschdidaktischen Unterrichtsforschung für verschiedene Anwendungsfelder fruchtbar gemacht werden kann. Mit ihrer Hilfe lassen sich nicht nur die interaktiven Strukturen von Unterricht mikroanalytisch beschreiben, sondern auch Erkenntnisse zu interaktiven Mechanismen des Erwerbs von Sprach- und Diskurskompetenz (z.B. Gardner/Forrester 2010; Hausendorf/Quasthoff 2005) zu Tage fördern. In jüngster Zeit werden auf gesprächsanalytischem Wege auch die für das fachliche Lernen spezifischen sprachlich-diskursiven Anforderungen herauspräpariert (z.B. Harren 2015) und multimodale Ressourcen der Bedeutungskonstitution beschrieben (Pitsch 2006). Eine besondere Stärke des gesprächsanalytischen Zugangs ist u.E. darin zu sehen, dass er der Interaktivität und Kontextualisiertheit sprachlicher Praktiken Rechnung trägt und auf den ersten Blick unscheinbare Details sichtbar zu machen vermag, die für das Verständnis eines Phänomens (und seiner Herstellung) wesentlich sind.

Anhang: Transkriptionskonventionen GAT 2

Sequenzielle Struktur/Verlaufsstruktur

[]	Überlappungen und Simultansprechen
[]	
=	schneller, unmittelbarer Anschluss neuer Sprecherbeiträge oder Segmente (<i>latching</i>)

Ein- und Ausatmen

°h/h°	Ein- bzw. Ausatmen von ca. 0.2-0.5 Sek. Dauer
°hh/hh°	Ein- bzw. Ausatmen von ca. 0.5-0.8 Sek. Dauer
°hhh/hhh°	Ein- bzw. Ausatmen von ca. 0.8-1.0 Sek. Dauer

Pausen

(.)	Mikropause, geschätzt, bis ca. 0.2 Sek. Dauer
(-)	kurze geschätzte Pause von ca. 0.2-0.5 Sek. Dauer
(--)	mittlere geschätzte Pause v. ca. 0.5-0.8 Sek. Dauer
(---)	längere geschätzte Pause von ca. 0.8-1.0 Sek. Dauer
(0.5)	gemessene Pausen

Akzentuierung

akZENT	Fokusakzent
--------	-------------

Tonhöhenbewegung am Ende von Intonationsphrasen

?	hoch steigend
,	mittel steigend
—	gleichbleibend
;	mittel fallend
.	tief fallend

Sonstige segmentale Konventionen

und_äh	Verschleifungen innerhalb von Einheiten
äh öh äm	Verzögerungssignale, sog. „gefüllte Pausen“
=	schneller Anschluss neuer Sprecherbeiträge

: ; :: ; :::	Dehnung, je nach Längung (0.2-0.5, 0.5-0.8, 0.8-1.0 Sek.)
?	Abbruch durch Glottalverschluss

Lachen und Weinen

haha hehe hihi	silbisches Lachen
((lacht))	Beschreibung des Lachens
<<lachend> >	Lachpartikeln in der Rede, mit Reichweite
<<:-)> sooo>	„smile voice“

Rezeptionssignale

hm ja nein nee	einsilbige Signale
hm_hm ja_a	zweisilbige Signale
?hm?hm	mit Glottalverschlüssen, meistens verneinend

Sonstige Konventionen

((hustet))	para- und außersprachliche Handlungen und Ereignisse
<<hustend> >	sprachbegleitende para- und außersprachliche Handlungen und Ereignisse mit Reichweite
()	unverständliche Passage ohne weitere Angaben
(xxx), (xxx xxx)	ein bzw. zwei unverständliche Silben
(solche), (also/alo)	vermuteter Wortlaut, mögliche Alternativen
((...))	Auslassung im Transkript

Literatur

- Ayass, Ruth: Doing data. The status of transcripts in Conversation Analysis. In: Discourse Studies 2015, 17, 5, 505-528.
- Becker-Mrotzek, Michael/Vogt, Rüdiger: Unterrichtskommunikation. Linguistische Analysemethoden und Forschungsergebnisse. Tübingen: Niemeyer 2009.
- Bergmann, Jörg R.: Harold Garfinkel und Harvey Sacks. In: Flick, Uwe/Kardorff, Ernst/Steinke, Ines (Hrsg.): Qualitative Forschung. Ein Handbuch. Reinbek bei Hamburg: Rowohlt 2008a, 51-62.
- Bergmann, Jörg R.: Konversationsanalyse. In: Flick, Uwe/von Kardorff, Ernst/Steinke, Ines (Hrsg.): Qualitative Forschung. Ein Handbuch. Reinbek bei Hamburg: Rowohlt 2008b, 524-537.
- Couper-Kuhlen, Elizabeth: Some truths and untruths about final intonation in conversational questions. In: de Ruiter, Jan P. (Hrsg.): Questions. Formal, Functional and Interactional Perspectives. Cambridge: Cambridge University Press 2012, 123-145.
- Deppermann, Arnulf: Ethnographische Gesprächsanalyse: Zu Nutzen und Notwendigkeit von Ethnographie für die Konversationsanalyse. In: Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion 2000, 1, 96-124.
- Deppermann, Arnulf: Gespräche analysieren. Eine Einführung. Wiesbaden: VS 2008.
- Ehlich, Konrad/Rehbein, Jochen: Halbinterpretative Arbeitstranskriptionen (HIAT). In: Linguistische Berichte 1976, 45, 21-41.
- Gardner, Hilary/Forrester, Michael A. (Hrsg.): Analysing interactions in childhood. Insights from conversation analysis. Chichester: Wiley-Blackwell 2010.
- Garfinkel, Harold: Studies in ethnomethodology. Cambridge: Polity Press 1967.
- Günthner, Susanne/Knoblauch, Hubert: "Forms are the food of faith": Gattungen als Muster kommunikativen Handelns. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 1994, 46, 4, 693-723.
- Harren, Inga: Fachliche Inhalte sprachlich ausdrücken lernen. Sprachliche Hürden und interaktive Vermittlungsverfahren im naturwissenschaftlichen Unterrichtsgespräch in der Mittel- und Oberstufe. Radolfzell: Verlag für Gesprächsforschung 2015.
- Hausendorf, Heiko/Quasthoff, Uta M.: Konversations-/Diskursanalyse: (Sprach-) Entwicklung durch Interaktion. In: Mey, Günter (Hrsg.): Handbuch Qualitative Entwicklungspsychologie. Köln: Kölner Studien Verlag 2005, 585-618.
- Hausendorf, Heiko/Quasthoff, Uta M.: Sprachentwicklung und Interaktion. Eine linguistische Studie zum Erwerb von Diskursfähigkeiten. Radolfzell: Verlag für Gesprächsforschung 1996.
- Heath, Christian/Hindmarsh, Jon/Luff, Paul: Video in qualitative research. Analysing social interaction in everyday life. Los Angeles: Sage 2010.
- Heller, Vivien/Morek, Miriam: Unterrichtsgespräche als Erwerbskontext: Kommunikative Gelegenheiten für bildungssprachliche Praktiken erkennen und nutzen. In: Literalität im Schnittfeld von Familie, Frühbereich und Schule. Themenheft im *Leseforum.ch*, 2015, 3.
- Heller, Vivien: Gesprächsanalyse in der sprachwissenschaftlichen und sprachdidaktischen Unterrichtsforschung. In: Neumann, Astrid/Mahler, Isabelle (Hrsg.): Empirische Methoden der Deutschdidaktik. Audio- und videogRAFIERENDE Unterrichtsforschung. Baltmannsweiler: Schneider Hohengehren 2014, 122-150.

- Heller, Vivien: Kommunikative Erfahrungen von Kindern in Familie und Unterricht. Passungen und Divergenzen. Tübingen: Stauffenburg 2012.
- Heritage, John/Clayman, Steven: Talk in action. Interactions, identities, and institutions. Malden: Wiley-Blackwell 2010.
- Hutchby, Ian/Wooffitt, Robin: Conversation analysis. Cambridge: Polity Press 2008.
- Jefferson, Gail: Glossary of transcript symbols with an introduction. In: Lerner, Gene (Hrsg.): Conversation Analysis. Studies from the first generation. Amsterdam: Benjamins 2004, 13-31.
- Keppler, Angela: Schritt für Schritt. Das Verfahren alltäglicher Belehrungen. In: Soziale Welt (1989), 40, 4, 538-556.
- Levinson, Stephen C.: Pragmatik. Tübingen: Niemeyer 1990.
- Macbeth, Douglas: The relevance of repair for classroom correction. In: Language in Society (2004), 33, 703-736.
- Mazeland, Harrie: Sprecherwechsel in der Schule. In: Ehlich, Konrad/Rehbein, Jochen (Hrsg.): Kommunikation in Schule und Hochschule. Linguistische und ethnomethodologische Analysen. Tübingen: Narr 1983, 77-101.
- Mehan, Hugh: "What Time Is It, Denise?": Asking Known Information Questions in Classroom Discourse. In: Theory into Practice (1979), 18 4, 285-294.
- Morek, Miriam/Heller, Vivien/Quasthoff, Uta M.: Erklären und Argumentieren. Modellierungen und empirische Befunde zu Strukturen und Varianzen. In: Wyss, Eva Lia/Meißner, Iris (Hrsg.): Erklären und Argumentieren. Konzepte und Modellierungen in der Angewandten Linguistik. Tübingen: Stauffenburg 2017, 11-46.
- Morek, Miriam: Kinder erklären. Interaktionen in Familie und Unterricht. Tübingen: Stauffenburg 2012.
- Pitsch, Karola: Sprache, Körper, Intermediäre Objekte: Zur Multimodalität der Interaktion im bilingualen Geschichtsunterricht. Universität Bielefeld, Bielefeld 2010. Online verfügbar unter <http://d-nb.info/1008212989/34> (letzter Zugriff: 01.08.2018).
- Pomerantz, Anita/Heritage, John: Preference. In: Sidnell, Jack/Stivers, Tanya (Hrsg.): The handbook of conversation analysis. Chichester: Blackwell, 2013, 210-228.
- Quasthoff, Uta M./Heller, Vivien/Prediger, Susanne/Erath, Kirstin: Learning in classroom interaction from the perspectives of interactional discourse analysis and subject matter education. Eingereicht.
- Quasthoff, Uta M.: Diskurs- und Textfähigkeiten: Kulturelle Ressourcen ihres Erwerbs. In: Hoffmann, Ludger/Leimbrink, Kerstin/Quasthoff, Uta M. (Hrsg.): Die Matrix der menschlichen Entwicklung. Berlin: de Gruyter 2011, 210-251.
- Quasthoff, Uta M.: Erzählen in Gesprächen. Linguistische Untersuchungen zu Strukturen und Funktionen am Beispiel einer Kommunikationsform des Alltags. Tübingen: Narr 1980.
- Sacks, Harvey/Jefferson, Gail: Lectures on conversation. Oxford: Blackwell 1995.
- Sacks, Harvey/Schegloff, Emanuel A./Jefferson, Gail: A Simplest Systematics for the Organization of Turn-Taking for Conversation. In: Language (1974), 50, 4, 696-735.
- Sacks, Harvey/Schegloff, Emanuel A.: Two preferences in the organization of reference to persons in conversation and their interaction. In: Psathas, George (Hrsg.): Everyday language. New York: Center for the Study of EM and CA 1979, 15-21.

- Sacks, Harvey: On the Analyzability of Stories by Children. In: Gumperz, John J./Hymes, Dell (Hrsg.): Directions in sociolinguistics. The ethnography of communication. Oxford: Blackwell 1972, 325-345.
- Schegloff, Emanuel A.: Sequence organization in interaction. Cambridge: Cambridge University Press 2007.
- Schegloff, Emanuel A.: Sequencing in Conversational Openings. In: Gumperz, John J.; Hymes, Dell (Hrsg.): Directions in sociolinguistics. The ethnography of communication. Oxford: Blackwell 1972, 346-380.
- Schegloff, Emanuel A./Jefferson, Gail/Sacks, Harvey: The Preference for Self-Correction in the Organization of Repair in Conversation. In: Language (1977), 53, 2, 361-382.
- Schegloff, Emanuel A./Sacks, Harvey: Opening Up Closings. In: Semiotica (1973), 8, 289-327.
- Schiffrin, Deborah: Discourse markers. Cambridge: Cambridge University Press 1987.
- Schwab, Götz: Transana – ein Transkriptions- und Analyseprogramm zur Verarbeitung von Videodaten am Computer. In: Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion (2006), 7, 70-78.
- Schwarze, Cordula: Theoretische und methodische Überlegungen zur Praxis der gesprächsanalytischen Datensitzung. In: Schwarze, Cordula/Konzett, Carmen (Hrsg.): Interaktionsforschung: Gesprächsanalytische Fallstudien und Forschungspraxis. Berlin: Frank & Timme 2014, 161-175.
- Selting, Margret/Auer, Peter/Barth-Weingarten, Dagmar/Bergmann, Jörg/Bergmann, Pia/Birkner, Karin et al: Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). In: Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion 2009, 10, 353-402. [englisch: Dies.: A system for transcribing talk-in-interaction: GAT 2. In: Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion 2011, 12, 1-51.]
- Selting, Margret/Couper-Kuhlen, Elizabeth: Argumente für die Entwicklung einer ‚interaktionalen Linguistik‘. In: Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion (2000), 1, 76-95.
- Sidnell, Jack/Stivers, Tanya (Hrsg.): The handbook of conversation analysis. Chichester: Wiley-Blackwell 2013.
- Silverman, David: Harvey Sacks. Social science and conversation analysis. Cambridge: Polity Press 1998.
- Spranz-Fogasy, Thomas: Alles Argumentieren, oder was? Zur Konstitution von Argumentation in Gesprächen. In: Deppermann, Arnulf/Hartung, Martin (Hrsg.): Argumentieren in Gesprächen. Gesprächsanalytische Studien. Tübingen: Stauffenburg 2003, 27-39.
- Streeck, Jürgen: Konversationsanalyse. Ein Reparaturversuch. In: Zeitschrift für Sprachwissenschaft (1983), 2, 1, 74-104.
- Stukenbrock, Anja: Herausforderungen der multimodalen Transkription. In: Birkner, Karin/Stukenbrock, Anja (Hrsg.): Arbeit mit Transkripten in der Praxis. Forschung, Lehre und Fortbildung. Mannheim: Verlag für Gesprächsforschung 2009, 144-170.

Grounded Theory

How might it have been otherwise¹

1. Vorstellung der Grounded Theory

Die Grounded Theory wurde in den 1960er Jahren von Anselm Strauss und Barney Glaser entwickelt – in ihrem gemeinsamen Buch *The Discovery of Grounded Theory* (1967) stellen die beiden Wissenschaftler ihr methodisches Vorgehen vor. Es handelt sich um ein Vorgehen, das von den beiden Autoren für den Umgang mit Daten aus dem Kontext der qualitativ-empirischen Sozialforschung entwickelt wurde. Der Schwerpunkt liegt dabei auf der Auswertung verbaler Daten (Interviewtranskripte oder Beobachtungsnotizen), im Zentrum stehen bestimmte Verfahren des Kodierens.

Das Ziel der Grounded Theory ist die Entwicklung einer Theorie aus den Daten heraus, insofern handelt es sich um ein offenes, exploratives Vorgehen, das Hypothesen generierend oder erweiternd und nicht überprüfend vorgeht.

Auch wenn der Schwerpunkt auf der Auswertungsphase liegt, handelt es sich um ein Vorgehen, das den gesamten Forschungsprozess in den Blick nimmt. In diesem Sinne ist es sinnvoll von einem *Forschungsstil* und nicht nur von einzelnen Methoden zu sprechen, denn Glaser und Strauss schlagen mit der Grounded Theory eine bestimmte Art des Forschens vom Zeitpunkt der Konzeption über die Phase der Erhebung bis hin zur Auswertung der Daten vor.

1.1 Funktionsweise der Grounded Theory

Zur Veranschaulichung seines wissenschaftlichen Vorgehens stellt Anselm Strauss in seinem Buch *Grundlagen qualitativer Sozialforschung* (1996) Interviewmitschnitte aus seiner eigenen Lehre dar. Hier zeigt er, wie er versucht, Studierenden sein Verfahren nahezubringen. Dafür arbeitet er mit Datenmaterial aus seinem eigenen Forschungsprojekt über den Umgang mit Schmerzen im Krankenhaus. Am Ende einer Sitzung, in der der Dozent und die Studierenden gemeinsam verschiedene Möglichkeiten, das Material zu kodieren und theoretisch zu befragen, diskutiert haben, findet sich folgendes abschließendes Resümee:

¹ Ein für Anselm Strauss zentraler Lehrsatz seines akademischen Lehrers Everett C. Hughes, einem Pionier der soziologischen Feldforschung (vgl. Strübing 2007, 11).

Student: ‚So geht es also nicht um das Problem, die Wahrheit herauszufinden, sondern darum, welche Wahrheit es ist. Was immer man herausfindet, es wird wahr sein, und es wird gültig sein. Was man auslässt, ist vielleicht nicht interessant oder wichtig für einen selbst, kann aber später für jemand anderen wichtig sein.‘ Dozent: ‚Und wenn Ihre Theorie breit genug angelegt und ausreichend dicht ist – wie wir im *Discovery*-Buch sagen –, dann kann die Arbeit anderer Wissenschaftler genau in Ihre Theorie passen. [...] Wenn jemand einwendet: ‚Ich habe nicht gesehen, was Sie gesehen haben‘, dann sagen Sie: ‚Natürlich nicht, Sie sehen die Arbeit auch in einem anderen Bezugsrahmen oder Sie haben den Umgang mit dem Schmerz vielleicht auf einer anderen Station beobachtet. Wenn Sie mir aber Schritt für Schritt in meine Arbeit folgen, dann werden Sie mit Sicherheit das sehen, was ich gesehen habe‘. (Strauss 1998, 83)

Die Beispiele, die Anselm Strauss in seinem Grundlagenwerk gibt, sind aufgrund ihrer komplexen Denkbewegungen selbst nicht leicht zu verstehen. Ich versuche eine interpretative Annäherung an das Zitat: Die *Grounded Theory* kann zunächst einmal als eine Abgrenzungsbewegung gegen einen naiven Begriff von wissenschaftlicher Theorie verstanden werden. Sie grenzt sich gegen die Vorstellung ab, es gäbe eine ‚richtige‘ theoretische Beschreibung eines Phänomens zu entdecken – in dem Sinne, dass allein durch intensives Nachdenken und eine genaue Konzeption der Erhebung der oder die Wissenschaftlerin in die Lage versetzt werden könnte, die allumfassende und treffende theoretische Beschreibung eines Phänomens liefern zu können. Ein solches Denken führt in der Regel zu einer sehr großen Konzentration der Arbeitsenergie bei der Vorbereitung und Durchführung der Erhebungsmethode, meist noch ohne jeglichen (oder wenig) Kontakt zum Feld und den damit verbundenen Phänomenen. Dabei besteht allerdings die Gefahr, dass erst durch den Kontakt zum Feld für den Forschenden klar wird, welche Phänomene im Feld auftreten und welche davon überhaupt interessant sind. Diesen Prozess wissenschaftlichen Vorgehens versucht die *Grounded Theory* im gewissen Sinne umzudrehen: Ihr Ziel ist es, möglichst früh im Forschungsprozess in Kontakt mit den Phänomenen selbst zu kommen, diese intensiv auszuwerten und das eigene forschende Vorgehen den Beobachtungen anzupassen. Es soll ein Wechselspiel zwischen Erhebung und Auswertung initiiert werden, damit sich theoretische Beschreibungen und das Untersuchen der Phänomene wechselseitig beeinflussen. Am Ende steht im besten Falle eine Theorie, die möglichst dicht und nahe an den Phänomenen ist und diese trotzdem theoretisch neu erschließt.

Durch die zirkuläre Struktur des Forschungsprozesses werden kontinuierliche Weichenstellungen durch die Forschenden nötig: Welche Phänomene sind interessant? Unter welchem Aspekt sollen die Phänomene weiter untersucht werden? Wie können bestimmte Aspekte im nächsten Schritt noch genauer untersucht werden, damit sich bestimmte Annahmen überprüfen lassen? Es geht immer wieder um Selektion und neue Rahmung der Ergebnisse, um dann erneut zu untersuchen. Diese Schritte sind dann für alle Außenstehenden nachvollziehbar – ‚dann werden Sie mit Sicherheit das sehen, was ich gesehen habe‘ –, verweisen aber nicht auf jene eine Wahrheit, die es ‚herauszufinden‘ gilt.

Peter Alheit (1999) stellt in seinem einführenden Vortrag heraus, dass für die Grounded Theory das Handlungskonzept des amerikanischen Pragmatismus leitend sei: Ein Konzept, das Handeln als natürlicherweise diffus teleologisch auffasst: „d.h. im Regelfall handeln wir zwar mit einer bestimmten Grundintention, die konkreten Zwecke der Handlung werden jedoch zumeist erst im Handlungsvollzug deutlich und können durchaus zur Revision der ursprünglichen Intention führen“ (Alheit 1999, 5). Hier wird deutlich, dass Erfahrungen im Handlungsprozess denselben verändern können. Darauf zielt auch der von Glaser und Strauss praktizierte Forschungsstil: Ein wissenschaftliches Handeln, das sich durch die im Forschungsprozess gemachten Erfahrungen verändert und sich dem anpasst, was es erforscht.

1.2 Ausgangsdaten und Ergebnisse

Grundsätzlicher Anspruch der Grounded Theory ist es, den *gesamten* (theoriegenerierenden) Forschungsprozess zu strukturieren, von der Phase der Datenerhebung und des Samplings bis hin zur Auswertung mit bestimmten Kodierverfahren. Dabei tragen die Kodierverfahren der Grounded Theory insbesondere der Individualität von Phänomenen Rechnung und ermöglichen es, darin Strukturen zu entdecken. Deshalb eignen sie sich am besten zur Auswertung von wenig strukturierten verbale Daten, z.B. für die Auswertung von narrativen Interviews (s. Lindow in diesem Band), Laut-Denken-Protokollen (s. Dannecker in diesem Band) oder Beobachtungsnotizen.²

Zunächst möchte ich einige grundlegende, mögliche Kodierverfahrensschritte der Grounded Theory vorstellen, um dann im Anschluss daran zentrale Qualitäten des Forschungsprozesses selbst darzulegen. Dabei beschränke ich mich auf die Vorschläge zum Kodierverfahren von Strauss und Corbin, der im deutschsprachigen Raum am meisten verbreiteten Schule der Grounded Theory.

Strauss und Corbin unterscheiden in ihrem Lehrbuch (Strauss/Corbin 1996) drei grundlegende Schritte des Kodierens: das offene, axiale und selektive Kodieren. Die Schritte des offenen und axialen Kodierens können abwechselnd eingesetzt werden und sich gegenseitig befruchten – sie brechen die Daten auf und systematisieren sie. Das selektive Kodieren dient der abschließenden Theoriebildung über die Daten. Im Folgenden stelle ich den Prozess des Kodierens linear dar, auch wenn es sich im Vollzug nicht um vollständig voneinander getrennte Phasen handelt.

Der erste Schritt, um z.B. ein Interviewtranskript auszuwerten, besteht darin, die Aussagen im Interview offen zu kodieren. Gemeint ist damit eine Zeile-für-Zeile oder Abschnitt-für-Abschnitt Zuordnung des Textes zu Labels oder Schlüsselwörtern, den Codes. Die Codes werden vom Forschenden ‚frei‘ entwickelt, in

² Daten, die selbst schon eine starke Struktur mit sich bringen, wie z.B. die Ergebnisse von Leitfrageninterviews, lassen sich einfacher und zeitsparender mit der qualitativen Inhaltsanalyse untersuchen (s. Heins in diesem Band).

dem Sinne, dass hier noch eine sehr große Vielfalt möglicher Kodes besteht. Kodes zu entwickeln, bedeutet zunächst einmal ganz trivial, die Vorkommnisse in den Daten miteinander zu vergleichen und ähnliche Vorkommnisse aufgrund bestimmter Gemeinsamkeiten mit demselben Kode zu versehen.

Der oder die Forschende nutzt für die Generierung von Kodes sowohl Anregungen aus dem Material selbst (natürliche Kodes) als auch sein eigenes theoretisches Hintergrundwissen (soziologisch konstruierte Kodes). Muckel spricht in dieser Phase von einer „Methode des freien Assoziierens“ (Muckel 2011, 343). Wichtig ist allerdings schon in dieser Phase, dass die Kodes bereits Ansätze zur Konzeptionalisierung der Daten liefern sollen. Sie stellen also nicht eine bloße Zusammenfassung des Inhaltes dar, sondern bieten eine konzeptionelle Erschließung der Daten:

Darunter versteht man das Benennen von Phänomenen mithilfe von Begriffen, die abstrakter und prägnanter sind als oberflächliche, allgemein gehaltene Beschreibungen. Um diese Unterscheidung zu verdeutlichen, möchte ich ein Beispiel geben. Eine Beobachtung kann ich beschreiben als: „Zwei Menschen sprechen miteinander.“ Wenn ich jedoch nach einer präziseren und zugleich konzeptualisierenden Benennung suche, könnte ich sagen: „Ein Mensch berät einen anderen.“ „Beraten“ ist eine konzeptualisierende Benennung im Unterschied zum Miteinander-Sprechen, weil das Beraten bereits eine bestimmte Rollenverteilung impliziert und so Fragen nach der Beziehung der beiden Personen zueinander, dem Gegenstand ihrer Beratung, der Qualität der Beratung etc. eröffnet. (ebd., 338)

Die Vergabe von konzeptionellen Kodes ist also ein erster Schritt, die Daten auf einem abstrakteren Niveau zu repräsentieren.

Während das offene Kodieren das Material in der Breite erschließt, könnte man sagen, dass das axiale Kodieren auf Tiefenbohrungen in den Daten zielt. Beim axialen Kodieren werden bestimmte, für die eigene Fragestellung als besonders relevant angesehene Stellen mit Hilfe eines Kodierparadigmas genauer analysiert. Strauss und Corbin stellen in ihrem Lehrbuch ein Kodierparadigma vor, das sie für besonders hilfreich bei der Untersuchung soziologischer Phänomene halten. Das vorgestellte Kodierparadigma ist im Grunde eine Art und Weise, wie man systematisch Fragen an das Material stellen kann. Strauss und Corbin befragen die Phänomene nach den in ihnen repräsentierten Ursachen, Kontexten, Bedingungen und Konsequenzen menschlichen Handelns sowie den darin sichtbar werdenden Handlungen und Strategien der involvierten Akteure (vgl. Strauss/Corbin 1996, 78ff.). Durch die Arbeit mit einem Kodierparadigma wird es möglich, Strukturzusammenhänge in den Phänomenen zu erkennen und zu beschreiben.

Das von Strauss und Corbin vorgeschlagene Kodierparadigma ist soziologisch ausgerichtet, da hier im Zentrum des Forschungsinteresses das menschliche Handeln steht. Solch ein Kodierparadigma eignet sich gut zur vertieften Analyse von Datenmaterial, das sich phänomenologisch auf Situationen bezieht, in denen Menschen handeln. Insofern ist es zur Analyse von Unterrichtssituationen sinnvoll, nicht aber zur Untersuchung von Laut-Denken-Protokollen, in denen sich

Verstehensprozesse abbilden. Hier ist es angemessener, ein Kodierparadigma zu verwenden oder selbst zu entwickeln, das dem Untersuchungsgegenstand angepasst ist (vgl. z.B. Stark 2012, 155).

In der den Forschungsprozess abschließenden Phase des selektiven Kodierens werden die in den vorherigen Arbeitsschritten gesammelten Erkenntnisse zu einer Theorie integriert – Strauss und Corbin sprechen in diesem Zusammenhang davon, dass eine Geschichte, „[e]ine beschreibende Erzählung oder Darstellung über das zentrale Phänomen der Untersuchung“ (Strauss/Corbin 1996, 94) erzählt werden muss. Diese Aussage von Strauss und Corbin verweist darauf, dass der letzte Arbeitsschritt nicht im Sinne einer analytischen Ableitung aus den entwickelten Kategorien erfolgt, sondern es vielmehr um eine Systematisierung der Erkenntnisse unter einer bestimmten Perspektive geht. Denn mit denselben Fakten können je nach Perspektive immer wieder unterschiedliche Geschichten erzählt werden. So entwickelt der Forschende im Prozess des selektiven Kodierens probenhalber eine Kernkategorie, die seiner Meinung nach das zentrale Thema der Untersuchung organisieren könnte und prüft durch Rekodieren des Materials, ob sich die anderen Kategorien darauf beziehen lassen. Wenn dies der Fall ist, handelt es sich tatsächlich um eine Kernkategorie. Natürlich lassen sich immer mehrere Kernkategorien finden, die das Material unter verschiedenen Aspekten oder Schwerpunkten organisieren – es liegt also in der Sicht des Forschenden begründet, was für ihn oder sie das zentrale Phänomen der Untersuchung darstellt.

Neben den bisher beschriebenen Arbeitsschritten gibt es einige prozessorientierte Merkmale, die das Arbeiten im Stil der Grounded Theory kennzeichnen. Dazu gehören: das theoretische Sampling, das Schreiben von Memos und die Arbeit im Team.

Innerhalb des Forschungsprozesses ist es notwendig, kontinuierlich Auswahlentscheidungen zu treffen. Zum einen muss zu Beginn der Erhebung und bei jeder weiteren Erhebung geklärt werden, wie viel und welches Material erhoben werden soll. Zum anderen muss während des Untersuchungsprozesses immer wieder entschieden werden, welche Teile des Datenmaterials untersucht werden. Das theoretische Sampling organisiert diese Auswahlentscheidungen, indem es ein Auswahlkriterium angibt: Die Auswahlentscheidungen sollen vor dem Hintergrund der sich entwickelnden Theorie getroffen werden. Damit zielt die Arbeit nicht wie bei statistischen Samplingverfahren auf eine Repräsentativität der Stichprobe für eine bestimmte Grundgesamtheit. Vielmehr müssen sich die Forschenden die Frage stellen, welche Daten sie zum gegebenen Zeitpunkt brauchen, um ihre theoretischen Ideen anzureichern oder zu überprüfen. Dabei spielen minimale und maximale Vergleiche innerhalb des Materials die zentrale Rolle. Mit Hilfe des ständigen Vergleichens gelingt es den Forschenden, die Phänomene auszudifferenzieren und zu ordnen. Es ist klar, dass die Auswahlentscheidungen, da sie „auf der Basis der analytischen Fragen erfolgen, die der bisherige Stand der Theoriebildung am konkreten Projekt aufwirft“ (Strübing 2008, 30), getroffen werden, erst im Laufe des Projektes immer spezifischer und genauer werden und somit die Eindeutigkeit im eigenen Vorgehen zunimmt.

Ein weiteres Merkmal der Arbeit im Stil der Grounded Theory ist das fortwährende Schreiben von Memos: In Form von spontanen Notizen hält der Forschende fest, was ihm bei der Arbeit durch den Kopf geht – es sind schriftliche Kommentare, die sich auf alle Ebenen des Forschungsprozesses beziehen können: auf die Kodes selbst, auf bestehende Theorien, auf die eigene im Entstehen begriffene Theorie. Es kann sich also um Kommentare handeln, die alternative Ideen zum vergebenen Kode enthalten, die Querverweise im Material bemerken, die Ideen für nächste Schritte festhalten, die Ähnlichkeiten oder Abweichungen zu bestehenden Forschungsarbeiten und deren Ergebnissen notieren usw. Das Material der Memos kann dann besonders genutzt werden, wenn neue Phasen im Forschungsprozess anstehen oder neue Kategorien entwickelt werden und am Ende bei der Suche nach der Kernkategorie.

Ein letztes wesentliches Merkmal der Arbeit im Stil der Grounded Theory stellt meiner Meinung nach das Arbeiten im Team dar. Den Hinweis auf die Bedeutung von Teamarbeit im Zusammenhang mit Forschungen im Stil der Grounded Theory findet sich nicht explizit in den Lehrbüchern der Grounded Theory. Beschäftigt man sich aber mit der Forschungsarbeit der Gründungsväter der Grounded Theory, so wird deutlich, dass es sich hierbei um ein weiteres – besonders für Anfängerinnen und Anfänger – wesentliches Moment des Arbeitens handelt. Gerade aufgrund der Offenheit des Verfahrens – allein schon die erste Phase des offenen Kodierens stellt hierbei eine große Herausforderung dar – ist es schwierig, alleine gut voranzukommen. Strauss bemerkt dazu,

dass Studenten aus unseren Forschungsseminaren manchmal einem besonders schwierigen Problem gegenüberstehen. Wenn sie nämlich versuchen, ihre Materialien allein zu kodieren, d.h. ohne Unterstützung durch die Analysediskussion im Seminar, dann haben sie das Gefühl, dass ihr Kodieren nicht annähernd so mühelos oder ‚tiefschürfend‘ ist wie das im Seminar, und sie haben vielleicht kein ausreichendes Vertrauen in ihre Arbeit. (Strauss 1998, 70)

Das wechselseitige Kodieren und der gemeinsame Austausch darüber stellt ein wesentliches Moment dar, um die eigenen Erkenntnisse zu validieren und dadurch Vertrauen innerhalb eines Prozesses zu gewinnen, in dem das Ergebnis des eigenen Arbeitens offen ist. Strauss selbst hat immer in kooperativen – nicht hierarchisch organisierten – Teams gearbeitet, bei denen sich die Arbeit von mehr oder weniger erfahrenen Wissenschaftlerinnen und Wissenschaftlern gegenseitig ergänzte und befruchtete. Zielsetzung war es, dass sowohl gute Ergebnisse produziert werden als auch „die Kreativität der einzelnen Projektteilnehmer zur Entfaltung kommt“ (ebd., 68) – also ein explizit forschungsethisches Anliegen, dessen große Strahlkraft Personen, die mit Strauss zusammengearbeitet haben, betonen (vgl. Riemann 2011).

1.3 Vorteile und Nachteile der Grounded Theory

Die Vorteile des Verfahrens liegen insbesondere in der Flexibilität des Vorgehens. Da sich das Vorgehen immer wieder schrittweise an die neu gewonnenen Erkenntnisse anpassen lässt, kann aus fehlgeschlagenen Vermutungen der nächste neue Erkenntnisschritt gewonnen werden. Die enge Verzahnung von theoriegeleiteter Vermutung und empirischer Überprüfung verhindert es, dass zu viel Arbeit in ein Erhebungskonzept investiert wird, das die wesentlichen Phänomene nicht erfasst und somit entweder keine oder triviale Ergebnisse liefert – sich dann aber aufgrund des fortgeschrittenen Forschungsprozesses nicht mehr korrigieren lässt. Im Forschungsprozess der Grounded Theory kann immer wieder nachjustiert werden und der Kontakt zwischen Theoriebildung und empirischem Phänomen geht nicht verloren.

Nachteile dieses Vorgehens liegen in der Offenheit des Verfahrens. Um sinnvoll die Entscheidungen treffen zu können, die im Laufe des Forschungsprozesses nötig werden, müssen die Forschenden sich sehr genau darüber bewusst sein, auf welchen Ebenen Phänomene analysiert werden können, um zwischen diesen verschiedenen Ebenen bewusst wählen zu können. Strauss und Glaser sprechen in diesem Zusammenhang von theoretischer Sensitivität, dem „Gespür dafür, wie man über Daten in theoretischen Begriffen nachdenkt“ (Strauss 1998, 6). Gerade dies ist aber eine Eigenschaft, die man als Anfänger/in vielleicht allererst noch entwickeln muss. Hier kann Teamarbeit eine große Hilfe sein, da innerhalb der Kommunikation verschiedene Sichtweisen durchgespielt und gegeneinander abgewogen werden können. Neben diesem Problem auf der theoretischen Ebene gibt es allerdings auch praktische Probleme bei der Arbeit mit der Grounded Theory: Nicht immer ist es möglich, im Laufe des Forschungsprozesses wiederholt Zugang zum Feld zu haben und entsprechend Erhebung, Auswertung und Theoriebildung wie gefordert miteinander zu verzahnen. Darüber hinaus hat das Verfahren einen hohen Anspruch: Das Ziel ist es, eine theoretische Sättigung der zu entwickelnden Theorie zu erreichen. Das heißt, dass die Theoriebildung erst dann als abgeschlossen zu betrachten wäre, wenn weitere Erhebungen keine neuen Erkenntnisse mehr produzierten – also *alle* möglichen Daten sich mit Hilfe der entwickelten Theorie erklären ließen und keine ‚unerklärlichen‘ Phänomene mehr enthielten.

1.4 Gütekriterien für den Einsatz der Grounded Theory

Was soll eine gute Grounded Theory ausmachen? Der zentrale Anspruch [...] besteht in der Erarbeitung einer Theorie, die soziale Prozesse erklären und insofern mit Einschränkungen (s.o.) auch vorhersagen kann. [...] Angestrebt wird soziologische Theoriebildung nicht um ihrer selbst willen, sondern mit dem Ziel einer verbesserten Handlungsfähigkeit der Akteure im Untersuchungsbereich. (Strübing 2008, 85)

Dementsprechend, so Strübing, erweist sich die Güte einer Theorie in der Praxis: Sie dient der Verhaltenssteuerung der in einem Feld Handelnden, indem sie Voraussagen über das Auftreten von Phänomenen macht. Auf die Schulwirklichkeit

übertragen, bedeutete dies, dass die Güte einer Theorie sich in ihrer Wirksamkeit für die Unterrichtspraxis erweisen würde. Trotz dieses pragmatischen Ziels, in dem sich auch deutlich der Wahrheitsbegriff des Pragmatismus widerspiegelt, stellen sich natürlich auch die Vertreterinnen und Vertreter der Grounded Theory dem Anspruch einer auf den wissenschaftlichen Diskurs bezogenen Prüfung der Güte von Forschungsergebnissen. Wie alle qualitativen Verfahren können mit der Grounded Theory keine Ergebnisse produziert werden, die den Kriterien der quantitativen Forschung wie Reliabilität, Repräsentativität und Validität genügen, sondern es werden analoge Kriterien formuliert. Dass Wiederholbarkeit von wissenschaftlichen Ergebnissen im Bereich des sozialen Handelns nicht im exakten Sinne herstellbar ist, liegt im Wesen der Sache begründet. Die Grounded Theory strebt aber im Sinne einer in gewissen Grenzen Vorhersehbarkeit von sozialen Prozessen Ergebnisse an, die mit Einschränkungen reproduzierbar sind. In Bezug auf das Kriterium der Repräsentativität wird an die Stelle einer statistischen Repräsentativität, die auf eine breite Population als Bezugsgröße zielt, die theoretische Repräsentativität gesetzt: Es soll ein bestimmtes Phänomen spezifiziert werden – dieses soll in den Daten repräsentiert sein und zwar möglichst umfassend, sodass die Theorie über das Phänomen gesättigt ist. Somit ist der Begriff des theoretischen Samplings (also der theoriegeleiteten Auswahl des Datenmaterials; s. 1.2) eng mit dem Begriff der theoretischen Sättigung verknüpft: Eine Theorie über ein Phänomen gilt dann als gesättigt, wenn sich durch zusätzliche Daten oder durch weitere Auswertung der vorhandenen Daten keine neuen Aussagen über dieselben gewinnen lassen. Die unterschiedliche Güte von Theorien, die auf der Grundlage des methodischen Vorgehens der Grounded Theory gewonnen werden, lässt sich also danach beurteilen, wie gesättigt eine Theorie ist, und wie groß die Reichweite einer Theorie ist, d.h. welche Varianz von Phänomenen mit Hilfe der Theorie erklärt werden können. Das Kriterium der Validität prägt das Vorgehen der Grounded Theory in dem Sinne, dass es idealerweise im Prozess des Forschens immer wieder zu einer Anpassung der Erhebungsinstrumente an die Daten kommt, insofern der Forschungsprozess als solcher zyklisch angelegt ist (s. 1.1). Dieses Kriterium nach außen sichtbar zu machen, stellt ein weiteres Gütekriterium beim Einsatz von Verfahren der Grounded Theory dar: Die Dokumentation von Entscheidungsprozessen und Sichtbarmachung der Genese der Forschungsergebnisse kennzeichnet eine gute Grounded Theory – je transparenter und nachvollziehbarer die Entscheidungen im Forschungsprozess für andere Forschende ist, desto höher die Güte einer Theorie.

1.5 Literaturempfehlungen

Strauss (1998): Grundlagen qualitativer Sozialforschung.

Einführendes Grundlagenbuch, das anhand von vielen ausführlichen Beispielen aus der Forschungsarbeit des Autors versucht, die Tätigkeiten eines Grounded-Theory-Forschenden zu veranschaulichen. Die Beispiele sind nicht unbedingt leicht zu verstehen. Es hat den Charakter eines Werkstatt-Buches.

Strauss/Corbin (1996): Grounded Theory.

Einführendes Lehrbuch, das systematisch Begriffe und Forschungsschritte der Grounded Theory in der Tradition von Strauss und Corbin erläutert. Es vermittelt bestimmte Techniken, stellt die Grounded Theory stark komprimiert dar.

Strübing (2008): Grounded Theory.

Sehr gut lesbare Einführung, die insbesondere den wissenschaftstheoretischen Hintergrund und somit den ‚Geist‘, der die Grounded Theory trägt, deutlich macht.

2. „und dann scheint er irgendwie ertrunken zu sein“ – Wie Schülerinnen und Schüler die Kurzprosa *Brudermord im Altwasser* verstehen

Im Folgenden werde ich einige Aspekte aus einem eigenen Forschungsprojekt³ zu literarischen Rezeptionsprozessen von jugendlichen Leserinnen und Lesern auf der Grundlage des Textes *Brudermord im Altwasser* von Georg Britting⁴ darstellen. Dabei ist es nicht das Ziel, das Forschungsprojekt in Gänze zu entfalten, sondern die Aspekte des Arbeitens zu erläutern, die durch die Orientierung an der Grounded Theory geprägt wurden. Die folgenden Ausführungen sind als Schlaglichter auf die Entwicklung eines Forschungsprojektes mit Hilfe der Grounded Theory-Methodologie gedacht. Auf eine breitere Einbettung in die bestehende Forschung zu literarischen Rezeptionsprozessen wird deshalb verzichtet.

2.1 Diffus teleologisches Vorgehen

Das Forschungsvorhaben hat sich durch den Kontakt zu den Phänomenen im Feld deutlich verändert. Zunächst wurde mit einem klar definierten Untersuchungsfokus auf ein bestimmtes Phänomen gestartet, der sich dann im Laufe des Prozesses deutlich verschoben hat, weil andere Phänomene in den Blick gerieten.

Ausgangspunkt der Planungen war die Lektüre und Auseinandersetzung mit Ergebnissen der Forschungsarbeiten von Tanja Janssen (2006; 2010), die literarische Verstehensprozesse mit der Methode des Lauten Denkens untersucht: Bei dieser Methode werden die Rezipienten während der Lektüre gebeten, alles, was ihnen durch den Kopf geht, zu verbalisieren, um daraus Rückschlüsse auf die bei der Lektüre aktiven Verarbeitungsprozesse zu gewinnen.⁵

Janssen hat zusammen mit Braaksma und Rijlaarsdam zwei Studien durchgeführt, die für die Konzeption meines eigenen Vorhabens zentral waren: 2006 untersuchte sie die Unterschiede zwischen starken und schwachen Schülerinnen und Schülern beim Lesen von literarischen Texten und 2010 den Effekt kreativen

³ Auch in diesem Forschungsprojekt stellt Teamarbeit im Sinne der Grounded Theory einen wesentlichen Faktor dar: Sowohl die Erhebung als auch die Kodierung und Auswertung der Daten wurde in Teamarbeit durchgeführt, diskutiert und weiterentwickelt. Das Team bestand aus drei Studierenden (Henrike Witte, Gregor Schipp und Alexandra Malukas) und mir.

⁴ Bewusst habe ich mich hier für eine Textgrundlage entschieden, die in der deutschsprachigen Erforschung von Rezeptionsprozessen schon mehrmals Beachtung fand (vgl. Winkler 2007; Kämper/Pieper 2008). Damit verbindet sich die Hoffnung, die eigenen Ergebnisse in bestehende Erkenntnisse einbetten zu können. Zudem eignet sich der Text *Brudermord im Altwasser* (Britting 1987, 20 [Erstveröffentlichung 1929]) aufgrund seiner Kürze und seiner Beschaffenheit, die genügend Probleme beim Lesen auf unterschiedlichsten Ebenen aufweist, besonders gut für eine Erhebung mit der Methode des Lauten Denkens (s. Dannecker in diesem Band).

⁵ Vgl. zur ausführlichen Darstellung der Chancen und Risiken dieser Erhebungsmethode: s. Dannecker in diesem Band und Stark 2010.

Schreibens auf das literarische Verstehen von Schülerinnen und Schülern. In beiden Fällen ist das methodische Vorgehen gleich: Es wird eine Gruppe von 50-60 Schülerinnen und Schülern im Alter von 15-16 Jahren dazu aufgefordert, parallel zu der Lektüre von kurzen erzählenden Texten laut zu denken. Die Äußerungen werden transkribiert und mit Hilfe eines neunstufigen Kodierschemas, das zuvor validiert wurde, quantitativ ausgewertet. Janssen vergleicht dann die Häufigkeit der gezeigten Leseaktivitäten zwischen den Gruppen – also einmal zwischen den starken und schwachen Schülerinnen und Schülern sowie das andere Mal zwischen der Gruppe von Probandinnen und Probanden, die zum Beginn der Lektüre einen kreativen Text zum Geschichtenanfang verfasst haben, und einer Gruppe von Probandinnen und Probanden, die den Text lediglich lesen.

Mein Ziel war es, eine ähnliche Forschungsarbeit zu konzipieren, diese allerdings in einem qualitativen Paradigma anzusiedeln und die Aussagen mit Hilfe von induktiv entwickelten Kodes im Stil der Grounded Theory auszuwerten. Das Forschungsvorhaben sollte sich an Janssens Studie zum Einfluss kreativen Schreibens auf den Leseprozess anschließen, da der Einsatz kreativer Methoden in der Schule zwar verbreitet, deren spezifischer Lerneffekt auf das Leseverstehen empirisch aber nicht belegt ist.

Wir reduzierten die Anzahl der Probandinnen und Probanden auf sechs Schülerinnen und Schüler eines Göttinger Gymnasiums im Alter von 15 Jahren, teilten sie in zwei Gruppen, von denen eine Gruppe eine kreative Schreibaufgabe bewältigte, bevor sie zur Lektüre der Erzählung übergang, und untersuchten im Anschluss daran die Laut-Denken-Protokolle, in der Erwartung, dass sich deutliche Unterschiede in den Verstehensprozessen zeigen würden, die sich dann mit Hilfe eines qualitativen Paradigmas genauer beschreiben ließen. Es kam aber anders.

Was uns überraschte, war die individuelle Bandbreite, die sich in den Laut-Denken-Protokollen zeigte. Die Probanden und Probandinnen äußerten jeweils sehr individuelle Verarbeitungsweisen beim Lesen des Textes. So reichten die mentalen Reaktionen z.B. auf den Textanfang von stark assoziativen Äußerungen und emotionalen Reaktionen

K2: „von wasserjungefern übersurt – bei wasserjungern ist dann irgendwie so ne meerjungfrau mit so roten haaren [...] eigentlich ne ganz schöne landschaft“

über Lesarten, die besonders die sprachliche Struktur wahrnahmen und bewerteten

E1: „langweilig beschrieben [...] also es wird ein ort auf jeden fall beschrieben – hm – also das ist der anfang einer – also die einleitung von einer geschichte“

bis zu Verarbeitungsweisen, in denen die lokale Ebene fast nicht verlassen wurde, und Wortbedeutungen thematisiert wurden

E 3: „mhm wasserjungfern ist mir nicht so bekannt – übersurrt – auch nicht so alltag [...] ich kenn mich mit fischen nicht so aus – gut dass es beschrieben wird“.⁶

Die Textgrundlage bietet für alle drei Leseweisen Anlässe: Zum einen wird in der Erzählung zu Beginn eine sehr unheilvoll-faszinierend-geheimnisvolle Atmosphäre aufgebaut und mit einer starken Bildlichkeit gearbeitet, die alle Sinne anspricht. Zum anderen werden in diesem einleitenden Teil noch keine Figuren eingeführt und der Satzbau ist hypotaktisch. Gerade diese sprachliche Struktur und das Fehlen einer Handlung machen das Lesen für Jugendliche evtl. schwierig und langweilig. Außerdem werden eine Reihe sehr spezieller Wörter verwendet, die sich auf das Sachgebiet des Wassers und der Fische beziehen. Die Textgrundlage konnte also bei den Schülerinnen und Schülern sehr unterschiedliche Reaktionen auslösen. Gleichzeitig – und das war es, was uns so überraschte – waren die gezeigten Reaktionen über die gesamte Länge der Laut-Denken-Protokolle sehr stabil. Da wir die Schreibaufgabe erst nach Auftreten der Figuren im Text angesetzt hatten, konnten wir sehr gut sehen, ob die Schülerinnen und Schüler sich vor und nach der Schreibintervention anders verhielten. Aber entgegen unserer Annahme änderte sich der Verarbeitungsstil kaum: Wer sich zuvor assoziativ geäußert hatte, blieb dabei – ebenso, wenn sich eine Person vor allem auf die sprachliche Struktur des Textes oder nur auf die semantische Oberfläche bezogen hatte (vgl. Malukas 2014, 61).

Nach dieser Erfahrung im Feld entschloss ich mich, meine Fragestellung zu ändern, und nicht mehr nach den Auswirkungen einer Schreibintervention auf den Rezeptionsprozess zu fragen, sondern zunächst mich den Phänomenen zuzuwenden, die zum Fragen Anlass gaben, nämlich den individuellen Zugangsweisen, die sich so stark unterscheiden und zugleich so habitualisiert zu sein scheinen, dass sie von einem kreativen Schreibimpuls wenig verändert werden.⁷ Genauer herauszufinden, worin sich diese Zugangsweisen unterscheiden, könnte dabei eine wesentliche Hilfe zur individuellen Modellierung von Verstehensprozessen im Literaturunterricht sein. Deshalb veränderten wir das Forschungsdesign und führten eine weitere Erhebung durch, in der wir auf den quasi-experimentellen Aufbau verzichteten und die Schülerinnen und Schüler lediglich zur Textgrundlage laut denken ließen. Insgesamt war unser eigenes Vorgehen im Sinne der Grounded Theory durch die im Feld gemachten Erfahrungen verändert worden.

⁶ Malukas 2014, Anhang. Zur besseren Lesbarkeit wurde bei allen Zitaten aus Transkripten auf die Wiedergabe von Pausenzeichen des Transkriptformates verzichtet.

⁷ Auch Janssen kann nur leichte Effekte nachweisen: „Our findings partly support the first hypothesis. [...] In particular, students who wrote a ‚story guess‘ showed more emotional engagement [...] than students who did not write. [...] Unlike we expected, writing as a prereading activity did not lead to more comprehension monitoring activities during reading.“ (Janssen/Braaksma 2010, 13)

2.2 Kontinuierliche Arbeit bei der Entwicklung von Kodes

Ein weiteres wichtiges Merkmal der Grounded Theory ist die Art und Weise, wie mit Kodes gearbeitet wird.⁸ Die Entwicklung von sinnvollen konzeptionellen Kodes stellt einen zentralen Arbeitsschritt in der Grounded Theory-Tradition dar und beansprucht viel Zeit, zählt sich aber insofern aus, als hier ein wesentliches Moment der Erkenntnisgewinnung liegt.⁹

Auch in unserem Forschungsprojekt kam der Entwicklung von Kodes und Kategorien eine zentrale Stellung zu. In einer ersten Phase wandten wir uns dem gesamten Material zu und kodierten es offen. Jedes Mitglied des Teams entwickelte seine bzw. ihre eigenen Kodes. Diese speisten sich aus dem Material und aus dem Vorwissen der Kodierenden. Strauss spricht in diesem Zusammenhang von natürlichen und soziologisch konstruierten Kodes – auf unser Projekt bezogen ist es sinnvoll, von natürlichen und lesepsychologisch konstruierten Kodes zu sprechen. So ging in die Kodeentwicklung ein, was wir durch die gängigen lesepsychologischen Modellierungen von z.B. van Dijk und Kintsch über den Prozess des Lesens wussten (vgl. Dijk/Kintsch 1983; Christmann 2010). Im Anschluss an diese Phase wurden aus den Kodes allgemeinere Kategorien entwickelt, die im Team diskutiert und so lange verändert wurden, bis sie zur Erfassung des gesamten Materials passend erschienen. Dann wurde das Datenmaterial erneut mit Hilfe der Kategorien kodiert. Das Ergebnis waren folgende Kodierkategorien¹⁰:

⁸ Hier unterscheidet die Grounded Theory-Methodologie sich deutlich von der qualitativen Inhaltsanalyse. In der Inhaltsanalyse geht es darum, einen möglichst hohen Grad an Objektivität zu erreichen, indem die Daten inhaltsnah reduziert und abstrahiert werden. In der Grounded Theory wird dagegen eine konzeptionalisierte Erschließung der Daten angestrebt, welche ein interpretatives Element beinhaltet (vgl. 1.2).

⁹ Vgl. auch Stark: „Insbesondere für *theorieentwickelnde* oder *hypothesengenerierende* Arbeiten ist die Entwicklung eines angemessenen Kategoriensystems eher das Ergebnis der Analyse als etwa eine ‚Vorarbeit‘ der eigentlichen Datenauswertung.“ (Stark 2010, 67).

¹⁰ Zusätzlich zu den genannten Kategorien gab es noch die Kategorie ‚Reste‘ für Äußerungen, die sich ansonsten nicht einordnen ließen, was allerdings äußerst selten der Fall war.

Bezeichnung	Kurzbeschreibung
1 Intertextualität	Bezugnahme auf andere Texte, Filme o.ä.
2 Lexikon	Klärung oder Äußerung von semantischen Fragen
3 Gattung	Bezugnahmen auf Gattungen
4 Fachsprache	Ausdrücke wie ‚Motiv‘ oder ‚Metapher‘ werden benutzt
5 Strukturen	Hypotaxe, Wiederholungen o.ä. werden bemerkt
6 Theory of mind	Äußerungen das Figurenbewusstsein betreffend
7 Emotion	Lachen, Seufzen
8 Erlebniserinnerung	Nennung von Erlebnissen aus dem eigenen Leben
9 Sinneswahrnehmungen	Visuelle, akustische, olfaktorische, haptische Assoziationen zum Text
10 Prozessregulation	Der eigene Leseprozess wird thematisiert.
11 Handlungsantizipation	Der Handlungsverlauf wird vorweggenommen.
12 Framework	Es werden holistische Deutungshypothesen aufgestellt.
13 Wiedergabe	Der Text wird z.T. mit eigenen Worten paraphrasiert.
14 Bewerten	Eine unter den Kategorien 1-13 gemachte Äußerung wird zusätzlich mit einer Bewertung verbunden.
15 Vergleichen	Zwei unter den Kategorien 1-13 gemachten Äußerungen werden in einem Vergleich aufeinander bezogen.
16 Lesen	Lautes Lesen

Tab. 4: Kodierkategorien „Brudermord im Altwasser“

Im Folgenden sollen kurz zwei Kodiersysteme vergleichend herangezogen werden, um zu zeigen, welche Überlegungen bei der Entwicklung unseres Kategoriensystems leitend waren.

Janssen setzt zur Analyse ihrer Daten ein Kategoriensystem ein, das auf Arbeiten von Andringa (1995) basiert: Retelling, Making inferences, Detecting problems, Making associations, Analysing, Evaluating, Responding emotionally, Responding metacognitively, Other activity (vgl. Janssen/Braaksma/Rijlaarsdam 2006, 40). Die von Janssen verwendeten Kategorien bilden Verstehensprozesse ab, d.h. sie beziehen sich auf die vermuteten mentalen Prozesse, die den im Laut-Denken-Protokoll gemachten Aussagen zugrunde liegen könnten.

Cornelia Rosebrock nutzt in ihrer Untersuchung zum *Erlkönig* eine ganz andere Art von Kategoriensystem. Sie greift auf die Klassifikation von Verstehensoperationen, die Grzesik (1983) entwickelt hat, zurück. Kennzeichnend ist hier die fast vollständige Beschränkung auf die Rekonstruktion der semantischen Ebene (vgl. Rosebrock 2008, 100) und das Ziel, verschiedene Komplexitätsgrade von Verstehensoperationen unterscheidbar zu machen.

Klasse 3: Vereinzelt Bedeutungen werden generiert [...] Klasse 5: Bedeutungen werden miteinander verknüpft [...] Klasse 6: Aus verschiedenen Bedeutungen werden weitere generiert [...] Klasse 7: Punktuelle Synthetisierung bzw. Zusammenhänge verschiedener Bedeutungen [...]. (ebd.).

Die Vorteile der hier verwendeten Kategorien sind der hohe Abstraktheitsgrad, die Möglichkeit, Verstehensleistungen zu hierarchisieren und die Beschränkung auf in den Äußerungen sichtbare semantische Marker. Gleichzeitig wird auch deutlich, dass bestimmte Phänomene in den Daten innerhalb eines solchen Kategoriensystems nicht sichtbar gemacht werden können (was auch nicht der Anspruch desselben ist), da emotionale oder assoziative Äußerungen hier keine eigene Kategorie zugewiesen bekommen. Janssens Kategorien bilden dagegen viel stärker Verstehensoperationen unterschiedlichster Art ab, insofern sie sowohl emotionale, metakognitive als auch rein abstrakt-verbale Prozesse sichtbar machen. Das für unser Forschungsprojekt entwickelte Kategoriensystem zeigt eine deutliche Nähe zum System von Janssen, da in unserer ersten Begegnung mit den Phänomenen im Feld assoziative und emotionale Reaktionen auf den Text oder das Fehlen derselben besonders hervorstechend waren. Insofern war es notwendig, ein Kategoriensystem zu entwickeln, das diesen Bereich von Leserreaktionen auch sichtbar und somit erforschbar macht. Deshalb wurde der assoziative Kategorienbereich für unser Forschungsvorhaben sogar noch weiter ausdifferenziert, indem hier auch persönliche Erfahrungen, die zum Text assoziiert werden, in der Kategorie ‚Erlebniserinnerungen‘ erkennbar werden. Insgesamt zeigt sich, dass das von uns entwickelte Kategoriensystem noch sehr differenziert und vielfältig ist. Im folgenden Arbeitsschritt ging es deshalb im Sinne der Grounded Theory darum, sich auf bestimmte Phänomene und den damit verbundenen Kategorien zu beschränken, um das Material, nachdem es in seiner Breite erschlossen worden war, nun durch ‚Tiefenbohrungen‘ zu untersuchen. Zunächst aber mussten die Ergebnisse unserer Kodierung ausgewertet werden.

2.3 Erste Ergebnisse und Ausblick

Bei der Auswertung der Daten zeigte sich, dass tatsächlich nur einige der Kategorien für unseren Untersuchungsfokus, der Frage nach individuellen Zugangsweisen und deren signifikanten Merkmalen, wichtig waren.

Als ersten Auswertungsschritt sahen wir uns nur die Verteilung der gezeigten Leseaktivitäten im Material an.¹¹ Wir werteten für jede Person die Häufigkeit der von ihr verwendeten Kategorien von Leseaktivitäten aus und verglichen sie miteinander, um zu sehen, ob sich Gemeinsamkeiten und Unterschiede zwischen den Probandinnen und Probanden zeigten.

Da die Menge der Äußerungen bei Laut-Denken-Protokollen sehr schwankt, ist die Aussagekraft der Daten pro Fall unterschiedlich groß. Bei Personen, die sich

¹¹ An dieser Stelle wurden Elemente der qualitativen Inhaltsanalyse (Einsatz eines Kodierleitfadens und Auszählung der Häufigkeit) in die Auswertung der Daten integriert.

viel geäußert hatten (60-110)¹², konnten wir leicht sehen, wo die Schwerpunkte des individuellen Leseverhaltens lagen, während unsere Beobachtungen betreffend der Personen, die sich sehr sparsam äußerten (20-40), eher extrapolierenden Charakter haben.

Eingedenk dieser Tatsache und der sehr kleinen Zahl an Probandinnen und Probanden ist es nur begrenzt möglich, mit ‚Häufigkeiten‘ zu arbeiten. Insofern dienen uns die beobachteten Häufigkeiten nicht so sehr als *Beweis von*, sondern vielmehr als *Hinweis auf* bestimmte Qualitäten im Leseprozess, die es dann genauer zu untersuchen gilt. Im Folgenden stelle ich zwei von vier Gruppen vor. Alle Fallgruppen haben gemeinsam, dass sie zwei bestimmte Kategorien als die häufigsten zeigen und gleichzeitig, dass sie andere – bei den anderen Gruppen vertretene Kategorien – gar nicht oder sehr selten aufweisen.

Eine erste Gruppe zeichnet sich dadurch aus, dass sie den höchsten Wert in der Kategorie ‚Wiedergabe‘ hat, kombiniert mit sehr hohen Werten in der Kategorie ‚Theory of mind‘. Eine zweite Gruppe von Probanden erreicht in den Kategorien ‚Emotion‘ und ‚Bewerten‘ die höchsten Werte. Interessant dabei ist, dass beide Gruppen sehr trennscharf sind: Während die Schülerinnen Nina, Naomi und Carolin¹³ zwischen 10-18 Äußerungen in der Kategorie ‚Emotion‘ zeigen – was genau der Menge von Äußerungen entspricht, die die Jungen (Leon, Christian, Tim) in der Kategorie ‚Wiedergabe‘ machen – äußern die drei Jungen keine einzige beobachtbare emotionale Reaktion (kein Lachen, kein Seufzen, kein Erschrecken).

Beispielhaft lassen sich die Merkmale der beiden Gruppen an den folgenden Reaktionen der Probanden Christian und Naomi auf eine Textstelle verdeutlichen. In der Textstelle wird erzählt, wie der kleinste von drei Brüdern beim gemeinsamen Spiel auf einem Weiher von den Brüdern erschreckt wird, ins Wasser fällt und ertrinkt.

Christian: „und dann scheint er irgendwie ertrunken zu sein [...] was ähm [...] vorhersehbar war anhand des Titels“

Naomi: „OOOH NEIN OOH NEIN DAS FIND ICH JETZ BÖSE NEEIN [seufzt] NEEIN DAS WILL ICH NICH [...] nein da denk ich an soon äähm [...] diverse kindergeschichten wo irgendwelche kinder ins wasser fallen und dann und dann liegen sie da so wie so wie beim toten mann dann da so im wasser [...] und dann haare so im wasser [...] dann so mit dem gesicht nach unten im wasser und das wasser is so ganz dunkel und ganz verlassen [...] OH NEEIN UND ER SCHRIE NICH MEHR OH NEIN DAS FIND ICH FÜRCHTERLICH“

¹² Die Zahlen beziehen sich auf die Anzahl von Äußerungen pro Laut-Denken-Protokoll. Die Äußerungen bildeten die Grundeinheit für die Zuordnungen von Kodes, wobei wir eine Äußerung im Sinne der Logik bzw. Linguistik als Proposition auffassten, d.h. als kleinste Wissenseinheit, die eine selbständige Aussage bilden kann (vgl. Anderson 2007, 175).

¹³ Die Namen aller Schülerinnen und Schüler wurden anonymisiert.

Während Christian tatsächlich nicht mehr zu dem Textabschnitt sagt, als das obige Zitat wiedergibt, sieht man bei Naomis Äußerung, dass sie nicht nur emotional und wertend auf die Textstelle reagiert, sondern auch die Szene vor ihrem inneren Auge imaginiert. Vermutlich ist ihre emotionale Reaktion eine Reaktion auf das innere Bild, das sie vor sich sieht.

Hier wird zugleich ein weiterer wichtiger Unterschied deutlich: Die Laut-Denken-Protokolle der drei Jungen weisen entweder keine (Christian und Leon) oder sehr wenige (Tim) Äußerungen auf, die in die Kategorien ‚Sinneswahrnehmungen‘ oder ‚Erlebniserinnerung‘ eingeordnet werden können, während die Protokolle von Naomi und Nina hier ebenfalls hohe Werte erreichen.¹⁴ Wollte man die Art von Naomi und Nina zu lesen auf eine zusammenfassende Formel bringen, könnte man davon sprechen, dass sie oft zum Text assoziieren, entweder so, dass sie sich Szenen sinnlich vorstellen oder sie mit eigenen Erlebnissen verbinden.

Christian, Leon und Tim dagegen entfernen sich mit ihren Äußerungen wenig vom Text: Sie haben entweder schon mit Problemen auf der lokalen Ebene zu kämpfen oder bringen ihre Leseindrücke auf eine nur wenig abstrahierende Deutungsformel und sehr allgemeine Lebensweisheiten, die nah an einer Paraphrase der Textoberfläche sind. So äußern sich Leon und Tim zu der oben genannten Textpassage des Unfalltodes des kleinsten Bruders:

Tim: „die beiden brüder sind noch jung begreifen erst später was sie getan haben“

Leon: „ja jetzt hört man ja nichts mehr von den das heißt wahrscheinlich dass sie unbewusst sind oder vielleicht sogar auch schon tot oder ertrunken“

An Leons Äußerung wird deutlich, dass ihm nicht klar ist, dass hier nur von einer Person die Rede ist, die ertrinkt, und er damit beschäftigt ist, überhaupt ein basales Textverständnis herzustellen. Tim dagegen verbindet den Textinhalt mit seinem allgemeinen Weltwissen über Kinder, um sich das erschreckende Verhalten der zwei Brüder, die ihrem kleinen Bruder nicht aus dem Wasser helfen, zu erklären. Seine Äußerung klingt sehr lapidar und allgemein, fast schon etwas zu abgeklärt für einen 15-jährigen. Ebenso lapidar äußert sich Christian: „was ähm [...] vorsehbar war anhand des Titels“. Es wirkt fast so, als würde hier das Verstörend-Unerklärliche der Geschichte etwas vorschnell domestiziert, indem es in gängige Schemata eingeordnet wird.

Das Forschungsprojekt steht nun an einer Stelle, an der wieder Entscheidungen im Sinne der Grounded Theory getroffen werden müssen. Es stellt sich die Frage, unter welcher Perspektive die unterschiedlichen Zugangsweisen genauer befragt werden sollen. Denn mit der Perspektive verbindet sich die Auswahl bzw. Entwicklung eines passenden Kodierparadigmas für die ‚Tiefenbohrungen‘ im Material.

¹⁴ Carolin unterscheidet sich hier von Naomi und Nina, da sie nur eine einzige Sinneswahrnehmung äußert.

Mit Blick auf die oben geschilderten Beispiele kann gesagt werden, dass durchaus divergierende Fragestellungen sinnvoll wären: So fällt die Genderproblematik innerhalb der beschriebenen Gruppen auf. In der ersten Gruppe sind nur Jungen, in der zweiten Gruppe nur Mädchen vertreten. Die Äußerungen passen gut zu stereotypen gesellschaftlichen Vorstellungen darüber, wie Jungen bzw. Mädchen auf emotionale Situationen reagieren sollten. Es könnte also nach der Rolle von geschlechterstereotypischen Momenten bei den individuellen Zugangsweisen gefragt werden. Als nächstes zeigt sich, dass in den Leseaktivitäten der ersten Gruppe gerade die Verstehensoperationen gehäuft vorkommen, die Janssen in ihrer Studie von 2008 als typisch für schwache Leserinnen und Leser anführt, nämlich die hohe Zahl an Retelling-Operationen, während sich in der zweiten Gruppe die Äußerungen häufen, die laut Janssen typisch für starke Leserinnen und Leser sind (,Emotion‘ und ,Bewerten‘) (vgl. Janssen/Braaksma/Rijlaarsdam 2006, 42). Es könnte also auch der Frage nachgegangen werden, ob und wie sich hier verschiedene Niveaus von Textrezeption ausdrücken. Als Letztes fällt auf, dass sich die beiden Gruppen darin unterscheiden, wie stark sich die Äußerungen von der Textoberfläche entfernen und wie stark diese Entfernung in den Äußerungen markiert ist. Es könnte gefragt werden, ob sich darin unterschiedliche mentale Operationen bzw. Strategien der Verarbeitung ausdrücken. Aus jeder der drei Fragerichtungen würde sich ein anderes Kodierparadigma für die weitere Arbeit ergeben und somit im Sinne der Grounded Theory eine andere theoretische Rahmung für die eigene Untersuchung.

Literatur

- Anderson, John R.: Kognitive Psychologie. Heidelberg: Springer 2007.
- Alheit, Peter: Grounded Theory. Ein alternativer Rahmen für qualitative Forschungsprozesse. (1999). http://www.global-systems-science.org/wp-content/uploads/2013/11/On_grounded_theory.pdf (letzter Zugriff: 01.08.2018).
- Andringa, Els: Strategien bij het lezen van literatuur. Spiegel (1995), 13, 3, 7-33.
- Britting, Georg: Brudermord im Altwasser. In: Ders: Sämtliche Werke. Prosa. Hrsg. von Wilhelm Haef. Bd. 3/2. München: List Verlag (1987), 20.
- Christmann, Ursula: Lesepsychologie. In: Kämper-van den Boogaart, Michael/Kaspar H. Spinner (Hrsg.): Deutschunterricht in Theorie und Praxis (DTP). Bd. 11/1: Lese- und Literaturunterricht. Teil 1: Geschichte und Entwicklung. Konzeptionelle und empirische Grundlagen. Baltmannsweiler: Schneider Verlag Hohengehren 2010, 148-200.
- Dijk, Teun van/Kintsch, Walter: Strategies of discourse comprehension. New York: Academic Press 1983.
- Grzesik, Jürgen: Methoden zur Deskription des Textverstehens im Unterricht: Zum Problem der Kategorisierung von Textverstehensleistungen im Verlauf und im Resultat des Unterrichts. In: Spiel (1983), 2, 1, 73-99.
- Janssen, Tanja/Braaksma, Martine/Gert Rijlaarsdam: Literary reading activities of good and weak students: A think aloud study. In: European Journal of Psychology of Education (2006) 21, 1, 35-52.

- Janssen, Tanja/Braaksma, Martine: Effects of creative writing on students' literary response to short stories. Paper presented at the 12th International Conference of the EARLI Special Interest Group on Writing, Heidelberg 8-10 september 2010.
- Kämper-van den Boogaart, Michael/Pieper, Irene: Literarisches Lesen. In: Didaktik Deutsch Sonderheft (2008), 46-65.
- Malukas, Alexandra: Verbessert kreatives Schreiben das Leseverstehen? Eine qualitative Untersuchung zum Einfluss des kreativen Schreibens auf das Leseverstehen bei Schülerinnen und Schülern einer 10. Klasse. Masterarbeit ‚Master of Education‘ Lehramt an Gymnasien an der Georg-August- Universität Göttingen. Göttingen 2014. (unveröffentlicht)
- Mey, Günter/Mruck, Katja (Hrsg.): Grounded Theory Reader. Köln: VS Verlag 2011.
- Riemann, Gerhard: Grounded theorizing als Gespräch: Anmerkungen zu Anselm Strauss, der frühe Chicagoer Soziologie und der Arbeit in Forschungsstätten. In: Mey, Günter/Mruck, Katja (Hrsg.): Grounded Theory Reader. Köln: VS Verlag 2011. 405-426.
- Rosebrock, Cornelia: Literarische Erfahrung mit dem Erklären. Ein Blick auf die Prozessebenen des Lesens. In: Härle, Gerhard/Rank, Bernhard: „Sich bilden, ist nichts anders, als frei werden.“ Sprachlich und literarische Bildung als Herausforderung des Deutschunterrichts. Baltmannsweiler: Schneider Verlag Hohengehren 2008, 89-110.
- Stark, Tobias: Zum Perspektivenverstehen beim Verstehen literarischer Texte: Ausgewählte Ergebnisse einer qualitativen Untersuchung. In: Pieper, Irene/Wieser, Dorothee (Hrsg.): Fachliches Wissen und literarisches Verstehen. Studien zu einer brisanten Relation. Frankfurt a.M.: Peter Lang 2012, 153-169.
- Strauss, Anselm/Corbin, Juliet: Grounded Theory: Grundlagen Qualitativer Sozialforschung. Weinheim: Beltz 1996.
- Strauss, Anselm L.: Grundlagen qualitativer Sozialforschung. Datenanalyse und Theoriebildung in der empirischen soziologischen Forschung. München: Fink 1998.
- Strübing, Jörg: Anselm Strauss. Konstanz: UVK 2007.
- Strübing, Jörg: Grounded Theory. Zur sozialtheoretischen und epistemologischen Fundierung des Verfahrens der empirisch begründeten Theoriebildung. Wiesbaden: VS Verlag 2008.
- Winkler, Iris: Welches Wissen fördert das Verstehen literarischer Texte? Zur Frage der Modellierung literarischen Wissens für den Deutschunterricht. In: Didaktik Deutsch (2007), 13, 22, 71-88.

Dokumentarische Methode

Einführung in eine Methodologie

1. Vorstellung der dokumentarischen Methode

Ziel dieser Ausführung ist es, die dokumentarische Methode als Methodologie vorzustellen. Damit verbunden ist die Position, dass es sich bei der dokumentarischen Methode nicht um ein Werkzeug handelt, das immer auf dieselbe Weise eingesetzt wird. Vielmehr liefert die dokumentarische Methode grundlegende Interpretationsschritte, die auf die jeweilige Forschungsfrage und auf das vorliegende Datenmaterial angepasst werden müssen. Daher ist die dokumentarische Methode weniger als ‚Werkzeug‘ und viel mehr als ‚Rezept‘ zu verstehen. Diese Form des nicht-standardisierten Vorgehens ist für eine Methode aus dem Bereich der qualitativ-rekonstruktiven Sozialforschung (Bohnsack/Marotzki/Meuser 2011, 140ff.; Bohnsack 2008) typisch.

Die dokumentarische Methode hat sich seit über zwanzig Jahren erfolgreich in der empirischen Sozialforschung und damit einhergehend in der bildungswissenschaftlichen Forschung etabliert. Die Forschungsgebiete sind vielfältig und decken unter anderem Jugendforschung, Schulforschung, fachdidaktische Forschung, Milieuforschung, Gender-/Beziehungs- und Paarforschung, Organisationsforschung, Erwachsenenbildung und vieles mehr ab.

Die dokumentarische Methode lässt sich den Methoden rekonstruktiv-qualitativer Sozialforschung (s. hierzu auch Heller/Morek und Pflugmacher in diesem Band) zuordnen. Rekonstruktive Verfahren zielen darauf ab, vom explizit vorliegenden empirischen Material auf dahinterliegende implizite Muster zu schließen, bzw. diese rekonstruieren zu wollen. Mit ihr verschiebt sich das Forschungsinteresse: Der Inhalt (‚Was‘) tritt in den Hintergrund, wodurch sogenannte Orientierungsmuster (‚Wie‘) in den Blick genommen werden können. Ein Wechsel vom Was zum Wie ist typisch für rekonstruktive Verfahren qualitativer Sozialforschung. Die dokumentarische Methode bezieht sich hierbei vor allem auf die ‚Wissenssoziologischen Überlegungen‘ von Mannheim. Diese Muster sind der Rahmen für Handlungen und lassen sich methodisch kontrolliert zu Typen verdichten. So können mit der Methode durch mehrdimensionale, komparative Vergleiche Typen herausgearbeitet werden. Hierzu werden Vergleiche sowohl innerhalb der einzelnen Fälle wie auch zwischen den Fällen angestellt. Grundlegend hierfür ist die

Unterscheidung zwischen kommunikativem und konjunktivem Wissen (Mannheim 1952; 1996). Kommunikatives Wissen entspricht explizitem, abrufbarem (lexikalischem) Wissen. Im Gegensatz dazu entspricht konjunktives Wissen einem implizit und konjunktiv geteilten Wissen. Dieses konjunktive Wissen ist den einzelnen Personen meist nicht bewusst und ergibt sich aus einer gemeinsam geteilten Handlungspraxis.¹ Teilen sich Personen konjunktives Wissen, verstehen sie sich (im Hinblick auf diesen Wissensbereich) auf dieser Grundlage. Im Terminus der dokumentarischen Methode teilen sich diese Personen konjunktive Erfahrungsräume. Im Fall der dokumentarischen Methode geht es darum, handlungsleitende Orientierungsmuster, die sich im vorliegenden empirischen Material *dokumentieren*, zu rekonstruieren. In Anlehnung an Bourdieu (1970, 125ff.) wird oft auch von der ‚Rekonstruktion eines Habitus‘ gesprochen. Ziel ist es, durch komparative Vergleiche Typen zu rekonstruieren, mit denen beschreibbar wird, auf welche kollektiv-konjunktiven Erfahrungsräume die Probandinnen und Probanden zurückgreifen, um sich zu orientieren.

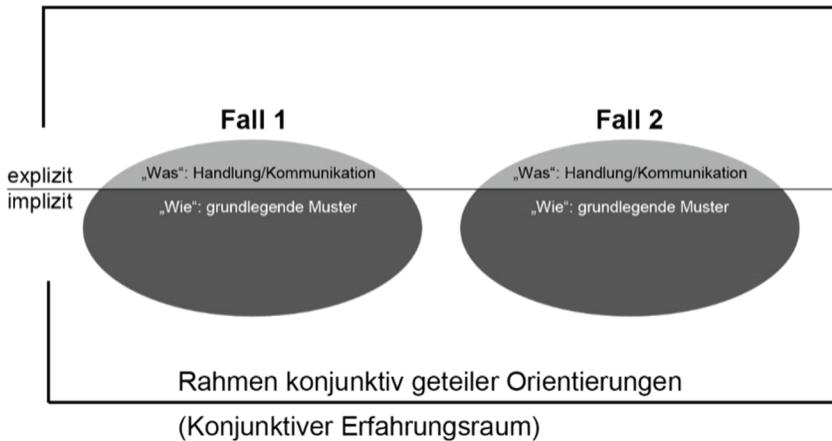


Abb. 1: Konjunktive Erfahrungsräume als impliziter Rahmen
(Quelle: eigene Abbildung)

Entwickelt wurde die Methode maßgeblich von Bohnsack, der in den 1980er Jahren zusammen mit Mangold vor der Herausforderung stand, Daten von Gruppendiskussionen auszuwerten (Mangold/Bohnsack 1988).² Aus diesem Forschungskontext heraus wurden die der Methode zugrunde liegenden Paradigmen entwickelt (vgl. hierzu Bohnsack 1989). Anfänglich wurden vorwiegend die Daten aus Gruppendiskussionen mittels dieses methodischen Repertoires analysiert. Schnell

¹ So kann angenommen werden, dass z.B. Mütter über ein solches konjunktives Wissen verfügen, dass ihnen in der Regel nicht bewusst ist, das aber dazu führt, dass sie sich untereinander in bestimmten Bereichen verstehen (dort wo andere interpretieren müssten).

² Bereits davor wurde das Verfahren der Gruppendiskussion in der deutschen Debatte der empirischen Sozialforschung maßgeblich von Mangold (1960; 1973) geprägt (s. auch Scherf in diesem Band).

entwickelte sich die Methode weiter und konnte so für die Interpretation von Interviews eingesetzt werden. Seit ein paar Jahren wird mit der dokumentarischen Methode ebenfalls visuelles und audiovisuelles Material interpretiert (Bohnsack 2009). Prinzipiell kann jede Form empirischen Materials mit der dokumentarischen Methode interpretiert werden. Von Bedeutung hierbei ist sowohl, dass die Methode an das Material angepasst werden muss und dass kritisch zu prüfen ist, ob eine Anwendung sinnvoll ist, bzw. durch die Form der Datenerhebung auch Daten generiert werden können, die sich interpretieren lassen (beispielsweise ein propositionaler Gehalt nicht vorgegeben, sondern von der Probandin bzw. vom Probanden eingebracht werden kann). Das methodische Vorgehen unterscheidet sich dabei, je nachdem auf welche Art des Materials zurückgegriffen wird. So verläuft die Interpretation von Gruppendiskussionen anders als die von Interviews. Ebenso kommen neue Interpretationsschritte bei Bild- und Videointerpretationen hinzu. Wie dynamisch die Methode immer noch weiterentwickelt wird, zeigt sich auch in einer neuen Form der Typenbildung, die Nohl (2014) mit der relationalen Typenbildung vorgelegt hat.

Heute gilt die dokumentarische Methode als eines der elaboriertesten Verfahren der qualitativ-rekonstruktiven Sozialforschung. Auch deshalb nimmt sie innerhalb der erziehungswissenschaftlichen Forschung heute eine wichtige Rolle ein. Die dokumentarische Methode wird zunehmend auch in der empirischen Forschung in den Fachdidaktiken verwendet. Auch wenn aus der Deutschdidaktik bisher nur einzelne Arbeiten hervorgegangen sind³, lässt sich mit Blick auf die Tendenzen innerhalb der Nachbardisziplinen vermuten, dass auch innerhalb der empirischen deutschdidaktischen Forschung Arbeiten mit der dokumentarischen Methode zunehmen werden (vgl. hierzu Kap. 2.2).

1.1 Funktionsweise der dokumentarischen Methode

Eine wichtige Aufgabe des Forschenden ist es, eine klare Passung zwischen Fragestellung und methodischem Vorgehen herzustellen. Hierzu kann es notwendig sein, wie bei der dokumentarischen Methode, das methodische Vorgehen an die Forschungsfrage und das empirische Material anzupassen. Die dokumentarische Methode kann daher nicht nur als Methode (im Sinne eines Werkzeugs), sondern auch als Methodologie verstanden werden. Als Methodologie (Theorie über das methodische Vorgehen) werden die grundlegenden Ausgangspunkte (Paradigmen) skizziert, auf deren Grundlage sich ein angemessenes forschungsmethodisches Vorgehen ableiten lässt. Die Datenauswertung der dokumentarischen Methode verläuft in mehreren Schritten. Mittels einer *formulierenden Interpretation* wird der explizite Gehalt des Gesagten zusammenfassend wiedergegeben. Hier wird die Sache, das ‚was‘ gesagt wird, fixiert. Die *reflektierende Interpretation* beschreibt, ‚wie‘ ein Thema behandelt wird. Hierzu wird z.B. bei der Interpretation von Gruppendiskussionen der Verlauf der Diskussion mit einer Diskursbe-

³ Für einen Überblick über die aktuellen Arbeiten innerhalb der Deutschdidaktik vgl. die Zusammenschau von Ballis et al. (2014).

schreibung nachgezeichnet. Schließlich erfolgen verschiedene Arten der *Typenbildung*. Die Basis hierfür ist immer eine *sinngenetische Typenbildung*. Darüber hinaus sind auch *soziogenetische und relationale Typenbildungen* möglich. Die Interpretationsschritte sind im praktischen Auswertungsprozess zirkulär miteinander verschränkt. Es kann also sein, dass nach der reflektierenden Interpretation einer Passage, der Blick auf andere (bis dahin weniger beachtete) Passagen fällt und diese zuerst formulierend und dann reflektierend interpretiert werden. So arbeitet sich die dokumentarische Methode durch Passagen des empirischen Materials, bis eine Typenbildung soweit möglich ist, dass weitere Passagen diesen Typen zugeordnet werden können. Im Folgenden werden die Grundlagen der methodischen Schritte dargestellt.

Formulierende Interpretation

Die formulierende Interpretation reformuliert den „wörtlichen oder, 'immanenten' Sinngehalt“ (Bohnsack 2008, 134). Sie legt somit zum einen das Textverständnis des Forschenden offen. Zum anderen bildet sie den Ausgangspunkt, um sich im zweiten Interpretationsschritt (der reflektierenden Interpretation) vom kommunikativen Wissen zu entfernen und den Fokus auf die Rekonstruktion des konjunktiven Wissens zu legen. Die formulierende Interpretation beinhaltet folgende Schritte:

- a) Der thematische Verlauf wird durch die Identifizierung von Überschriften (Ober- und Unterthemen) beschrieben (Beispiel hierzu in: Bohnsack/Nentwig-Gesemann/Nohl 2007, 314).
- b) Anschließend werden die Passagen ausgewählt, die für die reflektierende Interpretation verwendet werden. Kriterien für die Auswahl sind die thematische Relevanz, aber auch die interaktive Dichte im Diskurs und metaphorische Dichte des Materials (Bohnsack 2007, 233; 2008, 140).

Reflektierende Interpretation

Im zweiten Interpretationsschritt wird der ‚immanente Sinngehalt‘ vom ‚dokumentarischen Sinngehalt‘ (Bohnsack/Nentwig-Gesemann/Nohl 2007, 14) abgegrenzt. Die reflektierende Interpretation umfasst zwei Schritte. Zum einen wird durch die Analyse der Diskursorganisation der propositionale Gehalt identifiziert. Zum anderen zielt sie auf eine sinngenetische Typenbildung. Die Diskursorganisation beschreibt die „intertextuelle Einheit aus Proposition, Elaboration und Konklusion“ (Vogd 2005, 34). Mit ihr verschiebt sich die Interpretationsperspektive vom explizit Geäußerten hin zu dahinterliegenden impliziten Erfahrungsräumen. Mit der reflektierenden Interpretation wird der „(Orientierungs-)Rahmen“ (Bohnsack 2008, 135) rekonstruiert. Dieser Rahmen ist die Grundlage für die sinngenetische Typenbildung und wird durch die Identifikation von Vergleichshorizonten mithilfe eines sequenzanalytischen Vorgehens beschreibbar. Diese Suche nach „Regelhaftigkeit“ (Bohnsack/Nohl 2007, 303) richtet sich dabei nach „homologen, funktional äquivalenten [...] Reaktionen“ (ebd., 304) innerhalb der erhobenen Fälle. Die reflektierende Interpretation legt den Grundstein für die sinngenetische Typenbildung.

Typenbildung

Es lassen sich unterschiedliche Formen der Typenbildung bei der dokumentarischen Methode anstreben. Der Vergleich ist in allen Fällen das primäre methodische Instrument. „Der Kontrast in der Gemeinsamkeit ist fundamentales Prinzip der Generierung einzelner Typiken und ist zugleich die Klammer, die eine ganze Typologie zusammenhält“ (Bohnsack 2008, 143). Durch die Zusammenschau möglichst unterschiedlicher Fälle werden Typen generiert. Mit der dokumentarischen Methode kann auf drei verschiedene Arten eine Typenbildung vollzogen werden: sinngenetische Typenbildung, soziogenetische Typenbildung und komparative Typenbildung. Generell folgt bei der Typenbildung auf der Grundlage einer „Ausgangs- oder Basistypik“ (Bohnsack 2007, 237) eine fallübergreifende Analyse. Hier wird Material aus verschiedenen Fällen für die Interpretation hinzugezogen. Die fallübergreifende Analyse wird um eine fallinterne Analyse ergänzt. Dabei werden entsprechend Stellen innerhalb eines Falls für die Interpretation verwendet. Sowohl im fallinternen, wie auch im fallexternen Vergleich ist das Ziel, durch den komparativen Vergleich auf ein ‚Tertium Comperationis‘⁴ (Nohl 2007, 263ff.) zu schließen. Die *sinngenetische Typenbildung* beschreibt die konjunktiven Orientierungsmuster im Material. Diese Typen entstehen aus dem Material heraus. Mit der *soziogenetischen Typenbildung* werden die zugrunde liegenden kollektiven Erfahrungsräume mithilfe gesellschaftlicher Kategorien rekonstruiert. Diese Typen werden anhand sozialer Kategorien (Alter, Geschlecht, usw.) gebildet, die an das Material herangelegt werden. Durch die *relationale Typenbildung* wird deutlich, in welchem „systematischen Zusammenhang“ (Nohl 2013, 61) konjunktive Erfahrungsräume und Orientierungen stehen. Der relationale Vergleich innerhalb des Datenmaterials kann dazu führen, dass sich aus der Zusammenschau nicht nur von verschiedenen Fällen, sondern von mehrdimensionalen Falldimensionen Typen rekonstruieren lassen. In der Regel zielt die Interpretation zumindest auf die *Sinngese* ab. Klassischerweise kann daran eine *Soziogenese* angeschlossen werden. Hier besteht die Herausforderung, genug Ressourcen für mögliche Nacherhebungen aufwenden zu können um mögliche Kontrastfälle für die soziographischen Kategorien zu sammeln. Alternativ (oder ergänzend) kann eine relationale Typenbildung erfolgen. Sie ist vor allem dann angebracht, wenn zu vermuten ist, dass sich die konjunktiven Erfahrungen (z.B. auf Grund kultureller Distanz) vom Interpretierenden nicht durch Kategorien (wie bei der Soziogenese) an das Material heranlegen lassen können.

⁴ ‚Tertium Comperationis‘ bezeichnet das Gemeinsame der Vergleichsfälle und ist somit immer auch abhängig von der gewählten Perspektive.

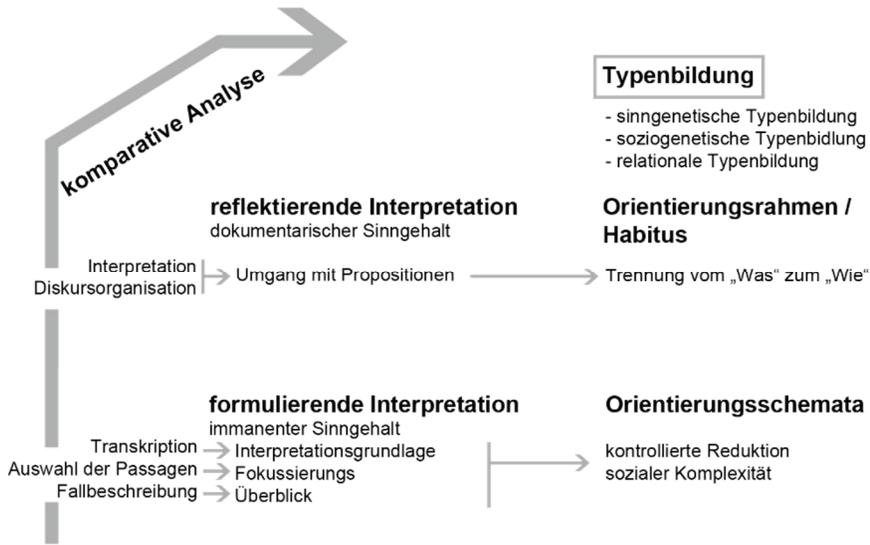


Abb. 2: Grundlegendes Muster der Interpretation mit der dokumentarischen Methode
(Quelle: eigene Abbildung)

Die Grafik stellt die Interpretationsschritte modellhaft dar. Sie zeigt, welche Ziele (rechte Seite) die formulierende und die reflektierende Interpretation verfolgen und welche Teilschritte (Mitte, kleine Pfeile) hierzu angewendet werden. Die Grafik entwickelt sich also von unten nach oben. Der große Pfeil auf der linken Seite verweist darauf, dass die komparative Analyse den gesamten Interpretationsprozess bis hin zur Typenbildung begleitet. Zu kurz kommt in dieser Abbildung a) dass es sich bei der Interpretation immer um einen zirkulären Prozess handelt, bei dem zwischen den verschiedenen Schritten gewechselt wird und b) dass die komparative Analyse Teil aller Interpretationsschritte ist (Bohnsack 2007, 253).

Ein Beispiel von Barbara Asbrand für die Illustration der einzelnen Interpretationsschritte am konkreten Material ist im Online-Fallarchiv der Universität Kassel zu finden.⁵

1.2 Ausgangsdaten und Ergebnisse

Wenn die vorangegangenen grundlegenden Schritte der Interpretation berücksichtigt werden, kann die dokumentarische Methode auf verschiedene Formen von Daten angewandt werden. Klassisch ist die Auswertung von Daten, die mit Gruppendiskussionen (Bohnsack/Przyborski/Schäffer 2010) erhoben wurden. Von zentraler Bedeutung sind hierbei zwei Faktoren. Zum einen sollte es sich bei den Gruppen um sogenannte natürliche Gruppen handeln. D.h. die Gruppen werden nicht künstlich (z.B. wie bei Experimenten) zusammengestellt, sondern die

⁵ <http://www.fallarchiv.uni-kassel.de/lernumgebung/dokumentarische-methode/>

Teilnehmenden der Gruppe haben auch außerhalb der Datenerhebungssituation etwas miteinander zu tun. Dies kann z.B. dadurch gefördert werden, dass sich die Lehrkräfte aus einem Kollegium selbst zu einer Gruppe für die Diskussion zusammenschließen. Zum anderen sollte auf die Selbstläufigkeit der Diskussion geachtet werden. Hierbei gilt es, als Diskussionsleitender Gesprächspausen auszuhalten, sich nicht in die Diskussion einzumischen und möglichst keine Fragen mit propositionalem Gehalt an die Teilnehmenden zu stellen (s. Scherf in diesem Band).

Auch die Interpretation von Interviews hat sich mittlerweile für Projekte etabliert, die mit der dokumentarischen Methode arbeiten (Nohl 2006). Meist wird hierzu, in Anlehnung an Schütz (1987) auf die Erhebung biographisch-narrativer Interviews (s. Lindow in diesem Band) zurückgegriffen. Für die Auswertung der Daten ist es wichtig, bereits bei der Datenerhebung darauf zu achten, dass die Interviewten in der Erhebungssituation möglichst frei erzählen können und möglichst keine Propositionen vom Interviewenden eingebracht werden. Bei der Interpretation von Interviews verändert sich die reflektierende Interpretation in dem u.a. eine Textsortentrennung in diesem Interpretationsschritt durchgeführt wird.

Seit ein paar Jahren werden zunehmend auch Bilder und Videos mittels der dokumentarischen Methode ausgewertet. Hierzu wird als zusätzliche Interpretationshilfe die Rekonstruktion der Planimetrie⁶ (Bohnsack 2009, 60f.) im Rahmen einer ikonischen Interpretation herangezogen. Welchen konjunktiven Erfahrungsräumen gilt das Forschungsinteresse? Dem des Abgebildeten, des Bildproduzenten oder einer möglichen Bildredaktion? Hierbei stellen sich Fragen der Abhängigkeit: Inwieweit kann der Abgebildete (z.B. eine Politikerin auf einem Foto als Schlagzeile einer Zeitung) die Art der Präsentation (Perspektivität, Bildschnitt, Fokus usw.) beeinflussen, bzw. wie unabhängig vom Abgebildeten lässt sich ein solches Bild inszenieren?⁷ Je nachdem richtet sich die Forschungsfrage z.B. eher auf die Darstellung von Macht durch die Medien oder aber auf die Selbstinszenierung von Politikerinnen und Politikern. Wenn visuelle oder audio-visuelle Daten selbst erhoben werden, ist es wichtig sicherzustellen, dass der Forschende methodisch kontrolliert die Aufnahme durchführt und so in einem möglichst geringen Maß seine eigene Perspektivität einbringt.

Auf Grund der paradigmatischen Grundlagen der dokumentarischen Methode erscheint es allerdings wenig sinnvoll, Daten für die Interpretation heranzuziehen, die klare propositionale Vorgaben machen (s. Schmidt in diesem Band⁸), bzw. die

⁶ Die Planimetrie bezeichnet die Analyse der Formalstruktur von Bildern anhand geometrischer Strukturen.

⁷ Der Prozess der Bildproduktion kann als Kommunikation (vergleichbar mit der Kommunikation in einer Gruppendiskussion) verstanden werden. Somit kann die Interpretation auf die konjunktiven Erfahrungen aller an der Bildproduktion Beteiligten abzielen.

⁸ Eine Ausnahme liegt dann vor, wenn erzählgenerierende Fragen gestellt werden (vgl. Nohl 2006, 23).

sich konjunktiven Erfahrungsräume von Beobachtendem und Beobachtetem nicht klar trennen lassen (s. Schwinghammer in diesem Band).

Mit der dokumentarischen Methode lassen sich unterschiedliche Formen empirischer Daten interpretieren. Die Methode ist daher vielfältig einsetzbar. Ergebnisse der dokumentarischen Methode werden als Typen dargestellt.

1.3 Vorteile und Nachteile des Vorgehens der dokumentarischen Methode

Das entscheidende Kriterium bei der Wahl für die dokumentarische Methode ist das Forschungsinteresse. Geht es darum, Orientierungen, bzw. habitualisierte Muster (kurz gesagt: konjunktive Erfahrungsräume) zu rekonstruieren, um davon ausgehend eine praxeologische Theorie⁹ zu generieren, ist die dokumentarische Methode ein adäquater Interpretationsansatz.

Die Methode ist trotz ihrer aufwändigen Art der Interpretation sehr effizient. Die Suchstrategien für die relevanten Passagen der Interpretation führen dazu, dass nicht das gesamte empirische Material für die Interpretation herangezogen wird und dennoch aus dem Material alle relevanten Stellen verwendet werden. Für eine nichtstandardisierte-rekonstruktive Methode kann die dokumentarische Methode als elaboriert gelten, da sich die Interpretation methodisch in hohem Maße selbst kontrolliert (vgl. Kap. 1.4). Auch führt die Kontrolle durch eine Forscherpeer-group zu einer vergleichsweise hohen Qualität der Interpretation. Hinzu kommt, dass die spezifische Sichtweise der dokumentarischen Methode eine alternative empirische Sicht auf soziale Wirklichkeit und den Strukturen menschlichen Handelns ermöglicht. Damit verbunden ist die Chance für eine Weiterentwicklung des pädagogisch-didaktischen Erkenntnisinteresses.

Allgemein können folgende vier Herausforderungen benannt werden, auf die es sich (unabhängig von der Forschungsfrage) einzulassen gilt, wenn mit der dokumentarischen Methode gearbeitet wird. Die Analyse des Materials mit der dokumentarischen Methode ist a) aufwändig. Die einzelnen Interpretationsschritte erweisen sich als recht umfangreich. Hinzu kommt, dass sich ausgehend von einer interpretierten Sequenz ein neuer Blick auf das Material ergibt, mit dessen Hilfe die nächste Interpretationsstelle gesucht wird. Dieses zirkuläre Vorgehen ist verhältnismäßig zeitintensiv. Das konkrete methodische Vorgehen ist in dieser Art kaum darstellbar und so werden die Interpretationen in Forschungsberichten (wie z.B. Qualifikationsarbeiten) nur exemplarisch illustriert – nicht etwa belegt. Methodenlehrbücher versuchen, diesen blinden Fleck zu beleuchten, doch gelingt dies unterschiedlich nachvollziehbar. Es lässt sich daher b) attestieren, dass das Erlernen des methodischen Vorgehens mit Literaturstudium (wie bei einigen anderen Methoden auch) nicht zu Genüge erlernt werden kann. Arbeitet man noch nicht lange mit rekonstruktiven Methoden, stellt sich c) der Perspektivwechsel

⁹ Die dokumentarische Methode verzichtet darauf, Theorien als Grundlagen für die Interpretation heranzuziehen. Die praxeologische Ausrichtung führt dazu, dass aus dem empirischen Material heraus Theorien konstruiert werden.

(auf das Implizit-Konjunktive) als ungewöhnlich dar. Notwendig daher ist d) die Rückbindung an eine Forscher-Community. Eine intersubjektive Kontrolle ist das zentrale Qualitätsmerkmal für eine Interpretation mit der dokumentarischen Methode. Hierbei ist die Teilnahme an einer Forschungswerkstatt (zur dokumentarischen Methode) zwingend notwendig. Aber auch andere Interpretationsgruppen, sowie die Teilnahme an Methodenworkshops und -tagungen sollten zusätzlich für die Validierung der eigenen Interpretation genutzt werden.

1.4 Gütekriterien für den Einsatz der dokumentarischen Methode

Eine Herausforderung für die qualitativ-rekonstruktive Sozialforschung ist, dass Gütekriterien zur Bestimmung der Qualität kaum explizit formuliert vorliegen. Allerdings entwickeln rekonstruktive Methoden (wie die dokumentarische Methode) aus ihrer Praxis heraus komplexe Standards zur Qualitätssicherung. Bohnsack (2005) stellt hierzu neun Thesen für Standards in der nichtstandardisierten Forschung auf (vgl. hierzu Bohnsacks Thesen zu Standards nicht-standardisierter Forschung; Bohnsack 2005, 63ff.).

Weitere vier spezifische Aspekte der Qualitätssicherung sind innerhalb der dokumentarischen Methode folgende:¹⁰

Zirkulärer Prozess

Die einzelnen Interpretationsschritte der dokumentarischen Methode vollziehen sich nicht strikt nacheinander. Ausgehend von einer Interpretationspassage und der sich daraus andeutenden Zwischenergebnisse wird innerhalb des Materials nach Kontrastfällen gesucht. Wo finden sich vergleichbare Orientierungen (Minimalkontraste) und wo völlig entgegengesetzte (Maximalkontraste)? Sowohl innerhalb eines Falls wie auch zwischen den einzelnen Fällen wird so nach der nächsten Passage für die weitere Interpretation gesucht. Daher wird im eigentlichen Interpretationsprozess immer wieder zwischen den Interpretationsschritten gewechselt. Der Interpretationsprozess tastet sich von einer relevanten Passage zur Nächsten und greift dabei zirkulär auf die bereits entwickelte Interpretation zurück. Die Vorteile dieses Vorgehens liegen auf der Hand. Zum einen wird das vorliegende Material nur soweit interpretiert, bis eine Sättigung vorliegt. Zum anderen kann so die Interpretation, durch die ständige Rückbindung an das Material durch die empirischen Daten selbst kontrolliert werden.

Interpretationsgruppen

Die Qualität der Interpretation hängt (wie bei vielen anderen Auswertungsverfahren) maßgeblich von dem Forschungsumfeld ab. Die regelmäßige Teilnahme an einer Forschungswerkstatt zur dokumentarischen Methode, oder einer Interpretationsgruppe, ist daher eine wesentliche Stellschraube der Qualitätssicherung im Forschungsprozess. Schon bei der Suche nach einer Forschungsfrage, der Planung und Gestaltung der Datenerhebung und letztlich für die intersubjektive Rück-

¹⁰ Die meisten Aspekte treffen auch auf andere rekonstruktive Auswertungsverfahren zu.

kopplung der Interpretationsschritte ist ein solches Arbeitsumfeld sinnvoll. Zumindest stellenweise sollte die eigene Interpretation in einem solchen Rahmen kritisch überprüft werden, um der Gefahr zu entgehen, seine eigene Orientierung in das Material hineinzuinterpretieren. Alternativ kann an Methodentagungen teilgenommen werden. Solche Veranstaltungen sind auch ideal, um Kontakte zu knüpfen und möglicherweise Anschluss an eine bestehende Forschungswerkstatt zu finden.

Praxeologische Theoriegenese

Bei der dokumentarischen Methode entwickelt sich die Theoriebildung aus dem vorliegenden Datenmaterial heraus. Diese Vorannahmen beinhaltet die Einschränkung, bei der Interpretation möglichst lange darauf zu verzichten, Fachtermini einfließen zu lassen. Würde z.B. im Rahmen der reflektierenden Interpretation bereits das Aushandeln bestimmter Propositionen einem Horizont pädagogischer Professionalität zugeschrieben, würden sich gleich zwei analytische Kurzschlüsse ergeben. Zum einen wird vorschnell eine normative Perspektive an das Material heran gelegt, zum anderen beinhaltet der Begriff ‚pädagogische Professionalität‘ bereits (je nach dem an welchem Autor sich orientiert wird) bestimmte Theorien und Modelle, die letztlich damit dann deduktiv an das Material heran gelegt werden würden. Ziel der dokumentarischen Methode ist es vielmehr aus dem Material heraus die handlungspraktischen und konjunktiv geteilten Theorien zu rekonstruieren und zu verdichten. Wird jedoch bei der Interpretation bereits auf wissenschaftliche Vorstellungen zurückgegriffen, besteht die Gefahr damit die im Datenmaterial vorhandenen Muster zu überdecken.

Typenbildung

Ziel der dokumentarischen Methode ist es, aus dem Material heraus Typen zu rekonstruieren. Eine sinngenetische Typenbildung ist somit das Minimalziel für die Interpretation mit dieser Methode. Eine soziogenetische Typenbildung erweist sich in den meisten Fällen als sehr umfassend in der Datenerhebung, weil die Suche nach Kontrastfällen auf Grundlage der soziographischen Kategorien ausgeweitet werden muss. Daher kann diese Form der Typenbildung für alleinstehende Qualifikationsarbeiten meist auch nur angedeutet werden. Sinnvoll wird eine Soziogenese meist erst bei größeren Forschungskoooperation, Projekten und Habilitationsarbeiten.

Fortbildungsmöglichkeiten (Auswahl):**Centrum für qualitative Evaluations- und Sozialforschung (CES)**

<http://www.ces-forschung.de>

In den kommerziell angebotenen Workshops wird in die methodischen Grundlagen eingeführt und intensiv in kleinen Gruppen an (von den Teilnehmerinnen und Teilnehmern eingebrachtem) Material gearbeitet (Die Teilnahmegebühr beträgt ca. 70-130 Euro).

Zentrum für Sozialforschung und Methodenentwicklung (ZSM)

<http://www.zsm.ovgu.de/>

Die Methodenworkshops des ZSM in Magdeburg bieten eine Vielfalt an Möglichkeiten an Interpretationsgruppen verschiedener Methoden teilzunehmen und so einen Einblick in das konkrete Vorgehen zu bekommen. Aktive Teilnehmerinnen und Teilnehmer können hier ihr empirisches Material als Arbeitsgrundlage vorlegen. Darüber hinaus bieten sich Vernetzungsmöglichkeiten.

Berliner Methodentreffen

<http://www.qualitative-forschung.de/methodentreffen/>

Eine der großen Methodentagungen qualitativer Forschung und eine hervorragende Möglichkeit, um im deutschsprachigen Kontext namentlich bekannte Autoren persönlich kennenzulernen.

DGfE-SummerSchool

http://www.dgfe.de/tagungen_workshops/dgfe-summer-school.html

Die von der Deutschen Gesellschaft für Erziehungswissenschaft organisierte Methodentagung für Nachwuchswissenschaftlerinnen und -wissenschaftler findet in der Regel Ende Juli statt. Das Programm kann drei bis vier Monate vorher online auf der Homepage der DGfE eingesehen werden.

1.5 Literaturempfehlungen

Für alle empirischen Auswertungsmethoden gilt, dass ihre Anwendung nicht ausschließlich auf der Basis eines theoretisch-literaturbasierten Studiums erlernt werden kann. Hierzu braucht es praktische Erfahrung, Anleitung sowie den Austausch innerhalb einer Interpretationsgruppe. Insofern sind die folgenden Literaturempfehlungen als Einstiegshilfe in die Forscher-Community zu verstehen.

Bohnsack (2008): Rekonstruktive Sozialforschung.

Als ein grundlegendes Nachschlagewerk mit dem Schwerpunkt auf rekonstruktive Methoden allgemein und dokumentarische Methode insbesondere kann Bohnsacks (2008) „Rekonstruktive Sozialforschung“ genannte werden. Es eignet sich, um einen Überblick über die methodischen Grundlagen zu erhalten und gezielt einzelne Begriffe nachzuschlagen.

Bohnsack/Nentwig-Gesemann/Nohl (2007): Die dokumentarische Methode und ihre Forschungspraxis.

Speziell für die Veranschaulichung der Interpretationsschritte kann das Arbeitsbuch von Bohnsack, Nentwig-Gesemann und Nohl (2007) „Die dokumentarische

Methode und ihre Forschungspraxis“ hilfreich sein. Anhand von praktischen Beispielen werden hier die Möglichkeiten der Interpretation vorgestellt.

Przyborski (2004): Gesprächsanalyse und dokumentarische Methode.

Przyborski (2004) legt mit ihrem Buch „Gesprächsanalyse und dokumentarische Methode“ ebenfalls eine sehr anschauliche Darstellung der Interpretationsschritte vor. Sie illustriert besonders ausführlich die Diskursorganisation.

Nohl (2006): Interview und dokumentarische Methode.

Viele Beispielinterpretationen beziehen sich auf Daten, die mittels Gruppendiskussionen erhoben wurden. Bei der Interpretation von Interviews gilt es jedoch nicht nur bei der Datenerhebung, sondern auch bei der Datenauswertung Besonderheiten zu beachten. Diese stellt Nohl (2006) in seiner Anleitung „Interview und dokumentarische Methode“ ausführlich und nachvollziehbar vor.

Bohnsack (2009): Qualitative Bild- und Videointerpretation.

Die Interpretation von Bildern und Videos gewinnt zunehmend an Bedeutung (nicht nur) für die Forschungsprojekte, die mit der dokumentarischen Methode arbeiten. Auf die Besonderheiten der Auswertungen von visuellen Daten geht Bohnsack (2009) mit dem Werk „Qualitative Bild- und Videointerpretation“ ein.

Eine umfassende Auflistung der Grundlagenliteratur zur dokumentarischen Methode wird vom Centrum für qualitative Evaluations- und Sozialforschung (CES) unter <http://www.ces-forschung.de/index.php/grundlagenliteratur> zur Verfügung gestellt. Dankenswerterweise wird auch unter <http://www.dokumentarischmethode.de> eine umfassende internationale Literaturliste aller Publikationen zur dokumentarischen Methode zugänglich gemacht und jährlich aktualisiert. Hier werden nicht nur forschungsmethodische Publikationen, sondern auch Forschungsergebnisse aus verschiedenen Bereichen der Bildungs- und Sozialforschung aufgelistet.

2. Beispiel für den Einsatz der dokumentarischen Methode

Im Folgenden werden Projekte vorgestellt, die mit der dokumentarischen Methode gearbeitet haben. In Kapitel 2.1 werden kurz die Forschungsfrage, das Vorgehen und die Ergebnisse eines konkreten Forschungsprojekts skizziert. Anschließend wird in Kapitel 2.2 ein Überblick über fachdidaktische Arbeiten aus verschiedenen Disziplinen geboten. Neben Arbeiten aus der Deutschdidaktik werden auch Projekte aus anderen Fachdidaktiken vorgestellt.

2.1 Die Orientierung von Lehrenden im Themenfeld Heterogenität

Grundlage für das Forschungsinteresse liegt in der Beobachtung des Diskurses über Heterogenität. Sowohl im erziehungswissenschaftlichen Theorie-, wie auch im pädagogischen Alltagsdiskurs spiegeln sich vor allem die normativen Horizonte von Heterogenität. Beide Diskurse scheinen größtenteils unabhängig voneinander zu bestehen. Zumindest finden sich kaum Überschneidungen und Bezugnahmen. Diese Annahme lieferte den Ausgangspunkt der jüngst abgeschlossenen Arbeit (Schieferdecker 2016), die im empirischen Teil der Frage nachgeht, woran sich der Alltagsdiskurs der Lehrenden orientiert. Ziel war es, Muster zu rekonstruieren, die Rückschlüsse darauf ermöglichen, in welcher Weise das Phänomen Heterogenität durch die Bildungswissenschaften thematisiert werden kann, um die Anschlussfähigkeit an die Orientierungen der pädagogischen Praktiker zu gewährleisten. Hierzu wurde zwei Fragen nachgegangen:

- An welchen kollektiven Themen orientiert sich die Kommunikation von Lehrerinnen und Lehrern? (*Themen*)
- Welche konjunktiven Erfahrungsräume lassen sich bei Lehrerinnen und Lehrern rekonstruieren? (*Soziogenetische Typenbildung*)

Einhergehende Erkenntnisse können z.B. bei der Konzeption von Angeboten für Lehrkräfte mit einbezogen werden, um solche Angebote (aus der (Fach-) Wissenschaft) noch anschlussfähiger an die Orientierungen von Lehrenden zu gestalten. Der Fokus liegt auf der Funktion der den Handlungen zugrundeliegenden Orientierungsmuster. Diese Orientierungen helfen den Lehrkräften Interaktionsräume übersichtlich zu halten und damit Handlungen (in einer komplexen pädagogischen Praxis) zu ermöglichen.

Themen

Innerhalb aller durchgeführten Gruppendiskussionen kehren vier Perspektiven wieder, welche Lehrkräfte einnehmen, um Heterogenität zu beschreiben. Es handelt sich dabei um die Wahrnehmung der Schülerinnen und Schüler, die Wahrnehmung von Schule, die Wahrnehmung von Eltern und schließlich um die Selbstsicht der Lehrkräfte. Diese vier Perspektiven bieten für die Teilnehmenden aller Gruppendiskussionen einen grundlegenden Rahmen, auf den sie sich beziehen, um sich über Heterogenität auszutauschen.

Wahrnehmung ...	Themen
... der Schülerinnen und Schüler	Homogenisierung durch Individualisierung und Vereinheitlichung
... von Schule	Verhältnis von Gesellschaft und Schule Aufgaben der Schule
... der Eltern	Gegenüberstellung (Eltern – Lehrer) Problematisierung (Werteunterschiede)
Selbstsicht der Lehrkräfte	Besonderheit der Situation der Lehrkräfte Orientierung an Hierarchie

Tab. 1: Perspektive der Orientierung (Quelle: eigene Darstellung)

Die Wahrnehmungen der Lehrenden stellen thematische Perspektiven dar, auf die Lehrerinnen und Lehrer zurückgreifen, um über Heterogenität zu sprechen. Diese lassen sich in allen Gruppendiskussionen finden. Schülerinnen und Schüler werden zumeist homogenisierend wahrgenommen. Dies geschieht, in dem auf ein komplexitätsreduzierendes Verständnis von Individualisierung oder auch vereinheitlichende Logiken zurückgeschlossen wird. Die Schule sehen die Lehrenden in einem Spannungsverhältnis zur Gesellschaft. Daher schrieben sie der Schule auch unterschiedliche Aufgaben (z.B. die Reparatur oder die Diagnose von Gesellschaft) zu. Eltern sehen Lehrerinnen und Lehrer zumeist als Störfaktoren ihrer pädagogischen Arbeit an. Die Verschiedenheit und Vielfalt mit der sie im schulischen Kontext zu tun haben, nehmen Lehrende so wahr, dass diese von den Eltern und den Familien in die Schule hineingetragen werden. Umfassend beschreiben die Lehrenden sich in den Gruppendiskussionen selbst. Hierbei betonen sie durchgängig die Besonderheit ihrer spezifischen Situation (auf Grund des Einzugsgebiets, des Stufenschwerpunkts ihrer Schule usw.). Lehrkräfte orientieren sich durchgängig an Hierarchie. Dies schlägt sich unter anderem da nieder, wenn sie Berufsgruppen (die Putzfrau und den Herrn Professor) gegenüberstellen, wenn sie gesellschaftliche Unterschiede analog zu der schulischen Dreigliedrigkeit sortieren, wie auch in ihrer Rolle gegenüber den Schülerinnen und Schülern.

Soziogenetische Typenbildung

In der Arbeit werden konjunktive Erfahrungsräume von Lehrkräften rekonstruiert und zu Typen verdichtet. Hierbei wird beschreibbar, dass Lehrkräfte auf Muster zurückgreifen, die es ihnen ermöglichen das Spannungsverhältnis zwischen normativen Erwartungen und strukturellen Vorgaben für ihre konkrete Handlungspraxis aufzulösen. Für die Arbeit wurden Gruppendiskussionen (Loos/Schäffer 2001; s. Scherf in diesem Band) mit Lehrerinnen und Lehrern durchgeführt. Drei Typen konnten mit der dokumentarischen Methode rekonstruiert werden. Alle drei verfolgen unterschiedliche Strategien, um mit dem Spannungsverhältnis (von normativen Erwartungen und strukturellen Vorgaben) umzugehen. Die Wahrnehmung von Heterogenität wird hierbei geprägt von Homogenisierung. Dies führt zu einem Umgang mit Heterogenität, der als Ablenken beschrieben werden kann. Durch den komparativen Vergleich der Fälle lässt sich ein gemeinsames Muster

innerhalb der Typen beschreiben.¹¹ Die folgende Tabelle stellt die drei rekonstruierten Typen dar im Hinblick auf ihre Orientierung im Diskurs, die Wahrnehmung der Heterogenität, sowie den Umgang mit Heterogenität. Neben diesen Unterschieden werden zusätzlich auch die Gemeinsamkeiten aller Typen dargestellt.

Typ	Orientierung im Diskurs	Wahrnehmung der Heterogenität	Umgang mit Heterogenität
Gemeinsamkeit aller Gruppen	Polarisierung	problematisierend	homogenisieren ablenken
Löffelschmeißen	Utopie – Machbarkeit Erfolg – Versagen	kategorisierend	resignieren kapitulieren
Panzern	hart – weich Lehrer – Eltern	homogenisierend	abwehren
Festhalten an Strukturen	Lehrer – Schulsystem Wandel – Beständigkeit	relativierend	konservieren

Tab. 2: Zusammenschau der drei Typen (Quelle: Schieferdecker 2016, 186)

Bei der Zusammenschau der Fälle wird deutlich, dass sich trotz verschiedener Diskussionsverläufe bestimmte Orientierungsmuster in allen Gruppen wiederholen. Als grundlegende Gemeinsamkeiten (grau hinterlegt) fällt auf, dass sich alle Typen an polarisierenden Gegenüberstellungen orientieren. Sie nehmen Heterogenität durchgängig problematisierend wahr. Mit Heterogenität gehen sie insofern um, dass sie diese homogenisieren, bzw. von ihr ablenken. Hier wird deutlich, dass sowohl Polarisation wie auch Problematisierung dem homogenisierenden Umgang mit Heterogenität zuspitzen.

Neben einer Orientierung an einer Verfallsemantik zeigt sich, dass sich Frustration durch Machtlosigkeit als konjunktive Semantik etabliert hat. Insbesondere da versucht wird, Schülerinnen und Schüler als Individuen wahrzunehmen und sich die Teilnehmenden der Gruppendiskussion dabei in hohem Maße an idealisierten normativen Vorgaben orientieren, scheint Frustration an den begrenzten Handlungsmöglichkeiten im Alltag der Normalfall zu sein. Die Thematisierung der eigenen Machtlosigkeit führt interessanterweise nicht zur Aufgabe ihrer Lehrtätigkeit. Resignation und Kapitulation führen zu einer Legitimation relativierender Zielvorgaben. Die utopischen Ideale der Gesellschaft (und die der Lehrperson an sich selbst) müssen nicht mehr angestrebt werden, weil sich diese ja (so die Überzeugung) nicht erreichen lassen. Neben dieser Machtlosigkeit zeigen sich ebenso quer durch die Diskussionen Machbarkeits- bzw. Machtvorstellungen der Lehrenden. Der Schülerfolg der Schülerinnen und Schüler wird demnach ausschließlich von den Lehrenden bestimmt. Die Lehrerinnen und Lehrer sind es, die dem-

¹¹ Für eine ausführliche Erläuterung der Typenbildung vgl. Schieferdecker 2016, 178ff.

nach den einen zum Chirurgen machen oder einen anderen von der Schule ausschließen. Aus dieser Perspektive beschreiben sich die Lehrerinnen und Lehrer als Souveräne über Bildungsbiographien. Ihre Machbarkeitsvorstellung widerspricht der (gleichfalls beschriebenen) Unverfügbarkeit gelingenden Unterrichts. Macht und Machtlosigkeit bleiben nebeneinander stehen, ohne dabei zu Irritationen zu führen (s. im Detail: Schieferdecker 2016).

2.2 Ausblick: Einsatz der dokumentarischen Methode in fachdidaktischer Forschung

In dem jungen Gebiet der deutschdidaktisch-empirischen Forschung wurden bisher nur wenige Arbeiten vorgelegt, die sich explizit auf die dokumentarische Methode beziehen. Daher lohnt sich für spezifisch fachdidaktische Perspektiven auf die Methode ein Blick in andere Fachdidaktiken.¹² Die ersten Arbeiten innerhalb der Deutschdidaktik wurden von Wieser (2008), Lindow (2013) und Scherf (2013) vorgelegt. Mittlerweile sind weitere Arbeiten hinzugekommen u.a. Kamzela (2015) sowie Schmidt (2015). Allen Arbeiten gemein ist die forschungsmethodische Ausrichtung. So erhoben diese Untersuchungen ihre Daten (vorwiegend) mit Interviews von Lehrkräften. Die Arbeiten lassen sich im weitesten Sinne somit der Professionsforschung zuordnen. Ziel war es dabei, die Orientierungen von Deutschlehrerinnen und Deutschlehrern (teilweise auch von Referendarinnen und Referendaren) zu beschreiben. Dass mittels der dokumentarischen Methode auch andere Bereiche der Schul- und Unterrichtsforschung berücksichtigt werden können, zeigen u.a. Arbeiten aus anderen Sprachdidaktiken (exempl. Bonnet 2004; 2009; Bracker 2012). Hier werden neben den Orientierungen von Lehrenden, Schülerinnen und Schülern auch Unterrichtsinteraktionen empirisch beschrieben. Im Vergleich hierzu wird deutlich, dass deutschdidaktisch-empirische Forschung bei weitem noch nicht alle Möglichkeiten der dokumentarischen Methode ausgeschöpft hat. Insbesondere Bild- und Videointerpretationen könnten mittelfristig den empirischen Diskurs innerhalb der Fachdidaktik nachhaltig bereichern. Auch ein verstärkter Blick auf die Orientierungen von Schülerinnen und Schülern birgt das Potential, die didaktische ‚Brille‘ zu schärfen.

Literatur

- Ballis, Anja/Penzold, Michael/Scherf, Daniel/Schieferdecker, Ralf: Die dokumentarische Methode und ihr Potenzial für Forschungen (nicht nur) in der Fachdidaktik Deutsch. In: *Didaktik Deutsch* (2014), 37, 92-104.
- Bohnsack, Ralf/Marotzki, Winfried/Meuser, Michael (Hrsg.): *Hauptbegriffe Qualitativer Forschung*. Opladen/Farmington Hills: Barbara Budrich 2010.
- Bohnsack, Ralf/Nentwig-Gesemann, Iris/Nohl, Arnd-Michael (Hrsg.): *Die dokumentarische Methode und ihre Forschungspraxis. Grundlagen qualitativer Sozialforschung*. Wiesbaden: VS Verlage 2007, 225-253.

¹² Für eine ausführliche Zusammenschau vgl. Ballis et al. (2014, 92ff.).

- Bohnsack, Ralf/Przyborski, Aglaja/Schäffer, Burkhard (Hrsg.): Das Gruppendiskussionsverfahren in der Forschungspraxis. Opladen/Farmington Hills: Barbara Budrich 2010.
- Bohnsack, Ralf: Generation, Milieu und Geschlecht – Ergebnisse aus Gruppendiskussionen mit Jugendlichen. Opladen: Leske+Budrich 1989.
- Bohnsack, Ralf: Rekonstruktive Sozialforschung. Einführung in qualitative Methoden. Opladen: Barbara Budrich 2008.
- Bohnsack, Ralf: Standards nicht-standardisierter Forschung in den Erziehungs- und Sozialwissenschaften. In: Zeitschrift für Erziehungswissenschaft 8 (2005) 4, 63-81. Abrufbar unter: http://www.forschungsnetzwerk.at/downloadpub/bohnsack_standards1.pdf (letzter Zugriff: 01.08.2018).
- Bohnsack, Ralf: Typenbildung, Generalisierung und komparative Analyse: Grundprinzipien der dokumentarischen Methode. In: Bohnsack, Ralf/Nentwig-Gesemann, Iris/Nohl, Arnd-Michael (Hrsg.): Die dokumentarische Methode und ihre Forschungspraxis. Grundlagen qualitativer Sozialforschung. Wiesbaden: VS Verlage 2007, 225-253.
- Bonnet, Andreas: Chemie im bilingualen Unterricht. Kompetenzerwerb durch Interaktion. Opladen: VS Verlag 2004.
- Bonnet, Andreas: Die Dokumentarische Methode in der Unterrichtsforschung. Ein integratives Forschungsinstrument für Strukturrekonstruktion und Kompetenzanalysen. In: Zeitschrift für Qualitative Forschung 10 (2009) 2, 219-240.
- Bonnet, Andreas: Von der Rekonstruktion zur Integration. Wissenssoziologie und dokumentarische Methode in der Fremdsprachenforschung. In: Doff, Sabine (Hrsg.): Fremdsprachenunterricht empirisch erforschen. Grundlagen – Methoden – Anwendung. Tübingen: Narr 2012, 286-305.
- Bourdieu, Pierre: Der Habitus als Vermittlung zwischen Struktur und Praxis. In: Ders.: Zur Soziologie symbolischer Formen. Frankfurt a.M.: Suhrkamp 2014, 125-158.
- Bracker, Elisabeth: Anwendungsbeitrag: Die dokumentarische Methode als Instrument zur Analyse von literarischer Anschlusskommunikation. In: Doff, Sabine (Hrsg.): Fremdsprachenunterricht empirisch erforschen. Grundlagen – Methoden – Anwendung. Tübingen: Narr 2012, 306-315.
- Bracker, Elisabeth: Fremdsprachliche Literaturdidaktik: Plädoyer für die Realisierung bildender Erfahrungsräume im Unterricht. Wiesbaden: Springer VS 2015.
- Kamzela, Kerrin: Die Perspektive von Lehrenden auf basale Lesefähigkeiten zu Beginn der Sekundarstufe I. In: Bräuer, Christoph/Wieser, Dorothee (Hrsg.): Die Lehrkraft im Blick – empirische Lehrerforschung in der Deutschdidaktik. Wiesbaden: VS Verlag 2015.
- Lischke-Eisinger, Lisa: Sinn, Werte und Religion in der Elementarpädagogik. Religion, Interreligiosität und Religionsfreiheit im Kontext der Bildungs- und Orientierungspläne. Wiesbaden: VS Verlag 2012.
- Luhmann, Niklas: Soziale Systeme. Grundriß einer allgemeinen Theorie. Frankfurt a.M.: Suhrkamp 1987.
- Mangold, Werner/Bohnsack, Ralf: Kollektive Orientierung in Gruppen Jugendlicher. Bericht für die Deutsche Forschungsgemeinschaft. Erlangen: Universität Erlangen 1988.

- Mangold, Werner: Gegenstand und Methode des Gruppendiskussionsverfahrens. Aus der Arbeit des Instituts für Sozialforschung. Frankfurt a.M.: Europäische Verlagsanstalt 1960.
- Mannheim, Karl: Ideologie und Utopie. Frankfurt a.M.: Suhrkamp 1952.
- Mannheim, Karl: Wissenssoziologie. Auswahl aus dem Werk. Neuwied/Berlin: Luchterhand 1964.
- Nentwig-Gesemann, Iris: Die Unterrichtssequenz Parallelogramm II. Fallrekonstruktion mit der Dokumentarischen Methode. In: Pieper, Irene/Frei, Peter/Hauenschild, Katrin/Schmidt-Thieme, Barbara (Hrsg.): Was der Fall ist. Beiträge zur Fallarbeit Bildungsforschung, Lehramtsstudium, Beruf und Ausbildung. Berlin: VS Verlag 2014, 123-138.
- Nohl, Arnd-Michael: Interview und dokumentarische Methode. Anleitung für die Forschung. Wiesbaden: VS Verlag 2006.
- Nohl, Arnd-Michael: Relationale Typenbildung: Neue Wege der dokumentarischen Methode. VS Verlag 2013.
- Scherf, Daniel: Leseförderung aus Lehrersicht. Eine qualitativ-empirische Untersuchung professionellen Wissens. Wiesbaden: VS Verlag 2013.
- Schieferdecker, Ralf: Die Orientierung von Lehrerinnen und Lehrern im Themenfeld Heterogenität. Eine rekonstruktive Analyse. Opladen/Farmington Hills: Budrich Verlag 2016.
- Schmidt, Frederike: Den diagnostischen Blick schärfen – Vorstellungen und Orientierungen von Deutschlehrern zur Diagnose von Lesekompetenz. In: Bräuer, Christoph/Wieser, Dorothee (Hrsg.): Die Lehrkraft im Blick – empirische Lehrerforschung in der Deutschdidaktik. Wiesbaden: VS Verlag 2015.
- Schütz, Fritz: Das narrative Interview in Interaktionsfeldstudien. Studienbrief der Universität Hagen. Teil I. Hagen 1987.
- <http://www.ces-forschung.de> (letzter Zugriff: 01.08.2018).
- <http://www.ces-forschung.de/index.php/grundlagenliteratur> (letzter Zugriff: 01.08.2018).
- http://www.dgfe.de/tagungen_workshops/dgfe-summer-school.html (letzter Zugriff: 01.08.2018).
- <http://www.dokumentarischmethode.de> (letzter Zugriff: 01.08.2018).
- <http://www.fallarchiv.uni-kassel.de/lernumgebung/dokumentarische-methode/>(letzter Zugriff: 01.08.2018).
- <http://www.qualitative-forschung.de/methodentreffen/>(letzter Zugriff: 01.08.2018).
- <http://www.zsm.ovgu.de/>(letzter Zugriff: 01.08.2018).

Objektive Hermeneutik

Sequenzanalytische Rekonstruktion von Unterrichtsprozessen

1. Vorstellung der Objektiven Hermeneutik

Die Objektive Hermeneutik ist eine Forschungsmethode der klinischen Soziologie und wurde von Oevermann und Mitarbeitern Ende der 1970er Jahre entwickelt (Oevermann et al. 1979). Alleinstellungsmerkmal dieser Methode ist ihr Fokus auf die *sequenzanalytische* Rekonstruktion vor allem natürlicher Protokolle sozialer Praxis, welche in der Regel als Transkriptionen vorliegen. Sie dient insbesondere dazu, die Regeln und Normen einer bestimmten Praxis vollumfänglich herauszuarbeiten und versteht sich als mikrologische, aus der Praxisanalyse heraus Theorie generierende Grundlagenforschung. Als Verfahren der Fallarbeit wird die Objektive Hermeneutik zunehmend in der Lehrprofessionalisierung eingesetzt.

1.1 Funktionsweise der Objektiven Hermeneutik

Mit Objektiver Hermeneutik wird scheinbar vertrautes soziales Handeln unvertraut und dabei systematisch einem theoretischen sowie kritischen Blick zugänglich. Sie ermöglicht die Explikation des Regelwissens, welches wir, zumeist unbewusst, erworben haben. Es handelt sich um eine soziologische Forschungsmethode, welche aber in beeindruckender Weise die Logik pädagogischen Handelns gerade auch in ihrer (antinomischen) Widersprüchlichkeit aufklären kann (vgl. Gruschka 2005; 2011b). Sie basiert methodologisch auf der Einsicht in die Sinnstrukturiertheit bzw. Regelgeleitetheit der sozialen Welt, in der wir auf wechselseitige Interpretationsleistungen in der Interaktion angewiesen sind. Folglich muss es einen gemeinsamen Rahmen geben, vor dem Handeln sinnvoll erscheint. Und es muss zugleich Handlungsspielräume geben, die Varianten sinnvollen Handelns erlauben und von sinnlosem (regelverletzendem) Handeln unterscheidbar sind. Ohne solche Handlungsspielräume wäre eine *individuelle* Lebenspraxis¹ nicht denkbar, die durch den Rückgriff auf bewährte ausgewählte Handlungsrou-

¹ Zur Objektiven Hermeneutik als Sozialisationstheorie bzw. Bildungstheorie vgl. Garz/Raven 2015. Dort ist auch ein umfangreiches Glossar zur Terminologie der Objektiven Hermeneutik zu finden.

tinen gekennzeichnet ist: Es gäbe keine Bildungsprozesse, keine Zukunftsoffenheit, sondern ein mechanisch anmutendes Reiz-Reaktions-Handeln. Ein Verweis auf das morgendliche Grußritual oder die Lehrerfrage kann dies illustrieren: Beide Handlungsmuster haben im schulischen Kontext ganz eigene, nur dort geltende Regeln, die von beiden Seiten erworben und verinnerlicht wurden. Dennoch kommt es immer wieder zu kommunikativen Krisen: Der Morgengruß, durchaus noch zur Mittagszeit, soll wiederholt werden (Wo außerhalb von Schule ist so etwas denkbar?), und der Lehrende fragt im zweiten Fall etwas, was er offenkundig weiß (Abseits der Schule wäre er deshalb ein schwer erträglicher Besserwisser).

Mit der Objektiven Hermeneutik werden die Regeln rekonstruiert, welche eine spezifische soziale Praxis konstituieren. Diese Regeln sind bestimmbar: sie sind *bewusstseinsfähig*, aber nicht *bewusstseinspflichtig*. Wir haben sie im Verlauf unseres Sozialisations- und Bildungsprozesses erworben und folgen ihnen *routiniert* – und bemerken oftmals erst, dass es sie gibt, wenn es zu einer *Krise* kommt und wir vor einer Wahl stehen, also sprachhandelnd eine Entscheidung treffen müssen. Ein Lehrer, der sich nicht richtig zurückgegrüßt empfindet, ist herausgefordert zu reagieren. Die Zahl seiner sinnvollen Reaktionsmöglichkeiten ist grundsätzlich begrenzt. Sie sind vorgegeben von *latenten Sinnstrukturen*, die Geltung haben unabhängig davon, ob sie dem Handelnden als *intentionale Sinnstrukturen* bewusst sind oder nicht.²

Hermeneutik im traditionellen Sinn ist die Lehre des Verstehens eines Anderen, seiner Subjektivität und Individualität. Dieses Verstehen erfolgt eben durch ein zweites Subjekt, dessen Verstehensleistungen vor dem Hintergrund seiner jeweils ebenfalls subjektiv erworbenen Lebenserfahrungen erfolgt. Die Objektive Hermeneutik rekonstruiert hingegen nicht innerpsychische Sinngehalte, sondern *strukturellen Sinn*: Es geht ihr gerade nicht um die Bestimmung, was jemand tun will oder tun wollte, sondern darum, was er objektiv getan bzw. gemeint hat – im Rahmen der individuellen Selektion sinnvoller Handlungsmöglichkeiten. Beides kann zwar durchaus übereinstimmen, was einem Ideal gelungener, selbstbewusster, *autonomer Lebenspraxis* entspreche, dem ist aber im empirischen Regelfall meist nicht so. Objektive Hermeneutik ist ein Verfahren zur Explikation der Regeln, auf die wir beim Handeln in bestimmten sozialen Kontexten zurückzugreifen gelernt haben. Objektiv sind diese Regeln, weil sie in der *sinnstrukturierten Welt* unser Handeln erst ermöglichen, also für alle gelten, auch wenn wir sie nicht bewusst kennen. Objektiv ist die Objektive Hermeneutik über ihren Erkenntnis-

² Hier unterscheidet sich die Objektive Hermeneutik von Verfahren, welche das Lehrwissen durch Befragung erheben. Latente Sinnstrukturen werden dabei nicht erschlossen, interpretierbar wird nur, was die Lehrperson für thematisierungswürdig hält. Ina Lindow hat eine lesenswerte Dissertation geliefert, in der sie das Fallwissen von Deutschlehrenden erforscht (Lindow 2013). Alles, was die Lehrenden selbst nicht für problematisch halten, fällt damit unter den Tisch. Genau dafür interessieren sich aber die objektiven Hermeneuten.

gegenstand hinaus als Kunstlehre, weil das regelgeleitete Verfahren der *Gruppeninterpretation* eine *falsifizierbare*, intersubjektiv nachvollziehbare Strukturrekonstruktion absichert.

Damit unterscheidet sich die Objektive Hermeneutik gleichermaßen von qualitativen wie von quantitativen Forschungsmethoden, die beide *subsumtionslogisch* vorgehen und die vorgefundene Wirklichkeit in eine vorher entworfene Sinnstruktur einpassen, anstelle diese Sinnstruktur zu *rekonstruieren*. Der Vorwurf einer unreflektierten Normativität, der gerne den qualitativen Verfahren gemacht wird, nicht selten zu Recht, trifft für die Objektive Hermeneutik deshalb nicht zu. Ihr geht es vielmehr darum, die in der Praxis selbst geltenden Normen sinnvollen Handelns zu rekonstruieren: Sie ist *keine normative Theorie* der Praxis, sondern eine *Theorie der Normen* der Praxis. Diese Normen erwerben die Handelnden im Laufe ihres Bildungsprozesses durch Handlungskrisen, in denen die bislang verfügbaren Routinen versagen und durch neue ersetzt werden müssen. Oevermann unterscheidet deshalb die *Strukturtransformation*, die bezogen auf das Subjekt den Normalfall darstellt – sonst gäbe es keine Bildungsprozesse –, von der *Strukturreproduktion*, die den Sonderfall darstellt. Dennoch begegnet dem empirisch Forschenden wesentlich häufiger die routinehafte Strukturreproduktion³ als die krisenhafte Strukturtransformation: Bildungsprozesse sind, aus hier nicht zu diskutierenden Gründen, schwer bzw. selten zu beobachten.⁴

Als Alleinstellungsmerkmal der Objektiven Hermeneutik kann gelten, dass sie zu einer *rekonstruktiven Sequenzanalyse* führt. Die Sequenzanalyse basiert darauf, dass die individuelle Lebenspraxis sich in zuvor krisenhaft gewonnenen Auswahlmustern von *Handlungsentscheidungen* ausdrückt, die nunmehr zu *Routinen* gewonnen sind. Die Feststellung der Routinehaftigkeit einer Praxis ist für das Verfahren der Objektiven Hermeneutik von grundlegender Bedeutung.

Schon früh hat die Forschung von Oevermann et al. gezeigt, dass es in der Lebenspraxis meist nicht mehr als zwei bis vier wirklich sinnstrukturell distinkte *sinnvolle Reaktionsmöglichkeiten* gibt. Welche davon jemand wählt, ist Ergebnis und Ausdruck seines individuellen Bildungsprozesses. Es ist unabdingbar für das Verfahren der Objektiven Hermeneutik, auch die nicht gewählten sinnvollen Handlungsmöglichkeiten restlos herauszuarbeiten und diese auch zu begründen. Dies erfolgt im Rahmen einer Schritt-für-Schritt-Analyse eines Interaktionsprotokolls. Dabei ist es wichtig, die tatsächlich gewählte Reaktionsform nicht bereits zu kennen, damit sie mit der Interpretationsgruppe zuvor gedankenexperimentell entworfen und vor allem begründet werden kann. Angestrebt ist damit einerseits, die *Fallstruktur* eines Handlungsmusters bestimmen zu können als gewissermaßen individualdiagnostisches Ziel, aber eben auch andererseits das *latente Feld* (die *Struktur*) der maximalen Handlungsmöglichkeiten. Dies ist sehr wichtig, da viele

³ Wie später gezeigt wird, ist die Strukturreproduktion jedoch unabdingbar für das Verfahren der Objektiven Hermeneutik.

⁴ Vgl. zur Methodologie und Terminologie der Objektiven Hermeneutik Oevermann 2000 und Garz/Raven 2015.

der bislang empirisch rekonstruierten Handlungsmuster nicht sinnvoll begründbar sind und daher als Verletzungen gelungenen sozialen Handelns erscheinen. Auch diese haben eine spezifische Struktur, folgen einer bestimmbar Logik und sind nicht bloß als Abweichungen zu verstehen (s.u.).

Bei der sequenzanalytischen Vorgehensweise wird an jeder Sequenzstelle gefragt, wie eine Sprachhandlung objektiv motiviert ist, wie es demnach weitergehen, wie reagiert werden könnte. Diese möglichen Anschlüsse müssen ausführlich begründet werden. Auf diese Weise erschließt man neben der individuellen Fallstruktur (zunächst als *Fallstrukturhypothese*) auch die anderen möglichen sinnvollen Fallstrukturen – wie es hätte weitergehen können. In einem Fall sind also stets mehrere Fälle gelagert. Hat man sie bestimmt, verfügt man über eine *Strukturtheorie* der sozialen Handlungsmöglichkeiten im spezifischen Feld. Empirisch gesättigt kann man diese Theorie dann nennen, wenn man sich auf die *empirische Suche* nach den alternativen Fällen, insbesondere von *kontrastiv* denkbaren Fällen gemacht hat. Auf diese Weise kontrolliert man die utopische Dimension der Fallbestimmung und befolgt das Primat der Dignität der Praxis vor jeder Theorie: Die Empirie dient nicht der Bestätigung einer vorgelagerten Theorie, sondern die Theorie der Praxis wird aus dem Fall heraus selbst entwickelt.

Im Rahmen der Kunstlehre der Objektiven Hermeneutik benötigt man dazu die rekonstruierte Beobachtung einer Strukturreproduktion. Sobald bzw. erst wenn sich im Fall eine Struktur wiederholt hat, gilt die Fallstrukturhypothese als belegt – will sagen, der Interaktant hat seine routinierte Auswahlentscheidung wiederholt, wodurch sie als Routine, also als in seinem Bildungsprozess entwickelte, geronnene Krisen(bewältigungs)struktur sichtbar geworden ist (Oevermann 2000, 95).

Die Ergebnisse objektiv-hermeneutischer Analysen führen zu einer Theorie der Praxis, d.h. die Praxis wird über ihre Strukturlogik aufgeklärt. Man kann dann begründen, weshalb eine bewährte Praxis sinnvoll ist und (wie sie) funktioniert. Darüber hinaus kann aber auch festgestellt werden, weshalb eine bewährte Praxis nicht sinnvoll ist, obwohl sie mitunter ebenso eingespielt funktioniert. Es wird auch in der Deutschdidaktik noch oft übersehen, dass es sinnvoll ist, die Handlungslogik eines dysfunktionalen Deutschunterrichts aufzuklären, anstelle immer nur innovationslogisch vorzugehen: Noch immer beschäftigen sich fachdidaktisch Forschende nahezu ausschließlich mit der Bestimmung guter Praxis, anstelle die Logik der realen Praxis aufzuklären – ein professionstheoretisches Defizit, da die Ausblendung der Unterrichtswirklichkeit in der fachdidaktischen Theoriebildung zur Kontinuität der bestehenden Praxis beiträgt.

Basierend auf den hier skizzierten methodologischen Prämissen gibt es in der Objektiven Hermeneutik fünf Maximen, die sich bei der Sequenzanalyse bewährt haben: *Sequenzialität*, *Kontextfreiheit*, *Wörtlichkeit*, *Extensivität*, *Sparsamkeit*.⁵ Das Prinzip der Sequenzialität verpflichtet – um es narratologisch auszudrücken

⁵ Genauere Erläuterungen zu den Maximen sind in Wernet (2005) und Gruschka (2005) nachzulesen.

– die Interpretationsgruppe darauf, eine personale und nicht eine auktoriale Perspektive einzunehmen: Man verzichtet, anders als in der traditionellen Hermeneutik, auf den hermeneutischen Zirkel. Eine Aussage soll also keinesfalls dadurch plausibel gemacht werden, indem man im Transkript vorblättert und dann göttergleich weiß, worauf der Lehrende eigentlich hinaus wollte. Das wäre eben klassische (subjektive) Hermeneutik. Das sequenzielle Vorgehen sichert ab, dass Schritt für Schritt die Handlungsalternativen der Interaktionspartner bewusst und expliziert werden. Wüsste man das Ergebnis schon vorab, wären die Handlungsalternativen keine echten Alternativen mehr, der rekonstruierte reale Möglichkeitsraum wäre deformiert. Deshalb gilt auch: Springen darf man im Transkript erst dann, wenn man einmal die Fallstruktur bestimmt hat und die Fallstrukturhypothese bestätigen will. Die Maxime sichert ab, dass die Interpreten sich künstlich in eine analoge Situation zu den fallinternen Rezipienten begeben. Denn diese müssen ja ebenfalls deuten, was das Gesagte für sie bedeutet und wie man darauf sinnvoll reagieren kann. Die einzigen Unterschiede zu den Beteiligten bestehen darin, dass sie alle denkbaren Anschlussmöglichkeiten benennen und dass sie handlungsentlastet die Sinnhaftigkeit der Anschlüsse begründen müssen.

Kontextfreiheit ist die zweitwichtigste Regel der Objektiven Hermeneutik. Denn die Erklärung des Handelns durch den Kontext führt ebenfalls zu Zirkelschlüssen. Man liefert gewissermaßen eine *Kontextanalyse* und keine *Textanalyse*. Auch dieses zweite Prinzip führt zu einer künstlichen Naivität, durch die erst die Handlungslogik sichtbar wird. Ein Experte würde im Gegenteil immer Kontexte heranziehen, um ein Geschehen plausibel erklären zu können. Verstanden wird es dadurch jedoch dann nicht mehr. Um die besondere Logik des schulischen Morgengrußes zu erschließen, ist es sinnvoll, ihn zu dekontextualisieren, also vom unterrichtlichen Kontext zunächst abzusehen. Nur so können Fragen an den Interakt gestellt werden, die man bei Kenntnis des Kontextes nicht stellen würde (vgl. Gruschka 2005, 61). Dazu erfindet man Geschichten, in denen der Sprechakt sinnvoll verwendet wird, und bestimmt anschließend die allgemeinen Regeln dieser Sinnhaftigkeit. Nun kann man antizipieren, wie man darauf ebenfalls sinnvoll reagieren könnte. Durch die Rekontextualisierung wird letztendlich sichtbar, ob die allgemeinen Regeln auch in der spezifischen Situation gültig sind und befolgt werden oder ob der jeweilige besondere Kontext besondere Regeln hervorbringt. Dies gilt sowohl für den Morgengruß als auch für die Lehrerfrage, aber beispielsweise auch für die Frage nach fachspezifischen Ergebniskulturen.

Das Wörtlichkeitsprinzip verpflichtet darauf, wirklich das Gesagte und nicht das womöglich Gemeinte zu interpretieren. Letzteres würde Tür und Tor öffnen für Heilungsversuche: ‚Der Lehrer habe es doch gar nicht so gemeint, als Deutschlehrer wisse man doch...‘. Die Heilung übernehmen oftmals die Schülerinnen und Schüler. Unterricht läuft reibungslos ab, selbst wenn die Interpreten sich an jeder Sequenzstelle wundern, dass es noch immer weiter geht. Das Wörtlichkeitsprinzip bedeutet, dass man beispielsweise Aufgabenstellungen buchstäblich nimmt: Sollen die Schülerinnen und Schüler eine Figurenentwicklung darstellen, ist dies etwas ganz anderes, als wenn sie das Figurenverhalten oder die Figurenwirkung

skizzieren sollen. Verspricht ein Lehrender, an den Eindrücken der Schülerinnen und Schüler zu arbeiten, sollte dies als Versprechen ernst genommen werden. Würde man nicht wörtlich interpretieren, wären eher die Normalvorstellungen des Interpreten Gegenstand der Interpretation als die Regeln der rekonstruierten sozialen Praxis. Zudem geht es darum, auf alle, noch so unscheinbaren Details zu achten, und, anders als in der Lebenspraxis, diese nicht geflissentlich zu überhören (vgl. ebd., 62).

Extensivität bedeutet, dass man erst dann in der Interpretation fortschreiten sollte, wenn man alle Elemente analysiert und alle möglichen plausiblen Lesarten formuliert und geprüft hat. Überhastete, unvollständige Strukturrekonstruktionen, die „das ggfs. schwer erträglich lange Ausinterpretieren der ersten Stellen“ (ebd., 63) abkürzen wollen, führen regelmäßig zurück über Los.

Die Sparsamkeitsregel bedeutet, dass man keine Kontexte heranziehen soll, die sich im Fall selbst nicht belegen lassen: Der Lehrende hatte einen schlechten Tag, die Klasse kommt von der Klassenfahrt wieder, nach Mathe müssen die Schülerinnen und Schüler im Deutschunterricht halt die Sau rauslassen, etc. Damit vermeidet man, vorschnell dem Unterricht einen pathologischen sinnfreien Verlauf zu unterstellen. Grundsätzlich müssen die Interpreten bis zum etwaigen Beweis des Gegenteils durch den Fall selbst davon ausgehen, dass die Beteiligten sinnvolle Interaktionsbeiträge liefern.

Die Objektive Hermeneutik unterscheidet sich von semantischen und pragmatischen Gesprächsanalysen durch den Fokus auf die sequenziell vorgehende Strukturrekonstruktion eines Falles. Untersucht wird dabei, welche Anschlussmöglichkeiten ein Interakt eröffnet und zugleich beschließt. Erfahrungsgemäß gibt es zu meist nicht mehr als 3-4 sinnvolle Reaktionsmöglichkeiten. Welche davon gewählt wird, ist Ausdruck eines individuellen Bildungsprozesses. Dessen Resultat lässt sich bestimmen, wenn sich ein einmal gewähltes und in der Interpretation bestimmtes Handlungsmuster im Rahmen der Analyse reproduziert. Dann wird aus der Fallstrukturhypothese eine Fallstruktur. In der fallspezifischen Ausdrucksgestalt zeigen sich, über die Erschließung der nicht gewählten sinnvollen Anschlüsse, allgemeine Strukturmerkmale der jeweiligen sozialen Praxis.

Im Rahmen der Umsetzung der fünf Maximen hat sich das dekontextualisierende Erfinden von Geschichten zu dem jeweiligen Interakt als sinnvolles heuristisches Instrument erwiesen, dessen Strukturlogik zu erschließen. Dazu wird der fokussierte Interakt eingebettet in von den Interpreten erfundene Kommunikationssituationen, in denen er sinnvoll erscheint. In einem zweiten Schritt muss begründet werden, weshalb der Interakt in dieser Geschichte einen sinnvollen Anschluss darstellt bzw. sinnvolle Anschlüsse hervorruft. Um diesen Schritt zu vollziehen, müssen die Interpreten ihr latentes Regelwissen explizieren. Bei der anschließenden Rekontextualisierung lässt sich feststellen, ob der Interakt bzw. seine Anschlüsse in der spezifischen Situation angemessen gewählt sind oder nicht. Forschungspraktisch ist es ebenfalls möglich, kontrastiv zu eruieren, in welchen Kontexten die Aussage völlig inakzeptabel wäre. Auch können vorübergehend allzu

kontextualisierende Begriffe ersetzt oder ausgelassen werden (Wernet 2006, 92), um die Strukturlogik in anderen Handlungsbereichen zu erschließen.

1.2 Ausgangsdaten und Ergebnisse

Die Begründer der Objektiven Hermeneutik haben zunächst bevorzugt mit natürlichen Protokollen der Lebenspraxis gearbeitet. Das sind insbesondere Transkriptionen von naturwüchsiger Kommunikation. Hinzu traten klinische Interviews. Seit einigen Jahren wird die Objektive Hermeneutik auch eingesetzt, um Paratexte, Texte, aber auch Bilder zu interpretieren. Besonders geeignet erscheint sie aber für die Rekonstruktion komplexer und zugleich hochgradig normativ regulierter Kommunikationssituationen wie z.B. Unterricht, da sie den Normen der Praxis mimetisch nachspürt, anstelle die Praxis aus normativer Sicht zu beurteilen.⁶

In der Diskussion um Videographie (s. Maak in diesem Band) versus Audiographie von Unterricht wird vielfach vertreten, das Videodatum sei objektiver, umfassender, authentischer als das Audiodatum, überführt in die Form der Transkription (s. Exkurs in Heller/Morek in diesem Band). Dem muss entgegnet werden, dass die für die Sequenzanalyse notwendige intermittierende Langsamkeit des Voranschreitens mit der realzeitlichen Videobetrachtung kaum erfüllt werden kann. Im Transkript kann jeder Interakt einzeln (kontextfrei) betrachtet werden, sodass man den künftigen Verlauf, insbesondere den tatsächlich gewählten Anschluss nicht voreilig beobachten kann. Auch mit Folienpräsentationen lässt sich die Zukunftsoffenheit der Interaktion forschungspraktisch gut inszenieren, indem immer nur ein Interakt sichtbar ist und erst dann fortgeschritten wird, wenn alle denkmöglichen sinnvollen Anschlüsse von der Interpretationsgemeinschaft expliziert worden sind. Zudem bietet das Transkript den Vorteil der Verfremdung von allzu Vertrautem, während der rasche Mitvollzug einer Videoaufzeichnung vor schnell zu einer Identifikation des Interpreten mit der zu interpretierenden Situation führen kann – was die rekonstruktionslogische Analyse be- oder gar verhindert (vgl. Pflugmacher/Twardella 2009).

Die Konzentration auf das gesprochene Wort basiert auf der extensiven Auslegungspraxis sowie der Fokussierung auf das wörtlich Gesagte und nicht das ‚eigentlich‘/‚vermeintlich‘/‚mutmaßlich‘ Gemeinte. Da Bildungsprozesse in der Verständigung über die Sache genauso wie Erfahrung an Sprache gebunden sind, ist ein Unterrichtsvideo bei einem sorgfältig angefertigten Transkript selten notwendig und kann allenfalls zur Verifikation herangezogen werden. Manche Forschende behalten auch die Audiodateien, um sie bei Zweifelsfällen heranziehen zu können (vgl. ebd.).

⁶ Vgl. Pflugmacher (2010; 2014). Helmuth Feilke wünscht sich eine Umstellung der Forschungsoptik in der Deutschdidaktik und möchte die vorherrschende Outcome-Forschung durch eine Bildungsprozessforschung ersetzen bzw. ergänzen (Feilke 2014). Einschlägige Unterrichtsprozessforschung mit Objektiver Hermeneutik findet man insbesondere in den Arbeiten von Andreas Gruschka et al.

Es gibt unterschiedliche Verfahren der Transkription. Welches gewählt wird, hängt vom jeweiligen Forschungsinteresse ab. Ein praktikables, im Vergleich mit aufwändigen und schwer lesbaren Partiturtranskriptionen nicht zu komplexes Transkriptionsverfahren hat Pollmanns entwickelt, welches insbesondere die Notation von Sprechpausen inkl. ihrer Länge und gleichzeitiges Sprechen mit berücksichtigt (vgl. Pollmanns 2010).⁷ Ausgangsdaten müssen aber nicht unbedingt selbst erhoben werden. Denn mittlerweile gibt es zahlreiche Fallarchive, von denen einige nach Anmeldung online zugänglich sind.⁸ Wer hier recherchiert, findet eine Fülle von Transkripten und anderen Dokumenten und kann mit Hilfe von Suchparametern auf Grund der Beschreibungsdaten spezifizierte Fundstellen erhalten. In den Transkripten selbst lässt sich häufig nicht recherchieren, da es sich meist um gescannte pdf-Dokumente handelt, die nicht von einer OCR-Software erfasst wurden.

In der Regel wird in der Objektiven Hermeneutik die Interpretation mit einer sozialen Eröffnungsszene begonnen – einem Anfang. Fachdidaktische Forschungsfragen bzw. Anfänge müssen aber nicht unbedingt mit der Rekonstruktion von Unterrichtseröffnungen beginnen, in der die pädagogische Sozialität hergestellt wird. Der Anfang der Rekonstruktion hängt vielmehr von der Frage ab, „Was der Fall ist?“ (Pieper et al. 2014; Wernet 2006). Eine Unterrichtsstunde enthält immer mehrere, bisweilen zahlreiche Fälle. Mitunter sind sie miteinander verknüpft (z.B. anerkennungs- bzw. adressierungstheoretische Aspekte mit der didaktischen Rahmung, mit der Bildungsbewegung, die Erziehungsdimension mit der Bildungsdimension und der Vermittlungsebene, etc.). Der fachdidaktische Fall beginnt in der Regel mit einer Hinwendung zum Gegenstand (Erwähnung, Anknüpfung, Aufgabenstellung, etc.). Aber wann ist eine Transkriptanalyse beendet? Wenn sich eine Fallstruktur mindestens einmal reproduziert hat, kann im Prinzip auf die weitere Auswertung des Transkripts verzichtet werden. Im Idealfall erfolgt erst dann eine zweite Erhebungsphase, in der maximal kontrastive Fälle zum ersten Fall gesucht werden. Zur Bestimmung der Fallstrukturgesetzmäßigkeiten im Rahmen einer bestimmten Forschungsfrage reichen in der Regel ein halbes Dutzend bis Dutzend Fälle aus. Oevermann spricht von maximal 12-14 Fällen in komplexen Forschungsfragen (Oevermann 2000, 99), da das Ziel ja nicht eine empirische Generalisierung, sondern die Strukturgeneralisierung ist.⁹

Die Ergebnisse einer solchen Erhebung lassen sich kaum abgekürzt in Aufsatzform darstellen. Die objektiv-hermeneutische Sequenzanalyse überzeugt durch

⁷ Weitere Hinweise zum Transkribieren in Dittmar (2009) und Fuß/Karbach (2014).

⁸ Z.B. Frankfurt (Apaek) und Kassel (Fallarchiv). Im Kasseler Archiv sind auch Lehrvideos vorhanden, welche die Methode der Objektiven Hermeneutik vorführen.

⁹ Wobei es aus forschungspraktischer Sicht zu empfehlen ist, immer mehr Stunden aufzunehmen, als später ausgewertet werden: Mal ist es die Technik, die versagt (Batterien sind unerwartet schnell leer, eine Datei stürzt beim Datentransfer ab), mal sind es in der Eile des Geräteaufbaus Fehler der Aufnehmenden, die die Aufnahme unbrauchbar machen (Gruppenmikrofon wurde nicht eingeschaltet), oder die Nutzungsrechte werden von einer aufgenommenen Person widerrufen, etc.

die Vorführung der bezwingenden Nachvollziehbarkeit der Analyse. Dazu bedarf es hinreichenden Raums – etwa im Umfang einer Dissertation. Gleichwohl sollte die Verschriftlichung der Fallanalyse kein Protokoll der Analyse sein: Dies würde zu einem schier unlesbaren Text führen. Gruschka (2005, 67ff.) schlägt für die einzelne Fallanalyse eine Darstellungsweise vor, die auf Pointierungen ohne Zuspitzung setzt und zugleich die intersubjektive Nachvollziehbarkeit der Rekonstruktion erlaubt.

Die Analyse eines einzelnen Falles kann den Umfang eines Aufsatzes haben. Dies führt aber rasch zu Darstellungsproblemen durch Verkürzungen und durch fehlenden Raum, die einzelnen Interpretationsschritte vorzuführen. Die Darstellungsform reicht aus, um auf ein bestimmtes Strukturphänomen im spezifischen Feld aufmerksam zu machen (z.B. „Üben des Unverstandenen“) und eröffnet eine Forschungsperspektive. Als mittleres Format sind Fallanalysen in der Reihe „Fallanthologien“ konzipiert, die einen einzelnen Fall in besonderer Ausführlichkeit (ca. 100 Seiten) analytisch vorführen und dadurch überzeugender wirken können als die Kurzdarstellungen (z.B. Gruschka 2010). Umfassende strukturtheoretische Bestimmungen eines Phänomens bedürfen letztlich der Darstellung im Rahmen einer Publikation im Dissertationsformat oder als umfangreiche Monographie, welche diverse Fälle entfaltet und in Beziehung zueinander setzt. Gegenwärtig wird diskutiert, wie die umfassenden Transkriptdaten den Rezipientinnen und Rezipienten (für Lehrzwecke, für Reanalysen) verfügbar gemacht werden können.

1.3 Vorteile und Nachteile der Objektiven Hermeneutik

Bildungsprozesse sind notwendigerweise mit der krisenhaften Überwindung bisheriger Routinen im Ich-Welt-Verhältnis verbunden. Die Objektive Hermeneutik ist, wie gezeigt wurde, als sequenzanalytisch rekonstruierende Forschungsmethode bestens für eine bildungsbezogen-fachdidaktische Unterrichtsforschung geeignet (vgl. Gruschka 2009). Sichtbar werden Inszenierungsmuster pädagogischen Handelns und die Strukturlogik ihrer Bearbeitung durch Lehrkraft und Schülerinnen und Schüler. Beobachtbar werden die der Unterrichtskommunikation zugrundeliegenden Normen (Normalitätsregeln) von Lehrkraft, Schülerinnen und Schüler und der Sache (im Sinne von notwendigen Verstehensvoraussetzungen), die zusammen eine *Normentrias* bilden. Diese Normen können konvergent zusammenspielen, was aber eher der Ausnahmefall ist. Sehr häufig sind (zweiseitige oder gar dreiseitige) *Normendivergenzen* Ergebnis der Strukturrekonstruktion. Aufgeklärt werden kann mithin, nach welchen Regeln ein Unterricht abläuft, ganz gleich, ob er funktional oder dysfunktional erscheint. Dabei wird die (Dys-)Funktionalität nicht nach äußeren Normen bemessen, sondern nach den *möglichen* Bildungsprozessen, die der Fall selbst auf Grund der didaktischen Entscheidungen im Vorfeld sowie der kommunikativen Entfaltung *in situ* entwirft, die von der rekonstruktiven Sequenzanalyse aufgeklärt werden müssen.

Die Aufklärung dysfunktionaler Strukturen sollte aus meiner Sicht derzeit dringlich verfolgt werden. Mit ihr könnte sowohl ein Beitrag zu einer *Realistischen*

Deutschdidaktik (Pflugmacher 2014) geliefert werden, als auch ein Beitrag zur Professionalisierung angehender Deutschlehrkräfte – indem die „Schmuddel-ecken“, also die nicht wünschenswerten Aspekte der Literaturvermittlungspraxis, nicht ausgeblendet bleiben, sondern untersucht, hinterfragt, analysiert werden. Dies stellt eine Alternative zur Klärung von Gelingensbedingungen der Literaturvermittlung dar, die dem Modell „best practice-Forschung“ folgt. Was spricht dagegen, dieser eine „bad-practice-Forschung“ entgegen zu setzen, also die Strukturrekonstruktion von Unterricht, der offenkundig nicht seine selbst gesteckten Ziele erreicht? Dafür spräche, einer Position zu entgegnen, gelingende Praxis sei wohlstrukturiert, misslingende Praxis hingegen chaotisch und deshalb nicht erforschenswert. Gezeigt werden müsste, dass beide Struktur haben. Herauszuarbeiten wären darüber hinaus auch Antinomien pädagogischen Handelns im Kontext des Unterrichtsfaches (Lindow/Wieser 2013), insofern didaktisch-methodische Lösungen von Vermittlungsproblemen eben auch mit Schwierigkeiten verbunden sind, die sie unvermeidlich auf der Hinterbühne produzieren.

Objektiv-hermeneutische Forschung ist zugleich Grundlagenforschung und angewandte Forschung. Sie führt zu Aussagen über den jeweiligen Fall. Dieser stellt aber auf Basis der Strukturrekonstruktion zugleich viele Fälle dar, aus denen eine Strukturtheorie der Handlungsmöglichkeiten in einer bestimmten Praxis entfaltet werden kann. Das damit verbundene *professionalisierende Potenzial* für fachdidaktisches Lernen – Denken in konkreten Handlungsalternativen, begründetes didaktisches Entscheiden – kann kaum unterschätzt werden, zumal man an realen Unterrichtssituationen arbeitet. Genau diese Fähigkeiten sind bei der handlungsentlasteten Sequenzanalyse von Unterricht auf der mikrologischen Ebene gefordert.

Bildungsprozesse als innerpsychische Vorgänge bilden sich nicht automatisch kommunikativ ab. Vor diesem Problem stehen viele Forschungsmethoden. Die Lösung, pädagogisch-psychologische *Outcomeforschung* zu betreiben und den Erkenntnisgewinn an den Schülerlösungen abzulesen, führt zu einem vorschnellen Verzicht, die Unterrichtsprozesse und die darin stattfindenden Bildungsprozesse selbst aufzuklären (vgl. Gruschka 2009). Und mit der experimentellen Methode des *Lauten Denkens* (s. Dannecker in diesem Band) kommt man zwar den Bildungsprozessen deutlich näher als mit der Unterrichtsprozessanalyse, da man Irritationen, Krisen und Bearbeitungsmustern folgen kann. Aber auch hier fehlt der Blick auf die bildungsbezogenen Unterrichtsprozesse, denn die Daten werden experimentell und nicht naturwüchsig generiert (Laborsituation). Die objektiv-hermeneutische Unterrichtsprozessforschung kann aber immerhin aufklären, in welchen Inszenierungsmustern die Beteiligten agieren und wie die Beteiligten mit der unvermeidlichen Verunsicherung, die mit Erkenntnisgewinn verbunden ist, umgehen. Um diese Muster zu bestimmen, bedarf es einer Interpretationsgruppe, die in der einzelnen Fachdidaktik im Regelfall nicht verfügbar ist. Häufiger gibt es erziehungswissenschaftliche oder soziologische Fall- bzw. Forschungswerkstätten vor Ort oder andernorts, in denen man das eigene Material vorstellen und interpretieren kann.

Ein Problem der Analyse von Unterrichtskommunikation besteht darin, dass diese Sonderform von zahlreichen Abbrüchen geprägt ist: Fragen verhallen ungehört, Nebengespräche treten auf, in Gruppenarbeitsphasen kann nicht die gesamte Kommunikation im Klassenraum dokumentiert, geschweige denn abgebildet und interpretiert werden. Das bedeutet bei der Auswertung, dass mitunter antizipierend mögliche Anschlüsse entworfen werden, im transkribierten Unterrichtsverlauf aber gar kein wohlgeformter Anschluss festgehalten ist oder erst an viel späterer Stelle erfolgt. Dennoch wird ein Transkript immer genug Stellen enthalten, die sich sequentiell entfalten und stellvertretend für die Logik des Ganzen stehen. Denn Unterricht, der nur aus Abbrüchen bestünde, wäre offenkundig ein pathologischer Grenzfall.

Da die Objektive Hermeneutik eine Methode der pädagogischen Grundlagenforschung ist, ist der unkontrollierte Rückgriff auf fachdidaktische Terminologie im Prozess der Analyse oftmals eher problematisch: Diese ist Ausdruck des Wunschkens, wie *Deutschunterricht sein sollte*, und verstellt deshalb bei der Rekonstruktion dessen, was in ihm tatsächlich geschieht, eher den Blick. Den im Studium antrainierten subsumierenden, kategorisierenden und wertenden Blick auf die rekonstruierten Phänomene im Unterricht abzulegen ist ein schwieriger, mitunter längerfristiger Lernprozess – obwohl die Objektive Hermeneutik eigentlich ein recht einfaches, lernbares Interpretationsverfahren darstellt.

Der Sachbezug ist eine lösbare Herausforderung für die sequenzanalytische fachdidaktische Unterrichtsforschung (vgl. Gruschka 2009; 2010). Da sich die sequenzanalytische fachdidaktische Unterrichtsforschung momentan in einer Aufbauphase befindet, sind deshalb fachdidaktische Nachwuchsforscherinnen und -forscher oft auf Workshops und Kolloquien in der Erziehungswissenschaft oder der Soziologie angewiesen. Hier gilt es, auch auf Grund der dort knapp bemessenen Zeit, konsequent das Forschungsthema als Fall vorzustellen und nicht etwa den Transkriptbeginn mit dem Begrüßungsritual ausgiebig zu analysieren. Geklärt werden muss rechtzeitig: Was ist der fachdidaktisch relevante Fall – in Konkurrenz zu allgemeindidaktisch-schulpädagogischen Fällen? Ungeachtet dessen ist die gemeinsame Analyse mit Nichtfachdidaktikerinnen und -didaktikern eine wichtige und bereichernde Erfahrung.

Denn die rekonstruktionslogische Sequenzanalyse erfordert und inszeniert die Bereitschaft zur Verfremdung des vermeintlich Vertrauten. Die Auswertung erfolgt extrem verlangsamt, geradezu in Zeitlupe. Vertrautes wird systematisch in Frage gestellt, um dessen Gelingensbedingungen zu explizieren und beschreibbar zu machen. Die Beteiligung von Nichtexperten an der Auswertung ist daher nicht problematisch, sondern geradezu willkommen. Sehr kurze Textauszüge des Materials führen dabei insgesamt zu äußerst umfangreichen Interpretationsprotokollen und Auswertungen und zu der Frage, inwieweit sie im Rahmen der Darstellung der Ergebnisse dokumentiert werden können. Dies ist ein sich wiederholendes strukturelles Lesbarkeitsproblem: Wie den Umfang minimieren und dabei doch die Nachvollziehbarkeit sichern?

1.4 Gütekriterien

Abschließend seien hier die wichtigsten Gütekriterien für den Einsatz des Verfahrens der Objektiven Hermeneutik zusammengefasst:

Notwendig ist eine Interpretationsgemeinschaft, um extensiv Lesarten zu bilden, zu prüfen und zu verwerfen. Das Auswertungsverfahren soll sich an den fünf Maximen der Kunstlehre der Objektiven Hermeneutik orientieren: Sequenzialität, Wörtlichkeit, Extensivität, Kontextfreiheit und Sparsamkeit (s. 1.1). Die Interpretation erfolgt nicht theoriesprachlich subsumierend, sondern in der Sprache des Falles. Das Verfahren beruht auf dem Anspruch der Falsifizierbarkeit. Der Geltungsanspruch der Fallrekonstruktion ist ein besonderer und allgemeiner zugleich: Anstelle der empirischen Generalisierung der quantifizierenden Unterrichtsforschung führt die Objektive Hermeneutik zu einer Strukturgeneralisierung. Erst die Strukturreproduktion erlaubt es, ausgehend von der im Fall entwickelten Fallstrukturhypothese zu einer Fallstrukturgesetzmäßigkeit zu kommen. Dafür ist eine Methodentriangulation nicht notwendig. Wichtig ist allerdings die Sicherung der Nachvollziehbarkeit der Ergebnisse in der Darstellung der Forschungsergebnisse.

1.5 Literaturempfehlungen

Oevermann (2000): Die Methode der Fallrekonstruktion in der Grundlagenforschung sowie der klinischen und pädagogischen Praxis.

Dies ist ein jüngerer methodologischer Grundlagentext zur Objektiven Hermeneutik als Methode der Fallrekonstruktion. Alternativ und kürzer zu lesen ist Oevermanns Manifest zur Objektiven Hermeneutik, welches online verfügbar ist. Für Dissertationsvorhaben ist der Aufsatz zu bevorzugen.

Gruschka (2005): Auf dem Weg zu einer Theorie des Unterrichtens – die widersprüchliche Einheit von Erziehung, Didaktik und Bildung in der allgemeinbildenden Schule.

In der umfangreichen Vorstudie (die Fallanalysen wurden gemeinsam mit Ulrich Oevermann durchgeführt) zum Projekt „Pädagogische Rekonstruktion des Unterrichtens (PAERDU)“ werden zahlreiche Fragen zur Transkription, zur Methode sowie zu den Möglichkeiten und Fallstricken der Ergebnisdarstellung pädagogischer Kasuistik mit Objektiver Hermeneutik aufgegriffen und beantwortet. Für fachdidaktische Forschung ist die Integration der Bildungsdimension und damit die der Sache in das Forschungsprojekt von erheblicher Bedeutung, die in Gruschkas Erkenntnisbuch 2009 sowie in die Analyse einer Deutschstunde im Rahmen der Fallanthologiereihe (2010) münden.

Gruschka (2011): Pädagogische Unterrichtsforschung als Erforschung der Pädagogik.

Umfangreich wird vorgeführt, inwiefern die soziologische Methode der Objektiven Hermeneutik ein ideales Instrument für eine *pädagogische* Bildungsforschung darstellt.

Wernet (2000): Einführung in die Interpretationstechnik der Objektiven Hermeneutik.

Wernet führt kleinschrittig anhand von Mikrosequenzen in die Interpretationstechnik der Objektiven Hermeneutik ein, behandelt aber – anders als Gruschka – ausschließlich erziehungswissenschaftliche Fragestellungen. Für Einsteiger die erste Wahl zur Orientierung über das Verfahren.

Pflugmacher (2019): Literatur lehren.

Pflugmacher zeigt auf Basis einer Reihe von Fallanalysen auf, inwiefern eine literaturdidaktische Forschung mit dem Verfahren der Objektiven Hermeneutik zahlreiche zentrale Handlungsprobleme der Literaturvermittlung herausarbeiten kann. Die Ergebnisse sind gleichermaßen bedeutsam für eine Theorie der Literaturvermittlung wie für eine literaturpädagogische Professionalisierung angehenden Deutschlehrender durch Fallarbeit.

2. Beispielprojekt: Realistische Literaturdidaktik

Basierend auf Erfahrungen aus der schulpädagogischen Frankfurter Forschungsgruppe „Pädagogische Rekonstruktion des Unterrichtens“ (PAERDU) soll eine empirisch gehaltvolle Theorie der Literaturvermittlung begründet werden. Einerseits ist der fachdidaktische Diskurs traditionell idealistisch-normativ und von der alltäglichen Unterrichtspraxis weitgehend abgekoppelt, wenn seine Protagonisten *innovationslogisch* neue Vermittlungsgegenstände und -verfahren generieren. Andererseits sollte die stark normierende empirische Bildungsforschung Ergänzung finden durch eine Forschungsrichtung, die einen ebenso kritischen wie wertschätzenden Blick auf die Praxis richtet und deren Eigenlogik aufklärt, anstelle sie an einer äußeren Richtschnur zu messen und dafür schon aus methodischen Gründen das Outcome heranzieht und nicht die Bildungsprozesse verfolgen kann.

In der *pädagogischen Professionsforschung* haben sowohl etablierte wie neue Forscherinnen und Forscher seither eine beachtliche Produktivität an den Tag gelegt und konnten die Bedeutung qualitativer Forschung und kasuistisch-professionalisierender Arbeit in der Lehrerausbildung aufzeigen.¹⁰ Auch in der Deutschdidaktik gibt es mittlerweile eine wachsende fachspezifische Professionsforschung – wenn auch zumeist im Rahmen einer Lehrerwissenschaft. Die Idee, eine Theorie der Literaturvermittlung auf Basis der Rekonstruktion alltäglichen Unterrichts zu formulieren, ist hingegen noch immer ein Desiderat. Dies liegt vermutlich daran, dass in einem verbreiteten wissenschaftstheoretischen Verständnis Theorie als Werkzeug gesehen wird, mit dem Praxis aus einer bestimmten Perspektive beschreibbar gemacht wird. Der mit der Objektiven Hermeneutik verbundene Anspruch, gewissermaßen entgegengesetzt aus der mikrologischen Rekonstruktion der Praxis selbst die Theorie ihrer Entfaltungsmöglichkeiten und -beschränkungen zu generieren, dürfte Fachdidaktikerinnen und -didaktikern vor diesem Hintergrund seltsam erscheinen.

In der explorativen Phase, die noch nicht abgeschlossen ist, eröffnete jede einzelne Analyse einer Literaturunterrichtsstunde den Blick auf ein neues und allgemeines Strukturproblem der Literaturvermittlung, welches im literaturdidaktischen Diskurs wenig bis gar nicht wahrgenommen oder diskutiert wurde. Dabei handelt es sich um Einzelfallstudien, die jede für sich im Rahmen beispielsweise eines Dissertationsvorhabens durch kontrastive Erhebungen im oben beschriebenen Verfahren erweitert werden müssten. Die Publikation von etwa sechs bis acht Fällen soll vorführen, dass eine deutschdidaktische Unterrichtsforschung möglich ist, welche nicht aus forschungsmethodischen Gründen von vorneherein die Komplexität der Praxis reduziert, wie etwa in Ratingverfahren (s. Kleinbusch in diesem Band), sondern ihre strukturelle Aufklärung sequenzanalytisch verfolgt. Die Analysen sollen dabei zeigen, dass bestimmte fachdidaktische Phänomene weiterer

¹⁰ Dass die beiden Wissenschaftskulturen weiterhin von wechselseitiger Ignoranz geprägt sind, kann man im ‚Handbuch der Forschung zum Lehrerberuf‘ (Terhart/Bennett/Rothland 2014) nachlesen.

Aufklärung lohnen: Dies umfasst sowohl typische Inszenierungsmuster, weiterhin fachspezifische Dilemmasituationen, und nicht zuletzt typische Fehlformen der Literaturvermittlung, die der Zunft zwar mitunter bekannt sind, aber gerade weil sie als nicht wünschenswert markiert werden, nicht als Herausforderungen für eine literaturdidaktische Theoriebildung betrachtet wurden. Explorativ-rekonstruktiv aufgeklärt wurden unter anderem folgende Bereiche:

- Inszenierungsmuster ästhetischer Erfahrung (vgl. Pflugmacher 2011)
- Literaturwissenschaftsdidaktik als Überformung von Literaturdidaktik
- Umgang mit didaktischen Fertigmaterialien (vgl. Pflugmacher 2007)
- Herausforderungen stellvertretender Deutung des Weltwissens von Schülerinnen und Schülern im interkulturellen Kontext (vgl. Pflugmacher 2015)
- Umgang mit literarischer Textschwierigkeit
- Inszenierungsmuster von Ergebnissen
- Umgang mit literarischen Figuren
- Was ist und wie erklärt man Spannung?
- Entwicklung einer Fragehaltung zum Text
- Umgang mit Nichtlesestrategien von Schülerinnen und Schülern

Exemplarisch soll hier als Referenzbeispiel meine Fallanalyse zu Inszenierungsmustern ästhetischer Erfahrung in knapper Form vorgestellt werden. Die mittellange Fassung ist in Pflugmacher 2011 nachzulesen, die Langfassung im Rahmen eines größeren Projekts steht noch aus.

Ausgangspunkt der Analyse war ein Transkript einer Unterrichtsstunde, in der erste Lesebegegnungen mit der Novelle „Kleider machen Leute“ von Gottfried Keller ausgewertet werden sollten.

Der Lehrer inszeniert sich zunächst analog zu einem erlebnispädagogischen Führer, der seinen Klienten einen ersten Einstieg wie auf einer Entdeckungstour verspricht. Dann führt er fort, dass er in der Stunde an den Eindrücken seiner Schülerinnen und Schüler zu „Kleider machen Leute“ weiter arbeiten wolle – was im Sinne einer Theorie ästhetischer Erfahrung unter schulischen Bedingungen sehr präzise das pädagogisch-professionelle Verhältnis beschreibt, nämlich die Spuren zu verfolgen, zu bearbeiten und zu vertiefen, die der ästhetische Gegenstand selbst schon hinterlassen hat (vgl. Oevermann 1996). Damit liefert er ein eindrückliches Beispiel für die normsetzende Eröffnung einer Sequenz, die immanente Normativität, zu deren Bestimmung auch die Befolgung der Prinzipien Kontextfreiheit und Wörtlichkeit beigetragen hat. Denn letztlich gibt der Lehrer mit seiner Inszenierung das Versprechen, das ästhetische Erleben der Schülerinnen und Schüler mit seiner Hilfe in ästhetische Erfahrung zu überführen. Die Gegenlesart, dass der Lehrer die Eindrücke abschleifen, also vereinheitlichen will, wurde als unsinnig verworfen.

Nach diesem Generalversprechen strukturiert er den weiteren Stundenverlauf methodisch. Die Schülerinnen und Schüler sollen sich in Gruppengesprächen darüber austauschen, was sie einerseits an der Lektüre positiv überrascht, andererseits was sie dabei abgeschreckt hat. Er verbleibt also im Bereich ästhetischer

Erfahrungsmodi, die in Gruppenarbeit versprachlicht werden sollen, andererseits werden diese schon im Vorfeld durch die Benennung der emotionalen Extrempole Überraschung und Abschreckung didaktisch zugespitzt. Diese Ambivalenz setzt sich fort: Die Schülerinnen und Schüler sollen sich „völlig offen ausfahren“ und haben dafür „7 Minuten“ Zeit. Im Anschluss sollen die Erfahrungen vorgetragen und „gebündelt“ werden. Der Lehrer zeigt hier unfreiwillig, dass die versprochene unlimitierte (bildsame) *Entfaltung* der Erfahrung so nicht umgesetzt werden kann. Sie wird fortlaufend von didaktischen *Verdichtungsoperationen* überformt, ohne dass expliziert wird, welche Funktion die Verdichtungen haben sollen.

Der Lehrer, so zeigt sich hier, betreibt eine Art Verweigerungshaltung gegenüber antinomischen Entscheidungen. Er will gewissermaßen beides: Entfaltung und Verdichtung zugleich. Dafür hat er aber keinen Modus, und es ist fraglich, ob es hinsichtlich der Anbahnung ästhetischer Erfahrung im schulischen Kontext überhaupt einen solchen gibt. Fraglich ist dann aber, wie es weiter gehen kann. Was machen seine Schülerinnen und Schüler in Anbetracht der unlösbaren Herausforderung? Sie entscheiden für ihn: „Wer schreibt auf?“ Es wird gebündelt. „Was hat uns beeindruckt?“ wird als gemeinsame Frage formuliert. Ein Schüler ist irritiert und fragt: „Uns?“ Ihm ist klar, dass individuelle Eindrücke nicht als kollektive zusammengefasst werden können. Andere Schüler treiben voran, solche Irritationen stören bei der Ergebnisproduktion in der knapp bemessenen Zeit. Die individuellen Eindrücke werden ausgetauscht und sofort evaluiert hinsichtlich der Frage, ob man das überhaupt aufschreiben könne, und welches Antwortspektrum der Lehrer wohl entfaltet habe. Zensurversuche finden statt: Ein Leseindruck, der die angenehme Kürze des Textes würdigt, wird von Mitschülern zunächst abgelehnt, denn es gehe dem Lehrer doch wohl um den ‚Inhalt‘ und die ‚Sprache‘. Indem ein Schüler überhaupt aufschreibt, obwohl dies nicht Teil der gestellten Aufgabe zum Austausch war, zeigen die Schülerinnen und Schüler, dass vertraute didaktische Normen (Ergebnissicherung) für sie selbst dann gelten, wenn sie gar nicht expliziert wurden und bei näherer Betrachtung auch kontraproduktiv erscheinen müssten. Denn es wird nicht gesammelt, sondern bereits verdichtet, weil festgelegt wird, was aufgeschrieben wird und was nicht.

Im Plenum wird dann weiter verdichtet, wenn der Lehrende z.B. Schüleraussagen zusammenfassend umformuliert, anstelle nachzuhaken, worauf die geschilderte Erfahrung basiert.

Aus diesem Fall ließ sich schon nach den ersten Interakten eine Vierfeldermatrix zu den immanent-normativen Bedingungen ästhetischer Erfahrung im Literaturunterricht entwickeln:

	Entfaltung schülerseits (als Anspruch/Artikulation)	Verdichtung schülerseits
Entfaltung lehrerseits (als Versprechen, Verlangen, Anmutung)	Fall A: Schüler nehmen das Angebot ästhetischer Erfahrung an, lassen sich irritieren und spüren der eigenen Irritation nach, der im Unterrichtsetting entsprechend Raum gegeben wird.	Fall B: Schüler transformieren das Angebot ästhetischer Erfahrung in „Normal“-Unterricht um und behandeln den ästhetischen Gegenstand als Lerngegenstand.
Verdichtung lehrerseits	Fall C: (Einzelne) Schüler beharren gegen den Lerngegenstand Literatur auf der Artikulation (eigensinniger) ästhetischer Erfahrung.	Fall D: Schüler machen den „Normal“-Unterricht bereitwillig mit.
Ergänzungsstrukturen	Fall E: lehrerseits ein Sowohl-als-auch	Fall F: lehrerseits ein Weder-noch

Tab. 1: Fallstrukturen zur ästhetischen Erfahrung im Literaturunterricht

Bemerkenswert im Fall ist, dass die aus dem Fall selbst heraus entwickelte Vierfeldermatrix zu den strukturell erwartbaren Handlungsmustern wider Erwarten ergänzt werden musste. Denn der Lehrende verhielt sich, wie die Fallanalyse gezeigt hat, konsequent inkonsequent, er formulierte unablässig mischend Entfaltungs- und Verdichtungsangebote, die sich wechselseitig ausschlossen (Fall E). Daraus folgte, dass nicht er entschied, sondern die Schülerinnen und Schüler herausgefordert waren, wie sie mit der Herausforderung, ästhetische Erfahrung artikulieren zu *sollen*, umgehen könnten. Sie mussten die Entscheidungen des Lehrenden an seiner Stelle treffen. Die Operation der stellvertretenden Deutung fand gewissermaßen umgekehrt statt. Nicht der Lehrende versetzte sich in die Schülerinnen und Schüler hinein, sondern die Schülerinnen und Schüler in den Lehrenden. Möglicherweise auf Grund ihrer fachkulturellen Erfahrungen entschieden sie sich für Verdichtungsoperationen und verzichteten darauf, ihr ästhetisches Erleben zu entfalten. Sie konnten das im Raum stehende Problem, wie man unter schulischen Rahmenbedingungen ästhetische Erfahrung im Umgang mit Literatur machen soll, nicht stellvertretend für den Lehrenden lösen.

Gezeigt werden konnte, wie an einem Fall auch die anderen möglichen Fälle sichtbar wurden. Darüber hinaus wurde deutlich, dass die Wirklichkeit auch in ihren Abweichungen von idealer Kommunikation immer noch strukturell beschreibbar ist. Dies ist enorm wichtig, um nicht vorschnell die Praxis als pathologisch zu diffamieren.

Die professionstheoretischen Einsichten liegen hier in ungewöhnlicher Deutlichkeit auf der Hand: Man kann sich im Prozess des didaktischen Handelns nicht *nicht* entscheiden. Hier steckt der Lehrende sprichwörtlich den Kopf in den Sand und die Schülerinnen und Schüler übernehmen mit einer Art Ersatzdidaktik das Ruder.

Die skizzierten Fallstrukturen können als Inszenierungsmuster ästhetischer Erfahrung im Literaturunterricht aufgefasst werden. Im Anschluss an diese Pilotstudie sollen kontrastive Fälle erhoben und untersucht werden. Dadurch sollen zum einen die aus dem Fall heraus generierten theoretischen Fälle empirisch nachgewiesen werden. Zum anderen sollen Rahmenbedingungen geprüft werden – ob ästhetische Erfahrung im außerunterrichtlichen schulischen Lernen eher möglich ist, ob in der Oberstufe oder im Primarbereich andere Normen gelten, die ästhetische Erfahrungskommunikation erleichtern, etc. Dies ist aber nur im Rahmen eines Drittmittelprojekts bzw. eines Dissertations- oder Habilitationsvorhabens zu leisten.

Literatur

- Archiv für pädagogische Kasuistik (Apaek). URL: <http://www.apaek.de>.
- Dittmar, Norbert: Transkription. Ein Leitfaden mit Aufgaben für Studenten, Forscher und Laien. Wiesbaden: VS 2009.
- Feilke, Helmuth: Überarbeiten! Überlegungen zu Bildungsstandards, Textkompetenz und Schreiben. In: *Didaktik Deutsch* (2014), 37, 6-9.
- Fuß, Susanne/Karbach, Ute: Grundlagen der Transkription: Eine praktische Einführung. Leverkusen: Budrich 2014.
- Garz, Detlef/Raven, Uwe: Theorie der Lebenspraxis. Einführung in das Werk Ulrich Oevermanns. Wiesbaden: VS 2015.
- Gruschka, Andreas/Pflugmacher, Torsten/Twardella, Johannes/Rosch, Jens: PAERDU. Pädagogische Rekonstruktion des Unterrichtens. (2006). <http://www.uni-frankfurt.de/fb/fb04/forschung/paerdu.html> (letzter Zugriff 12.04.2011).
- Gruschka, Andreas: Auf dem Weg zu einer Theorie des Unterrichtens – Die widersprüchliche Einheit von Erziehung, Didaktik und Bildung in der allgemeinbildenden Schule. *Frankfurter Beiträge zur Erziehungswissenschaft, Forschungsberichte Band 5*. Frankfurt a.M.: Fachbereich Erziehungswissenschaften 2005.
- Gruschka, Andreas: Die Grenzen des Unterrichtens. Leverkusen: Budrich 2010.
- Gruschka, Andreas: Erkenntnis in und durch Unterricht. Wetzlar: Büchse der Pandora 2009.
- Gruschka, Andreas: Pädagogische Unterrichtsforschung als Erforschung der Pädagogik. Leverkusen: Budrich 2011.
- Gruschka, Andreas: Verstehen lehren. Plädoyer für guten Unterricht. Stuttgart: Reclam 2011.
- Lindow, Ina/Wieser, Dorothee: Literaturunterricht als Fall philologischer Praxis: Antinomien und Lösungsversuche. In: *Zeitschrift für Germanistik* (2013), 2, 390-404.
- Lindow, Ina: Literaturunterricht als Fall. Kasuistisches Wissen von Deutschlehrenden. Wiesbaden: VS 2013.

- Oevermann, Ulrich/Allert, Tilman/Konau, Elisabeth/Krambeck, Jürgen: Die Methodologie einer „objektiven Hermeneutik“ und ihre allgemeine forschungslogische Bedeutung in den Sozialwissenschaften. In: Soeffner, Hans-Georg (Hrsg.): Interpretative Verfahren in den Sozial- und Textwissenschaften. Stuttgart: Metzler 1979, 352-434.
- Oevermann, Ulrich: Die Methode der Fallrekonstruktion in der Grundlagenforschung sowie der klinischen und pädagogischen Praxis. In: Kraimer, Klaus/Garz, Detlef (Hrsg.): Die Fallrekonstruktion. Frankfurt a.M.: Suhrkamp 2000, 58-156.
- Oevermann, Ulrich: Klinische Soziologie auf der Basis der Methodologie der objektiven Hermeneutik – Manifest der objektiv hermeneutischen Sozialforschung. (2002). URL: http://www.ihsk.de/publikationen/Ulrich_Oevermann-Manifest_der_objektiv_hermeneutischen_Sozialforschung.pdf (letzter Zugriff: 01.08.2018).
- Oevermann, Ulrich: Skizze einer revidierten Theorie professionalisierten Handelns. In: Combe, Arno/Helsper, Werner (Hrsg.): Pädagogische Professionalität. Untersuchungen zum Typus pädagogischen Handelns. Frankfurt a.M.: Suhrkamp 1996, 70-182.
- Oevermann, Ulrich: Krise und Muße. Struktureigenschaften ästhetischer Erfahrung aus soziologischer Sicht. Vortrag am 19.06.1996 in der Städelschule. URL: <https://dnb.info/974364967/34> (letzter Zugriff: 01.08.2018).
- Online-Fallarchiv Schulpädagogik der Universität Kassel. <http://www.fallarchiv.uni-kassel.de/lernumgebung/lehrfilm/> (letzter Zugriff: 01.08.2018).
- Online-Fallarchiv Schulpädagogik der Universität Kassel. URL: <http://www.fallarchiv.uni-kassel.de/lernumgebung/objektive-hermeneutik/> (letzter Zugriff: 01.08.2018).
- Pflugmacher, Torsten/Gruschka, Andreas/Twardella, Johannes/Rosch, Jens: Vom Nutzen einer pädagogischen Unterrichtsforschung. In: Beiträge zur Lehrerbildung. Zeitschrift zu Theorie und Praxis der Aus- und Weiterbildung von Lehrerinnen und Lehrern (2009), 3, 372-284.
- Pflugmacher, Torsten/Twardella, Johannes: Beantwortung häufig gestellter Fragen zum Projekt PAERDU (2009). http://www.uni-frankfurt.de/fb/fb04/download/Paerdu_FAQ.pdf (letzter Zugriff: 12.04.2011).
- Pflugmacher, Torsten: „Und vor allen Dingen an Euren Eindrücken auch ein bisschen arbeiten.“ Eine Fallrekonstruktion zu ästhetischer Erfahrung im Literaturunterricht. In: Kirschenmann, Johannes/Spinner, Kaspar H./Richter, Christian (Hrsg.): Reden über Kunst. München: kopaed 2011, 121-135.
- Pflugmacher, Torsten: Deutschunterricht und Didaktikindustrie. Kritische Theorie nach ihrer empirischen Wende. In: Bönnighausen, Marion/Michael Baum (Hrsg.): Kulturtheoretische Kontexte der Literaturdidaktik. Baltmannsweiler: Schneider Hohengehren 2010, 47-61.
- Pflugmacher, Torsten: Literatur lehren. Eine Einführung in die Literaturdidaktik mit Fallbeispielen. Paderborn: Schöningh 2019 (i.V.).
- Pflugmacher, Torsten: Möglichkeiten und Grenzen der Fallarbeit in der DeutschlehrerInnenausbildung. In: Pieper, Irene/Frei, Peter/Hauenschild, Katrin/Schmidt-Thieme, Barbara (Hrsg.): Was der Fall ist: Fallarbeit in Bildungsforschung und Lehrerbildung. Wiesbaden: VS 2014.
- Pflugmacher, Torsten: Try Pattern and Drill Error. Zwei Fallanalysen zur Fast-Food-Didaktik mit didaktischen Fertigmaterialeien. In: Pädagogische Korrespondenz.

Zeitschrift für kritische Zeitdiagnostik in Pädagogik und Gesellschaft (2007), 37, 81-106.

- Pflugmacher, Torsten: Verstehen verstehen verstehen als literaturpädagogische Kompetenz. In: Bräuer, Christoph/Wieser, Dorothee (Hrsg.): Die Lehrkraft im Blick – empirische Lehrerforschung in der Deutschdidaktik. Wiesbaden: VS 2015, 129-156.
- Pieper, Irene/Frei, Peter/Hauenschild, Katrin/Schmidt-Thieme, Barbara (Hrsg.): Was der Fall ist: Fallarbeit in Bildungsforschung und Lehrerbildung. Wiesbaden: VS 2014.
- Pollmanns, Marion/Belak, Jan: Leitfaden für die gerätegestützte Aufzeichnung von Unterricht. URL: <https://www.uni-frankfurt.de/58404447/aufnahmeleitfaden.pdf> (letzter Zugriff: 01.08.2018).
- Pollmanns, Marion: Leitfaden für die Transkription von Unterricht. <http://www.apaek.uni-frankfurt.de/angebote/sps/leitfaeden/transkription.html> (letzter Zugriff: 20.11.2015).
- Scherf, Daniel: Leseförderung aus Lehrersicht. Eine qualitativ-empirische Untersuchung professionellen Wissens. Wiesbaden: VS 2013.
- Terhart, Ewald/Bennewitz, Hedda/Rothland, Martin: Handbuch der Forschung zum Lehrerberuf. Münster: Waxmann 2014.
- Wernet, Andreas: Einführung in die Interpretationstechnik der Objektiven Hermeneutik. Wiesbaden: VS 2006.
- Wieser, Dorothee: Literaturunterricht aus Sicht der Lehrenden. Eine qualitative Interviewstudie. Wiesbaden: VS 2008.

Qualitative Inhaltsanalyse

Mit dem Rüstzeug der QIA neue Wege gehen

1. Vorstellung der qualitativen Inhaltsanalyse

Die Qualitative Inhaltsanalyse (QIA) ist ein qualitatives Auswertungsverfahren der Sozialforschung, das heute auf einen breiten Einsatzbereich in unterschiedlichen Disziplinen blicken kann und eine lange Tradition besitzt. Durch die kommunikationswissenschaftliche Fundierung groß angelegter quantitativ orientierter Massenmedienanalysen in den 20er und 30er Jahren des 20. Jahrhunderts wurden die Grundlagen für ihr systematisches Vorgehen geschaffen. Die Aufspaltung der Inhaltsanalyse (IA) in zwei „polare Methodik-Konkretisierungen“ (Groeben/Rustemeyer 2002, 237) geht auf die Kritik Kracauers gegen die Fokussierung manifester Informationen sowie vorschneller Quantifizierungen zurück und umfasst die Forderung nach Berücksichtigung latenter Bedeutungsdimensionen.

1.1 Funktionsweise der qualitativen Inhaltsanalyse

Die Funktionsweise der QIA vorzustellen, erscheint vor der folgenden Einschätzung Schreiers als nicht ausführbar: „Die qualitative Inhaltsanalyse gibt es nicht, und es besteht kein Konsens darüber, was qualitative Inhaltsanalyse ausmacht“ (Schreier 2014, Abs. 4). Anstelle in eklektischer Manier einige konkrete Ablaufschemata vorzustellen, die letztlich für die Anwendung in eigenen Forschungsvorhaben zu abstrakt bleiben oder den Blick in unnötiger Weise verengen¹, sollen im Folgenden Merkmale skizziert werden, die als Orientierungspunkte „im Dickicht der Begrifflichkeiten“ (Schreier 2014) einer kaum mehr zu überblickenden QIA dienen. Die QIA ist (a) ein datenreduzierendes, (b) kategorienorientiertes, (c) auch latente Bedeutungen berücksichtigendes, (d) systematisches und regelgeleitetes Vorgehen zur Analyse von bedeutungshaltigem Material.

(a) Die QIA ist ein datenreduzierendes (vgl. Groeben/Rustemeyer 2002, 234; Früh 2011, 42) Verfahren zur Erfassung von Textbedeutungen. Ihr Anspruch besteht nicht darin, alle möglichen, sondern „immer nur ausgewählte Bedeutungs-

¹ Konkrete Ablaufschemata zu unterschiedlichen inhaltsanalytischen Techniken bzw. Basismodellen finden sich z.B. bei Mayring (2010) oder Kuckartz (2012).

aspekte“ (Schreier 2006, 428) zu fokussieren, um einen „[k]omplexitätsreduzierende[n] Überblick (über Textbedeutungen)“ (Groeben/Rustemeyer 2002, 235) zu gewähren. Groeben und Rustemeyer (2002) sehen in der Datenreduzierung aber keine Beschränkung der QIA, sondern vielmehr eine Notwendigkeit empirischer Methoden generell. Sie konstatieren: „[E]in systematisch-methodisches Vorgehen wird immer bestimmte Merkmale eines Gegenstandes herausheben müssen, um sie möglichst präzise ‚abzubilden‘“ (ebd., 247). Hat man beispielsweise transkribierte Aushandlungsgespräche von Lernenden aus der Bearbeitung von Aufgaben zu einem literarischen Text zur Auswertung vorliegen, so wäre es denkbar, herauszuarbeiten, (i) welche Aspekte des zugrundeliegenden literarischen Textes die Lernenden berücksichtigen, (ii) über den Vollzug welcher Verstehensoperationen die Verstehensresultate Auskunft geben oder, im ersten Moment ggf. überraschend, (iii) worüber die Lernenden parallel zur Aufgabenbearbeitung sprechen. Das Anliegen, all diese Bedeutungsaspekte in der Auswertung zu berücksichtigen, muss am eigenen Anspruch scheitern. In der QIA werden mithin in Abhängigkeit zur konkreten Forschungsfrage Bedeutungsaspekte bestimmt, auf die das Datenmaterial reduziert werden soll.² Dementsprechend kann nicht jede Forschungsfrage inhaltsanalytisch ausgewertet werden. Eine Passung besteht nur dann, wenn sich die Fragestellung und die in ihr enthaltenen Variablen in inhaltsanalytischen Kategorien operationalisieren lassen (vgl. Groeben/Rustemeyer 2002, 240; Hussy et al. 2013, 256).

(b) Die Kategorienorientierung unterscheidet die QIA von sequenzanalytischen Verfahren wie z.B. der Objektiven Hermeneutik (s. Pflugmacher in diesem Band) oder der Diskurs- oder Konversationsanalyse (s. Heller/Morek in diesem Band). Im Zentrum der QIA steht das Kategoriensystem, dessen Bezeichnung als „Herzstück der Inhaltsanalyse“ (Schreier 2014, Abs. 4) in der folgenden Charakterisierung durch Berelson verständlich wird:

Die Inhaltsanalyse steht und fällt mit ihren Kategorien, [...] da die Kategorien die Substanz der Forschung enthalten, kann eine Inhaltsanalyse nicht besser sein als ihr Kategoriensystem. (Berelson 1952, 147 zit. n. Kuckartz 2012, 40)

Expliziert werden in dem Kategoriensystem die interessierenden Bedeutungsaspekte. Folgerichtig bezeichnen Groeben und Rustemeyer (2002) das Kategoriensystem als „explizite Systematisierung“ einer „unvermeidbaren Merkmalsfokussierung“ (Groeben/Rustemeyer 2002, 248) und Früh charakterisiert Kategorien als Suchstrategien: Die interessierenden Bedeutungsaspekte werden systematisch aus dem Material herausgefiltert und der Informationsgehalt des Gegenstandes nach Maßgabe des Forschungsinteresses reduziert (vgl. Früh 2011, 124). Die folgenden Bestandteile sind obligatorisch in einem Kategoriensystem: Kategorien-

² Darin unterscheidet sich die QIA von sequenzanalytischen Verfahren, die das Datenmaterial „vermehren und in exegetischer Absicht interpretieren“ (Kuckartz 2012, 76). In hermeneutischer Tradition steht dabei der einzelne Text bzw. Gegenstand im Fokus, mithin diese Verfahren für den Merkmalsvergleich größerer Textmengen weniger geeignet erscheinen (vgl. Hussy et al. 2013, 255f.; Früh 2011, 64ff.).

bezeichnung, Definition und Beispiel zur Illustration der Kategorie. Angenommen das Erkenntnisinteresse besteht darin, aus den Transkripten von Aufgabebearbeitungsprozessen herauszuarbeiten, worüber die Lernenden parallel noch sprechen, und angenommen die Lernenden sprechen über kommende Wochenendaktivitäten, dann sieht ein Kategoriensystem in einfachster Form wie folgt aus (Tabelle 1).

Bezeichnung	Definition	Beispiel
Kommendes Wochenende	Diese Kategorie ist zu kodieren, wenn in den Gruppen über das kommende Wochenende und nicht über die Aufgabebearbeitung gesprochen wird.	„Was machst du am Wochenende? Ich treffe mich mit Freunden aus dem Jugendcamp.“ „Ich werde im Internet Spiele zocken.“

Tab. 1: Beispiel Kategoriensystem

In aller Regel setzt sich ein Kategoriensystem aus mehreren nebeneinander bestehenden Oberkategorien zusammen, die wiederum in Unterkategorien oder Ausprägungen ausdifferenziert sein können. Zur oben angeführten Kategorie ließen sich die Unterkategorien ‚Freunde treffen‘ (‚Freunde aus dem Jugendcamp‘) oder ‚Online Spielangebote nutzen‘ (‚im Internet zocken‘) bilden. Das Kategoriensystem ist dann hierarchisch organisiert.³ Wie viele Ebenen von Unterkategorien angemessen sind, lässt sich nur anhand einer konkreten Forschungsfrage bestimmen. An dieser Stelle soll aber auf einen Fallstrick der QIA hingewiesen werden: Es ist mitunter zu beobachten, dass extrem ausdifferenzierte Kategoriensysteme entwickelt werden, in denen einzelne Aspekte in drei und mehr Unterkategorieebenen unterschieden sind. Inwiefern das Kategoriensystem schlussendlich zur Beantwortung der Forschungsfrage beitragen kann, bleibt aber unklar. Weil Oberkategorien auch dann noch in Unterkategorien ausdifferenziert werden können, wenn sich ihre besondere Bedeutung im Forschungsprozess zeigt, hat es sich als sinnvoll erwiesen, mit einem geringen Differenzierungsgrad zu beginnen. Nutzt man ein Programm zur computerunterstützten Analyse (beispielsweise MAXQDA⁴), ist es möglich, sich alle kategorisierten Textstellen zu einer Kategorie anzeigen zu lassen, diese zu sortieren und auf diesem Wege Systematisierungen auf der Ebene der Unterkategorien zu bilden. Leitend muss dabei immer die Frage sein, inwiefern die Unterkategorien zur Generierung von Erkenntnissen

³ Bei den Beispielkategorien handelt es sich um inhaltlich-strukturierende Kategorien. Kuckartz unterscheidet darüber hinaus Fakten-Kategorien, inhaltliche Kategorien, analytische Kategorien, natürliche Kategorien, evaluative Kategorien und formale Kategorien (vgl. Kuckartz 2012, 43f.). Die Art der Kategorie hat einen wesentlichen Einfluss auf die Art der Ergebnisse der QIA.

⁴ Siehe zur Einführung in die computergestützte Analyse qualitativer Daten Kuckartz 2010.

dienlich sind. Ist ein solcher Zusammenhang nicht erkennbar, dann ist die Unterkategorisierung ggf. nicht notwendig und steigert unnötig die Komplexität des Kategoriensystems.

Kuckartz (2012) bemängelt jedoch, dass trotz der zentralen Bedeutung der Kategorien in zahlreichen Forschungsarbeiten die Herkunft der genutzten Kategorien nicht ausreichend dokumentiert würde (vgl. Kuckartz 2012, 20f.). Drei Wege zur Gewinnung von Kategorien werden unterschieden: (i) deduktive, (ii) induktive und (iii) gemischt induktiv-deduktive Kategorienbildung (vgl. im Folgenden ebd., 63ff.).

(i) Deduktive Kategorien bestehen schon, bevor das Material gesichtet wurde. Sie werden aus bestehenden Konzepten und Theorien deduziert oder aus anderen Forschungsvorhaben übernommen und forschungsfragenspezifisch adaptiert. Im Falle der Auswertung leitfadengestützter Interviews kann die Strukturvorgabe der Fragen in Oberkategorien überführt werden. Die problematische Folge einer ausschließlich deduktiven Kategorienbildung kann eine nicht hinreichende Passung zum Material sein: Alle Bedeutungsaspekte, die nicht vorab in den Kategorien aufgenommen sind, bleiben unweigerlich unberücksichtigt. Die Gefahr ist, auf dem ‚Materialauge‘ blind zu sein und nur zu entdecken, was schon vorher bekannt war.⁵

(ii) Ein genau entgegengesetztes Risiko besteht in der induktiven Kategorienbildung. Bei diesem Vorgehen werden die Kategorien direkt am Material gebildet. Zur Illustration wird das Beispielinteresse an Nebengesprächen in der Aufgabebearbeitung wieder aufgegriffen. Zur induktiven Kategorienbildung wird das Material bis zur ersten Fundstelle durchgegangen: Die Lernenden sprechen beispielsweise über Wochenendaktivitäten. In Abhängigkeit vom angestrebten Abstraktionsgrad der herauszubildenden Kategorien (vgl. ebd., 63; Mayring 2010, 85) wird direkt eine Kategorie gebildet – z.B. ‚Kommendes Wochenende‘. Aus folgenden Fundstellen, die sich nicht einer bereits gebildeten Kategorie subsumieren lassen, werden ebenfalls neue Kategorien gebildet. Auf diese Weise ist das Kategoriensystem maximal an das Material angepasst, läuft aber Gefahr, Kategorien ‚neu zu bilden‘, über deren Bedeutung in einem Forschungsfeld bereits Einigkeit besteht. Die induktive Kategorienbildung ist mithin auf dem Auge für existierende Systematisierungen und Forschungsergebnisse blind.

(iii) Die gemischt deduktiv-induktive Kategorienbildung stellt einen Weg dar, der beide Richtungen miteinander verbindet und vermehrt anzutreffen ist (vgl. Schreier/Groeben 1999, 47). Häufig wird das Vorgehen zur induktiven Ausdifferenzierung von deduktiven Oberkategorien genutzt (vgl. Kuckartz 2012, 69).⁶ Ein

⁵ Die skizzierte Einschränkung gilt vornehmlich für explorativ ausgerichtete Untersuchungszusammenhänge. In hypothesenprüfenden Zusammenhängen kann eine Nicht-Passung Hinweis auf Falsifizierung einer Hypothese sein (vgl. Schreier 2006, 431; Schreier/Groeben 1999, 48).

⁶ Es sind aber auch andere Kombinationsmöglichkeiten eines gemischt deduktiv-induktiven Vorgehens denkbar, die Schreier (2014, Abs. 12) zusammenfasst.

Vorteil besteht darin, dass auf diese Weise sowohl bestehende Konzepte und Erkenntnisse berücksichtigt und gleichzeitig darüber hinausgehende Kategorien entwickelt werden können.

Welcher Weg der Kategorienbildung angemessen erscheint, ist abermals nur in Abhängigkeit von einer konkreten Forschungsfrage zu bedenken, weil die Antwort auch von den interessierenden Bedeutungsaspekten abhängt, die inhaltsanalytisch erfasst werden sollen.

(c) Die Berücksichtigung auch latenter Bedeutungsaspekte unterscheidet die QIA von der klassischen IA, in der vornehmlich manifeste Textmerkmale Aufnahme finden. Demgegenüber werden in der QIA die Interpretativität hervorgehoben und weitgehende Inferenzen auf latente Bedeutungsaspekte als zulässig bzw. sogar merkmalsgebend angesehen (vgl. Lamnek/Krell 2010, 463). Die QIA stellt demzufolge

eine interpretative Form der Auswertung dar, hier werden Codierungen aufgrund von Interpretation, Klassifikation und Bewertung vorgenommen: die [sic] Textauswertung und -codierung ist hier also an eine menschliche Verstehens- und Interpretationsleistung geknüpft. (Kuckartz 2012, 39)

Durch die Erfassung latenter Bedeutungsaspekte ist der Bedeutungsbereich der Gegenstandserkenntnis umfangreich, da ausgehend von manifesten Aspekten Rückschlüsse auf „individuelle und gesellschaftliche, nicht-sprachliche Phänomene“ (Mayntz et al. 1974, 151) zulässig sind. So können auch Absichten, Einstellungen, Situationsdeutungen, Wissen und stillschweigende Annahmen der Probandinnen und Probanden aus dem Datenmaterial herausgearbeitet werden (vgl. ebd.)⁷: Die Interpretativität stellt keine zu vermeidende Störung dar (vgl. Schreier 2014, Abs. 4). Allerdings sind damit auch erhöhte Anforderungen an die Systematik und Regelgeleitetheit des qualitativ-inhaltsanalytischen Vorgehens gestellt.

(d) Regelgeleitetheit und Systematik sind Aspekte, die auch in der klassischen IA von Bedeutung sind. Allerdings folgt diese einem festen Regelkanon, wohingegen sich die QIA durch Regelvariabilität auszeichnet: Die Regeln werden gegenstandsangemessen entwickelt, um eine größtmögliche Passung zum Gegenstand bzw. zum Erkenntnisinteresse zu gewährleisten. Aufgrund der Anpassung von Verfahrensregeln zugunsten der Gegenstandsangemessenheit wird der QIA von Vertretern einer quantitativen IA mitunter eine übertriebene Gegenstandsprävalenz vorgeworfen, durch die methodische Systematik und Sicherheit der Erkenntnis aufgegeben würden (vgl. Groeben/Rustemeyer 2002, 245). Der Gegenwurf von qualitativer Seite lautet: Der feste Regelkanon einer klassischen IA führe zwar zu methodisch sauberen, aber an inhaltlicher Substantialität kranken Ergebnissen (vgl. ebd.). Groeben und Rustemeyer (2002) resümieren ausgleichend, dass beide Extremformen nicht in der Lage seien, „zugleich (möglichst)

⁷ Merten resümiert: „Inhaltsanalyse ist eine Methode [...], bei der von Merkmalen eines manifesten Textes auf Merkmale eines nicht manifesten Kontextes geschlossen wird.“ (Merten 1983, 57)

substantielle und sichere Erkenntnisse zu generieren“ (ebd., 246), und sehen es als anzustrebendes Ziel an, in einer Methodik-Gegenstand-Interaktion ein Maximum an methodischer Sicherheit und inhaltlicher Substantialität zu erreichen:

Das bedeutet z.B., dass ‚Offenheit‘ im Sinne der Transparenz und (begründenden) Explikation (von theoretisch wie methodischen Festlegungen) sinnvoll, berechtigt und anzustreben ist; das gilt allerdings nicht für ‚Offenheit‘ im Sinne einer uneingeschränkten Variabilität des Vorgehens, durch die methodische Systematik praktisch ausgeschlossen wird. (ebd.)

Drei Punkte sind in diesem Zusammenhang als zwingend notwendig hervorzuheben: (1) Alle Entscheidungen, Festlegungen und Setzungen sind lückenlos transparent zu machen. Die Kritik Kuckartz' an der häufig fehlenden Offenlegung der Herkunft von Kategorien setzt an dieser Stelle an. Und auch das Kategoriensystem, in dem expliziert wird, unter welchen Bedingungen von manifesten Aspekte auf latente Bedeutungen geschlossen wird, ist ein zentraler Bestandteil, um die Regelgeleitetheit des qualitativ-inhaltsanalytischen Vorgehens zu gewährleisten. Darüber hinaus ist die Einheitenfestlegung ein weiterer Bestandteil, der offenzulegen ist, da diese einen erheblichen Einfluss auf die Reliabilität der Kategorienzuweisung hat. Kuckartz (2012) unterscheidet (i) Auswahl-, (ii) Analyse-, (iii) Kodier- und (iv) Kontexteinheit (Kuckartz 2012, 46ff).⁸ Insbesondere die Transparenz der Kodier- und die Kontexteinheiten spielt eine zentrale Rolle für die Intra- oder Interkodierreliabilität, da es einen erheblichen Unterschied macht, auf welche Weise Textabschnitte bestimmt werden, die einer Kategorie zugewiesen, und mithilfe welcher weiteren Textstellen uneindeutige Textstellen zur Kategorisierung monosemiert werden dürfen. Der Aspekt der Transparenz macht folglich eine ausführliche Dokumentation des methodischen Vorgehens notwendig. (2) Die Transparenz ermöglicht darüber hinaus eine systematische Zuordnung von Textelementen zu Kategorien und verhindert eine „impressionistische Ausdeutung“ (Mayring 2010, 12) des Materials. Groeben und Rustemeyer (2002) sprechen von einer „inferenzexplikative[n] Regelanpassung“ (Groeben/Rustemeyer 2002, 246), die es (3) ermöglicht, Kodierungen „intersubjektiv nachvollziehbar und damit auch reproduzierbar, kommunizierbar und kritisierbar“ (Früh 2011, 40 Hervorh. getilgt) zu machen.

⁸ (i) Die Auswahlinheit bezeichnet den prinzipiellen Einschluss oder Ausschluss potentieller Untersuchungsobjekte. (ii) Analyseeinheit sind die Teile einer Auswahlinheit, die in die inhaltsanalytische Auswertung einbezogen werden. (iii) Die Kodiereinheit legt fest, wie groß eine Textstelle ist, die einer Kategorie zugeordnet wird. Mit am wichtigsten aber auch am komplexesten ist (iv) die Kontexteinheit. Diese bestimmt, auf welche größeren Textteile zur Entscheidung der Kategorienzuweisung zurückgegriffen werden darf, d.h. sie gibt an, welche zusätzlichen Informationen zum Ziele der Monosemierung des beim Verstehen immer gegebenen Interpretationsspielraums (Groeben/Rustemeyer 2002, 239; Schreier/Groeben 1999, 46) dienen dürfen.

Die Flexibilität der QIA besteht also nicht in der Abwesenheit eines regelgeleiteten Vorgehens, sondern in der gegenstands- und erkenntnisinteressensangemessenen Entwicklung, transparenten Darstellung und systematischen Durchführung derselben.

An jedem die QIA charakterisierenden Orientierungspunkt sind Entscheidungen in Abhängigkeit zur konkreten Forschungsfrage zu treffen⁹, sodass eine Doppellorientierung erforderlich wird: einerseits ‚Wo will ich hin?‘ (Orientierung im Material) und andererseits ‚Welche Wegmarken bietet meine Methode mir für diesen Weg?‘ (Orientierung im Verfahren).

1.2 Ausgangsdaten und Ergebnisse

Die QIA ist ein Auswertungsverfahren, d.h. sie wird auf bereits bestehende Daten angewandt. Einschränkungen zur Art des Datenmaterials gibt es kaum, in aller Regel wird es aber ein Kommunikationsmaterial im weitesten Sinne sein: Texte (Interviews, Gruppendiskussionen und schriftliche Erzeugnisse aller Art), aber auch musikalisches, bildliches oder plastisches Material. Erforderlich ist nur, dass das Material in manifester Form besteht (vgl. Mayring 2004, 468). Auch in diesem Punkt erweist sich die QIA als sehr flexibel.

Was für Ergebnisse man durch die QIA erhält, ist nicht einfach zu beantworten. Die im Anschluss an die Kategorisierung häufig gestellte Frage: ‚Und was mache ich jetzt mit den Kodierungen?‘ ist symptomatisch dafür: Zum einen wird dem Umgang mit den Kodierungsergebnissen in der einführenden Literatur nur wenig Raum zugestanden¹⁰, zum anderen wird häufig die Forschungsfrage bei der Kategoriengewinnung und -anwendung aus dem Blick verloren. In beiden Fällen steht man am Ende ratlos vor einem Berg von Kategorienzuweisungen.

Die QIA leistet eine Datenaufbereitung: Durch die Kategorisierung erhält man strukturierte und systematisierte Datenmengen.¹¹ Aber – und dieser Aspekt ist m.E. besonders wichtig: Die Kategorisierung ist der Weg zum Ziel, sie ist ein Instrument, um sich eine ‚Schneise durch das Datendickicht‘ zu bahnen. Aus sich heraus jedoch besitzt die Kategorisierung kaum Aussagekraft. Demzufolge spricht Schreier (2006) von der Notwendigkeit einer ‚sekundären Auswertung‘¹² (Schreier 2006, 435), die Kuckartz (2012) treffend als ‚eigentliche Auswertung‘ (Kuckartz 2012, 93) betitelt. Auf der Basis des aufbereiteten Datenmaterials wird es möglich, das Vorkommen und die Ausprägung unterschiedlicher Bedeutungs-

⁹ Zur Zentralität der Forschungsfrage s. Boelmann in diesem Band oder Kuckartz (2012, 21ff.).

¹⁰ Kaum Hinweise finden sich bei z.B. Hussy et al. (2013) oder Mayring (2010; 2004).

¹¹ An diesem Punkt wird der Einfluss der Kategorien-Art auf die Ergebnisse einsichtig: Evaluative Kategorien arbeiten eine Strukturierung entlang der Einschätzungsdimensionen heraus, wohingegen thematische Kategorien inhaltlich strukturierende Aufbereitungen ermöglichen.

¹² Die Kategorisierung stellt die primäre Auswertung dar.

aspekte gezielt zu betrachten bzw. das gemeinsame Auftreten verschiedener Bedeutungsaspekte zu bestimmen. Insbesondere fallvergleichende Analysen, die ohne einen datenreduzierenden Überblick nahezu unmöglich sind, werden nach Maßgaben der Kategorien vorbereitet. Welche Form der sekundären Auswertung zielführend ist, kann abermals nur anhand einer konkreten Fragestellung entschieden werden. Was aber mit Entschiedenheit zu betonen ist: Die QIA ermöglicht mehr als ein einfaches Auszählen von Kodierhäufigkeiten. Sie kennt auch in diesem Punkt kaum Grenzen, wenn die entwickelten sekundären Auswertungsschritte transparent gemacht werden und ein systematisches und regelgeleitetes Vorgehen ermöglicht wird.¹³

1.3 Vorteile und Nachteile der qualitativen Inhaltsanalyse

Auch die QIA ist kein Königsweg, den man ohne Nachteile bzw. Einschränkungen bis ans Ziel beschreiten kann. Die QIA besitzt unbestritten einige Vorzüge, gleichwohl auch Nachteile, die aber nicht zwingend als solche auftreten müssen und dementsprechend als negative Potenziale bezeichnet werden sollen. Die Systematik und Regelgeleitetheit bei gleichzeitiger Flexibilität in der primären und sekundären Auswertung ist offensichtlich ein Vorteil der QIA, der aber auch als Inflexibilität im Anschluss an die Festlegung eines Ablaufschemas empfunden werden kann. Ähnlich ambivalent ist der Vorteil einer ökonomischen Anwendbarkeit auf größere Datenmengen, der auch als Nachteil aufgefasst werden kann, da durch die Zergliederung des Datenmaterials und der Datenkomplexität in der Kategorisierung beispielsweise individuellen Bedeutungsaspekte verloren zu gehen drohen. Als ein Vorteil – insbesondere gegenüber einer quantitativen IA – muss die Berücksichtigung latenter Bedeutung hervorgehoben werden. Vertreter einer (strengen) interpretativen Sozialforschung merken jedoch an, dass die Kategorisierung nicht dazu geeignet sei, tatsächlich Tiefenstrukturen zu rekonstruieren, sondern dass die QIA einen Hang zur Oberflächlichkeit aufweise (Rosenthal 2014, 213ff.).

Wie über diese Vor- und Nachteile zu befinden ist, kann nur in Abhängigkeit zu einer konkreten Forschungsfrage entschieden werden. Was in dem einen Fall zum Nachteil gereichen kann, ist im anderen Fall ein methodologischer Vorteil der QIA.

¹³ Bei Kuckartz (2012, 93ff./108ff./130f.) und Schreier (2006, 435ff.) finden sich sehr hilfreiche Ideen und Beispiele für sekundäre Auswertungsverfahren. Ein für die deutschdidaktische Forschung interessantes Beispiel wird im zweiten Teil des Beitrags vorgestellt.

1.4 Gütekriterien für den Einsatz der qualitativen Inhaltsanalyse

Die Angemessenheit der klassischen Gütekriterien Validität, Objektivität und Reliabilität für einen qualitativen Forschungsprozess ist umstritten (vgl. Mayring 2010, 116ff.; Groeben/Rustemeyer 2002, 246ff.; Hussy et al. 2013, 258). Die Diskussion kann an dieser Stelle nicht referiert werden und auch die unterschiedlichen verfahrensspezifischen Gütekriterien können nicht besprochen werden.¹⁴ Bezieht man die klassischen Gütekriterien aber nicht auf den gesamten Forschungsprozess – denn dieser ist tatsächlich mit den an der „naturwissenschaftlichen Logik des Messens“ (Kuckartz 2012, 24) orientierten Kriterien der psychologischen Testtheorie nur unzureichend zu beurteilen und müsste „modifiziert und erweitert werden“, um den „prozeduralen Charakter qualitativer Forschung stärker zu berücksichtigen“ (ebd., 167) –, sondern legt diese nur an das Kategoriensystem an, dann erscheinen sie die Anforderungen angemessen zu bestimmen. Ein Kategoriensystem ist valide, wenn es die relevanten Bedeutungsaspekte auch tatsächlich erfasst.

Eine hohe Besetzung der Restkategorie oder (vor allem bei einem induktiven Kategoriensystem) eine überproportional hohe Besetzung einer Unterkategorie im Vergleich zu anderen Unterkategorien kann einen Hinweis auf unzureichende Validität darstellen: Die hohe Besetzungshäufigkeit lässt vermuten, dass dieser Bedeutungsaspekt zu undifferenziert abgedeckt wird. (Hussy et al., 258)

Das Ziel der Objektivität und Reliabilität besteht im inhaltsanalytischen Vorgehen in der „Überschreitung subjektiven und potentiell verzerrten Verstehens“ (ebd.) durch Regelgeleitetheit, Systematik sowie Transparenz und somit durch „Nachvollziehbarkeit als Plausibilitätsvermittlung“ (Groeben/Rustemeyer 2002, 246f.). Gewährleistet wird damit auch die Nachprüfbarkeit im Sinne methodischer Kontrollierbarkeit.

Da die Objektivität bei der Inhaltsanalyse nur intersubjektiv über die Reliabilität abgeschätzt werden kann, fallen bei diesem Verfahren Objektivität und Reliabilität zusammen, so dass eine hinreichende Interkodier-Übereinstimmung zugleich auch eine hinreichende Objektivität des Kategoriensystems indiziert. (Schreier/Groeben 1999, 50)

Interkodier-Reliabilität ist ein Maß der Übereinstimmung von Kodierungen durch zwei unabhängige Kodierer, über das die „Brauchbarkeit des entwickelten Kategoriensystems“ (Groeben/Rustemeyer 2002, 241) zu beurteilen ist.¹⁵

¹⁴ Siehe dazu Mayring (2010, 116ff.), Kuckartz (2012, 165ff.) oder Rustemeyer (1992, 104ff.). Insbesondere die Gütekriterien ‚Exhaustion‘, ‚Saturiertheit‘ und ‚Disjunktheit‘ erscheinen im Falle einer Passung durch Methodik-Gegenstands-Interaktion als nicht mehr tragfähig.

¹⁵ Möglich ist aber auch, die Übereinstimmung der Kategorienzueweisung eines Kodierers zu zwei Zeitpunkten zu bestimmen. Die Intrakoder-Reliabilität bestimmt die „Stabilität der Codierung“ (Hussy et al. 2013, 258; siehe auch Groeben/Rustemeyer 2002, 247).

1.5 Literaturempfehlungen

Kuckartz (2012): Qualitative Inhaltsanalyse.

Mayring (2010): Qualitative Inhaltsanalyse.

Beide Bücher liefern eine überblicksartige Einführung in die QIA. Auf der Basis theoretischer Fundierungen werden Techniken bzw. Basismodelle der QIA unterschieden. Für Mayring gilt in stärkerem Maße, dass die Ablaufschemata nicht als festgefügte Schrittfolgen misszuverstehen sind.

Mayring/Gläser-Zikuda (2008): Die Praxis der qualitativen Inhaltsanalyse.

Anhand von Anwendungsbeiträgen wird der große Einsatzbereich anschaulich gemacht. Die Unterschiedlichkeit der Forschungsvorhaben bietet zahlreiche Anregungen zur Adaption der QIA für eigene Fragestellungen.

Schreier (2014): Varianten qualitativer Inhaltsanalyse.

Schreier unterscheidet in instruktiver und konziser Weise Spielarten und Ausprägungen der QIA. Der Beitrag schließt mit einem Werkzeugkasten-Modell, das dem Schematismus einer zu eng verstandenen QIA entgegenwirken soll.

2. Lenkungsgrade im Literaturunterricht

Der QIA wird häufig nachgesagt, dass mit ihr keine komplexen Fragestellungen zu bearbeiten sind: Sie sei zu unflexibel, oberflächlich und ungeeignet zur Rekonstruktion von Prozessen. Die folgende Darstellung zum methodischen Vorgehen der Datenauswertung in einem aktuellen Projekt der Lernaufgabenwirkungsforschung im Literaturunterricht „Zum Einfluss stark und gering lenkender Aufgabensets auf das Textverstehen von Lernenden mit unterschiedlichen Lernvoraussetzungen“ soll ein Gegenbeispiel zu diesem Vorurteil liefern und gleichzeitig die Ausführungen des ersten Teils illustrieren.

2.1 Fragestellung und Forschungsdesign

Die übergreifende Fragestellung im benannten Forschungsprojekt, in dem es um die aufgabenbasierte Förderung des literarischen Textverstehens geht, lautet: ‚Inwieweit ist ein erhöhtes Ausmaß an instruktionaler Unterstützung in Aufgaben zu komplexen Problemstellungen lernförderlich?‘ Zur Untersuchung der instruktionalen Unterstützung wurden zwei Aufgabensets entwickelt: Das stark lenkende Aufgabenset modellierte einen strukturierten Verstehensprozess, indem die zentralen Verstehensherausforderungen der Geschichte gezielt durch neun Teilaufgaben fokussiert und deren Bewältigung angeregt werden. Unmittelbar ohne Vorstrukturierungen des Verstehensprozesses setzt die Variante einer geringen Lenkung bei den komplexen Verstehensherausforderung an. Dieses Aufgabenset umfasst bloß die Aufgaben a. und b. (siehe Abb. 1).¹⁶

Aufgabensets								
1	2	3	4	5	6	7	8	9
							a.	b.

Abb. 1: Aufgabensets der Untersuchung

(starke Lenkung = Teilaufgaben 1-9 vs. geringe Lenkung = Teilaufgaben a. und b.)

Erhoben wurden sowohl schriftliche Produktdaten als auch Prozessdaten einer kooperativen Bedeutungsaushandlung in Kleingruppen. Der übergreifenden Fragestellung wird in zwei Teilfragen nachgegangen:

¹⁶ Der Untersuchung liegt die Geschichte „Wie man eine Hilfe findet“ von Jürg Schubi-ger (1995) zugrunde. Die Verstehensherausforderungen können wie folgt zusammengefasst werden: Ein Mädchen ist einsam und hilflos auf der Welt und sucht dringend nach Hilfe. Sie fragt nacheinander einen wilden Wolf, einen starken Stier und eine große Frau um Hilfe gegen Orientierungslosigkeit, Feuer und Überflutung. Aber keine der gefragten Figuren kann Hilfe geben. In einem Gewitter kumulieren die Gefahren. Die Figuren rücken zusammen und sind gemeinsam hilflos den Gefahren ausgeliefert. Nachdem das Gewitter vorüber ist, beschließen die Figuren, sich im Falle der Hilfsbedürftigkeit erneut zu treffen. Dass die Gemeinschaft selber die gefundene Hilfe ist, wird in der Geschichte nicht manifest ausgesagt und ist Resultat parabolischen Verstehens.

- (1) Welche Unterschiede bestehen in den *inhaltsbezogenen* Verarbeitungsweisen zwischen Aufgaben geringer und starker Lenkung in der Bearbeitung durch Lernende mit guten und weniger guten Lernvoraussetzungen?
- (2) Welche Unterschiede bestehen in den *prozessbezogenen* Verarbeitungsweisen zwischen Aufgaben geringer und starker Lenkung in der Bearbeitung durch Lernende mit guten und weniger guten Lernvoraussetzungen?

Weil die Entwicklung eines gegenstands- und erkenntnisinteressenangemessenen Vorgehens nur bei der Forschungsfrage ansetzen kann, wird diese von Kuckartz (2012) als „Dreh- und Angelpunkt jedes Forschungsprojektes“ (Kuckartz 2012, 21) bezeichnet. Von ihr sind vielfältige methodische Entscheidungen abhängig, schließlich besteht die Funktion einer Auswertungsmethode darin, Erkenntnisse über den beforschten Gegenstand zu erlangen. Im Folgekapitel werden ferner weitere Aspekte des Forschungsdesigns eingeflochten.

2.2 Kategorien und primäre Datenauswertung

Ausgehend von diesen Teilforschungsfragen wurde ein Kategoriensystem als datenreduzierendes Raster entwickelt, in welchem die Variablen der Forschungsfrage operationalisiert sind. Drei Kodierdimensionen sind aus den Teiluntersuchungsfragen abgeleitet worden:

Aufgabensets								
1	2	3	4	5	6	7	8	9
							a.	b.
Kodierdimension 1: Konstruktionsgegenstände								
Kodierdimension 2: Verstehensleitende Konzepte								
Kodierdimension 3: Verarbeitungsprozesse								

Abb. 2: Strukturelle Anbindung der Kodierdimensionen an die Aufgabensets

Zur strukturierenden Aufbereitung des Datenmaterials in Bezug auf Teilfrage (1) wurden zwei Kodierdimensionen gebildet: Kodierdimension 1 umfasst die ‚Konstruktionsgegenstände‘, d.h. die Kategorien bilden ab, welche inhaltlichen Aspekte der Geschichte von den Lernenden im Verstehensprozess aufgegriffen oder konstruiert werden, um die Aufgaben zu bearbeiten (vgl. Stark 2010, 120). Die Kodierdimension 2 ‚Verstehensleitende gedankliche Konzepte‘ dient dazu, die wissensseitige Basis der Bedeutungszuschreibung aus dem Material herauszufiltern. Mithilfe der Kodierdimension 3 ‚Verarbeitungsprozesse‘ wird das Datenmaterial bezogen auf Teilfrage (2) aufbereitet. Die Kategorien dieser Kodierdimen-

sion umfassen die kognitiven Operationen, deren Resultate die inhaltlichen Verstehensergebnisse darstellen.¹⁷ Ein Beispiel soll den Zusammenhang der Kodierdimensionen illustrieren. Die erste Aufgabe im stark lenkenden Aufgabenset lautet:

*Aufgabe 1: „Das Mädchen würde sogar Hilfe stehlen, wenn es nur wüsste wo.“
Was erfährt man daraus über die Gefühle des Mädchens?*

In der Bearbeitung dieser Aufgabe, die die Rekonstruktion der Ausgangssituation zum Ziel hat, kommt eine Gruppe zu dem folgenden Verstehensergebnis:

„Das Mädchen fühlt sich gut, wenn sie Geld bekommt“.

Wie in Abbildung 2 durch die vertikalen Linien angezeigt, werden die Verstehensresultate zu jeder einzelnen Aufgabe parallel in den drei Kodierdimensionen kodiert. Die Kategorienzuweisung zu obenstehendem Verstehensresultat lautet:

- Kodierdimension 1 ‚Konstruktionsgegenstände‘: Oberkategorie = ‚Ausgangssituation‘ und Unterkategorie = ‚Mädchen braucht Geld‘ – eine monetäre Hilfsbedürftigkeit ist der Gegenstand der Verstehenskonstruktion;
- Kodierdimension 2 ‚Verstehensleitende Konzepte‘: Kategorie = ‚konkrete Hilfe‘ – der Konstruktion der monetären Mangelsituation liegt die Vorstellung einer konkreten, d.h. dinglichen Hilfe zugrunde¹⁸;
- Kodierdimension 3 ‚Verarbeitungsprozesse‘: Kategorie = ‚lokale Ausweitung‘ – für die monetäre Mangelsituation gibt es in der Geschichte keine manifeste Referenz, sondern die Textbasis wird um Welt- und Erfahrungswissen angereichert.¹⁹

Auf diese Weise wurde das gesamte Datenmaterial parallel in den drei Dimensionen kodiert und entsprechend strukturiert.

Wie aber wurden die Kategorien in der jeweiligen Kodierdimension gewonnen? Die Konstruktionsgegenstände der Kodierdimension 1 standen durch die Teilaufgaben fest: In einer Aufgabe, die beispielsweise die Rekonstruktion der Ausgangssituation fordert, ist die Ausgangssituation auch zwingend der Konstruktionsgegenstand. Folglich konnten die Oberkategorien deduktiv aus den Vorgaben des Aufgabensets abgeleitet werden. Wie aber die Ausgangssituation jeweils von den Lernenden rekonstruiert wird, kann nur anhand der konkreten Aufgabenresultate bestimmt werden, mithin die Unterkategorien der jeweiligen Konstruktionsgegenstände induktiv ausdifferenziert wurden. Die Kategorien der Kodierdimension 2 wurden induktiv gebildet, insofern sich ein verstehensleitendes Kon-

¹⁷ Die Kategorien der jeweiligen Kodierdimensionen sind hierarchisch in Ober-, Unter- und teilweise sogar Subkategorien der Unterkategorien differenziert. An dieser Stelle muss eine einfache schematische Darstellung ausreichen (s. genauer Heins i.V.).

¹⁸ Es wird eben gerade nicht von psychischer Hilfe gegen die Hilflosigkeit ausgegangen, wie sie die Geschichte ins Bild setzt (s. dazu Fußnote 16).

¹⁹ Die Gruppe konstruiert einen inhaltlichen Aspekt und macht diesen zum Gegenstand des Verstehens (monetäre Hilfsbedürftigkeit = Kodierdimension 1).

zept nur anhand der tatsächlichen Verstehensergebnisse bestimmen lässt. Ausschließlich deduktiv wurden die Kategorien der Kodierdimension 3 aus den Erkenntnissen der kognitionspsychologischen Textverstehensforschung abgeleitet (vgl. Christmann/Groeben 1999; Grzesik 2005). In vorgestellter Weise konnte die Flexibilität der QIA dazu beitragen, ein dem Gegenstand und Erkenntnisinteresse entsprechend valides Kategoriensystem zu entwickeln.

2.3 Sekundäre Auswertung und Ergebnisse

Während die primäre Auswertung einer QIA in der Kategorisierung der Ausgangsdaten besteht, umfasst die sekundäre Auswertung die analytische Weiterarbeit mit den kategorisierten Daten. Die sekundäre Auswertung nutzt die systematisierte und strukturierte Datenbasis zur Generierung von Erkenntnissen zur Forschungsfrage.

Aus der sekundären Auswertung der zugrundeliegenden Studie soll der zweite Auswertungsschritt vorgestellt werden. Dieser bestand darin, für jeden untersuchten Fall die Kategorienzuweisung im Verlauf der Aufgabenbearbeitung zu analysieren und Muster der Kodierung zu identifizieren (im stark lenkenden Aufgabenset Muster der Verstehensentwicklung zwischen den Aufgaben 1-9 und im gering lenkenden Aufgabenset zwischen den Aufgaben a. und b.; s. Abb. 3). Aus methodischer Sicht ist dieser Schritt insofern bedenkenswert, als er beispielgebend dafür sein kann, wie auch mit der QIA Prozessstrukturen rekonstruiert und erforscht werden können – ein oft für sequenzanalytische Verfahren erhobener Anspruch. Wie dies möglich ist, soll anhand ausgewählter Datenbeispiele illustriert werden.

Am Beispiel von Verstehensresultaten zu den Aufgaben 1, 3, 6 und 8 einer Gruppe mit weniger guten Lernvoraussetzungen soll im Folgenden der Verlauf der aufgabenbasierten Verstehensentwicklung anhand der Kategorienzuweisungen nachgezeichnet werden (die Beispielaufgaben sind in Abb. 3 hervorgehoben). Um den Weg der Erkenntnisgewinnung so anschaulich wie möglich darstellen zu können, wird die Komplexität der Kategorienzuweisung auf die Kodierdimensionen 2 und 3 reduziert. An diesen Kodierdimensionen lässt sich der Wert inhaltsanalytisch strukturierter und systematisierter Daten plastisch aufzeigen.

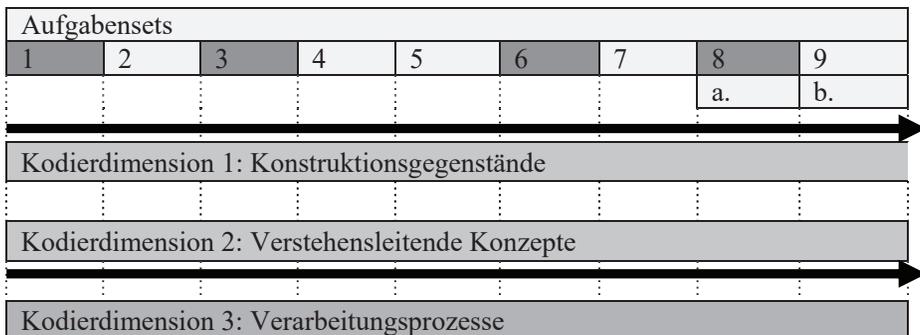


Abb. 3: Sekundäre Auswertung:
Analyse der Kategorienzuweisungen im Verlauf der Verstehensentwicklung

Beispiel 1: Intendiertes Ziel der Teilaufgabe 1 besteht in der Rekonstruktion der Ausgangssituation des Mädchens zu Beginn der Erzählung (für die konkrete Aufgabenstellung s.o.). Die Gruppe kommt zum folgenden Verstehensergebnis:

das Mädchen fühlt sich schlecht, wenn es es arm ist und wenn es Geld stiehlt, hat sie ja Geld und muss nicht auf der Straße wohnen, sondern kann sich ein Haus oder Wohnung kaufen (-) so hat sie wieder Luxus, wenn sie es mal hatte. Sie fühlt sich fröhlich, also wenn sie n haus hat, dann kann sie ja irgendwie warm und muss nicht im Winter irgendwo rumsitzen.

(NI_ARIN: 90-102)

Diesem Verstehensresultat sind die folgenden Kategorien zugewiesen: 2. Kodierdimension „Verstehensleitendes Konzept“ = ‚konkrete Hilfe‘: Es wird eine Mangelsituation rekonstruiert, die mittels monetärer Hilfe zu beheben ist. 3. Kodierdimension „Verarbeitungsprozesse“ = ‚lokale Ausweitung‘: Die Geschichte bietet keine Anhaltspunkte für eine so gestaltete Ausgangssituation, sodass das Verstehen durch eine top-down Wissensintegration erfolgt und die Textbasis ausweitet.

Beispiel 2: Im Zentrum der Teilaufgabe 3 steht die Rekonstruktion der Hilflosigkeit aller Figuren.

Aufgabe 3: Was meinst du: Können der Wolf, der Stier und die Frau überhaupt Hilfe gegen die Gefahren haben? Begründe deine Antwort!

Die Gruppe hingegen verhandelt Folgendes:

ME ja aber man kann man kann jemand mit nem Stock nehm
 KI ja schwimm schwimm schwimm schwimm
 ME ja man kann schwimm schwimm machen(-) das ist auch eine Hilfe oder man oder plötzlich kommt einer vorbei
 KI oder Stock Stock Stock ((verbunden mit Grabebewegungen))
 ME ja genau
 ME [und dann festhalten]
 KI [Loch Loch Loch]
 ME wieso Loch?
 KI ein Loch und dann fließt das Wasser darein
 ME nein also man man nimmt da einfach einen Stock nä? hält das dann darüber und schon kommt der wieder raus.

(NI_ARIN: 481-500)

Auch diesem Aushandlungsprozess sind die gleichen Kategorien der Kodierdimensionen 2 und 3 zugewiesen: ‚konkrete Hilfe‘ und ‚lokale Ausweitung‘. Die Entwicklung konkreter Hilfemaßnahmen gegen die Gefahren (im vorliegenden Beispiel gegen die Überflutung) überlagert die Wahrnehmung der Hilflosigkeit, die im Text manifest ist. Ferner wird erneut die Textbasis um Erfahrungswissen ergänzt.

Beispiel 3: Die gleiche Kategorienstruktur weist das Verstehensresultat zur Teilaufgabe 6 auf, welche die Rekonstruktion der konkreten Hilflosigkeit im Gewitter intendiert.

Aufgabe 6: In Zeile 45 fragt das Mädchen: „... – was dann?“ und alle denken nach, was dann zu tun sei. Aber können sie denn überhaupt eine Antwort auf die Frage „was dann?“ finden?

Auf die Frage, ob die Figuren im Gewitter eine Hilfe haben können – was auf der Ebene der manifesten Textinformationen zu verneinen wäre –, handelt die Gruppe folgendes Verstehen aus:

KI ja man KAnn helfen
 AR ja okay ja (-) weil:(...)
 ME ja [ankreuzen ja]
 KI [ja ankreuzen]
 AR ja (-) weil:
 KI und weil (-) man kann eine Erdgrube bei dem Ufer machen.
 (NI_ARIN: 595-607)

Auch in diesem Verstehensresultat setzen sich das gedankliche Konzept ‚konkreter Hilfe‘ und die ‚lokale Ausweitung‘ einer Hilfsmaßnahme fort. Wieder aufgenommen wird die konkrete Hilfemaßnahme ‚Erdgrube gegen Überflutung‘, die bereits in Teilaufgabe 3 wissensbasiert ergänzt wurde.

Beispiel 4: Die Hilfe in der Gemeinschaft, die sich im Zusammenrücken während der Gewittersituation zeigt, kann schlussendlich von der Gruppe vor ihrem bisher entwickelten Verstehen nicht erkannt werden.

Aufgabe 8: Am Ende der Geschichte heißt es, dass die Figuren sich wieder treffen wollen, wenn wieder jemand in einer Situation ist, in der er keine Hilfe hat. Geh die Geschichte noch einmal von Anfang an durch. Gibt es eine Stelle, an der die Figuren das Gefühl haben, Hilfe zu finden?

Auf die Frage, ob es in der Geschichte eine Situation gibt, in der die Figuren das Gefühl haben, Hilfe zu finden, hält die Gruppe fest:

Man kann hilfe finden: 1. Wegen denn Erdgruben. | 2. Mann kann die Feuerwehr rufen. | 3. Jemand fragen, wo mann ist.
 (NI_ARIN_schriftl: 43-48)

Auch das Verstehensresultat zur Teilaufgabe 8 ist mit den Unterkategorien ‚konkrete Hilfe‘ und ‚lokale Ausweitung‘ kodiert.

Die Analyse der Kategorienzuweisungen im Verlauf der Verstehensentwicklung macht erkennbar: In dieser Gruppe bleibt ein einmal aktiviertes gedankliches Konzept im Verstehens- und Aufgabenbearbeitungsprozess verstehensleitend: In den Verstehensresultaten zu allen Aufgaben ist es das gedankliche Konzept ‚konkreter Hilfe‘, das die inhaltliche Verstehenweise beeinflusst. Die Text- und Aufgabenwahrnehmung wird dadurch erheblich eingeschränkt, weil rekursive Verstehensprozesse ausbleiben: Der eingeschlagene Verstehensweg wird beharrlich bis zum Ende durchgegangen. Ferner wird deutlich, dass auch die vollzogenen textverstehenden Operationen unverändert die gleichen bleiben – sie sind mit dem Deutungskonzept verknüpft. Die mit der instruktionalen Unterstützung intendierte Aktivierung komplexer Verstehensoperationen und die Fokussierung auf

Verstehensherausforderungen zur Bildung umfänglicher semantischer Relationen gelingt in dieser Gruppe augenscheinlich nicht.²⁰

Wenngleich an dieser Stelle die These nur anhand weniger Beispiele gestützt werden kann, wird deutlich ersichtlich, dass Entwicklungsverläufe innerhalb der Kodierdimensionen rekonstruiert und interpretiert werden können. Die Einschränkung auf Verstehensentwicklungen innerhalb einer Kodierdimension – die selbstredend in Verbindung mit der Verstehensentwicklung innerhalb der jeweils anderen beiden Kodierdimensionen zu analysieren ist – stellt keine unverhältnismäßige Verengung dar, wenn die Kodierdimensionen so entwickelt wurden, dass in ihnen die bedeutungsrelevanten Aspekte abgebildet werden. Dann nämlich sind mit der QIA auch komplexe Prozessrekonstruktionen auf der Ebene nicht unmittelbar einsehbarer Verstehensoperationen und gedanklicher Konzepte möglich.

2.4 Fazit und Aufruf

Die QIA ist keine festgefügte Methode mit vorgezeichneten Erkenntniswegen in unflexiblen Ablaufschemata. Vielmehr war es das Ziel der Darstellung, aufzuzeigen, dass vier methodologische Orientierungspunkte bei der Suche nach eigenen Wegen zur Erkenntnisgenerierung hilfreiche Wegmarken sein können: Die QIA ist (a) ein datenreduzierendes, (b) kategorienorientiertes, (c) auch latente Bedeutungen berücksichtigendes, (d) systematisches und regelgeleitetes Vorgehen zur Analyse von bedeutungshaltigem Material. Behält man diese Punkte und seine Forschungsfrage beharrlich im Blick und versucht, eine größtmögliche Passung zu erzielen, dann können mit der QIA neue Wege gegangen werden.

Wie dies möglich ist, wurde anhand eines aktuellen Projekts zur Lernaufgabenwirkungsforschung im Literaturunterricht skizziert. Dabei konnte die Flexibilität der Kategorienbildung und -zuweisung bei gleichzeitiger Systematik und Regelgeleitetheit aufgezeigt werden. Mit Nachdruck wurde hervorgehoben, dass die QIA nicht mit der Kategorisierung endet. Die Kategorisierung stellt eine strukturierende Datenaufbereitung dar, der zwingend eine sekundäre Auswertung folgen muss. Und dass in der sekundären Auswertung für die deutschdidaktische Forschung noch ungenutztes Potenzial liegt, konnte anhand der Möglichkeit, auch mit einem kategorisierenden Verfahren Prozessstrukturen zu rekonstruieren, illustriert werden.

²⁰ Für weitere und ausführliche Falldarstellungen s. Heins (2017). Wie dort gezeigt wird, kann in anderen Gruppen mit weniger guter Lernvoraussetzungen eine positive Verstehensentwicklung unterstützt werden, wenn ein stärker lenkendes Aufgabenset bearbeitet wird, wohingegen die geringe Lenkung bei Lernenden mit weniger guten Lernvoraussetzungen nicht zu intendierten Verstehensergebnissen führt. Für Lernende mit guten Lernvoraussetzungen ist die positive Verstehensentwicklung nicht bzw. nur geringfügig von der Lenkung der Aufgaben abhängig.

Literatur

- Berelson, Bernard: Content analysis in communication research. Glencoe: Free Press 1952.
- Christmann, Ursula/Groeben, Norbert: Psychologie des Lesens. In: Franzmann, Bodo/Jäger, Georg (Hrsg.): Handbuch Lesen. München: Saur 1999, 145-223.
- Früh, Werner: Inhaltsanalyse. Theorie und Praxis. Konstanz: UVK-Verlagsgesellschaft 2011.
- Groeben, Norbert/Rustemeyer, Ruth: Inhaltsanalyse. In: König, Eckard/Zedler, Peter (Hrsg.): Qualitative Forschung. Grundlagen und Methoden. Weinheim: Beltz 2002, 233-258.
- Grzesik, Jürgen: Texte verstehen lernen. Neurobiologie und Psychologie der Entwicklung von Lesekompetenzen durch den Erwerb von textverstehenden Operationen. Münster: Waxmann 2005.
- Heins, Jochen: Lenkungsgrade im Literaturunterricht. Zum Einfluss stark und gering lenkender Aufgabensets auf das Textverstehen. Wiesbaden: Springer VS 2017.
- Hussy, Walter/Schreier, Margrit/Echterhoff, Gerald: Forschungsmethoden in Psychologie und Sozialwissenschaften für Bachelor. Berlin/Heidelberg: Springer 2013.
- Kuckartz, Udo: Einführung in die computergestützte Analyse qualitativer Daten. Wiesbaden: VS Verlag für Sozialwissenschaften 2010.
- Kuckartz, Udo: Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung. Weinheim: Beltz Juventa 2012.
- Lamnek, Siegfried/Krell, Claudia: Qualitative Sozialforschung. Lehrbuch. Weinheim: Beltz 2010.
- Mayntz, Renate/Holm, Kurt/Hübner, Peter: Einführung in die Methoden der empirischen Soziologie. Opladen: Westdeutscher Verlag 1974.
- Mayring, Philipp/Gläser-Zikuda, Michaela: Die Praxis der qualitativen Inhaltsanalyse. Weinheim: Beltz 2008.
- Mayring, Philipp: Qualitative Inhaltsanalyse. Grundlagen und Techniken. Weinheim: Beltz 2010.
- Mayring, Philipp: Qualitative Inhaltsanalyse. In: Flick, Uwe/Kardorff, Ernst von/Steinke, Ines (Hrsg.): Qualitative Forschung. Ein Handbuch. Reinbek bei Hamburg: Rowohlt-Taschenbuch-Verlag 2004, 468-475.
- Merten, Klaus: Inhaltsanalyse. Einführung in Theorie, Methode und Praxis. Opladen: Westdeutscher Verlag 1983.
- Rosenthal, Gabriele: Interpretative Sozialforschung. Eine Einführung. Weinheim: Beltz Juventa 2014.
- Rustemeyer, Ruth: Praktisch-methodische Schritte der Inhaltsanalyse. Eine Einführung am Beispiel der Analyse von Interviewtexten. Münster: Aschendorff 1992.
- Schreier, Margrit/Groeben, Norbert: Leitfadeninterview und Inhaltsanalyse: 2. Inhaltsanalyse. In: Siegener Periodicum zur internationalen Empirischen Literaturwissenschaft (SPIEL) (1999), 18, 1, 43-54.

- Schreier, Margrit: Qualitative Auswertungsverfahren. In: Groeben, Norbert/Hurrelmann, Bettina (Hrsg.): Empirische Unterrichtsforschung in der Literatur- und Lesedidaktik. Ein Weiterbildungsprogramm. Weinheim: Juventa 2006, 421-441.
- Schreier, Margrit: Varianten qualitativer Inhaltsanalyse: Ein Wegweiser im Dickicht der Begrifflichkeiten. [59 Absätze]. In: Forum Qualitative Sozialforschung/Forum: Qualitative Social Research (2014), 15, 1, Art. 18, <http://nbn-resolving.de/urn:nbn:de:0114-fqs1401185> (letzter Zugriff 01.08.2018).
- Schubiger, Jürg: Als die Welt noch jung war. Weinheim: Beltz/Gelberg 1995.
- Stark, Tobias: Zur Interaktion von Wissensaktivierung, Textverstehens- und Bewertungsprozessen beim literarischen Lesen –Erste Ergebnisse einer qualitativen empirischen Untersuchung. In: Winkler, Iris/Abraham, Ulf/Masanek, Nicole (Hrsg.): Poetisches Verstehen. Literaturdidaktische Positionen – empirische Forschung – Projekte aus dem Deutschunterricht. Baltmannsweiler: Schneider Verlag Hohengehren 2010, 114-132.

Quantitative Inhaltsanalyse

Quantifizieren um jeden Preis?!

1. Vorstellung der quantitativen Inhaltsanalyse

Die quantitative Inhaltsanalyse hat ihren Ursprung in der amerikanischen Sozialwissenschaft der dreißiger bzw. vierziger Jahren des 20. Jahrhunderts. Inhaltsanalysen wurden zu dieser Zeit verwendet, um Zeitungen nach ideologischer Prägung und propagandistischen Inhalten zu untersuchen (so z.B. Lasswell 1927). Ein wichtiges und prägendes Lehrbuch der quantitativen Inhaltsanalyse wurde bereits 1952 von dem Sozialwissenschaftler Bernard Berelson verfasst. Im Zuge der siebziger Jahre geriet die quantitative Inhaltsanalyse vermehrt in die Kritik (generell zur Geschichte der Inhaltsanalyse vgl. Mayring 1994, 160ff.). Kritisiert wurde die methodische Einseitigkeit; hinterfragt wurde in diesem Zusammenhang, ob sich komplexe menschliche Sachverhalte wie z.B. (politische) Einstellung über ein reines Auszählen von Datenmaterial ermitteln lassen (für eine detaillierte Darstellung der einzelnen Kritikpunkte vgl. Lamnek 2010, 7ff.). Seit dieser Kritik werden in modernen Forschungsdesigns rein quantitative Verfahren in der Regel mit qualitativen Verfahren kombiniert (wie in dem im zweiten Teil vorgestellten Dissertationsprojekt).

Wenn ein Forschungsvorhaben zunächst das Ziel hat, sprachliche Daten zur weiteren Überprüfung fachdidaktischer Fragestellungen zu quantifizieren, kann es sinnvoll sein, eine quantitative Inhaltsanalyse einzusetzen. Der Einsatz der Methode ist allerdings an bestimmte Bedingungen geknüpft, die im Folgenden erläutert werden sollen.

1.1 Funktionsweise der quantitativen Inhaltsanalyse

Untersuchungsgegenstand einer quantitativen Inhaltsanalyse sind in der Regel Texte. Die über die Methode gewonnenen Daten werden also nicht wie bei einem Interview oder einem Fragebogen erfragt, „sondern aus Texten gewonnen, die die Untersuchungseinheiten selbst hervorgebracht haben oder diese repräsentieren“ (Lamnek 2010, 454). Ziel einer quantitativen Inhaltsanalyse ist es, „Wortmaterial hinsichtlich bestimmter Aspekte (stilistische, grammatische, inhaltliche, pragmatische Merkmale) zu quantifizieren“ (Bortz/Döring 2006, 149). Die Methode ist

also eher eine Datenerhebungsmethode als eine Methode zum Auswerten von Daten. So schreiben Bortz und Döring (ebd.): „Fasst man den Text als ‚Untersuchungsobjekt‘ auf, so erscheint die Inhaltsanalyse tatsächlich als Datenerhebungsmethode, weil sie angibt, wie Eigenschaften des Textes zu messen sind“. In einem Forschungsdesign bildet die quantitative Inhaltsanalyse folglich immer nur einen ersten Schritt. Die quantitative Inhaltsanalyse wird in diesem ersten Schritt zur Operationalisierung von Daten eingesetzt, d.h. sie dient dazu, Variablen oder Merkmale konkret messbar, d.h. auszählbar, zu machen.

Der Vorteil der Methode ist, dass sie beim richtigen Einsatz auf objektivem, validem und reliablem Weg Daten liefert, die dann in einem Folgeschritt zur Hypothesenprüfung genutzt werden können: „Die Inhaltsanalyse wird im quantitativen Forschungsprozess als Technik zur Erhebung von Daten verwendet, die dann die Grundlage für Versuche der Hypothesenfalsifikation durch den Forscher sind“ (Lamnek 2010, 454). Zu dieser Hypothesenfalsifikation können bestimmte inferenzstatistische Verfahren wie Chi-Quadrat-Techniken, Regressionen oder Korrelationen genutzt werden. Hiermit ist eine wichtige Besonderheit der quantitativen Inhaltsanalyse angesprochen: Sie ist eine Methode, die im Kontext hypothesenprüfender Designs eingesetzt werden kann; in hypothesengenerierenden Kontexten ist die Methode nicht einsetzbar. Dies liegt darin begründet, dass die in der Inhaltsanalyse erhobenen Merkmale immer auf Grundlage einer bereits bestehenden Forschungshypothese gebildet werden. Die Forschungshypothese ist also Ausgangslage für die Festlegung einzelner Merkmale, die dann über die quantitative Inhaltsanalyse erhoben werden sollen.

Festgehalten werden die einzelnen Merkmale in einem „Kategoriensystem“, das sozusagen das „Herzstück“ jeder quantitativen Inhaltsanalyse bildet (vgl. hierzu auch Bortz/Döring 2006, 151: „Kern jeder quantitativen Inhaltsanalyse ist das Kategoriensystem, das festlegt, welche Texteeigenschaften durch Auszählen ‚gemessen‘ werden sollen“). An dieser Stelle wird bereits deutlich, wie wichtig es für das Gelingen einer quantitativen Inhaltsanalyse ist, dass Hypothesen theoriegeleitet begründet werden und dass die Herleitung einzelner Kategorien fachwissenschaftlich bzw. fachdidaktisch abgesichert erfolgt. Denn nur diese Vorarbeit führt zu einem theoretisch „sauberen“ Forschungsdesign und gewährleistet, dass die Inhaltsanalyse auch das erhebt, was sie erheben soll (zur Validität der Methode siehe weiter unten).

Zur Illustration eines Kategoriensystems soll an dieser Stelle ein Kategoriensystem vorgestellt werden, das in einer Studie (Uhl 2015) verwendet wurde, um die Textqualität von Erzählungen zu erheben. In der Forschungsliteratur wird Textqualität als latente Variable aufgefasst. Latente Variablen sind nicht unmittelbar beobachtbar. Hierzu schreiben Bortz und Döring:

Während etwa das Alter einer Person ein eindeutig definiertes und auch formal registriertes (Geburtsurkunde, Ausweis) manifestes Merkmal ist, sind latente Merkmale, wie Studienmotivation, Neurotizismus oder Gewaltbereitschaft, sehr viel ‚schwammiger‘ und müssen im Zuge der Operationalisierung konkretisiert werden. (Bortz/Döring 2006, 3)

Zur Darstellung des Konstrukts *Textqualität* in einem Kategoriensystem wurden theoretisch fundiert mehrere Merkmale hergeleitet, von denen anzunehmen ist, dass sie dieses Konstrukt abbilden. Es ergeben sich somit mehrere *Merkmale narrativer Vertextung*, die in das Kategoriensystem übertragen werden können (s. Abb. 1, hier ist das Kategoriensystem überblicksartig aufgeführt). An oberster Stelle steht im Kategoriensystem die Forschungshypothese.¹ Eine Ebene darunter sind die Variablen der Hypothese angeführt. Die Variable *Merkmale narrativer Vertextung* gliedert sich in mehrere Merkmale, die über die quantitative Inhaltsanalyse erhoben werden.

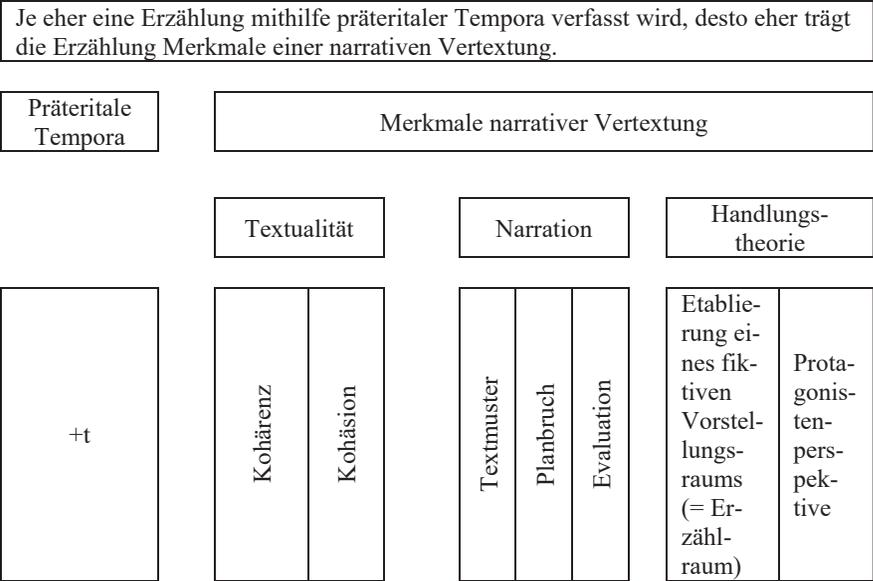


Abb. 1: Beispiel für ein Kategoriensystem

1.2 Ausgangsdaten und Ergebnisse

Auf der Grundlage eines solchen Kategoriensystems kann in einem Folgeschritt die konkrete Kodierung vorbereitet werden. Unter Kodierung versteht man den Vorgang, wenn Einheiten eines Textes einer Kategorie zugeordnet werden („Die Zuordnung von Textteilen zu Kategorien nennt man Kodierung“ Bortz/Döring 2006, 153). Hierbei ist es hilfreich, auf korpuslinguistische Computer-Tools zurückzugreifen. So kann die Kodierung z.B. mithilfe des Computerprogramms EXMARALDA² vorgenommen werden (Abb. 2). Mithilfe des Programms EXAKT³

¹ Die Forschungshypothese und die damit verbundene Studie (Uhl 2015) werden im zweiten Teil dieses Beitrags dargestellt.
² Kostenfrei herunterzuladen im Internet unter: <http://www.exmaralda.org>.
³ EXAKT = „EXMARALDA Analyse und Konkordanztool“, ebenfalls kostenfrei unter <http://www.exmaralda.org> verfügbar.

kann dann in einem Folgeschritt das Korpus durchsucht werden (s. hierzu Abb. 3):

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
ez01 [written]	Die	Entführung	der	Prinzessin	Es	war	einmal	ein	schöner	Sommertag	in	der	Stadt	Terabitija	Es	war	eine	wunderschöne	Stadt
ez01 [verb_crp]						prät										prät			
ez01 [verb_pp]																			
ez01 [ko_wrf]																			
ez01 [kohasion]																			
ez01 [evaluation]									eva_adj1									eva_adj1	
ez01 [verba_decendi]																			
tm_ort																			
ez01 [etablierung_ER]						cta_zeit					cta_ort					cta_ort			
ez01 [perspck]																			
						crp													

Abb. 2: Kodierung mithilfe des Programms EXMARaLDA

Wie in der Abbildung erkennbar, wird zur Kodierung jede Erzählung des Untersuchungskorpus (in dem Beispiel Erzählung 1) in EXMARaLDA eingepflegt (s. hierzu die Spur: ez01 [written]). Im Anschluss kann der Text mit mehreren Annotationen versehen werden, die auf dem Kategoriensystem fußen. Es können somit pro Annotationsspur bestimmte tags vergeben werden (so steht z.B. der tag tm_ori in der Annotationsspur Textmuster für die narrative Textmusterphase Orientierung oder der tag ko_para steht in der Annotationsspur Kohäsion für Satzkonnektor parataktisch. Der Vorteil eines solchen Vorgehens ist, dass die Auszählung der tags nicht mehr händisch erfolgen muss – dies übernimmt das Programm EXAKT:

The screenshot shows the EXMARaLDA EXAKT 1.1 interface. The main window displays a concordance search for the tag 'kohasion' with the regex 'ko'. The results are shown in a table with columns: Communication, Speaker, Left Context, Match, Right Context, and kohasion. The text in the concordance is a German story about a prince and a princess. On the left, there are panels for 'Word lists' and 'Concordances'. On the right, there are search filters and statistics.

Abb. 3: Korpusanalyse mit EXAKT

Beim Auszählen der *tags* unterscheidet das Programm automatisch zwischen *types* und *token* (s. Abb. 3): In dem Beispiel wurden die expliziten Satzkonnectoren (*und, dann, weil, etc.*) von 45 Erzählungen ausgewertet. Wenn *token* gezählt werden, dann wird ausgewertet, wie oft ein Satzverknüpfer verwendet wurde. In den 45 Erzählungen haben die Kinder insgesamt 436 mal Satzverknüpfer genutzt. Bei der Zählung nach *types* wird angegeben, wie viele verschiedene Typen von Satzverknüpfern gebraucht wurden. Hier werden dann z.B. alle *unds*, alle *danns* usw. zu einem Typ zusammengefasst. In den 45 Schülererzählungen wurden 52 verschiedene Gruppen von Satzverknüpfern verwendet. Eine Übersicht über diese verschiedenen Typen kann man sich über das Programm anzeigen lassen. Die über die quantitative Inhaltsanalyse erhobenen Daten können dann zur Datenauswertung verwendet werden. Hierzu können die Daten in Excel oder SPSS übertragen werden, wo verschiedene statistische Verfahren zur Hypothesenprüfung durchgeführt werden können.

1.3 Vorteile der quantitativen Inhaltsanalyse (mit Blick auf Gütekriterien)

Die Vorteile des Verfahrens liegen in einer objektiven, validen und reliablen Datenerhebung. Die Validität wird dadurch gewährleistet, dass alle Merkmale des Kategoriensystems fachwissenschaftlich bzw. fachdidaktisch begründet werden. Bei der Erstellung des Kategoriensystems ist also eine enge Anbindung an den Forschungsstand wichtig; dies gilt besonders bei einer Operationalisierung von latenten Variablen (wie z.B. Textqualität, emotionale Involvierung, s.o.). Wichtige Fragen bei der Erstellung eines Kategoriensystems könnten also lauten: Welche Studien gibt es bereits, die zu ähnlichen Fragestellungen geforscht haben? Wie wurden latente Variablen in diesen Studien messbar gemacht? Lässt sich die von mir entworfene Kategorie fachwissenschaftlich/fachdidaktisch begründen? Außerdem ist es bei der Erstellung eines Kategoriensystems wichtig, die einzelnen Kategorien immer aus der zu untersuchenden Forschungshypothese abzuleiten. Der erste Schritt zur Erstellung eines Kategoriensystems geht demnach immer über die konkrete Forschungsfrage.

Eine hohe Reliabilität wird dadurch erreicht, dass die Kodierung der einzelnen Kategorien klar und eindeutig beschrieben wird. Auch dies sollte fachdidaktisch bzw. fachwissenschaftlich begründet erfolgen. Will man z.B. kodieren lassen, wie viele Satzverknüpfer verwendet werden, muss man dem Kodierer eine klare Anweisung an die Hand geben, was Satzverknüpfer kennzeichnet und wie mit unterschiedlichen Arten von Satzverknüpfern umgegangen werden kann. Man spricht hier auch von einem Kodierleitfaden, der beinhaltet, welche Merkmale wie genau zu messen sind. Zur Erhöhung der Reliabilität empfiehlt sich mit Beispielen, sogenannten *Benchmarks*, zu arbeiten. Zur Prüfung der Reliabilität ist es ratsam, das Korpus mehrfach kodieren zu lassen. Über einfache statistische Verfahren (Rangkorrelationen) lässt sich ein Zusammenhangsmaß der Kodierungen berechnen. Wenn das Zusammenhangsmaß zu niedrig ist (Empfehlungen finden sich bei Bortz/Döring 2006, 273), sollte die Kodierung überarbeitet werden.

Da die einzelnen Merkmale eines Textes nicht wie bei Ratings eingeschätzt, sondern einfach nur ausgezählt werden, ist die Objektivität des Verfahrens sehr hoch. Zusätzlich erhöht wird die Objektivität, wenn mit computergestützten Auswertungsprogrammen gearbeitet wird (s. o.); dies vermindert das Auftreten von Fehlern beim Auszählen. Gerade wenn das Untersuchungskorpus aus vielen Texten besteht, empfiehlt es sich, computergestützt zu arbeiten.

1.4 Nachteile des Verfahrens

Wie bereits erwähnt, „dient [die quantitative Inhaltsanalyse] eher als Erhebungstechnik, denn als Auswertungstechnik wie im qualitativen Paradigma“ (Lamnek 2010, 454). Zur Datenauswertung sind also weitere Schritte nötig. In der Regel bestehen diese weiteren Schritte darin, dass die Werte, die die quantitative Inhaltsanalyse erbracht hat, zur Hypothesenprüfung verrechnet werden (beispielsweise über Korrelationen, Regressionen, etc.). Zu bedenken ist hierbei, dass über die quantitative Inhaltsanalyse Daten erhoben werden, die lediglich anhand der Textoberfläche identifiziert worden sind. Zur Erhebung bestimmter Merkmale, die kontextsensitiv sind (in dem Sinne, dass die Merkmale sehr stark vom Kontext abhängen oder mit diesem interagieren, wie z.B. bei Ironie), ist die quantitative Inhaltsanalyse nicht geeignet. Diesbezüglich schreiben Bortz und Döring:

In der Tat wird die quantitative Inhaltsanalyse versagen, wenn man aus der Häufigkeit einzelner Textelemente z.B. auf die Logik einer Argumentation, auf ironische Absichten oder mangelnden Sachverstand des Urhebers eines Textes schließen will. Der ‚pragmatische Kontext‘, der die inhaltliche Bedeutung eines Textes prägt, kommt bei einer Analyse zu kurz, die Wörter oder Satzteile aus dem Kontext löst und auszählt. (ebd., 149f.)

In diesem Sinne empfiehlt es sich zur Beantwortung einer Fragestellung nicht alleine auf inhaltsanalytische Verfahren zu vertrauen, die rein quantitativ ausgerichtet sind. Das im Anschluss vorgestellte Dissertationsprojekt gibt ein Beispiel für ein Mixed-Methods-Design und zeigt Wege auf, wie quantitative und qualitative Verfahren miteinander kombiniert werden können.

1.5 Literaturempfehlungen

Bortz/Döring (2006): Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler.

Das Buch bietet eine generelle Einführung in wichtige empirische Forschungsmethoden. Zu der quantitativen Inhaltsanalyse findet sich hier ein empfehlenswertes Kapitel, das viele praktische Anwendungsbeispiele enthält.

Evert/Fitschen (2001): Textkorpora.

Wer sich mehr für die technische Seite des *Taggens* und *Kodierens* interessiert und mehr wissen möchte über Methoden zur Analyse und Auswertung von Korpusdaten, kann bei Evert und Fitschen (2001) nachschlagen.

Lamnek (2010): Qualitative Sozialforschung.

Auch dieses Buch bietet eine Einführung in empirische Forschungsmethoden. Inhaltsanalytische Verfahren spielen hierbei eine große Rolle; lesenswert ist vor allem der historische Abriss, der Einblick in die Entstehung der Methode gibt.

Lemnitzer/Zinsmeister (2006): Korpuslinguistik.

Lemnitzer und Zinsmeister legen mit diesem Buch eine kurze Einführung in die Korpuslinguistik vor. Wer plant, mit größeren Forschungskorpora zu arbeiten und wissen möchte, wie man diese verwaltet und wie mit Metadaten umgegangen werden sollte, kann hier nachschlagen.

Mayring (1994): Qualitative Inhaltsanalyse.

Mayring gibt in diesem Beitrag einen kompakten Überblick über die Entstehung der qualitativen Inhaltsanalyse. Hierbei stellt Mayring die qualitative Inhaltsanalyse der quantitativen gegenüber; es wird aber auch auf Gemeinsamkeiten beider Verfahren aufmerksam gemacht.

2. Tempus – Narration – Medialität: Ein Beispiel für den Einsatz der quantitativen Inhaltsanalyse

In meinem Dissertationsprojekt (Uhl 2015) habe ich untersucht, wie sich bei Grundschulkindern die Fähigkeit entwickelt, schriftliche Erzählungen zu schreiben. Hierzu habe ich 45 Schülererzählungen von Drittklässlern analysiert. Als zentrale Forschungshypothese habe ich angenommen, dass die Verwendung des Präteritums mit der Textqualität zusammenhängt. Für einen solchen Zusammenhang spricht zum einen die theoretisch ausgerichtete Forschungsliteratur (Weinrich 1964; Thieroff 1992; Bredel/Töpler 2007; Topalović/Uhl 2014; Hoffmann 2014); zum anderen deutet sich dieser Zusammenhang in empirisch ausgelegten Studien an (Augst et al. 2007, 70; Steinig et al. 2009, 211).

Zur Operationalisierung des Konstrukts Textqualität wurden mehrere *Merkmale narrativer Vertextung* aus dem Forschungsdiskurs abgeleitet – beispielsweise das Gestalten eines Textes mittels narrativer Textmusterphasen (nach Augst 2010, 70) oder das Verwenden evaluativer Adjektive (nach Boueke et al. 1995, 108). Die Ausprägung dieser Merkmale wurde mit Hilfe einer quantitativen Inhaltsanalyse erhoben (Bortz/Döring 2006, 149). Zur Prüfung der Forschungshypothese wurde die Ausprägung dieser Merkmale dann mit der Präteritumsverwendung im jeweiligen Text korreliert. Es kann somit statistisch belegt werden, ob – und wenn ja, wie stark – die Verwendung des Präteritums mit dem Vorhandensein des jeweiligen *Merkmals narrativer Vertextung* zusammenhängt.

Im Folgenden sollen einige ausgewählte Ergebnisse dieser Studie vorgestellt werden. Eine komplette Darstellung der Ergebnisse kann aus Platzgründen nicht erfolgen (für eine komplette Darstellung vgl. Uhl 2015). Anstelle einer vollständigen Ergebnispräsentation liegt der Fokus des Beitrags eher auf einer Reflexion des Einsatzes der Methode in dem Kontext des Forschungsprojektes.

2.1 Kategoriensystem und Kodierung

Wie im ersten Teil skizziert, ist es für die Arbeit mit einer quantitativen Inhaltsanalyse elementar, dass zunächst die einzelnen zu messenden Kategorien in ein Kategoriensystem überführt werden. Gemäß der Forschungshypothese *Je mehr präteritale Finita eine Erzählung enthält, desto mehr Merkmale narrativer Vertextung enthält die Erzählung* lassen sich zwei Variablen ausmachen: Das Vorhandensein präteritaler Tempora und die Merkmale narrativer Vertextung. Das Vorhandensein präteritaler Tempora ist eine manifeste Variable, die sich mit der quantitativen Inhaltsanalyse einfach auszählen lässt. Die Merkmale narrativer Vertextung hingegen müssen erst hergeleitet werden. Hierbei ergeben sich zunächst drei Bereiche (Textualität, Narratologie und Handlungstheorie), die aus dem Forschungsdiskurs hergeleitet werden können. Innerhalb dieser drei Bereiche lassen sich dann Kategorien bilden, die in das Kategoriensystem überführt werden können (zum Kategoriensystem s. auch Abb. 1).

Merkmale narrativer Vertextung			
	Schriftliche Erzählungen sind/werden...		
	gekennzeichnet durch...	theoretisch fundiert durch	Merkmale für das Kategoriensystem
Textualität	besondere textspezifische Merkmale wie Kohärenz und Kohäsion; schriftliches Erzählen wird verstanden als eine komplexe Form des sprachlichen Handelns, die durch Speicherung und Weitergabe über das Medium Schrift <i>verdauert</i> wird.	Ehlich (1983a), (1983b) u. (1994) Vater (1992) Becker-Mrotzek/Böttcher (2006) Fix (2006) Brinker (2010)	- Kohärenz - Kohäsion
Narratologie	erzähltheoretische Merkmale wie den Planbruch, evaluative Elemente und eine spezifische Phasen- bzw. Musterstruktur.	Labov/Waletzky (1967 [1973]) Quasthoff (1980) Rehbein (1980) u. (1984) Boueke et al. (1995) Becker (2001) bzw. (2002)	- Textmuster - Planbruch - Evaluation
Handlungstheorie	eine Origoverschiebung, mittels derer die Situationsorientierung von Sprachproduzent und Sprachrezipient vom Wahrnehmungsraum in den Erzählraum versetzt wird; dies dient der Erfüllung der Tiefenfunktion des Erzählens.	Ehlich (1983b) Rehbein (1980) u. (1984) Fienemann (2006) Bredel (2008)	- Etablierung Erzählraum - Protagonistenperspektive

Tab. 1: Herleitung der Merkmale narrativer Vertextung

Erst auf der Grundlage des Kategoriensystems können dann Überlegungen zur Kodierung vorgenommen werden. Hierbei gilt es zu klären, wie genau kodiert werden kann. Die meisten Kategorien des Kategoriensystems werden über eine Häufigkeitsanalyse (vgl. Bortz/Döring 2006, 151) erhoben⁴; dies betrifft alle Kategorien, die klar anhand der Textoberfläche (d.h. anhand der Mikrostruktur des Textes) identifizierbar sind und somit einfach ausgezählt werden können. Folgende Tabelle zeigt eine Übersicht über die Kategorien, die mittels einer Häufigkeitsanalyse erhoben wurden:

⁴ Wie genau die einzelnen Kategorien des Kategoriensystems operationalisiert und kodiert wurden, kann an dieser Stelle aus Platzgründen nicht detailliert aufgeführt werden. Der interessierte Leser/die interessierte Leserin findet in der Studie (Uhl 2015) eine ausführliche Darstellung.

Kategorie	kodiert durch ...
Kohäsive Mittel	Variation der Satzkonnectoren
Evaluation ⁵	Anzahl expressiver Verben, Adjektive und Adverbien in Relation zur Wortanzahl
Textmuster	Anzahl der einzelnen narrativen Phasen (Orientierung, Komplikation, Auflösung)
Planbruch	Anzahl des skizzierten Handlungsplans und Planbruchs
Etablierung Erzählerperspektive	Anzahl der Elemente (z.B. Ort, Person, Zeit), die zu Beginn der Erzählung genannt werden
Protagonistenperspektive	Anzahl der Redebeiträge und Protagonisten, die sich unterhalten, zusätzlich: Anzahl der Protagonisten in Relation zur Wortanzahl

Tab. 2: Kodierrichtlinien zur Erhebung der einzelnen Kategorien

Die Kategorie *Kohärenz*⁶ ist nicht anhand der Textoberfläche auszählbar, sondern kann nur in Bezug auf die Makrostruktur des Textes ermittelt werden. Diese Kategorie wird von daher über eine Valenz- und Intensitätsanalyse ermittelt; hierzu wird „mit ordinal- und intervallskalierten Variablen, die durch Schätzerurteile quantifiziert werden“ (ebd., 152) gearbeitet (generell zum Rating als Datenauswertungsverfahren s. Kleinbub in diesem Band). Zur Auswertung der Kategorie *Kohärenz* wird daher ein ordinalskaliertes Rating mit zwei Beurteilern genutzt. Den Beurteilern stehen beim Einschätzen der Kategorie *Kohärenz* folgende Ränge zur Verfügung:

Beschreibung
<u>1. Niveau:</u> Keine Gesamtidee erkennbar: In der Erzählung werden verschiedene Ereignisse beschrieben, ohne dass ein logischer Zusammenhang zwischen den Ereignissen deutlich wird.
<u>2. Niveau:</u> Gesamtidee ist in Ansätzen erkennbar: Der Aufbau der Erzählung und die Ereignisfolge ist aber überwiegend unlogisch und nicht nachvollziehbar.
<u>3. Niveau:</u> Gesamtidee ist prinzipiell ausmachbar, wird aber nicht konsequent befolgt: Es werden in der Erzählung noch Ereignisse angeführt, die nicht in Bezug auf eine Gesamtidee interpretierbar sind.
<u>4. Niveau:</u> Gesamtidee ist klar erkennbar: Alle in der Erzählung angeführten Ereignisse sind in Bezug zu einer Gesamtidee interpretierbar. Die Erzählung enthält einen Anfangs- und Endzustand, die zueinander in Bezug stehen.

Tab. 3: Ratingskala zur Bestimmung der Kategorie Kohärenz

⁵ Unter Evaluation bzw. evaluativen Elementen werden expressive Verben/Adjektive und Adverbien gefasst, die der Erzählung eine affektive Komponente verleihen.

⁶ Nach Fix (2006) bezeichnet Kohärenz „die satzübergreifende thematische Beziehung in einem Text. Das Gewebe (lat. *textus*) soll zusammenhängen (lat. *cohaerere*), der ‚Sinn‘, der ‚rote Faden‘ vom Leser in eine stimmige mentale Repräsentation umgesetzt werden“ (Fix 2006, 75).

Zusammenfassend ergibt somit folgendes Kategoriensystem. Die eben skizzierten Kodierrichtlinien wurden in diese Übersicht integriert:

Hypothese		
Je eher eine Erzählung mithilfe präteritaler Tempora verfasst wird, desto eher trägt die Erzählung Merkmale einer narrativen Vertextung .		
Anzahl präteritaler Tempora		
Kategorien d. quantitativen Inhaltsanalyse	Kodierung	Kodiereinheit
Anzahl präteritaler Verbformen	Auszählen; absolut u. prozentual (Häufigkeitsanalyse)	Mikrostruktur
Merkmale narrativer Vertextung		
Kategorien d. quantitativen Inhaltsanalyse	Kodierung	Kodiereinheit
Beurteilung der Textkohärenz	Rating	Makrostruktur
Variation der Satzkonnectoren	Auszählen (Häufigkeitsanalyse)	Mikrostruktur
Anzahl der narrativen Textmusterphasen	Auszählen (Häufigkeitsanalyse)	Mikrostruktur
Verwendung eines Planbruchs	Auszählen (Häufigkeitsanalyse)	Mikrostruktur
Anzahl der evaluativen Markierungen (<i>token</i> und <i>token</i> zu Wortanzahl)	Auszählen (Häufigkeitsanalyse)	Mikrostruktur
Anzahl der sprachlich korrekt eingeführten Elemente im Erzählraum (<i>types</i>)	Auszählen (Häufigkeitsanalyse)	Mikrostruktur
Anzahl der Redebeiträge (<i>token</i>), Anzahl Aktanten, die sich in der Protagonistenperspektive unterhalten (= <i>types</i>), sowie <i>types</i> durch Wortanzahl	Auszählen (Häufigkeitsanalyse)	Mikrostruktur

Tab. 4: Kategoriensystem und Kodierrichtlinien

Zur Veranschaulichung, welche Daten mit dieser Form der quantitativen Inhaltsanalyse generiert werden, soll im Folgenden beispielhaft eine Erzählung analysiert werden. In der folgenden Transkription sind die Tempora hervorgehoben.

Es **war** einmal vor vielen Jahren ein verlassenes Schloss. Dort **wohnten** eine Hexe und ein Geist. Der Geist **fragte** die Hexe. „Was machst du da?“ Die Hexe **antwortete** „ich zaubere an der Zauberkugel.“ Dann **sagte** der Geist ich schwebte ein bisschen durch [das] Schloss und erschrecke die Fledermäuse. Plötzlich **kam** auf einmal ein riesiger Schatten über ihn. Der Geist **schwebte** weg in eine Ecke da **sah** er dass ein riesengroßer Riese vor ihm **stand** schnell **schwebte** er zur Hexe und **sagte** „in dem Schloss ist ein Riese“ **sagte** der Geist stotternd vor Angst „und der Himmel brennt bestimmt ist er so groß dass er den Himmel sogar anzünden kann.“ Da **erschrak** die Hexe und **machte** schnell einen Zaubertrank um den Riesen zu verschrecken und den Himmel zu löschen. Das **hat** sie dann ganz schnell und die Hexe **hat** es geschafft den Riesen zu verschrecken und den Himmel gelöscht und die Hexe **sagte** gut dass du mich gewarnt hast so konnten wir den Riesen besiegen.“ Dann **war** wieder alles friedlich der Geist **konnte** dann wieder in Ruhe rumspuken.

Anhand der quantitativen Inhaltsanalyse ergibt sich zunächst für die Variable ‚Tempusgebrauch‘:

Finite Verbformen der Erzählerperspektive	19
davon präterital	17
prozentual	89,47%

Tab. 5: Auswertung der Variable Tempusgebrauch

Für die ‚Merkmale narrativer Vertextung‘ können folgende Daten mittels der quantitativen Inhaltsanalyse erhoben werden:

Raterurteile			
	Rater 1	Rater 2	
Kohärenz	4	4	
Analyseurteile			
	<i>token</i>	<i>types</i>	
Wortanzahl	177	/	
Kohäsion	14 (= <i>dort, dann*4, plötzlich, da*2, dass, und*5</i>)	6	<i>type-token-Ratio:</i> 0,42
Evaluation	11 (= <i>verlassenes Schloss, riesiger Schatten, riesengroß, stotternd vor Angst, groß, erschrecken*2, friedlich, schnell*3</i>)	0,062 Token auf Wortanzahl	
Etablierung Erzählerperspektive	4 (Zeit = <i>es war einmal vor vielen Jahren</i> ; Ort = <i>ein verlassenes Schloss</i> ; Person = <i>Hexe</i> ; Person = <i>Geist</i>)		
Protagonistenperspektive	6 Redebeiträge	2 Protagonisten (<i>Geist, Hexe</i>)	<i>types</i> bezogen auf Wortanzahl: 0,011
Textmuster	3		
Planbruch	Handlungsplan: <i>Dann sagte der Geist ich schwebte ein bisschen durch [das] Schloss und erschreckte die Fledermäuse.</i>	Planbruch (sprachlich exponiert): <i>Dann <u>kam auf einmal</u> ein riesiger Schatten über ihn. Der Geist schwebte weg in eine Ecke da sah er dass ein riesengroßer Riese vor ihm stand.</i>	

Tab. 6: Auswertung der Variable Merkmale narrativer Vertextung

2.2 Ergebnisse der Studie – mit quantitativem Blick

Die über die quantitative Inhaltsanalyse erhobenen Daten können nun für hypothesenprüfende Verfahren verwendet werden. Hierzu wird gemäß der Forschungshypothese die Verwendung des Präteritums mit der Ausprägung der Merkmale narrativer Vertextung korreliert. Die Ergebnisse liefern zum einen Informationen über den Zusammenhang von Tempusgebrauch und der Ausprägung der Merkmale narrativer Vertextung: Dieser Zusammenhang wird über den Korrelationskoeffizienten (r) ersichtlich. Der Korrelationskoeffizient gibt einen Wert von 0 bis 1 an. Der Wert 0 bedeutet, dass kein Zusammenhang vorliegt, der Wert 1 deutet auf einen absoluten Zusammenhang hin. In der Forschungsliteratur gilt

ein Wert von $r < 0,5$ als geringe Korrelation, $r < 0,7$ als mittlere Korrelation und ein Wert von $r < 0,9$ als hohe Korrelation (Bühl 2008, 346). Neben dem Korrelationskoeffizienten geben Korrelationen aber zum anderen auch Auskunft über die Irrtumswahrscheinlichkeit (p): Der p -Wert liegt ebenfalls zwischen 0 und 1. Hier gilt, dass bei einem Wert von 0 ein Irrtum absolut auszuschließen ist. Ein Wert von 1 hingegen bedeutet, dass zu 100% ein Irrtum vorliegt. Für (p) gilt in der Forschung ein Wert von $p < 0,05$ als signifikant, d.h. ein Irrtum kann mit einer Wahrscheinlichkeit von 5% ausgeschlossen werden (vgl. Bühl 2008, 121; Bortz 2005, 114). Ist $p < 0,01$ gilt der Wert als hochsignifikant, d.h. hier ist die Irrtumswahrscheinlichkeit auf 1% minimiert. Die Korrelationen werden nach Pearson berechnet, wenn die Daten metrisch skaliert sind. Dies trifft bei allen Kategorien des Kategoriensystems mit einer Ausnahme zu: Die Kategorie *Kohärenz* wurde mithilfe von ordinalskalierten Ratings erhoben und kann daher nur mittels Rangkorrelationen nach Spearman verrechnet werden (vgl. hierzu Wirtz/Casper 2002, 123ff.).

Bei der Verrechnung der beiden Variablen der Forschungshypothese ergibt sich, dass je mehr präteritale Tempora eine Erzählung aufweist,

- desto eher werden satzkohäsive Mittel variiert: $r=0,43$ (Pearson), $p < 0,01$;
- desto eher werden Adjektive, Adverbien und expressive Verben verwendet, um die Erzählung zu evaluieren: $r=0,33$ (Pearson), $p < 0,05$;
- desto eher wird ein narratives Textmuster bestehend aus den Phasen der Orientierung, Komplikation und Auflösung genutzt: $r=0,57$ (Pearson), $p < 0,01$;
- desto eher wird die Erzählung mithilfe eines Planbruchs gestaltet, der einen Handlungsplan eines Aktanten infrage stellt: $r=0,34$ (Pearson), $p < 0,05$;
- desto detaillierter wird zu Beginn der Erzählung ein Erzählraum etabliert: $r=0,46$ (Pearson), $p < 0,01$;
- desto weniger Protagonisten werden in die Erzählung eingeführt, die sich in Form der wörtlichen Rede unterhalten: $r=-0,38$ (Pearson), $p < 0,01$;⁷
- desto eher ist in den Erzählungen eine Gesamtidee zu erkennen: $r=0,60$ (Spearman-Rho), $p < 0,01$.

Der Vorteil dieses Vorgehens ist, dass über die Korrelationen belastbare Daten generiert werden. D.h. es kann statistisch belegt davon ausgegangen werden, dass das Verwenden des Präteritums mit dem Vorhandensein eines bestimmten Merkmals narrativer Vertextung zusammenhängt. Allerdings geben Korrelationen immer nur Auskünfte über einen Zusammenhang, nicht aber über Kausalität (d.h. es kann anhand der Daten „nur“ gezeigt werden, dass die Variablen zusammenhängen, es kann aber nichts über ein „Ursache-Wirkungs-Verhältnis“ ausgesagt werden). Als Ergänzung zu dieser eben beschriebenen quantitativen Untersuchung ist daher ein qualitativer Blick auf die Daten sehr hilfreich und erkenntnisgewinnend. Dies soll im nächsten Unterkapitel dargelegt werden.

⁷ Dieses Ergebnis überrascht, kann aber an dieser Stelle nicht weiter kommentiert werden (s. hierzu Uhl 2015, 217ff.).

2.3 Streudiagramme interpretieren: Vom quantitativen zum qualitativen Zugang

In der Studie (Uhl 2015) wurden immer im Anschluss an die Berechnung der Korrelation das entsprechende Streudiagramm mit qualitativem Blick analysiert (für ein Beispiel siehe unten). Es können somit ähnliche Erzählungen identifiziert werden, die sich in einem Folgeschritt zu Subgruppen zusammenfassen lassen. Diese Subgruppen können dann mit einer qualitativ ausgerichteten Fragestellung untersucht werden (gefragt wurde in der Studie: Welche Gemeinsamkeiten finden sich in den einzelnen Erzählungen einer Subgruppe?). Des Weiteren können aber auch auffällige Erzählungen, die der Forschungshypothese widersprechen, sichtbar gemacht werden. Es kann somit mit qualitativem Blick der Frage nachgegangen werden, ob diese Ausnahmen nicht weiter begründbare Ausreißer darstellen, oder aber, ob diesen Ausnahmen eventuell eine Systematik zu Grunde liegt. Die Betrachtung des Streudiagramms fungiert an dieser Stelle also als eine Art Screening, mit dessen Hilfe das Korpus geordnet werden kann und einzelne, besondere Texte identifiziert werden (vgl. hierzu im Detail Uhl 2015). Dieses Vorgehen möchte ich exemplarisch anhand des Zusammenhangs der Variablen *Präteritumsverwendung* und *Evaluation* demonstrieren:

Der Zusammenhang zwischen der Verwendung des Präteritums und der Verwendung von evaluativen Adjektiven, Adverbien und expressiven Verben ist wie dargestellt (s.o.) relativ gering ($r=0,33$ (Pearson)). Mit Blick auf das entsprechende Streudiagramm, das diesen Zusammenhang abbildet, werden mehrere auffällige Erzählungen sichtbar (zu dem Streudiagramm: In dem Streudiagramm befindet sich links auf der Y-Achse der Wert für die Verwendung evaluativer Ausdrücke (dieser ist durch die Wortanzahl dividiert, um den Wert in Relation zur Textlänge zu setzen). Auf der X-Achse befindet sich die Verwendung des Präteritums (als Prozentwert)).

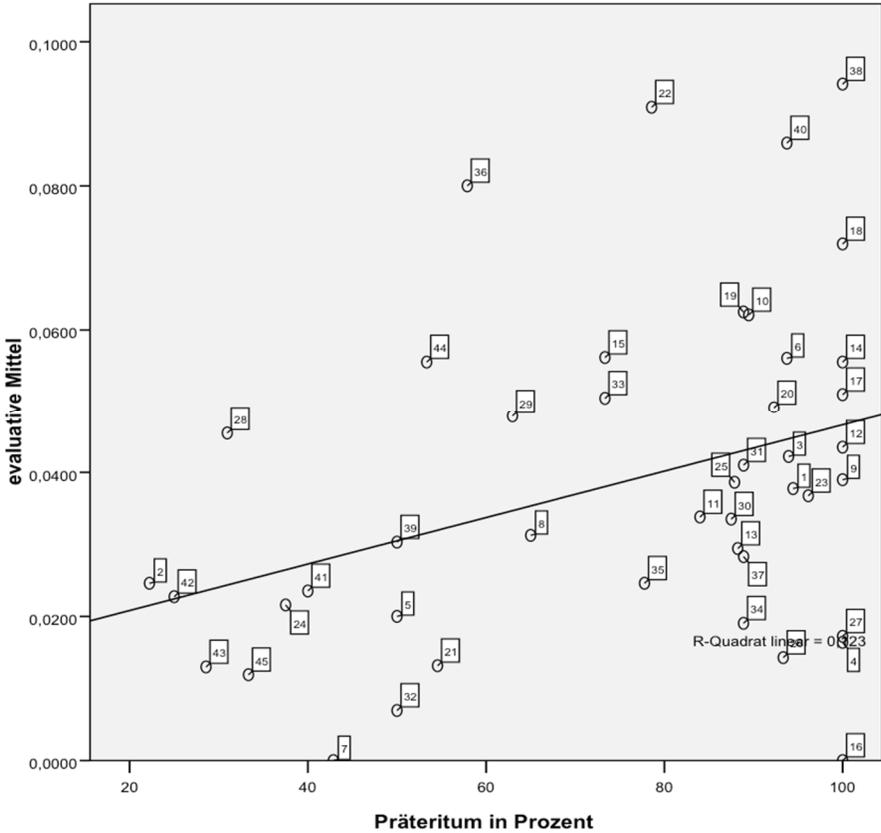


Abb. 4: Streudiagramm Präteritumsverwendung und evaluative Mittel (Uhl 2015, 163)

Eine mögliche Subgruppe bilden beispielsweise die Erzählungen, bei denen die Autorinnen und Autoren wenig Präteritum verwenden und wenig evaluative Mittel genutzt werden (Erzählung 2, 24, 41, 42, 43, 45). Eine weitere Subgruppe könnten die Erzählungen bilden, bei denen sowohl die Verwendung des Präteritums als auch die Verwendung evaluativer Ausdrücke hoch sind (3, 6, 9, 10, 12, 14, 17, 18, 19, 20, 38, 40). Diese beiden Subgruppen entsprechen der Forschungshypothese. Um aber zu analysieren, was diese Erzählungen für Gemeinsamkeiten besitzen und ob die analysierten Subgruppen vielleicht noch weiter subdifferenziert werden können/sollten, ist ein qualitativer Blick auf die Texte nötig. Aus Platzgründen können die Ergebnisse der qualitativen Analyse allerdings an dieser Stelle nicht vorgestellt werden. Zum Abschluss dieses Beitrags möchte ich stattdessen nochmals ein Blick auf die verwendete Methode lenken: Hierzu möchte ich eine Erzählung genauer analysieren, die eigentlich mithilfe der quantitativen Inhaltsanalyse als Ausnahme ausgewiesen wurde (Erzählung 4). Somit soll zum Abschluss nochmals mithilfe eines konkreten Beispiels auf Schwächen des Verfahrens aufmerksam gemacht werden.

Erzählung 4

Die Suche der Krone

Eines Tages suchte ein König seine Krone. Erst suchte er in seinem [seinen] Schloss aber da war sie nicht. Danach ging er zur Hexe. Die Hexe fragte was willst du hier? Ich möchte mal in deine Zauberkugel gucken. Aber warum denn? Ich such meine Krone. Aber der König sah sie nicht. Er suchte noch in der Wüste im Wald und auf der Wiese. Aber auch da war sie nicht. Auf einmal kam der Riese. Ich weiß, du suchst deine Krone. Sie liegt auf dem [den] Fliegenpilz. Er ging gucken. Die Krone war auch da aber sie lag zu hoch. Er ging zurück ins Schloss eine Leiter holen. Mit der Leiter nahm der König die Krone. Danach ging er glücklich und zufrieden in sein Schloss.

Erzählung 4 trägt viele Merkmale narrativer Vertextung und ist gemäß der Forschungshypothese im Präteritum verfasst. Mit der quantitativen Inhaltsanalyse können jedoch nur zwei evaluative Ausdrücke ermittelt werden (in der Transkription unterstrichen). Diese geringe Verwendung evaluativer Markierungen bedeutet aber nicht, dass der Text per se keine emotionale Färbung trägt: Anstatt in der Erzählung evaluative Markierungen explizit sprachlich deutlich zu machen, z.B. wie in anderen Erzählungen durch das Adverb *überall* (etwa *Er suchte überall nach seiner Krone*), wird in diesem Text eine emotionale Färbung inhaltlich zum Ausdruck gebracht: *Aber der König sah sie nicht. Er suchte noch in der Wüste im Wald und auf der Wiese. Aber auch da war sie nicht.* Gerade durch diese repetitiven Satzanfänge wird das Bestreben des Protagonisten, unbedingt seine Krone zu finden, auf eine implizite Art und Weise erkennbar. Das bedeutet also, dass auch in der Erzählung evaluiert wird; nur vollzieht sich diese Evaluation sehr stark inhaltlich und nicht durch das Verwenden explizit sprachlicher Mittel. Diese sprachliche Besonderheit von Erzählung 4 wäre bei rein quantitativer Betrachtung verborgen geblieben. Hier werden also Grenzen der Methode sichtbar.

2.4 Die Arbeit mit quantitativen Inhaltsanalysen: ein Fazit

Ziel dieses Beitrags war es, die quantitative Inhaltsanalyse als Methode vorzustellen. Mithilfe eines konkreten Forschungsprojektes im zweiten Teil des Beitrags sollten die Wirkungsweise sowie Vor- und Nachteile der Methode verdeutlicht werden. Als Vorteil der Methode kann angeführt werden, dass durch ihren Einsatz in einem validen, objektiven und reliablen Maße Daten erhoben werden können. Voraussetzung hierfür ist allerdings, dass zuvor ein Kategoriensystem ausgearbeitet wurde, in dem die einzelnen zu untersuchenden Kategorien fachwissenschaftlich bzw. fachdidaktisch präzise hergeleitet werden. Gezeigt wurde im Kontext eines Forschungsprojektes, dass es sinnvoll ist, rein quantitative Verfahren durch qualitative Verfahren zu ergänzen. Dieser Beitrag war somit ein Plädoyer dafür, die Wahl und die Auslegung einer Forschungsmethode immer unmittelbar an das Erkenntnisinteresse des Forschungsprojektes zu binden.

Literatur

- Augst, Gerhard/Disselhoff, Katrin/Henrich, Alexandra/Pohl, Thorsten/Völzing, Paul-Ludwig: Text-Sorten-Kompetenz. Eine echte Longitudinalstudie zur Entwicklung der Textkompetenz im Grundschulalter. Frankfurt a.M.: Peter Lang 2007.
- Augst, Gerhard: Zur Ontogenese der Erzählkompetenz in der Primar- und Sekundarstufe. In: Pohl, Thorsten; Steinhoff, Torsten (Hrsg.): Textformen als Lernformen. Duisburg: Gilles & Francke 2010, 63-96.
- Becker, Tabea: Kinder lernen Erzählen. Zur Entwicklung der narrativen Fähigkeiten von Kindern unter Berücksichtigung der Erzählform. Baltmannsweiler: Schneider Verlag Hohengehren 2001.
- Becker, Tabea: Mündliches und schriftliches Erzählen. Ein Vergleich unter entwicklungs-theoretischen Gesichtspunkten. In: Didaktik Deutsch (2002), 12, 23-38.
- Becker-Mrotzek, Michael/Böttcher, Ingrid: Schreibkompetenz entwickeln und beurteilen. Berlin: Cornelsen Scriptor 2006.
- Berelson, Bernard: Content Analysis in Communication Research. Glencoe: The Free Press 1952.
- Bortz Jürgen/Döring, Nicola: Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Heidelberg: Springer 2006.
- Bortz, Jürgen: Statistik für Human- und Sozialwissenschaftler. Heidelberg: Springer 2005.
- Boueke, Dietrich/Schüle, Frieder/Büscher, Hartmut/Terhorst, Evamaria/Wolf, Dagmar: Wie Kinder erzählen. Untersuchungen zur Erzähltheorie und zur Entwicklung narrativer Fähigkeiten. München: Fink 1995.
- Bredel, Ursula/Töpler, Cäcilia: Das Verb. In: Hoffmann, Ludger (Hrsg.): Deutsche Wortarten. Berlin/New York: De Gruyter 2007, 823-901.
- Bredel, Ursula: Deixis in Mündlichkeit und Schriftlichkeit. In: Ahrenholz, Bernt/Bredel, Ursula/Klein, Wolfgang/Rost-Roth, Martina/Skiba, Romuald (Hrsg.): Empirische Forschung und Theoriebildung. Beiträge aus Soziolinguistik, Gesprochene-Sprach- und Zweitspracherwerbsforschung. Festschrift für Norbert Dittmar zum 65. Geburtstag. Frankfurt a.M.: Peter Lang 2008, 129-138.
- Brinker, Klaus: Linguistische Textanalyse: Eine Einführung in Grundbegriffe und Methoden. Berlin: Erich Schmidt Verlag 2010.
- Bühl, Achim: SPSS 16. Einführung in die moderne Datenanalyse. München: Pearson Studium 2008.
- Ehlich, Konrad: Alltägliches Erzählen. In: Sanders, Willy/Wegenast, Klaus (Hrsg.): Erzählen für Kinder – Erzählen von Gott. Mainz: Verlag W. Kohlhammer 1983b, 128-150.
- Ehlich, Konrad: Funktion und Struktur schriftlicher Kommunikation. In: Günther, Hartmut/Ludwig, Otto (Hrsg.): Schrift und Schriftlichkeit. Writing in Use. Ein interdisziplinäres Handbuch internationaler Forschung. An Interdisciplinary Handbook of International Research. Berlin/New York: De Gruyter 1994, 18-41.
- Ehlich, Konrad: Text und sprachliches Handeln. Die Entstehung von Texten aus dem Bedürfnis der Überlieferung. In: Assmann, Aleida/Assmann, Jan/Hardmeier, Christof (Hrsg.): Schrift und Gedächtnis. München: Fink 1983a, 24-43.

- Evert, Stefan/Fitschen, Arne: Textkorpora. In: Carstensen, Kai-Uwe/Ebert, Cornelia/Endriss, Christian/Jekat, Susanne/Klabunde, Ralf/Langer, Hagen (Hrsg.): *Computerlinguistik und Sprachtechnologie. Eine Einführung*. Heidelberg: Spektrum Akademischer Verlag 2001, 369-376.
- Finemann, Jutta: *Erzählen in zwei Sprachen. Diskursanalytische Untersuchungen von Erzählungen auf Deutsch und Französisch*. Münster et al.: Waxmann 2006.
- Fix, Martin: *Texte schreiben – Schreibprozesse im Deutschunterricht*. Paderborn: Schöningh UTB 2006.
- Hoffmann, Ludger: *Deutsche Grammatik. Grundlagen für Lehrerbildung, Schule, Deutsch als Zweitsprache und Deutsch als Fremdsprache*. Berlin: Erich Schmidt Verlag 2014.
- Labov, William/Waletzky, Joshua: *Erzählanalyse: Mündliche Versionen persönlicher Erfahrung*. In: Ihwe, Jens (Hrsg.): *Literaturwissenschaft und Linguistik, Bd. 2*. Frankfurt a.M.: Fischer-Athenäum 1967/1973, 78-126.
- Lamnek, Siegfried: *Qualitative Sozialforschung: Lehrbuch*. Weinheim et al.: Beltz 2010.
- Lasswell, Harold D.: *The Theory of Political Propaganda*. In: *American Political Science Review* (1927), 21, 627-631.
- Lemnitzer, Lothar/Zinsmeister, Heike: *Korpuslinguistik: Eine Einführung*. Tübingen: Gunter Narr Verlag 2006.
- Mayring, Philipp: *Qualitative Inhaltsanalyse*. In: Boehm, Andreas/Mengel, Andreas/Muhr, Thomas (Hrsg.): *Texte verstehen – Konzepte, Methoden, Werkzeuge*. Konstanz: Universitätsverlag 1994, 159-176.
- Quasthoff, Uta M.: *Erzählen in Gesprächen. Linguistische Untersuchungen zu Strukturen und Funktionen am Beispiel einer Kommunikationsform des Alltags*. Tübingen: Gunter Narr Verlag 1980.
- Rehbein, Jochen: *Beschreiben, Berichten und Erzählen*. In: Ehlich, Konrad (Hrsg.): *Erzählen in der Schule*. Tübingen: Gunter Narr Verlag 1984, 67-124.
- Rehbein, Jochen: *Sequentielles Erzählen. Erzählstrukturen von Immigranten bei Sozialberatungen in England*. In: Ehlich, Konrad (Hrsg.): *Erzählen im Alltag*. Frankfurt a.M.: Suhrkamp 1980, 64-108.
- Steinig, Wolfgang/Betzel, Dirk/Geider, Franz Josef/Herbold, Andreas: *Schreiben von Kindern im diachronen Vergleich*. Münster et al.: Waxmann 2009.
- Thieroff, Rolf: *Das Finite Verb im Deutschen. Tempus – Modus – Distanz*. Tübingen: Gunter Narr Verlag 1992.
- Topalović, Elvira/Uhl, Benjamin: *Linguistik des literarischen Erzählens*. In: *Zeitschrift für Germanistische Linguistik* (2014), 40, 26-49.
- Uhl, Benjamin: *Tempus, Narration und Medialität. Eine Studie über die Entwicklung schriftlicher Erzählfähigkeit an der Schnittstelle zwischen Grammatik und Schreiben*. Baltmannsweiler: Schneider Verlag Hohengehren 2015.
- Vater, Heinz: *Einführung in die Textlinguistik. Struktur, Thema und Referenz in Texten*. München: Fink 1992.
- Weinrich, Harald: *Tempus – Besprochene und erzählte Welt*. Mainz: Verlag W. Kohlhammer 1964.
- Wirtz, Markus/Caspar Franz: *Beurteilerübereinstimmung und Beurteilerreliabilität. Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen: Hogrefe 2002.

Kriteriengeleitetes Rating

Ein Datenerhebungsverfahren zur Einschätzung von Unterrichtsmerkmalen

1. Vorstellung des Verfahrens: Das Rating als Instrument zur systematischen Beobachtung

Das Ratingverfahren – man spricht auch von Einschätzungsverfahren, Schätz-, Einstufungs- oder Beurteilungsverfahren (engl. to rate: *schätzen*) – ist ein hochinferentes Datenerhebungsverfahren und dadurch gekennzeichnet, dass die Messwerte durch Schätzurteile von Beurteilern zustande kommen. Es zählt zu den in der angewandten psychologischen und sozialwissenschaftlichen Forschung am häufigsten verwendeten, gleichzeitig aber auch umstrittensten Erhebungsverfahren, das dazu dient, komplexe verhaltensmäßige, soziale und emotionale Vorgänge zu messen. In der empirischen Unterrichtsforschung gilt das Beobachterrating u.a. als eine der zentralen Methoden zur Messung von Unterrichtsqualität im Rahmen von Videostudien.¹ Es nutzt die Fähigkeit des menschlichen Gehirns zur Integration einer Vielzahl von Einzelindikatoren (Indikatorenverschmelzung). Es erfolgt eine Gewichtung und Bedeutungsbeimessung der einzelnen Anhaltspunkte in Abhängigkeit vom Gesamtkontext, wodurch eine überschaubare Auswahl bereits aufbereiteter Daten entsteht. Diese sind dann gebräuchlichen statistischen Auswertungsverfahren zugänglich.

1.1 Funktionsweise des Beobachterratings

Die Beobachtung ist eine Art der visuellen Wahrnehmung, die sich dadurch auszeichnet, dass sie zielgerichtet und von einer planvollen, selektiven Suchhaltung bestimmt ist. Aus einem Ablauf von Ereignissen wird etwas aktiv zum Objekt der

¹ In diesem Artikel wird mit Bezug zum zweiten Kapitel insbesondere auf Ratings als Beobachtungsverfahren von Lehr-Lern-Prozessen eingegangen. Mit Ratings lassen sich im Rahmen der empirischen Unterrichtsforschung aber auch andere „Messobjekte“ wie z.B. Lernprodukte hinsichtlich verschiedener Merkmale einschätzen (Kürzinger und Pohlmann-Rother 2014 schätzen die Qualität von Schülertexten ein; Boelmann 2015 schätzt Schüleräußerungen nach dem Gehalt literarischer Kompetenz ein).

Aufmerksamkeit gemacht, um das Beobachtete im Anschluss auszuwerten. Die Selektion impliziert, dass Entscheidungen darüber zu treffen sind, was ins Zentrum der Aufmerksamkeit rücken soll und wie das Beobachtete zu interpretieren ist. Demnach ist der Beobachtung eine gewisse Subjektivität inhärent, die es so weit wie möglich einzuschränken oder zu kontrollieren gilt. Dies ist laut Bortz und Döring möglich, wenn „Regeln angegeben werden, die den Beobachtungsprozess so eindeutig festlegen, dass die Beobachtung zumindest theoretisch nachvollzogen werden kann“ (Bortz/Döring 2006, 263ff.). Das Ratingverfahren ist ein Instrument, das diese Möglichkeit bietet und daher oftmals zur systematischen Beobachtung genutzt wird.

Der Ratingvorgang ist durch mehrere Vorgänge gekennzeichnet: Es ist ein Messobjekt (z.B. eine Unterrichtsstunde) vorhanden; jedoch nicht alles, was dieses Objekt charakterisiert, ist von Interesse. Das, was gemessen werden soll, ist das Merkmal (z.B. *Strukturiertheit der Unterrichtsstunde*). Durch dieses Merkmal wird ein Beobachtungsauftrag an den Rater formuliert, der seine Wahrnehmung und Erlebnisempfindlichkeit bahnt. Entsprechend der Beobachtung geht vom Objekt eine Eindruckswirkung auf den Rater aus. Der Rater gibt über das subjektiv erlebte Mehr oder Minder der Eindruckswirkung Auskunft in Form eines Urteils auf einer abgestuften Skala. Mit der Abstufung ist eine zahlenmäßige Zuordnung verbunden. Das Ergebnis dieses Messvorganges ist eine Zahl, die das Objekt in dem betreffenden Merkmal charakterisiert (vgl. Langer/Schulz von Thun 2007, 14).

1.2 Ergebnisse des Ratings

Die Skalen geben (z.B. durch Zahlen: 1 – 2 – 3 – 4 oder verbale Beschreibungen: *trifft nicht zu – trifft eher nicht zu – trifft eher zu – trifft zu*) markierte Abschnitte eines Merkmalkontinuums vor, die die Rater als gleich groß bewerten sollen; man geht also davon aus, dass die Stufen der Ratingskala eine Intervallskala bilden. Die Stufenanzahl kann geradzahlig oder ungeradzahlig sein. Ungeradzahlige Ratingskalen enthalten eine neutrale Mittelkategorie und ermöglichen bei unsicheren Urteilen das Ausweichen auf diese Neutralkategorie. Geradzahlige Skalen verzichten auf diese neutrale Kategorie und fordern damit vom Rater ein (zumindest tendenziell) in eine Richtung weisendes Urteil. Dies empfiehlt sich, wenn mit Verfälschungen der Urteile durch eine zentrale Tendenz der Urteiler zu rechnen ist. Die Anzahl der Skalenstufen ist ebenfalls sorgfältig zu erwägen. Mit steigender Anzahl nimmt die Differenzierungsfähigkeit einer Skala zu, bis schließlich die Differenzierungskapazität der Rater ausgeschöpft ist (vgl. Bortz/Döring 2006, 176ff.). Die Daten, die so generiert werden, lassen sich mit den gebräuchlichen statistischen Verfahren auswerten.

1.3 Vorteile und Nachteile des Ratings

Beim Rating handelt sich um ein hoch-inferentes Verfahren: Inferenz bezieht sich hierbei auf das Ausmaß an Schlussfolgerungen und Interpretation, das von den Ratern bei der Beobachtung verlangt wird.

Ein Vorteil dieses Verfahrens ist, dass es sehr nah an die Erlebnisrealität von Menschen und an deren alltäglich zur Lebensbewältigung und Entscheidungsbildung praktiziertes Einordnen ihrer Umwelt heranreicht. Ein weiteres Plus liegt darin, dass mit großer Einfachheit in der Handhabung eine überschaubare Auswahl bereits aufbereiteter Daten entsteht (vgl. Langer/Schulz von Thun 2007, 20ff.).

Der wesentliche Kritikpunkt gegen diese Art der Datenerhebung gründet sich auf die Subjektivität und Undifferenziertheit des hoch-inferenten Verfahrens und dem damit verbundenen Zweifel an dessen Zuverlässigkeit (Reliabilität) und Gültigkeit (Validität). Dies bedeutet, dass das Ergebnis einer Messung nicht nur vom gemessenen Objekt, sondern auch von der Beschaffenheit des verwendeten Messinstrumentes (Ratinginventar) und von der des Reizverarbeitungssystems des Raters abhängt. Da beim Ratingverfahren die Merkmale nicht auf leicht zählbare Einzelereignisse reduziert werden, wie dies bei niedrig-inferenten Verfahren der Fall ist², sondern der Rater als „lebendes Reizdeutungs- und -verarbeitungssystem“ (Langer/Schulz von Thun 2007, 22) eine Vielzahl von Anhaltspunkten integriert und interpretiert, kann es je nach Wahrnehmungsakzentuierung und intuitiver Gewichtung zu verschiedenen Messwerten kommen; beispielsweise indem nicht die Qualität des Unterrichts, sondern subjektive Theorien ‚guten Unterrichts‘ des Beobachters erfasst werden.

Die Brauchbarkeit von Urteilen, die über Ratingskalen gewonnen werden, kann darüber hinaus durch systematische Urteilsfehler eingeschränkt sein. Hierzu zählt beispielsweise der sog. Haloeffekt, ein Versäumnis des Urteilers, konzeptuell unterschiedliche und potenziell unabhängige Merkmale im Urteil zu differenzieren. Haloeffekte treten verstärkt auf, wenn das einzuschätzende Merkmal ungewöhnlich, nur schwer zu beobachten oder schlecht definiert ist. Der Milde-Härte-Fehler besagt, dass die zu beurteilenden Ereignisse systematisch entweder zu positiv oder zu negativ eingestuft werden. Der Urteilsfehler der zentralen Tendenz bezeichnet eine Tendenz, alle Urteilsobjekte im mittleren Bereich der Urteilsskala einzustufen und extreme Ausprägungen zu vermeiden. Bei der Rater-Ratee-Interaktion können Urteilsverzerrungen in Abhängigkeit von der Position des Urteilers auf der zu beurteilenden Dimension entstehen. Urteiler mit extremer Merkmalsausprägung können die Merkmalsausprägungen anderer in Richtung der eigenen Merkmalsausprägung (Ähnlichkeitsfehler) oder in Richtung auf das gegensätzliche Extrem verschätzen (Kontrastfehler). Der Primacy-Recency-Effekt bezeichnet Urteilsverzerrungen, die mit der sequenziellen Position der zu beurteilenden Objekte zusammenhängen. Werden Objekte mit extremer Merkmalsausprägung

² Zu grundlegenden Unterscheidungsaspekten von hoch- und niedrig-inferenten Verfahren (vgl. z.B. Lotz et al. 2013b, 359).

z.B. zu Anfang beurteilt, können die nachfolgenden Beurteilungen von den ersten Beurteilungen abhängen.

Derzeit gibt es nur wenige Untersuchungen, die sich dezidiert mit der Eignung von Beobachterratings zur Messung von Unterrichtsqualität beschäftigen (s. z.B. Clausen et al. 2003; Lotz et al. 2013b). Wie Clausen (2002) sowie Kunter und Baumert (2006) zeigen konnten, sind Schätzurteile abhängig von der eingenommenen Perspektive. So bestehen mitunter nur geringe Übereinstimmungen zwischen Unterrichtseinschätzungen von Lehrerinnen und Lehrern, Schülerinnen und Schülern sowie externen Beobachterinnen und Beobachtern.

Praetorius (2014) untersuchte gezielt, wie reliabel und valide Beobachterratings zur Messung von Unterrichtsqualität sind. Hierbei stützte sie sich insbesondere auf die Qualitätsdimensionen *Klassenführung*, *Schülerorientierung* und *kognitive Aktivierung*, also generische (d.h. fachunabhängige) Merkmale der Unterrichtsqualität. Für fachdidaktisch-spezifizierte Merkmale stehen derzeit Analysen dieser Art noch aus. In fünf Teilstudien zeigte sich, dass hoch-inferente Beobachterratings in einem nicht unerheblichen Ausmaß durch Rater-Bias verzerrt sind. In ihrer Untersuchung waren etwa 20 Prozent der Varianz in den Unterrichtsqualitätsdimensionen *Klassenführung* und *Schülerorientierung* auf Rater-Effekte zurückzuführen. Auch stellte sich heraus, dass Unterrichtseinschätzungen geschulter Beobachter nicht wesentlich reliabler und valider waren als Einschätzungen ungeschulter Beobachter. Darüber hinaus waren Unterrichtsratings geschulter Beobachter durch situationale Merkmale beeinflusst: Die Begründungen der Rater für ihre Urteile über die Messzeitpunkte hinweg schwankten zum Teil deutlich. Auch deckte Praetorius Probleme im Urteilsprozess der Rater auf: Hier zeigten sich insbesondere Unstimmigkeiten bezüglich der Iteminterpretation sowie der Indikatorensuche und -auswahl; die Rater gewichteten Indikatoren unterschiedlich und es gab Probleme bei der Wahl der Antwortkategorien. Was die Stabilität der Merkmale der Unterrichtsqualität über Unterrichtsstunden hinweg betrifft, so zeigten sich deutliche Unterschiede im Hinblick auf die untersuchten Dimensionen, sodass in der Folge unterschiedlich viele Stunden pro Lehrkraft nötig sind, um das betreffende Merkmal reliabel zu erfassen. Praetorius zufolge lassen sich die Dimensionen *Klassenführung* und *Schülerorientierung* schon anhand einer Unterrichtsstunde mit einer Reliabilität von .70 erfassen, während für die Dimension *kognitive Aktivierung* neun Unterrichtsstunden für eine entsprechende Messgenauigkeit nötig sind (vgl. ebd., 274f.).

1.4 Gütekriterien für den Einsatz von Ratings

Selbst angesichts der vielfältigen, hier lediglich skizzierten Problemfelder, die Beobachterratings mit sich bringen, ist ihr Einsatz in der empirischen Unterrichtsforschung derzeit alternativlos. Anvisiert werden sollte jedoch eine Erhöhung von Validität (Gültigkeit) und Reliabilität (Zuverlässigkeit).³ In Bezug auf Ersteres

³ Argumente für Ratings im Allgemeinen lassen sich u.a. bei Saal, Downey und Lahey (1980) und bei Berliner (2005) finden.

sollte geprüft werden, ob mit der Beobachtung tatsächlich das infrage stehende Merkmal gemessen wird. Hier geht es also um eine Übereinstimmung zwischen dem Merkmal, das gemessen werden soll, und dem tatsächlich gemessenen Merkmal. So könnte beispielsweise eine fehlerhafte Vorstellung eines Konzeptes die Richtigkeit der Beurteilungen beeinträchtigen.

In Bezug auf Letzteres lässt sich feststellen, dass Urteile umso zuverlässiger (reliabler) sind, je mehr verschiedene Rater des gleichen Stundenausschnittes in ihrem Urteil über ein Unterrichtsmerkmal übereinstimmen und je mehr die Urteile zwischen verschiedenen Sachverhalten variieren. Die Höhe der Reliabilität hängt zum einen vom zugrunde liegenden Instrument und zum anderen von der Urteilskompetenz der Rater ab. Je mehr Ermessensspielraum die Beschreibung eines Merkmals erlaubt, desto größer ist die Gefahr, dass die Urteile der Rater streuen, d.h. mehr oder weniger weit voneinander abweichen. Insbesondere bei hoch-inferenten Merkmalen, die eine Wertung voraussetzen, ist dies eine Quelle für Differenzen zwischen den Urteilern.

So finden sich in der Fachliteratur einige Empfehlungen zur Qualitätssicherung bei Ratingverfahren. Empirische Überprüfungen ihres Nutzens mit Hinweisen zur adäquaten Ausgestaltung stellen derzeit jedoch noch ein Desiderat dar.

1.4.1 Operationalisierung der Merkmale

Hinter den Kritikpunkten am Rating als Beobachtungsverfahren steht die Forderung, dass das Vorliegen von Sachverhalten objektiv, d.h. mit intersubjektiver Übereinstimmung, feststellbar sein muss. Es gibt in der Tat Möglichkeiten der Vereinheitlichung der Wahrnehmung und des Eindrucksurteils und damit der Erhöhung der intersubjektiven Übereinstimmung sowie der Übereinstimmung mit der Merkmalsdefinition. Diese sollen durch die Wahl eines sog. *konzeptorientierten Ratings* ausgeschöpft werden, das sich wesentlich von dem sog. *intuitiven Rating* unterscheidet. Bei Letzterem ist es dem Rater überlassen, welches Merkmalsverständnis er seinem Urteil zugrunde legt; d.h., er ist nicht weisungsgebunden, was er unter den einzuschätzenden Merkmalen zu verstehen hat.

Bei der ausgereifteren Form des konzeptorientierten Ratings hingegen wird das Merkmal, das auf Grund theoretischer Überlegungen und/oder empirischer Vorforschungen für wichtig oder zielgemäß erachtet wird, definiert, präzise beschrieben und in seinen einzelnen Abstufungen durch Erläuterungen und prägnante Anker-Beispiele operationalisiert. Diese sollen den angehenden Rater schulen, ihn auf ein festgelegtes Merkmalsverständnis verpflichten und seine Wahrnehmung normieren. Darüber hinaus wird das Merkmal dadurch öffentlich und die Beurteilung nachvollziehbar. Grundlegend ist hierbei eine adäquate Operationalisierung der Merkmale durch geeignete Indikatoren beobachtbaren (Schüler- und/oder Lehrer-)Verhaltens. Sie werden i.d.R. in Ratingmanualen expliziert und dokumentiert.

1.4.2 Rater-Training und Ratingmanual

Die Objektivität ist ein wesentliches Gütekriterium, das beschreibt, dass Datenerhebung, -analyse und -auswertung unabhängig von der Person des Durchführenden sind. Ist die Vorgehensweise also standardisiert, ist das Erhebungsinstrument weniger anfällig für Verzerrungen durch verschiedene Beobachter (Moosbrugger/Kelava 2007, 8ff.). Dies kann durch Rater-Training und den Einsatz eines Ratingmanuals erfolgen.

Die Wiederholbarkeit durch andere sowie die Anwendung der Ergebnisse machen den Einsatz von mehreren Ratern erforderlich. Wie bereits erwähnt, sind die Rater beim konzeptorientierten Rating weisungsgebunden, was sie unter dem jeweiligen Merkmal zu verstehen haben. Ein solchermaßen festgelegtes Merkmalsverständnis, eine entsprechend konzeptgetreue Wahrnehmung und ein konzeptgetreues Verhalten bei der Skalenhandhabung können bei externen Ratern nicht als gegeben vorausgesetzt werden; vielmehr müssen sie in einem entsprechend wirksamen Lehrgang kommuniziert und erläutert werden (vgl. Langer/Schulz von Thun 2007, 126). Ratingmanuals mit Indikatoren für einzelne Items, Ankerbeispielen, Erläuterungen der Skalenabstufungen ggf. mit Extrembeispielen u.Ä., die den Ratern bereitgestellt werden, dienen in diesem Zusammenhang als Anleitung und Hilfestellung (vgl. z.B. Lotz et al. 2013a).

Dieses – dem Ratingvorgang vorgeschaltete – Rater-Training verfolgt neben der Qualifizierung der Rater weitere Ziele: Zum einen sollen Effekte unterschiedlichen Hintergrundwissens der Beurteiler minimiert werden, sodass sie in der Lage sind, übereinstimmend zu urteilen; zum anderen sollen sie befähigt werden, übereinstimmend richtig – d.h. konzeptgemäß – zu urteilen. Während Ersteres im Hinblick auf die Reliabilität der Messwerte von Bedeutung ist, spielt Letzteres insbesondere im Hinblick auf die (interne) Validität eine Rolle. Es wäre nämlich durchaus möglich, dass die Rater ein gemeinsames Merkmalsverständnis entwickeln, das jedoch an der Konzeptdefinition vorbeigeht (Helmke 2012 nennt dies „kollektiven Irrtum“). Es wäre jedoch unrealistisch zu erwarten, dass infolgedessen die Inter-Rater-Streuung komplett verschwindet. Darüber hinaus weist Praetorius (2014, 279f.) zu Recht darauf hin, dass der Einsatz von Rater-Trainings und Ratingmanuals nicht automatisch die Qualität von Ratings erhöhe. Mitunter sei auch denkbar, dass diese Hilfsmittel die Rater verunsichern und daher nur unzureichend zur intendierten Professionalisierung beitragen. Über die geeignete Dauer und Konzeption von Rater-Trainings und die ideale Gestaltung von Manualen gibt es derzeit noch wenige empirische Befunde.

1.4.3 Wahl der Rater

Die Rater sind die ausführenden Personen des Ratings; ihre Eindrücke sind ergebnisbestimmend, weshalb bei ihrer Auswahl Sorgfalt geboten ist. Es gilt daher im Vorfeld einer Untersuchung, Rater-Merkmale zu bestimmen, wonach die Rater ausgewählt werden (z.B. Motivation, Gewissenhaftigkeit, deklaratives und prozedurales Wissen in Bezug auf den einzuschätzenden Gegenstand).

Wie Studien aus der Expertiseforschung zeigen, sind mehrere tausend Stunden intensiver Auseinandersetzung mit einem Beurteilungsgegenstand nötig, um die erwünschte Expertise aufzubauen. Ein Untersucher kann sich i.d.R. jedoch keine übergebührlich lange Ausbildung seiner Rater erlauben. Damit ein kurzes spezifisches Rater-Training für sie ausreicht, um der Aufgabe gerecht zu werden, sollten solche Personen gewählt werden, die hinreichende wahrnehmungsmäßige Vorausstattung besitzen und längere Erfahrungen mit dem Objektbereich haben, aus dem die untersuchten Objekte stammen.

Idealerweise sollen die Rater nach Möglichkeit das Kriterium der Repräsentativität erfüllen und – aus einer Gruppe von geeigneten Personen (s.1.4.4) – zufällig ausgewählt werden. Sie sollen neutral sein, um Motive für absichtliche Verfälschungen zu vermeiden. In Anbetracht dessen, dass ein Rating mit vorangestelltem Training eine zeit- und arbeitsintensive Unternehmung ist, für die man mitunter auf freiwillige Teilnehmerinnen und Teilnehmer angewiesen ist, muss hier jedoch oft ein Kompromiss zwischen Geeignetheit und Verfügbarkeit eingegangen werden.

Bezüglich der Anzahl der benötigten Rater nennen Langer und Schulz von Thun (2007, 60) einen Erfahrungswert von mindestens fünf, da mit zunehmender Raterzahl die Güte der Einschätzungen i.d.R. steigt.

1.4.4 Qualitätsmerkmale für Ratings

Das verbreitetste Qualitätsmerkmal für Ratings in der empirischen Unterrichtsforschung ist die Inter-Rater-Übereinstimmung (vgl. Wirtz/Caspar 2002). Beim Einsatz mehrerer Rater werden deren Protokolle nach der Beobachtung verglichen, um das Ausmaß an Subjektivität von Beobachtungen zu kontrollieren. Diese rein mathematische Qualitätsbestimmung über die zentrale Tendenz der Rater einschätzungen kann jedoch kritisch hinterfragt werden (wie z.B. von Praetorius 2014 mit Verweis auf Murphy/DeShon 2000 und Hill et al. 2012). Ein Nachteil dieser Methode ist beispielsweise der, dass die Übereinstimmungen zwar geteilte Wahrnehmungen widerspiegeln, einen kollektiven Irrtum der Beobachter allerdings nicht ausschließen können. Praetorius (2014) empfiehlt daher, zusätzlich zur Inter-Rater-Reliabilität auch prozessorientierte Kriterien wie z.B. Tests zu deklarativem Wissen und zur Art und Weise der Informationsverarbeitung der Rater heranzuziehen. Lotz et al. (2013b) prüften beispielsweise, ob sich hoch- und niedrig-inferent gewonnene Beobachtungsdaten gegenseitig validieren lassen. Allerdings stehen derzeit vergleichende Studien zu unterschiedlichen Kriterien zur Bestimmung der Ratingqualität noch aus, sodass keine Angaben darüber gemacht werden können, welche Kriterien adäquat und ggf. der Inter-Rater-Reliabilität an Aussagekraft und Genauigkeit überlegen sind (vgl. Praetorius 2014, 277f.).

1.5 Literaturempfehlungen

Langer/Schulz von Thun (2007): Messung komplexer Merkmale in Psychologie und Pädagogik: Ratingverfahren.

Dieses in der Reprintreihe erschienene Buch gilt als *das* Standardwerk zu Ratingverfahren. Es behandelt auf gut verständliche Weise zahlreiche Themen und lässt kaum eine Frage unbeantwortet.

Lotz et al. (2013b): Technischer Bericht zu den PERLE-Videostudien.

In diesem technischen Bericht sind Ratinginventare dokumentiert und expliziert, die zur Veranschaulichung dienen: z.B. zu Klassenführung und Unterrichtsklima in drei Fächern und zur Qualität angeleiteter Bilderbuchrezeption.

Praetorius (2014): Messung von Unterrichtsqualität durch Ratings.

Diese empirische Studie prüft als eine der wenigen im deutschsprachigen Raum, ob das Ratingverfahren zur Messung von Unterrichtsqualität geeignet ist. Hier werden Fehlerquellen aufgedeckt sowie Schlussfolgerungen für qualitativ hochwertige Ratings abgeleitet.

2. Das Rating am Beispiel der Studie *VERA – Gute Unterrichtspraxis*

In diesem Kapitel sollen die vorangehenden Ausführungen anhand eines lesedidaktisch spezifizierten Ratings veranschaulicht werden, das im Rahmen der Studie *VERA – Gute Unterrichtspraxis* entstand. Es handelt sich hierbei um die erste Videostudie zur Lesekompetenz, die in deutschen Grundschulklassen durchgeführt wurde (vgl. Kleinbub 2010).

2.1 Ziele und Forschungsfragestellungen

Ziele der Videostudie *VERA – GU* waren die Identifikation von Faktoren, die guten Unterricht in der Grundschule ausmachen, die Konstruktion von Instrumenten und die Nutzung von Methoden, mit denen sie sich erfassen lassen, sowie das Ableiten von Maßnahmen, mittels derer Unterricht verbessert werden kann. Mit dem Einsatz von Videographie und Beobachterring sollte ein umfassendes Bild vom Ist-Zustand und somit den Merkmalen des aktuell praktizierten Leseunterrichts dokumentiert werden. Der Vergleich der real ablaufenden Unterrichtsprozesse mit dem Bild nahezu „idealen“ Unterrichts bot die Chance, Stärken zu erkennen und Schwachstellen aufzudecken, um somit eine Grundlage für die Ableitung von Maßnahmen zur Optimierung zu liefern. Ziel war es, einen Beitrag zur Unterrichtsentwicklung zu leisten und Impulse für deutschdidaktische und pädagogisch-psychologische Unterrichtsforschung zum Lesen zu geben.

Aufgrund der geringen Erkenntnisse, die die Deutschdidaktik über Leseunterricht hat, wurden im Rahmen des Projekts *VERA – GU* drei weite Fragenkomplexe ausgewiesen. Zum einen stellte sich mit dem Ziel der Deskription des Leseunterrichts die Frage, welche Lernangebote im Leseunterricht der Grundschule gemacht werden. Zum zweiten erfolgte neben der „Inventarisierung“ auch eine Einschätzung der Angemessenheit dieser Angebote mit dem Ziel der Evaluation. Hierzu wurde der Unterricht hinsichtlich verschiedener Merkmale der Prozessqualität geratet (*Kompetenzorientierung, Aktivierung, Adaptivität, Konkretisierung, Motivierung, Strukturierung*). Zum dritten wurde mit dem Ziel der Identifikation leistungsrelevanter Unterrichtscharakteristika untersucht, welche dieser Unterrichtsmerkmale Zusammenhänge mit hoher Leseleistung der Schülerinnen und Schüler aufweisen.

2.2 Untersuchungsplan und Stichprobe

Das Projekt dockte im September/Oktober 2005 mit einer Teilstichprobe an die übliche *VERA*-Kompetenzmessung an und verwendete die dort eingesetzten Lesetests als ersten Messzeitpunkt. Im Juli 2006 wurde in derselben Stichprobe eine weitere Lernstandstestung mit *VERA*-Instrumenten durchgeführt (Messzeitpunkt 3). Die Leseleistungen der Schülerinnen und Schüler wurden demnach zu Beginn und zum Ende der vierten Klasse erhoben. Im Zeitraum dazwischen von Mai bis Juni 2006 wurde der Unterricht u.a. im Fach Deutsch videographiert (Messzeitpunkt 2). Zur Analyse des Leseunterrichts wurden schließlich Videographien von

41 Klassen und 862 Schülerinnen und Schülern aus 37 Schulen aus Rheinland-Pfalz genutzt. Pro Klasse standen ein bis zwei Stunden Leseunterricht für die Analyse zur Verfügung.

2.3 Gründe für die Wahl eines Ratings bei *VERA – GU*

Mit der Wahl des Untersuchungsbereichs „Leseunterricht“ wurde ein komplexes Feld betreten. Um dem, was in Lehr-Lern-Prozessen passiert, gerecht zu werden und fachdidaktischen wie methodischen Kriterien bei der Unterrichtsbeobachtung Genüge zu leisten, bedurfte es eines angemessenen und effizienten Instruments. Es musste vor allem auf fachspezifischen Untersuchungsergebnissen basieren, eine fundierte theoretische Grundlage besitzen und darüber hinaus in der Lage sein, Prozesse in der Praxis abzubilden. Des Weiteren sollte es das „Was“ ebenso wie das „Wie“ des Unterrichts berücksichtigen, strukturiert sein und kognitive ebenso wie affektive Unterrichtsangebote erfassen. Zudem sollte es bestimmten Minimalanforderungen an die methodische Qualität genügen. Bisher existierten zwar einige Analyseinstrumente, die bei der Untersuchung von Unterrichtspraxis auf large-scale Basis eingesetzt werden konnten, allerdings orientierten sie sich meist an generischen Unterrichtsmerkmalen und ließen eine fachdidaktische Spezifizierung weitgehend vermissen. Daher musste für die Studie eigens ein Erhebungsinstrument zur Analyse der Unterrichtsvideos konstruiert werden.

Das Ratingverfahren bot sich zur Datenerhebung für die Beschreibung der Lernangebote und für deren Evaluation im Hinblick auf verschiedene Qualitätsmerkmale an. Insbesondere der zweite Aspekt – die Beurteilung der Qualität des Unterrichts – verlangte nach komplexen Einschätzungen, die nicht mit niedrig-inferenten Verfahren gemessen werden konnten. Daher wurde ein umfangreiches, fachdidaktisch orientiertes Ratinginventar entwickelt, das eine Vielzahl textanalytischer, hermeneutisch-interpretierender und produktionsorientierter Angebote abbildet. Neben der Deskription von Unterrichtsangeboten ermöglicht das konzipierte Ratinginventar auch die Evaluation hinsichtlich verschiedener Qualitätskriterien. Darüber hinaus können auf Grund der Strukturierung verschiedene Analyseebenen (einzelne Unterrichtsangebote, inhaltliche und methodische (Teil-) Bereiche, Unterrichtsphasen, gesamte Unterrichtsstunde) gewählt werden. Mit Hilfe des Ratinginstruments gelang es, die Unterrichtsstunden detailliert zu beschreiben und vergleichbar zu machen.

Da die Videostudie der angewandten Feldforschung zuzuordnen ist, dürfte die externe Validität der internen Validität überlegen sein. Empirische Studien im Schnittfeld von Pädagogischer Psychologie und Deutschdidaktik, die reale Unterrichtsprozesse zum Gegenstand haben, sind allerdings (noch) eher selten, weswegen Einschränkungen der internen Validität in Kauf genommen und nicht als ausreichender Grund angesehen wurden, die Durchführung der Untersuchung zu verwerfen.

2.4 Konstruktion des Ratinginventars

Die Konstruktion des Inventars orientiert sich an den „Grundkategorien zur Reflexion und Gestaltung des Unterrichts“ (Jank/Meyer 2005, 51). Das Inventar verfügt über eine differenzierte Architektur, die Auswertungen zur Deskription und Evaluation einzelner Unterrichtsangebote, verschiedener Unterrichtsphasen sowie der gesamten Unterrichtsstunde ermöglichte.⁴

Da die Videostudie u.a. eine Deskription des Leseunterrichts zum Ziel hatte, die wiederum Grundlage für weitere Auswertungen war, wurde ein umfangreicher Katalog an Lernangeboten, wie sie im Leseunterricht der vierten Klasse möglich sind, abgebildet. Aufgenommen wurden potenziell *wünschenswerte* Angebote, d.h., es erfolgte eine Positivauswahl auf der Grundlage empirischer und deutschdidaktischer Empfehlungen. Diese hat keinen Anspruch auf Vollständigkeit (kein Instrument könnte dies leisten), bildet jedoch einen großen Teil möglicher textanalytischer, hermeneutisch-interpretierender und produktionsorientierter Unterrichtsangebote ab. Diese wurden dann zur Itemformulierung genutzt. In der Tabelle sind beispielhaft Items aufgelistet, die mögliche Aufgaben zum Lesetext abbilden:

<i>Deskription</i>	<i>Evaluation</i>
D1. Die Lehrkraft regt das Detailverständnis an, indem sie die Lerner auffordert, explizite Informationen im Text zu lokalisieren (<i>scannen</i>).	E 1. Kompetenzorientierung
	E 2. Adaptivität
	E 3. Konkretisierung
	E 4. Aktivierung
D2. Die Lehrkraft regt das Detailverständnis an, indem sie die Lerner auffordert, implizite Informationen zu ermitteln (<i>spotting</i>).	E 5. Kompetenzorientierung
	E 6. Adaptivität
	E 7. Konkretisierung
	E 8. Aktivierung

Tab. 1: Beispielitems zu Aufgaben zum Lesetext

Den deskriptiven Items (in der Tab. D1 und D2) waren jeweils evaluative Items zugeordnet (in der Tab. E1 bis E8). Stellten die Rater bei der Deskription fest, dass ein Lernangebot in einer Unterrichtsstunde gemacht wurde (z.B. fordert die Lehrperson tatsächlich die Lerner auf, explizite Informationen im Text zu lokalisieren), konnte dieses Angebot im evaluativen Teil hinsichtlich der zuvor festgelegten und beschriebenen Qualitätsmerkmale ebenfalls eingeschätzt werden. Wurde ein Angebot im Unterricht jedoch nicht dargeboten, waren auch keine qualitativen Urteile möglich.

Neben Items zu didaktischen Inhalten (s. Tab. 1) wurden auch Items zu Medienauswahl und -einsatz, zum Umgang mit Heterogenität und zur Unterstützung

⁴ Die Konstruktion und Architektur des Instruments sowie die theoretischen Hintergründe können in diesem Artikel nicht detailliert dargelegt werden (vgl. hierzu Kleinbub 2010).

durch die Lehrperson formuliert. Insgesamt entstand so ein lesedidaktisch fokussiertes, konzeptgeleitetes Ratinginstrument bestehend aus 80 deskriptiven und 270 evaluativen Items, die auf einer 4-stufigen Likert-Skala eingeschätzt wurden. Auf dieser Grundlage beobachteten Experten-Rater den videographierten Lesunterricht und nahmen komplexe, hoch-inferente Einschätzungen vor. Ein Beispiel aus dem Ratingmanual soll die Skalenabstufungen bei der Evaluation erläutern (z.B. *Kompetenzorientierung: trifft zu – trifft eher zu – trifft eher nicht zu – trifft nicht zu*):

Um einzuschätzen, ob eine Lernaufgabe zu einem Lesetext dazu dienen kann, die Lesekompetenz der Schülerinnen und Schüler zu fördern, musste ein Rater folgendes bedenken:

Zur Einschätzung der Kompetenzorientierung einer Lernaufgabe muss beurteilt werden, ob sie deutlich darauf abzielt, das Textverständnis der Lerner anzuregen. Hierbei steht nicht die theoretische Aufgabe an sich im Mittelpunkt, sondern ihre Einbettung in den Reflexionsprozess und die Tiefe ihrer Behandlung im Rahmen der Textarbeit. Demnach kann eine Aufgabe wie *Male ein Bild der Hauptperson* eine kreative Beschäftigung mit oder ohne Einfluss auf das Textverständnis der Lerner sein, je nachdem, ob sie die Textarbeit intensiviert und die mehrmalige Lektüre erforderlich macht, indem beispielsweise verlangt wird, dass der Zeichnung die im Text genannten Informationen zugrunde liegen müssen. Kompetenzorientierung *trifft zu* wird dann als adäquates Schätzurteil angesehen, wenn deutlich wird, dass eine intensive Auseinandersetzung mit dem Text (z.B. durch mehrfache Lektüre, Anwenden von Lesestrategien, Austausch mit Mitschülerinnen und Mitschülern, Notieren von Informationen) zur Lösung der Aufgabe erforderlich ist und dies von der Lehrkraft eingefordert wird. *Trifft eher zu* beschreibt ein Unterrichtsgeschehen, bei dem Abstriche bei o.g. Szenario zu machen sind, d.h. bei dem die Aufgabe eine weniger intensive Auseinandersetzung mit dem Text fordert. Bei dem Urteil *trifft eher nicht zu* bearbeiten die Lerner die Aufgabe mit wenig Bezug zum Text. Dies wird von der Lehrkraft auch nur punktuell eingefordert. *Trifft nicht zu* wird dementsprechend als adäquat angesehen, wenn eine Aufgabe ohne jegliche Auseinandersetzung mit dem Text bearbeitet wird und die Texterschließung zur Lösung der Aufgabe nicht nötig ist, sodass die Konstruktion einer umfassenden Textbedeutung nicht gefördert werden kann.

Der folgende Transkriptauszug zeigt eine Aufgabenstellung, bei der die Schülerinnen und Schüler aufgefordert werden, explizite Informationen im Text zu finden. Sie werden angehalten, ihre Ergebnisse zu nennen und mit Textstellen und Schlüsselwörtern zu belegen. Die Rater wählten hier beispielsweise für das Merkmal *Kompetenzorientierung* die Skalenstufe *trifft zu*:

<i>Sprecher</i>	<i>Text</i>
L	So, wir fangen an mit der Gruppe eins: „Böse Taten des Bischofs Hatto.“ Womit könnten wir denn anfangen? Was unterstreichen wir? Sagt auch bitte, in welcher Zeile das ist, damit die anderen das auch finden können. – Ines (S6).
S6	In der ersten...in der ersten Zeile, hm(über) der zweite Satz: „Er war reich, aber geizig und spendete lieber den Segen als Almosen zu verteilen.“(o)
L(S7)	Mhm(ja). Welches Wort ist da besonders wichtig? Wo man sehen kann, dass es 'n böser Mensch ist? – Fabienne (S7).
S7	Geizig.
L(S7)	Gut, das nehmen wir schon mal.
L	Geizig.
C	(L schreibt „geizig“ an die Tafel.)
L	Wie geht's weiter mit dem Bischof? Was hat er noch alles auf'm Kerbholz? Auch die, die jetzt nicht in der Gruppe sind, die gucken mit...in welcher Zeile das ist. – Sabine (S9).
S9	Zeile 6: „Hatto beschimpfte sie.“ (o)

Transkriptauszug: Explizite Informationen suchen, hohe Kompetenzorientierung

Im Hinblick auf die Forschungsfragen war eine Strukturierung des Ratinginventars angebracht. Da die in der didaktischen Literatur diskutierten Phasenschemata zahlreich sind, wenig Konsens besteht und Bezüge zum Thema *Lesen* unbefriedigend sind, wurde eine Strukturierung, die sich am institutionalisierten Leseprozess orientiert, bevorzugt. Hierzu wurden die Phasen *Rezeption (Vorbereitung auf die Lektüre, Lektüre)* und *Anschlusskommunikation (Reflexion, Produktion, Präsentation)* unterschieden. Dies hatte auch den Vorteil, dass eine Vielzahl inhaltlich heterogener Unterrichtsstunden mit der Phasenstruktur des Instruments nachvollzogen werden konnte. Zudem gelang es so, das komplexe Instrument überschaubarer und damit auch handlicher zu gestalten. Aufgrund der so geschaffenen Architektur des Instruments konnten verschiedene Auswertungsebenen angelegt werden.

Je nach Urteilsbereich war es angebracht, unterschiedlich große Beurteilungsgegenstände zu wählen. So wurde jedes Unterrichtsangebot hinsichtlich der Merkmale *Kompetenzorientierung, Adaptivität, Konkretisierung* und *Aktivierung* beurteilt. Bei den Merkmalen *Motivierung* und *Strukturierung* hingegen schien es angebracht, die Beurteilungsgegenstände weiter zu fassen. Somit wurde für eine gesamte Unterrichtsphase ein Urteil zur *Motivierung*, für die gesamte Unterrichtsstunde ein Urteil zur *Strukturierung* getroffen.

2.5 Angaben zur Güte des Ratings

Den Anforderungen an die Objektivität der Unterrichtsbeobachtung wurde durch die im Benutzerleitfaden dokumentierten Instruktionen entsprochen. Um die Validität zu erhöhen und die Gefahr fehlerhafter Vorstellungen eines Konzeptes zu minimieren, wurden darüber hinaus verschiedene Maßnahmen getroffen: Besonders Augenmerk wurde auf die Wahl der Rater gelegt. Da es sich um ein deutschdidaktisch orientiertes Ratinginventar handelte, dessen konzeptgetreue Anwendung lesedidaktischer Expertise bedurfte, waren Wahrnehmungsfähigkeit und Erfahrung mit dem Objektbereich unabdingbar. Neben der Autorin konnten daher vier Personen gewonnen werden, die Experten auf dem Gebiet der Unterrichtsbeobachtung waren. Bei den Ratern handelte es sich um Fachleiterinnen und Fachleiter aus einem Staatlichen Studienseminar sowie um eine Schulleiterin einer Grundschule. Alle fünf Personen verfügten über mehrjährige Erfahrungen als Fachlehrkräfte und als Ausbilder von Lehramtsanwärtern im Fach Deutsch. Unterrichtsbeobachtung, -analyse und -reflexion im Rahmen von Unterrichtsbesuchen, Lehrproben und dienstlichen Überprüfungen stellten daher den Kern ihrer beruflichen Tätigkeit dar. Sie standen dem Untersuchungsgegenstand neutral gegenüber und waren für mögliche systematische Urteilsfehler sensibilisiert. Auch waren sie mit Ergebnissen der Unterrichtsforschung bestens vertraut. Die Verfügbarkeit ergab sich infolge persönlicher Kontakte. Das Kriterium der Repräsentativität wurde nicht uneingeschränkt erfüllt, da keine zufällige Auswahl aus einer beliebig großen Gruppe erfolgen konnte.

Des Weiteren wurde die Form des konzeptorientierten Ratings gewählt, bei dem die Rater an ein festgelegtes Merkmalsverständnis gebunden waren. Um zu verhindern, dass die Rater ein gemeinsames Merkmalsverständnis entwickelten, das an der Konzeptdefinition vorbeigeht, fand ein Rater-Training statt. Aufgrund des großen Umfangs des Ratinginventars wurde den Ratern vor Beginn des Trainings das Manual zur Lektüre zur Verfügung gestellt.

Zu Beginn der Trainingssitzung wurden die Rater in ihre Arbeit eingewiesen. Hierzu wurde das Forschungsprojekt detailliert erläutert und die Forschungsfragenstellungen dargelegt. Da die Rater mit deutschdidaktischen Konzepten wie dem Lesekompetenzbegriff und Determinanten der Lesekompetenz vertraut waren, wurden theoretische Grundpositionen lediglich kurz in Erinnerung gerufen. Es folgte eine eingehende Diskussion der Abstufungen der Merkmale in den Bereichen der Deskription und der Evaluation jeweils mit Beispielen. Darüber hinaus wurde die Handhabung des Ratinginventars vorgeführt.

Im Anschluss an die Instruktion folgten Diskriminations- und Konzept-Unterscheidungsübungen anhand von Filmmaterial, das ebenfalls im Rahmen von *VERA – GU* videographiert, allerdings nicht in die Hauptuntersuchung aufgenommen worden war. Das Trainingsprinzip bei beiden Übungsformen bestand in der wiederholten Herstellung einer Verknüpfung von beobachtetem Verhalten mit dem konzeptgemäßen Skalenwert durch die übenden Rater. Das Diskriminations-training diente dem Erkennen von Ausprägungsunterschieden eines Merkmals.

Hierzu wurde den Ratern eine Unterrichtssequenz vorgespielt, in der sie zunächst das Merkmal identifizieren und einer Ausprägung auf der Skala zuordnen sollten. Beispielsweise mussten sie das Unterrichtsangebot *Die Lehrkraft regt das Detailverständnis an, indem sie die Lerner auffordert, implizite Informationen zu ermitteln (spotting)* in einer Filmsequenz identifizieren und den Grad der geschätzten Kompetenzorientierung auf der Skala (*trifft zu – trifft eher zu – trifft eher nicht zu – trifft nicht zu*) einordnen. Beim Konzept-Unterscheidungstraining wurden die Unterrichtssequenzen hinsichtlich mehrerer Merkmale beurteilt. Dies stellte die Anforderung an die Rater, dieselben Objekte nacheinander unter verschiedenen Betrachtungsgesichtspunkten wahrzunehmen. So wurde beispielsweise geübt, ein Unterrichtsangebot sowohl deskriptiv als auch evaluativ zu beurteilen, beide Urteile jedoch unabhängig voneinander zu treffen. Des Weiteren wurde das Unterrichtsangebot im Bereich der Evaluation unter verschiedenen Gesichtspunkten – wie der *Kompetenzorientierung, Adaptivität, etc.* – beurteilt.

Während dieses Trainings schätzten die Teilnehmerinnen und Teilnehmer dasselbe Datenmaterial ein, verglichen und diskutierten ihre Ergebnisse, um bei der Anwendung der Konzepte Konsens zu erreichen. In diesem Zusammenhang wurde auch die Gefahr systematischer Urteilsfehler thematisiert. Die Rater wurden darüber hinaus ermutigt, die gesamte Stufenweite entsprechend ihrer Eindrücke zu nutzen.

Zur Prüfung der Reliabilität wurde bei zehn Unterrichtsvideographien mit jeweils unterschiedlichen Lehrpersonen und Schulklassen die Inter-Rater-Übereinstimmung (fünf Rater bearbeiteten dieselben zehn Filme) anhand des Kennwerts Krippendorffs Alpha (α) bestimmt. Das Ergebnis zeigte insgesamt zufriedenstellende Übereinstimmungen zwischen .46 und 1.00.

2.6 Resümee der Arbeit mit einem Rating bei *VERA – GU*

Abschließend bleibt zu diskutieren, wie angemessen das eingesetzte Ratinginstrument in der Lage war, die Unterrichtsangebote zu erfassen.⁵

Die eigene Konstruktion des Ratinginventars erlaubte eine lesedidaktische Fokussierung im Hinblick auf die Forschungsfragestellungen. Anhand der Strukturierung des Untersuchungsinstruments wurde zudem eine Datenerhebung auf mehreren Analyseebenen möglich. Die erhobenen Messdaten konnten mittels diverser statistischer Verfahren ausgewertet werden.

⁵ Einige methodische Diskussionspunkte, die in der Studie *VERA – GU* auftraten, erwachsen primär aus der Tatsache, dass es sich um eine Videostudie handelte (wie z.B. Fragen nach Reaktivitätseffekten, nach der Aussagekraft einer Unterrichtsstunde einer Lehrperson für Unterrichtsqualität im Allgemeinen, nach der Anzahl und Terminierung der Messzeitpunkte etc.; s. auch Maak in diesem Band). Diese Aspekte werden in diesem Artikel vernachlässigt zugunsten von Themen, die eher das Ratingverfahren als Erhebungsmethode betreffen – wenngleich sie in einem engen Zusammenhang gesehen werden sollten.

Die Unterscheidung nach deskriptiven und evaluativen Items ermöglichte eine differenzierte Betrachtung der Unterrichtsstunden. Zur Deskription der Unterrichtsangebote bewährte sich der ausführliche und umfangreiche Item-Katalog. Die Evaluation jedes einzelnen Unterrichtsangebots war auf Grund der geringen inhaltlichen Vergleichbarkeit der Unterrichtsstunden (den Lehrpersonen stand die Gestaltung der Lesestunden völlig frei) jedoch wenig ergiebig. Aussagekräftiger war hier die Charakterisierung von Unterrichtsphasen. Geht es um die Beurteilung von Merkmalen der Unterrichtsqualität, so empfiehlt sich bei nicht-standardisierten Unterrichtsstunden zur Sequenzierung wohl eher die Wahl einer größeren Beurteilungseinheit wie z.B. Zeitintervalle oder eher fachdidaktisch motivierte Einheiten wie Phaseneinteilungen.

Der Befund, dass sich die Phasen einer Unterrichtsstunde durch unterschiedliche Ausprägungen der Merkmale charakterisieren ließen, machte zudem deutlich, dass eine Sequenzierung der Unterrichtsstunden und eine detaillierte Untersuchung mehrerer Qualitätsmerkmale einen Erkenntnisgewinn darstellen. Es ist anzunehmen, dass weder die Beurteilung nur eines Qualitätsmerkmals noch die ausschnittshafte Begutachtung einer einzigen kurzen Teilsequenz des Unterrichts angemessene Schlussfolgerungen erlauben. Globalurteile, bei denen für eine gesamte Unterrichtsstunde nur eine Einschätzung getroffen wird, können für bestimmte Merkmale (wie z.B. *Strukturierung der Stunde*) sinnvoll sein; für andere Merkmale können sie jedoch einen Erkenntnisverlust bedeuten, da sich hohe und niedrige Ausprägungen von Unterrichtsmerkmalen in verschiedenen Phasen gegenseitig ausgleichen können. Vor diesem Hintergrund wird es schwierig, exzellente Stunden zu identifizieren, die in allen Phasen und über alle Merkmale hinweg Höchstwerte erzielen müssten, um sich sehr deutlich von einem Durchschnitt abzuheben. Im umgekehrten Fall gilt dies auch für schlechte Stunden, die demzufolge in allen Phasen und allen Merkmalen nahezu Minimalwerte erzielen müssten.

Besonders hilfreich war der Einsatz von Experten-Ratern, der es ermöglichte, ein effektives und zeitökonomisches Rater-Training und einen detaillierten Ratingvorgang durchzuführen. Die Untersuchung profitierte sowohl von der Vertrautheit der Beurteiler mit fachdidaktischen Konzepten als auch von deren Expertise bei der Unterrichtsbeobachtung. Die detaillierten Beschreibungen der Beurteilungskategorien im Ratingmanual machten eine konzeptgetreue Handhabung des Instruments möglich und trugen ihrerseits zu hohen Beurteilerübereinstimmungen bei. Aus finanziellen und zeitlichen Gründen konnten jedoch nur 10 der 41 Unterrichtssequenzen von mehreren Ratern beurteilt werden. Auch war die Auswahl der Rater nicht zufällig.

Zwar wurde versucht, durch die Wahl von Experten-Ratern beste Bedingungen zu schaffen, dennoch sind bei hoch-inferenten Schätzurteilen Beobachtungsfehler, beispielsweise durch Halo-Effekte u.ä., nicht auszuschließen. Zudem kann ein Merkmal wie *Motivierung*, das Schülerinnen und Schüler als Zielgruppe hat, auch von Experten nur aus einer Fremdperspektive beurteilt werden. Für ein valides

Gesamturteil wäre es daher von Vorteil, über unterschiedliche Perspektiven zu verfügen.

Des Weiteren lag der Ratingskala die theoretische Annahme zugrunde, dass die höchste Ausprägung der Merkmale wünschenswert sei. Im pädagogischen Kontext ist höchstmögliche *Strukturierung*, *Motivierung* etc. auf Grund einer möglichen differenziellen Wirksamkeit jedoch nicht immer die bestmögliche Ausprägung, vielmehr gilt das vage Maß der „Angemessenheit“.

2.7 Fazit

Aufgrund der Komplexität des Forschungsgegenstands muss sich die empirische Deutschdidaktik einer Vielzahl von Herausforderungen stellen. Hierzu zählt neben der Beantwortung inhaltlicher Fragen, dass sie sich empirischer Methoden bedient, die über ihren eigenen Geltungsbereich hinaus nachvollziehbar sind. Forschungsinstrumente wie Ratings müssen demnach die Standards empirischer Forschung erfüllen und fach- ggf. sogar bereichsspezifisch sein, um die Besonderheiten des jeweiligen Gegenstands zu erfassen. Von zentraler Bedeutung ist es daher, eine Kompatibilität von Forschungsfragen, -gegenständen und -methoden herzustellen und dabei Leistungen und Grenzen zu reflektieren.

Literatur

- Berliner, David C.: The near impossibility of testing for teacher quality. In: Journal of Teacher Education, (2005), 56, 205-213.
- Boelmann, Jan M.: Literarisches Verstehen mit narrativen Computerspielen: Eine empirische Studie zu den Potenzialen der Vermittlung von literarischer Bildung und literarischer Kompetenz mit einem schüleraffinen Medium. München: kopaed 2015.
- Bortz, Jürgen/Döring, Nicola: Forschungsmethoden und Evaluation. Für Human- und Sozialwissenschaftler. Heidelberg: Springer 2006.
- Clausen, Marten/Reusser, Kurt/Klieme, Eckhard: Unterrichtsqualität auf der Basis hochinferenter Unterrichtsbeurteilungen: Ein Vergleich zwischen Deutschland und der deutschsprachigen Schweiz. In: Unterrichtswissenschaft (2003), 31 (2), 122-141.
- Clausen, Marten: Unterrichtsqualität: eine Frage der Perspektive? Empirische Analysen zur Übereinstimmung, Konstrukt- und Kriteriumsvalidität. Münster: Waxmann 2002.
- Helmke, Andreas: Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts. Seelze-Velber: Klett Kallmeyer 2012.
- Hill, Heather C./Charalambous, Charalambos Y./Kraft, Matthew A.: When rater reliability is not enough: Teacher observation systems and a case for the generalizability study. In: Educational Researcher (2012), 41, 56-64.
- Jank, Werner/Meyer, Hilbert: Didaktische Modelle. Berlin: Cornelsen Scriptor 2005.
- Kleinbub, Iris: Unterrichtsqualität im Leseunterricht. Eine videobasierte Analyse in vier Klassen. Trier: Wissenschaftlicher Verlag 2010.

- Kunter, Mareike/Baumert, Jürgen: Who ist he expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research* (2006), 9, 231-251.
- Kürzinger, Anja/Pohlmann-Rother, Sanna: Kriterienkatalog Textkorpus. Ein Instrument zur Bestimmung von Textqualität in Klasse 1. Bamberg: University of Bamberg Press 2014.
- Langer, Inghard/Schulz von Thun, Friedemann: Messung komplexer Merkmale in Psychologie und Pädagogik: Ratingverfahren. Münster: Waxmann 2007.
- Lotz, Miriam/Gabriel, Katrin/Lipowsky, Frank: Niedrig- und hoch-inferente Verfahren der Unterrichtsbeobachtung – Analysen zu deren gegenseitiger Validierung. In: *Zeitschrift für Pädagogik* (2013a), 50 (3), 357-380.
- Lotz, Miriam/Lipowsky, Frank/Faust, Gabriele (Hrsg.): Technischer Bericht zu den PERLE-Videostudien. Materialien zur Bildungsforschung, Band 23/3, Frankfurt a.M.: GFPF 2013b. http://www.pedocs.de/volltexte/2013/7702/pdf/Mat-Bild_Bd23_3.pdf (letzter Zugriff: 01.08.2018).
- Moosbrugger, Helfried/Kelava, Augustin (Hrsg.): Testtheorie und Fragebogenkonstruktion. Berlin Heidelberg: Springer 2007.
- Murphy, Kevin R./DeShon, Richard: Interrater correlations do not estimate the reliability of job performance ratings. In: *Personnel Psychology* (2000), 53, 873-900.
- Praetorius, Anna-Katharina: Messung von Unterrichtsqualität durch Ratings. Münster: Waxmann 2014.
- Saal, Frank, E./Downey, Ronald G./Lahey, Mary Anne: Rating the ratings: Assessing the psychometric quality of rating data. In: *Psychological Bulletin* (1980), 88, 413-428.
- Wirtz, Markus/Caspar, Franz: Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen. Göttingen: Hogrefe 2002.

Grammatische Textanalyse

Ein korpusanalytisches, quantitatives Untersuchungsverfahren

1. Vorstellung der grammatischen Textanalyse

Die Auswertung (text-)grammatischer Strukturen in Lernertexten hat in der Erst-, Zweit- und Fremdsprachdidaktik/-erwerbs-Forschung eine eigenständige Tradition und baut auf text- und korpuslinguistischen Methoden auf. Ausgehend von einer linguistischen Theorie zu einem grammatischen Erwerbs- oder Entwicklungsgegenstand oder ausgehend von einem sprachdidaktischen Modell von Schreibkompetenz wird ein auszuwertender Merkmalskatalog erstellt, der auf ein zuvor erhobenes oder ein bereits verfügbares Textkorpus bestehend aus Lernertexten angewendet wird. Die Merkmale werden annotiert, ausgezählt und je nach Fragestellung deren Vorkommenshäufigkeit in Abhängigkeit von individuellen Voraussetzungen (Alter, Erstsprache) der Probandinnen und Probanden verglichen.

So wird a) die Analyse von Lernerkorpora vor allem in der Zweit- und Fremdspracherwerbsforschung genutzt, um den Spracherwerbsstand bzw. besondere Schwierigkeiten des Zweit- oder Fremdspracherwerbs zu identifizieren (vgl. Granger 2008, 259ff.). Mit b) dem im Folgenden im Vordergrund stehenden Aufbau von Schreibkompetenz beschäftigt sich die empirische Schreib(entwicklungs)forschung, innerhalb der der (textsortenspezifische) Verlauf von Schreibentwicklung durch die Auswertung von textgrammatischen Merkmalen (z.B. Augst et al. 2007) oder auch Schreibprozesse, d.h. das Zusammenspiel von Textplanung-, -formulierung und -überarbeitung, nachgezeichnet werden (Weinzierl/Wrobel 2017).¹

1.1 Funktionsweise der grammatischen Textanalyse

Die empirische Schreibentwicklungsforschung gilt als gut entwickelter Forschungszeitweig (vgl. Pohl 2014, 101), in welchem eine Reihe verschiedener Auswertungsverfahren entwickelt wurden. Die korpusbasierten Auswertungsverfahren

¹ Dieser Ansatz steht in enger Beziehung zu dem aus der empirischen Sozialforschung stammenden Quantitativen Inhaltsanalyse (s. Uhl in diesem Band).

ren von (text-)grammatischen Merkmalen in Schülertexten hier unter der Bezeichnung *grammatische Textanalyse* zusammengefasst werden. Diese Verfahren folgen der Annahme, dass Textprodukte als Indikatoren von Schreib- bzw. Sprachkompetenz von Individuen genutzt werden können (vgl. Steinhoff 2017, 354). In der deutschsprachigen Forschung beginnt eine solche produktbasierte Auswertung grammatischer Merkmale von Lernertexten in den 70er und 80er Jahren mit Studien wie Messelken (1971) und Augst/Faigel (1986). Eines der bekanntesten Analysemodelle stellt das Züricher Textanalyseraster nach Nussbaumer (1996) dar, das versucht umfassende und textsortenübergreifende Einblicke in die Sprachfähigkeiten von Schülerinnen und Schülern zu ermöglichen, ohne jedoch konkrete grammatische Merkmale als Analyse Kriterien zu bestimmen.

Die Erstellung eines konkreten (text-)grammatischen Analysemodells geht von einer funktionalen, operationalisierbaren Definition eines zu untersuchenden grammatischen oder textlinguistischen Gegenstandes aus, dessen verschiedene Realisierungen Erwerbs- bzw. Entwicklungsschritte zeigen. Dabei ist die textsortenspezifische Einbettung des Gegenstands von besonderer Bedeutung, da verschiedene Textsorten sich durch unterschiedliche grammatische Merkmale auszeichnen und unterschiedliche Teilkompetenzen zur Produktion voraussetzen.² Ein Beispiel für einen funktionalen Ausgangspunkt für die Erstellung eines textsortenspezifischen Analysemodells ist das Konzept der Textprozeduren nach Feilke (2012a). Mit dem Terminus *Textprozeduren* bezeichnet Feilke textsortentypische grammatische Konstruktionen für textsortentypische Teilhandlungen, z.B. die Positionierung als textsortentypische Handlung des Argumentierens.

Positionierungsprozeduren

Ich finde...

Ich bin (da)für/(da)gegen...

Ich bin der Meinung, dass...

Ich meine, dass...

Ich halte + NP+ für + Adj.

Meiner Meinung nach...

Nominale

Meines Erachtens...

Konstruktionen

M.E. ...

Gätje/Rezat/Steinhoff (2012) untersuchen formale Realisierungen von Positionierungsprozeduren an einem Korpus von schriftlichen Argumentationen von Schülerinnen und Schülern der 1. bis 12. Jahrgangsstufe und vergleichen die Auftretenshäufigkeit der unterschiedlichen grammatischen Varianten. Dabei stellen sie eine Zunahme nominaler Konstruktionen (vgl. Abb. 1) mit höherem Jahrgang

Abb. 3 Positionierungsprozeduren nach Gätje/Rezat/Steinhoff (2012, 143)

² Schülertexte lassen sich jedoch nicht problemlos in textlinguistische Textsortenklassifikationen einordnen, da Schülertexte, die in Lehr-Lern-Kontexten entstanden sind, vor allem zu Bewertungszwecken produziert werden und weniger aus einem echten Kommunikationsanlass. Feilke (2012b, 168) bezeichnet die typischen schulischen Textsorten ‚Erörterung‘, ‚Inhaltsangabe‘ etc. als textdidaktische Gattungen, deren Kennzeichen die „Fiktivität der Kommunikationssituation“ sei.

fest, die sie als Indiz für konzeptionell schriftlichen Sprachgebrauch bzw. die Zunahme literaler Formen interpretieren (vgl. ebd., 143f.).

Unter dem Stichwort „Literalität“ (i.S. von Maas 2010) oder „konzeptioneller Schriftlichkeit“ (i.S. von Koch/Österreicher 1985) können grammatische Merkmale von Texten erfasst werden, die den Zugang von Schülerinnen und Schülern zur textuellen Handlungsfähigkeit in der Modalität der Schriftlichkeit unter Berücksichtigung der zerdehnten Kommunikationssituation anzeigen können. Auch auf Basis dieser theoretischen Grundlage können Analysemodelle für Texte entwickelt werden. Ein Beispiel stellt die Studie von Siekmeyer (2013) dar, die verschiedene Attributtypen (Partizipialattribut, Relativsatzattribut etc.) in Bezug auf ihre Literalität diskutiert und deren Gebrauch in mündlichen und schriftlichen Schülererzählungen vergleicht. Häufig wird auch als grammatischer Indikator für konzeptionelle Schriftlichkeit syntaktische Integration angeführt (vgl. z.B. Augst/Faigel 1986), wie sie auch für das Modell von Junktion definiert werden kann, das im späteren Beispielprojekt als Auswertungsmodell vorgestellt wird.

Korpusanalytische Verfahren werden auch in der Zweit- und Fremdspracherwerbsforschung angewandt, die die systematische Auswertung von Lernerkorpora nach dem Auftreten von bzw. spezifischen Gebrauchsformen bestimmter grammatischer Strukturen zur Bestimmung von Erwerbsstadien oder Erwerbswegen nutzen (vgl. z.B. Dollnick 2013; Petersen 2014; Binanzer 2017).

1.2 Ausgangsdaten und Ergebnisse

Die Auswertung der (text-)grammatischen Merkmale erfolgt an einem Korpus von Schülertexten, das zuvor erhoben wird, sofern man nicht auf bestehende Korpora zurückgreifen möchte, bzw. wenn bestehende Korpora die Untersuchungsziele nicht ausreichend abdecken können. Der Vorteil der Erhebung eines Korpus ist, dass man die Eigenschaften des Korpus entsprechend der gewählten Fragestellung anpassen kann (z.B. die Auswahl der Klassenstufen oder von Lernenden bestimmter Herkunftssprachen). Vor der eigentlichen Erhebung sind auf Basis der Fragestellung eine Reihe von Vorentscheidungen bezüglich der Auswahl und Größe der Stichprobe an Schülerinnen und Schülern zu treffen, die Texte für das Korpus produzieren sollen. Das Vorgehen zur Auswahl der untersuchten Stichprobe sowie die Größe entscheiden über die Repräsentativität der Untersuchungsergebnisse zu der betreffenden Grundgesamtheit (vgl. Eckey/Kosfeld/Dreger 2000, 15). Diese Phase einer empirischen Untersuchung bezeichnet Becker-Mrotzek (1997, 98) als Phase der Datenplanung. Sie stellt eine ebenso wichtige Phase dar, wie die anschließende Datenerhebung, da nur mit einer sorgfältig geplanten Erhebung auch valide Ergebnisse entstehen.

Zusätzlich zu den Schreibprodukten sind personenbezogene Merkmale zu erheben. Dazu können neben Alter, Geschlecht, Schulart, Klassenstufe, Muttersprache auch durch Fragebögen ergänzende Informationen zum Leseverhalten oder zur Mediennutzung erhoben werden, um diese anschließend mit den sprachlichen Daten zu korrelieren. Diese Merkmale sollten im Korpus zu möglichst gleichen Anteilen vertreten sein, um repräsentative Ergebnisse zu erzielen.

Die Auswirkungen der Schreibsituation und Schreibaufgabe auf die Schreibmotivation und den Sprachgebrauch beschreiben u.a. Ossner (1996), Bachmann (2002) und Becker-Mrotzek/Bachmann (2010) sowie auch Binanzer (i.d.B.). Die Schreibaufgabe sollte eine kommunikative Funktion und den Adressat bzw. die Adressaten der Texte für die Probandinnen und Probanden erkennbar machen (vgl. Becker-Mrotzek/Bachmann 2010, 195). Zudem sollte sichergestellt werden, dass alle Schülerinnen und Schüler über die gleichen Wissensvoraussetzungen zu dem gewählten Thema der Texte verfügen. Die Schreibaufgabe indiziert außerdem die Textsorte der entstehenden Texte. Die Auswahl der Aufgabenstellung kann auch in Anlehnung an bestehenden Untersuchungen erfolgen, die sich in der Forschung bereits als bewährt erwiesen haben. Verschiedene Aufgabenstellungen für Argumentation, Bericht, Erzählung, Instruktion und Beschreibung haben z.B. Augst et al. (2007) für ihre Studie zum Vergleich der „Text-Sorten-Kompetenz“ entwickelt.

Nach der Datenerhebung folgt die Phase der Datenaufbereitung. Zuerst sollten die Texte in digitaler Form verfügbar gemacht werden, d.h. die handschriftlichen Texte digitalisiert und anschließend nach den gewählten grammatischen Analyse-kriterien annotiert werden.³ Die Annotation kann mit einem Annotationsprogramm wie EXMARaLDA vorgenommen werden oder auch in SPSS. Da SPSS ausschließlich Daten in Form von Zahlen auswerten kann, sind alle Analyse-kategorien zu kodieren. Hierzu wird jeder Kategorie einer Kategorisierung bzw. jeder Merkmalsausprägung der zu untersuchenden Merkmale ein beliebiger Wert zu gewiesen. Es handelt sich bei den Werten um nominalskalierte Daten, sodass der Abstand zwischen den Werten keine Bedeutung hat.

Bei der quantitativen Auswertung und dem Vergleich der Häufigkeit des untersuchten grammatischen Phänomens ist zu berücksichtigen, dass die Textlänge das Ergebnis in hohem Maße beeinflusst (vgl. Langlotz 2014, 75ff.): Schülerinnen und Schüler in höherem Jahrgang schreiben längere Texte als Schülerinnen und Schüler der niedrigen Jahrgänge, Mädchen schreiben signifikant längere Texte als Jungen. Um die Häufigkeit eines Phänomens unabhängig von diesen Einflussfaktoren auszuwerten, ist es möglich, die Auftretenshäufigkeit auf eine einheitliche Textlänge von beispielsweise 1000 Wörtern hochzurechnen (vgl. Augst/Faigel 1986) oder die relationale Verteilung der Merkmalsausprägungen zu vergleichen (vgl. Langlotz 2014). Die ausgewerteten Häufigkeiten der Merkmalsausprägungen können dann mit den personenbezogenen Daten korreliert werden, um festzustellen, ob ein signifikanter Zusammenhang zwischen den gewählten Variablen besteht, d.h. ein bestimmtes grammatisches Merkmal nicht zufällig bei Personen mit bestimmten Eigenschaften häufiger auftritt. Weitere statistische Auswertungsmöglichkeiten werden hier nicht näher erläutert, einen guten Überblick bieten Eckey/Kosfeld/Dreger (2000) sowie Bühl/Zöfel (2005).

³ *Annotieren* bezeichnet die Anreicherung des Textes mit linguistischen Informationen (vgl. Bubenhofer 2006-2015).

Die Ergebnisse liegen im ersten Schritt der Auswertung als Zahlenwerte vor (z.B. Häufigkeitsverteilungen und Mittelwerte), die im Rahmen der Forschungsliteratur interpretiert werden können. Anhand der Daten werden die aufgestellten Hypothesen verifiziert (bestätigt) oder falsifiziert (widerlegt), auch die Falsifikation ist ein wichtiges Ergebnis für die Forschung – ebenso das Nicht-Vorkommen von erwarteten Phänomenen, das bei der Interpretation der Ergebnisse berücksichtigt werden sollte.

1.3 Vorteile und Nachteile der grammatischen Textanalyse unter Berücksichtigung der Gütekriterien

Das hier vorgeschlagene Vorgehen, Schülertexte nach funktionalen grammatischen Auswertungskategorien zu analysieren, bietet gegenüber einer automatischen, rein formalen Auswertung den Vorteil, dass alle Realisierungsformen eines Phänomens erfasst werden können, auch wenn es sich um nicht normgerechte bzw. um kreative Varianten handelt. Gerade diese nicht vorhersehbaren Elemente in Schülertexten können Hinweise auf den Entwicklungsverlauf geben. Gätje/Rezat/Steinhoff (2012, 145) bezeichnen solche Varianten als Lernerformen (Beispiel: *Meiner Meinung nach finde ich...*). Zudem ist zu berücksichtigen, dass häufig orthografisch abweichende Varianten in Texten von Lernenden zu finden sind, wenn z.B. der Gebrauch von *weil* ausschließlich über eine automatische Suchfunktion ausgewertet wird, werden damit abweichende Varianten wie (*weill, wail, wall*) nicht erfasst.

Der Nachteil bei der Auswertung von Texten durch eine einzige Person ist die mangelnde Auswertungsobjektivität, die neben der Durchführungsobjektivität und der Interpretationsobjektivität als wesentliches Gütekriterium zu berücksichtigen ist. Das Problem kann jedoch durch eine möglichst genaue Definition des linguistischen Gegenstandes vermieden werden: Je genauer der Untersuchungsgegenstand definiert ist, desto eher kommen verschiedene auswertende Personen zu dem gleichen Ergebnis. Für ein grammatisches Phänomen gibt es meistens verschiedene Definitionen. Beispielsweise kann ein Nebensatz als untergeordneter Teilsatz definiert werden, aber auch als durch die Verbletzstellung und den Subjunktor als Einleitewort gekennzeichnet (vgl. Fabricius-Hansen 1992). Abhängige Hauptsätze, wie sie in Formen der direkten Redewiedergabe vorliegen können, fallen zwar unter die erste, jedoch nicht unter die zweite Definition. Man käme je nach angewandter Definition zu unterschiedlichen Ergebnissen darüber, wie viele Nebensätze in einem Schülertext auftreten. Die Auszählung von Nebensätzen wird in Studien der empirischen Bildungsforschung häufig als Kriterium eingesetzt (z.B. bei Kürzinger et al. 2013; Steinig 2009). Die quantitative Auswertung von Sprache birgt immer den Nachteil, dass man sich von dem eigentlichen Phänomen in seinem funktionalen Kontext entfernt. Zudem hat ein Text als Ganzes eine bestimmte ästhetische oder kommunikative Wirkung, die bei der Zerlegung in Zahlen verloren geht. Die Quantität eines grammatischen Merkmales sagt nichts über dessen Qualität aus, deshalb sind die untersuchten Formen immer in Bezug zur Funktionalität und Konventionalität zu setzen. Um

diesen Nachteil auszugleichen, ist eine Kombination quantitativer und qualitativer Verfahren ratsam, vor allem die Ergänzung der qualitativen Analyse einzelner Texte (vgl. Dannerer 2012, 65ff.). Dies trägt zudem zu einer höheren Interpretationsobjektivität der sprachlichen Daten bei, da sie durch die qualitative Analyse rückgebunden an den realen Gebrauchskontext begründet werden können.

Eng zusammen mit der Frage nach der Interpretationsobjektivität hängt die Frage nach der Konstruktvalidität und der Frage danach, was das Vorhandensein bestimmter sprachlicher Konstruktionen in einem Schülertext über die bestimmten kognitiven Fähigkeiten des Lernenden aussagt. Die Studie von Becker-Mrotzek et al. (2014) nimmt genau diese Frage in den Fokus und weist einen sehr hohen Zusammenhang von sprachlichen und kognitiven Fähigkeiten empirisch nach.

1.4 Literaturempfehlungen

Becker-Mrotzek/Grabowski/Steinhoff (Hrsg.) (2017): Forschungshandbuch empirische Schreibdidaktik.

Das Handbuch bietet einen Überblick über verschiedene Arbeitsgebiete schreibdidaktischer Forschung und enthält ebenfalls methodische Artikel. Der Beitrag von Steinhoff fokussiert ebenfalls die Auswertung von Textkorpora.

Untersuchungen, die das hier beschriebenen Auswertungsvorgehen anwenden, sind beispielsweise:

Augst et al. (2007): Text – Sorten – Kompetenz. Eine echte Longitudinalstudie zur Entwicklung der Textkompetenz im Grundschulalter.

In dieser Studie wird ein echtes Longitudinalkorpus systematisch auf die Entwicklung der Fähigkeit zur textsortenspezifischen Gestaltung der typischen schulischen Textsorten Beschreibung, Bericht, Erzählung, Instruktion und Argumentation gegeben.

Binanzer (2017): Genus – Kongruenz und Klassifikation. Evidenzen aus dem Zweitspracherwerb.

Diese Studie stellt ein Beispiel zum Erwerb einer grammatischen Kategorie im Zweitspracherwerb dar. Hier werden schriftliche Erzählungen ein- und mehrsprachiger Kinder in Bezug auf den Gebrauch genuscodierender Ausdrücke ausgewertet.

Siekmeyer (2013): Sprachlicher Ausbau in gesprochenen und geschriebenen Texten.

Siekmeyer entwickelt ein grammatisches Analysemodell der verwendeten Attributformen, um die modalitätsspezifische Gestaltung von mündlichen und schriftlichen Erzählungen zu evaluieren.

2. Beispielprojekt „Junktion und Schreibentwicklung“

Am Beispiel meines Dissertationsprojektes „Junktion und Schreibentwicklung“ möchte ich nun eine mögliche Umsetzung der beschriebenen Vorgehensweise erläutern (zum Folgenden vgl. Langlotz 2014). Der Ausgangspunkt des Projektes war die Frage, wie Schülerinnen und Schüler unterschiedlicher Klassenstufe der Sekundarstufe I grammatischer Mittel der Aussagenverknüpfung (Junktoren) in Erzählungen und Argumentationen nutzen, um dabei einerseits die Fähigkeit der textsortenspezifischen Gestaltung narrativer und argumentativer Texte festzustellen und um andererseits zu prüfen, ob über den Nachweis einer zunehmenden syntaktischen Integration auch die Zunahme von Literalität bzw. konzeptioneller Schriftlichkeit in den Texten nachgewiesen werden kann.

Die Definition des Analysegegenstandes *Junktion* ist an Raible (1992) angelehnt, der darunter die sprachliche Darstellung von semantischen Relationen zwischen Propositionen versteht. Hierbei handelt es sich um eine übereinzelsprachlich geltende Definition, im Deutschen werden in dieser Funktion v.a. Kon- und Subjunktoren (z.B. *denn* bzw. *weil*), bestimmte Adverbien (z.B. *deswegen*) oder Präpositionen (z.B. *wegen*) genutzt. An einer Junktion sind immer zwei Propositionen und ein Junktor beteiligt, der eine semantische Relation (z.B. kausal, adversativ etc.) zwischen den Propositionen herstellt (vgl. Ágel 2010, 905). Neben dem Terminus *Junktor* existieren eine Reihe von Termini, die auf ähnliche Konzepte referieren, z.B. *Konjunktionen* oder *Konnektoren*. Die Konzepte unterscheiden sich durch die definierten syntaktischen Merkmale.

Die Basis für das entwickelte Analysemodell stellt das Junktionsmodell von Ágel (2010) dar, der die Junktionstechniken des Deutschen auf einer Skala zwischen Aggregation und Integration systematisiert und eine Operationalisierung entwirft, die eine Auswertung von Junktion als intervallskaliertes Merkmal möglich macht. An Beispielen aus den von mir erhobenen Schülertextkorpus erläutere ich vereinfacht die wesentlichen Merkmale aggregativer und integrativer Junktionstechniken und diskutiere anschließend die Bedeutung von Junktion als Indikator für Schreibentwicklung.

In der folgenden Tabelle sind die wichtigsten Junktionstechniken nach steigender Integration angeordnet. Beispiel 3 zeigt somit die aggregativste Junktionstechnik:

Junktionstechnik	Beispielbeleg
I. Koordination	
Koordination durch Konjunktor:	Beispiel 3: 5106 ⁴ In der Grundschule haben sich die Hausaufgaben langsam gesteigert <u>und</u> jetzt muss ich auch noch abends im Bett lernen.
Koordination durch AP-Junktor:	Beispiel 4: 5222 Ich sitze manchmal sogar den ganzen Tag, bis zum Abendessen, an den Hausaufgaben. <u>Deshalb</u> habe ich keine Zeit, um mich zu verabreden.
Koordination durch Konjunktor + AP-Junktor:	Beispiel 5: 5101 Denn ich komme ungefähr um 13.40 Uhr nach Hause, ruhe mich ein bisschen aus <u>und dann</u> gehe ich essen.
II. Subordination	
Subordination durch Partizipialkonstruktion:	Beispiel 6: 5213 <u>Im Klassenzimmer angekommen</u> , legten wir unsere Ranzen an den Plätzen von uns ab.
Subordination durch Subjunktorsatz/ Verbzweitsatzeinbetter:	Beispiel 7: 7423 <u>Hätte</u> man aber mehr Zeit, könnte man sich das Lernen besser einteilen. Beispiel 8: 9301 <u>Angenommen</u> , ich komme donnerstags um 16 Uhr nach hause, dann bin ich doch relativ erschöpft.
Subordination durch Subjunktor:	Beispiel 9: 5106 Jetzt muss ich auch noch abends im Bett lernen, <u>weil</u> die Hausaufgaben einfach zu lange dauern.
Subordination durch Infinitivkonstruktion:	Beispiel 10: 9104 Außerdem sollte ebenfalls ein Limit gesetzt werden, <u>um</u> zu lange Aufgaben zu ersparen.
III. Inkorporation	
Inkorporation	Beispiel 11: 5106 Außerdem sollte die Schulzeit <u>durch</u> das Verkürzen oder sogar Abschaffen der Hausaufgaben nicht verlängert werden.

Tab. 1: Übersicht Junktionstechniken

Mit Aggregation bezeichnet Raible (1992) die maximale syntaktische Unabhängigkeit von zwei Propositionen, d.h. die beiden verknüpften Propositionen sind jeweils hauptsatzförmig realisiert, wie es bei den Formen der Koordination der Fall ist. Innerhalb der Techniken der Koordination ist die Koordination durch Konjunktor aggregativer als die durch Adverb-Partikel-Junktor (AP-Junktor), da der AP-Junktor in die syntaktische Struktur einer der Propositionen integriert wird. Maximale Integration liegt bei der Inkorporation vor. Mit Inkorporation bezeichnet Ágel (2010, 907) die Junktionstechnik, bei der eine der beiden Propositionen nominalisiert realisiert ist und durch den präpositionalen Junktor registriert wird. Auf der Skala der Integration zwischen Koordination und Inkorporation befinden sich die verschiedenen Subordinationstechniken, bei denen die integrierte

⁴ Bei dieser vierstelligen Ziffernfolge handelt es sich um den Schülercode, den ich jedem Schüler zugewiesen habe, um die Daten anonymisiert zu verarbeiten. Die erste Ziffer steht für den Jahrgang, die zweite für die Klasse, die ich innerhalb des Jahrgangs durchnummeriert habe. Die letzten beiden Ziffern entsprechen der Nummerierung der Schüler innerhalb einer Klasse, die ich willkürlich vorgenommen habe.

Proposition noch über ein (meist finites) verbales Element verfügt, jedoch syntaktisch subordiniert ist.

2.1 Junktion als Kompetenzindikator

Die Anwendung von Junktionsanalysen als Instrument der Schreibentwicklungsforschung setzt voraus, dass die unterschiedlichen syntaktischen, semantischen und lexikalischen Merkmale von Junktoren als Kompetenzindikatoren fungieren können. Dies scheint offensichtlich, da Junktoren ein Kohäsionsmittel sind und als Indikator für die Fähigkeit zur Produktion kohärenter Texte stehen können. Es gibt bereits Untersuchungen zur Schreibentwicklung, deren Fragestellung auf den Zusammenhang der Häufigkeit von Junktoren und der Kohärenzfähigkeit von Schülerinnen und Schülern abzielen (z.B. Bachmann 2002).

Ziel der vorgestellten Studie war nicht ein Vergleich der Häufigkeit des Auftretens, sondern die Merkmalsausprägungen der syntaktischen Integration, der semantischen Relationen und der lexikalischen Merkmale auf Kompetenz abzubilden. Zunehmende syntaktische Integration haben bereits Augst und Faigel (1986) als Kompetenzindikator interpretiert:

Erst mit zunehmendem Alter lernt das Kind den Satz als den Baustein und die Integration der Propositionen als die syntaktischen Hauptformen der Schriftlichkeit kennen und anwenden. (Augst/Faigel 1986, 105).

In Bezug auf die semantischen und lexikalischen Ausprägungen wird ein Mehr an verschiedenen semantischen Relationen und Junktionsausdrücken als Kompetenzmerkmal gesehen (z.B. ebd.). Das Auftreten bestimmter semantische Relationen, z.B. der Konzessivität, wird als Bestandteil der Fähigkeit zur Rezipientenorientierung interpretiert (vgl. Rezat 2011).

Verschiedene Junktionsausdrücke können außerdem als eher typisch mündlich oder typisch schriftsprachlich eingeordnet werden. So wird beispielsweise das häufige Auftreten von *und dann* von Schmidlin (1999, 189) und Augst et al. (2007, 51) als Mündlichkeitsmerkmal von Erzählungen eingestuft. Als weiteres Merkmal von konzeptioneller Mündlichkeit gilt das Auftreten von *weil* als Konjunktor (vgl. Hennig 2006, 142f.). Der kausale Konjunktor *denn* und der Subjunktor *da* werden in der gesprochenen Sprache kaum verwendet und gelten tendenziell als konzeptionell schriftlich (Zifonun 1997, 2304; Selting 1999, 170).

Diese Forschungsergebnisse bilden den Ausgangspunkt für die Hypothesen der hier dargestellten Untersuchung:

- Hypothese 1: Die syntaktische Integration der Junktion nimmt mit höherem Jahrgang zu.
- Hypothese 2: Es findet in höheren Jahrgangsstufen eine Ausdifferenzierung der semantischen Relationen und des Ausdrucks von Junktion statt. Es wird mit höherem Jahrgang eine größere Vielfalt an Types zum Ausdruck von Junktion verwendet.

- **Hypothese 3:** Es lassen sich Unterschiede in Bezug auf konzeptionelle Mündlichkeit und Schriftlichkeit anhand von Junktion nachweisen. Texte in höheren Jahrgängen zeigen mehr Merkmale konzeptioneller Schriftlichkeit, Texte in niedrigeren Jahrgängen mehr Merkmale konzeptioneller Mündlichkeit.
- **Hypothese 4:** Argumentation und Erzählung zeigen signifikante Unterschiede in der Entwicklung von Junktion.

2.2 Korpuserhebung

Ausgehend von diesen Hypothesen wurde bei der Erhebung des Korpus vor allem die gleichmäßige Verteilung der erhobenen Texte auf die verschiedenen Jahrgangsstufen beachtet. Für mein Korpus habe ich Texte aus drei verschiedenen Jahrgängen der Sekundarstufe I in gleichmäßigem Abstand zueinander erhoben, es haben jeweils drei Klassen des 5., 7. und 9. Jahrgangs eines Gymnasiums teilgenommen. Es handelt sich um eine Quasi-Longitudinalstudie, da alle Texte zu einem Zeitpunkt von verschiedenen Individuen produziert wurden. Für eine echte Longitudinalstudie hätten dieselben Individuen zu unterschiedlichen Zeitpunkten an der Erhebung teilnehmen müssen. In die Auswertung sind ca. 120 Texte pro Jahrgang eingegangen. Texte von Schülerinnen und Schülern, die nicht Deutsch als ihre Erstsprache angegeben haben, wurden nicht berücksichtigt. Das Korpus besteht aus 382 Schülertexten, die 202 verschiedene Schülerinnen und Schüler verfasst haben. Dazu hat (fast) jeder Lernende zwei Texte nach den folgenden Aufgabenstellungen verfasst:

Aufgabenstellung Argumentation

Professor Heuer von der Universität Kassel führt zurzeit ein Forschungsprojekt durch, mit dem er herausfinden möchte, ob Hausaufgaben sinnvoll sind oder besser abgeschafft werden sollten. Dafür interessiert ihn auch die Meinung der Schüler. Schreibe einen Brief an Professor Heuer, indem du ihm deine Meinung zu dem Thema Hausaufgaben darstellst.

Aufgabenstellung Erzählung

Deine Klasse verfasst gemeinsam ein Geschichtenbuch, für das jeder von euch eine Geschichte schreiben darf. Erfinde für das Geschichtenbuch eine lustige/spannende/gruselige Geschichte zu einem Ereignis, das in deiner Schule vorgefallen sein könnte.

Eine Argumentation pro und contra Hausaufgaben ist ein bewährter Schreibanlass aus Schreibeentwicklungsstudien (z.B. auch bei Augst/Faigel 1986). Die Aufgabenstellung enthält einen direkten Adressaten mit relativ hoher Distanz zu den Schülerinnen und Schülern und macht eine kommunikative Funktion der Texte erkennbar; die Schülerinnen und Schüler sollen ihre Position überzeugend darlegen. Auch mit der Aufgabenstellung für den narrativen Text soll durch das Entstehen eines Geschichtenbuchs eine Funktion der Texte für die Schülerinnen und Schüler erkennbar werden.

2.3 Ergebnisse

Ich möchte nun einige konkrete Ergebnisse der Auswertung a) der Junktionstechniken und b) des Junktionsausdrucks zeigen und anschließend die Interpretation der Zahlen in Bezug auf die aufgestellten Hypothesen erläutern.

a) Junktionstechniken und syntaktische Integration

Die folgende Tabelle zeigt die relationale Verteilung der Merkmalsausprägungen der Junktionstechniken innerhalb eines Jahrgangs. Alle Vorkommen von Junktion innerhalb eines Jahrgangs bilden 100%. In Klammern stehen die absoluten Zahlen, die unabhängig von der relationalen Verteilung im 9. Jahrgang am höchsten sind, da in diesem Jahrgang die längsten Texte und somit auch die meisten Junktoren produziert wurden.

Junktionstechniken	5. Jahrgang	7. Jahrgang	9. Jahrgang
<i>Koordination</i>			
Koordination durch Konjunktoren:	49,6% (527)	50,0% (684)	45,6% (801)
Koordination durch AP-Junktoren:	14,9% (158)	11,7% (160)	16,2% (285)
Koordination durch Konjunktoren + AP-Junktoren:	6,6% (70)	5,2% (71)	3,6% (5)
<i>Subordination</i>			
Subordination durch Partizipialkonstruktion:	0,2% (2)	0,1% (1)	0,3% (5)
Subordination durch Subjunktorersatz/Verbzweitsatzeinbettung:	0	0,4% (5)	0,6% (10)
Subordination durch Subjunktoren:	26,5% (282)	30,2% (412)	28,5% (501)
Subordination durch Infinitivkonstruktion:	2,0% (22)	1,7% (23)	3,5% (62)
<i>Inkorporation</i>	0,2% (2)	0,7% (9)	1,3% (23)
<i>Gesamt</i>	100% (1063)	100% (1365)	99,6% ⁵ (1692)

Tab. 2: Ergebnisse Junktionstechniken

Bei einem Blick auf die unteren Spalten der Tabelle zeigt sich die Hypothese 1 nur zum Teil als bestätigt: Im 9. Jahrgang haben sowohl die integrativste Junktionstechnik *Inkorporation* wie auch die Infinitivkonstruktion den höchsten Anteil. Im 7. Jahrgang haben jedoch Infinitivkonstruktionen einen geringeren Anteil als im 5. Jahrgang. Die aggregativste Junktionstechnik *Koordination durch Konjunktoren* hat zudem im 7. Jahrgang den höchsten Anteil. Die Entwicklung verläuft insgesamt nicht kontinuierlich.

b) Junktionsausdruck der kausalen Relation

Die folgende Tabelle zeigt, zu welchem Anteil welche Junktoren genutzt werden, um Kausalität innerhalb eines Jahrgangs auszudrücken. Die Ergebnisse sind nach Argumentation und Erzählung aufgeteilt und zeigen deutliche Unterschiede zwischen den beiden Textsorten.

⁵ Die zu 100% fehlenden Belege treten in der hier nicht vorgestellten Junktionsklasse Unifikation auf.

signifikanter Zusammenhang zwischen Jahrgang und höherer syntaktischer Integration nachgewiesen werden, die Effektstärke des Jahrgangs hat sich jedoch als sehr gering herausgestellt. Eine offene Fragestellung für anschließende Untersuchungen bleibt also, welche weiteren Einflussfaktoren Auswirkungen auf das Auftreten syntaktisch integrativer Formen bzw. den Sprachgebrauch der Schülerinnen und Schülern insgesamt haben.

Die Hypothesen 2, 3 und 4 können dagegen eher bestätigt werden: Der Junktionsausdruck der kausalen Relation zeigt beispielsweise die Entwicklungstendenz zum Gebrauch eher der konzeptionellen Schriftlichkeit entsprechenden Ausdrucksformen. Die syntaktische Integration kann somit nur bedingt als Merkmal konzeptioneller Schriftlichkeit gelten, da bestimmte Ausdrucksformen unabhängig von ihrer syntaktischen Integration als eher konzeptionell mündlich oder schriftlich eingestuft werden können. Ergänzend zeigen die qualitativen Einzelanalysen von Texten eine globale Strukturierung sowie textsortentypische Gestaltung in höherem Jahrgang:

Beispieltext 1: Argumentation 5. Jahrgang

5101 Ich bin für die Abschaffung der Hausaufgaben, weil ich lange an den Hausaufgaben sitze. Denn ich komme ungefähr um 13.40 Uhr nach Hause, ruhe mich ein bisschen aus und dann gehe ich essen. Um ungefähr 14.10 Uhr fang ich an mit den Hausaufgaben. Um ungefähr 15.30 Uhr bin ich mit den Hausaufgaben fertig und dann muss ich noch z.B. für eine Arbeit oder sonstiges üben und manchmal will ich mich verabreden und dann ist es schon zu spät.

Beispieltext 2: Argumentation 9. Jahrgang

9120 Es gibt natürlich negative und positive Aspekte in der Hinsicht auf Hausaufgaben. Positiv zu bemerken wäre, dass man den Schulstoff wiederholt und so besser lernt. Das wiederum hat zur Folge, dass man für Klassenarbeiten weniger lernen muss. Außerdem kommen die Lehrkräfte viel schneller mit dem Unterrichtsstoff durch, so lernt man in der Schule mehr, da man mehr Zeit hat. Besonders für Schüler der G8 Jahrgänge ist dies ein Segen. Negative Seiten der Hausaufgaben sind, z.B. die jeden Tag verlorene Zeit, da man ja am nächsten Tag wieder genau dieses Schulfach hat. Dies löst in den Schülern ein Stressgefühl aus und da Stress bewiesen schlecht zum Lernen ist, schlägt dieser Aspekt sehr zu Buche. Ich persönlich wäre für die Lösung eine Wochen-Hausaufgabe (2-Wochen-Hausaufgabe), da dies das Beste für die Lehrer sowie die Schüler wäre.

Der Schüler des 9. Jahrgangs kündigt die Struktur seines Textes im ersten Satz an und setzt diese anschließend um, indem er zunächst die positiven Aspekte und anschließend die negativen Aspekte aufzählt. Abschließend positioniert er sich mit einem Kompromissvorschlag. Auch der Schüler des 5. Jahrgangs beherrscht bereits textsortentypische Positionierungsprozedur „ich bin für“. Anschließend folgt eine Begründung der Position mit einem eher deskriptiv bzw. narrativ aufgebauten Textteil, der durch den häufigen *und dann*-Gebrauch gekennzeichnet ist. Insgesamt zeigen die Ergebnisse, dass die Bedeutung der Textsorte für das Auftreten von Kompetenzindikatoren nicht unterschätzt werden darf.

Die Studie zeigt, dass sich durch die Analyse grammatischer Merkmale in Schülertexten deutliche Unterschiede zwischen den verschiedenen Textsorten und

Jahrgängen nachweisen lassen. Anhand dieser Ergebnisse lassen sich Entwicklungsschritte der Schreibentwicklung nachzeichnen, eine solche empirische Erfassung von Entwicklungsschritten bildet eine sinnvolle Basis für didaktische Modelle eines kompetenzfördernden Schreibunterrichts.

Literatur

- Ágel, Vilmos: Explizite Junktion. Theorie und Operationalisierung. In: Ziegler, Arne/Braun, Christian (Hrsg.): Historische Textgrammatik und Historische Syntax des Deutschen. Traditionen, Innovationen, Perspektiven. Bd. 2: Frühneuhochdeutsch, Neuhochdeutsch. Berlin: De Gruyter 2010, 897-936.
- Augst, Gerhard/Disselhoff, Katrin/Henrich, Alexandra/Pohl, Thorsten/Völzing, Paul-Ludwig: Text – Sorten – Kompetenz. Eine echte Longitudinalstudie zur Entwicklung der Textkompetenz im Grundschulalter. Frankfurt a.M.: Peter Lang 2007.
- Augst, Gerhard/Faigel, Peter: Von der Reihung zur Gestaltung. Untersuchung zur Ontogenese der schriftsprachlichen Fähigkeiten von 13-23 Jahren. Frankfurt a.M.: Peter Lang 1986.
- Bachmann, Thomas: Kohäsion und Kohärenz: Indikatoren für Schreibentwicklung. Zum Aufbau kohärenzstiftender Strukturen in instruktiven Texten von Kindern und Jugendlichen. Innsbruck: Studienverlag 2002.
- Becker-Mrotzek, Michael: Schreibentwicklung und Textproduktion. Der Erwerb der Schreibfertigkeit am Beispiel der Bedienungsanleitung. Opladen: Westdeutscher Verlag 1997.
- Becker-Mrotzek, Michael et al.: Adressatenorientierung und Kohärenzherstellung im Text. Zum Zusammenhang kognitiver und schriftlich realisierter Teilkomponenten von Schreibkompetenz. In: Didaktik Deutsch (2014), 37, 21-43.
- Becker-Mrotzek, Michael/Bachmann, Thomas: Schreibaufgaben situieren und profilieren. In: Pohl, Thorsten/Steinhoff, Torsten (Hrsg.): Textformen als Lernformen. Duisburg: Gilles & Francke 2010, 191-210.
- Becker-Mrotzek, Michael/Grabowski, Joachim/Steinhoff, Torsten (Hrsg.): Handbuch empirische Schreibdidaktik. Münster: Waxmann 2017.
- Binanzer, Anja: Genus – Kongruenz und Klassifikation. Evidenzen aus dem Zweitspracherwerb des Deutschen. Berlin: De Gruyter 2017.
- Brinker, Klaus: Linguistische Textanalyse: Eine Einführung in Grundbegriffe und Methoden. Berlin: Erich Schmidt 2010.
- Bubenhofer, Noah: Einführung in die Korpuslinguistik: Praktische Grundlagen und Werkzeuge. (2006-2015) <http://www.bubenhofer.com/korpuslinguistik/>. (letzter Zugriff: 01.08.2018).
- Bubenhofer, Noah: Sprachgebrauchsmuster: Korpuslinguistik als Methode der Diskurs- und Kulturanalyse. Berlin: De Gruyter 2009.
- Bühl, Achim/Zöfel, Peter: SPSS 12. Einführung in die moderne Datenanalyse unter Windows. München: Pearson 2005.

- Dannerer, Monika: Narrative Fähigkeiten und Individualität. Mündlicher und schriftlicher Erzählerwerb im Längsschnitt von der 5. bis zur 12. Schulstufe. Tübingen: Stauffenburg 2012.
- Dollnick, Meral: Konnektoren in türkischen und deutschen Texten bilingualer Schüler. Eine vergleichende Langzeituntersuchung zur Entwicklung schriftsprachlicher Kompetenzen. Frankfurt a.M.: Lang 2013.
- Eckey, Hans-Friedrich/Kosfeld, Reinhold/Dreger, Christian: Statistik. Grundlagen – Methoden – Beispiele. Wiesbaden: Gabler 2000.
- Ehlich, Konrad: Zum Textbegriff. In: Rothkegel, Anneli/Sandig, Barbara (Hrsg.): Text – Textsorten – Semantik. Hamburg: Buske 1984, 9-25.
- Fabricius-Hansen, Cathrine: Subordination. In: Hoffmann, Ludger (Hrsg.): Deutsche Syntax: Ansichten und Aussichten. Berlin: De Gruyter 1992, 458-483.
- Feilke, Helmuth: Was sind Textroutinen? Zur Theorie und Methodik des Forschungsfeldes. In: Feilke, Helmuth/Lehnen, Katrin (Hrsg.): Schreib- und Textroutinen. Theorie, Erwerb und didaktisch-mediale Modellierung. Frankfurt a.M.: Peter Lang 2012a, 1-33.
- Feilke, Helmuth: Schulsprache – Wie Schule Sprache macht. In: Günthner, Susanne/Imo, Wolfgang/Meer, Dorothee/Schneider, Jan Georg (Hrsg.): Kommunikation und Öffentlichkeit. Sprachwissenschaftliche Potentiale zwischen Empirie und Norm. Berlin: De Gruyter 2012b, 149-175.
- Gätje, Olaf/Steinhoff, Torsten/Rezat, Sara: Positionierung. Zur Entwicklung des Gebrauchs modalisierender Prozeduren in argumentativen Texten von Schülern und Studenten. In: Feilke, Helmuth/Lehnen, Katrin (Hrsg.): Schreib- und Textroutinen. Frankfurt a.M.: Peter Lang 2012, 125-153.
- Glück, Helmut: Metzler-Lexikon Sprache. Metzler: Stuttgart 2005.
- Granger, Sylviane: Learner Corpora. In: Lüdeling, Anke/Kytö, Merja (Eds.): Corpus linguistics. HSK 29.1. Berlin: Mouton de Gruyter 2008, 259-275.
- Hennig, Mathilde: Grammatik der gesprochenen Sprache in Theorie und Praxis. Kassel: University Press 2006.
- Koch, Peter/Oesterreicher, Wulf: Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. In: Romanistisches Jahrbuch 36 (1985), 15-43.
- Kürzinger, Anja et al.: Auswertung der Lucybriefe: Perspektivenübernahme und Schreibkompetenz. In: Lotz, Miriam/Lipowsky, Frank/Faust, Gabriele (Hrsg.): Dokumentation der Erhebungsinstrumente des Projekts „Persönlichkeits- und Lernentwicklung von Grundschulkindern“ (PERLE). 3. Technischer Bericht zu den PERLE-Videostudien. Frankfurt a.M.: GFPP 2013, 255-296.
- Langlotz, Miriam: Junktion und Schreibentwicklung. Eine empirische Untersuchung narrativer und argumentativer Schülertexte. Berlin: De Gruyter 2014.
- Langlotz, Miriam: Zur Bedeutung von Junktoren als narrative und argumentative Textprozeduren in Schülertexten. In: Behrens, Ulrike/Gätje, Olaf (Hrsg.): Mündliches und schriftliches Handeln im Deutschunterricht. Wie Themen entfaltet werden. Frankfurt a.M.: Peter Lang 2016, 169-191.

- Maas, Utz: Orat und literat. Grundbegriffe der Analyse geschriebener und gesprochener Sprache. In: Grazer linguistische Studien 73 (2010), 21-150.
- Messelken, Hans: Empirische Sprachdidaktik. Heidelberg: Quelle und Meyer 1971.
- Nussbaumer, Markus: Lernerorientierte Textanalyse - Eine Hilfe zum Textverfassen? In: Feilke, Helmuth/Portmann, Paul (Hrsg.): Schreiben im Umbruch. Schreibforschung und schulisches Schreiben. Stuttgart: Klett 1996, 96-112.
- Ossner, Jakob: Gibt es Entwicklungsstufen beim Aufsatzschreiben? In: Feilke, Helmut/Portman, Paul (Hrsg.): Schreiben im Umbruch. Schreibforschung und schulisches Schreiben. Stuttgart: Klett 1996, 74-85.
- Pohl, Thorsten: Entwicklung der Schreibkompetenzen. In: Feilke, Helmuth/Pohl, Thorsten: Schriftlicher Sprachgebrauch – Texte verfassen. Baltmannsweiler: Schneider 2014, 101-141.
- Raible, Wolfgang: Junktion: Eine Dimension der Sprache und ihre Realisierungsformen zwischen Aggregation und Integration. Heidelberg: Winter 1992.
- Schmidlin, Regula: Wie Deutschschweizer Kinder schreiben und erzählen lernen. Textstruktur und Lexik von Kindertexten aus der Deutschschweiz und aus Deutschland. Tübingen/Basel: Francke 1999.
- Selting, Margret: Kontinuität und Wandel der Verbstellung von ahd. wanta bis gwd. weil. Zur historischen Syntax der weil-Konstruktionen. In: Zeitschrift für germanistische Linguistik (1999), 27, 167-204.
- Siekmeyer, Anne: Sprachlicher Ausbau in gesprochenen und geschriebenen Texten. Zum Gebrauch komplexer Nominalphrasen als Merkmale literater Strukturen bei Jugendlichen mit Deutsch als Erst- und Zweitsprache in verschiedenen Schulformen. Dissertation 2013. Verfügbar unter: <https://d-nb.info/1053681577/34> (letzter Zugriff 25.07.2018)
- Steinhoff, Torsten: Wissenschaftliche Textkompetenz. Sprachgebrauch und Schreibentwicklung in wissenschaftlichen Texten von Studenten und Experten. Berlin: De Gruyter 2007.
- Steinig, Wolfgang: Schreiben von Kindern im diachronen Vergleich: Texte von Viertklässlern aus den Jahren 1972 und 2002. Münster: Waxmann 2009.
- von Polenz, Peter: Deutsche Satzsemantik: Grundbegriffe des Zwischen-den-Zeilen-Lesens. Berlin: De Gruyter 2008.
- Weinzierl, Christian/Wrobel, Arne: Schreibprozesse untersuchen. In: Becker-Mrotzek, Michael/Grabowski, Joachim/Steinhoff, Torsten (Hrsg.): Handbuch empirische Schreibdidaktik. Münster: Waxmann 2017. 221-238.
- Zifonun, Gisela: Subordination. In: Zifonun, Gisela/Hoffmann, Ludger/Strecker, Bruno (Hrsg.): Grammatik der deutschen Sprache. Berlin: De Gruyter 1997, 2233-2358.

Interaktionsanalysen

Veranschaulichung des Analyseverfahrens anhand einer Studie zur Wirkung der globalen Textkohäsion auf das Textverständnis

1. Vorstellung der Interaktionsanalyse

Die multiple Regression ist ein Analyseverfahren, das sich für eine Prüfung von Wechselwirkungen, sogenannte Interaktionen, eignet (vgl. Saunders 1956). Ein Beispiel für eine Interaktion wäre, wenn die Wirkung eines Lesestrategietrainings auf das Textverständnis von der allgemeinen Lesefähigkeit der Rezipienten abhängt. Interaktionsanalysen stammen aus der Psychologie und den Sozialwissenschaften, aber auch für empirische Arbeiten in der deutschdidaktischen Forschung erweisen sie sich als erkenntnisgenerierendes Auswertungsverfahren.

1.1 Funktionsweise der multiplen Regression und der Interaktionsanalyse

Die multiple Regression wird grundsätzlich dazu verwendet, um Beziehungen zwischen einer abhängigen Variable (Kriterium – z.B. das Textverständnis) und mehr als zwei erklärenden Variablen (Prädiktoren – z.B. die Lesefähigkeit und das inhaltliche Vorwissen) in Form von Streuungen zu analysieren. Für eine zuverlässige Vorhersage, bei der alle vorhergesagten Werte auf einer Geraden liegen sollten, ist ein linearer Zusammenhang der abhängigen Variable mit den betrachteten unabhängigen Variablen bedeutsam. Ist eine solche Linearbeziehung gegeben, lässt sich mit der multiplen Regression für jede der erklärenden Variablen eine spezifische Gewichtung (der sogenannte Regressionskoeffizient) schätzen. Das Ergebnis der Regressionsanalyse besteht in einer Lineargleichung, womit die Werte einer abhängigen Variable durch Prädiktoren vorhergesagt werden. Es lässt sich beispielsweise ermitteln, wie stark die Lesefähigkeit und das Vorwissen das Textverständnis vorhersagen. Die für die Interpretation bedeutsamsten Koeffizienten lassen sich Bortz und Schuster (2010, 342) entnehmen.

Zunächst soll das zugrunde liegende Prinzip der multiplen Regression veranschaulicht werden, um anschließend auf die Vorgehensweise der Analyse von Wechselwirkungen überzuleiten. Erneut soll davon ausgegangen werden, dass

zwei Prädiktoren (A_i = Lesefähigkeit und B_i = Lesemotivation) das Textverständnis (Y_i) vorhersagen. Die Regressionsgleichung mit zwei unabhängigen Variablen lautet (vgl. Field 2013, 297):

$$Y_i = (b_0 + b_1A_i + b_2B_i) + e_i$$

Abb. 1: Regressionsgleichung mit zwei Prädiktoren (eigene Darstellung)

Die Gleichung besagt, dass beispielsweise das Textverständnis Y_i durch den Regressionskoeffizienten b_1 des Prädiktors A_i (Lesefähigkeit) sowie durch den Koeffizienten b_2 des Prädiktors B_i (Lesemotivation) vorhergesagt wird. Bei diesen Koeffizienten handelt es sich um sogenannte Haupteffekte. Sie bestimmen mit spezifischen Steigungswerten den Anstieg der Regressionsgeraden (vgl. Bortz/Schuster 2010, 343). Die Koeffizienten prognostizieren somit, um wie viele Einheiten (Standardabweichungen) sich die abhängige Variable verändert, wenn sich die Prädiktoren ebenfalls um eine Einheit ändern. Außerdem beinhaltet die Regressionsschätzung die Konstante b_0 , die den Mittelwert der abhängigen Variable auf der y-Achse unter Berücksichtigung beider unabhängigen Variablen angibt. Auf das vorherige Beispiel bezogen bedeutet dies, es wird das mittlere Textverständnis geschätzt, wenn die Lesefähigkeit und das Vorwissen zugleich berücksichtigt werden. Der Fehlerterm e_i umfasst sämtliche Abweichungswerte, die sich nicht exakt auf der Regressionsgeraden befinden. Ob die Regressionskoeffizienten (b) die Kriteriumsvariable statistisch bedeutsam vorhersagen, verdeutlichen die Signifikanzprüfungen mittels t -Tests. Hingegen verweist der Determinationskoeffizient (R^2) auf die Höhe der Varianzaufklärung der abhängigen Variable. Mit anderen Worten: Wie bedeutsam sind der gemeinsame Einfluss von Lesefähigkeit und Vorwissen für die Vorhersage des Textverständnisses? Für ausführliche Informationen zu den Teststatistiken sei auf Field (2013) verwiesen. Ferner sei darauf aufmerksam gemacht, dass die Höhe der Koeffizienten von den Wissenschaftsdisziplinen unterschiedlich interpretiert wird. Im Rahmen deutschdidaktischer Fragestellungen, die auf empirischen Informationen beruhen, welche aus weitaus unkontrollierteren Bedingungen stammen als es im psychologischen Experiment üblich ist, sind die Koeffizienten etwas weniger streng zu interpretieren.

Neben den oben aufgezeigten linearen Zusammenhangsanalysen lassen sich mittels der multiplen Regression Interaktionen schätzen. Es könnte sich beispielsweise um eine Interaktion handeln, wenn die Wirkung eines Lesestrategietrainings von der allgemeinen Lesefähigkeit der Leserinnen und Leser beeinflusst (moderiert) wird. Abbildung 2 veranschaulicht das Grundprinzip einer Moderation.

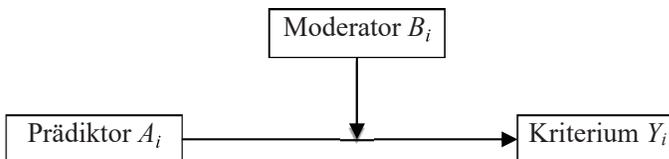


Abb. 2: Prinzip eines Moderatoreffekts (eigene Darstellung)

Eine Moderation liegt folglich vor, wenn eine Drittvariable (z.B. der Moderator Lesefähigkeit) den Zusammenhang zwischen zwei anderen Variablen (z.B. dem Lesestrategietraining (Prädiktor A) und dem Textverständnis bedeutsam verändert (vgl. Bortz/Schuster 2010, 357). Für die statistische Modellierung eines Interaktionseffektes eignet sich eine moderierte Regressionsanalyse. Die entsprechende Regressionsgleichung inkludiert neben dem Prädiktor A_i (Lesestrategietraining) und dem Moderator B_i (Lesefähigkeit) die Wechselwirkung zwischen diesen zwei Variablen in Form eines Interaktionsprädiktors (vgl. Abb. 3).

$$Y_i = (b_0 + b_1A_i + b_2B_i + b_3AB_i) + e_i$$

Abb. 3: Regressionsgleichung mit Interaktionsterm (eigene Darstellung)

Die lineare Regressionsgleichung beruht bei einer Interaktionsanalyse somit auf einer zentralen Veränderung zu Abbildung 2: Die Wechselwirkung aus den Kriteriumsvariablen A_i und B_i wird durch einen Produktterm ($A \times B_i$) zusätzlich berücksichtigt. Mit einer solchen Modellierung lässt sich analysieren, ob ein Interaktionseffekt vorliegt, der den Zusammenhang zwischen einem Kriterium und einer abhängigen Variable Y_i verändert. Es kann folglich ermittelt werden, bei welchem Ausmaß an Lesefähigkeit das Lesestrategietraining effektiv ist. Aufschluss über die statistische Bedeutsamkeit von Interaktionstermen gibt die Signifikanzprüfung mittels eines t -Tests (vgl. ebd.).

Liegt eine statistisch bedeutsame Wechselwirkung zwischen zwei Variablen vor, so ist der Einfluss des Prädiktors A_i auf das Kriterium Y_i betroffen. Mit der multiplen Regression kann das dahinterliegende Muster erschlossen werden (vgl. Aiken/West 1991). Auf das vorherige Beispiel bezogen ließe sich aufklären, ob sich das Lesestrategietraining als effektiver bei mehr versus weniger Lesefähigkeit erweist (vgl. Richter 2007, 119). Um die Effektstärken einer kategorialen Variable (Training vorhanden in einer Experimentalgruppe vs. Training nicht vorhanden in einer Kontrollgruppe) zu ermitteln, ist die Moderatorvariable (die Lesefähigkeit beispielsweise in Form einer Skala von 1-10 Punkten) zu skalieren. Diese Skalierung kann sich darauf beziehen, dass Abweichungswerte von einer oder mehreren Standardabweichungen über sowie unter dem Stichprobenmittelwert gebildet werden (vgl. ebd., 121). Mit diesen neuen Skalen können anschließend komplementäre multivariate Regressionen durchgeführt werden, um die Wirkung des Lesestrategietrainings bei unterschiedlichen Ausprägungen an Lesefähigkeit zu schätzen. Betrachtet werden könnte somit, wie das Training bei einer unterdurchschnittlichen und einer überdurchschnittlichen Lesefähigkeit wirkt.

Das Ergebnis der Interaktionsanalyse besteht folglich in der Aufklärung möglicher Einflüsse einer Drittvariable auf den linearen Zusammenhang von zwei Variablen. Bedeutsam für die inhaltliche Interpretation ist, dass bei Vorliegen einer signifikanten Interaktion dieser Effekt zu betrachten ist und die Haupteffekte der an der Interaktion beteiligten Variablen zu vernachlässigen sind. Inhaltlich bedeutet dies, dass der reine Haupteffekt des Lesestrategietrainings hinfällig wird, so-

fern eine Interaktion mit der Lesefähigkeit vorliegt. Das Training hat in Abhängigkeit vom Vorwissen keinen allgemeinen Effekt auf das Textverständnis. Mit anderen Worten: Das Lesestrategietraining wirkt nicht bei allen Schülerinnen und Schülern in gleichem Ausmaß auf das Textverständnis (in diesem Fall wäre der Haupteffekt bedeutsam), sondern wirkt in Abhängigkeit vom Vorwissen differenziell. Nur dieses letztgenannte Abhängigkeitsverhältnis sollte im Falle einer signifikanten Wechselwirkung betrachtet werden, was durch den Interaktionseffekt ausgedrückt wird.

1.2 Vorbereitung der Ausgangsdaten: Kodierung und Skalierung von Variablen

Für das oben beschriebene Auswertungsverfahren der multiplen (moderierten) Regressionsanalyse gilt, dass die abhängige Variable, das Kriterium, ein metrisches Skalenniveau aufweist. Metrische Skalen sind durch Verhältnis- und Intervallniveaus gekennzeichnet, wie unter anderem Temperaturskalen, Gewichts- und Altersangaben (vgl. Bortz/Schuster 2010, 13). Sollen abhängige Variablen mit kategorialen Skalenniveaus einbezogen werden, sei auf die logistische Regression verwiesen (vgl. Field 2013). Angesichts der Modellierung der Prädiktoren sind neben metrischen Variablen kategoriale Skalenniveaus zulässig. Kategoriale Variablen umfassen Nominal- und Ordinalskalenniveaus. Bei Nominalskalen handelt es sich beispielsweise um Kategorisierungen, wie der Familienstand oder das Geschlecht einer Person, wohingegen Ordinalskalen rangverteilte Daten, wie Schulabschlüsse, repräsentieren.

Hinsichtlich der Analyse von Wechselwirkungen mittels moderierter Regression sind die Prädiktoren zu transformieren (vgl. Richter 2007, 118). Um kategoriale Variablen als Prädiktoren einzubeziehen, sollten sie mit einer Dummykodierung als Designvariablen transformiert werden (vgl. Field 2013, 419). Im Hinblick auf das oben beschriebene experimentelle Lesestrategietraining würde den Probandinnen und Probanden der Gruppe, die das Training erhalten hat, der Wert eins zugewiesen. Die Kontrollgruppe erhielte hingegen den Wert null. Auch bei Untersuchungsdesigns, in denen kategoriale Prädiktoren mit mehreren Ausprägungen einbezogen werden, würde eine Dummykodierung mit mehreren Ausprägungen erfolgen (vgl. zur ausführlichen Vorgehensweise: Richter 2007, 118). Ferner bietet es sich an, kategoriale Prädiktoren angesichts deren Verteilung auf ungleiche Gruppengrößen durch eine gewichtete Effektkodierung zu transformieren. Die Personen der ersten Gruppe erhielten den Wert eins und die der zweiten Gruppe einen negativen gewichteten Wert ($-N_1/N_2$). Eine solche gewichtete Effektkodierung repräsentiert die Abweichung vom theoretischen Gesamtmittelwert unter der Annahme gleicher Gruppengrößen (vgl. ebd.). Weitere Skalierungsvarianten für kategoriale Prädiktoren lassen sich bei Cohen et al. (2003) recherchieren.

Sofern mehrere metrische Prädiktorvariablen (z.B. Lesefähigkeit *und* Vorwissen) für die Vorhersage einer abhängigen Variable einbezogen werden und diese in

einem engen empirischen Zusammenhang stehen (deutlich miteinander korrelieren), liegt Multikollinearität vor. „Mit Multikollinearität ist die „Instabilität“ der Regressionskoeffizienten gemeint, welche sich aus Abhängigkeiten zwischen Prädiktorvariablen ergeben“ (Bortz/Döring 2010, 354). Diese Abhängigkeit äußert sich beispielsweise darin, dass mit einer besseren Lesefähigkeit ein größeres Vorwissen einhergehen kann. Da die Prädiktorvariablen in solchen Fällen nicht unabhängig voneinander sind und eine gemeinsame Erklärungskraft für die abhängige Variable (z.B. das Textverständnis) besitzen, können die Haupteffekte der Prädiktoren und ihre Standardfehler fehlerbehaftet sein. Um einer Multikollinearität vorzubeugen, sollten die metrischen Prädiktoren an ihrem Stichprobenmittelwert zentriert werden. Die Zentrierung erfolgt durch die Verwendung von Abweichungswerten vom Gesamtmittelwert der Variablen ($X_i - \bar{X}$). Der Stichprobenwert der Variablen repräsentiert nach der Zentrierung den Wert null und der Regressionskoeffizient eines kategorialen Prädiktors ist bei einer mittleren Ausprägung des metrischen Prädiktors zu interpretieren (Richter 2007, 118). In Bezug auf das vorherige Beispiel würde der Effekt des Lesestrategietrainings bei einer mittleren Ausprägung an Lesefähigkeit und Vorwissen interpretiert werden, wenn die Lesefähigkeit und das Vorwissen in zentrierter Form berücksichtigt werden.

1.3 Gütekriterien für die multiple Regressionsschätzung

Die zuverlässige Vorhersage einer abhängigen Variable basiert auf der Berücksichtigung spezifischer Gütekriterien. Als Ausgangsbasis ist festzulegen, welche unabhängigen Variablen der Prognostizierung der abhängigen Variable dienen sollen. Der Einschluss einer beliebigen Anzahl an erklärenden Variablen ist nicht zielführend. Eine Entscheidungshilfe für die Selektion von erklärenden Variablen bieten sowohl der theoretische und empirische Forschungsstand als auch eine Analyse der Zusammenhänge (Korrelationen) der Prädiktoren untereinander und mit dem Kriterium.

Neben diesen grundsätzlichen Vorüberlegungen sollte durch einen Kolmogorov-Smirnov-Verteilungstest die Normalverteilung der Variablen überprüft werden. Eine Normalverteilung besitzt einen glockenförmigen, symmetrischen Verlauf. Abbildung 4 veranschaulicht eine Häufigkeitsverteilung anhand eines Balkendiagramms und eine darüber liegende Normalverteilungskurve.

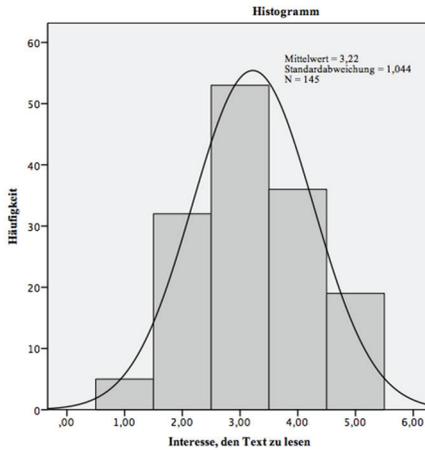


Abb. 4: Balkendiagramm und Normalverteilung (eigene Darstellung)

Deutlich wird, dass die Skala „Interesse, den Text zu lesen“ mit 145 gültigen Antworten geringfügig von einer Normalverteilung abweicht. Würde eine perfekte Normalverteilung vorliegen, wäre die glockenförmige Linie exakt mit Werten ausgefüllt und es lägen keine überstehenden sowie darunterliegenden Verteilungen vor. Für dieses exemplarische Beispiel lässt sich allerdings schlussfolgern, dass trotz der nicht vorliegenden perfekten Normalverteilung ebenjene Daten für eine Regressionsanalyse genutzt werden können, da die Abweichungen geringfügig sind. Drastisch abweichende Verteilungsformen finden sich in Bortz und Schuster (2010, 73). Normalverteilte Daten stellen eine zentrale Voraussetzung für die Signifikanzprüfung dar (F -Test zur Signifikanzprüfung des Determinationskoeffizienten und t -Test zur Signifikanzprüfung der einzelnen Steigungskoeffizienten). Der K-S-Verteilungstest klärt für jede betrachtete Variable auf, ob die Verteilung statistisch bedeutsam von einer Gauß'schen Normalverteilung abweicht. Für die Annahme normalverteilter Daten ist ein nicht-signifikantes Ergebnis ($p > .05$) erforderlich, das besagt, dass die Daten nicht von einer Normalverteilung abweichen. Allerdings gilt dieses Testverfahren als stichprobensensibel. Dies bedeutet, dass mit einem steigenden Umfang an Probandinnen und Probanden signifikante Ergebnisse auf Grund des verringerten Standardfehlers resultieren können, die sich bei kleineren Stichproben nicht ergeben würden. Statistische Transformationen mithilfe derer die Schiefe oder Wölbung von Skalen der Form einer Normalverteilung angeglichen werden kann, werden in Bortz und Schuster (2010) behandelt.

Ein weiteres Kernkriterium für eine robuste Regressionsschätzung besteht in der linearen Beziehung der erklärenden Variable mit der abhängigen Variable. Linearität bedeutet, dass sich Kriterium und Prädiktor nur in einer konstanten Position verändern. Steigt z.B. das Vorwissen, so steigt in der Regel auch das Textverständnis. Eine Unterstellung eines linearen Richtungszusammenhangs ist für eine präzise Vorhersage der Werte der abhängigen Variable konstitutiv. Mittels eines

Streudiagramms lässt sich sehr anschaulich ermitteln, ob ein linearer Zusammenhang zwischen den zu untersuchenden Variablen vorliegt. Hierbei werden die Werte der abhängigen Variable auf der y-Achse und die standardisierten Werte der unabhängigen Variable auf der x-Achse in einem Koordinatensystem abgetragen. Abbildung 5 stellt einen tendenziell linearen Zusammenhang zwischen thematischem Vorwissen und Textverständnis dar, der mittels der Statistiksoftware SPSS ermittelt wurde.

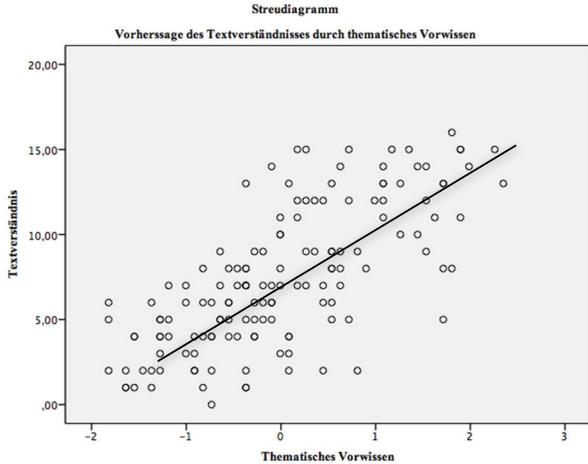


Abb. 5: Exemplarisches Streudiagramm (eigene Darstellung)

Die Linearität der beiden Variablen wird dadurch deutlich, dass die Punktwolke sich tendenziell auf einer Linie von links unten nach rechts oben entwickelt. Je mehr Vorwissen vorliegt, desto besser erweisen sich die Mittelwerte im Textverständnis. Perfekte Linearität läge vor, wenn sich die Punkte alle auf der Linie befänden, was jedoch mit empirischen Daten häufig nur schwer zu erreichen ist. Weicht der Zusammenhang deutlich von einer linearen Beziehung ab, ist zu entscheiden, ob die Variablen transformiert werden können oder ob sie aus der Vorhersage exkludiert werden sollten (vgl. Bortz/Schuster 2010, 200f.). Linearität wird jedoch nicht ausschließlich durch eine Linie von links unten nach rechts oben abgebildet, sondern dieser Zusammenhang liegt ebenfalls vor, wenn sich die Punktwolke von links oben nach rechts unten bewegt. Als ein prägnantes Beispiel für einen solchen Zusammenhang lässt sich Folgendes anführen: Je höher die Kindersterblichkeit in einem Land ausfällt, desto geringer ist die Lebenserwartung. Weitere Zusammenhänge, die andere Formen als die der Linearität aufweisen, lassen sich in Bortz und Schuster (ebd., 198) recherchieren. Mithilfe eines Streudiagramms kann außerdem überprüft werden, ob Ausreißer in den empirischen Daten vorliegen. Um einen Ausreißer handelt es sich, wenn ein Wert einer Probandin bzw. eines Probanden nicht in eine Messreihe passt. Solche Werte können die Schätzung von Regressionskoeffizienten stark verzerren und sind mittels einer fallweisen Diagnose aus dem Datensatz zu eliminieren (vgl. Field 2013, 305).

Die Unabhängigkeit der erklärenden Variablen untereinander gilt als weitere zentrale Bedingung für eine zuverlässige Regressionsanalyse. Bei empirischen Untersuchungen besteht jedoch häufig ein gewisser Grad an Abhängigkeit (Multikollinearität) zwischen den Prädiktoren (vgl. Bortz/Schuster 2010, 354). Eine solche Abhängigkeit liegt u.a. dann vor, wenn eine bessere Lesefähigkeit mit umfangreichem Vorwissen einhergeht und diese beiden Prädiktoren miteinander korrelieren. Die lineare Unabhängigkeit lässt sich anhand einer Kollinearitätsdiagnose spezifizieren: Je höher die Korrelation der Variablen ausfällt, desto stärker erweist sich die Kollinearität. Als Folgen der Kollinearität gelten eine Erhöhung der Standardfehler der Regressionskoeffizienten und eine Unzuverlässigkeit der Schätzung. Die in Abschnitt 1.2 erläuterte Zentrierung von Variablen am Stichprobenmittelwert stellt eine Möglichkeit dar, der Multikollinearität entgegenzuwirken.

1.4 Vorteile und Nachteile des Verfahrens

Die multiple Regression stellt insgesamt ein recht komplexes Verfahren dar, dessen Aneignung sich aber auf Grund der aussagekräftigen Berechnungen und des Vorteils des Einschlusses mehrerer Variablen sowie Skalenniveaus durchaus lohnt.

Sollen Vor- und Nachteile der multiplen moderierten Regression diskutiert werden, so ist ein Vergleich mit dem Verfahren der Varianzanalyse unerlässlich. Wechselwirkungen zwischen zwei oder mehr Variablen werden überwiegend mittels Varianzanalysen berechnet. Jedoch gerät dieses Analyseverfahren bei der Modellierung von Wechselwirkungen mit unterschiedlichen Skalenniveaus (metrisch *und* kategorial) und der Analyse des spezifischen Interaktionsmusters an seine Grenzen, da Varianzanalysen auf kategoriale Skalenniveaus ausgerichtet sind. Für die Analyse des Effekts eines kategorialen Prädiktors bei unterschiedlichen Ausprägungen eines metrischen Prädiktors ist Letzterer im Rahmen einer Varianzanalyse mittels Gruppenbildung zu kategorisieren. Eine solche Teilung einer kontinuierlichen Skala, z.B. an ihrem Mittelwert, führt zu neuen Teilstichproben, die in der Varianzanalyse in Form von Kategorien einbezogen werden können. Man könnte eine Skala, die einen fiktiven Mittelwert von etwa 5.4 besitzt zu zwei Gruppen transformieren: Personen der Gruppe A, die einen Mittelwert < 5.4 besitzen und Personen der Gruppe B, die einen Mittelwert ≥ 5.4 aufweisen. Mit dieser künstlichen Gruppenbildung lassen sich anschließend die Effekte des kategorialen Prädiktors für beide Gruppen bestimmen. Die beschriebene Skalentransformation führt jedoch zu einem Verlust der ursprünglich enthaltenen Informationen und reduziert die Messgenauigkeit. Verzerrte Schätzungen von Interaktionseffekten und falsche Entscheidungen zugunsten von Alternativhypothesen sind nicht auszuschließen (vgl. Richter 2007, 117). Entscheidungen wie jene, dass sich ein Lesestrategietraining positiv auf das Textverständnis auswirkt, könnten somit fehlerbehaftet sein.

Die moderierte Regression besitzt im Gegensatz zur Varianzanalyse den Vorteil, mit einer sensiblen Bildung von Abweichungswerten vom Stichprobendurch-

schnitt auf eine künstliche Gruppenbildung zu verzichten und das Muster des Interaktionseffektes robuster zu modellieren (vgl. ebd.). Eine Berücksichtigung von kategorialen Variablen mit mehr als zwei Ausprägungen hingegen führt im Rahmen der multiplen Regression wiederum zu dem Nachteil, dass für den Kontrast von kategorialen Effekten eine Dummykodierung genutzt werden muss und mehrere Regressionen für den Vergleich der jeweiligen kategorialen Ausprägungen zu modellieren sind. Die Varianzanalyse hingegen eignet sich für einen Einbezug von kategorialen, mehrstufigen Skalenniveaus in besonderer Weise, da Mehrfachvergleiche für kategoriale Gruppen durchgeführt werden können. Angesichts dieser spezifischen Voraussetzungen ist bereits bei der Konzeption von Testinstrumenten und somit vor der Erhebung der Daten zu berücksichtigen, welches statistische Auswertungsverfahren sich für die Beantwortung der Forschungsfragen eignet und welche Grenzen sowie Möglichkeiten hiermit verbunden sind.

Abschließend sollen pragmatische Aspekte der Datenerhebung und der Datenauswertung skizziert werden. Für eine zuverlässige Regressionsanalyse ist es bedeutsam, dass Fragebögen oder Testinstrumente zielgerichtet konstruiert werden und dass sie den Gütekriterien der Reliabilität sowie der Validität genügen. Nur wenn ein Testinstrument eine Variable zuverlässig misst, sind Aussagen über deren Bedeutung für ein Kriterium möglich und Regressionsanalysen zuverlässig. Daher sollten Items oder Aufgaben von Testinstrumenten einer Voruntersuchung unterzogen werden. Bedeutsam ist es ferner, dass die Stichprobengröße a priori bestimmt wird, beispielsweise mit dem Programm g-Power (Faul et al. 2009). Die Datenanalyse lässt sich mit gängiger Statistiksoftware wie IBM SPSS Statistics umsetzen. SPSS besitzt im Gegensatz zu Mplus eine anwenderfreundliche Benutzeroberfläche und bietet durch zahlreiche Handbücher mit illustrierten Dialogfeldern Hilfestellungen.

1.5 Literaturempfehlungen

Aiken/West (1991): Multiple regression: Testing and interpreting interactions.

Ein umfassendes Grundlagenwerk zur Funktionsweise multipler (moderierter) Regressionsanalysen.

Bortz/Schuster (2010): Statistik für Human- und Sozialwissenschaftler.

Dieses Lehrbuch vermittelt fundiertes Grundlagenwissen zu verschiedenen statistischen Auswertungsverfahren.

Field (2013): Discovering statistics using IBM SPSS Statistics.

Ein anwendungsorientiertes Lehrbuch zur Systematik der multiplen Regression und zur praktischen Anwendung des Verfahrens mit der Statistiksoftware SPSS.

Richter (2007): Wie analysiert man Interaktionen von metrischen und kategorialen Prädiktoren? Nicht mit Median-Splits!

Eine theoretische und anwendungsorientierte Erläuterung des Analyseverfahrens mithilfe eines Beispiels aus der Medienpsychologie.

2. Die Wirkung von globaler Textkohäsion auf das Textverständnis in Abhängigkeit vom thematischen Vorwissen

In der nachfolgend erläuterten experimentellen Studie wurde untersucht, ob eine erhöhte globale Kohäsionsdichte in einem Sachtext das Textverständnis unterstützen kann und welche Schülerinnen und Schüler von dem Angebot der Textkohäsion profitieren (Interaktionsanalyse) (vgl. Schmitz/Gräsel 2016). Die globale Textkohäsion bezieht sich auf sprachliche Verknüpfungen eines Textes über benachbarte Sätze hinweg, innerhalb von Absätzen und im Textganzen. 154 Schülerinnen und Schüler der neunten Jahrgangsstufe lasen entweder einen global kohäsiven oder einen weniger kohäsiven Sachtext. Deren Textverständnis wurde mit einem Verständnistest nach dem Lesen des Textes erfasst. Ferner wurden kognitive Verständnisvoraussetzungen erhoben, die den Einfluss der globalen Textkohäsion beeinflussen können. Hierzu zählen das thematische Vorwissen und die Lesefähigkeit (vgl. Ozuru et al. 2009). Kennzeichen der Datenlage waren folglich ein kategorialer Prädiktor (zwei Versionen eines Sachtextes) und zwei Lernvoraussetzungen mit metrischem Skalenniveau (thematisches Vorwissen und Lesefähigkeit). Im Folgenden wird insbesondere auf die Wechselwirkung der Textkohäsion mit dem thematischen Vorwissen fokussiert. Der Einfachheit halber wird eine Wechselwirkung der globalen Kohäsion mit der Lesefähigkeit nicht modelliert, sondern die Lesefähigkeit wird in Form eines Haupteffekts einbezogen.

2.1 Forschungsfragen

In der Studie wurde den Forschungsfragen nachgegangen, ob die globale Textkohäsion eine allgemeine förderliche Wirkung auf das Textverständnis entfaltet oder ob sie mit dem thematischen Vorwissen interagiert und wie sich diese Wechselwirkung auf das Textverständnis der Schülerinnen und Schüler auswirkt.

2.2 Stichprobe

Mithilfe des Bildungsservers des Landes Nordrhein-Westfalen wurden 50 Gesamtschulen und Gymnasien in städtischem Einzugsgebiet kontaktiert. Es wurden Schülerinnen und Schüler beider Schulformen befragt, um eine Varianz in den kognitiven Lesevoraussetzungen zu erzielen. Für die vorliegende Studie wurden ein Gymnasium (3 Klassen) und zwei Gesamtschulen (4 Klassen) per Zufallsziehung ausgewählt. Es nahmen 154 Schülerinnen und Schüler der neunten Jahrgangsstufe teil (74 Mädchen, 71 Jungen). 55% der Schülerinnen und Schüler besuchten ein Gymnasium und 45% eine Gesamtschule. Das mittlere Alter betrug $M=15.3$ Jahre ($SD=0.60$).

2.3 Methode

2.3.1 Konstruktion eines Lesetextes

Als unabhängige Variable wurden zwei Sachtexte konstruiert, die sich in ihrem globalen Kohäsionsgrad voneinander unterschieden. Dieser Sachtext könnte im Fach Deutsch sowie in gesellschaftswissenschaftlichen Unterrichtsfächern zur

Anwendung kommen (vgl. Schmitz 2016). Das Thema des Textes war die wirtschaftliche Entwicklung Afrikas und die Notwendigkeit von Bildungsinvestitionen sowie von Entwicklungshilfen. Der Text beinhaltete keine diskontinuierlichen Textteile, um den Effekt der Textmanipulation auf die globale Textkohäsion rückführen zu können. Letztere wurde in Anlehnung an theoretisch fundierte und empirisch erprobte Kriterien von McNamara und Kintsch (1996) und den Kohäzenzbildungshilfen nach Schnotz (1994) manipuliert. Die Manipulation der globalen Textkohäsion erfolgte zur Verknüpfung nicht direkt-benachbarter Sätze und umfasste fünf Merkmale: (1) Erzeugen nominaler Referenz durch Argumentüberlappungen; Ersetzen von Pronomina durch Nomen, die über Sätze hinweg ambig wirken könnten; (2) Einfügen von expliziten globalen Topic-Strukturen durch Überschriften und Textdeixis; (3) Markierung von Textsegmenten und Segmentgrenzen durch Sinnabschnitte und Aufzählungen; (4) Betonung übergeordneter Wissenseinheiten u.a. durch Relevanzindikatoren und Gradpartikeln; (5) Explikation funktionaler Relationen durch Konjunkionaladverbien und Herausstellung hierarchiehoher Relationen. Abbildung 6 veranschaulicht die Textmanipulation durch Verweise auf die zuvor benannte Nummerierungen und die unterstrichenen Textstellen.

[2. 3] Schattenseiten: kaum Industrie auf dem Sorgenkontinent

Der [2]oben beschriebene Handel ist [5]problematisch für [1]Afrika. Da viele Industriestaaten, wie einst europäische Kolonialstaaten, [4]nur an Rohstoffgewinnung interessiert sind und man sich auf den Abbau der Ressourcen konzentriert, trägt die [2]oben beschriebene [1]Wirtschaftslage nicht zur Industrialisierung bei. [1]Afrika stellt kaum eigene Produkte her. Die OECD sieht [5]3 Probleme: [3]1) [1]Afrika macht sich [1]vom Rohstoffexport abhängig. [3]2) [4]Nur eine Minderheit profitiert. [3]3) Ein Preisverfall der Bodenschätze hätte katastrophale [5]Folgen für [1]Afrika. Der Kontinent hat zudem mit infrastrukturellen, sozialen und politischen [5]Problemen zu kämpfen. [4]Insgesamt zeigt sich, dass [1]Afrika [4]nur teilweise vom Rohstoffexport profitiert und sich industriell nicht entwickelt.

Abb. 6: Exemplarische Manipulation des Sachtextes

Angesichts dieser experimentellen Manipulation entstanden zwei Lesetexte (global kohäsiv und global weniger kohäsiv), die im Datensatz mit den Ausprägungen null und eins kodiert wurden. Jede Probandin bzw. jeder Proband las einen der beiden Texte (Within-Design).

2.3.2 Operationalisierung von thematischem Vorwissen und Lesefähigkeit

Um zu erfassen, ob die globale Textkohäsion in Abhängigkeit von thematischem Vorwissen unterschiedlich stark wird, wurde diese Lesevoraussetzung durch neun Aufgaben erfasst (drei Multiple-Choice-Fragen und sechs halb-offene Fragen). Die Aufgabenkonstruktion basierte auf einer Wissensstrukturanalyse des Sachtextes, um das für das Verständnis des Textes erforderliche Wissen zu erfassen. Die Auswertung der Stichworte zu den halb-offenen Aufgaben basierte auf einem a priori definierten Erwartungshorizont, womit die Antworten als korrekt oder falsch bewertet wurden. Pro Aufgabe konnten drei Punkte erzielt werden.

Die Schülerinnen und Schüler erreichten im Durchschnitt $M=10.04$ ($SD=5.52$) Punkte von 18 Punkten. Die korrigierten Trennschärfen der Items umfassten $r_{it}=.38-.50$. Die interne Konsistenz dieser metrischen Skala erzielte einen guten Cronbachs Alpha-Wert von $\alpha=.74$.

Um die allgemeine Lesefähigkeit der Schülerinnen und Schüler als metrischen Prädiktor zu erheben, wurde der standardisierte Lesegeschwindigkeits- und -verständnisstest (LGVT 6-12) verwendet (vgl. Schneider et al. 2007 – zu standardisierten Lesetests s. Dube in diesem Band). Der Test beinhaltet einen Text mit 1.727 Wörtern. In diesem befinden sich 23 Wortlücken, die anhand eines geeigneten Wortes aus drei konkurrierenden Wortalternativen in Klammern während des Lesens zu ergänzen sind. Der LGVT 6-12 misst mittels eines globalen Screenings die Lesefähigkeit auf der Satzebene, indem die Rezipienten die syntaktischen und semantischen Informationen des Satzes mit der zu ergänzenden Wortlücke komplettieren müssen. Neben der syntaktischen und semantischen Integration auf der Satzebene sind Fähigkeiten zum sinnentnehmenden Lesen auf der Textebene (lokale und globale Kohärenzbildung) bedeutsam. Gleichzeitig überprüft der Test mit der Anzahl der gelesenen Wörter die Lesegeschwindigkeit. Es wurde der Summenwert der Satzergänzungen als Maß für die Lesefähigkeit genutzt, da Verfälschungstendenzen mit einer Ratekorrektur entgegengewirkt werden konnte. Jedes richtig ausgewählte Wort erhielt zwei Punkte, eine falsche Auswahl erhielt einen Minuspunkt. Ausgelassene Bearbeitungen wurden nicht bewertet (Schneider et al. 2007, 17-19). Die Schülerinnen und Schüler ergänzten im Durchschnitt ca. sieben Satzergänzungen richtig ($M=14.22$, $SD=8.98$).

2.3.3 Operationalisierung der Skala Textverständnis

Für die Erfassung des Textverständnisses in Form einer metrischen, abhängigen Variable wurden 17 Aufgaben konstruiert. Es wurden drei Multiple-Choice-Fragen formuliert, die vier Antwortoptionen mit drei Disktraktoren, also falschen Wahlmöglichkeiten, beinhalteten. Weiter wurden sechs Verifikationsfragen entwickelt, bei welchen die Schülerinnen und Schüler vorgegebene Sätze hinsichtlich ihrer inhaltlichen Richtigkeit beurteilen sollten. Um die Ratewahrscheinlichkeit zu reduzieren, konnte bei den Verifikationsfragen und den Multiple-Choice-Fragen die Option „weiß ich nicht“ ausgewählt werden. Zudem wurden fünf halb-offene sowie drei offene Aufgaben gestellt. Bei vier der fünf halb-offenen Fragen reichte als Antwort ein Stichwort aus. Eine halb-offene Aufgabe, bei der ein Sachverhalt begründet werden sollte, wurde durch zwei Rater beurteilt, deren Übereinstimmung einen Kappa-Wert von $\kappa=.73$ aufwies. Die drei offenen Antworten konnten mit zwei bis drei Sätzen beantwortet werden. Zwei unabhängige Rater kodierten die Schülerantworten zu den drei offenen Aufgaben anhand eines vorab formulierten Erwartungshorizonts. Die Interrater-Reliabilität für die Aufgabenbewertungen ergab eine gute Übereinstimmung ($\kappa=.68 - .71$). Für jede richtig beantwortete Aufgabe wurde ein Punkt vergeben. Die Schülerinnen und Schüler erzielten im Durchschnitt $M=6.96$ ($SD=3.92$) von 17 Punkten. Die korrigierten Trennschärfen der Items umfassten $r_{it}=.32 - .65$. Die interne Konsistenz der Skala erzielte einen guten Cronbachs-Alpha Wert von $\alpha=.83$.

2.4 Vorgehensweise der Datenanalyse

Um eine Wechselwirkung des kategorialen Prädiktors globale Textkohäsion mit dem metrischen Prädiktor thematisches Vorwissen sowie Haupteffekte der globalen Kohäsion, des thematischen Vorwissens und der Lesefähigkeit auf die abhängige Variable Textverständnis zu schätzen, wurden moderierte Regressionsanalysen berechnet. Die zuvor durchgeführten Normalverteilungs- und Linearitätsanalysen werden an dieser Stelle aus Platzgründen ausgespart, sie lassen sich jedoch in Schmitz (2016) recherchieren.

In einem vorbereitenden Schritt wurden die Prädiktorvariablen globale Textkohäsion, thematisches Vorwissen und Lesefähigkeit skaliert. Der kategoriale Prädiktor globale Textkohäsion wurde auf Grund ungleicher Gruppengrößen (74 Schülerinnen und Schüler lasen den inkohäsiven Text, 71 Probandinnen und Probanden lasen den kohäsiven Text) durch eine gewichtete Effektkodierung transformiert. Die gewichtete Effektkodierung dieser Variable repräsentiert die Abweichung vom theoretischen Gesamtmittelwert unter der Annahme gleicher Gruppengrößen (vgl. Richter 2007, 118). Um eine Multikollinearität der unabhängigen Variablen zu reduzieren, wurden die untereinander korrelierten Skalen thematisches Vorwissen und Lesefähigkeit zentriert. Die Zentrierung erfolgte durch die Verwendung von Abweichungswerten vom Gesamtmittelwert der Variablen ($X_i - \bar{X}$). Anschließend wurden mit diesen transformierten Skalen Haupteffekte und Interaktionseffekte auf das Konstrukt Textverständnis regrediert. Abbildung 7 veranschaulicht die lineare Regressionsgleichung mit Haupt- und Interaktionstermen:

$$Y_i = (b_0 + b_1A_i + b_2(X_i - X)B_i + b_3(X_i - X)C_i + b_4AB_i) + e_i$$

Abb. 7: Regressionsanalyse

b_0 stellt die additive Konstante dar, die das mittlere Textverständnis nach dem Lesen des global weniger kohäsiven Textes ausdrückt. b_1 repräsentiert den Haupteffekt des kategorialen, dichotomen Prädiktors globale Textkohäsion. b_2 umfasst den Haupteffekt des zentrierten, metrischen Prädiktors (thematisches Vorwissen) und b_3 beinhaltet den Haupteffekt der zweiten zentrierten, metrischen Variable (Lesefähigkeit). b_4 verdeutlicht den Interaktionsterm aus globaler Textkohäsion multipliziert mit thematischem Vorwissen. Zudem enthält die Gleichung den Residualterm e_i , der die Abweichungen von der Regressionsgeraden widerspiegelt.

2.5 Ergebnis der multiplen moderierten Regression

Für die Beantwortung der Fragestellungen wurde die Interaktion aus globaler Textkohäsion und thematischem Vorwissen sowie die Haupteffekte der globalen Textkohäsion, des thematischen Vorwissens und der Lesefähigkeit auf die abhängige Variable Textverständnis geschätzt (vgl. Tab. 1).

	<i>B</i>	<i>SE B</i>	β	<i>t</i>	Sig <i>p</i>
Modell zur Schätzung der Interaktionsanalyse					
Konstante	7.43	0.22			
Globale Kohäsion	0.69	0.21	.17	3.20	.002
Thematisches Vorwissen	0.33	0.06	.44	5.14	.000
Lesefähigkeit	0.14	0.03	.31	4.59	.000
Globale Kohäsion x thematisches Vorwissen	0.17	0.08	.15	2.07	.041
$R^2=.78.$					

Anmerkungen. *B* = unstandardisiertes Beta-Gewicht; *SE B* = Standardfehler des Beta-Gewichts, β = standardisiertes, um Messeinheiten befreites Beta-Gewicht, *t* = Wert der *t*-Statistik; Sig *t* = Signifikanzniveau des *t*-Wertes.

Tab. 1: Multiple Regressionsanalyse

Die letzte Zeile der Tabelle verdeutlicht, dass ein signifikanter Interaktionseffekt der globalen Textkohäsion mit dem thematischen Vorwissen vorliegt ($t(1, 140)=2.07, p=.041$). Der *t*-Wert für die Interaktion aus globaler Textkohäsion und thematischem Vorwissen umfasst bei 140 Freiheitsgraden ein β -Gewicht von .15 und eine Ausprägung 2.07. Der *p*-Wert beträgt gerundet .04 und ist somit kleiner als das Signifikanzniveau von .05 und folglich signifikant. Dies bedeutet, dass der Interaktionseffekt statistisch bedeutsam ist und dass es sich bei einer Wahrscheinlichkeit von 95% um keinen zufälligen Effekt handelt. Ferner lässt sich entnehmen, dass der Mittelwert im Textverständnis nach dem Lesen des global inkohäsiven Textes $M=7.43$ mit einem Standardfehler von $SD=0.22$ umfasst, wenn das Vorwissen und die Lesefähigkeit als zentrierte Variablen (Stichprobendurchschnitt) sowie der Interaktionsterm einbezogen werden. Die Gesamtaufklärung des Textverständnisses durch die Lesefähigkeit, das Vorwissen und die Interaktion lässt sich am Determinationskoeffizienten $R^2=.78$ ablesen. Folglich wird die abhängige Variable Textverständnis von den erklärenden Variablen zu 78% prognostiziert, wohingegen 22% der abhängigen Variable unaufgeklärt bleiben. Die Haupteffekte des thematischen Vorwissens und der globalen Textkohäsion sind nicht zu interpretieren, da der Interaktionsterm signifikant ausfällt: Dies bedeutet, die globale Textkohäsion wirkt nicht bei allen Lernenden allgemein förderlich auf das Textverständnis ungeachtet des thematischen Vorwissens, sondern hat einen differenziellen Effekt in Abhängigkeit vom thematischen Vorwissen. Nur bestimmte Schülerinnen und Schüler konnten in Abhängigkeit ihres Vorwissens einen Nutzen aus der globalen Textkohäsion ziehen. Aufgrund dieser Wechselwirkung kann der durch den Haupteffekt ausgedrückte allgemein förderliche Effekt nicht interpretiert werden, sondern es muss die Interaktion betrachtet werden. Der Haupteffekt der Lesefähigkeit beträgt ($t(1, 140) = 4.59, p < .001$). Dies bedeutet, dass die Lesefähigkeit das Textverständnis im Vergleich zum Interaktionseffekt noch deutlicher vorhersagt, da das β -Gewicht der Lesefähigkeit mit .31 höher ausfällt, wohingegen der Interaktionseffekt $\beta=.15$ beträgt. Mit anderen

Worten: Steigt die Lesefähigkeit um eine Standardabweichung, so erhöht sich der Mittelwert im Textverständnis um .31 Standardabweichungen.

Um das Muster der signifikanten Wechselwirkung zwischen globaler Textkohäsion und thematischem Vorwissen zu untersuchen, wurden die Haupteffekte des kategorialen Prädiktors globale Textkohäsion für unterschiedliche Ausprägungen an thematischem Vorwissen geschätzt. Aufgeklärt werden sollte hiermit, wie die globale Textkohäsion auf das Textverständnis wirkt, wenn die Schülerinnen und Schüler unterschiedlich viel Vorwissen zum Text aufweisen. Dazu wurden Abweichungswerte von einer Standardabweichung über (*Mittelwert + 1 SD*) bzw. unter dem Stichprobenmittelwert (*Mittelwert - 1 SD*) der Skala thematisches Vorwissen gebildet und komplementäre multivariate Regressionen durchgeführt (Richter 2007, 120). Tabelle 2 stellt die Wirkung der globalen Textkohäsion bei einer Standardabweichung mehr sowie bei einer Standardabweichung weniger thematischem Vorwissen als der Stichprobendurchschnitt dar. Die additive Konstante repräsentiert das mittlere Textverständnis nach dem Lesen des global weniger kohäsiven Textes.

Lesevoraussetzung	Additive Konstante (SE)	B (SE)	β	<i>t</i>	<i>P</i>
Thematisches Vorwissen (<i>M - 1 SD</i>)	4.06 (0.47)	.072 (0.66)	.087	1.08	.282
Thematisches Vorwissen (<i>M + 1 SD</i>)	9.44 (0.44)	2.03 (0.66)	.244	3.01	.003

Tab. 2: Haupteffekte der globalen Kohäsion in Abhängigkeit von thematischem Vorwissen

Schülerinnen und Schüler mit mehr thematischem Vorwissen profitierten von der globalen Textkohäsion im Hinblick auf ihr erreichtes Textverständnis ($p=.003$). Sie erzielten mit dem Lesen des global kohäsiven Textes 2.03 Standardabweichungen mehr richtige Punkte im Verständnistest. Hingegen zeigte sich bei den Schülerinnen und Schülern mit weniger thematischem Vorwissen kein signifikanter Effekt der globalen Textkohäsion auf das Textverständnis ($p=.282$). Der Unterschied nach dem Lesen des global kohäsiven Texts beträgt 0.07 Standardabweichungen mehr korrekte Antworten. Somit konnten lediglich Schülerinnen und Schüler mit mehr thematischem Vorwissen als der Stichprobendurchschnitt das Angebot der globalen Textkohäsion für sich nutzen (vgl. Schmitz/Gräsel 2016).

Insgesamt lässt sich resümieren, dass die in dieser Studie fokussierten globalen Textmerkmale das Textverständnis einiger Schülerinnen und Schüler verbessern konnten. Allerdings ist nicht zu vernachlässigen, dass hiermit Leistungsunterschiede vergrößert wurden (Matthäus-Effekt), da lediglich leistungsstärkere Schülerinnen und Schüler profitierten. Damit auch schwächere Schülerinnen und Schüler die textseitigen Hilfen stärker für sich nutzen können, bedarf es vermutlich einer gezielten Unterstützung seitens der Lehrperson. Eine Möglichkeit stellt ein systematischer, auf sprachliche Strukturen ausgerichteter Leseunterricht dar,

in dem die globale Textkohäsion als Mittel der Texterschließung explizit thematisiert und ein Bewusstsein für diese Kohärenzbildungshilfe geschaffen wird. Um solche skizzierten Implikationen für die Unterrichtspraxis zu konkretisieren, bedarf es jedoch zunächst einer Replikation der Befunde (vgl. Schmitz 2016).

Literatur

- Aiken, Leona S./West, Stephen G.: Multiple regression: Testing and interpreting interactions. Thousand Oaks. CA: Sage Publications 1991.
- Bortz, Jürgen/Schuster, Christof (Hrsg.): Statistik für Human- und Sozialwissenschaftler. Berlin: Springer 2010.
- Cohen, Jacob/Cohen, Patricia/West, Stephen G./Aiken, Leona S.: Applied multiple regression/correlation analyses for the behavioral sciences. Hillsdale, NJ: Lawrence Erlbaum 2003.
- Faul, Franz/Erdfelder, Edar/Buchner, Axel/Lang, Albert-Georg: Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behavior Research Methods (2009), 41, 1149-1160.
- Field, Andy: Discovering statistics using IBM SPSS Statistics. Los Angeles. CA: Sage Publications 2013.
- McNamara, Danielle S./Kintsch, Walter: Learning from text: Effects of prior knowledge and text coherence. Discourse Processes (1996), 22, 247-288.
- Ozuru, Yasuhiro/Dempsey, Kyle/McNamara, Danielle S.: Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. Learning and Instruction (2009), 19, 228-242.
- Richter, Tobias: Wie analysiert man Interaktionen von metrischen und kategorialen Prädiktoren? Nicht mit Median-Splits! In: Zeitschrift für Medienpsychologie (2007), 3, 116-125.
- Saunders, David R.: Moderator variables in prediction. Educational and Psychological Measurement (1956), 16, 209-222.
- Schmitz, Anke/Gräsel, Cornelia: Bei welchen Lernenden fördert globale Textkohäsion das Verstehen von Sachtexten? Eine Studie zu Wechselwirkungen zwischen globaler Textkohäsion und kognitiven Verständnisvoraussetzungen. In: Unterrichtswissenschaft Unterrichtswissenschaft (2016), 44(3), 267-281.
- Schmitz, Anke: Verständlichkeit von Sachtexten. Eine empirische Studie zur Wirkung der globalen Textkohäsion auf das Textverständnis von Schülerinnen und Schülern. Wiesbaden: VS Verlag für Sozialwissenschaften 2016.
- Schneider, Wolfgang/Schlagmüller, Matthias/Ennemoser, Marco: Lesegeschwindigkeits- und -verständnistest für die Klassen 6-12. Göttingen: Hogrefe 2007.
- Schnotz, Wolfgang: Aufbau von Wissensstrukturen. Untersuchungen zur Kohärenzbildung beim Wissenserwerb mit Texten (Fortschritte der psychologischen Forschung 20). Seelze-Velber: Beltz Psychologie Verlags Union 1994.

Kurz-Glossar

Der im Folgenden aufgeführte Kurz-Glossar soll insbesondere für empirie-unerfahrene Leserinnen und Leser dieses Bandes eine Orientierungshilfe darstellen und geläufige Begriffe der empirischen Bildungsforschung erläutern. Hierbei ist der Gedanke leitend, Verständnisschwierigkeiten vorzubeugen, wenngleich die hier gewählte Form eine Verringerung der Komplexität und Vielschichtigkeit der beschriebenen Begriffe nach sich zieht.

Benchmark

„Benchmark“ stellt einen Bezugswert oder auch Maßstab dar, welcher bei der vergleichenden Analyse von Prozessen oder Produkten verwendet wird. Das Ziel des Benchmarkings ist es dabei, Verbesserungsmöglichkeiten von Anderen zu übernehmen und von bereits gemachten Erfahrungen zu profitieren.

Deduktiv

Allgemein versteht man unter „Deduktion“ das logische Schließen auf einen Sachverhalt. Bei der Deduktion schließt man vom Allgemeinen zum Bestimmten. Auch bekannt ist es als das Schließen von der Theorie zum Einzelfall. Im Gegensatz dazu: → *induktiv/Induktion*.

Häufigkeitsverteilungen

Die „Häufigkeitsverteilung“ ist ein statistisches Verfahren, mit dem angegeben wird, wie oft ein Wert in einer Erhebung vorkommt. Man kann die „Häufigkeitsverteilung“ grafisch (in Form von Diagrammen), tabellarisch oder in einer Gleichung darstellen. Der Vorteil gegenüber der bloßen Errechnung von Mittelwerten ist, dass durch die „Häufigkeitsverteilung“ die Bandbreite der Messergebnisse ersichtlich wird.

Implementationsstudie

Eine „Implementationsstudie“ untersucht qualitativ oder quantitativ welche Effekte eine durchgeführte Maßnahme erzielt. Dazu erheben die Forschenden den spezifischen Leistungsstand einer Gruppe vor Durchführung der Maßnahme (→ Pretest), führen die Implementation durch und erheben den Leistungsstand erneut (→ Posttest). Anhand der Veränderungen lässt sich unter Berücksichtigung von Umgebungsvariablen der Effekt der Maßnahme beschreiben.

Induktiv

Die „Induktion“ ist die ursprünglichste Erkenntnisform der Naturwissenschaft. Durch (systematische) Einzelbeobachtungen wird eine allgemeingültigkeitsbeanspruchende Theorie entwickelt. Man versteht deshalb „Induktion“ auch als das Schließen vom Bestimmten zum Allgemeinen. Im Gegensatz dazu: → *deduktiv/Deduktion*.

Inferenz

Unter der ‚Inferenz‘ versteht man die abschließenden Schlussfolgerungen innerhalb einer Forschungsfrage. Der Grad der Inferenz gibt dabei an, inwieweit die Forschungsfrage allgemeingültigen Ansprüchen genügt. Man spricht von einer hohen ‚Inferenz‘, wenn das Ergebnis auf eine größere Grundgesamtheit übertragen werden kann. Eine niedrige ‚Inferenz‘ liegt hingegen vor, wenn die Konklusion lediglich in einer spezifischen Situation zutrifft.

Intersubjektivität

‚Intersubjektivität‘ meint das, was zwischen zwei Subjekten vorhanden, d.h. das, was ihnen beiden einsichtig, ist. ‚Intersubjektiv‘ kann z.B. Wissen durch Sprache vermittelt werden. Ein Sachverhalt wird also einem Subjekt durch ein anderes Subjekt mitgeteilt. Der konstruktivistischen Theorie nach kann aber niemals geklärt werden, ob das Wissen von beiden Personen gleich aufgenommen und verstanden wurde (kritische Medientheorie). Abzugrenzen ist ‚Intersubjektivität‘ von Objektivität (das, was unabhängig gegeben ist) sowie von Subjektivität (das, was nur für ein Subjekt gegeben ist).

Items

Der Begriff des ‚Item‘ bezeichnet eine als Frage oder Urteil formulierte Äußerung, zu der Probandinnen und Probanden in → standardisierten oder nicht standardisierten Verfahren Stellung nehmen sollen. So besteht beispielsweise ein Test aus mehreren ‚Items‘ zu einem oder mehreren Themen.

Kategoriensystem

Das ‚Kategoriensystem‘ ist ein zentraler Punkt der quantitativen und qualitativen Inhaltsanalyse sowie von kriteriengeleiteten Ratings. Im ‚Kategoriensystem‘ können die in der Analyse vorgenommenen Auswahlentscheidungen veranschaulicht und → intersubjektiv nachvollziehbar gemacht werden. Hierdurch werden die Ergebnisse vergleichbar, was sich positiv auf die → Reliabilität auswirkt. Die Vorgehensweise zur Definition der Kategorien muss anhand des Materials entschieden werden – beispielsweise können die Kategorien induktiv aus dem Material extrahiert oder vorab deduktiv festgelegt werden.

Kodierleitfaden

Der ‚Kodierleitfaden‘ ist die Grundlage für die → Kodierung. Der ‚Kodierleitfaden‘ umfasst für jede Analysekategorie eine Definition, ein Ankerbeispiel aus den Daten und Grundsätze, welche für eine Textsequenz gelten, um der Kategorie zugeordnet zu werden. Durch diesen ‚Kodierleitfaden‘ können alle Rezipientinnen und Rezipienten die Analyse des Materials nachvollziehen.

Kodierung

Die ‚Kodierung‘ bezeichnet den Prozess der Zuordnung des Datenmaterials zu den einzelnen Kategorien. Dabei wird der Rezipientin bzw. dem Rezipienten die Fundstelle und die Begründung für die Zuordnung offengelegt (→ Kodierleitfaden).

Korpus

Der für die Analyse des Forschungsprojekts relevante Datensatz (etwa Transkripte der erhobenen Interviews, alle Lehrpläne für das Fach Deutsch oder andere Zusammenstellungen von Medien) wird ‚Korpus‘ genannt. Diese Daten werden in der Auswertungsphase analysiert.

Korrelation

Die ‚Korrelation‘ steht für den Zusammenhang zwischen zwei oder mehreren Merkmalen, welche sich wechselseitig bedingen und damit in Abhängigkeit zueinander stehen. Dieses Verhältnis wird mit dem Korrelationskoeffizienten r angegeben, welcher die Werte von -1 über 0 bis 1 annehmen kann. $r=0$ bedeutet, dass kein Zusammenhang zwischen den untersuchten Merkmalen besteht, wohingegen $r=1$ auf eine absolute Abhängigkeit verweist (d.h. beide Merkmale treten immer gemeinsam auf). $r=-1$ drückt hingegen aus, dass ein absoluter negativer Zusammenhang besteht. Allgemein kann man bei $r < 0,5$ von einer geringen, bei $0,51 < r < 0,7$ von einer mittleren und bei $0,71 < r < 0,9$ von einer hohen ‚Korrelation‘ sprechen. Die Zuverlässigkeit der ‚Korrelationen‘ drückt sich zudem durch ihre \rightarrow Signifikanz aus.

Methodologie

Die ‚Methodologie‘ ist die Lehre von den Methoden. Oftmals bestimmt der Forschungsgegenstand die passende Methode. Die ‚Methodologie‘ definiert also Kriterien, die festlegen, wann welche Methode angebracht ist. Man könnte bildlich gesprochen auch sagen, Methoden sind Werkzeuge und die ‚Methodologie‘ ist die Bedienungsanleitung dieser Werkzeuge.

Normierung

Daten, die in standardisierten Tests erhoben werden, können (sofern die Stichprobe (\rightarrow Normierungsstichprobe) hinreichend groß ist) als repräsentativ für eine spezifische Gruppe gelten. Dies führt zu einem Vergleichswert, der als Norm für andere Erhebungen gelten kann. So liegen etwa für standardisierte Lesetests wie dem *SLS 5-8* normierte Vergleichsskalen zu verschiedenen Jahrgangsstufen und Schultypen vor, anhand derer die Leistung anderer Schülerinnen und Schüler verglichen werden kann.

Normierungsstichprobe

‚Normierungsstichproben‘ treten bei standardisierten Testverfahren auf und sind für deren Eichung entscheidend. Hierfür wird der ausgewählte Test einer möglichst großen Personenanzahl vorgelegt, deren Merkmale (z.B. Alter, Geschlecht, Leistungsstand) mit denen der späteren Testzielgruppe weitestgehend übereinstimmen. Damit erhält der Forschende normierte Werte, die es ihm ermöglichen, die Leistungen zukünftiger Testteilnehmerinnen und -teilnehmer einzuordnen (\rightarrow Normierung).

Objektivität

Die ‚Objektivität‘ ist neben \rightarrow Validität und \rightarrow Reliabilität eines der Hauptgütekriterien der empirischen Forschung. ‚Objektivität‘ sorgt hierbei dafür, dass die Versuchsergebnisse weitestgehend subjektunabhängig sind. Subjektunabhängigkeit meint die Unveränderbarkeit der Ergebnisse bei Durchführung des Experiments durch verschiedene Versuchsleiterinnen und -leiter.

Pilotierung

Die ‚Pilotierung‘ kann als Vortest bereits konzipierter Erhebungsinstrumente verstanden werden, welcher vor der eigentlichen Erhebung erfolgt. Hierbei soll beispielsweise die Verständlichkeit oder Verwendbarkeit eines Fragebogens durch Testpersonen (welche nach Möglichkeit der Stichprobe ähneln) überprüft werden. Die ‚Pilotierung‘ dient damit der Wirksamkeitskontrolle eines Vorhabens und kann zu einer Modifizierung bzw. Verbesserung des Erhebungsinstruments führen.

Pretest

Der ‚Pretest‘ ist der Test vor der Durchführung einer Implementation (→ Implementationsstudie) zur Ermittlung des Lern-/bzw. Leistungsstandes der Probandengruppe.

Posttest

Ein ‚Posttest‘ kann als Nachtest einer bereits durchgeführten Erhebung angesehen werden und dient der Erfolgskontrolle der Maßnahme (→ Implementationsstudie). Durch dessen Durchführung können Informationen über das Veränderungsmaß der beteiligten Probandinnen und Probanden gewonnen werden, welche zur Reflexion und gegebenenfalls Modifikation des Vorgehens beitragen.

Qualitative Forschung

Die ‚qualitative Forschung‘ zeichnet sich insbesondere dadurch aus, dass sie das einzelne Subjekt fokussiert. Im Rahmen der ‚qualitativen Forschung‘ gibt es für einen Forschungsgegenstand eine Vielzahl an verschiedenen Betrachtungsmöglichkeiten. Verwendet werden kommunikative und dialogische Methoden, da die Grundannahme besteht, dass die Umwelt durch soziale Aushandlungsprozesse und Interaktionen entsteht.

Quantitative Forschung

Die ‚quantitative Forschung‘ beschäftigt sich mit dem systematischen Messen und numerischen Auswerten zur Überprüfung aufgestellter Hypothesen bei großen Stichproben. Anders als in der qualitativen Forschung werden meist standardisierte Verfahren verwendet. Es sollen möglichst objektiv Zusammenhänge entdeckt und Schlussfolgerungen verifizierbar oder falsifizierbar gemacht werden, sodass sich die Ergebnisse auf eine Grundgesamtheit übertragen lassen.

Regression

Die ‚Regression‘ bezeichnet den Zusammenhang einer abhängigen Variable mit (mindestens) einer unabhängigen Variable. Die abhängige Variable ändert sich → korrelativ zur Veränderung der unabhängigen Variable. Andererseits ändert sich die unabhängige Variable aber nicht korrelativ zur Veränderung der abhängigen Variable. Mithilfe der unabhängigen Variablen ist es also möglich, Prognosen für die Veränderung der abhängigen Variablen zu berechnen.

Rekonstruktion

Die ‚Rekonstruktion‘ beschreibt den Vorgang sowie das Ergebnis eines erneuten Nachvollziehens von Erlebnissen der Probandengruppe mit dem Ziel der Generierung von (subjektiven) Theorien oder von bereits in der Vergangenheit erarbeiteten Theorien, Objekten oder Werken. Die Lehr-Lernforschung legt dabei den Fokus auf die wechselseitige Abstimmung von fachlichem Wissen und Alltagsvorstellungen (etwa der Schülerinnen und Schüler oder auch Lehrerinnen und Lehrer), um einen möglichst lernförderlichen Unterricht zu konzipieren. Eine weitere Betrachtungsweise liefert die empirisch-analytische Wissenschaftstheorie. In ihrem Sinne wird die Rekonstruktion als das Nachvollziehen einer Theorie durch logische Analysen und deduktives Schließen aus Axiomen (anerkannte Annahmen, welche keine weitere Begründung benötigen) verstanden.

Rekonstruktive Sozialforschung

Im Prozess der ‚rekonstruktiven Sozialforschung‘ wird die soziale Wirklichkeit verstehend und interpretierend analysiert (→ soziale Wirklichkeit). Der Prozess wird als rekon-

struktiv beschrieben, da die Wirklichkeit von den in ihr handelnden Subjekten vorstrukturiert wird. Diese Konstrukte sollen im Rahmen der ‚rekonstruktiven Sozialforschung‘ interpretiert werden.

Reliabilität

Die ‚Reliabilität‘ ist neben \rightarrow Validität und \rightarrow Objektivität eines der Hauptgütekriterien der empirischen Forschung und gibt an, wie zuverlässig eine durchgeführte Messung ist. Die Zuverlässigkeit bedeutet, dass bei Untersuchungswiederholungen, unter Beibehaltung der Erhebungsmethode und möglichst gleichbleibenden Rahmenbedingungen, die Forschungsergebnisse stabil bleiben oder übereinstimmen.

Signifikanz

Die ‚Signifikanz‘ beschreibt, mit welcher Wahrscheinlichkeit ein in einer Hypothese formulierter Zusammenhang (\rightarrow Korrelation) dem Zufall geschuldet ist. Die ‚Signifikanz‘ wird als p-Wert (Irrtumswahrscheinlichkeit) angegeben. Die Beziehung zwischen zwei Merkmalen ist dann signifikant, wenn der p-Wert unter einer Überschreitungswahrscheinlichkeit liegt. Die Überschreitungswahrscheinlichkeit liegt in den meisten Fällen bei 5%. Ist dies der Fall, so ist die Abhängigkeit zwischen den untersuchten Merkmalen anzunehmen, da die Zuschreibung für ein zufälliges Auftreten zu gering ausfällt. Bei einem Wert von $p < 0,01$ gilt der Zusammenhang als hochsignifikant. Liegt der p-Wert über der Überschreitungswahrscheinlichkeit, besteht keine Verbindung zwischen den untersuchten Merkmalen. In diesem Fall muss eine Alternativhypothese aufgestellt werden.

Skalenniveaus

Das ‚Skalenniveau‘ wird auch als ‚Messniveau‘ bezeichnet. Skalen geben an, in welchen Beziehungen bestimmte Aussagen stehen können. Sie werden entweder bei der Analyse erhobener Daten von Seiten des Forschenden oder bereits von dem Befragten zur Gewichtung seiner Aussagen herangezogen. Die meistverwendete Klassifikation der Skalenniveaus geht auf Stevens (1946) zurück:

- Nominal-Skala: Zahlenwert steht für die Einteilung in eine bestimmte Klasse, wobei jedes Objekt lediglich einer Klasse zugeordnet werden kann (bspw. Geschlecht).
- Ordinal-Skala: Objekte werden nicht nur klassifiziert, sondern auch in eine Rangordnung gebracht (bspw. Tabellenplatz, Schulnoten).
- Intervall-Skala: Objekte werden nicht nur klassifiziert und geordnet, sondern auch die Differenz zwischen ihnen bestimmt (bspw. Intelligenzquotient).
- Relations-Skala: Objekte werden nicht nur klassifiziert, geordnet und bestimmt, sondern auch in Verhältnis zueinander gesetzt (bspw. Body-Mass-Index).

Soziale Erwünschtheit

Die ‚Soziale Erwünschtheit‘ kennzeichnet sich durch das Handeln auf Grund einer angenommenen gesellschaftlichen oder personalen Norm. Ein Individuum tätigt beispielsweise eine Äußerung nicht wahrheitsgemäß, sondern angesichts einer scheinbaren Vorgabe der gewünschten Antwort durch den gesellschaftlichen oder personellen Kontext. Aussagen, die durch ‚soziale Erwünschtheit‘ beeinflusst werden, stellen damit ein Problem der wissenschaftlichen Datenerhebung dar.

Soziale Wirklichkeit

Die ‚soziale Wirklichkeit‘ entsteht aus menschlichen Handlungen und Auseinandersetzungen mit anderen Individuen im Alltag. Sie ist die vom Menschen erfahrbare Wirklichkeit. Folgt man der soziologischen Argumentation, ist sie der Bestandteil der Wirklichkeit, welche durch empirische Erhebungen abgebildet werden kann.

SPSS

‚IBM SPSS‘ ist eine Statistik- und Analysesoftware. Sie kann genutzt werden, um quantitativ erhobene Daten auszuwerten. Dabei haben die Nutzenden die Möglichkeit, die gängigsten statistischen Verfahren auf die erhobenen Daten anzuwenden. Zusätzlich kann man die ausgewerteten Daten graphisch aufbereiten.

Standardabweichung

Bei der ‚Standardabweichung‘ (SD/SA) handelt es sich um einen statistischen Kennwert. Die ‚Standardabweichung‘ zeigt an, wie stark die durchschnittliche Abweichung der einzelnen Ergebnisse vom Mittelwert ausfällt. Eine hohe ‚Standardabweichung‘ deutet auf eine breite Streuung der Ergebnisse hin, während eine geringe ‚Standardabweichung‘ auf eine Verteilung der Werte nahe des Mittelwerts verweist.

Standardisiert – nicht Standardisiert

Die Begriffe ‚standardisiert‘ und ‚nicht standardisiert‘ werden zumeist im Kontext der Methodengroßformen ‚Befragung‘ und ‚Beobachtung‘ verwendet. Der Grad der ‚Standardisierung‘ gibt an, in welchem Maß Befragungen oder Beobachtungen nach festgelegten oder freien Kriterien gestaltet werden können. Bei ‚standardisierten mündlichen Befragungen‘ werden etwa die Probandinnen und Probanden zu zuvor festgelegten Fragen in einem standardisierten Wortlaut von einem Testleitenden befragt, wohingegen eine ‚nicht standardisierte mündliche Befragung‘ aus einem Gespräch zwischen Testperson und Testleitenden erwächst. Unterstützend kann hierzu ein Leitfaden verwendet werden.

t-Test

Der ‚t-Test‘ wird auch als ‚Mittelwertdifferenztest‘ bezeichnet und dient der Überprüfung von aufgestellten Hypothesen. Mit dem ‚t-Test‘ wird die Wahrscheinlichkeit berechnet, ob eine gefundene Mittelwertdifferenz zwischen zwei Stichproben auf die Forschungsfrage und nicht auf den Zufall zurückzuführen ist. Grundlegend ist die Annahme, dass die Stichproben aus einer normalverteilten Grundgesamtheit stammen.

Theoretische Sättigung

Die ‚Theoretische Sättigung‘ beschreibt einen Zustand im Rahmen des qualitativen Forschungsprozesses, in dem die Hypothese nicht weiter modifiziert bzw. durch weitere Untersuchungen spezifiziert werden muss, da der Einbezug neuer Fälle keinen Anhaltspunkt für andere Einflussfaktoren bieten würde.

Transkript

In einem ‚Transkript‘ werden aufgenommene verbale, je nach Transkriptionsvorschrift auch nonverbale, Äußerungen verschriftet. Die Namen der aufgenommenen Personen werden anonymisiert. Die Regeln der Verschriftung, das bedeutet welche Zeichen für Pausen etc. verwendet werden, müssen im Voraus festgelegt werden. Besondere Auffälligkeiten während des Interviews werden ebenfalls im ‚Transkript‘ vermerkt (s. hierzu den Exkurs bei Heller/Morek in diesem Band).

Treatment

Unter ‚Treatment‘ versteht man den Umgang mit Versuchspersonen während einem Experiment. Mit dem ‚Treatment‘ soll eine Einstellung oder Handlung erzeugt, hervorgerufen und/oder verändert werden, die man dann im Experiment beobachten kann. Daraufhin werden die aufgestellten Hypothesen geprüft.

Triangulation

Der Ausdruck der ‚Triangulation‘ bezeichnet die vielfältige Betrachtung eines Forschungsgegenstandes. Verschiedene Methoden sollen bei der Annäherung an einen Gegenstand genutzt werden, um dessen Facetten ganzheitlich erfassen zu können. Heute versteht man unter ‚Triangulation‘ jedoch nicht nur die Verwendung und Verbindung mindestens zweier Methoden, sondern ebenso eine multiperspektivische Betrachtung der Forschungsmaterie. So können neben verschiedenen Theorien, auch mehrere Datenquellen oder Forscherteams zur vieldimensionalen Wahrnehmung herangezogen werden.

Typik/Typenbildung

Es handelt sich bei der ‚Typik‘/‚Typenbildung‘ um einen Arbeitsschritt im Rahmen von inhaltsanalytischen Verfahren. Bei dieser Form werden bedeutende Sequenzen des Datenmaterials hervorgehoben und genauer analysierend beschrieben. Aus diesen Sequenzen werden Typen gebildet. Diese Typen gelten dann als Prototyp einer Kategorie und werden beschrieben, definiert und interpretiert. Das birgt die Gefahr, dass die Pluralität des Datenmaterials nicht mehr beachtet wird.

Validität

Die ‚Validität‘ ist neben → Reliabilität und → Objektivität eines der Hauptgütekriterien der empirischen Forschung. Die Beurteilung der ‚Validität‘ kann eine Aussage darüber machen, ob eine empirische Erhebung die Merkmale, Konstrukte etc. erhebt, welche sie zu erheben vorgibt. Nachgewiesen wird die ‚Validität‘ über die Korrelation zwischen Merkmal und Erhebungsinstrument.

Variable

Unter dem Begriff der ‚Variable‘ versteht man die möglichen Ausprägungen spezifischer Eigenschaften, obgleich sie durch diverse Gesichtspunkte unterschieden werden können. Ein Beispiel hierfür ist die Bezeichnung der manifesten oder der latenten ‚Variable‘. Direkt beobachtbare Eigenschaften, wie z.B. die Körpergröße, werden als manifest bezeichnet. Latente ‚Variablen‘ meinen hingegen nicht direkt beobachtbare Größen wie das Abstraktionsvermögen.

Varianzanalysen

Unter dem Begriff der ‚Varianzanalyse‘ werden eine Vielzahl an Varianten dieser Form gefasst. Allen gemeinsam ist, dass sie ein Verfahren beschreiben, welches die errechneten Mittelwerte mehrerer Erhebungsdurchläufe vergleicht. Mit diesem Verfahren soll die mögliche Beeinflussung und Zusammenwirkung verschiedener unabhängiger Variablen (werden in den verschiedenen Erhebungen variiert) auf eine abhängige Variable (bleibt gleich in allen Erhebungsdurchläufen) überprüft werden.

Was ist das richtige empirische Verfahren für mein Forschungsprojekt? Der vorliegende Band liefert eine einsteigerfreundliche, umfassende Einführung in die Erhebung und Auswertung empirischer Daten im Rahmen deutschdidaktischer Forschung. Hierbei werden in 21 thematischen Kapiteln verschiedene Erhebungs- und Auswertungsverfahren vorgestellt und anschließend ein Einblick in die praktische Arbeit mit diesem Verfahren anhand eines aktuellen Forschungsprojekts geboten.

Ein Grundlagenwerk für Studierende und Promovierende.



Dr. Jan M. Boelmann ist Professor für Literatur- und Mediendidaktik.

Seine Arbeits- und Forschungsschwerpunkte liegen im medienübergreifenden literarischen Lernen, der Modellierung literarischer Kompetenz sowie in der Kinder- und Jugendliteratur.



Schneider Verlag Hohengehren