

# **Erfassung von Experimentierkompetenz im universitären Kontext**

Entwicklung und Validierung eines Experimentiertests zum  
Themenbereich Optik

**Dissertation**

zur Erlangung des Doktorgrades  
der Mathematisch-Naturwissenschaftlichen Fakultät  
der Christian-Albrechts-Universität zu Kiel

vorgelegt von

**Jan Heidrich**

**Kiel, 2017**

**Erster Gutachter:**

Prof. Dr. Knut Neumann

**Zweiter Gutachter:**

Prof. Dr. Hendrik Härtig

**Tag der mündlichen Prüfung:**

16.12.2016

## **Zusammenfassung**

Der Erwerb von Experimentierkompetenz stellt ein wichtiges Ziel der universitären Ausbildung im Fach Physik dar. Dennoch wird bei Untersuchungen zum Lernerfolg im Physikstudium gerade dieser Aspekt vernachlässigt. Dies liegt insbesondere daran, dass eine valide Erfassung von Experimentierkompetenz bisher nur eingeschränkt gelungen ist. Das Ziel der vorliegenden Arbeit ist die Entwicklung eines Instruments zur validen Erfassung von Experimentierkompetenz.

In einem ersten Schritt wurde ermittelt, welche Fähigkeiten und Fertigkeiten repräsentativ für die experimentelle Ausbildung sind. Dazu wurde ein Modell mit notwendigen Fähigkeiten und Fertigkeiten beim Experimentieren adaptiert und Experimentierkompetenz modelliert.

Darauf aufbauend wurden Ansätze zur Diagnostik von Experimentierkompetenz analysiert und bzgl. der Testgüte beurteilt. Aus dieser Analyse wurden Anforderungen an ein Testinstrument zur reliablen und objektiven, vor allem aber validen Erfassung von Experimentierkompetenz abgeleitet. In einem zweiten Schritt wurde basierend auf einer Fachinhaltsanalyse von  $N=10$  deutschsprachigen Praktika unter Berücksichtigung der Anforderungen an Testinstrumente, ein inhaltlich valider Experimentiertest mit sechs Aufgaben entwickelt. Mit dem Experimentiertest soll sich der Zuwachs von Experimentierkompetenz im Bereich Optik erfasst lassen. Anschließend wurde mit einer Think Aloud Studie mit  $N=16$  Probanden geprüft, inwieweit die kognitiven Prozesse der Probanden durch die Auswertung des Laborhefts (Produkt) bzw. die zeitaufwändigere Auswertung des Vorgehens beim Experimentieren mittels Videoanalyse (Prozess) valide abgebildet werden können. Aus den Erkenntnissen der Think Aloud Studie wurde ein ökonomisches und valides Auswerteverfahren zur Beurteilung von Experimentierkompetenz entwickelt. Im letzten Schritt wurde geprüft, inwieweit das entwickelte Testinstrument den Ansprüchen einer validen, reliablen und objektiven Diagnostik von Experimentierkompetenz genügen kann. Dazu wurde eine Feldstudie im physikalischen Praktikum mit  $N=58$  Studierenden durchgeführt. Der Schwerpunkt dieser Untersuchung lag auf der strukturellen- und externen Validität. Dazu wurde in Zusammenarbeit mit dem KiL Projekt ein Fachwissenstest mit 15 Aufgaben entwickelt und Testinstrumente zur Selbstwirksamkeit und des Strategiewissens eingesetzt.

Im Rahmen der Arbeit ist es gelungen Probleme bei der Erfassung von Experimentierkompetenz zu identifizieren und Lösungsansätze zu erarbeiten. Es konnte ein objektives Testinstrument entwickelt werden, das Experimentierkompetenz als Gesamtkonstrukt, sowie die Teilkomponenten Richtigkeit und Zielorientiertheit mit einer zufriedenstellenden Reliabilität abbilden kann. Die Erkenntnisse der Analyse des Testinstruments lassen den Schluss zu, dass

die Erfassung von Experimentierkompetenz inhaltlich valide und im Bereich Optik auf deutschsprachige Praktika verallgemeinerbar ist. Es konnten ausreichende Kennwerte für die interne Konsistenz von authentischen Realexperimentiertests, auch für einzelne Subskalen (zwei von drei) nachgewiesen werden. Zudem wurden erste Hinweise zur Unabhängigkeit der Teildimensionen der Experimentierkompetenz gefunden. Bei der Analyse der externen Validität konnte bedingt durch fehlende, zur Verfügung stehende Testinstrumente nur unzureichende Erkenntnisse gesammelt werden. Für das zeitaufwendige prozessorientierte Auswerteverfahren mittels Videoanalyse und lautem Denken konnte kognitive Validität bestätigt werden. Es liegen Hinweise vor, dass auch das darauf aufbauend entwickelte ökonomische Auswerteverfahren kognitiv valide ist. Eine Bestätigung dieser Annahme konnte im Rahmen der Arbeit nicht realisiert werden. Gleiches gilt für die prädiktive Validität, deren Analyse eine Langzeitstudie voraussetzt.

## **Abstract**

The acquisition of the ability to conduct scientific investigations is an important goal of university education in the field of physics. However, this aspect is neglected in researches on the learning success in physics. This is because of the fact that a valid measuring of the ability to conduct scientific investigations has only been possible to a limited extent so far. The aim of this thesis is the development of an instrument for the valid measurement of the ability to conduct scientific investigations.

In the first step representative skills and abilities to conduct scientific investigations in university education were identified. For this purpose, a model with the necessary skills and abilities to conduct scientific investigations was adapted and the ability to conduct scientific investigations was modeled. Based on this approaches to the diagnostics of the ability to conduct scientific investigations were analyzed and assessed with regard to the test quality. This analysis results in requirements for a test instrument to reach a reliable, objective and valid diagnostic. The focus was especially on the valid assessment of the ability to conduct scientific investigations.

In a second step, a hands on performance test with six tasks was developed. It is based on a content analysis of  $N = 10$  German-speaking laboratory and the requirements for test instruments. These development lead to the conclusion that content based validity is given.

The purpose of this performance assessment is to measure the increase of the ability to conduct scientific investigations in the field of optics. By use of a Think Aloud study with  $N = 16$  participants the fitting of the cognitive processes was analyzed. Therefore the evaluation based on the laboratory book (product) and the more time-consuming evaluation of the procedure during experimentation by video analysis (process) were compared. The results of the Think Aloud study lead to an economic and valid evaluation procedure for the assessment of the ability to conduct scientific investigations.

The last step examined the extent to which the developed performance assessment can meet the requirements of a valid, reliable and objective diagnosis of the ability to conduct scientific investigations. For this purpose, a field study was conducted in the physical laboratory with  $N = 58$  students, whose focus was the investigation of structural and external validity. In collaboration with the KiL project, a knowledge test with 15 tasks was developed and test instruments for self-efficacy and strategic knowledge were used.

Within the scope of the work, problems in the measurement of the ability to conduct scientific investigations have been identified and solution approaches have been developed. An objective test instrument could be developed, which can be used as an overall construct, as well as

the subcomponents can represent correctness and goal-orientation with a satisfactory reliability.

The cognitions of the analysis of the test instrument lead to the conclusion that the comprehension of the ability to conduct scientific investigations is valid in terms of content and generalizable to the field of optics on German-speaking laboratories. Sufficient parameters for the internal consistency of the overall construct and individual subscales (two out of three) have been demonstrated. First indications of the independence of the partial dimensions of the ability to conduct scientific investigations could be found. In the analysis of the external validity, insufficient cognitions could be gathered due to the lack of available test instruments. Cognitive validity has been confirmed for the time-consuming evaluation. There are indications that the developed, economical evaluation method is cognitively valid too. A confirmation of this assumption could not be realized within the scope of the work. The same applies to the predictive validity, whose analysis requires a long-term study.

## **Erklärung**

Die vorgelegte Arbeit ist – abgesehen von der Beratung durch meinen Betreuer – nach Inhalt und Form meine eigene Arbeit. Weiterhin wurde die Arbeit an keiner anderen Stelle zum Prüfverfahren vorgelegt. Teile der Arbeit wurde in Fachzeitschriften veröffentlicht. Bei der Anfertigung der Arbeit wurden die Regeln guter wissenschaftlicher Praxis nach der Deutschen Forschungsgemeinschaft nach meinem besten Wissen und Gewissen eingehalten.

*Jan Heidrich*

# Inhalt

1 Einleitung .....	11
2 Modellierung von Experimentierkompetenz .....	14
2.1 Fähigkeiten, Fertigkeiten und Handlungen im Experimentierprozess .....	15
2.1.1 Fragestellung bzw. Problemstellung entwickeln oder klären .....	18
2.1.2 Hypothesen generieren/ Klären der theoretischen Grundlage .....	19
2.1.3 Experiment planen .....	20
2.1.4 Versuch fehlerfrei aufbauen .....	21
2.1.5 Messen, Beobachten und Dokumentieren .....	21
2.1.6 Daten analysieren und Verallgemeinerbarkeit diskutieren .....	22
2.1.7 Sachgerechte Schlüsse ziehen und diese kommunizieren .....	22
2.1.8 Anordnung der Handlungen im Experimentierprozess .....	23
2.2 Qualitätsmaße der Experimentierkompetenz .....	24
2.2.1 Richtigkeit .....	25
2.2.2 Strukturiertheit .....	26
2.2.3 Zielorientiertheit .....	28
2.3 Zusammenführung des Modells der Experimentierkompetenz .....	30
3 Erfassung von Experimentierkompetenz .....	32
3.1 Aspekte der Erfassung von Experimentierkompetenz .....	33
3.1.1 Bewertungsgegenstand: Produkt vs. Prozess .....	34
3.1.2 Aufgabenformat: Praktisch vs. Theoretisch .....	37
3.1.3 Modellierung: Einzelne Handlungen vs. Gesamter Prozess .....	38
3.2 Klassifizierung von Testinstrumenten .....	40
3.2.1 Authentische Testinstrumente .....	41
3.2.2 Eher authentische Testinstrumente .....	43
3.2.3 Eher ökonomische Testinstrumente .....	46
3.2.4 Ökonomische Testinstrumente .....	47
3.3 Güte von Testinstrumenten .....	48
3.3.1 Objektivität .....	50
3.3.2 Reliabilität .....	52
3.3.3 Validität .....	53
3.3.4 Ökonomie .....	59
3.4 Forschungsdefizit .....	59
4 Forschungsziel .....	61
5 Anforderungen an ein Testinstrument zur Erfassung von Experimentierkompetenz .....	62



5.1 Inhaltsvalidität .....	62
5.2 Generalisierbarkeit .....	64
5.3 Strukturelle Validität .....	65
5.4 Externe Validität .....	66
5.5 Prädiktive Validität.....	67
5.6 Kognitive Validität .....	68
6 Studie I – Fachinhaltsanalyse .....	69
6.1 Forschungsfrage .....	69
6.2 Studiendesign .....	69
6.3 Kategoriensystem .....	70
6.4 Datenanalyse .....	71
6.5 Ergebnisse: Fachinhaltsverteilung in deutschsprachigen physikalischen Anfängerpraktika .....	72
6.6 Diskussion .....	75
7 Entwicklung eines Testinstruments zur Erfassung von Experimentierkompetenz .....	76
7.1 Testumfeld.....	76
7.2 Aufgaben.....	77
7.2.1 Experimentiertest.....	81
7.2.2 Beispielaufgabe .....	82
7.3 Auswerteverfahren .....	85
7.3.1 Bewertungseinheiten .....	86
7.3.2 Maße der Experimentierkompetenz.....	93
8 Studie II – Qualitative Untersuchung mittels Lautem Denken .....	103
8.1 Forschungsfragen zu den qualitativen Aspekten der Testgüte .....	103
8.1.1 Objektivität.....	103
8.1.2 Reliabilität .....	103
8.1.3 Inhaltsvalidität.....	104
8.1.4 Kognitive Validität .....	104
8.2 Studiendesign .....	105
8.3 Stichprobe .....	107
8.4 Ergebnisse Studie II.....	108
8.4.1 Objektivität.....	108
8.4.2 Reliabilität .....	108
8.4.3 Inhaltsvalidität.....	113
8.4.4 Kognitive Validität .....	120
8.5 Diskussion .....	130

8.6 Überarbeitung des Testinstruments .....	134
9 Studie III – Quantitative Untersuchung mittels Feldtest.....	137
9.1 Forschungsfragen zu den quantitativen Aspekten der Testgüte .....	137
9.1.1 Objektivität.....	137
9.1.2 Reliabilität und Strukturelle Validität .....	137
9.1.3 Externe Validität.....	138
9.2 Studiendesign .....	139
9.3 Stichprobe .....	141
9.4 Instrumente.....	141
9.5 Ergebnisse Studie III .....	144
9.5.1 Objektivität.....	144
9.5.2 Reliabilität und Strukturelle Validität .....	144
9.5.3 Externe Validität.....	146
9.6 Diskussion .....	149
10 Fazit .....	152
11 Literaturverzeichnis.....	160
12 Abbildungsverzeichnis .....	169
13 Tabellenverzeichnis.....	171
A.1 Experimentierkompetenz: Fähigkeiten und Fertigkeiten nach vorliegenden Modellierungen .....	172
A.2 Klassifizierung von Experimentiertests .....	183
A.3 Kategorienbildung Fachinhaltsanalyse .....	188
A.4 Experimentiertes.....	192
A.4.1 Theoretische Hinweise.....	192
A.4.2 Inventarliste .....	199
A.4.3 Experimentiertest Version 1 .....	201
A.4.4 Experimentiertest Version 2 .....	220
A.4.5 Laborheft-App Konzept.....	233
A.4.6 Kodiermanual produktorientierte Auswertung.....	238
A.4.7 Kodiermanual prozessorientierte- und Referenzauswertung.....	257
A.4.8 Kodiermanual Protokollierung .....	275
A.4.9 R-Skript zur Auswertung.....	295
A.4.10 Manual zur Testeinführung Think Aloud .....	299
A.4.11 Manual zur Testeinführung Feldstudie .....	302
A.4.12 Globalmaße zur Experimentierkompetenz.....	304
A.5 Fachwissenstest .....	305

A.6 Lautes Denken: Gesamtleistung der Probanden über alle Auswertungsverfahren .....	313
A.7 Lautes Denken Studie: Handlungsverteilung über alle Aufgaben.....	314
A.8 Forschungsdefizite.....	332

## 1 Einleitung

Zentrales Ziel der naturwissenschaftlichen Ausbildung von Schülerinnen und Schülern ist, dass diese lernen *erfolgreich zu experimentieren*, kurz Experimentierkompetenz erlangen (Gesellschaft für Didaktik des Sachunterrichts, 2002, Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005c, Lunetta, 1998; National Research Council, 2012). Dadurch sollen die Schülerinnen und Schüler unter anderem befähigt werden, in der weiteren beruflichen Laufbahn den Anforderungen naturwissenschaftlicher Berufe gerecht zu werden (Rauner & Spöttl, 2002; Frank, 2014). Gleichzeitig stellt das Experiment die zentrale Methode naturwissenschaftlicher Erkenntnisgewinnung dar (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005a; Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005b; Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005c). Als mündige Bürger sollen Lernende nach der naturwissenschaftlichen Ausbildung in der Lage sein eigenständig Erkenntnisse zu gewinnen, sich (auf diesem Weg) Wissen anzueignen und die Güte generierter Evidenz zu bewerten (Ministeriums für Bildung und Wissenschaft, 2014).

Untersuchungen zeigen, dass Schülerinnen und Schüler im Physikunterricht wenig Gelegenheit zum eigenständigen Experimentieren erhalten (Lunetta, 1998; Schecker, 2001; Tesch, 2005), so dass sie häufig gerade keine Experimentierkompetenz erwerben (Hofstein & Lunetta, 2004). Umso wichtiger ist es, dass die Lernenden in der universitären naturwissenschaftlichen Ausbildung entsprechende Lerngelegenheiten erhalten. Nach Welzel (1998) stellt das Erlernen bzw. Vertiefen der Experimentierkompetenz das zentrale Ziel der praktischen universitären Ausbildung in der Naturwissenschaft dar.

In den Studienordnungen für Naturwissenschaften ist deutschlandweit festgeschrieben, dass Studierende selbstständig mit Experimenten interagieren müssen. Dies findet vornehmlich im Rahmen von Experimentierpraktika<sup>1</sup> statt. In der universitären Ausbildung in Physik ist das (physikalische) Anfängerpraktikum bzw. Grundlagenpraktikum üblicherweise das erste Experimentierpraktikum, an dem Studierende teilnehmen. Allerdings wird seit Jahrzehnten kritisch hinterfragt, ob die Studierenden durch die Experimentierpraktika Experimentierkompetenz in ausreichendem Maße erwerben (Woolnough, 1983; Lunetta, 1998; für eine Übersicht vgl. Neumann, 2004). Als Reaktion hierauf wurden verschiedene Ansätze zur Verbesserung der

---

<sup>1</sup> In der gängigen Literatur wird nur von Praktika gesprochen. Zur Abgrenzung von Schulpraktika wird hier der Begriff der Experimentierpraktika verwendet.

Experimentierpraktika entwickelt (Theyßen, 1999; Neumann, 2004; Schumacher & Planinšič, 2007). Allerdings liegen bisher keine gesicherten Erkenntnisse zur Entwicklung der Experimentierkompetenz vor. Dies gilt sowohl für herkömmliche Experimentierpraktika als auch für die Ansätze zur Verbesserung.

Hauptursache für das Defizit an fundierten Erkenntnissen zur Entwicklung von Experimentierkompetenz ist ein Mangel an geeigneten Testinstrumenten. Zwar liegen eine Vielzahl von Testinstrumenten zur Erfassung von Experimentierkompetenz vor (z.B. Hammann, Phan, Ehmer, & Grimm, 2008; Maiseyenko, Nawrath, & Schecker, 2011; Schreiber, 2012; Emden, 2011; Baxter & Shavelson, 1994; Shavelson, Solano-Flores, & Ruiz-Primo, 1998; Gut, Labudde, & Ramseier, 2010; Olson, Martin, Mullis, & Arora, op. 2008), diese sind jedoch im Wesentlichen aus zwei Gründen nicht geeignet. Erstens fokussieren fast alle Testinstrumente auf den schulischen Kontext. Zweitens zeigen sich bei der Erfassung von Experimentierkompetenz mit den jeweiligen Testinstrumenten unabhängig vom Kontext verschiedene Probleme (Shavelson R. J., Gao X., & Baxter G. P., 1993; Shavelson, Ruiz-Primo, & Wiley, 1999; Emden, 2011), wie z.B. eine unzureichende interne Konsistenz, hohe nicht aufgeklärte Varianzen oder unerklärbare Zusammenhänge zwischen einzelnen Konstrukten. Durch diese Probleme ergeben sich ungenaue und sogar gegenläufige Erkenntnisse. Es lässt sich analysieren, dass die Testgüte bzgl. der Reliabilität und Validität nicht ausreichend ist. Eine reliable und valide Datengenerierung ist jedoch notwendig, um entsprechende Schlussfolgerungen zur Entwicklung von Experimentierkompetenz ziehen zu können.

In dieser Arbeit soll analysiert werden, wodurch die Probleme bei der Erfassung von Experimentierkompetenz entstehen. Darauf aufbauend sollen Ansätze zur Lösung der identifizierten Probleme entwickelt werden. Basierend auf diesen Ansätzen wird ein Testinstrument zur Erfassung von Experimentierkompetenz entwickelt und die Ansätze werden auf ihre Wirksamkeit untersucht.

Die Grundlage der Testentwicklung bildet die Modellierung des Konstrukts – der Experimentierkompetenz. Je unkonkreter die Modellierung ausfällt, desto geringer ist die Messgenauigkeit eines darauf aufbauenden Testinstruments. Um ein möglichst hohes Maß an Konkretisierung bei der Modellierung zu erreichen, wird in Kapitel 2 ausgehend von dem Begriff der Handlungskompetenz in drei Schritten ein Modell zur Experimentierkompetenz entwickelt. Im ersten Schritt wird das Experimentieren als (iterativer) Prozess näher beleuchtet. Im zweiten Schritt werden die konkreten Handlungen beim Experimentieren, die zur erfolgreichen Bewältigung des Experimentierprozesses benötigt werden, identifiziert. In diesem Rahmen

wird geklärt, welche Fähigkeiten und Fertigkeiten diese Handlungen für erfolgreiches Experimentieren voraussetzen. Die Qualität des Experimentierens zeigt sich jedoch nicht nur darin, inwieweit die Handlungen adäquat umgesetzt werden können. Ein Experimentator muss die Handlungen in einer sinnvollen Abfolge durchlaufen, auf mögliche Fehler und Probleme eingehen und eine übergeordnete Strategie für sein Vorgehen wählen. Im dritten Schritt werden daher zur Bewertung der Qualität der Experimentierkompetenz Qualitätskriterien zur Richtigkeit der durchgeführten Handlungen, der Strukturierung der Handlungen und der Zielorientierung im Vorgehen entwickelt. Aus den Handlungen, Fähigkeiten und Fertigkeiten sowie den Qualitätskriterien ergibt sich ein Modell der Experimentierkompetenz, welches als Grundlage zur Analyse vorliegender Testinstrumente und der Testentwicklung in dieser Arbeit genutzt wird.

In Kapitel 3 werden vorliegende Testinstrumente analysiert. Im Rahmen dieser Analyse werden die Testinstrumente bzgl. ihrer Eigenschaften klassifiziert. Anschließend wird untersucht, ob bei ähnlichen Testinstrumenten auch ähnliche Schwierigkeiten bei der Testgüte auftreten. Aus den gewonnenen Erkenntnissen werden konkrete Probleme bei der Erfassung von Experimentierkompetenz destilliert und Ansätze zur Lösung dieser Probleme diskutiert.

Zur Analyse der Wirksamkeit der Ansätze ergibt sich in Kapitel 4 das Forschungsziel für den empirischen Teil. Um die Wirksamkeit zu analysieren werden auf Grundlage der Erkenntnisse aus den Kapiteln 2 und 3 in Kapitel 5 konkrete Anforderungen an ein Testinstrument herausgearbeitet. Als Grundlage zur Testentwicklung wird in Kapitel 6 eine (Fachinhalts)Analyse von Fachinhalten in deutschsprachigen physikalischen Praktika im Bereich Optik durchgeführt. Basierend auf den Anforderungen, den identifizierten Ansätzen und der Fachinhaltsanalyse entsteht in Kapitel 7 ein Testinstrument zur Erfassung von Experimentierkompetenz im Fachinhaltsbereich Optik für das physikalische Anfängerpraktikum.

Die Untersuchung der Forschungsfragen enthält qualitative- und quantitative Aspekte. In Kapitel 8 werden die qualitativen Aspekte mittels einer Video-Studie mit der Methode des lauten Denkens mit  $N=16$  Probanden untersucht. Basierend auf der generierten Evidenz wird das Testinstrument angepasst. Anschließend werden in Kapitel 9 mit einem Feldtest mit  $N=58$  Studierenden im physikalischen Anfängerpraktikum in Kiel die quantitativen Aspekte analysiert.

In Kapitel 10 werden die Erkenntnisse der qualitativen und quantitativen Studie vor dem Hintergrund der Einsetzbarkeit des Testinstruments diskutiert. Die Wirksamkeit der Ansätze zur Verbesserung der Testgüte wird bewertet und weitere Schritte zur Optimierung der Testgüte vorgeschlagen.

## 2 Modellierung von Experimentierkompetenz

In der gesamten naturwissenschaftlichen Ausbildung ist eines der zentralen Ziele zu erlernen, wie man experimentiert (in der schulischen Ausbildung Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005c; Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005a; Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005b; "National Science Education Standards," 1996; National Research Council, 2012; in der universitären Ausbildung Welzel et al., 1998, Praktikums- und Studienordnungen; und in der berufsorientierten Ausbildung Rauner & Spöttl, 2002; Frank, 2014). Im Output-orientierten Bildungswesen wird der Erfolg von Bildung daran gemessen, welche Kompetenzen Lernende erworben haben (z.B. Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005c). Nach Weinert (2001) umfasst Kompetenz diejenigen kognitiven Fähigkeiten und Fertigkeiten sowie die motivationalen, volitionalen und sozialen Bereitschaften, die notwendig sind um spezifische Probleme zu lösen (Weinert, 2001). Der Prozess des Experimentierens umfasst im Allgemeinen eine Vielzahl an spezifischen Problemen. Um effizient experimentieren zu können, müssen alle Probleme bearbeitet und dafür verschiedene Handlungen vollzogen werden (z.B. Klahr & Dunbar, 2000; Hofstein & Lunetta, 2004). Die einzelnen Handlungen lassen sich beobachten und bewerten. Grundlage zur Durchführung der Handlungen sind jedoch verschiedene Fähigkeiten und Fertigkeiten. Diese einzelnen Fähigkeiten und Fertigkeiten lassen sich nicht direkt beobachten. Die Integrierung verschiedener Kompetenzen zur erfolgreichen Realisierung von Handlungen wird als Handlungskompetenz bezeichnet (z.B. Nickolaus, Retelsdorf, Winther, & Köller, 2013; Erpenbeck, 2007). Nach dieser Darstellung lässt sich Experimentieren als Handlungskompetenz verstehen und wird in der vorliegenden Arbeit als Experimentierkompetenz<sup>2</sup> bezeichnet.

In Modellierungen von Experimentierkompetenz wird meistens entweder von Handlungen oder von Fähigkeiten bzw. Fertigkeiten gesprochen. Die beschreibenden Begriffe der Handlungen bzw. der dafür notwendigen Fähigkeiten bzw. Fertigkeiten sind jedoch größtenteils vergleichbar. Das gilt sogar für alle Phasen der naturwissenschaftlichen Ausbildung (vgl. Emden, 2011). Daraus lässt sich ableiten, dass die Beschränkung der Modellierung auf eine einzelne Phase der Ausbildung nicht sinnvoll ist. Gleichzeitig zeigen sich bedingt durch normative Setzungen oder gesellschaftliche bzw. historische Strebungen teilweise große Unterschiede

---

<sup>2</sup> Im weiteren Verlauf der Arbeit wird der Begriff Experimentierkompetenz als Synonym für die verschiedenen ähnlichen Konstrukte verwendet.

in der Gewichtung der Handlungen bzw. Fähigkeiten und Fertigkeiten im Experimentierprozess (z.B. Mayer; Schreiber, 2012; National Research Council, 2012).

Im ersten Abschnitt dieses Kapitels werden basierend auf einer Darstellung des Experimentierprozesses, die relevanten Handlungen extrahiert und durch die notwendigen Fähigkeiten und Fertigkeiten beschrieben und voneinander abgegrenzt. Durch die Zusammenführung der ausdifferenzierten Schwerpunkte aller vorliegenden Modellierungen entsteht ein möglichst konkretes, umfassendes und konsensfähiges Modell.

Neben der Klärung und Abgrenzung der Handlungen des Konstrukts durch die Fähigkeiten und Fertigkeiten, umfasst Kompetenz auch eine Beschreibung der Qualität bzw. der Ausprägung von der Kompetenz (Nickolaus et al., 2013)). Zur Klärung der Frage, wann ein Experimentator erfolgreich oder gut experimentieren kann, also ein hohes Maß an Experimentierkompetenz besitzt, wird im zweiten Abschnitt des Kapitels analysiert, woran die Güte bzw. Qualität beim Experimentieren bemessen werden kann. Ausgehend davon werden zur Bemessung der Ausprägung von Experimentierkompetenz die Gütemaße *Richtigkeit*, *Strukturiertheit* und *Zielorientiertheit* entwickelt.

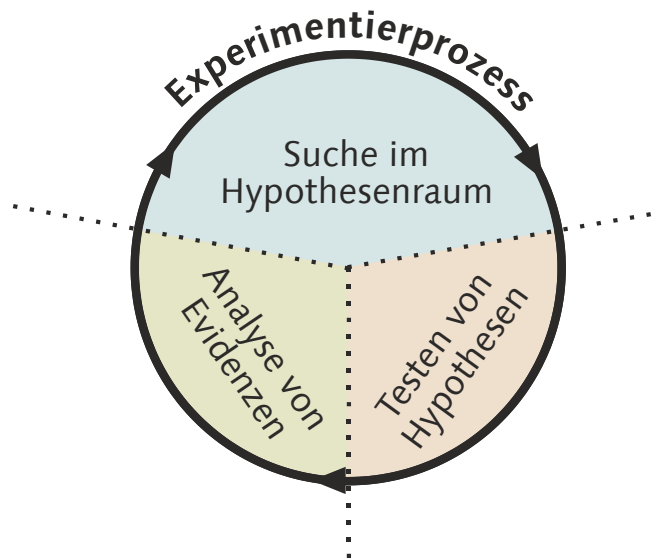
Im letzten Abschnitt des Kapitels wird aus den Handlungen, Fähigkeiten und Fertigkeiten auf der einen Seite und den Gütemaßen auf der anderen Seite ein zweidimensionales Modell der Experimentierkompetenz entwickelt.

## **2.1 Fähigkeiten, Fertigkeiten und Handlungen im Experimentierprozess**

Experimentieren bedeutet, dass Frage- oder Problemstellungen unter Verwendung von Evidenz, generiert durch Experimente, beantwortet bzw. gelöst werden (Klahr & Dunbar, 2000; Emden, 2011; Schreiber, 2012). Nach Klahr und Dunbar (2000) lässt sich der Prozess des Experimentierens in drei Phasen gliedern: Die erste Phase wird als *„Suche im Hypothesenraum“* oder *„Vorbereitung/Planung“* bezeichnet (Klahr & Dunbar, 2000; Hammann & Hoi Phan, 2007 bzw. Schreiber, 2012; Maiseyenko et al., 2011; Emden, 2011). Diese Phase umfasst alle Schritte und die entsprechenden Fähigkeiten und Fertigkeiten, die notwendig sind, um ausgehend von einer Aufgabe oder einem Problem eine Planung zu entwerfen. Sie beginnt mit einer Fragestellung und führt über Hypothesen zur Planung eines Experiments. Die Planung ist die Grundlage zur sinnvollen Interaktion mit Versuchsmaterial. Die zweite Phase ist das *„Testen von Hypothesen“* oder die *„Durchführung“* (Klahr & Dunbar, 2000; Hammann & Hoi Phan, 2007 bzw. Schreiber, 2012; Maiseyenko et al., 2011; Emden, 2011). In dieser Phase interagiert die experimentierende Person mit dem Versuchsmaterial. Dies beinhaltet



sowohl den Aufbau von Versuchen als auch Messungen, inklusive der Dokumentation. Die dritte Phase nennt sich *“Analyse von Evidenzen“* oder *“Auswertung“* (Klahr & Dunbar, 2000; Hammann & Hoi Phan, 2007 bzw. Schreiber, 2012; Maiseyenko et al., 2011; Emden, 2011). Diese Phase dient der Evaluation der gewonnenen Daten oder Erkenntnissen mit einem Rückbezug auf die in Phase 1 formulierten Erwartungen. Ergebnisse der dritten Phase können dazu führen, dass Phase 1 unter Berücksichtigung der neuen Erkenntnisse erneut durchlaufen wird. Dies passiert vor allem dann, wenn die Problemstellung nicht vollständig gelöst wurde oder sich neue Fragestellungen ergeben. Damit entsteht ein potentiell iterativer Kreisprozess (siehe Abb. 1).



**Abbildung 1: Der Prozess des Experimentierens.**

Dieser iterative Prozess aus drei Phasen lässt sich weiter ausdifferenzieren. So werden verschiedentlich weitere Handlungen oder entsprechende Fähigkeiten oder Fertigkeiten innerhalb der drei Phasen differenziert (z.B. Schreiber, 2012; National Research Council, 2012; Wissenschaftliches Konsortium HarmoS Naturwissenschaften+, 2008; Hodson, 1996). Eine Unterscheidung zwischen Handlungen und Fähigkeiten bzw. Fertigkeiten aber auch zwischen den Begriffen Fähigkeiten und Fertigkeiten wird in den meisten Arbeiten nicht vorgenommen. Vielmehr werden die Begriffe Fähigkeiten, Fertigkeiten oder Handlungen teilweise simultan verwendet.

Zur Eindeutigkeit wurde zu Beginn des Kapitels bereits dargestellt, dass Handlungen die tatsächlich zu beobachtenden Aktivitäten der experimentierenden Personen sind, während die Fähigkeiten bzw. Fertigkeiten notwendige Grundlagen für die Handlungen darstellen. Die

Spezifizierung zwischen Fähigkeiten und Fertigkeiten wird in der vorliegenden Arbeit nach Emden (Emden, 2011) wie folgt definiert:

**Fertigkeiten** lassen sich als *Voraussetzung zur händischen Durchführung laborspezifischer Handlungsabläufe* bezeichnen (Emden, 2011, 2011, S. 15), während

**Fähigkeiten** die *planvolle Durchführung, Überwachung und Reflexion naturwissenschaftlicher Untersuchungen* darstellen (Emden, 2011, 2011, S. 15).

Das heißt, man braucht sowohl die Fertigkeit, einen Versuch aufzubauen und zu messen, als auch die Fähigkeit, logisch darauf aufbauend eine Schlussfolgerung zu ziehen und gegebenenfalls eine Hypothese zu verwerfen. Die jeweilige Aktivität der experimentierenden Person kann durch die zugehörige Handlung beobachtet werden. Fähigkeiten und Fertigkeiten umfassen deklaratives und prozedurales Wissen zu den jeweiligen Fähigkeiten bzw. Fertigkeiten (Shavelson & Ruiz-Primo, 1998). Im Folgenden werden die Handlungen im Experimentierprozess herausgearbeitet. Die Handlungen werden durch konkrete Fähigkeiten und Fertigkeiten beschrieben und voneinander abgegrenzt. Eine Übersicht der Handlungen gegliedert nach Arbeiten findet sich im Anhang unter A.1 (David Wetzel; Emden, 2011; Emden, Hübinger, & Sumfleth, Juli 2010; Gut et al., 2010; Hammann & Hoi Phan, 2007; Hammann et al., 2008; Hanauer, Hatfull, & Jacobs-Sera, 2009; Hodson, 1996; Hofstein & Lunetta, 2004; Kempa, 1986; Kipnis & Hofstein, 2008; Klahr & Dunbar, 2000; Klieme, Funke, Leutner, Reimann, & Wirth, 2001; Labudde, Metzger, & Gut, 2009; Lunetta, 1998; Maiseyenko et al., 2011; Mayer, Bauer, & Kattmann, 2003; Mayer, Grube, & Möller, 2008; Grube & Christiane, 2010; Murphy, 1996; National Research Council, 2012; "National Science Education Standards," 1996; Prenzel et al., 04.12.2007; Schreiber, 2012; Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005c; Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005a; Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005b; Tamir, Doran, Kojima, & Bathory, 1992; Tamir, Doran, & Chye, 1992; Tamir & Lunetta, 1981; Walpuski, 2006; Welzel et al., 1998; Wissenschaftliches Konsortium HarmoS Naturwissenschaften+, 2008).

### 2.1.1 Fragestellung bzw. Problemstellung entwickeln oder klären

Der Prozess des Experimentierens beginnt üblicherweise mit einer Frage- oder Problemstellung. Dazu gehört das einfache Stellen bzw. Formulieren, Identifizieren, Klären und Entwickeln von Forschungsfragen (z.B. Labudde et al., 2009; Mayer et al., 2008; David Wetzel; Maiseyenko et al., 2011; für eine vollständige Übersicht vgl. A.1). Die Entwicklung von Forschungsfragen ist Gegenstand realer Forschungsarbeiten und nicht von vorgeplanten Experimenten z.B. in Praktika. Dies geschieht, wenn theoretische oder empirische Hinweise zu einer Problemsituation führen, welche sich durch den bestehenden Wissensstand nicht erklären bzw. lösen lässt. Die Forschungsfrage erfüllt die Aufgabe einen theoretisch fundierten Erklärungsansatz zu fokussieren. Im weiteren Verlauf des Experimentierprozesses soll aufbauend auf der Forschungsfrage Evidenz generiert werden, mit der sich der Erklärungsansatz stützen oder widerlegen lässt. Ein Beispiel dafür ist der Welle-Teilchen-Dualismus von Licht und die Problemstellung von Thomas Young, die zum bekannten Doppelspalt-Experiment geführt hat<sup>3</sup>.

Die weniger wissenschaftliche Version des Entwickelns ist das Stellen bzw. Formulieren von Fragestellungen. Anders als beim Entwickeln ist das Vorgehen wenig wissenschaftsgeleitet und mehr intuitiv aus der Situation generiert. Beim Entwickeln von Fragestellungen gilt das verfügbare Wissen als Referenz. Wird eine Forschungsfrage entwickelt und geklärt, die bereits beantwortet ist, so gibt es keinen Erkenntnisgewinn für die Forschungsgemeinschaft. Beim Stellen bzw. Formulieren von Forschungsfragen gilt das Wissen des Lernenden als Referenz. Weiß der Lernende etwas nicht, wird die Forschungsfrage entsprechend formuliert. Darauf aufbauend generiert der Lernende Evidenz. Für ihn selbst entsteht auf diesem Weg ein Erkenntnisgewinn. Das "Fragestellen" bzw. "Formulieren" findet sich vor allem in Standards für die Schulausbildung (z.B. (National Research Council, 2012; Wissenschaftliches Konsortium HarmoS Naturwissenschaften+, 2008; für eine vollständige Übersicht vgl. A.1).

---

<sup>3</sup>Um 1800 war die vorherrschende Meinung, dass Licht ein Teilchen ist. Populärer Anhänger dieser Theorie war Isaac Newton, dessen Wort so ein Gewicht hatte, dass andere Ansätze (z.B. das Huygensche Prinzip) nicht ernsthaft verfolgt wurden. Trotzdem war Thomas Young davon überzeugt, dass Licht eine Welle ist. Er suchte nach einer Möglichkeit seine Ansicht durch Evidenz zu stützen. Dafür entwickelte er 1802 das bekannte Doppelspalt-Experiment. Basierend auf der Theorie nach der Licht ein Teilchen ist, müssten bei dem Experiment andere Ergebnisse beobachtet werden können, als nach der Theorie, wonach Licht eine Welle ist. Darauf aufbauend entwickelte er die Fragestellung, wie das Bild am Schirm hinter dem Doppelspalt aussähe.

Die Fragestellung wurde basierend auf zwei konkurrierenden Theorien entwickelt. Die Forschungsfrage ist zudem in der Lage die Forschungslücke (Hat Licht Welleneigenschaften?) zu klären. In dem vorliegenden Fall konnten die Welleneigenschaften von Licht gezeigt werden.

In expliziten Experimentiersituationen in der Schule und weiten Teilen der Universität sind die Frage- oder Problemstellungen vorgegeben und müssen identifiziert, geklärt oder verstanden und nicht selbst generiert werden. Bezogen auf eine Testsituation stellen das Identifizieren und Klären die einzig relevanten Aspekte der Frage- bzw. Problemstellung dar. Dies folgt aus Gründen der Vergleichbarkeit von Testergebnissen durch vorgegebene Aufgabenstellungen. Dazu gehört es zu verstehen, worum es in der Aufgabe geht und was das Ergebnis umfassen muss. Wenn die Problemstellung z.B. lautet „*Bestimmen Sie die Spannung über einem unbekanntem Widerstand in einer gegebenen Schaltung.*“ dann muss die experimentierende Person wissen, dass sie einen Widerstand untersuchen muss und das Ergebnis eine Spannung (gemessen in Volt) umfasst.

### **2.1.2 Hypothesen generieren/ Klären der theoretischen Grundlage**

Nachdem eine Frage bzw. eine Problemstellung vorliegt, werden Hypothesen gebildet (z.B. National Research Council, 2012; Tamir, Doran, Kojima et al., 1992; für eine vollständige Übersicht vgl. A.1). Der Begriff Hypothese steht laut Duden für *eine vom Widerspruch freie, aber zunächst unbewiesene Annahme*. Ein Beispiel für eine richtige Hypothese wäre angeknüpft an das Beispiel mit Thomas Young, dass nach dem Wellenmodell Licht auch in den Schattenbereich des Spaltes eindringen kann<sup>4</sup>. Allerdings beziehen sich die Beschreibungen der Fähigkeit zur Hypothesenbildung in den meisten der vorliegenden Arbeiten eher darauf, bekannte Zusammenhänge zu verifizieren. Zur Abgrenzung der tatsächlichen Hypothesenbildung ist eine schwächere Bezeichnung die Formulierung von Erwartungen oder möglichen Lösungen (Schreiber, 2012; David Wetzel; Klahr & Dunbar, 2000; Hammann & Hoi Phan, 2007). Wenn bei den vorliegenden Beispielen von einem ungewissen Ergebnis des Experiments gesprochen wird, bedeutet es eher, dass der Experimentator den Ausgang nicht kennt ihm als das entsprechende Fachwissen fehlt. So werden Experimente z.B. in der Schule zur Generierung eines kognitiven Konfliktes eingesetzt. Das Ergebnis ist nur unbekannt, weil den Lernenden die benötigten theoretischen Grundlagen fehlen. Der Lehrende hingegen kennt den intendierten Ausgang des Experiments. Zur Generierung von Hypothesen (vornehmlich in Forschungsarbeiten) ist umfassendes Wissen über das Forschungsfeld und dessen For-

---

<sup>4</sup>Da zwei konkurrierende Theorien vorliegen, gibt es auch zwei verschiedene mögliche Ergebnisse des Experiments. Die erste Hypothese wäre entsprechend, dass ein Beugungsbild entsteht (Welleneigenschaften), die zweite Hypothese wäre, dass zwei Lichtflecken entstehen (Teilcheneigenschaften). Beide Hypothesen sind basierend auf den ihnen zugrundeliegenden Theorien widerspruchsfrei und damit mögliche Ausgänge des Experiments.

schungsdefizite notwendig (Hanauer et al., 2009; Wissenschaftliches Konsortium HarmoS Naturwissenschaften+, 2008; Prenzel et al., 04.12.2007; Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005c; National Science Education Standards, 1996; National Research Council, 2012). Dazu gehört es Analogien und Modelle zu nutzen sowie Mathematisierungen anzuwenden (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005c). Beispielsweise müsste bei der Untersuchung der Bewegung eines Probekörpers auf einer schiefen Ebene, bei der Klärung der theoretischen Grundlage ein notwendiges Fachwissen zu der Newtonschen Mechanik, zu Winkelgesetzen und ggf. zu Reibung vorhanden sein. Zudem müsste der Experimentator die entsprechenden Formeln und die trigonometrischen Zusammenhänge beherrschen oder sich aneignen. Erst mit diesem Fachwissen wäre er in der Lage eine komplexere Hypothese zu entwickeln bzw. begründet eine Erwartung oder den Ausgang des Experiments zu formulieren. Bei trivialen oder didaktisch reduzierten Experimenten, beispielsweise in der Schule, kann auch mit einem geringeren Maß an Fachwissen eine einfache Hypothese entwickelt werden.

### **2.1.3 Experiment planen**

Wenn eine Frage- bzw. Problemstellung vorliegt, die theoretische Grundlage geklärt ist und ggf. Hypothesen bzw. Vermutungen geäußert wurden, wird im nächsten Schritt ein Experiment geplant, um die Frage bzw. das Problem zu untersuchen (z.B. National Research Council, 2012; Wissenschaftliches Konsortium HarmoS Naturwissenschaften+, 2008; Gut et al., 2010; Maiseyenko et al., 2011; Schreiber, 2012; für eine vollständige Übersicht vgl. A.1). Dazu gehört, dass die abhängigen und unabhängigen Variablen identifiziert und Strategien zur Kontrolle weiterer Variablen entwickelt werden (National Science Education Standards, 1996; National Research Council, 2012; Tamir, Doran, & Chye, 1992; Tamir, Doran, Kojima et al., 1992). Die Planung muss so umfassend sein, dass der erstellte Plan ohne weiteres abgearbeitet werden kann. Beispielsweise sieht ein Lösungsweg zur Wellenlängenbestimmung mit optischem Gitter wie folgt aus: Der Proband muss einen Versuch planen, bei dem der Laserstrahl auf das optische Gitter und anschließend auf den Schirm trifft. Der Strahlengang muss Gitter und Schirm im rechten Winkel treffen. Die abhängige Variable ist die Wellenlänge des Lichts. Die unabhängigen Variablen sind Spaltabstand des optischen Gitters, Abstand vom Gitter zum Schirm und der Abstand von der optischen Achse zum n-ten Maximum. Dazu müssen die notwendigen Formeln und die Rechenvorschrift bezeichnet werden.

#### **2.1.4 Versuch fehlerfrei aufbauen**

Nach der Planung müssen dieser entsprechend Komponenten identifiziert und gewählt sowie der Versuch aufgebaut werden (z.B. Kempa, 1986; Gut et al., 2010; für eine vollständige Übersicht vgl. A.1). Bestandteil des Aufbaus ist es, Fehler zu erkennen und zu beheben (Maiseyenko et al., 2011; Welzel et al., 1998). Am Ende des erfolgreichen Versuchsaufbaus steht ein fertig justierter Versuch, der geeignet ist, die in der Planung beschriebenen abhängigen und unabhängigen Variablen zu bestimmen. Beispielsweise müsste ein Experimentator zur Messung des Spannungsabfalls über einem Widerstand eine Schaltung wie vorgegeben aufgebaut und ein Spannungsmessgerät parallel zum Widerstand geschaltet haben. Dazu muss er die verschiedenen Komponenten identifizieren.

#### **2.1.5 Messen, Beobachten und Dokumentieren**

Auf Grundlage der Planung und des aufgebauten Versuchs ergibt sich als nächster Schritt im Experimentierprozess die Durchführung der Messung (z.B. Hodson, 1996; Tamir, Doran, Kojima et al., 1992; Tamir & Lunetta, 1981; Emden, 2011; Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005c; für eine vollständige Übersicht vgl. A.1). Dabei sind Sicherheitshinweise zu beachten (National Research Council, 2012). Bei der Durchführung werden mit geeigneten Messinstrumenten Messungen durchgeführt, bei denen gezielt Werte für die abhängigen und unabhängigen Variablen erhoben werden (z.B. Kempa, 1986; Schreiber, 2012; Maiseyenko et al., 2011; National Research Council, 2012; für eine vollständige Übersicht vgl. A.1). Zum Messen gehört es, die Qualität der Messung zu beurteilen bzw. sicherzustellen (National Research Council, 2012; Hofstein & Lunetta, 2004). Alternativ zu Messungen werden Beobachtungen angestellt, bei denen qualitative Eigenschaften beschrieben werden (z.B. Hanauer et al., 2009; Hofstein & Lunetta, 2004; National Research Council, 2012; für eine vollständige Übersicht vgl. A.1). Die jeweiligen Messwerte bzw. Beobachtungen müssen korrekt und nachvollziehbar dokumentiert werden (z.B. Kempa, 1986; National Research Council, 2012; für eine vollständige Übersicht vgl. A.1). Am Beispiel der Bewegung verschiedener Massen an der schiefen Ebene müssen als Messinstrument ein Winkelmesser, eine Waage und ggf. ein Geschwindigkeitsmesser genutzt werden. Mit diesen Instrumenten müssen die in der Planung beschriebenen Gewichte, Winkel und Bewegungen gemessen werden. Dokumentiert werden die jeweiligen Größen für Gewicht, Winkel und Bewegung.

### **2.1.6 Daten analysieren und Verallgemeinerbarkeit diskutieren**

Nach der Generierung von Daten kann oftmals nicht ohne weiteres eine Aussage bzgl. der Frage- bzw. Problemstellung getroffen werden. Die Daten werden im Regelfall aufbereitet (z.B. Gut et al., 2010; KMK, 2005c; für eine vollständige Übersicht vgl. A.1). Beides lässt sich als Analyse oder Auswertung der Daten zusammenfassen (z.B. Hodson, 1996; Hammann & Hoi Phan, 2007; Emden, 2011; für eine vollständige Übersicht vgl. A.1). Ein relevanter Bestandteil ist es, die Verallgemeinerbarkeit zu diskutieren, z.B. indem Fehlerrechnungen durchgeführt werden (z.B. Hofstein & Lunetta, 2004; National Research Council, 2012; für eine vollständige Übersicht vgl. A.1).

Für das Beispiel einer Messung zur Aktivität eines radioaktiven Probekörpers müsste die Anzahl der detektierten ionisierten Teilchen mit der Zeit ins Verhältnis gesetzt werden. Anschließend sollte der Einfluss der Totzeit des Detektors, ggf. des Abstandes und des Öffnungswinkels auf die Ergebnisse diskutiert werden.

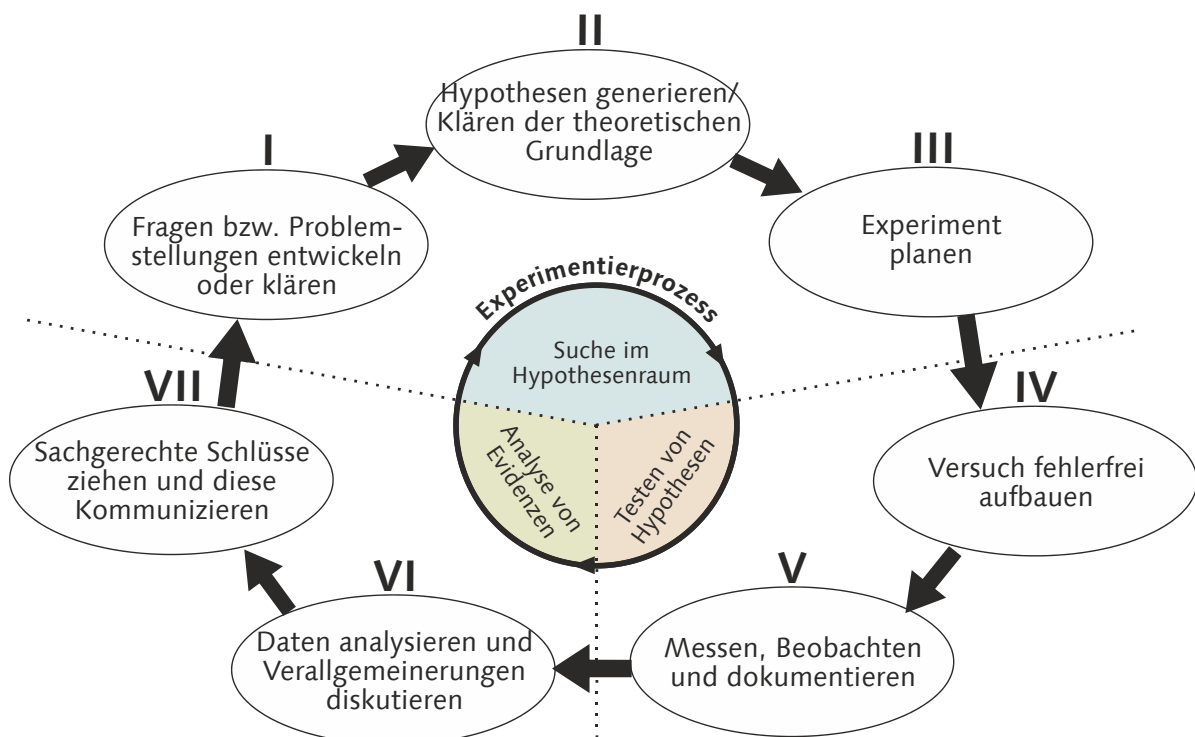
### **2.1.7 Sachgerechte Schlüsse ziehen und diese kommunizieren**

In diesem Schritt müssen die vorliegenden aufbereiteten und ggf. durch Mathematisierung berechneten Daten bzgl. der Frage- bzw. Problemstellung diskutiert und interpretiert werden (z.B. Schreiber, 2012; Murphy, 1996; Kempa, 1986; für eine vollständige Übersicht vgl. A.1). Dazu müssen die generierten Daten in Bezug gesetzt werden zu Hypothesen, Erwartungen bzw. den theoretischen Grundlagen, welche als Voraussetzung zum Experiment dienten (z.B. Hofstein & Lunetta, 2004; Hanauer et al., 2009; Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005c; für eine vollständige Übersicht vgl. A.1). Auf dieser Grundlage muss eine Schlussfolgerung gezogen bzw. ein Ergebnis formuliert werden (z.B. Mayer et al., 2008; Kipnis & Hofstein, 2008; Murphy, 1996; für eine vollständige Übersicht vgl. A.1). Die gewonnenen Erkenntnisse müssen kommuniziert werden (z.B. Gut et al., 2010; Hanauer et al., 2009; Hofstein & Lunetta, 2004; für eine vollständige Übersicht vgl. A.1). Die Kommunikation reicht vom Verfassen eines wissenschaftlichen Artikels über die Präsentation der Erkenntnisse bis hin zum Notieren eines begründeten Antwortsatzes im Schulheft.

Am Beispiel der Wellenlängenbestimmung bedeutet dies: Die Wellenlänge sollte nach einer ersten Schätzung im Bereich von 400-900 nm liegen. Dieser Bereich kann durch die Farbe des Lasers weiter spezifiziert werden. Abhängig von der Passung der bestimmten Wellenlänge kann das Ergebnis als richtig angenommen oder ein Fehler im Experiment vermutet werden. Die jeweilige Erkenntnis muss nachvollziehbar formuliert werden.

### 2.1.8 Anordnung der Handlungen im Experimentierprozess

Zu Beginn des Kapitels 2.1 wurde erläutert, dass sich der Experimentierprozess als Folge unterschiedlicher Klassen von Handlungen beschreiben lässt (vgl. Emden, 2011; Schreiber, 2012). Die spezifischen Handlungen mit den dafür notwendigen Fähigkeiten und Fertigkeiten wurden identifiziert. Der zu Beginn des Kapitels 2.1 in Abbildung 1 dargestellte Experimentierprozess kann entsprechend um die Handlungen erweitert werden. Daraus ergibt sich das in Abbildung 2 dargestellte Modell. Die ersten drei Handlungen (I Fragestellung bzw. Problemstellung entwickeln oder klären; II Hypothesen generieren/ Klären der theoretischen Grundlage; III Experiment planen) lassen sich der ersten Phase, der Suche im Hypothesenraum zuordnen. Das Testen von Hypothesen wird durch die nächsten beiden Handlungen (IV Versuch fehlerfrei aufbauen; V Messen, Beobachten und Dokumentieren) beschrieben. Die dritte Phase zur Analyse von Evidenzen enthält die zwei Handlungen VI *Daten analysieren und Verallgemeinerbarkeit diskutieren* sowie VII *Sachgerechte Schlüsse ziehen und diese kommunizieren*. Durch diese Erweiterung des Modells bleibt der bereits beschriebene potentiell iterative Kreisprozess bestehen. Es zeigt sich aber auch deutlich, dass Experimentierkompetenz mehr ist als eine einzelne Fähigkeit bzw. Fertigkeit "zu können".



**Abbildung 2: Handlungen des potentiell iterativen Kreisprozess des Experimentierens.**

Ausgehend von einem abgeschlossenen idealen Durchlauf beim Experimentieren- also von Handlung I bis Handlung VII- kann ein neuer Durchlauf initialisiert werden. Die Handlungen



eines Durchlaufs werden als Zyklus bezeichnet. Im Regelfall werden oftmals nicht alle Handlungen von I bis VII durchlaufen und trotzdem ein neuer Durchlauf initialisiert. Allgemein werden im weiteren Verlauf der Arbeit die Handlungen eines (auch nicht vollständigen) Durchlaufs bis zur Initialisierung eines neuen Durchlaufs als Zyklus im Experimentierprozess bezeichnet. Der iterative Kreisprozess setzt sich aus mehreren Zyklen zusammen.

## **2.2 Qualitätsmaße der Experimentierkompetenz**

Einem Experimentator wird nach der Darstellung der Handlungskompetenz eine hohe Ausprägung von Experimentierkompetenz zugeschrieben, wenn er (komplexe) Problem- bzw. Fragestellungen korrekt lösen kann. Eine Bewertung ausschließlich auf Basis der Korrektheit der Lösung in einer vorgegebenen Bearbeitungszeit ist naheliegend (z.B. Prenzel et al., 04.12.2007; Bybee, McCrae, & Laurie, 2009). Den gesamten Prozess des Experimentierens nur auf ein einzelnes Merkmal zu beziehen, lässt diverse Chancen zur Differenzierung und gezielten Förderung außer Acht. Zu einer stärker differenzierenden Bewertung werden Teilergebnisse analysiert. Dabei wird bewertet, ob einzelne Handlungen richtig durchgeführt wurden (z.B. Maiseyenko et al., 2011; Hammann et al., 2008).

Die bisher vorgestellten Bewertungen zielen ausschließlich auf die korrekte Durchführung von Handlungen. Es wird aber postuliert, dass vor allem die Abfolge der Handlungen die Qualität des Experimentators beschreibt (vgl. Shavelson et al., 1998; Shavelson & Ruiz-Primo, 1998; Klieme et al., 2001). Bei Messinstrumenten basierend auf diesem Ansatz wird vor allem die Passung der Handlungen zum aktuellen Stand des Experimentierprozesses und vorheriger Erkenntnisse bewertet (z.B. Emden, 2011). Ein weiterer Ansatz unterscheidet Novizen und Experten beim Experimentieren durch die Wahl der Lösungsstrategie (Klahr & Dunbar, 2000). Hochwertige Lösungsstrategien konvergieren schneller und führen zu einer kürzeren Bearbeitungszeit, was auf ein höheres Maß an Experimentierkompetenz zurückgeführt wird.

Die drei dargestellten Bewertungsansätze basieren auf ganz unterschiedlichen Aspekten des Experimentierprozesses. Alle drei Aspekte beschreiben nachvollziehbar einen Qualitätsaspekt im Experimentierprozess und setzen die vollzogenen Handlungen voraus. In der vorliegenden Arbeit wird daher postuliert, dass Qualität beim Experimentieren bzw. die Ausprägung der Experimentierkompetenz sich in drei Teilaspekte gliedert: das Gütemaß der Richtigkeit bei der Durchführung von Handlungen, das Gütemaß der Strukturiertheit bei der Abfolge von Handlungen und das Gütemaß der Zielorientiertheit einer gewählten Strategie zur Steuerung der Handlungen. Im Folgenden werden die drei Gütemaße konkret entwickelt.

### 2.2.1 Richtigkeit

Typischerweise wird bei Testinstrumenten zum Experimentieren bewertet, ob ein Proband an einem bestimmten Punkt das richtige Ergebnis erreicht hat. Dabei kann es sich um das Endergebnis einer Aufgabe oder um Zwischenergebnisse nach bestimmten Handlungen in Teilaufgaben handeln (z.B. Prenzel et al., 04.12.2007; Bybee et al., 2009; Martin, 2008; Mullis, 2007; Wissenschaftliches Konsortium HarMoS Naturwissenschaften+, 2008; Hammann & Hoi Phan, 2007). Bei den zitierten Arbeiten werden einzelne Handlungsschritte bewertet und dann ein Maß für Experimentierkompetenz als Summenscore oder Mittelwert aus diesen Bewertungen gebildet. Dabei wird üblicherweise keine Gewichtung der Ergebnisse oder Handlungen vorgenommen. Für den Fall, dass alle Handlungen beim Experimentieren durch dieselbe Anzahl an Aufgaben geprüft werden, sind alle Handlungen gleich gewichtet. Allerdings sind die meisten Testinstrumente ungleichmäßig gewichtet. Das bedeutet, dass einige Handlungsschritte im Experimentierprozess durch mehr Aufgaben bzw. Teilaufgaben abgedeckt werden als andere Handlungsschritte. In diesem Fall haben die verschiedenen Handlungen einen unterschiedlichen Einfluss auf das Testergebnis und damit auf die vorhergesagte Experimentierkompetenz. Um dieses Problem zu beheben, wird bei Testentwicklung im Allgemeinen auf eine homogene Aufgabenverteilung in den Subskalen geachtet (z.B. Kleickmann et al., 2014). Das ist beim Experimentieren in einer freien Umgebung allerdings nicht immer zu realisieren, denn die Auswahl der Handlungen ist in diesem Fall von dem spezifischen Vorgehen der einzelnen Probanden abhängig. Probanden, die eine ausführliche Planung zum Experiment durchführen und sich detailliert mit den theoretischen Grundlagen beschäftigen, werden vor allem in der ersten Phase einen Schwerpunkt von Handlungen aufweisen. Entsprechend wird die Testleistung zur Richtigkeit dann vor allem durch die Handlungen der ersten Phase beschrieben. Es gibt nun zwei Möglichkeiten damit umzugehen. Eine Möglichkeit wäre, dem Probanden zu attestieren, dass seine inhomogene Wahl an Handlungen einen bewussten oder unbewussten Abgleich möglicher Lösungswege mit seinem Fähigkeitsprofil<sup>5</sup> darstellt. Entscheidet sich der Proband für einen Lösungsprozess, der durch sein Fähigkeitsprofil am besten durchlaufen werden kann, so spricht das für seine Kompetenz. Mögliche Probleme bei Handlungen, die ihm mehr Schwierigkeiten bereiten, werden durch für ihn einfachere Handlungen behoben. Die zweite Möglichkeit besteht darin, die einzelnen Handlungen unabhängig von der Häufigkeit der Handlungen gleich zu gewichten.

Eine Alternative zur Bewertung von Teil- bzw. Endergebnissen als Maß für die richtige Durchführung von Handlungen stellt das Liverating oder die Videoanalyse dar (Janík, 2009;

---

<sup>5</sup>Als Fähigkeitsprofil werden die handlungsspezifischen Fähigkeiten und Fertigkeiten zu jeder Handlung verstanden.

Walpuski & Sumfleth, 2009; Thek & Hoole, 2006). Diese Methode des Liveratings wird in einigen Praktika angewendet. Ein Assistent oder Praktikumsleiter beobachtet die Studierenden beim Experimentieren. Er bewertet, basierend auf seinen Eindrücken, ob die Handlungen beim Experimentieren richtig durchgeführt wurden. Die Videoanalyse wird vor allem in der Forschung genutzt.

Diese Ansätze zur Beurteilung der Qualität beim Experimentieren lassen sich als Qualitätsmaß der Richtigkeit zusammenfassen. Gegenstand der Beurteilung sind die einzelnen Handlungen. Es wird bewertet, inwieweit die Handlungen richtig durchgeführt wurden.

### 2.2.2 Strukturiertheit

In der Einleitung wurde dargestellt, dass sich die Qualität des Experimentierens vor allem im Prozess, also der Abfolge von Handlungen zeigt (vgl. Shavelson et al., 1998; Shavelson & Ruiz-Primo, 1998; Klieme et al., 2001). Emden (2011) hat darauf aufbauend ein Testinstrument entwickelt. Beispielsweise ist es sinnvoll, einen Versuch wie geplant aufzubauen und anschließend durchzuführen. Nicht sinnvoll ist es dagegen, wenn ein anderer Versuch als geplant aufgebaut wird. Relevant ist also nicht allein, ob die Handlung richtig durchgeführt wurde, sondern ob die Handlung zur aktuellen Experimentiersituation passt und damit zur Lösung der Problemstellung beiträgt- also den Erkenntnisgewinn voranbringt.

Demnach brauchen Experimentatoren ein übergeordnetes, konzeptuelles Wissen, um die Handlungen logisch sinnvoll miteinander zu verknüpfen (Hodson, 1996, S. 129ff; Shavelson et al., 1998). In Bildungsstandards finden sich Hinweise zur Abfolge von Handlungen. So umfasst der Begriff Inquiry die Beschreibung *use of critical and logical thinking* ("National Science Education Standards," 1996, S. 20). *Conceptual understanding* und *understanding of the nature, processes and methods of science* sind zentrale Ziele der National Curriculum (National Research Council, 2012, K1-2: S. 3f; K3: S. 2f). Auch in den praxisnahen Beschreibungen finden sich ähnliche Aussagen zu Lernzielen: *Wissenschaftliches Denken kennenlernen* (Welzel et al., 1998), *Make and justify decisions regarding the methodology; Demonstrate robust understanding of fundamental science concepts (not simply articulating isolated facts and using mathematical algorithms to solve relatively meaningless problems); Set objectives, make decisions, exhibit analytical and reflective thinking, and self-evaluate while inquiring and investigating;* bzw. *Retrieve and use current scientific concepts during authentic inquiry* (Lunetta, Hofstein, & Clough, 2007, S. 419).

Reflektiertes Handeln, Selbstreflektion und konzeptuelles Wissen lassen sich nach diesen Beschreibungen als Voraussetzungen zur Strukturierung der Handlungen im Experimentierpro-

zess verstehen. Alle dargestellten Aspekte haben gemein, dass eine beachten bzw. Berücksichtigung zu einer Verbesserung der Struktur im Experimentierprozess führen sollte. Hat ein Experimentator ein hohes konzeptuelles Wissen, dann sollte er erkennen, was im Experiment (passend zur Planung) als nächstes zu tun ist. Für eine spezifische Situation im Experimentierprozess sollte er entsprechend eine sinnvolle nächste Handlung wählen können. Um die spezifische Situation adäquat zu erkennen, muss er sein Handeln reflektieren und die durchgeführten Handlungen evaluieren können. Bei der Richtigkeit wird bewertet, ob verschiedene Handlungen richtig durchgeführt werden. Bei der Strukturiertheit wird bewertet, ob die Wahl der nachfolgenden Handlung in der spezifischen Experimentiersituation sinnvoll ist. Im Folgenden werden typische sinnvolle Übergänge identifiziert.

### **Idealtypische Abfolge**

In den verschiedenen vorgestellten Modellierungen wird implizit eine Ordnung der Handlungen vorgegeben (Maiseyenko et al., 2011; Schreiber, 2012; Emden, 2011; Gut et al., 2010; vgl. auch Abb. 2). Gehen Experimentatoren entsprechend dieser Reihenfolge die einzelnen Handlungen durch, so folgen sie dem idealtypischen Prozess des Experimentierens. Dabei müssen die jeweiligen Handlungen dem Lösungsweg vorheriger Handlungen entsprechen (Emden, 2011). Am Beispiel der Experimentierkompetenz bedeutet das, wenn nach der Handlung (III) „Experiment planen“ die Handlung (IV) „Versuch fehlerfrei aufbauen“ mit einem zur Planung passenden Versuch durchgeführt wird, dann ist der Handlungsübergang sinnvoll. Kommt nach der Handlung (III) „Experiment planen“ jedoch die Handlung (VI) Daten analysieren und Verallgemeinerungen diskutieren“, dann ist der Handlungsübergang nicht sinnvoll.

### **Kontrolle**

Das Vorgehen kann aber auch strukturiert sein, wenn es nicht strikt dem idealtypischen Vorgehen folgt. Eine Struktur zeigt sich auch, wenn der Experimentator sein Handeln kontrolliert (vgl. Selbstregulation von Arbeitstätigkeiten, z.B. Hacker, 1986; Oesterreich, 1981; Wirth, 2004). In diesem Fall springt er zurück, um die Konsistenz seines Handelns zu analysieren. Dieses Vorgehen kann gerade in aufwendigen Experimenten die Effizienz steigern. Bevor z.B. eine aufwendige Messung durchgeführt wird, prüfen Experimentatoren, ob der Aufbau richtig ist und die Hypothese durch die geplante Messung verifiziert bzw. widerlegt werden kann. Diese Übergänge sind weder falsch noch richtig. Sie helfen dem Experimentator Zeit zu sparen, indem geprüft wird, ob die Durchführung einer Handlung notwendig ist.

## **Fehlerkorrektur**

Ein Ergebnis der Kontrolle kann das Finden von Fehlern sein. Die Struktur wird gesteigert, wenn ein Fehler erkannt und korrigiert wurde (Wiederholung einer Handlung). Ein strukturierter Experimentiervorgang zeigt sich darin, dass gezielt Feedback zu einzelnen Handlungen erzeugt und genutzt wird (Shavelson et al., 1998; Klieme et al., 2001), um zukünftige Handlungen zu planen (Lunetta et al., 2007, S. 419). Die Reflexion der Handlungen beim Experimentieren kann als gelungen erachtet werden, wenn Fehler gefunden und behoben werden. Ein adäquater Umgang mit Messfehlern lässt sich zum Beispiel als Qualitätsmerkmal beim Experimentieren beschreiben (Heinicke, Rieß, & Heering, 2012). Hat ein Proband seine Messung falsch durchgeführt und dies erkannt (Evaluation durch die Erzeugung von Feedback bzw. Kontrolle), so ist es sinnvoll die Messung zu wiederholen (Qualitätssicherung durch Nutzen des erzeugten Feedbacks). Gleichzeitig ist es nicht sinnvoll, mit den falschen Messwerten eine Auswertung zu beginnen. Ein Handlungsübergang zur Korrektur eines Fehlers ist immer als sinnvoll zu beurteilen.

### **2.2.3 Zielorientiertheit**

Nachdem sich die Aspekte der Richtigkeit und der Strukturiertheit direkt auf einzelne Handlungen, beziehungsweise auf die Übergänge zwischen einzelnen Handlungen bezogen haben, beschreibt der Aspekt der Zielorientiertheit eine übergeordnete, den ganzen Prozess charakterisierende Strategie (Shavelson & Ruiz-Primo, 1998). Als Beispielsollen zwei Handlungsfolgen bei der Bestimmung der Wellenlängen zweier unterschiedlicher Laser dienen: Ein Proband führt zunächst alle Messungen und anschließend eine vollständige Auswertung durch. Der andere Proband misst zunächst einen Laser aus und wertet die Ergebnisse aus. Anschließend geht er analog für den zweiten Laser vor. Gehen beide Probanden korrekt und strukturiert vor, so erhalten beide dieselbe (richtige) Lösung. Allerdings benötigt der erste Proband weniger Zeit. Er bestimmt in einem Forschungszyklus die Wellenlänge beider Laser. Der zweite Proband bestimmt im ersten Zyklus die Wellenlänge des ersten Lasers und im zweiten Zyklus die Wellenlänge des zweiten Lasers. Bezogen auf den idealtypischen Experimentierprozess führt der erste Proband einen vollständigen Zyklus durch, während der zweite Proband zunächst einen vollständigen Zyklus durchläuft und dann zurück springt, um einen verkürzten zweiten Zyklus zu durchlaufen. Während dieses Beispiel relativ trivial ist, können gerade bei komplexen Experimenten eine große Anzahl von Zyklen benötigt werden. Durch Vorüberlegungen bzw. die Wahl einer Strategie können die Zyklen reduziert und damit die Bearbeitungszeit minimiert werden. Die Strategien dienen dazu, einzelne Zyklen zu verknüpfen

(Shavelson & Ruiz-Primo, 1998). Klahr (2000) beschreibt als Strategien *Ausprobieren (Trial and Error)*, *Bergsteigen (Hill climbing)*, *Planung (Planning)* und *Analogie* (Klahr & Dunbar, 2000). Unter *Ausprobieren* versteht er das wahllose Durchführen von Zyklen, bis die Fragestellung befriedigend gelöst wurde. Entsprechend findet keine Interaktion zwischen den einzelnen Zyklen statt. Dennoch müssen die Aspekte der Richtigkeit und der Strukturiertheit in ausreichender Qualität durchgeführt werden, damit die einzelnen Zyklen eine Erkenntnis liefern. Einfach gesagt kann diese Erkenntnis sein: das Problem wurde gelöst oder eben nicht gelöst. Die Strategie *Bergsteigen* steht synonym für das Besteigen eines Berggipfels. Eine Möglichkeit den Gipfel zu besteigen besteht darin, immer den höchsten möglichen Anstieg zu nehmen. Bezogen auf das Experimentieren bedeutet dies, dass immer das größte Problem bearbeitet wird. Dieses löst die experimentierende Person durch einen Experimentierzyklus und wählt dann erneut das größte Problem. Dieses Vorgehen zeigt der zweite Proband aus dem Beispiel. Er hat zwei Probleme, er kennt die Wellenlänge von Laser 1 und Laser 2 nicht. In diesem Fall wären die Probleme gleich groß. Er würde eines der beiden Probleme lösen und die Wellenlänge von Laser 1 bestimmen. Zu diesem Zeitpunkt wäre er noch nicht am Ziel (bzw. auf dem Gipfel), also würde er das nächste Problem lösen. In diesem Fall würde er die Wellenlänge von Laser 2 bestimmen. Damit hätte er alle Probleme gelöst und hätte seine Aufgaben erfolgreich beendet (wäre auf dem Gipfel angekommen). Bei der Strategie *Planung* würde sich der Experimentator vor dem ersten Experimentierzyklus überlegen, wie er die Aufgabe vollständig lösen kann und welche Handlungen durchgeführt werden müssen. Anschließend würden die Handlungen in einem Zyklus abgearbeitet. Diese Strategie zeigt der erste Proband. Er muss für jeden Laser eine Messung durchführen. Dies tut er in einem Zyklus, bevor er zur nächsten Handlung übergeht.

Die letzte Strategie *Analogie* setzt voraus, dass vorhandenes Wissen zu einem identischen Problem oder einer analogen Fragestellung vorhanden ist. Wenn ein Experimentator ein vergleichbares Problem und den entsprechenden Lösungsweg kennt, so kann er ihn auf die angegebene Fragestellung transferieren und umsetzen. Ein Beispiel dieses Vorgehens sind Praktikumsleiter. Praktikumsleiter müssen nicht explizit über die Lösung einer Aufgabe nachdenken. Sie *wissen* wie es geht, da sie Strategien bzw. Lösungswege für analoge oder identische Probleme kennen. Ein Beispiel dafür ist das Justieren einer Linse mit dem Brennpunkt auf einem Schirm mit einer optischen Bank. Ein Praktikumsleiter (Experte) berücksichtigt die Brennweite der Linse und achtet instinktiv auf einen senkrechten Strahlengang. Ein weniger erfahrener Experimentator, beispielsweise ein Assistent oder Student im Praktikum denkt explizit über das relevante Vorgehen nach- wählt also die Strategie *Planung*. Er bemerkt die

Notwendigkeit eines senkrechten Strahlengangs und überlegt, welche Position geeignet ist. Nach der Planung lässt sich die Handlung umsetzen. Nehmen wir als Beispiel für das *Bergsteigen* einen Studienanfänger. Sein erstes Problem könnte sein, die Linse in die Halterung zu bekommen. Ist dies gelungen, so gibt es möglicherweise eine horizontale oder vertikale Verschiebung der Abbildung auf dem Schirm. Der Studienanfänger würde den Strahlengang kontrollieren und anpassen. Abschließend wird die Unschärfe des Brennpunktes durch eine Veränderung der Position auf der optischen Bank hergestellt. Für das *Ausprobieren* wäre ein Schüler in der Mittelstufe ein geeignetes Beispiel. Er wird die Linse in die Halterung bekommen. Doch er kann die einzelnen Probleme nicht erkennen und für sich nutzen. Entsprechend *spielt* er mit der Linse herum, bis er zufällig die richtige Position findet.

Nach Klahr (2000) nutzen Novizen eher die Strategie *Ausprobieren*, während mit steigender Qualität des Experimentierens hochwertigere Strategien ausgehend von *Bergsteigen* über *Planung* bis hin zu *Analogien* genutzt werden (vgl. auch Shavelson & Ruiz-Primo, 1998). Gleichzeitig steigt ausgehend von *Analogien*, über *Planung* und *Bergsteigen* bis hin zum *Ausprobieren* die notwendige Anzahl an Zyklen, weil der Experimentator einen größeren Anteil der Problemlösung in einem einzelnen Zyklus realisieren kann. Dies führt zu dem Schluss, dass Novizen mehr Zyklen brauchen als Experten. Entsprechend kann die Anzahl an Zyklen als Maß für die Qualität der Zielorientiertheit genutzt werden.

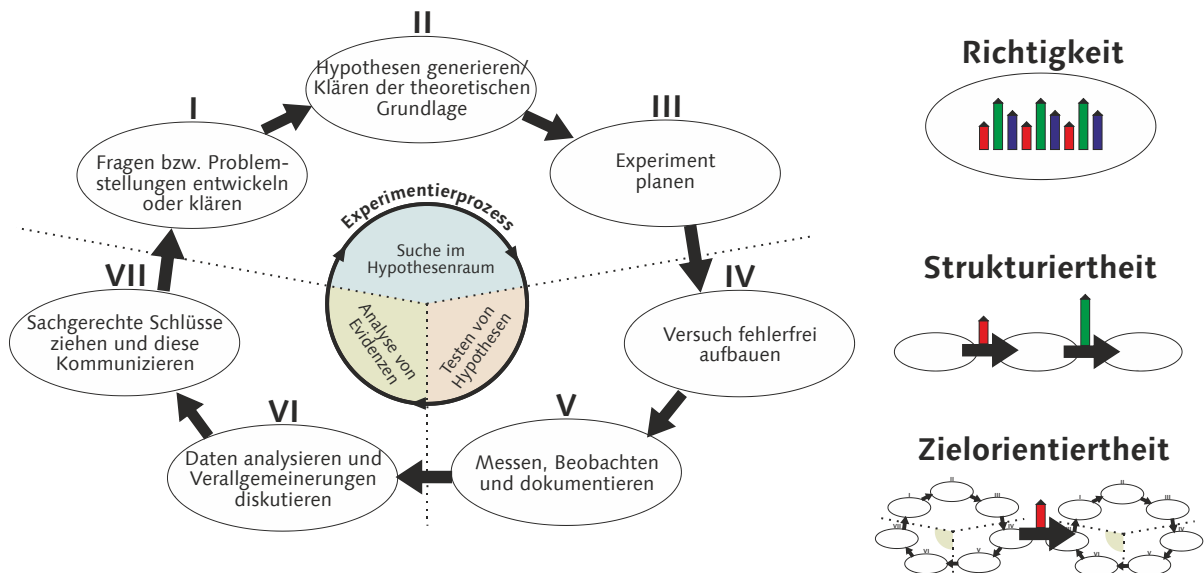
### **2.3 Zusammenführung des Modells der Experimentierkompetenz**

In Kapitel 2 wurde beschrieben, dass Kompetenz konkrete Fähigkeiten und Fertigkeiten voraussetzen, die Grundlage zur Durchführung von Handlungen sind. Der Experimentierprozess umfasst sieben Handlungen, damit umfasst Experimentierkompetenz die Fähigkeiten und Fertigkeiten diese Handlungen durchzuführen. Die konkreten Handlungen wurden in Kapitel 2.1 durch die Fähigkeiten und Fertigkeiten beschrieben und abgegrenzt.

Die Ausprägung von Experimentierkompetenz ist gemeinhin als die mittlere Ausprägung aller Fähigkeiten bzw. Fertigkeiten modelliert. In Kapitel 2.2 wurde jedoch gezeigt, dass ein alternatives Maß zur Bewertung des Experimentierprozesses denkbar ist. Das Qualitätsmaß der Richtigkeit beschreibt die Qualität einzelner durchgeführter Handlungen. Als Qualitätsmaß für die spezifische Steuerung des Experimentierprozesses dient die Strukturiertheit. Dabei spielt das Generieren von Feedback, das Erkennen von Fehlern und die Kontrolle der Handlungen eine Rolle. Die Zielorientiertheit beschreibt die Qualität einer globalen Strategie, die einzelne Zyklen im iterativen Kreisprozess miteinander verknüpft. Diese Aspekte zur zielorientierten

Lösung von Problemen bzw. Fragestellungen stellen, wie in Kapitel 2 dargestellt, eine Handlungskompetenz dar.

Durch die Adaption der Qualitätskriterien auf den durch spezifische Handlungen ausdifferenzierten potentiell iterativen Kreisprozess des Experimentierens ergibt sich das in Abbildung 3 dargestellte Modell der Experimentierkompetenz.



**Abbildung 3: Modell der Experimentierkompetenz mit Handlungen und Gütemaßen.**

In dem dargestellten Modell sind der ausdifferenzierte Kreisprozess (links) und die Qualitätsmaße (rechts) dargestellt. Die Darstellung des ausdifferenzierten Kreisprozesses wird in Kapitel 2.1 beschrieben. Die Darstellung der Qualitätsmaße zeigt oben die Richtigkeit. Bewertet wird die Richtigkeit der einzelnen Handlungen. Die einzelnen Handlungen sind als Pfeile dargestellt. Als Maß der Richtigkeit dient die Farbe und Höhe der Pfeile. Die Bewertung der Strukturiertheit basiert auf dem Übergang der Handlungen. Handlungen sind als Ellipsen dargestellt. Bewertet wird ob der Übergang von einer Handlung zur nächsten sinnvoll ist. Die Qualität des Übergangs ist erneut mithilfe von Pfeilen dargestellt. Ihre Höhe und Farbe beschreiben die Qualität der Übergänge. Bei der Zielorientiertheit wird die Abfolge ganzer Zyklen bewertet. Bewertet wird, ob die Abfolge der Zyklen zu einer Problemlösung konvergiert. Analog beschreibt der Pfeil die Qualität der Konvergenz.



### **3 Erfassung von Experimentierkompetenz**

Im vorherigen Kapitel wurde ausgehend von bestehenden Arbeiten ein Modell der Experimentierkompetenz hergeleitet. In diesem Kapitel wird der Stand der Forschung zur Diagnose von Experimentierkompetenz aufgearbeitet. Die Diagnostik ist charakterisiert durch das Spannungsfeld, dass durch den Wunsch nach einer möglichst authentischen Testsituation einerseits und dem Streben nach Ökonomie bei der Testdurchführung und Auswertung andererseits aufgespannt wird. In einer möglichst authentischen Testsituation verhält sich ein Proband natürlich und die diagnostizierte Leistung entspricht am ehesten seinen tatsächlichen Fähigkeiten (Harmon, 1997). Wird die Testsituation aus Gründen der ökonomischeren Testdurchführung, Auswertung oder Ähnlichem angepasst, so ändert sich auch die Wahrnehmung und ggf. das Verhalten des Probanden. Die Veränderungen der Testsituation beruhen auf der Hypothese, dass die diagnostizierte Leistung sich durch die Veränderung der Testsituation nur unbedeutend verändert. Es entsteht ein weniger aufwendiges, also ökonomischeres Verfahren, dass eine weniger authentische Testsituation rechtfertigt (z.B. Organisation for Economic Cooperation and Development, 2007). Die Veränderung der Testsituation kann aber auch zu einer nicht vernachlässigbaren Diskrepanz zwischen der diagnostizierten Leistung und den tatsächlichen Fähigkeiten des Probanden führen (Kane, 2001; Messick, 1989). Das populärste Beispiel ist die Testangst. Ein Schüler der gute Fähigkeiten hat und diese im Unterricht erfolgreich nutzt, kann in einem Test seine Fähigkeiten nicht abrufen. Entsprechend werden ihm unzureichende Fähigkeiten attestiert.

Die Notwendigkeit ökonomischer Testinstrumente ist unbestritten. Vor allem bei großen Vergleichsstudien wie PISA oder TIMSS wird deutlich, dass für jeden einzelnen Probanden nur ein begrenzter Aufwand bei der Testdurchführung und Auswertung aufgewendet werden kann. Doch auch in kleineren Studien wird versucht die Ökonomie zu verbessern. Durch eine verbesserte Ökonomie kann mit den verfügbaren Ressourcen eine größere Anzahl an Probanden untersucht werden. Mit steigender Stichprobenzahl steigen die Möglichkeiten in der Auswertung, es wird eine Verbesserung der Signifikanz erwartet und auch die Zulässigkeit von Verallgemeinerungen wächst.

Zur Untersuchung vom Einfluss der Anpassung einer Testsituation zur Verbesserung der Ökonomie auf die mögliche Diskrepanz zwischen diagnostizierter und tatsächlicher Performance werden in diesem Kapitel die nachfolgenden 23 Testinstrumente analysiert: FISS (Comber & Keeves, 1973), Test of Science Processes (Tannenbaum, 1971), TOES (Fraser, 1980), SISS P&P (IEA, 1988), SISS RE (Tamir, Doran, & Chye, 1992), TIMSS 95 (Beaton, 1996), TIMSS 99 (Martin, 2000), TIMSS 03 (Martin, 2004), TIMSS 07 (Martin, 2008),

TIMSS 11 (Martin, 2012), TIMSS-R (Harmon, 1997), NAW (Walpuski, 2006) & (Klos, 2008), PISA (Organisation for Economic Co-operation and Development, 2007), Hammann (Hammann & Hoi Phan, 2007), Hammann (Hammann et al., 2008), NEAPs (National Assessment Governing board U.S. Department of Education, 2008), NEAPs (National Assessment Governing board U.S. Department of Education, 2008), HarmoS (Wissenschaftliches Konsortium HarmoS Naturwissenschaften+, 2008), HarmoS Realexperimentiertest (Wissenschaftliches Konsortium HarmoS Naturwissenschaften+, 2008), AG Bremen (Maiseyenko et al., 2011), Emden (Emden, 2011), Schreiber (Schreiber, 2012).

Zur Analyse werden in Kapitel 3.1 Aspekte der Erfassung zur Einordnung der Testinstrumente im Spannungsfeld der Authentizität und der Ökonomie hergeleitet. Diese Aspekte werden an ausgewählten Testinstrumenten dargestellt. Mithilfe der Aspekte kann ein Testinstrument im Spannungsfeld der Ökonomie und Authentizität in Kapitel 3.2 eingeordnet werden. Typische Vertreter von ökonomischen und authentischen Testinstrumenten werden vorgestellt. In Kapitel 3.3 wird die Diskrepanz zwischen der diagnostizierten und der tatsächlichen Leistung von ökonomischen und authentischen Testinstrumenten fokussiert. Da die tatsächliche Leistung nicht zugänglich ist, wird die Testgüte als Maß für die postulierte Passung zwischen tatsächlicher und diagnostizierter Leistung genutzt. In den Sozialwissenschaften ist die Testgüte ein Maß dafür, wie adäquat das zu diagnostizierende Konstrukt durch das Testinstrument abgebildet wird und inwieweit Schlussfolgerungen, basierend auf generierter Evidenz, zulässig sind. Die Testgüte wird auf der Einordnung der Testinstrumente im Spannungsfeld zwischen Authentizität und Ökonomie bewertet.

Auf Grundlage der Bewertung werden durch die Anpassung der Testsituation zur Verbesserung der Ökonomie Probleme bzgl. der Testgüte identifiziert. Lösungsansätze zur Verbesserung der Testgüte sowie zur Verbesserung der Ökonomie bei gleichbleibender Testgüte werden diskutiert. Abschließend werden die aktuellen Forschungsdefizite für die Entwicklung eines adäquaten Testinstruments zur Erfassung von Experimentierkompetenz abgeleitet.

### **3.1 Aspekte der Erfassung von Experimentierkompetenz**

Um Experimentierkompetenz zu diagnostizieren, erscheint eine authentische Experimentiersituation, bei der differenziert einzelne Handlungen bewertet und zur Diagnostik der Experimentierprozess berücksichtigt werden, ideal (Harmon, 1997, Seite 5). Das Gütemaß der Richtigkeit kann aus der differenzierten Bewertung der einzelnen Handlungen gewonnen werden. Die Gütemaße der Strukturiertheit und Zielorientiertheit lassen sich auf Basis des Experimentierprozesses ableiten. Um eine solche Diagnostik zu erreichen, kann ein Experte die Hand-

lungen, Aufzeichnungen und das Vorgehen eines Probanden beobachten und bewerten (z.B. Emden, 2011). Dieses Vorgehen stellt einen enormen Aufwand dar und ist aus Gründen der Testökonomie selten umsetzbar. Oft werden deshalb theoretische Aufgaben, teilweise zu einzelnen Fähigkeiten bzw. Fertigkeiten gestellt und das Ergebnis als Produkt des gesamten Experimentierprozesses betrachtet (z.B. Hammann et al., 2008).

Die dargestellten Verfahren lassen sich im Gegenstand der Bewertung unterscheiden. Im ersten Fall wird das Vorgehen als Prozess bewertet, während im zweiten Fall das Endergebnis als Produkt die Grundlage der Bewertung darstellt. Ein weiteres Unterscheidungsmerkmal ist das Aufgabenformat. Die Aufgaben unterscheiden sich zwischen Aufgaben die praktisch, durch tatsächliches Experimentieren, bearbeitet werden und Aufgaben, die durch theoretische Überlegungen zum Experimentieren gelöst werden. Letztlich liegt eine Unterscheidung vor, wenn einzelne voneinander unabhängige Fähigkeiten bzw. Fertigkeiten abgefragt werden. In diesem Fall sind die Fähigkeiten und Fertigkeiten kein Bestandteil des Prozesses. Eine Modellierung ohne Bezug zum Prozess führt zu einem anderen Konstrukt des Experimentierens. Im Folgenden werden die beschriebenen Aspekte des Bewertungsgegenstandes, des Aufgabenformats und der Modellierung detailliert beschrieben.

### **3.1.1 Bewertungsgegenstand: Produkt vs. Prozess**

Ein Aspekt zur Unterscheidung der Testsituation ist der Bewertungsgegenstand. Die am weitesten verbreitete Art der Bewertung bei Testinstrumenten fokussiert auf ein Endergebnis oder auf Teilergebnisse (z.B. Comber & Keeves, 1973; Martin, 2004; 2008; Wissenschaftliches Konsortium HarmoS Naturwissenschaften+, 2008). Bewertet wird, ob das Endergebnis (als Produkt) bzw. ob Teilergebnisse (als Produkte) korrekt sind. Entsprechend ist diese Art des Bewertungsgegenstandes als produktorientiertes Auswerteverfahren bekannt (Schreiber, Theyßen, & Schecker, 2011b; Emden, 2011; Shavelson et al., 1999).

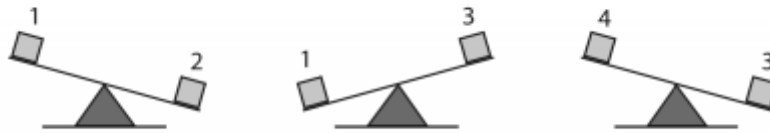
Die zugrunde liegende Hypothese ist, dass es ausreicht nur das Endergebnis oder bestimmte Teilergebnisse als Gegenstand der Bewertung zu nutzen. Ist das Ergebnis richtig, wird dem Probanden Experimentierkompetenz attestiert, ist es falsch, so wird dem Probanden keine Experimentierkompetenz attestiert. Dieses Verfahren ist wenig aufwendig und daher im Spannungsfeld als ökonomisch einzuordnen.

Als Beispiel für die Bewertung des Produkts ist in Abbildung 4 eine Aufgabe aus TIMSS 2015 dargestellt (Mullis & Martin, 2013). Die Aufgabe umfasst eine Beschreibung einer Situation (mit einer Abbildung) und eine Fragestellung. Die Aufgabe umfasst nach dem Modell der Experimentierkompetenz (vgl. II) Teile der Handlungen *V Messen, Beobachten und dokumen-*

tieren und VII Sachgerechte Schlüsse ziehen und diese Kommunizieren. Aufgaben zum Produkt sind üblicherweise, wie auch hier, im Multiple-Choice-Verfahren gestellt. Eine produktorientierte Auswertung lässt sich jedoch auch mit dem offenen Antwortformat erreichen.

Stephanie has a balance and four cubes (1, 2, 3, 4). The cubes are made of different materials.

She puts two cubes at a time on the balance and observes the following results.



What can she conclude about the weight of cube 2?

- A It is heavier than cubes 1, 3, and 4.
- B It is heavier than cube 1 but lighter than cubes 3 and 4.
- C It is heavier than cube 3 but lighter than cubes 1 and 4.
- D It is heavier than cube 4 but lighter than cubes 1 and 3.

**Abbildung 4: Beispielaufgabe einer produktorientierten Auswertung (aus Mullis & Martin, 2013).**

Mit Bezug auf die Modellierung der Experimentierkompetenz in Kapitel 2 kann das Gütemaß der Richtigkeit mit produktorientierten Aufgaben diagnostiziert werden. Die Gütemaße der Strukturiertheit und Zielorientiertheit lassen sich nicht explizit bewerten. Das Verfahren eignet sich vor allem für einzelne Handlungen im Experimentierprozess und ggf. kurze (Teil)Aufgaben.

Bei längeren Aufgaben ist die Bewertung von Endergebnissen weniger gut geeignet. Hat ein Proband alle Schritte im Lösungsprozess korrekt durchgeführt und nur an einer Stelle einen nicht korrigierten Fehler gemacht bzw. eine Handlung nicht hinbekommen, so wird er voraussichtlich zu einer falschen Antwort kommen. In diesem Fall wäre es nicht angebracht dem Probanden keine Experimentierkompetenz zu diagnostizieren. Das heißt, dass bei längeren Aufgaben nur die extremen Ausprägungen (1 und 0) korrekt aufgelöst werden können. Ein Proband, der zum richtigen Ergebnis kommt (ggf. auch mit fehlerhaften Handlungen die korrigiert wurden), erhält passenderweise einen Punkt. Ein Proband, der alles falsch macht oder gar nicht weiß, wie er das Problem angehen soll, erhält 0 Punkte. Eine Differenzierung zwischen den Schritten ist nicht möglich. Durch die Einbeziehung von Teilergebnissen (z.B. Schreiber, 2012) kann die Differenzierung verbessert werden. In diesem Fall würde der Lö-

sungsprozess durch mehrere Produkte angenähert und auf diese Weise diagnostiziert werden. Ein Beispiel dafür ist ein Aufgabenstamm mit mehreren Teilaufgaben. Als anderes Beispiel dienen die Praktika im universitären Bereich. Dort werden zur Beurteilung der Leistung der Studierenden neben einem Prüfungsgespräch die Protokollhefte als Bewertungsgegenstand genutzt (Neumann, 2004; Schumacher & Planinšič, 2007; Theyßen, 1999). Die Aufzeichnungen innerhalb der Protokolle sind Produkte zu den einzelnen Handlungen bei der Durchführung der Experimente. Auch diese Bewertungsverfahren basieren ausschließlich auf Produkten und werden entsprechend als produktorientierte Auswerteverfahren klassifiziert. Im Vergleich zur Bewertung ausschließlich des Endergebnisses können die Gütemaße der Strukturiertheit und Zielorientiertheit durch die Annäherung des Prozesses diagnostiziert werden.

Im zweiten Kapitel wurde dargestellt, dass das Experimentieren mehr als die Summe der einzelnen Handlungen ist. Es handelt sich um einen Prozess. Entsprechend kann für eine höhere Authentizität der Lösungsprozess, vor allem für die Gütemaße der Strukturiertheit und Zielorientiertheit, ein Bestandteil des Bewertungsgegenstandes sein.

Um den Prozess zu bewerten, können Liveratings durchgeführt werden. Liveratings werden üblicherweise für die Note zur mündlichen bzw. sonstigen Mitarbeit in der Schule oder in mündlichen Prüfungen angewendet. Alternativ kann der Prozess auch aufgezeichnet werden, beispielsweise mit Tonband, Videokamera oder im Falle eines Computers mit Logdaten. In diesem Fall kann der Prozess nachträglich bewertet werden (Emden, 2011, Seite 67ff; Schreiber, 2012). Die Einbeziehung des Experimentierprozesses als Bewertungsgegenstand erhöht die Authentizität der Messung. Gleichzeitig erfordert dieses Vorgehen eine Bewertung durch einen Beobachter. Wenn der Verlauf des Experimentierens jedes Probanden von einem Beobachter bewertet werden muss, entsteht ein enormer Arbeitsaufwand und entsprechend leidet die Ökonomie des Tests. Zwar ist es schon gelungen, diesen Aufwand durch automatische Analysen von Bildschirmdaten bei Simulationsexperimenten zu reduzieren, allerdings ist bisher noch nicht abschließend geklärt, ob mit Simulationen dasselbe Konstrukt erfasst werden kann, wie durch reale Experimentiersituationen (Schreiber, 2012; Shavelson & Ruiz-Primo, 1998, vgl. III.1.2). Das skizzierte Verfahren zur Auswertung wird als prozessorientiertes Verfahren bezeichnet. Der Unterschied zu den produktorientierten Auswerteverfahren ist, dass der Prozess durch die Vielzahl<sup>6</sup> an Produkten authentisch abgebildet werden kann.

---

<sup>6</sup> In Videoanalysen werden die Handlungen der Probanden je nach Arbeit in Intervallen von 10 Sekunden bewertet (z.B. Schreiber (2012)). Bei einer Testdauer von 15 Minuten ergeben sich damit 90 Produkte, die eine authentische Abbildung des Prozesses glaubhaft machen.

Die dargestellten Ergebnisse lassen sich wie folgt zusammenfassen. Zur Diagnose von Experimentierkompetenz kann als Bewertungsgegenstand das Produkt, für eine produktorientierte Auswertung genutzt werden oder durch eine Vielzahl an Produkten kann der Prozess für eine prozessorientierte Auswertung authentisch abgebildet werden. Die Diagnose basierend auf einer produktorientierten Auswertung als Bewertungsgegenstand ist weniger authentisch, dafür aber auch weniger aufwendig. Die produktorientierte Auswertung lässt sich im Spannungsfeld zwischen ökonomischen und authentischen Testinstrumenten als ökonomisches Merkmal eines Testinstruments bewerten. Die aufwendige prozessorientierte Auswertung wird hingegen als authentisches Merkmal eingestuft.

### **3.1.2 Aufgabenformat: Praktisch vs. Theoretisch**

Bei den Aufgabenmerkmalen von Testinstrumenten zur Erfassung von Experimentierkompetenz unterscheidet man, ob Probanden tatsächlich experimentieren oder theoretisch über das Experimentieren nachdenken (Shavelson R. J. et al., 1993; Shavelson & Ruiz-Primo, 1998). Dieses Merkmal eines Testinstruments wird als Testformat oder Aufgabenformat bezeichnet. Bei theoretischen Aufgaben wird dem Probanden eine Experimentiersituation präsentiert. Der Proband führt einzelne Handlungen ohne Interaktion mit Versuchsmaterial aus. Dies ist das gängige Format in Experimentiertests (z.B. Comber & Keeves, 1973; Tannenbaum, 1971; IEA, 1988; Beaton, 1996; National Assessment Governing board U.S. Department of Education, 2008). Das Beispiel zum Bewertungsgegenstand in Abbildung 4 dient auch als Beispiel für eine theoretische Experimentieraufgabe. Der Vorteil dieses Formats ist die relativ kurze Bearbeitungszeit und der geringe Aufwand während der Testdurchführung. Allerdings können Teile des Konstrukts der Experimentierkompetenz nicht erfasst werden. Dies sind aus dem Aspekt Richtigkeit beispielsweise die Handlungen *IV Versuch fehlerfrei aufbauen* und *V Messen, Beobachten und dokumentieren*. Zudem kann kein Feedback generiert werden und es kommt zu keiner Interaktion zwischen Handlungen oder Experimentierzyklen. Nach der Modellierung von Experimentierkompetenz aus Kapitel 2 wäre dies notwendig, um die Aspekte Strukturiertheit und Zielorientiertheit abbilden zu können. Daraus folgt, dass die Authentizität von Testinstrumenten, in denen es zu keiner Interaktion der Probanden mit Versuchsmaterial kommt, eingeschränkt ist.

Im Gegensatz dazu sind Experimentiertests entwickelt worden, in denen Probanden tatsächlich mit Versuchsmaterial interagieren. International werden diese Instrumente mit dem Begriff "*hands-on*" charakterisiert, im deutschsprachigen Raum werden sie als Realexperimentiertests bezeichnet (z.B. Klahr, Triona, & Williams, 2007; Tamir, Doran, & Chye, 1992;

Harmon, 1997; Emden, 2011). Bei den Realexperimentiertests ist es möglich, dass Probanden nur spezifische Handlungen durchführen müssen (Maiseyenko et al., 2011; Nawrath, Maiseyenko, & Schecker, 2011) oder den kompletten Prozess des Experimentierens durchlaufen (Emden, 2011). Das Format des Realexperiments erfordert grundsätzlich mehr Bearbeitungszeit. Zudem muss entsprechendes Material für jeden Probanden angeschafft werden. Bedingt dadurch können generell weniger Aufgaben bearbeitet werden und die Kosten für die Testentwicklung sind deutlich erhöht, gegenüber dem Entwickeln eines Papier-und-Bleistift-Tests. Allerdings gibt es Erkenntnisse, dass sich mit Tests verschiedener Formate nicht dasselbe Konstrukt erfassen lässt (Schreiber, 2012; Shavelson & Ruiz-Primo, 1998).

In den letzten Jahren gab es vermehrt Untersuchungen zu einem neuen Format, das den geringen Aufwand eines Papier-und-Bleistift-Tests und die Interaktion des Realexperimentiertests verbindet, die Computersimulation (z.B. Shavelson & Ruiz-Primo, 1998). Indes ist bisher noch nicht ausreichend geklärt, ob sich mit Simulationen dasselbe Konstrukt wie bei Experimentiertests erfassen lässt (Schreiber, 2012).

Es lässt sich zusammenfassen, dass zur Diagnose von Experimentierkompetenz das Aufgabenformat variiert werden kann. In theoretischen Aufgaben werden die Probanden in eine Experimentiersituation hineinversetzt und interagieren nicht mit Versuchsmaterial. Bei praktischen Aufgaben experimentieren die Probanden tatsächlich. Bei der Diagnose mit theoretischen Aufgaben ist weniger Authentizität gegeben, dafür sind sie ökonomischer. Entsprechend werden theoretische Aufgaben als ökonomisches Merkmal eingestuft. Praktische Aufgaben sind aufwendiger und authentischer, sie werden daher als authentisches Merkmal bei dem Aufgabenformat eingestuft.

### **3.1.3 Modellierung: Einzelne Handlungen vs. Gesamter Prozess**

Mit den meisten Testinstrumenten werden die Fähigkeiten und Fertigkeiten beim Experimentieren getrennt voneinander gemessen und daraus wird die Experimentierkompetenz berechnet (z.B. über einen Summenscore, vgl.: 1988; Tamir, Doran, & Chye, 1992; Walpuski, 2006; Klos, 2008; Hammann et al., 2008). Das in Abbildung 4 dargestellte Beispiel zum Bewertungsgegenstand umfasst Fähigkeiten und Fertigkeiten der Handlungen *V Messen, Beobachten und dokumentieren* und *VII Sachgerechte Schlüsse ziehen und diese Kommunizieren*. Mit anderen Aufgaben werden Fähigkeiten und Fertigkeiten zu verschiedenen weiteren Handlungen diagnostiziert. Als Gesamtmaß für Experimentierkompetenz werden die Bewertungen zu den einzelnen Aufgaben miteinander verrechnet. Die Bearbeitungsprozesse dieser konkreten

Aufgaben lassen sich, im Vergleich zu offenen Experimentieraufgaben, gut vorhersagen. Zudem kann durch die Aufgabenentwicklung leichter sichergestellt werden, dass alle Handlungen durchlaufen werden. In einer offenen Experimentieraufgabe kann ein Proband ggf. an einer Stelle scheitern und gar nicht in die Gelegenheit kommen Handlungen zur Auswertung zu zeigen. Gleichzeitig kann die Gewichtung der Handlungen in einem offenen Experiment deutlich schwerer gesteuert werden, als durch konkrete Aufgaben zu einzelnen Handlungen. Nach den Erkenntnissen von Kapitel 2 ist Experimentieren aber mehr als die Summe der einzelnen Handlungen. Ein Proband muss Handlungen in eine sinnvolle Abfolge bringen (Strukturiertheit) und mittels einer Strategie zur Lösung des Problems/ der Aufgabe konvergieren lassen (Zielorientiertheit). Diese Aspekte könnten nicht diagnostiziert werden, wenn der Proband nur einzelne, voneinander losgelöste Handlungen vollführt. Damit auch die Gütemaße der Strukturiertheit und Zielorientiertheit diagnostiziert werden können, ist bei einigen Testinstrumenten die Abfolge und Interaktion zwischen Handlungen relevant (z.B. National Assessment Governing board U.S. Department of Education, 2008; Emden, 2011; Schreiber, 2012). Dies wird realisiert, indem verschiedene Handlungen zur Bearbeitung einer Aufgabe benötigt werden. Beispielsweise muss ein Proband einen Versuch planen, diesen aufbauen, Messungen durchführen und am Ende durch eine Berechnung die Messdaten analysieren. Durch die Bearbeitung lässt sich Feedback aus den Erkenntnissen des Lösungsprozesses ziehen und ggf. für die weitere Bearbeitung nutzen. Ein externes Feedback zu den einzelnen Handlungen durch Zwischenlösungen oder einen Testleiter gibt es nicht. Zur Diagnose der Zielorientiertheit wird in einigen Arbeiten ein übergeordnetes konzeptuelles bzw. strategisches Verständnis einbezogen (Shavelson & Ruiz-Primo, 1998; Klahr & Dunbar, 2000; Schreiber, 2012). Dazu werden vor allem Aufgaben genutzt, in denen Probanden komplexere Problemstellungen selbstständig bearbeiten müssen. In diesen Fällen müssen alle Handlungen durchgeführt, aufeinander bezogen und durch eine Strategie verknüpft werden. Diese Aufgaben sind sehr authentisch. Allerdings ist es schwer die spezifischen Lösungsprozesse vorherzusagen. Verschiedene Probanden könnten komplett unterschiedliche Lösungswege wählen. Mit steigender Freiheit in der Bearbeitung ergeben sich kreative (und ungewollte) Lösungsstrategien. Es lässt sich nur schwer sicherstellen, dass alle Handlungen und Aspekte des Experimentierens in den verschiedenen Lösungswegen relevant sind und gleichzeitig adäquat abgebildet werden können. Entsprechend schwierig wird es, vergleichbare Aufgaben zu entwickeln. Bei kreativen Lösungsstrategien besteht zudem die Gefahr, dass andere Konstrukte einen nicht zu vernachlässigbaren Einfluss auf die Testleistung haben.



Als weitere Schwierigkeit sollen Probanden zur Diagnostik der Zielorientiertheit mehrere Experimentierzyklen durchlaufen können. Damit steigt die Bearbeitungsdauer einer Aufgabe. Insgesamt gibt es bei Testinstrumenten, in denen Probanden den gesamten Prozess des Experimentierens durchlaufen, die Gefahr, dass nicht alle Aspekte der Experimentierkompetenz relevant sind und abgebildet werden können, und dass konstruktferne Aspekte relevanten Einfluss auf die Testleistung haben. Zudem ist die Ökonomie eingeschränkt. Auf der anderen Seite können Testinstrumente mit eingeschränktem Experimentierprozess das Konstrukt nicht authentisch erfassen und die Aspekte der Strukturiertheit und Zielorientiertheit werden vernachlässigt.

### **3.2 Klassifizierung von Testinstrumenten**

Im ersten Teil von Kapitel 3 wurden drei Aspekte zur Einordnung von Testinstrumenten im Spannungsfeld zwischen Ökonomie und Authentizität identifiziert: der Bewertungsgegenstand (Produkt vs. Prozess), das Aufgabenformat (Theoretisch vs. Praktisch) und die Modellierung (unabhängige Handlungen vs. Gesamtprozess). Jede Entscheidung über einen dieser Aspekte kann nach der Beschreibung aus Kapitel 3.2 als (eher) authentisch oder (eher) ökonomisch beurteilt werden. Ein Testinstrument kann bzgl. aller drei Aspekte ökonomisch sein. Dies wird im Folgenden als ökonomisches Testinstrument beschrieben. Ist einer der drei Aspekte authentisch und die anderen ökonomisch, so wird er fortan als eher ökonomisch bezeichnet. Entsprechend folgt für ein Testinstrument mit zwei authentischen und einem ökonomischen Aspekt die Bezeichnung eher authentisch. Sind alle Aspekte authentisch, so wird das Testinstrument als authentisch bezeichnet. Die Klassifizierung im Spannungsfeld der Authentizität und der Ökonomie ist in Tabelle 1 dargestellt.

Insgesamt wurden 23 Testinstrumente analysiert. Eine vollständige Auflistung findet sich im Anhang unter A.2. Im Folgenden werden jeweils typische Testinstrumente als Repräsentanten für authentische-, eher authentische-, eher ökonomische- und ökonomische Testinstrumente vorgestellt. Im anschließenden Abschnitt dieses Kapitels wird die Testgüte von authentischen, eher authentischen, eher ökonomischen und ökonomischen Testinstrumenten innerhalb der Gruppierungen und global verglichen. Das Ziel ist es Vorhersagen darüber zu treffen, welchen Einfluss die verschiedenen Aspekte auf die Testgüte haben.

**Tabelle 1: Klassifizierung von Testinstrumenten im Spannungsfeld von Authentizität und Ökonomie.**

<b>Beurteilung des Testinstruments</b>	<b>Bewertungsgegenstand</b> (Produkt = ökonomisch; Prozess = authentisch)	<b>Aufgabenformat<sup>7</sup></b> (Papier-und-Bleistift-Test Papier-und-Bleistift-Test = ökonomisch; Realexperiment = authentisch)	<b>Modellierung</b> (Unabhängig = ökonomisch; Interaktion = authentisch)
<b>Authentisch</b>	Prozess	Realexperiment	Interaktion
<b>Eher authentisch</b>	Prozess	Realexperiment	Unabhängig
	Prozess	Papier-und-Bleistift	Interaktion
	Produkt	Realexperiment	Interaktion
<b>Eher ökonomisch</b>	Produkt	Realexperiment	Unabhängig
	Produkt	Papier-und-Bleistift	Interaktion
	Prozess	Papier-und-Bleistift	Unabhängig
<b>ökonomisch</b>	Produkt	Papier-und-Bleistift	Unabhängig

### 3.2.1 Authentische Testinstrumente

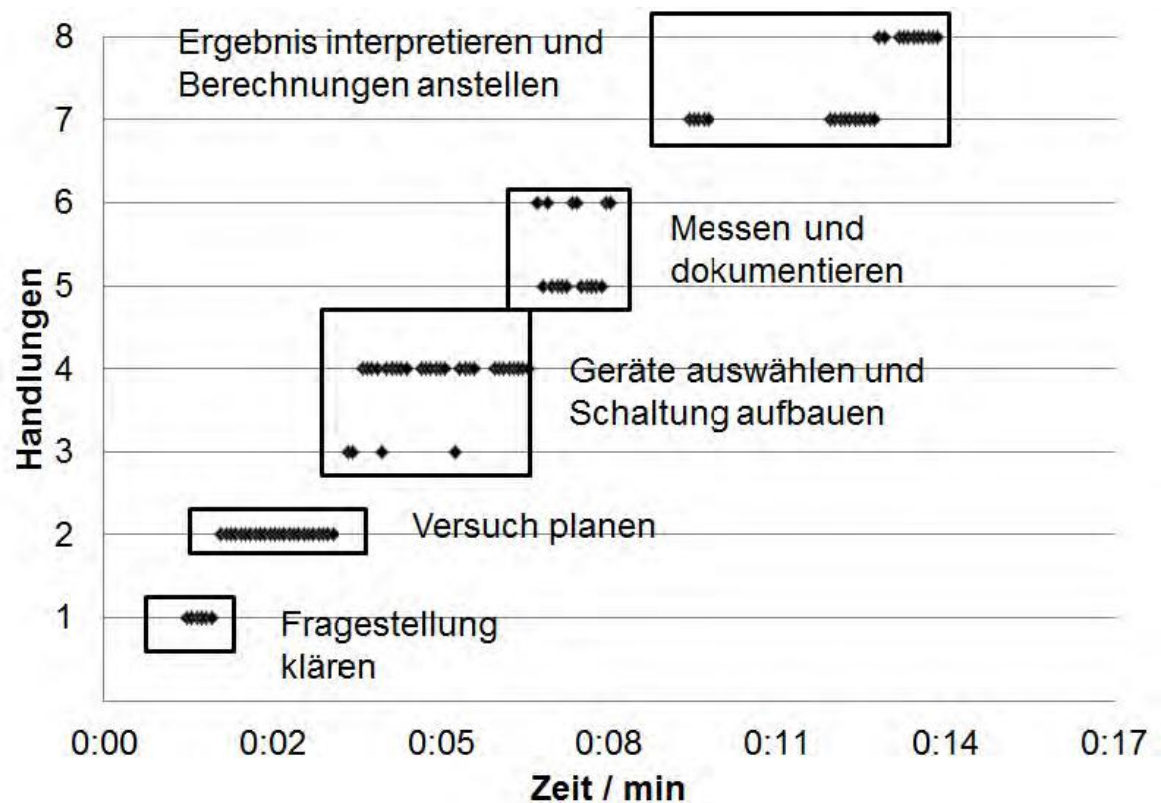
Am authentischsten wird Experimentierkompetenz gemessen, wenn Probanden reale Experimente bearbeiten (Aufgabenformat), dabei der Prozess in die Bewertung einbezogen wird (Bewertungsgegenstand) und die Interaktion bzw. Steuerung von Fähigkeiten und Fertigkeiten Bestandteil der Leistungsbestimmung ist (Modellierung). Bei den analysierten Testinstrumenten erfüllen alle diese Kriterien am ehesten die Lernprozessgrafiken mit Interaktionsboxen im NAW-Test (Walpuski, 2006; Walpuski & Sumfleth, 2009; Emden et al., Juli 2010; Emden, 2011) und der Test von Schreiber (Schreiber, Theyßen, & Schecker, 2009; Schreiber, Theyßen, & Schecker, 2010; Schreiber, Theyßen, & Schecker, 2011a; Schreiber et al., 2011b; Schreiber, 2012). In der aktuellsten Version der Lernprozessgrafiken von Emden (Emden, 2011) werden Ideen, Hypothesen, Experimente und Schlussfolgerungen beim Experimentierprozess identifiziert. Dabei wird bewertet, inwieweit die einzelnen Handlungen richtig sind und inwieweit sie in die Experimentiersituation passen. So wird z.B. unterschieden zwischen Schlussfolgerungen *mit* und *ohne* Rückbezug zur Hypothese (Emden, 2011, Seite 66). Die identifizierten Handlungen werden dann auf einer Zeitachse aufgetragen und bzgl. verschiedener Lösungsideen getrennt (Emden, 2011, Seite 65). Die Experimentierprozesse zu einer

<sup>7</sup> Computersimulationen wurden beim Aufgabenformat nicht aufgenommen. Mit den Simulationen sollte die Ökonomie von Papier-und-Bleistift-Tests mit der Authentizität von Realexperimenten verknüpft werden. Derzeit ist jedoch nicht geklärt, inwieweit bei Simulationen dasselbe Konstrukt wie beim Realexperiment diagnostiziert wird.

Lösungsidee werden anschließend bzgl. der Qualität des Experimentierens bewertet und daraus wird ein Leistungsmaß für die Experimentierkompetenz bestimmt (vgl. Experimentierpfade in Emden, 2011, Seite 68ff).

Einen vergleichbaren Ansatz verfolgt auch Schreiber. In seiner Arbeit werden die Handlungen mittels Videoanalyse identifiziert und zeitlich aufgetragen (Schreiber et al., 2011a, Seite 2; Schreiber, 2012, Seite 71). Ein Beispiel für die Darstellung des Handlungsbildes ist in Abbildung 5 dargestellt. Die Handlungen werden analysiert, um die Qualitätskontrolle bzw. Fehlerkorrektur zu berücksichtigen (Schreiber et al., 2011a, Seite 2; Schreiber, 2012, Seite 72ff). Zuletzt wird die Arbeitsqualität für den gesamten Experimentierprozess bewertet (Schreiber et al., 2011a, Seite 3; Schreiber, 2012, Seite 78f).

Die Bewertungseinheit in diesem Testinstrument sind verschiedene Handlungen über einen längeren Zeitraum. Im konkreten Beispiel (Abb. 5) wurden über einen Zeitraum von 14 Minuten über 100 Handlungen diagnostiziert. Das entspricht etwa einer Fähigkeit bzw. Fertigkeit diagnostizierten Fähigkeit bzw. Fertigkeit alle 8 Sekunden. Durch diese differenzierte Analyse kann der Experimentierprozess glaubhaft durch die Produkte abgebildet werden. Die Einnordung in eine authentische prozessorientierte Auswertung ist damit zulässig. Für den Realexperimentiertest von Schreiber (2012) wurden zwei Aufgaben aus dem Inhaltsbereich Elektrizitätslehre entwickelt. Die Aufgaben umfassen konkrete Handlungen mit dem Experimentiermaterial. Aus dem Beispiel in Abbildung 5 geht hervor, dass der Proband Geräte ausgewählt und eine Schaltung aufgebaut hat sowie Messungen angefertigt und dokumentiert wurden. Das Aufgabenformat entspricht authentischen praktischen Aufgaben. Die Modellierung des Testinstruments setzt sowohl verschiedene Handlungen innerhalb einer Aufgabe, als auch ein Zusammenspiel dieser Handlungen voraus. Dies zeigt sich z.B. im Bezug der Handlungen auf frühere Phasen im Experiment oder in der Bewertung der Qualität im Experimentierprozess. Die Modellierung entspricht der authentischen Modellierung des gesamten Experimentierprozesses.

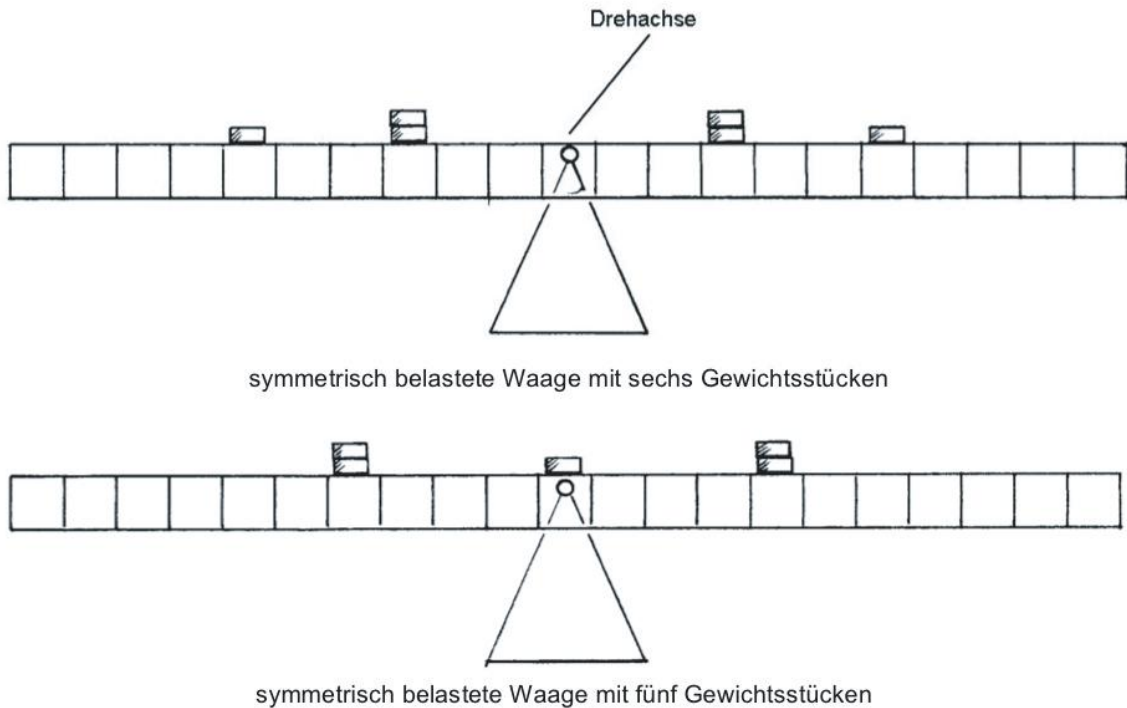


**Abbildung 5: Beispiel des zeitlichen Verlaufs von Handlungen bei Schreiber (Schreiber, 2012, Seite 71).**

### 3.2.2 Eher authentische Testinstrumente

Bei den eher authentischen Testinstrumenten wurde genau eine Vereinfachung zur authentischen Testsituation vorgenommen. Beispielsweise wird nicht der Prozess, sondern nur das Produkt bewertet, z.B. bei Realexperimentierteilen von NAEP (National Assessment Governing board U.S. Department of Education, 2008) und HarmoS Realexperimentiertest (Wissenschaftliches Konsortium HarmoS Naturwissenschaften+, 2008; Gut et al., 2010; Gut-Glanzmann, 2012) oder in den Arbeiten des Instituts für die Didaktik der Naturwissenschaften in Bremen (z.B. Maiseyken et al., 2011).

Als Beispiel eines eher authentischen Testinstruments wird der Test zum Realexperimentiertest des HarmoS Projekts vorgestellt (Gut et al., 2010; Gut-Glanzmann, 2012). Für den Test wurden 15 Experimente entwickelt die insgesamt 145 Teilaufgaben enthalten. Ein Experiment hat z.B. eine Balkenwaage als Gegenstand (Gut-Glanzmann, 2012, Seite 216ff, vgl. Abbildung 6). Zusätzlich zu der Waage werden Gewichte (sechs Schraubenmuttern) zur Verfügung gestellt, die auf der Waage an beliebigen Stellen positioniert werden können. Eine Skizze des Experiments ist in Abbildung 6 gezeigt.



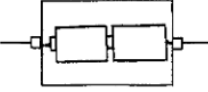
**Abbildung 6: Beispielaufgabe des Experimentiertests aus dem HarMoS Projekt (aus Gut-Glanzmann, 2012, Seite 216).**


Die Auflagefläche der Waage ist in 21 gleich große Felder eingeteilt, die ausgehend von 0 (in der Mitte) jeweils mit 1-10 auf beiden Seiten nummeriert sind. Mit diesen Markierungen können die Probanden die Position der Gewichte auf der Waage quantitativ beschreiben. Zu diesem Experiment wurden sechs Behauptungen als Teilaufgaben formuliert, die von den Probanden untersucht werden sollen. Eine Behauptung lautet z.B.: *„Eine symmetrisch belastete Waage befindet sich immer im Gleichgewicht. Führe zwei Experimente durch, um die Behauptung zu überprüfen.“* (Gut-Glanzmann, 2012, Seite 217). Die Probanden können zur Lösung der Aufgabe frei experimentieren und halten ihre Ergebnisse in einem vorgegebenen Schema fest. Bewertet wird für jede Aufgabe der Grad der Adäquatheit der Experimente und die Beurteilung der Gültigkeit der vorgegebenen Behauptung (Gut-Glanzmann, 2012, Seite 223). Bezogen auf den Bewertungsgegenstand werden die Endergebnisse der Teilaufgaben berücksichtigt. Die produktorientierte Auswertung wurde als ökonomisches Merkmal eines Testinstruments identifiziert. Die Probanden bearbeiten in den Aufgaben reale Experimente. Sie interagieren mit dem Versuchsmaterial und arbeiten praktisch. Das Aufgabenformat ist daher als authentisch einzustufen. Bei der Bearbeitung der Aufgaben haben die Probanden sämtliche Freiheiten. Es können beliebig viele Zyklen durchlaufen und die Handlungen aufeinander bezogen werden. Probanden können Fehler erkennen und diese korrigieren. Die Mo-


dellierung umfasst entsprechend den gesamten Experimentierprozess und ist als authentisch zu beschreiben.

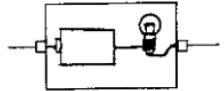
Ein zweiter Repräsentant eines eher authentischen Testinstruments ist der Realexperimentier-test im Rahmen der TIMS Studie (Tamir, Doran, & Chye, 1992; Harmon, 1997). Der Test umfasst 14 Aufgaben. Ein typisches Beispiel für eine Aufgabe, die in dieser Form mehrfach für andere Testinstrumente adaptiert wurde, ist in Abbildung 7 dargestellt.

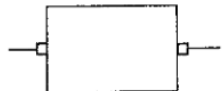
Find out what is in the six mystery boxes A, B, C, D, E and F. They have five different things inside, shown below. Two of the boxes will have the same thing. All of the others will have something different inside.

Two batteries: 

A wire: 

A bulb: 

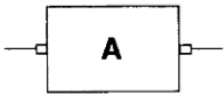
A battery and a bulb: 

Nothing at all: 

For each box, connect it in a circuit to help you figure out what is inside. You can use your bulbs, batteries and wires any way you like.

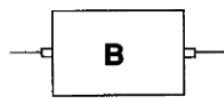
When you find out what is in a box, fill in the spaces on the following pages.

Box A: Has \_\_\_\_\_ inside.  
 Draw a picture of the circuit that told you what was inside BOX A:



How could you tell from your circuit what was inside BOX A ?  
 -----  
 -----

Box B: Has \_\_\_\_\_ inside.  
 Draw a picture of the circuit that told you what was inside BOX B:



How could you tell from your circuit what was inside BOX B ?  
 -----  
 -----

FIGURE 4 Hands-on Electric Mysteries investigation.

**Abbildung 7: Beispiel einer Aufgabe im Realexperiment Format von TIMSS (aus: Shavelson, Baxter, & Pine, 1991, Seite 354).**

Dem Proband werden zur Bearbeitung dieser Aufgabe verschiedene Komponenten aus dem Bereich der Elektrizitätslehre und zwei der in Abbildung 7 dargestellten Boxen zur Verfügung gestellt. Er muss mit diesen Komponenten Stromkreise aufbauen, Beobachtungen anstellen und einen Schluss ziehen, den er auf der rechten Seite der Aufgabe notiert. Das Ziel ist es herauszufinden, welche der auf der linken Seite dargestellten Boxen der Proband erhalten hat. Bewertet werden letztlich ein Ergebnis, sowie eine kurze Erklärung des Ergebnisses. Dies entspricht einer produktorientierten Auswertung. Der Bewertungsgegenstand ist ökonomisch. Zur Bearbeitung der Aufgaben müssen die Probanden praktisch mit dem Material experimentieren. Das Aufgabenformat lässt sich als authentisch einordnen. Beim Experimentieren wird

nicht auf voneinander unabhängige Handlungen fokussiert, sondern der gesamte Prozess ist Bestandteil der Aufgaben.

Eine andere Möglichkeit eines eher authentischen Testinstruments wäre, dass die Testleistung ausschließlich auf voneinander unabhängigen Bewertungen einzelner Fähigkeiten und Fertigkeiten basiert. Allerdings konnte kein entsprechendes Testinstrument zu dieser Klassifikation gefunden werden. Ganz verwunderlich ist dies aber nicht. Denn wenn ein Test den Prozess beim realen Experimentieren berücksichtigt (was relativ aufwendig ist), ist es wenig intuitiv als Grundlage zur Testleistung nicht auf diese Daten, sondern nur auf voneinander unabhängige Maße zu einzelnen Fähigkeiten und Fertigkeiten zurückzugreifen. Theoretisch wäre es aber möglich, die Einzelmaße aus einem Prozess der verschiedenen Handlungen zu der jeweiligen Fähigkeit bzw. Fertigkeit zu modellieren. Beispielsweise könnte das Maß zum *Beobachten und Messen* sich aus verschiedenen Messungen während der Bearbeitung der Aufgaben zusammensetzen und davon abhängen, inwieweit sich der Proband verbessert. Als letzte Methode in dieser Kategorie könnte als Vereinfachung die reale Experimentiersituation durch eine theoretische Aufgabenstellung ersetzt werden. Auch zu dieser Kategorie findet sich kein Testinstrument. Dies lässt sich dadurch begründen, dass die Prozessdaten und eine Qualitätssicherung nur schwer ohne eine reale Experimentiersituation zu realisieren sind.

### **3.2.3 Eher ökonomische Testinstrumente**

Eher ökonomische Testinstrumente sind dadurch charakterisiert, dass in zwei von den drei Aspekten (Bewertungsgegenstand, Aufgabenformat und Modellierung) ein ökonomisches Merkmal gewählt wird. Entsprechend fallen in diese Kategorie Testinstrumente, bei denen das Produkt (ökonomisch) bewertet wird und es keine Interaktion und Steuerung der Fähigkeiten gibt (ökonomisch), die Probanden aber in realen Experimentiersituationen sind (authentisch). Eine Alternative sind Testinstrumente, in denen Probanden nicht real experimentieren (ökonomisch) und die Produkte bewertet werden (ökonomisch), dafür aber eine Interaktion von Fähigkeiten und Fertigkeiten (authentisch) stattfindet. Eine weitere Möglichkeit sind Tests, in denen zwar der Prozess bewertet wird (authentisch), die Probanden aber nicht experimentieren (ökonomisch) und keine Interaktion von Fähigkeiten und Fertigkeiten stattfindet (ökonomisch). Zu allen beschriebenen Kombinationen von eher ökonomischen Testinstrumenten wurden keine Repräsentanten gefunden. Dies könnte dadurch begründet sein, dass es schwer umsetzbar ist, eine Interaktion oder Steuerung von Fähigkeiten und Fertigkeiten als Leistungsgrundlage bzw. eine Bewertung des Prozesses zu nutzen, wenn nicht real experimentiert wird. Wenn jedoch real experimentiert wird, dann werden die Handlungen der Probanden


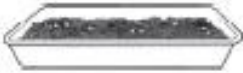
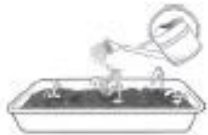

nicht eingeschränkt und der gesamte Experimentierprozess ist relevant. Es wäre möglich ein solches Testinstrument zu entwickeln. In diesem würden zu jeder Handlung der Experimentierkompetenz Aufgaben vorliegen. Diese müssten jedoch voneinander unabhängig sein. Das bedeutet konkret, dass Probanden Experimente aufbauen müssen, an diesen jedoch nicht weiter arbeiten. Aufgaben zur Durchführung von Experimenten müssten auf vorgegebenen Versuchsaufbauten beruhen.

### **3.2.4 Ökonomische Testinstrumente**

Bei ökonomischen Testinstrumenten wurden die Merkmale bei allen drei Aspekten (Bewertungsgegenstand, Aufgabenformat und Modellierung) als ökonomisch gewählt. Tests dieser Kategorie sind die am häufigsten verwendeten (z.B. Comber & Keeves, 1973; Tannenbaum, 1971; Fraser, 1980; IEA, 1988; Beaton, 1996; Martin, 2000; Martin, 2004; Martin, 2008; Martin, 2012; Organisation for Economic Co-operation and Development, 2007; Hammann & Hoi Phan, 2007; Hammann et al., 2008; National Assessment Governing board U.S. Department of Education, 2008; Wissenschaftliches Konsortium HarmoS Naturwissenschaften+, 2008; Gut et al., 2010). Ein Beispiel für ein ökonomisches Testinstrument findet sich in Abbildung 8. Dargestellt ist eine Aufgabe aus dem Testinstrument von Hammann (Hammann & Hoi Phan, 2007). Bei dieser Aufgabe wird im Multiple-Choice Format bewertet, ob der Proband die richtige Lösung angekreuzt hat oder nicht. Die Bewertungseinheit ist entsprechend das Produkt, was einer ökonomischen Diagnose entspricht. In den Aufgaben werden nur theoretische Überlegungen an vorgestellten Experimentiersituationen durchgeführt. Das Aufgabenformat ist ökonomisch. In den Aufgaben wird auf einzelne Fähigkeiten und Fertigkeiten fokussiert. Ein Prozess ist nicht zu erkennen. Die zugrunde liegende Modellierung ist ökonomisch.

Teilweise werden, vor allem in großen Studien verschiedene Testinstrumente genutzt, um unterschiedliche Aspekte zum Experimentieren getrennt voneinander abzubilden (z.B. NAEP und HarmoS). Bei NEAPs werden z.B. Multiple-Choice Tests für das Fachwissen, Concept-Maps für strukturelles Wissen und Experimentiertests für prozedurales Wissen eingesetzt (National Assessment Governing board U.S. Department of Education, 2008; vgl. auch Shavelson & Ruiz-Primo, 1998). Durch diese Trennung können z.B. die Vorzüge ökonomischer Testinstrumente bei der Analyse einzelner Fähigkeiten und Fertigkeiten zum Gütemaß der Richtigkeit genutzt werden. Zur Diagnose der anderen beiden Gütemaße wurden entsprechend Concept-Maps und letztlich Experimentiertests in kleinerem Umfang eingesetzt.



<b>Samenkeimung</b>	
<p>Andreas macht ein Experiment zur Samenkeimung. Er verwendet dafür zwei Töpfe mit Erde. Er sät Bohnensamen in die Töpfe aus und sorgt dafür, dass beide Töpfe im Licht bei einer Temperatur von 22°C stehen. Topf 2 erhält kein Wasser (siehe Abbildungen).</p>	
<p><b>Topf 1</b></p>  <p><b>Erde / Wasser / Licht / 22°C</b></p>	<p><b>Topf 2</b></p>  <p><b>Erde / kein Wasser / Licht / 22°C</b></p>
<p><b>Warum macht Andreas dieses Experiment? Kreuze die richtige Antwort an.</b></p>	
A Weil er Samen dazu bringen will, schneller auszukeimen.	<input type="checkbox"/>
B Weil er vermutet, dass Licht und Erde für die Samenkeimung notwendig sind.	<input type="checkbox"/>
C Weil er vermutet, dass Wasser und Wärme für die Keimung notwendig sind.	<input type="checkbox"/>
D Weil er vermutet, dass Wasser für die Samenkeimung notwendig ist.	<input type="checkbox"/>
<p>Nach einigen Tagen konnte Andreas Folgendes feststellen: Die Samen im Topf 1 waren gekeimt. Im Topf 2 waren die Samen nicht gekeimt (siehe Abbildung).</p>	
<p><b>Topf 1</b></p>  <p><b>Erde / Wasser / Licht / 22°C</b></p>	<p><b>Topf 2</b></p>  <p><b>Erde / kein Wasser / Licht / 22°C</b></p>
<p><b>Wie lautet die beste Erklärung für das Ergebnis?</b></p>	
A Das Experiment klappte nicht, weil die Samen im Topf 2 nicht keimten.	<input type="checkbox"/>
B Das Experiment zeigte, dass Samen Wasser brauchen, um zu keimen.	<input type="checkbox"/>
C Das Experiment zeigte, dass die Samen Licht und Erde brauchen, um zu keimen.	<input type="checkbox"/>
D Das Experiment zeigte, dass Samen Wasser und Wärme brauchen, um zu keimen.	<input type="checkbox"/>

**Abbildung 8: Beispiel einer Bewertung der Experimentierkompetenz durch ein Produkt (aus: Hammann & Hoi Phan, 2007, Seite 38).**

### 3.3 Güte von Testinstrumenten

Im vorherigen Abschnitt dieses Kapitels wurden die unterschiedlichen Aspekte bei der Erfassung von Experimentierkompetenz vorgestellt und im Spannungsfeld der Authentizität und der Ökonomie klassifiziert. Im Feld ist die gängige Einschätzung, dass durch eine authentische Diagnostik die tatsächliche Performance beim Experimentieren bestmöglich abgebildet werden kann (Harmon, 1997, Seite 5). Empirisch ist jedoch nicht geklärt, ob dem wirklich so ist. Im Folgenden soll analysiert werden, ob die einzelnen Testinstrumente "gut geeignet"

sind, um Experimentierkompetenz zu erfassen. Hierzu wird zunächst geklärt, woran sich die Güte von Testinstrumenten bemessen lässt.

In den Naturwissenschaften werden Testmethoden vor allem anhand ihrer Genauigkeit bewertet. In Lehrwerken findet sich zum Messen oder Experimentieren ein entsprechend umfangreicher Anteil zur Messunsicherheit bzw. Fehlerrechnung (z.B. Schenk & Kremer, 2011; Heinicke et al., 2012). Dies zeigt sich auch in der Forschung, bei der Entwicklung neuer Messmethoden. Neue Messmethoden weisen eine höhere Genauigkeit auf und lösen damit gängige ältere- weniger genaue- Methoden ab. Ein Beispiel dafür stellt die Messung der Lichtgeschwindigkeit dar. Bei den ersten Messungen von Galileo Galilei um 1620 konnte nur eine Abschätzung angegeben werden. Mithilfe darauf folgender Messmethoden zwischen 1667 und 1849 konnte eine konkrete Größenordnung bestimmt werden (z.B. James Bradley, 1728; Armand Fizeau, 1849). Bei Weiterentwicklungen der Messungen ab 1851 wurde zusätzlich zu einer höheren Genauigkeit auch ein Messfehler angegeben (z.B. Léon Foucault, 1851; Albert A. Michelson, 1926). Seit 1958 ist die Messgenauigkeit durch Keith Froome kleiner als 0,1 km/s und seit 1973 sogar kleiner als 0,001 km/s (Boulder-Gruppe) bevor sie durch die Definition der CGPM festgelegt wurde (vgl. Tipler, Baumgartner, Gerlich, & Jerke, 1994, Seite 1025ff). Die Angabe einer Messgenauigkeit ist etwas Fundamentales in den Naturwissenschaften. Tipler sagt: *Viele Zahlen, mit denen wir es in der Wissenschaft zu tun haben, sind das Resultat einer Messung damit nur bis zu einer bestimmten Messgenauigkeit bekannt. Das Ausmaß der Ungenauigkeit hängt sowohl von der Geschicklichkeit des Experimentators als auch von dem verwendeten Apparat ab und kann häufig nur abgeschätzt werden* (Tipler et al., 1994, Seite 3f). Die Abhängigkeit von der Messapparatur ist am Beispiel der Lichtgeschwindigkeitsbestimmung nachvollziehbar dargelegt. Das angesprochene *Geschick des Experimentators* lässt sich nach der vorliegenden Modellierung von Experimentierkompetenz auf die Handlung *V Messen, Beobachten und dokumentieren* zurückführen. Selbst beim geschicktesten Experimentator treten trotz einer idealen Messapparatur Messfehler auf. Diese lassen sich in systematische und zufällige Fehler aufgliedern. Systematische Fehler lassen sich oft korrigieren. Zufällige Fehler lassen sich hingegen nicht beseitigen. Bei der wiederholten Messung, verteilen sich die Messwerte um den wahren Wert. Diese Fehler entstehen durch nicht kontrollierbare Einflüsse, Ungenauigkeiten (wie z.B. Rundungsfehler) oder Zufälle. Bei der Radioaktivität ist der Einfluss vom Zufall leicht ersichtlich aber auch die Quantenmechanik stellt ein Beispiel hierfür dar. Zur Bewertung der Güte oder Qualität einer Messung, wird der Fehler in der Messung abgeschätzt. Unter Berücksichtigung des Fehlers soll der gemessene Wert

dem wahren Wert entsprechen. Die Qualität einer Messung steigt, wenn der Messfehler reduziert wird. Die Qualität der Messung wird in den Naturwissenschaften als Messgenauigkeit bezeichnet. Rückbezogen auf das Beispiel der Lichtgeschwindigkeit folgt z.B., dass die Boulder-Gruppe 1973 einen Wert für die Lichtgeschwindigkeit bestimmt hat, der höchstens um 0,001 km/s von dem tatsächlichen Wert der Lichtgeschwindigkeit abweicht.

In den Sozialwissenschaften wird die Güte einer Messung in gleicher Weise bewertet. Eine Messung soll eine möglichst hohe Genauigkeit aufweisen. Genau wie in den Naturwissenschaften treten systematische und zufällige Fehler auf. Die Genauigkeit der Messung, die zu einer Wiederholbarkeit der Ergebnisse führt, wird in den Sozialwissenschaften als Reliabilität bezeichnet. Anders als in den Naturwissenschaften gibt es für die Sozialwissenschaften jedoch keine Messgeräte, die an einen Probanden angeschlossen werden können und einen Wert ausgeben. Die Durchführung einer Diagnose und die Bewertung erfolgt immer durch einen Testleiter. Durch Bewertungsgrundlagen lassen sich Durchführung und Bewertung vereinheitlichen. Dennoch entsteht trotz aller Bemühungen eine Ungenauigkeit durch den Faktor Mensch. Diese durch den Testleiter entstandene Ungenauigkeit in der Messung wird als Objektivität bezeichnet.

Letztlich muss in den Sozialwissenschaften auch sichergestellt werden, was konkret gemessen wird. In den Naturwissenschaften lässt sich diese Frage zumeist problemlos beantworten. In den Sozialwissenschaften ist der Messgegenstand jedoch erneut ein Mensch. Sein Handeln und Denken ist individuell. Beobachtet werden können nur seine Handlungen als Performance. Bedeutsam sind jedoch meistens die zugrundeliegenden Fähigkeiten und Fertigkeiten sowie das Wissen des Probanden. Im Rahmen der Diagnose wird postuliert, dass bestimmte Handlungen des Probanden auf die Ausprägung von Fähigkeiten, Fertigkeiten und Wissen zurückgeführt werden können. Die Zulässigkeit dieser Hypothese und die Beschreibung, inwieweit die spezifischen Fähigkeiten und Fertigkeiten wie gewünscht tatsächlich diagnostiziert werden, werden in den Sozialwissenschaften als Validität bezeichnet.

Die drei dargestellten Gütekriterien Objektivität, Reliabilität und Validität werden im Nachfolgenden beschrieben. Anschließend wird die Güte von Experimentiertests an diesen drei Kriterien diskutiert.

### **3.3.1 Objektivität**

*„Objektivität meint [...] interpersonalen Konsens, d.h., unterschiedliche Forscher müssen bei der Untersuchung desselben Sachverhalts mit denselben Methoden zu vergleichbaren Resul-*

taten kommen können.“ (Bortz, Bortz-Döring, & Döring, ;2010, Seite 326; vgl. auch Wirtz & Caspar, 2002; Wild & Möller, 2009). In der empirischen (quantitativen) fachdidaktischen Forschung werden häufig die Durchführungsobjektivität, die Auswertungsobjektivität und die Interpretationsobjektivität unterschieden (Bortz et al., 2010, Seite 195). Bei der Durchführungsobjektivität ist die Unabhängigkeit der Testergebnisse von der Testsituation gemeint. Einen besonders zentralen Einfluss auf die Testsituation hat der Testleiter als mit dem Probanden interagierende Person. Zur Herstellung von Durchführungsobjektivität muss die Testsituation für alle Probanden identisch sein, das bedeutet insbesondere, dass sich der Testleiter bei allen Probanden identisch verhält. Entsprechend kann eine Durchführungsobjektivität durch die Standardisierung der Testsituation inklusive der Handlungen des Testleiters erreicht werden. Die Auswertungsobjektivität bezeichnet den Einfluss der Auswertung auf die Testleistung. Auch bei der Auswertung kommt den auswertenden Personen, kurz den Auswertern, bei der Bewertung der Testdaten eine zentrale Rolle zu. Um eine hohe Auswertungsobjektivität zu erzielen, muss die Auswertung standardisiert sein. Insbesondere müssen für eine perfekte Auswertungsobjektivität verschiedene Auswerter bei denselben Testantworten zu exakt derselben Punktzahl kommen (vgl. Bortz et al., 2010, Seite 195). Als Maß für die Auswertungsobjektivität wird die Beobachter- (oder Interrater-) Übereinstimmung Cohens Kappa als Kenngröße verwendet (Bortz et al., 2010, Seite 274; vgl. Cohen, 1960). Mit der Interpretationsobjektivität wird die Deutung bzw. Schlussfolgerung der Testleistung bezeichnet. Entsprechend müssen verschiedene Forscher zu denselben Schlussfolgerungen, basierend auf denselben Testdaten, kommen. Dazu dürfen keine individuellen Deutungen einfließen und die Interpretationen müssen sich an Vergleichswerten bzw. Normen orientieren (Bortz et al., 2010, Seite 195).

Allgemein lässt sich sagen, dass bei gründlicher Vorbereitung<sup>8</sup> die Objektivität „meist ein recht unproblematisches Testgütekriterium[ist], das auch bei Eigenkonstruktionen von Fragebögen und Tests leicht realisierbar ist“ (Bortz et al., 2010, Seite 195). Das gilt auch für alle Arten von Experimentiertests. Die größten Probleme ergeben sich bei authentischen Testinstrumenten. Allerdings können auch dort zufriedenstellende Kennwerte für die Objektivität erreicht werden (z.B.  $\kappa = .76 - .91$ , vgl. Schreiber, 2012, Seite 93).

---

<sup>8</sup> Eine gründliche Vorbereitung beinhaltet ein Manual zur Testdurchführung für die Testleiter, zur Standardisierung der Testbedingungen und möglicher Interaktionen von Probanden mit dem Testleiter. Zusätzlich wird ein Manual zur Kodierung benötigt, um eine objektive Interpretation der Daten zu realisieren.

### 3.3.2 Reliabilität

„Die Reliabilität (Zuverlässigkeit) gibt den Grad der Messgenauigkeit (Präzision) eines Instrumentes an. Die Reliabilität ist umso höher, je kleiner der zu einem Messwert  $X$  gehörende Fehleranteil  $E$  ist. Perfekte Reliabilität würde bedeuten, dass der Test in der Lage ist, den wahren Wert  $T$  ohne jeden Messfehler  $E$  zu erfassen ( $X=T$ ). Dieser Idealfall tritt in der Praxis leider nicht auf, da sich Fehlereinflüsse durch situative Störungen, Müdigkeit der Probanden, Missverständnisse oder Raten nie ganz ausschließen lassen.“ (Bortz et al., 2010, Seite 196; vgl. Wild & Möller, 2009). Zur Beurteilung der Reliabilität werden vier verschiedene Aspekte untersucht: Retestreliabilität, Paralleltestreliabilität, Testhalbierungsreliabilität und eine Analyse der internen Konsistenz (Bortz et al., 2010, Seite 196). Zur Beurteilung der Retestreliabilität bearbeiten dieselben Probanden denselben Test zweimal zu unterschiedlichen Zeitpunkten. Um eine hohe Retestreliabilität zu erreichen, sollte die Testleistung bei allen Aufgaben sehr ähnlich sein. Zur Analyse der Paralleltestreliabilität werden zwei Testinstrumente zum selben Konstrukt entwickelt. Probanden bearbeiten beide Tests (Bortz et al., 2010, Seite 197). Auch hier wird die Testleistung verglichen und sollte sehr ähnlich ausfallen. Zur Bewertung der Testhalbierung wird ein Test so halbiert, dass er als zwei eigenständige Testinstrumente zum selben Konstrukt verstanden werden kann. Entsprechend kann wie bei der Paralleltestreliabilität vorgegangen werden (Bortz et al., 2010, Seite 198). Ähnlich wie die Testhalbierungsreliabilität ist die Berechnung der internen Konsistenz angelegt. Dabei wird jede Aufgabe zum selben Konstrukt als eigener Test betrachtet. Entsprechend sollten alle Aufgaben dasselbe messen und die Testleistung zu allen Aufgaben sollte sehr ähnlich sein. Das heißt die Testleistungen zu den verschiedenen Aufgaben sollten jeweils hoch miteinander korrelieren. Als Maß für die Korrelation aller Aufgaben miteinander wird zumeist der Alphakoeffizient nach Cronbach berechnet (Cronbach, 1951; vgl. Bortz et al., 2010, Seite 198).

Bei ökonomischen Testinstrumenten zur Erfassung von Experimentierkompetenz wurden z.B. für das bereits dargestellte Testinstrument von Hammann Werte für Cronbachs Alpha für die Gesamtskala von  $\alpha = .80 - .90$  erreicht (Hammann & Hoi Phan, 2007; vgl. auch Tannenbaum, 1971). Diese Werte deuten auf eine gute interne Konsistenz und damit auf eine zufriedenstellende Reliabilität hin. Allerdings zeigen sich bei Subskalen Werte zwischen  $\alpha = .32 - .80$ . Während bei älteren Testinstrumenten wie dem *Test of Science Processes* von 1971 (Tannenbaum, 1971) relevante Subskalen wie *Experimentieren* mit einer Reliabilität von  $\alpha = .45 - .55$  nicht überzeugen konnten, können in aktuellen ökonomischen Tests mit  $\alpha = .56 - .78$  nahezu ausreichende oder ausreichende Reliabilitäten auch für vergleichbare Subskalen erzielt werden (Hammann & Hoi Phan, 2007). Allerdings ist auffällig, dass in den meisten Studien keine

Angaben zur Reliabilität gemacht werden. Das kann bei den großen Vergleichsstudien daran liegen, dass für die angestrebten Vergleiche zwischen Populationen keine Notwendigkeit besteht, dass die Aufgaben dasselbe messen. Entsprechend wurden zu eher ökonomischen Testinstrumenten keine Angaben zur Reliabilität gefunden.

Bei den eher authentischen Testinstrumenten konnten akzeptable interne Konsistenzen von  $\alpha = .73 - .86$  für die Gesamtscores erreicht werden (Klos, Henke, Kieren, Walpuski, & Sumfleth, 2008). Allerdings zeigen sich teilweise erhebliche Probleme mit der internen Konsistenz der Subskalen von  $\alpha = .21 - .74$  (Klos et al., 2008; Emden, 2011). Während sich bei Klos et al. (Klos et al., 2008) vor allem bei der 7. Jahrgangsstufe Probleme zeigten, waren die Kennwerte der 12. Jahrgangsstufe durchweg ausreichend. Bei Emden zeigten sich in der 6. Jahrgangsstufe Probleme, die auch durch Elimination von Aufgaben nicht beseitigt werden konnten (vgl. Emden, 2011, Seite 87). Bei den authentischen Testinstrumenten wie dem dargestellten von Emden und Schreiber gibt es bisher große Schwierigkeiten bei der internen Konsistenz. Schreiber (Schreiber, 2012, Seite 97ff) gibt keine Kennwerte an. Allerdings enthält das Testinstrument auch nur zwei Aufgaben. Bei Emden (Emden, 2011, Seite 89) liegt die interne Reliabilität der durch Eliminierung reduzierten Skala zwischen  $\alpha = .284$  und  $\alpha = .629$ .

Es lässt sich der Schluss ziehen, dass die Reliabilität ausgehend von ökonomischen Testinstrumenten ( $\alpha_{\text{gesamt}} = .84 - .91$ ;  $\alpha_{\text{sub}} = .32-.80$ ) über eher authentische Testinstrumente ( $\alpha_{\text{gesamt}} = .73 - .86$ ;  $\alpha_{\text{sub}} = .21-.75$ ) bis hin zu authentischen Testinstrumenten ( $\alpha_{\text{sub}} = .29-.63$ ) abnimmt. Die Reliabilität der Gesamtskalen ist für die dargestellten ökonomischen und eher authentischen Testinstrumente im ausreichenden bis guten Bereich. Bei den Subskalen zeigen sich jedoch über alle Testinstrumente Probleme bzgl. der Reliabilität.

Insgesamt wird die Reliabilität mit steigender Authentizität zu einem Problem.

### **3.3.3 Validität**

*„Die Validität (Gültigkeit) ist das wichtigste Testgütekriterium. Die Validität gibt an, ob ein Test das misst, was er messen soll bzw. was er zu messen vorgibt (d.h., ein Intelligenztest sollte tatsächlich Intelligenz messen und nicht z.B. Testangst). Ein Test kann trotz hoher Reliabilität unbrauchbar sein, weil er etwas anderes misst, als man vermutet. [...] Noch fraglicher ist es, ob allgemeine Intelligenz- und Leistungstests, die z.B. als Selektionsinstrumente in konkreten Auswahl-situationen in Schulen, Betrieben, Behörden, beim Arbeitsamt oder in Universitäten eingesetzt werden, tatsächlich die Informationen liefern, die man für derartige Ent-*

*scheidungen benötigt.*“ (Bortz et al., 2010, Seite 200, vgl. auch: Kane, 2001). Es gibt zwei fundamentale Bedrohungen der Validität: Erstens kann es passieren, dass nicht das gesamte Konstrukt gemessen wird, sondern nur ein Teil des Konstrukts relevant für die Testleistung ist (*constructunderrepresentation*); Zweitens kann es passieren, dass zusätzlich zum Konstrukt noch andere, konstrukt-irrelevante Wissensaspekte, Fähigkeiten oder Fertigkeiten einen Einfluss auf die Testleistung haben (*construct-irrelevant variance*, vgl. auch Shavelson R. J. et al., 1993; Shavelson et al., 1999). Es gibt verschiedene mögliche Gründe, die zu *constructunderrepresentation* oder *construct-irrelevant variance* führen können. Messick spricht von Bedrohungen an die Validität und identifizierte sechs Aspekte (Messick, 1989; Messick, 1995). Wird sichergestellt, dass alle sechs Aspekte bei der Generierung von Daten zu keiner *constructunderrepresentation* oder *construct-irrelevant variance* führen, so sind Schlussfolgerungen zulässig. Entsprechend muss zur Sicherung der Zulässigkeit von Schlussfolgerungen auch jeder der Aspekte untersucht und gesichert werden. Die verschiedenen Aspekte der Validität werden im Folgenden vorgestellt.

### **3.3.3.1 Inhaltsvalidität**

Bei dem Aspekt der inhaltlichen Validität (*content validity*, Messick, 1989; Messick, 1995) steht die Passung des Wissens, der Fähigkeiten und Fertigkeiten, die zur Bearbeitung des Tests benötigt werden, im Vordergrund. Beispielsweise sollte ein Fachwissenstest im Bereich Mathematik für die 5. Klasse nur Wissen aus der 5. Klasse voraussetzen. Wenn Aufgaben Bestandteil des Tests sind, für die Wissen aus höheren Jahrgangsstufen benötigt wird, dann würde durch dieses Wissen *construct-irrelevant variance* erzeugt werden. Der Test würde dann nicht das Fachwissen der 5. Klasse messen, sondern auch Fachwissen höherer Klassen. Gleichzeitig sollte das Wissen der 5. Klasse auch vollständig durch den Test abgebildet werden und relevant für die Testleistung sein. Ist dies nicht der Fall, so wird nicht das gesamte Konstrukt erfasst. Gleiches gilt für Fähigkeiten und Fertigkeiten.

Bezogen auf diese Forschungsarbeit sollte ein Testinstrument das Konstrukt Experimentierkompetenz (siehe Kapitel 2) abbilden. Gleichzeitig sollten keine oder so wenig wie möglich andere Fähigkeiten oder Fertigkeiten testrelevant sein. Damit das Konstrukt komplett abgebildet werden kann, müssen Probanden aber auch tatsächlich experimentieren und die Gelegenheit haben Feedback zu erzeugen und ihre Handlungen zu steuern. Testinstrumente, bei denen keine realen Experimentiersituationen vorliegen, sind weniger geeignet prozedurale Fähigkeiten und Fertigkeiten beim Experimentieren zu erfassen (Shavelson & Ruiz-Primo, 1998). Dies zeigt sich auch in Untersuchungen zur Übertragbarkeit zwischen Papier-und-Bleistift-Texts,

Simulationen und Realexperimenten ((Tamir, Doran, & Chye, 1992); (Shavelson & Ruiz-Primo, 1998; Schreiber, 2012; Shavelson R. J. et al., 1993; Shavelson et al., 1999; Schreiber, Theyßen, & Schecker; Gott & Duggan, 2002; vgl. Emden, 2011, Seite 43). Entsprechend sind Testinstrumente, bei denen keine realen Experimentiersituationen vorkommen, weniger geeignet Experimentierkompetenz nach der vorliegenden Modellierung valide zu erfassen, da es im Bereich der inhaltlichen Validität zu einer Konstrukt-Unterrepräsentation kommt.

Damit die Aspekte der Strukturiertheit und Zielorientiertheit der Experimentierkompetenz erfasst werden können, müssen Probanden verschiedene übergeordnete Strategien wählen (Zielorientiertheit) sowie sich Feedback von einzelnen Handlungen generieren können (Strukturiertheit). Die Abfolge von Handlungen muss durch die Probanden selbst bestimmt werden (Strukturiertheit). Sind durch ein Testinstrument keine entsprechenden Gelegenheiten gegeben, so können diese Aspekte der Experimentierkompetenz nicht diagnostiziert werden. Entsprechend würde eine valide Diagnostik nach dem Modell der Experimentierkompetenz nicht möglich sein, da es im Bereich der inhaltlichen Validität wiederum zu einer Konstrukt-Unterrepräsentation kommt.

Zusätzlich muss bei der Testentwicklung sichergestellt werden, dass alle Fähigkeiten und Fertigkeiten der Experimentierkompetenz in ähnlicher Weise testrelevant sind. Um dies zu erreichen, kann ein Aufgabenkonstruktionsmodell verwendet werden, das hilft verschiedene Aspekte der Aufgaben konstant zu halten bzw. mit dem gesteuert werden kann, welche Aspekte variiert werden sollen. Auf diese Weise kann die Aufgabenentwicklung standardisiert werden (Stecher et al., 2000). Zusätzlich muss sichergestellt werden, dass notwendiges Fachwissen und die Inhalte der Aufgaben für die Zielpopulation angemessen sind. Um entsprechende Aufgaben zu entwickeln und die Passung zu überprüfen, können Fachinhaltsanalysen genutzt werden (Mayring, 2008).

### **3.3.3.2 Generalisierbarkeit**

Beim Aspekt der Generalisierbarkeit (*generalizability*, Messick, 1989; Messick, 1995) geht es darum, inwieweit die Ergebnisse verallgemeinert werden können. Oft müssen in Testsituationen Einschränkungen vorgenommen werden, beispielsweise auf die Zielpopulation (vgl. Kapitel 3.1). Diese beeinflussen die Generalisierbarkeit der Ergebnisse in diesem Fall auf eine spezifische Population. Valide Schlussfolgerungen basierend auf den gewonnenen Testdaten, sind nur innerhalb der Einschränkungen zulässig. Sollen Schlussfolgerungen über die Einschränkungen hinaus gezogen werden, so muss die Zulässigkeit begründet werden. Probleme der Generalisierbarkeit werden vor allem bei Testwiederholungen mit veränderten Rahmenbedingungen bzw. Populationen sichtbar. Typische Limitationen der Schlussfolgerungen sind



die Zielpopulation und der Kontext. Ein Testinstrument, das für Zehnjährige valide ist, muss nicht auch für Achtzehnjährige geeignet sein (z.B. Fraser, 1980; vgl. auch Klos, 2008 für die 7. Jahrgangsstufe mit Validitätsproblemen und die 12. Jahrgangsstufe ohne Validitätsprobleme). Beim Kontext gibt es bisher keine Evidenz dazu, ob Experimentierkompetenz in einem Inhaltsbereich, z.B. der Mechanik, auf andere Inhaltsbereiche, z.B. Elektrizitätslehre oder sogar andere Domänen wie Biologie, übertragen werden kann.

Ein Testinstrument ist nur valide, wenn die in der quantitativen Forschung gewünschte Übertragbarkeit der Schlussfolgerungen über die Einschränkungen hinaus nachvollziehbar begründet bzw. belegt werden kann. Entsprechend muss die Diskussion thematisieren, inwieweit eine Generalisierbarkeit, basierend auf den Ergebnissen, vertretbar ist und in welchen Bereichen eine Einschränkung vorgenommen werden muss.

### **3.3.3.3 Strukturelle Validität**

Beim Aspekt der strukturellen oder internen Validität (*structural validity*, Messick, 1989; Messick, 1995) geht man davon aus, dass aus theoretischen Überlegungen bzw. vorliegender Evidenz gewisse Zusammenhänge innerhalb der Daten erwartet werden können. Diese Überlegungen fließen bei der Modellierung des Konstrukts ein. So wurden beim Modell der Experimentierkompetenz die drei Aspekte *Richtigkeit*, *Strukturiertheit* und *Zielorientiertheit* formuliert. Diese Aspekte sollten sich nun im Sinne der strukturellen Validität auch in den Daten finden.

Das setzt zunächst voraus, dass die einzelnen Skalen eine ausreichende interne Konsistenz aufweisen. Wie in Kapitel 3.3.2 beschrieben, besagt eine hohe interne Konsistenz, dass die einzelnen Aufgabendasselbe Konstrukt messen. Ist dies nicht der Fall, obwohl die Aufgabe so entwickelt wurden, dass sie dasselbe messen sollen, so gibt es zwei Möglichkeiten: Entweder misst nur ein Teil der Aufgaben das Konstrukt und der andere Teil der Aufgaben nicht oder alle Aufgaben bilden das Konstrukt nicht adäquat ab. In beiden Fällen liegt entweder Konstrukt-Unterrepräsentation und/ oder Konstrukt-irrelevante Varianz vor. Man muss dann davon ausgehen, dass der Test nicht valide ist. Diese Probleme zeigen sich mit steigender Authentizität von Testinstrumenten (z.B. Emden, 2011, vgl. Kapitel 3.3.2).

Neben den Erwartungen an die interne Konsistenz, mit der die einzelnen Aspekte erfasst werden, folgen aus der Modellierung der Experimentierkompetenz auch Erwartungen an die Zusammenhänge zwischen den drei Aspekten Richtigkeit, Strukturiertheit und Zielorientiertheit. Die einzelnen Aspekte sollen unterschiedliche Facetten beim Experimentieren erfassen. Entsprechend sollten die drei Aspekte nicht zu hoch miteinander korrelieren. Würden sie hoch miteinander korrelieren, so würden die drei Aspekte keine voneinander unabhängigen Kon-

strukture darstellen. Zur Prüfung, ob es sich um ein Konstrukt (Experimentierkompetenz) oder voneinander unabhängige Konstrukte (Richtigkeit, Strukturiertheit und Zielorientiertheit) handelt, lässt sich eine Faktorenanalyse durchführen. Bei einer Faktorenanalyse wird geprüft, inwieweit die Varianz der verschiedenen Aufgaben durch eine bestimmte Anzahl von Faktoren bzw. Konstrukten aufgeklärt werden kann. Um zu zeigen, dass die drei Teilkonstrukte *Richtigkeit*, *Strukturiertheit* und *Zielorientiertheit* das Konstrukt Experimentierkompetenz besser beschreiben als ein einzelnes Konstrukt, müsste bei einer konfirmatorischen Faktorenanalyse die 3-faktorielle Lösung (für die drei Teilkonstrukte) eine bessere Passung zu den Testdaten aufweisen als eine 1-faktorielle Lösung (für das Gesamtkonstrukt Experimentierkompetenz). Bei der 3-faktoriellen Lösung würden die Aufgaben zu einem Aspekt jeweils auf einen Faktor laden. Alternativ könnte mittels einer explorativen Faktorenanalyse nach Hinweisen für weitere Teilkonstrukte gesucht werden.

Neben der Konsistenz innerhalb eines Aspekts und der Trennbarkeit der drei Aspekte voneinander sollten weitere bekannte interne Strukturen oder Zusammenhänge durch die Daten bestätigt werden können. Dies könnten konkrete Zusammenhänge zwischen Konstrukten oder die Vorhersagekraft einiger Faktoren auf andere sein. Im Bereich der Experimentierkompetenz konnten bisher keine konsensfähigen Zusammenhänge gefunden werden. Allerdings gibt es verschiedene gängige Auswertemethoden, die dasselbe Konstrukt messen sollen (z.B. Checklisten oder qualitative Globalmaße). Die Testleistungen, basierend auf diesen Auswertemethoden, sollten durch die Experimentierkompetenz vorhergesagt werden. Abhängig von den spezifischen Methoden kann die Vorhersage vor allem durch einzelne Aspekte gegeben sein.

#### **3.3.3.4 Externe Validität**

Der Aspekt der externen Validität (*external validity*, Messick, 1989; Messick, 1995) beschreibt den Zusammenhang zu externen Maßen. Diese externen Zusammenhänge sollen für eine Bestätigung der externen Validität mit den theoretischen Erwartungen bzw. der generierten Evidenz aus anderen Untersuchungen übereinstimmen. Der Aspekt der externen Validität gliedert sich in konvergente- und divergente Validität (Messick, 1989; Messick, 1995). Bei der konvergenten Validität werden abhängig von bekannter Evidenz mittlere bis hohe Zusammenhänge zu einer anderen Testleistung erwartet. Dies ist z.B. dann der Fall, wenn dasselbe Konstrukt durch unterschiedliche Testinstrumente erfasst wird. In diesem Fall werden die Testinstrumente gegeneinander validiert. Dieses Verfahren ist auch aus der Fachwissenschaft bekannt (z.B. bei der Genauigkeit der Zeitmessung). Bei der divergenten Validität wer-

den höchstens geringe Zusammenhänge zu anderen, ähnlichen Konstrukten erwartet. Bei der Experimentierkompetenz ist dies vor allem das Fachwissen. In Begleiterhebungen bei anderen Untersuchungen wurden zusätzlich aber auch z.B. Interesse, Selbstkonzept und Unterrichtswahrnehmung erhoben (Klos et al., 2008; Emden, 2011). Bei der spezifischen Modellierung der Experimentierkompetenz würden sich auch Untersuchungen zu Konstrukten anbieten, die den Aspekten der Experimentierkompetenz ähnlich sind. Für den Aspekt der Zielorientiertheit wäre dies z.B. das Strategiewissen von Probanden bzgl. des Einschätzens von möglichen Lösungsstrategien bzw. Lösungswegen (Thillmann, 2007).

#### **3.3.3.5 Prädiktive Validität**

Der Aspekt der prädiktiven Validität (*consequential validity*, Messick, 1989; Messick, 1995) bezieht sich auf die zulässige Vorhersagekraft von Testinstrumenten. Testinstrumente werden teilweise als Auswahlkriterium bzw. Selektionstests genutzt (Bortz et al., 2010). In diesem Fall muss sichergestellt sein, dass durch das Testinstrument späterer Erfolg in diesem Bereich adäquat vorhergesagt werden kann. Dies lässt sich in der Regel nur über Langzeitstudien sicherstellen. Problematisch wäre es, wenn beispielsweise ein Studieneingangstest die Leistung der Probanden im Verlauf des Studiums oder dem späteren beruflichen Leben nicht vorhersagt. Lässt sich die Entwicklung des Konstrukts über einen Zeitraum (beispielsweise die universitäre Ausbildung) vorhersagen, so ist die Validität nicht eingeschränkt.

#### **3.3.3.6 Kognitive Validität**

Beim Aspekt der kognitiven Validität (*substantive validity*, Messick, 1989; Messick, 1995) steht die Passung der erwarteten Handlungen von Probanden und den tatsächlich durchgeführten Testhandlungen im Vordergrund. Bei Multiple-Choice Tests ist es z.B. möglich, dass Probanden erkennen, dass die längste Lösung immer die richtige ist. Entsprechend erreichen sie eine hohe Testleistung, allerdings misst der Test in diesem Fall nicht das Wissen der Probanden, sondern wie „clever“ die Probanden sind oder wie gut sie das Konstruktionsschemader Aufgaben erkennen. Der Test wäre in diesem Fall nicht valide, da im Aspekt der kognitiven Validität Konstrukt-irrelevante Fähigkeiten einen nicht zu vernachlässigenden Varianzanteil der Testleistung erklären würden. In diesem Fall wäre die Varianz sogar fast ausschließlich durch die „Cleverness“ bedingt.

Es reicht nicht aus, dass der Proband die erwarteten Handlungen zeigt. Seine Handlungen müssen durch das Auswerteverfahren auch adäquat abgebildet werden können. Wenn das Auswerteverfahren die Handlungen des Probanden nicht adäquat darstellen kann, dann weiß man streng genommen gar nicht, worauf die Testleistung letztlich basiert. Wird beispielsweise

in einem Experimentiertest nur das Endergebnis bewertet, so kann keine Aussage über die konkreten Handlungen des Probanden getroffen werden. Entsprechend können Konstrukt-irrelevante Einflüsse testrelevant werden oder Aspekte des Konstrukts nicht testrelevant sein. In beiden Fällen wäre eine Validität des Testinstruments nicht gegeben.

Um die Passung des Auswerteverfahrens und einen Abgleich der Lösungsprozesse zu realisieren, können Probanden die Intention ihrer Handlungen beschreiben. Dies kann sowohl bei der Durchführung (Eye-Tracker; Think Aloud) oder nachträglich (Interviews) erfolgen (Messick, 1995). Bei der Durchführung des Tests mit einem Eye-Tracker oder der Methode des Think-Alouds kann es zum kognitiven Overload bei Probanden kommen. Nachträgliche Interviews haben dieses Problem nicht, dafür kann es zu einer Verzerrung der tatsächlichen Prozesse kommen.

### **3.3.4 Ökonomie**

Die Ökonomie von Testinstrumenten ist ein wichtiges Merkmal, das entscheidend für die Einsatzfähigkeit und/ oder die Stichprobenzahl ist. Werden bei geplanten Untersuchungen große Probandenzahlen benötigt (beispielsweise um geringe Zusammenhänge oder Faktorladungen zu untersuchen), so darf die Testdurchführung und -auswertung nicht zu aufwendig sein. Auch wenn keine großen Probandenzahlen benötigt werden, wird kritisch hinterfragt, wie viel Aufwand für den Erkenntnisgewinn angemessen ist. Die Antwort ist einfach: Es ist soviel Aufwand angemessen, wie benötigt, um eine objektive, reliable und valide Diagnose zu ermöglichen und darauf aufbauend zulässige Schlussfolgerungen zu ziehen, die in einem Erkenntnisgewinn münden.

Zu Beginn dieses Kapitels wurde das Spannungsfeld zwischen der Authentizität und der Ökonomie beschrieben. Zum Ende des Kapitels erscheint der Sprachgebrauch eines Spannungsfeldes zwischen Testgüte und Ökonomie passender. Es lässt sich zusammenfassen, dass die Ökonomie einer Diagnose soweit reduziert werden sollte, wie es für eine objektive, reliable und valide Messung notwendig ist.

### **3.4 Forschungsdefizit**

Im theoretischen Teil der Arbeit wurde dargestellt, dass Experimentierkompetenz, basierend auf verschiedenen Handlungen, die Fähigkeiten und Fertigkeiten voraussetzen, modelliert werden kann. In der Modellierung wird die Qualität des Experimentierens als Ausprägung der Kompetenz in die drei Teilaspekte *Richtigkeit*, *Strukturiertheit* und *Zielorientiertheit* aufgliedert (vgl. II). Bei der Diagnose von Experimentierkompetenz zeigten sich bisher jedoch

verschiedene Schwierigkeiten (z.B. IEA, 1988, Seite 20, Seite 21, Seite 88; Emden, 2011; Shavelson R. J. et al., 1993; Shavelson et al., 1999; vgl. III). Die Schwierigkeiten lassen sich überwiegend als Einschränkung bzgl. ausgewählter Aspekte der Validität beschreiben. Bezogen auf das Modell der Experimentierkompetenz kann mit den dargestellten ökonomischen Testinstrumenten keine ausreichende inhaltliche Validität erreicht werden (z.B. IEA, 1988, vgl. III.2.3.1). Bei der strukturellen Validität zeigen sich Schwierigkeiten bei authentischen und eher authentischen Testinstrumenten (z.B. Walpuski, 2006; Klos, 2008; Emden, 2011; Schreiber, 2012; vgl. auch Shavelson R. J. et al., 1993; Shavelson et al., 1999). Generalisierbarkeit ist teilweise nicht gegeben (z.B. für die Übertragbarkeit von Populationen, vgl. Fraser, 1980; Klos, 2008). Zudem gibt es zu den vorliegenden Testinstrumenten kaum Untersuchungen zur kognitiven und prädiktiven Validität. Eine externe nachträgliche Einschätzung zu diesen Aspekten ist kaum zu realisieren.

Hinweise zu Problemen mit der Validität entsprechender Testinstrumente zur Erfassung von Experimentierkompetenz gibt es seit über 20 Jahren. In einigen Arbeiten wurden Probleme der Validität aufgezeigt und Lösungsansätze zu ihrer Behebung vorgeschlagen (z.B. Shavelson R. J. et al., 1993; Shavelson et al., 1999; Stecher et al., 2000). In anderen Arbeiten wurde bei der Entwicklung von Testinstrumenten einzelne Aspekte der Validität analysiert (Schreiber, 2012; Emden, 2011). Bisher wurde die Wirkung verschiedener Lösungsansätze zur Verbesserung der Validität jedoch nicht ausführlich untersucht. Auch fehlt eine detaillierte und vollständige Analyse der Validität von Testinstrumenten zur Erfassung von Experimentierkompetenz, um konkrete Problemstellen zu identifizieren.

Folglich gibt es kaum gesicherte Erkenntnisse darüber, welche Eigenschaften von Testinstrumenten bzw. welche Maßnahmen bei der Entwicklung von Testinstrumenten zu einer ausreichenden Güte der Validität führen. Durch die fehlenden Erkenntnisse zum Einfluss der Eigenschaften und Ansätze zur Diagnostik von Experimentierkompetenz auf die Testgüte fällt die Entwicklung eines geeigneten Testinstruments schwer. Aber erst durch Testinstrumente, die Datenerfassungen mit einer ausreichenden Validität ermöglichen, lassen sich aussagekräftige und zulässige Schlussfolgerungen ziehen.

Neben den generellen Schwierigkeiten einer validen Datengenerierung bleibt die Herausforderung einer ökonomischen Diagnostik bestehen. Verschiedene Ansätze zur Verbesserung der Ökonomie wurden in Kapitel 3 vorgestellt. Die angemessene Einordnung im Spannungsfeld zwischen Authentizität bzw. Testgüte und der Ökonomie der Testdurchführung setzt jedoch voraus zu wissen, welche ökonomischen Ansätze im Rahmen einer ausreichenden Testgüte vertretbar sind bzw. wie sich die Testgüte durch diese Ansätze verändert.

## 4 Forschungsziel

Im theoretischen Teil der Arbeit wurde dargestellt, dass der Erwerb von Experimentierkompetenz ein zentrales Ziel der physikalischen Ausbildung ist. Damit untersucht werden kann, ob die praktische Ausbildung bei der Entwicklung von Experimentierkompetenz erfolgreich ist und um die Qualität der praktischen Ausbildung zu verbessern, muss Experimentierkompetenz messbar gemacht werden. Dazu werden Messinstrumente benötigt, die Experimentierkompetenz mit einer ausreichenden Güte erfassen können. Ein Ziel dieser Arbeit ist es, ein solches Testinstrument zu entwickeln. Die Güte von Testinstrumenten in den Sozialwissenschaften wurde in Kapitel 3 anhand der Gütemaße Objektivität, Reliabilität und Validität diskutiert. Bei bisherigen Arbeiten zur Erfassung der Experimentierkompetenz haben sich wiederholt Probleme insbesondere mit der Validität gezeigt (vgl. Kapitel 3). Um diese Probleme zu lösen, wird bei der Testentwicklung umfassend, auf die in Kapitel 3 identifizierten Probleme und Ansätze zur Problemlösung eingegangen. Zur Prüfung der tatsächlichen Testgüte des entwickelten Testinstruments ergibt sich als zweites Ziel eine ausführliche Analyse der Testgüte bzgl. aller Aspekte der drei Gütemaße, vor allem der Validität.

Eine gezielte Analyse aller Aspekte der Validität und Untersuchungen zum Erfolg von Lösungsansätzen zu den Validitätsproblemen würden in einem Erkenntnisgewinn für das Feld resultieren. Darauf aufbauend könnten Testinstrumente entwickelt werden, die valide Schlussfolgerungen zulassen. Diese Instrumente und das in dieser Arbeit entwickelte Testinstrument könnten für Untersuchungen zur Entwicklung von Experimentierkompetenz und darauf aufbauend zur Förderung und Qualitätssicherung (z.B. von Praktika an der Universität) genutzt werden.

## **5 Anforderungen an ein Testinstrument zur Erfassung von Experimentierkompetenz**

Als Ziel dieser Arbeit wurde ein Testinstrument mit ausreichender Testgüte zur Erfassung von Experimentierkompetenz entwickelt (vgl. Kapitel 4). In Kapitel 3 wurden verschiedene Probleme der Testgüte, vor allem bzgl. der Validität identifiziert und in einigen Fällen Lösungsansätze vorgestellt. Aus den Problemen und Lösungsansätzen werden in diesem Kapitel Anforderungen an ein Testinstrument abgeleitet. Diese Anforderungen stellen die Grundlage zur Testentwicklung dar.

In Kapitel 3 wurde herausgearbeitet, dass die Testleistung vom Testformat abhängig ist. Als authentischstes Testformat wurde als zentrale Anforderung an ein Testinstrument zur Erfassung von Experimentierkompetenz die reale Experimentiersituation identifiziert. Daher wird als Format für das Testinstrument das Realexperiment gewählt.

Die Grundlage eines qualitativen Testinstruments ist die Objektivität (vgl. Kapitel 3). Mit Hilfe von Anleitungen bzw. Manualen kann die Testsituation und Testdurchführung durch einen Testleiter standardisiert werden. Dies führt zu einer ausreichenden Güte der Objektivität. Gleiches gilt für Manuale bei der Auswertung. Entsprechend muss bei der Testentwicklung auch die Testdurchführung und die Testauswertung durch konkrete Manuale standardisiert werden. Die Anforderungen und Probleme der Reliabilität wurden auf Aspekte der Validität zurückgeführt. In Kapitel 3 wurde beschrieben, dass eine ausreichende Güte der Validität in allen Aspekten zu einer ausreichenden Reliabilität führt. Im weiteren Verlauf dieses Kapitels werden daher Anforderungen zu den einzelnen Aspekten der Validität erarbeitet.

### **5.1 Inhaltsvalidität**

In Kapitel 2 wurde herausgearbeitet, dass Experimentierkompetenz durch verschiedene Fähigkeiten und Fertigkeiten operationalisiert werden kann. Diese Fähigkeiten und Fertigkeiten lassen sich sieben konkreten Handlungen beim Experimentieren zuordnen. Mit den Handlungen lässt sich der Prozess beim Experimentieren abbilden. Die Qualität des Experimentierprozesses wurde in drei Aspekte gegliedert: die Richtigkeit der Durchführung einzelner Handlungen, die Strukturiertheit in der Abfolge von Handlungen und die Zielorientiertheit ganzer Handlungssequenzen basierend auf einer übergeordneten Strategie. In Kapitel 3 wurde erläutert, dass Aufgaben nur dasselbe messen können, wenn zur Bearbeitung der verschiedenen Aufgaben auch wirklich dieselben Fähigkeiten bzw. Fertigkeiten benötigt werden. In diesem Fall können Probanden sowohl Handlungen zu allen Fähigkeiten bzw. Fertigkeiten als auch

alle Handlungsübergänge zeigen. Dies ist gerade vor dem Hintergrund der Forderung nach einer Interaktion zwischen den Fähigkeiten und Fertigkeiten notwendig (vgl. 3.3.3.1). Es ist aber nicht nur notwendig, dass in den Aufgaben dieselben Fähigkeiten und Fertigkeiten erforderlich sind, sondern auch, dass diese in ähnlichem Umfang erforderlich, also nicht unterschiedlich gewichtet sind. Liegt der Schwerpunkt einer Aufgabe z.B. im Aufbau und der Justierung eines aufwendigen Versuchsaufbaus und die anderen Fähigkeiten bzw. Fertigkeiten werden kaum benötigt, so misst die Aufgabe eher die Fertigkeit einen Versuch fehlerfrei aufzubauen. Diese Fertigkeit wird entsprechend einen überdurchschnittlich großen Einfluss auf die Testleistung zu dieser Aufgabe haben. Werden Aufgaben mit anderen Schwerpunkten entwickelt, so erfassen diese andere Schwerpunkte von Experimentierkompetenz und die Aufgaben messen entsprechend nicht dasselbe.

Bei der Entwicklung eines Testinstruments wird gefordert, dass bei allen Aufgaben auch alle Fähigkeiten und Fertigkeiten in vergleichbarem Umfang zur Bearbeitung notwendig sind. Gerade bei offeneren Aufgaben können sich verschiedene Lösungswege ergeben. Entsprechend müssen sogar bzgl. aller angemessenen Lösungswege alle Fähigkeiten und Fertigkeiten in einem vergleichbaren Umfang relevant sein.

Bezüglich der inhaltlichen Validität wurde in Kapitel 3.2.3.1 die Passung der Fachinhalte und der benötigten Fähigkeiten bzw. Fertigkeiten von Testinstrument zu Zielpopulation sowie die Passung des Konstrukts Experimentierkompetenz durch die Testaufgaben als notwendige Eigenschaft eines Testinstruments identifiziert. Bisher gibt es keine gesicherten Erkenntnisse dazu, ob Experimentierkompetenz unabhängig vom fachlichen Kontext ist. Um einen potentiellen Einfluss des fachlichen Kontextes zu vermeiden, wird der fachliche Kontext auf einen einzelnen Inhaltsbereich der Physik, die Optik, festgelegt. Optik ist ein zentraler Inhaltsbereich der Ausbildung, der gleichzeitig die Möglichkeit bietet, Aufgaben zu entwickeln, die einen hohen Aufwand im Aufbauen und Messen erfordern. Bei Versuchen auf dem Gebiet der Optik müssen Aufbauten justiert werden. Kleine Ungenauigkeiten führen schnell zu großen Abweichungen. Ein weiterer Vorteil der Optik ist, dass alle Veränderungen direkt beobachtet werden können. Die beiden Kriterien würde neben der Optik auch die Mechanik erfüllen. Die Wahl fiel auf die Optik, da es im Bereich der Mechanik schon eine ganze Reihe Testinstrumente gibt, während die Optik bisher eher vernachlässigt wurde. Inhalte aus der Optik sind z.B. Beugung, Brechung und Polarisation. Damit die Fachinhalte repräsentativ für das physikalische Anfängerpraktikum gewählt werden können, wird eine Fachinhaltsanalyse benötigt.



Als Zielpopulation für den Test wurden Studierende zu Beginn des Studiums gewählt, die das physikalische Anfängerpraktikum durchlaufen. Man kann davon ausgehen, dass Studierende wenigstens ein Mindestmaß an Experimentierkompetenz mitbringen. Bei Schülerinnen und Schülern ist nicht geklärt, ob sie in einem ausreichenden Maß über die angestrebten Fähigkeiten und Fertigkeiten verfügen. Im physikalischen Anfängerpraktikum müssen Probanden eigenständig experimentieren. Die Klärung der Zielpopulation stellt bzgl. der inhaltlichen Validität die Grundlage für eine Fachinhaltsanalyse und den Schwierigkeitsgrad der Aufgaben dar.

Der Einfluss von Fachwissen ist nicht nur im inhaltlichen Bereich zu erwarten. Auch Wissen über den Umgang mit optischen Komponenten und Messgeräten kann Varianz erzeugen. Dazu kommen manuelle Fähigkeiten und Erfahrungen (vgl. Kapitel 3). Basierend auf der Fachinhaltsanalyse müssen die Komponenten und Messgeräte vor diesem Hintergrund gewählt oder es müssen Hilfestellungen bereitgestellt werden.

## **5.2 Generalisierbarkeit**

Die in Kapitel 5.1 geforderte Fachinhaltsanalyse ist auch notwendig um herauszufinden, inwieweit die Fachinhalte in deutschsprachigen Praktika übereinstimmen. Sind die Praktika ähnlich, so kann ein Testinstrument die Fachinhalte abdecken und Erkenntnisse sind für alle Praktika zulässig. Unterscheiden sich die behandelten Fachinhalte in den Praktika jedoch zu stark voneinander, so kann kein Testinstrument entwickelt werden, das die Fachinhalte aller Praktika repräsentativ abbildet. Dann sind die Erkenntnisse nicht bzw. nur eingeschränkt verallgemeinerbar (vgl. Kapitel 3.3.3.2). Die Ergebnisse der Fachinhaltsanalyse beeinflussen zwar nicht direkt die Testentwicklung, allerdings kann durch die Erkenntnisse schon im Vorfeld die Limitation zulässiger Schlussfolgerungen herausgefunden werden.

Limitationen gibt es gemäß Kapitel 3.3.3.2 aber auch bei dem Testumfeld. Nur wenn die Testsituation vergleichbar mit einem physikalischen Anfängerpraktikum ist, können Erkenntnisse bedenkenlos auf die Praktika übertragen werden. Bei der Inhaltsvalidität wurde bereits das Realexperiment als Format gewählt. Dieses Format gilt als das authentischste Format und kommt dem Vorgehen im Praktikum am Nächsten. Es bleibt zu beachten, dass auch die Experimentiermaterialien vergleichbar mit den Materialien im Praktikum sind.

Als zweiter Punkt zur Testsituation müssen Probanden genug Zeit haben, um die verschiedenen Fähigkeiten und Fertigkeiten anzuwenden und auch mit weniger hochwertigen Strategien zur Lösung zu kommen. Allerdings muss der Test im Rahmen des physikalischen Anfängerpraktikums bearbeitet werden können. Damit ist die Testdauer beschränkt. Im physikalischen

Anfängerpraktikum haben Studierende meistens einen halben Tag mit drei bis vier Stunden Zeit zur Bearbeitung eines Versuches. Allerdings gibt es auch Teilaufgaben, die in weit kürzerer Zeit gelöst werden können. Bei einer Erprobung erster Aufgaben zeigte sich, dass eine Bearbeitungsdauer von 30 Minuten angemessen ist und Probanden alle Fähigkeiten und Fertigkeiten zeigen konnten. Dies führt zu einem Testumfang von sechs Aufgaben zzgl. Begleiterhebungen bei einem dem physikalischen Anfängerpraktikum vergleichbaren Aufwand. Damit ist eine Verallgemeinerbarkeit des Testumfeldes auf die Praktika gegeben. Der dritte Punkt an Limitationen bezieht sich auf die Population (vgl. III.3.3.2). An dieser Stelle muss eine Limitation der Erkenntnisse auf die Population des physikalischen Anfängerpraktikums im deutschsprachigen Bereich vorgenommen werden.

### **5.3 Strukturelle Validität**

Zur Inhaltsvalidität wurde gefordert, dass bei der Bearbeitung aller Aufgaben alle Fähigkeiten und Fertigkeiten der Experimentierkompetenz gleichgewichtet vorkommen. Die auf den Fähigkeiten und Fertigkeiten basierenden Handlungen lassen sich als Indikatoren detektieren. Gemäß der in Kapitel 2.3 festgelegten, voneinander unabhängigen Qualitätsmaße Richtigkeit, Strukturiertheit und Zielorientiertheit lassen sich für jeden Experimentierprozess- und damit für jede Testaufgabe- drei voneinander unabhängige Indikatoren festlegen. Experimentierkompetenz resultiert aus diesen drei Qualitätskriterien. In den Sozialwissenschaften spricht man bei einem Konstrukt, das durch verschiedene Indikatoren operationalisiert wird von einer *formativen Operationalisierung* bzw. von einem *formativen Messmodell* (z.B. Wirtz & Strohmer, 2013). Eine notwendige Voraussetzung eines formativen Messmodells besteht darin, dass alle Indikatoren des Konstrukts definiert und durch das Testinstrument abgedeckt werden. Als Anforderung für die Testentwicklung müssen nicht nur alle Fähigkeiten und Fertigkeiten gleichgewichtet abgedeckt werden. Bei der Auswertung müssen die drei Qualitätsmaße voneinander unabhängig bewertet werden. Die Ausprägung der Experimentierkompetenz muss sich aus den drei Qualitätsmaßen ergeben. Aufgrund fehlender Erkenntnisse wird eine Gleichgewichtung zwischen den drei Qualitätsmaßen angenommen.

Da die drei Gütemaße beim Experimentieren voneinander unabhängig definiert wurden (vgl. Kapitel 2), wird maximal ein geringer Zusammenhang zwischen den Maßen erwartet. Eine Prüfung der Unabhängigkeit von den drei Aspekten und damit eine Bestätigung des formativen Messmodells, lassen sich mit einer Faktorenanalyse realisieren. Eine 3-faktorielle Lösung, bzgl. einer Zuordnung der Indikatoren zu den drei Qualitätsmaßen müsste die beste Passung des Modells aufweisen.

Bei der Beschreibung der Generalisierbarkeit (vgl. Kapitel 3) wurde herausgestellt, dass die Aufgabenanzahl bedingt durch die Bearbeitungszeit limitiert ist. Eine Modellierung von Testaufgaben verschiedener Schwierigkeitsgrade führt zu der Herausforderung, dass Probanden ggf. nicht ausreichend Aufgaben bearbeiten können, um ihre Kompetenz angemessen zu diagnostizieren. Damit adäquat zwischen schlechten, mittleren und guten Probanden unterschieden werden kann, sollte ein anderer Ansatz gewählt werden. Als Lösung dieses Problems kann jede einzelne Aufgabe in der Lage sein, die Personenfähigkeit für alle Qualitätskriterien differenziert zu diagnostizieren. Dies kann realisiert werden, indem aus dem zeitaufwendigen Experimentierprozess eine größere Anzahl an Indikatoren generiert werden kann. Alternativ müssen die vorhandenen Messungen eine sehr ausdifferenzierte Bewertung liefern können (beispielsweise durch offene Antworten). In beiden Fällen müssen einzelne Aufgaben in der Lage sein, das Konstrukt der Experimentierkompetenz vollständig, inklusive aller drei Qualitätsmaße, abzubilden. Jede Aufgabe in dem Test lässt sich damit als Teil einer streng eindimensionalen Skala verstehen. Die einzelnen Aufgaben bilden auf diesem Weg ein *reflexives Messmodell*. Das bedeutet, dass sich die Leistung einer einzelnen Aufgabe durch die Experimentierkompetenz als prädiktivem Konstrukt vorhersagen lässt. Entsprechend wird eine hohe interne Konsistenz (vgl. Kapitel 3.3) zwischen den einzelnen Aufgaben erwartet. Prüfen lässt sich die interne Konsistenz mittels einer Korrelationsanalyse.

#### **5.4 Externe Validität**

Mit Blick auf die Inhaltsvalidität wurden die Fachinhalte auf die Domäne Optik beschränkt. Zur Bearbeitung von Experimentieraufgaben im Bereich der Optik wird entsprechendes Fachwissen benötigt. Es ist davon auszugehen, dass Probanden unterschiedliche Fachkenntnisse im Bereich der Optik aufweisen. *Die Klärung bzw. Aneignung von Fachwissen* ist Bestandteil der Experimentierkompetenz (vgl. Kapitel 2). Besitzt ein Proband jedoch größere fachinhaltliche Lücken, so kann sich bei der Aneignung von Fachwissen ein Schwerpunkt bei der Bearbeitung der Experimentieraufgaben entwickeln. Hat ein Proband beispielsweise das Thema Polarisation nicht verstanden, so sind 30 Minuten für die Bearbeitung einer Experimentieraufgabe nicht ausreichend, um sich das Fachwissen zur Polarisation anzueignen. Es besteht die Gefahr, dass der Proband innerhalb der limitierten Zeit nicht über diese Handlung hinaus kommt. Um die Varianz durch die Fachinhalte innerhalb der Optik zu minimieren, sollten Hilfestellungen zum Fachwissen in den Experimentiertests integriert werden. Damit können Probanden sich fehlendes Fachwissen gemäß des Modells der Experimentierkompe-

tenz aneignen (vgl. Kapitel 2.3.2). Zur Minimierung des zeitlichen Faktors sollten nur theoretische Hinweise zu den Lösungswegen der jeweiligen Aufgaben aufgenommen werden. Dies soll helfen, die benötigten kognitiven Ressourcen zu reduzieren.

## **5.5 Prädiktive Validität**

In Kapitel 3.2.3.6 wurde erörtert, dass Testinstrumente zur Leistungsmessung als Vorhersage von Erfolg und als Auswahlkriterium genutzt werden. Im Bereich der physikalischen Anfängerpraktika bewerten Assistenten oder Praktikumsleiter zum Teil die Experimentierkompetenz der Probanden. Diese Kompetenz sollte laut Studienordnung und Anforderungen des Praktikums relevant zum Bestehen der Praktika sein. Ohne das Praktikum zu bestehen, kann der Bachelor nicht erfolgreich abgeschlossen werden. Damit die Nutzung als Auswahlkriterium zulässig ist, muss das Testinstrument die Kompetenz des Probanden (auch in anderen Situationen und dem weiteren Verlauf der Karriere) vorhersagen. Das bedeutet wer im Praktikum gut ist, der sollte gut experimentieren können und im weiteren Verlauf des Studiums bzw. im Beruf in der Lage sein experimentelle Anforderungen zu erfüllen.

Zum Zusammenhang zwischen Testleistung und späterem Erfolg beim Experimentieren konnten keine verwertbaren Erkenntnisse gefunden werden. Dies erklärt sich dadurch, dass auch für Langzeitstudien Testinstrumente mit einer entsprechenden Güte notwendig sind, die es aktuell nicht gibt. Im Bereich der Vorhersagekraft zukünftigen Erfolgs besteht ein Forschungsdefizit, das im Rahmen dieser Arbeit nicht geklärt werden kann.

Ein zweiter Bestandteil eines Testinstruments als Auswahlkriterium beinhaltet die Fairness. Ein Testinstrument muss fair sein, das heißt keine Teilpopulation darf Vor- oder Nachteile haben. Ein typisches Beispiel ist der Gender-Effekt oder der Effekt des sozioökonomischen Umfeldes. Im Bereich der Praktika könnten aber auch Körpergröße oder Kraft für die adäquate Benutzung/Bedienung eines Experiments zu Vor- bzw. Nachteilen führen. Damit ein Test prädiktiv valide Daten generiert, sollte keine Subpopulation, die durch Konstrukt-fremde Kriterien gebildet wird, eine vom Standard abweichende Testleistung erbringen. Die Gesamtpopulation besteht aus Studierenden im Hauptfach Physik. Diese Population ist bereits sehr spezifisch. Hinweise auf Vor- bzw. Nachteile von Teilpopulationen sind nicht bekannt. Entsprechend gibt es keine Anforderungen bei der Testentwicklung.

## **5.6 Kognitive Validität**

In Kapitel 3 wurde erörtert, dass die Testleistung auf die Probandenfähigkeiten zum entsprechenden Konstrukt und nicht zu anderen Konstrukten wie der „Testcleverness“ zurückgeführt werden muss. Das bedeutet ein Proband muss die relevanten Handlungen vollführen und das Auswerteverfahren muss sowohl die Handlungen als auch die zugrunde liegende Ausprägung der drei Qualitätsmaße Richtigkeit, Strukturiertheit und Zielorientiertheit adäquat abbilden können. Grundlagen um alle Handlungen beim Experimentieren zeigen zu können, wurden in der Inhaltsvalidität diskutiert. Schwerpunkt in diesem Aspekt ist die Reduktion von alternativen, Konstrukt-fremden Lösungsstrategien. Beim Experimentieren kann ein Proband aus dem gegebenen Material Hinweise und Schlüsse zur Bearbeitung des Tests ziehen. Bei der Testentwicklung müssen materialbezogene Hilfestellungen minimiert werden.

Der zweite Aspekt der kognitiven Validität bezieht sich auf die adäquate Abbildung des Lösungsprozesses durch das Auswerteverfahren. Das Auswerteverfahren muss in der Lage sein die Fähigkeiten und Fertigkeiten bzw. kognitiven Prozesse adäquat abzubilden. In Kapitel 2 wurden diese auf einzelne Handlungen zurückgeführt. Das Auswerteverfahren muss daher die Handlungen der Probanden identifizieren und angemessen bei der Performance berücksichtigen. Im Kapitel 3.2.3.2 wurde beschrieben, dass es bisher keine Erkenntnisse zu Vorteilen bzw. Limitationen von verschiedenen produkt- oder prozessorientierten Auswerteverfahren gibt.

## 6 Studie I – Fachinhaltsanalyse

In Kapitel 5 wurde herausgearbeitet, dass eine Fachinhaltsanalyse zur Testentwicklung notwendig ist. Um sicherzustellen, dass der zu entwickelnde Test den Inhaltsbereich Optik angemessen abbildet (Inhaltsvalidität) und um Aussagen zur Verallgemeinerbarkeit bzgl. der Fachinhalte treffen zu können (Generalisierbarkeit), wird eine Fachinhaltsanalyse durchgeführt.

### 6.1 Forschungsfrage

Um Aufgaben für ein Testinstrument zu entwickeln, die im Bereich Optik in physikalischen Anfängerpraktika relevante Inhalte angemessen abdecken, muss zunächst geklärt werden, welche Fachinhalte in physikalischen Praktika behandelt werden. Zur Klärung der fachinhaltlichen Generalisierbarkeit ergibt sich:

*Forschungsfrage 1: Inwieweit unterscheiden sich die Verteilungen der Fachinhalte deutschsprachiger physikalischer Anfängerpraktika im Inhaltsbereich Optik?*

Hypothese 1a: Die Verteilung der Fachinhalte deutschsprachiger physikalischer Anfängerpraktika im Kontext Optik unterscheiden sich nicht signifikant voneinander.

### 6.2 Studiendesign

Um die Frage im Bereich der Generalisierbarkeit zu beantworten und die Grundlage zur Aufgabenentwicklung zu legen, wurde eine Fachinhaltsanalyse nach Mayring durchgeführt (Mayring, 2008). Im ersten Schritt des Ablaufs nach Mayring (2008) wurden als Untersuchungsgegenstand die Fachinhalte in physikalischen Anfängerpraktika festgelegt. In einer ersten Analyse konnten 59 Universitäten im deutschsprachigen Bereich identifiziert werden, die ein physikalisches Anfängerpraktikum anbieten. Die deutschsprachigen Anfängerpraktika gehen alle auf denselben Ansatz zurück. Es werden nur geringe standortspezifische Unterschiede erwartet. Daher wurden zur Analyse der Inhalte nicht alle Praktika analysiert.  $N=10$  von diesen 59 physikalischen Anfängerpraktika wurden für die Fachinhaltsanalyse als Stichprobe ausgewählt. Damit dennoch ein möglichst repräsentatives Bild gegeben ist, wurden  $n=5$  Praktika aus verschiedenen regionalen Standorten gewählt. Diese Standorte waren Bremen (Norddeutschland), Essen (Westdeutschland), Heidelberg (Süddeutschland), Jena (Ostdeutschland) sowie Kiel als lokale Universität. Zu diesen  $n=5$  Universitäten wurden weitere  $n=5$  Universitäten zufällig ausgewählt. Dabei handelt es sich um Aachen, Gießen, Hannover, Karlsruhe und Magdeburg.

Um die relevanten Fachinhalte der physikalischen Anfängerpraktika dieser Standorte zu untersuchen, wurden als formale Charakteristika des Materials die Versuchsanleitungen genutzt. Diese umfassen üblicherweise sowohl die theoretischen Grundlagen für die jeweiligen Versuche als auch konkrete Arbeitsanweisungen und Fragestellungen. Die Analyse beschränkte sich darauf, welche Fachinhalte tatsächlich relevant für die Handlungen während des Experimentierens sind. Die konkreten Arbeitsanweisungen sowie Fragestellungen zur Bearbeitung von Tätigkeiten im Praktikum wurden von  $N=2$  Ratern analysiert.

Zur Analyse wurde ein Kategoriensystem gebildet. Dazu wurden die Überschriften der Optik aus den Inhaltsverzeichnissen von vier Standardlehrwerken<sup>9</sup> (Tipler et al., 1994; Bergmann et al., 2004; Demtröder, 2009 und Halliday, 2009) als Repräsentanten für relevante Fachinhalte extrahiert. Ähnliche Fachinhalte wurden gruppiert und die Gruppen mit einem Oberbegriff versehen. Diese Oberbegriffe wurden als Kategorien gewählt. Die Zuordnung der Überschriften zu den Kategorien dient als Abgrenzung zwischen den Kategorien. Die Zuteilung der Fachinhalte auf das Kategoriensystem wurde von  $N=6$  Ratern wiederholt. Zur Analyse wurde die Verteilung der Fachinhalte von den physikalischen Anfängerpraktika verglichen.

### 6.3 Kategoriensystem

Durch das im Studiendesign beschriebene Vorgehen wurden zunächst 11 Kategorien zur Optik identifiziert. Die vollständige Liste der zugeordneten Überschriften findet sich im Anhang (A.3). Die 11 identifizierten fachspezifischen Kategorien für die Fachinhalte im Bereich Optik sind:

1. Geradlinige Ausbreitung von Licht
2. Reflexion
3. Brechung
4. Polarisierung
5. Licht als elektromagnetische Welle
6. Absorption
7. Spektren
8. Bildentstehung
9. Optische Instrumente
10. Interferenz
11. Beugung

---

<sup>9</sup> Diese Lehrwerke werden von verschiedenen physikalischen Fachschaften und Professoren für das physikalische Anfängerpraktikum empfohlen. Zudem belegt der Tipler bei der Google-Suche nach Standardlehrwerken den Spitzenplatz. Bei Amazon wird unter Physik, Optik der Demtröder empfohlen.

Die Objektivität bei der Kategorienbildung wurde kontrolliert, indem  $N=6$  Experten aus der Physikdidaktik die Einträge der Inhaltsverzeichnisse erneut den Kategorien zuordneten. Der Anteil an Zuordnungen zur Kategorie „Sonstiges“ war zu vernachlässigen. Dies legt den Schluss nah, dass die gebildeten Kategorien umfassend sind. Bei der Zuordnung durch die Experten wurde eine Übereinstimmung von durchschnittlich 83,32% von einem Rater zur Masterkodierung erzielt. Da bei 14 Kategorien<sup>10</sup> der Einfluss zufälliger Übereinstimmungen vernachlässigt werden kann, wurde keine weitere Analyse zur Verifizierung in Form von Cohens  $\kappa$  durchgeführt. Es kann angenommen werden, dass die Kategorienbildung objektiv und die Kategorien disjunkt sind.

#### **6.4 Datenanalyse**

Zur Analyse der Fachinhalte wurden die Versuchsanleitungen genutzt. Die Versuchsanleitungen enthalten die konkreten Arbeitsanweisungen und Fragestellungen, die zu einer Interaktion der Studierenden mit dem Versuchsmaterial im Rahmen des Praktikums führen. Zudem bieten sie weitere Informationen und theoretische Hinweise, die aber nicht zwingend zu Handlungen beim Experimentieren führen. Diese Inhalte wurden für die Analyse nicht berücksichtigt. In einem ersten Schritt wurden die konkreten Arbeitsanweisungen und Fragestellungen von  $N=2$  Ratern analysiert. Dabei wurden zu 94,2% dieselben Passagen innerhalb der Versuchsanleitungen als konkrete Arbeitsanweisung oder Fragestellung identifiziert. Die konkreten Arbeitsanweisungen bzw. Fragestellungen in den identifizierten Passagen wurden in einem zweiten Schritt von denselben Ratern dem Kategoriensystem zugeordnet. Die Übereinstimmung zwischen den Ratern lag bei 87,8%. Da das Kategoriensystem 14 Kategorien umfasst, wurde der Einfluss von zufälligen Übereinstimmungen vernachlässigt und es wurden keine weiteren Analysen gerechnet.

Um die Verteilung der relevanten Fachinhalte im physikalischen Anfängerpraktikum zu ermitteln, wurden alle Versuchsanleitungen zum Thema Optik zu den  $N=10$  physikalischen Anfängerpraktika analysiert.

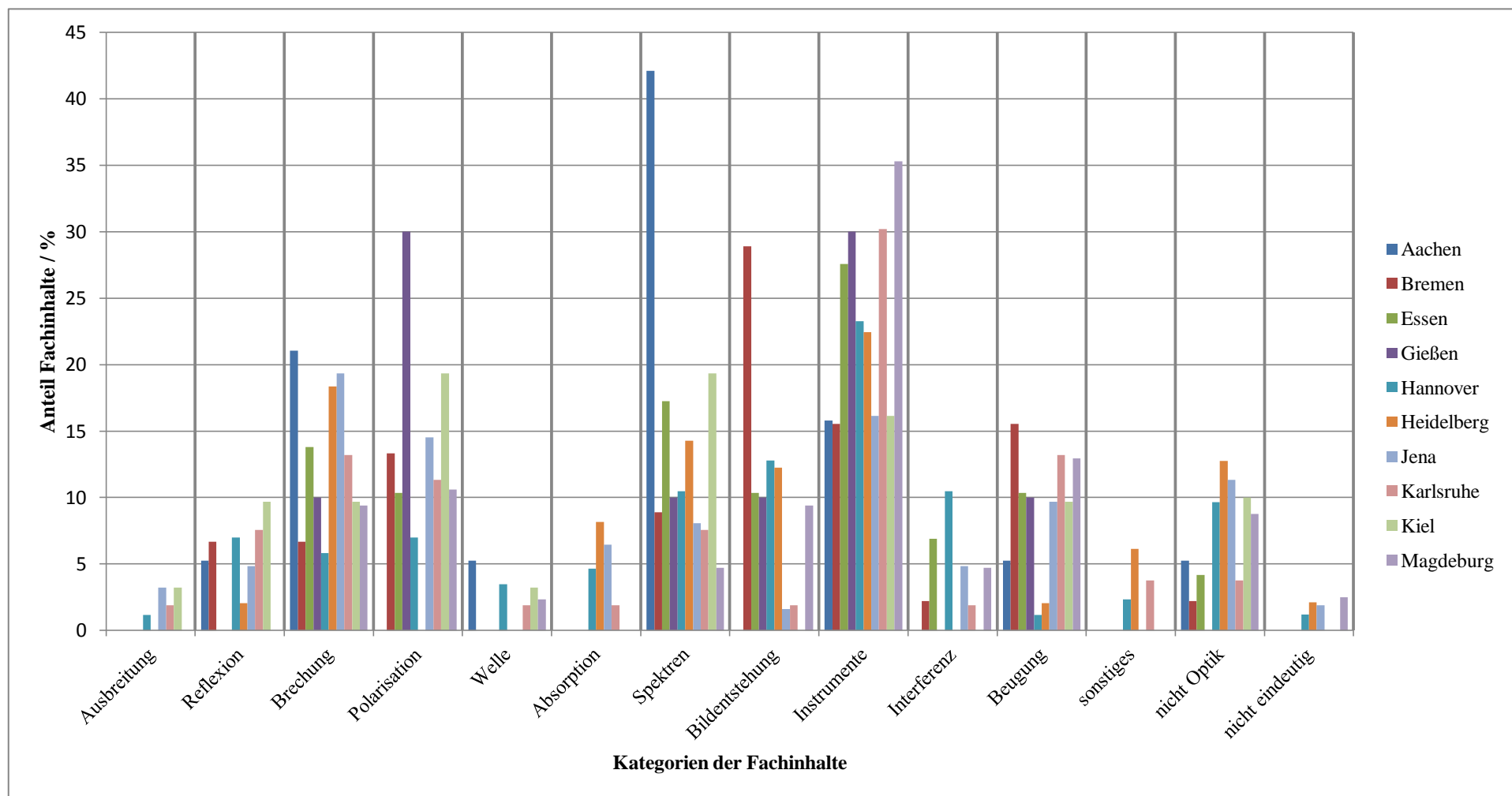
---

<sup>10</sup> Neben den 11 Kategorien zur Optik gab es die Kategorien „Sonstiges“, „nicht eindeutig“ und „Nicht Optik“. Zur Kategorie „nicht eindeutig“ wurden alle Inhalte zugeordnet, die mehr als einer der Kategorien zugeordnet werden konnten. Die Kategorie „Sonstiges“ umfasste alle Inhalte, die durch keine der 11 Fachinhaltskategorien abgedeckt wurden. Die Kategorie „Nicht Optik“ war notwendig, da einige Kapitel sowohl Optik als auch Anteile der Elektrizitätslehre oder Atomphysik enthielten.



## **6.5 Ergebnisse: Fachinhaltsverteilung in deutschsprachigen physikalischen Anfängerpraktika**

Die prozentuale Verteilung der relevanten Fachinhalte der zehn analysierten physikalischen Anfängerpraktika ist in Abbildung 9 dargestellt. Der Großteil der Arbeitsanweisungen und expliziten Fragestellungen konnte einer der 11 Fachinhalts-Kategorien zugeordnet werden. Es wurden bei vier Praktika Inhalte gefunden die "nicht eindeutig" zugeordnet werden konnten. Diese Inhalte entsprechen maximal 2,5% der analysierten Fachinhalte in den Praktika. Zur Kategorie "Sonstige" wurden nur in drei der Praktika Fachinhalte identifiziert. Der Anteil dieser Fachinhalte liegt bei maximal 6%. Einige Fachinhalte lassen sich nicht der Optik zuzuordnen. Dies ist vor allem bei Versuchen der Fall, die Optik und Atomphysik vermischen. Es wurde die gesamte Versuchsanleitung unter dem Schwerpunkt Optik analysiert, obwohl einige Teile eindeutig nicht der Optik zuzuordnen waren. Bei der Verteilung der Fachinhalte zeigt sich ein inhomogenes Bild über die zehn untersuchten Praktika. Nur Inhalte zu den Kategorien Brechung, Optische Instrumente und Beugung finden sich in allen untersuchten Praktika. Spektren finden sich in neun der zehn Praktika. In 80% der Praktika wurden Fachinhalte zu den Kategorien Polarisation und Bildentstehung gefunden. Die Kategorien Geradlinige Ausbreitung von Licht, Reflexion, Licht als elektromagnetische Welle, Absorption und Interferenz sind nur in 40-70% der Praktika von Relevanz. Allerdings ist der prozentuale Anteil jeweils bei maximal 10%. Um diese Gemeinsamkeiten bzw. Unterschiede der Fachinhaltsverteilung zwischen den Praktika genauer zu untersuchen, wurde eine Rangordnung gebildet. Dafür wurden die Fachinhalte für jedes Praktikum nach der prozentualen Ausprägung sortiert (siehe Tabelle 2, zeilenweise). Der Fachinhalt mit der größten prozentualen Ausprägung wurde auf den ersten Rang gesetzt, der Fachinhalt mit der zweitgrößten prozentualen Ausprägung auf den zweiten Rang usw. Wenn zwei Fachinhalte denselben prozentualen Anteil hatten, so wurden sie auf denselben Rang sortiert und der entsprechend nächste Rang wurde nicht vergeben. Zum Vergleich der Fachinhaltsverteilungen zwischen den Praktika wurden die Ränge eines Fachinhalts für alle Praktika verglichen (siehe Tabelle 2, spaltenweise). Die Kategorie Geradlinige Ausbreitung von Licht belegte in allen Praktika einen Rang zwischen 7 und 12. Der Durchschnittliche Rang liegt bei 9,4 und weist eine Standardabweichung von 1,5 auf. Daraus lässt sich schließen, dass die Geradlinige Ausbreitung von Licht über alle Praktika ein weniger relevanter Fachinhalt ist. Die Fachinhalte weisen Standardabweichungen von 0,8 – 3,52 auf. Daraus lässt sich schließen, dass die Relevanz der Fachinhalte über die verschiedenen Praktika ähnlich aber nicht verallgemeinerbar ist. Optische Instrumente gehören in allen Praktika zu den relevantesten Fachinhalten. Brechung, Spektren, Polarisation, Beugung und teilweise Bildentstehung lassen sich als weitere relevante Fachinhalte identifizieren.



**Abbildung 9: Fachinhaltsverteilung deutschsprachiger physikalischer Praktika im Bereich Optik.**

**Tabelle 2: Rangordnung der Fachinhalte in deutschsprachigen physikalischen Anfängerpraktika.**

Praktikumsstandort	Geradlinige Ausbreitung von Licht	Reflexion	Brechung	Polarisation	Licht als elektromagnetische Welle	Absorption	Spektren	Bildentstehung	Optische Instrumente	Interferenz	Beugung	Sonstiges	Nicht Optik	Nicht eindeutig
Aachen	8	4	2	8	4	8	1	8	3	8	4	8	4	8
Bremen	10	6	6	4	10	10	5	1	2	8	2	10	8	10
Essen	9	9	3	4	9	9	2	4	1	7	4	9	8	9
Gießen	7	7	3	1	7	7	3	3	1	7	3	7	7	7
Hannover	12	6	8	6	10	9	3	2	1	3	12	11	5	12
Heidelberg	11	9	2	11	11	6	3	5	1	11	9	7	4	8
Jena	9	8	1	3	13	7	6	12	2	8	5	13	4	11
Karlsruhe	9	5	2	4	9	9	5	9	1	9	2	7	7	14
Kiel	8	5	5	1	8	10	1	10	3	10	5	10	4	10
Magdeburg	11	11	4	3	9	11	7	4	1	7	2	11	6	9
Durchschnitt	9,4	7	3,6	4,5	9	8,6	3,6	5,8	1,6	7,8	4,8	9,3	5,7	9,8
Standardabweichung	1,5	2,1	2,06	2,94	2,28	1,5	1,95	3,52	0,8	2,04	3,12	1,95	1,62	1,99
Maximale Abweichung	2,6	4	4,4	6,5	5	2,6	3,4	6,2	1,4	4,8	7,2	3,7	2,3	4,2

## 6.6 Diskussion

Die Ergebnisse der Fachinhaltsanalyse bestätigen die Erkenntnisse bei der Bildung des Kategoriensystems im Hinblick darauf, dass die Kategorien "Sonstiges" und "Nicht eindeutig" zu vernachlässigen sind. Damit wird die Qualität des Kategoriensystems in Bezug auf eine umfassende und disjunkte Bildung der Kategorien gestärkt. Durch diese Grundlage und eine zufriedenstellende Übereinstimmung der Rater lassen sich die generierten Erkenntnisse als geeignet für Schlussfolgerungen einschätzen. Der Anteil der Fachinhalte, die als "Nicht Optik" kategorisiert wurden, ist für die Erkenntnisse der Studie nicht problematisch. Diese Bestandteile der Versuchsanleitungen lassen sich eindeutig anderen Themenbereichen wie beispielsweise der Atomphysik in themenübergreifenden Versuchen zuordnen. Entsprechend wurden sie für eine Analyse der relevanten Fachinhalte im Bereich der Optik ignoriert.

Die Studie liefert zwei Erkenntnisse. Zum einen sind die relevanten Fachinhalte im Bereich der Optik inhomogen zwischen den verschiedenen physikalischen Anfängerpraktika. Zum anderen lassen sich innerhalb dieser Inhomogenität fachinhaltliche Schwerpunkte finden, die in allen bzw. den meisten Praktika eine zentrale Rolle einnehmen. Durch die Inhomogenität ist eine Verallgemeinerbarkeit von Erkenntnissen zwischen den verschiedenen Praktika nicht ohne weiteres möglich. Hypothese 1a muss wiederlegt werden. Eine teilweise Verallgemeinerung könnte erreicht werden, wenn spezifisch untersucht wird, welche Praktika vergleichbare Fachinhaltsverteilungen aufweisen. Erkenntnisse wären im Bereich Optik über alle Praktika verallgemeinerbar, wenn die Fachinhalte in der Optik keinen Einfluss auf die Experimentierkompetenz der Probanden haben.

Die zweite Erkenntnis bezieht sich auf Schwerpunkte der relevanten Fachinhalte innerhalb der verschiedenen untersuchten physikalischen Anfängerpraktika. Diese sind "optische Instrumente", "Beugung", "Brechung", "Polarisation", "Spektren" und ggf. "Bildentstehung". Diese Erkenntnisse lassen sich als Grundlage zur Aufgabenentwicklung für die Auswahl von Fachinhalten nutzen.

## **7 Entwicklung eines Testinstruments zur Erfassung von Experimentierkompetenz**

In diesem Kapitel wird eine standardisierte Entwicklung des Testinstruments basierend auf den Ansätzen und Erkenntnissen aus Kapitel 3, den Anforderungen aus Kapitel 5 und der Fachinhaltsanalyse aus Kapitel 6 beschrieben. Die Standardisierung von Testumfeld, Testaufgaben und der Auswertung ist nach Stecher ein sinnvoller Weg, um die vielen Einflussfaktoren beim Experimentieren bei der Testentwicklung zu regulieren (Stecher et al., 2000). Damit soll der Einfluss von nicht-personenbezogenen Einflussfaktoren auf die Testleistung minimiert werden. Diese Einflussfaktoren sind Merkmale der Aufgaben (task vgl. Shavelson et al., 1999), Merkmale der Testsituation (occasion vgl. Shavelson et al., 1999) und vor allem Interaktionseffekte zwischen Person, Aufgabe und Testsituation ( $p*t*o$  vgl. Shavelson et al., 1999). Gelingt es durch die Standardisierung die Faktoren *Testsituation*, *Aufgaben* und *Interaktion von Person zu Testsituation und Aufgabe* zu minimieren, so bleibt für die Aufklärung der Varianz nur noch die Personenfähigkeit. Entsprechend sollte sich der bisher geringe Anteil der Varianz der Personenfähigkeit auf die Testleistung der Experimentierkompetenz maximieren (vgl. Shavelson et al., 1999).

Gegliedert wird das Kapitel in den konkreten Umsetzungen zum Testumfeld, zur Aufgabenentwicklung und abschließend zum Auswerteverfahren.

### **7.1 Testumfeld**

Aus den Anforderungen von Kapitel 5 wurden der Inhaltsbereich Optik als Themenbereich für den Experimentiertest gewählt (vgl. Kapitel 5.1). Zudem wurde als Format das Realexperiment als ideal geeignet identifiziert (vgl. Kapitel 3, 5.1, 5.2). Die Bearbeitungsdauer einer Aufgabe wurde auf 30 Minuten festgelegt und ein Gesamtrahmen des Tests auf sechs Aufgaben vorgeschlagen (vgl. Kapitel 5.2).

Im Bereich der Anforderungen bzgl. der Inhaltsvalidität und der externen Validität wurden Hilfestellungen zu den Fachinhalten und ggf. dem Testmaterial gefordert (vgl. Kapitel 5). Zur gezielten Aneignung von Fachwissen wurden theoretische Hinweise zu den in der Fachinhaltsanalyse (vgl. Kapitel 6) identifizierten Fachinhalten aufgenommen. Die theoretischen Hinweise sind reduziert auf die relevanten Fachinhalte zur Bearbeitung aller Lösungswege für die entwickelten Aufgaben. Auch zur Hilfestellung bzgl. der Komponenten wurde die Fachinhaltsanalyse genutzt. Im Bereich des physikalischen Anfängerpraktikums finden sich im Themenbereich Optik vor allem Lux-Meter, Maßband und Winkelmesser. Längen- und Win-

kelmessung ist aus dem Mathematikunterricht der schulischen Ausbildung ausreichend bekannt. Die Bedienung eines Lux-Meters ist einfach<sup>11</sup>. Es wird vermutet, dass die Bedienung innerhalb weniger Minuten verinnerlicht wird. Das bedeutet, dass die in der Optik typischen Messgeräte wenig Wissen bzw. Übung erfordern. Dies gilt ausdrücklich nicht für das Justieren von optischen Komponenten oder Aufbauten, z.B. Strahlengängen. In diesem Bereich ist viel Übung notwendig. Dies lässt sich aber keinem Fachwissen, sondern einem prozeduralen Wissen zum Messen und Beobachten zuordnen. Bezogen auf die Voraussetzungen zu den Messgeräten wird wegen des einfachen Einsatzes bzw. der Erfahrung aus der Schule bei Längen- und Winkelmessung eine minimale Varianz erwartet. Bezogen auf die Komponenten, z.B. Prismen, Gitter, Laser, Farb- und Polarisationsfilter, Spiegel, Blenden und Ähnlichem, müssen die Probanden aufgrund unterschiedlicher Erfahrungswerte aus der schulischen Ausbildung als inhomogene Gruppe angesehen werden (vgl. Kapitel 3.2.3.6). Die einzelnen Aufgaben müssen so entwickelt werden, dass begründet angenommen werden kann, dass der Varianzanteil durch die optischen Komponenten gering sein wird.

In Kapitel 5.4 wurde die Gefahr von Hilfestellungen durch konkrete Materialvorgaben dargestellt. Um die Varianzanteile durch Hilfestellungen von optischen Komponenten und materialbezogene Hilfestellungen zu minimieren, erhalten die Probanden zum Bearbeiten aller Aufgaben denselben Koffer mit Material. Dieser steht ihnen für den gesamten Experimentiertest zur Verfügung. Er enthält Messgeräte und optische Komponenten für alle Lösungswege zu den verschiedenen Aufgaben. Das bedeutet Probanden nutzen identisches Material für die unterschiedlichen Aufgaben. Gleichzeitig gibt es mehr Material als die Probanden beim Bearbeiten aller Aufgaben nutzen werden. Entsprechend gibt es keine Hilfestellungen und die Probanden können bei den Lösungswegen die Komponenten auswählen, zu denen sie Erfahrungswerte besitzen. Das bereitgestellte Material entspricht den typischen, in physikalischen Anfängerpraktika verwendeten Materialien (vgl. Kapitel 5.2).

## **7.2 Aufgaben**

In Kapitel 5 wurde herausgearbeitet, dass die Probanden bei der Bearbeitung jeder Aufgabe alle Handlungen im Experimentierprozess nutzen müssen. Gleichzeitig müssen die Aufgaben bzgl. der kognitiven Validität geeignet sein, die drei Qualitätsmaße Richtigkeit, Strukturiertheit und Zielorientiertheit ausdifferenziert diagnostizieren zu können (vgl. Kapitel 5.6). Im

---

<sup>11</sup> Ein typisches Lux-Meter enthält nur zwei Knöpfe. Den Schalter zum Aktivieren des Messgeräts mit einer Auswahl des Messbereichs und einen "Hold"-Button zum Sperren bzw. Entsperren des Messwertes.

Folgendes wird beschrieben, wie innerhalb einer einzelnen Aufgabe Experimentierkompetenz bzgl. der einzelnen Aspekte ausdifferenziert werden soll.

- (1) Für das Qualitätsmaß Richtigkeit wird jede Handlung als eigener Indikator betrachtet. Da bei der Bearbeitung der Aufgaben eine Vielzahl von Handlungen unabhängig voneinander stattfinden, gibt es viele Teilmessungen und eine adäquate Differenzierung sollte problemlos erreicht werden. Das Maß für die Richtigkeit wird im weiteren Verlauf dieses Kapitels unter 7.3.2.1 beschrieben.
- (2) Ähnlich problemlos stellt sich die Situation für das Qualitätsmaß Strukturiertheit dar, weil die Übergänge zwischen den Handlungen als Indikator genutzt werden. Wenn es  $n$  Handlungen gibt, so gibt es  $n-1$  Übergänge zwischen den Handlungen die bewertet werden können. Da es viele Handlungen gibt, gibt es auch viele Handlungsübergänge. Jeder Handlungsübergang kann unabhängig von den anderen Übergängen bewertet werden. Das daraus resultierende Maß für die Strukturiertheit wird in Kapitel 7.3.2.2 ausführlich vorgestellt.
- (3) Schwierig wird es bei dem Qualitätsmaß der Zielorientiertheit. Probanden nutzen nur eine Strategie für die gesamte Aufgabe. Daher müssen Probanden entsprechend ihrer Personenfähigkeit sowohl einfache Strategien wie Trial-and-Error aber auch anspruchsvolle Strategien wie planungsbasiertes oder analogiebasiertes Experimentieren zeigen können. Bei Aufgaben, in denen der erwartete Lösungsweg linear bzw. eindeutig ist, wird ein Großteil der Probanden genau diesen Lösungsweg zeigen. Um im oberen Bereich differenzieren zu können, müssen die Aufgaben Schwierigkeiten enthalten, an denen die Probanden ihre Qualität zeigen können. Dies lässt sich realisieren, wenn Schwierigkeiten eingebaut werden oder der Lösungsweg nicht linear ist. So zeigt sich Experimentierkompetenz gerade darin Feedback vom Experiment zu erzeugen und dieses zu nutzen (vgl. Kapitel 2). Gleichzeitig müssen alle Probanden die Aufgabe bearbeiten können. Fühlen sich Probanden von der Aufgabe überfordert oder wissen nicht was sie tun sollen, so kann es passieren, dass die Aufgabe nicht beantwortet wird. In diesem Fall kann die Aufgabe im unteren Bereich nicht differenzieren.

Die Schwierigkeit bei der Aufgabenentwicklung liegt also vor allem bei der Zielorientiertheit. Damit die Aufgaben zwischen Probanden im unteren Bereich der Personenfähigkeiten differenzieren können, wird den Probanden eine einfache Fragestellung gegeben. Diese wird durch einen Einleitungstext, der den fachlichen Rahmen eingrenzt, versehen. Die Probanden können

mit der Bearbeitung der Aufgabe beginnen, auch wenn nötiges Fachwissen fehlt. Entsprechend werden auch für diese Probanden Teilmessungen für die Aspekte der Richtigkeit und der Strukturiertheit erzeugt. Gleichzeitig nutzen Probanden eine einfache Strategie, die identifiziert und bewertet werden kann. Damit die Aufgaben gleichzeitig zwischen Probanden im hohen Bereich der Personenfähigkeiten differenzieren können, werden Schwierigkeiten in die Aufgaben eingebaut. Diese Schwierigkeiten zeigen sich für die meisten Probanden erst bei der Bearbeitung der Aufgabe. Wenn Probanden im unteren Bereich der Personenfähigkeit an diesen Problemen scheitern, so liegt wie bereits beschrieben eine Bearbeitung der Aufgabe vor, die bewertet werden kann. Probanden im mittleren Bereich der Personenfähigkeiten können diese Probleme lösen. Damit unterscheiden sie sich von den Probanden aus dem unteren Bereich. Gleichzeitig ist es aber möglich, die Probleme bei der Planung zu berücksichtigen. Dies setzt ein hohes Verständnis des Experimentierprozesses voraus. Besitzen Probanden ein derart hohes Verständnis und nutzen eine entsprechende Strategie, so können die Aufgaben dieses Vorgehen und damit die hohe Experimentierkompetenz abbilden.

**Tabelle 3: Durchschnittliche Rangordnung der physikalischen Anfängerpraktika und des entwickelten Experimentiertests.**

Fachinhalt	Durchschnittlicher Rang des physikalischen Anfängerpraktikums	Rang der entwickelten Aufgaben
Geradlinige Ausbreitung von Licht	11	11
Reflexion	7	5
Brechung	3	2
Polarisation	4	2
Licht als elektromagnetische Welle	10	10
Absorption	9	8
Spektren	2	4
Bildentstehung	6	7
Optische Instrumente	1	1
Interferenz	8	9
Beugung	5	6

Basierend auf den Ergebnissen der Fachinhaltsanalyse physikalischer Anfängerpraktika (vgl. Kapitel 6) wurden sechs Aufgaben für das Testinstrument entwickelt. Um sicherzustellen,



dass die Aufgaben die identifizierten Fachinhalte angemessen abbilden, wurden jeder Aufgabe die darin abgebildeten Fachinhalte nach dem Kategoriensystem von Studie I zugeordnet. Anschließend wurde analog zu Studie I eine Rangordnung für die Fachinhalte des Testinstruments gebildet. Diese Rangordnung und die durchschnittliche Rangordnung der analysierten physikalischen Anfängerpraktika sind in Tabelle 3 dargestellt.

Ein Vergleich zeigt, dass sich die Rangordnungen im Mittel nur um einen Rang unterscheiden. Die maximale Abweichung beträgt zwei Ränge.

Das bedeutet, dass die Fachinhalte des Experimentiertests Schwerpunkte der physikalischen Anfängerpraktika repräsentativ abdecken können. Damit kann die Forderung der Inhaltsvalidität nach der repräsentativen Passung der Fachinhalte und eingeschränkt auch die Generalisierbarkeit von Ergebnissen erfüllt werden.

### 7.2.1 Experimentiertest

Basierend auf den Ansätzen zur Validität und der Fachinhaltsanalyse wurden sechs Aufgaben aus der Inhaltsdomäne Optik im Format des Realexperiments für den Experimentiertest entwickelt. Die Bearbeitungszeit für eine Aufgabe liegt bei 30 Minuten. Die Inhalte der Aufgaben sind in Tabelle 4 dargestellt, eine vollständige Darstellung der Aufgaben und eine Musterlösung finden sich im Anhang (A.4.3).

**Tabelle 4: Entwickelte Aufgaben des Experimentiertests.**

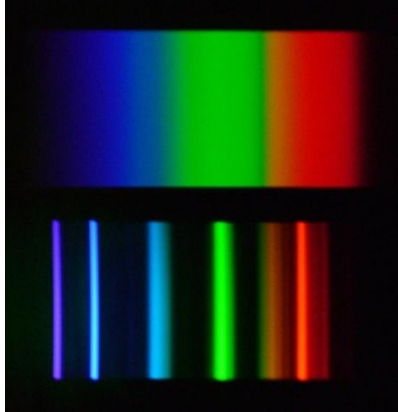
<b>Aufgabe</b>	<b>Aufgabenstellung und Inhalt</b>
<b>1</b>	<b>Vermessung einer Glühwendel</b> Mithilfe einer Abbildung durch eine Linse oder ein Pin-Hole kann die Glühwendel vermessen werden.
<b>2</b>	<b>Optische Dichte</b> Durch die Brechung verschiedener Laserstrahlen kann die wellenlängenabhängige Dispersion für ein gegebenes, unbekanntes Prisma bestimmt werden.
<b>3</b>	<b>Spektroskopie</b> Die Wellenlänge eines unbekanntes Lasers soll mithilfe eines Gitters oder eines bekannten Prismas bestimmt werden.
<b>4</b>	<b>Optische Aktivität von Zuckerwasser</b> Die Probanden sollen eine Konzentration von Zucker zu einer vorgegebenen Drehung der Polarisationssebene von linear polarisiertem Licht eines optisch aktiven Mediums untersuchen.
<b>5</b>	<b>Polarisationseigenschaften eines Prismas</b> Mithilfe von Reflexion und Doppelbrechung soll ein einfallender Lichtstrahl linear polarisiert werden. Die Probanden sollen die Eigenschaften der Polarisation untersuchen.
<b>6</b>	<b>Identifikation optischer Eigenschaften von unbekanntes Bauteilen</b> Die Probanden sollen verbaute Teile in einer Blackbox identifizieren. Verbaut sind in drei möglichen Positionen ein Gitter und ein Farbfilter.

## 7.2.2 Beispielaufgabe

Im Folgenden wird eine der Aufgaben beispielhaft vorgestellt. Die Aufgabenstellung findet sich in Abbildung 10.

**Spektroskopie**

Jede elektromagnetische Welle hat eine Frequenz die charakteristisch für die Welle ist. Liegt die Frequenz im vom menschlichen Auge wahrnehmbaren Bereich, so kann dem Licht eine Farbe zugeordnet werden. Überlagern sich elektromagnetische Wellen verschiedener Frequenzen so lässt sich Licht in einer Farbe beobachten. Mit Hilfe von Gittern oder Prismen lassen sich überlagerte elektromagnetische Wellen bzgl. ihrer Frequenz aufspalten. Dabei entstehen Spektren, die sich nutzen lassen um Stoffe oder einzelne Wellenlängen zu identifizieren.



**Bestimmen Sie experimentell die Wellenlänge des unbekanntes Lasers (Inventarnr. 5).**

**Abbildung 10 - Beispielaufgabe des Experimentiertests.**

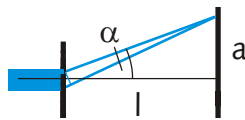
Die Aufgabe fokussiert auf die Wellenlängenbestimmung eines unbekanntes Laserpointers. Dies lässt sich durch Brechung am Prisma oder Beugung am Spalt bzw. Gitter realisieren. Sowohl Prisma als auch Gitter stehen zur Bearbeitung zur Verfügung. Die jeweiligen Versuchsaufbauten sind in Abbildung 11 dargestellt. Bei dem Lösungsweg mit dem Gitter muss zunächst die unbekanntes Gitterkonstante mit einem der beiden anderen Laserpointer mit gegebener Wellenlänge bestimmt werden.



**Abbildung 11 – Versuchsaufbau zu einer Beispielaufgabe a) mit einem Prisma (links) oder b) einem optischen Gitter (rechts) des Experimentiertests.**

Zur Bearbeitung werden alle Fähigkeiten bzw. Fertigkeiten der Experimentierkompetenz benötigt.

- (1) Es ist notwendig, dass die Fragestellung nachvollzogen wird.
- (2) Die theoretischen Grundlagen sowohl zu dem Versuchsaufbau mit Prisma als auch mit Gitter müssen geschaffen werden. Beim Aufbau mit dem Prisma gilt das Snelliusche Brechungsgesetz  $n_1 \cdot \sin(\alpha_1) = n_2 \cdot \sin(\alpha_2)$ ; Variable:  $n_1=1$ . Beim Versuch mit dem Gitter müssen zwei Dreiecke aufgestellt werden, die ähnlich zueinander sind (siehe Abb. 12). Zudem werden die folgenden Formeln und Variablen benötigt:  $n \cdot \lambda = g \cdot \sin(\alpha_n)$ ;  $\sin(\alpha_n) = a_i / l$ ;  $\tan(\alpha_n) = a_i / l'$ ;  $n_1=1$ ;  $\lambda_{blau}=430nm$ ;  $\lambda_{grün}=532nm$ .



**Abbildung 12: Theoretische Grundlage zu ähnlichen Dreiecken bei der Spektroskopie.**

- (3) Mithilfe dieser Grundlage kann ein Experiment (vgl. Abb. 11) geplant werden.
- (4) Dieses muss entsprechend und fehlerfrei aufgebaut werden.
- (5) Als Messwerte müssen Einfallswinkel und Ausfallwinkel bzw. die Abstände Gitter-Schirm und Optische Achse-Maximum bestimmt werden.
- (6) Mithilfe der Formeln aus den theoretischen Grundlagen und der Messwerte lassen sich Berechnungen anstellen bzw. Wellenlängen aus einer Referenzgrafik ablesen.
- (7) Letztlich müssen die Berechnungen bzw. abgelesenen Wellenlängen auf die Fragestellung bezogen werden. Dazu gehört, dass die Güte des Ergebnisses eingeschätzt wird.

Als Fachinhalte werden vor allem Beugung und Brechung, aber auch die Welleneigenschaften von Licht und der Umgang mit optischen Instrumenten berücksichtigt. Durch den Einleitungstext wird eine Fokussierung auf diese Inhalte erreicht. Mit Hilfe der theoretischen Hinweise (siehe Anhang A.4.1) sollen sich die relevanten Inhalte schnell identifizieren lassen. Es ist davon auszugehen, dass der Einstieg in die Aufgabe leicht ist und auch Probanden im unteren Fähigkeitsbereich mit der Bearbeitung beginnen können. Allerdings müssen die Probanden selbstständig entscheiden, welcher Aufbau (Gitter oder Prisma) besser durchzuführen ist. Beide Verfahren haben Vorteile, aber sie bergen auch Schwierigkeiten. Bei dem Aufbau mit dem Prisma muss sehr genau justiert und gemessen werden. Kleine Ungenauigkeiten bei Aufbau

und Messung führen zu Ergebnissen, die weit außerhalb der Fehlertoleranz liegen. Zudem muss die Wellenlänge aus einem Graphen identifiziert werden, der den Brechungsindex für verschiedene Wellenlängen angibt. Beim Gitter ergeben sich aus weniger ideal justierten Aufbauten nicht automatisch große Abweichungen vom Toleranzwert. Zudem müssen “nur“ einfache Rechnungen durchgeführt werden. Dafür ist die Gitterkonstante nicht bekannt und muss zuvor ermittelt werden. Alternativ ist es möglich mit Verhältnissen für verschiedene Wellenlängen zu rechnen. Entsprechend können für alle Lösungswegen Probleme bei der Bearbeitung auftreten. Probanden mit hoher Experimentierkompetenz können sich dadurch auszeichnen, dass diese Probleme im Vorfeld bedacht und berücksichtigt werden. Probanden mit niedriger oder mittlerer Experimentierkompetenz können darüber differenziert werden, wie sie mit den Problemen umgehen. Entsprechend werden diese Probanden zusätzliche Experimentierzyklen benötigen, um die Aufgabe erfolgreich bearbeiten zu können (vgl. Kapitel 2.4.3).

Das Material zur Bearbeitung dieser Aufgabe sowie das Material für die anderen fünf Aufgaben wird, wie im vorherigen Kapitel beschrieben, in einem Koffer zur Verfügung gestellt (siehe Abb. 13). Neben den optischen Komponenten sind Inventarliste und theoretische Hinweise enthalten (beides zu finden im Anhang A.4.1 und A.4.2). Zu den Komponenten zählen rutschfeste Metallplatten (siehe Abb. 14). Diese dienen als stabile Experimentierunterlage. Die einzelnen optischen Komponenten sind mit Magneten versehen und können damit fixiert werden.



**Abbildung 13: Rutschfeste, magnetische Arbeitsfläche und optische Komponenten vom Experimentiertest.**



**Abbildung 14: Experimentierkoffer mit Material.**

### **7.3 Auswerteverfahren**

Im theoretischen Teil wurde herausgearbeitet, dass Experimentierkompetenz bzgl. der drei Aspekte Richtigkeit, Strukturiertheit und Zielorientiertheit beurteilt werden kann. Um Experimentierkompetenz zu messen, müssen die entwickelten Aufgaben bezüglich der drei Aspekte Richtigkeit, Strukturiertheit und Zielorientiertheit bewertet werden. In Kapitel 3 wurde aufgezeigt, dass dazu produkt- bzw. prozessorientierte Auswerteverfahren genutzt werden. Produktorientierte Auswerteverfahren sind ökonomischer, weshalb sie für den Einsatz von Testinstrumenten bevorzugt genutzt werden. Doch auch wenn die Produkte auf Prozesse schließen lassen, so liefert ein Auswerteverfahren, mit welchem die Prozesse zusätzlich zu den Produkten direkt bewertet werden können, zuverlässigere Daten. Diese Auswerteverfahren werden prozessorientiert genannt und sind deutlich zeitaufwendiger. Bei der Analyse zur kognitiven Validität (3.3.3.6; 5.6) wurde hinterfragt, ob ein ökonomisches produktorientiertes Auswerteverfahren im Vergleich zu einem zeitaufwendigen prozessorientierten Auswerteverfahren ausreichend valide Daten generieren kann. Deshalb wurde, wie in Kapitel 5.6 ausgeführt, festgelegt, dass sowohl ein produkt- als auch ein prozessorientiertes Auswerteverfahren ver-

wendet werden soll. Zur Validierung des Testinstruments wird zudem ein Referenz-Auswerteverfahren genutzt. Dieses beruht auf der Methode des "lauten Denkens" (vgl. 3.3.3.6). Beim "lauten Denken" werden neben Produkt und Prozess auch die Gedanken der Probanden direkt zugänglich gemacht. Damit liefert dieses Verfahren die zuverlässigsten Daten, ist gleichzeitig aber auch am aufwendigsten durchzuführen.

Im Folgenden wird beschrieben wie durch das produkt- bzw. die prozessorientierten Auswerteverfahren die Produkte bzw. Handlungen der Probanden identifiziert werden können. Anschließend werden die konkreten Maße für die Aspekte Richtigkeit, Strukturiertheit und Zielorientiertheit vorgestellt.

### **7.3.1 Bewertungseinheiten**

Im theoretischen Teil wurde aufgezeigt, dass Experimentierkompetenz ein Konstrukt ist, welches verschiedene Fähigkeiten und Fertigkeiten im Experimentierprozess umfasst. Natürlich könnten Aufgaben zur Erfassung einzelner Fähigkeiten bzw. Fertigkeiten entwickelt werden. In diesem Fall lässt sich die Bewertung direkt bezogen auf die jeweiligen Fähigkeiten bzw. Fertigkeiten durchführen. Bei den Experimentieraufgaben sollen jedoch verschiedene Fähigkeiten und Fertigkeiten genutzt werden. Entsprechend ist es notwendig ein Verfahren zu entwickeln, mit dem sich die einzelnen Produkte bzw. Handlungen zu den einzelnen Fähigkeiten bzw. Fertigkeiten voneinander getrennt identifizieren lassen. Für die Strukturiertheit und die Zielorientiertheit ist es zudem notwendig eine Abfolge von Fähigkeiten bzw. Fertigkeiten abzubilden. Ist nicht bekannt, in welcher Reihenfolge ein Proband Handlungen durchgeführt hat, kann auch nicht bewertet werden, ob diese Abfolge logisch richtig ist oder inwieweit die Handlungen einer Strategie folgen. Im Folgenden wird jeweils ein produkt- bzw. prozessorientiertes Auswerteverfahren sowie ein Referenzauswerteverfahren basierend auf dem "lauten Denken" vorgestellt und es wird beschrieben, wie die Bewertungseinheiten ermittelt und in eine Reihenfolge gebracht werden. Nachdem die Reihenfolge vorliegt, wird in allen Auswerteverfahren simultan verfahren. Das konkrete Vorgehen wird zum Abschluss dieses Kapitels in 7.3.2 vorgestellt.

#### **7.3.1.1 Produktorientiertes Auswerteverfahren**

Bei einer produktorientierten Auswertemethode werden nur die Produkte von Probanden zur Bewertung genutzt (vgl. Kapitel 3). Beim Experimentieren sind Produkte entweder die Gesamtlösung der Aufgabe oder diverse Aufzeichnungen in sogenannten Laborheften. Diese werden bei experimentellen Arbeiten in Laboren für Aufzeichnungen beim Experimentieren,

beispielsweise von Messwerten oder Formeln und zur Nachvollziehbarkeit der Ergebnisse genutzt. Da eine Lösung der Aufgabe mit dem Ergebnis bzw. der Schlussfolgerung aus einem Laborheft übereinstimmt, bietet das Laborheft mehr Informationen als eine einfache Lösung. Gerade bezogen auf das Modell der Experimentierkompetenz, das auf sieben Fähigkeiten bzw. Fertigkeiten basiert und sich in drei Aspekte gliedert, liefert nur eine finale Lösung der Aufgabe zu wenig Informationen für eine adäquate Messung von Experimentierkompetenz. Entsprechend bietet sich das Laborheft als Grundlage zur produktorientierten Auswertung an. Den Probanden steht ein Laborheft mit ausreichend Platz zur freien Bearbeitung der Aufgabe zur Verfügung, in dem auch die Lösung notiert werden sollte. Ein Beispiel eines ausgefüllten Laborhefts findet sich in der Abbildung 15.



Bestimmen Sie experimentell die Wellenlänge des unbekanntes Lasers (Inventarnr. 5).

Platz für Ihre Lösung

$$n\lambda = g \sin \alpha$$

↓

$$\sin \alpha = \frac{a_n}{\ell} = \frac{0,001275126}{0,073059361}$$

$$\rightarrow n = 1$$

$$\rightarrow a_{n=1} = 3,2 \text{ cm}$$

$$\rightarrow \ell = 43,8 \text{ cm}$$

$$\rightarrow \lambda = 532 \text{ nm} = 0,000000532 \text{ m}$$

$$\rightarrow g = \frac{n \cdot \lambda}{\sin \alpha} = \frac{0,000000532}{0,000007282} \text{ m}$$

$$\lambda = \frac{g \sin \alpha}{n} = 665 \text{ nm}$$

$$\rightarrow n = 1$$

$$\rightarrow a_{n=1} = 4 \text{ cm}$$

$$\rightarrow \ell = 43,8 \text{ cm}$$

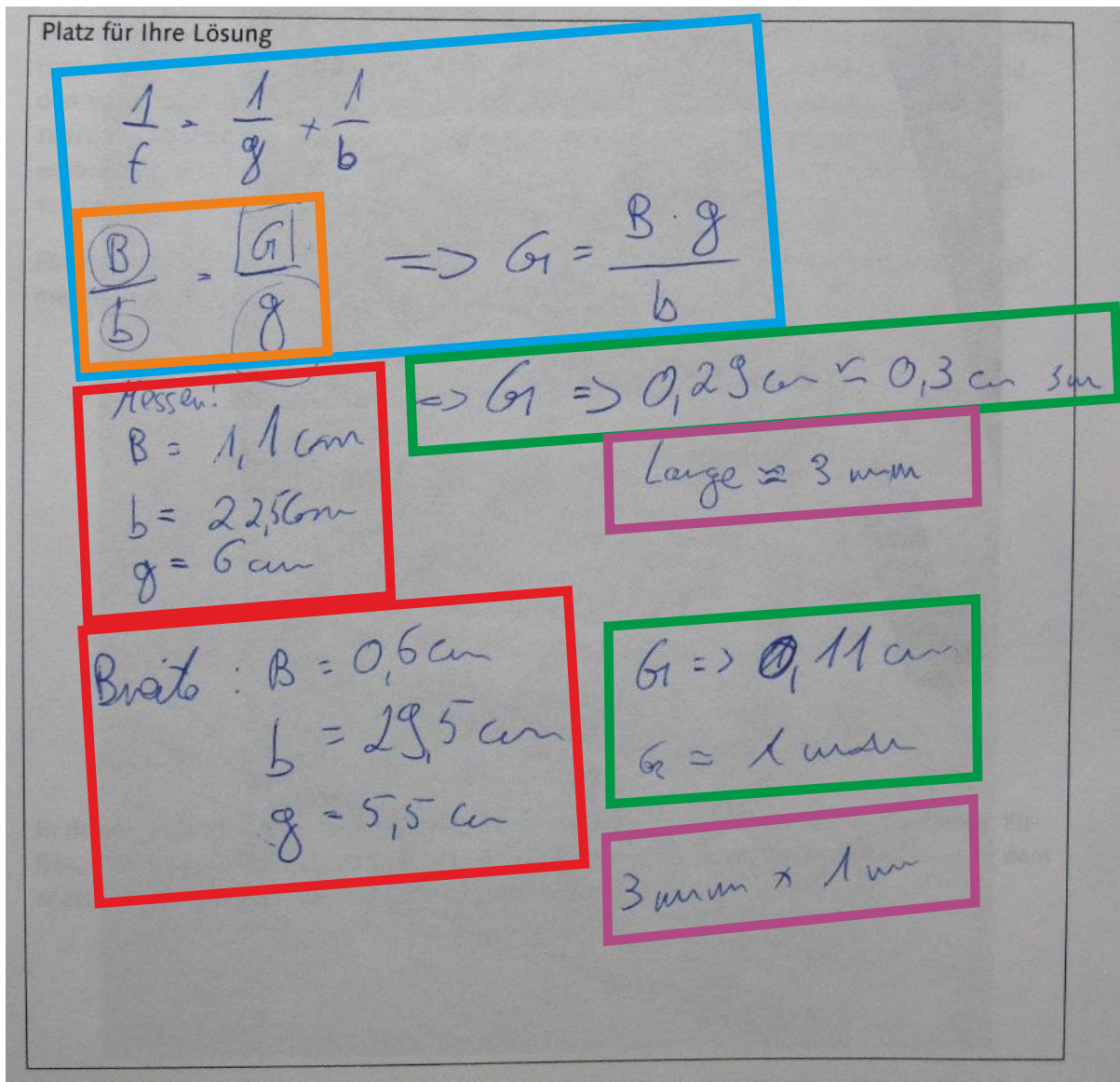
$$\rightarrow \sin \alpha = 0,091324201$$

Lösung: Die Wellenlänge des unbekanntes Lasers (Inventarnr. 5) beträgt: 665 nm

Abbildung 15 - Beispiel eines ausgefüllten Laborhefts.

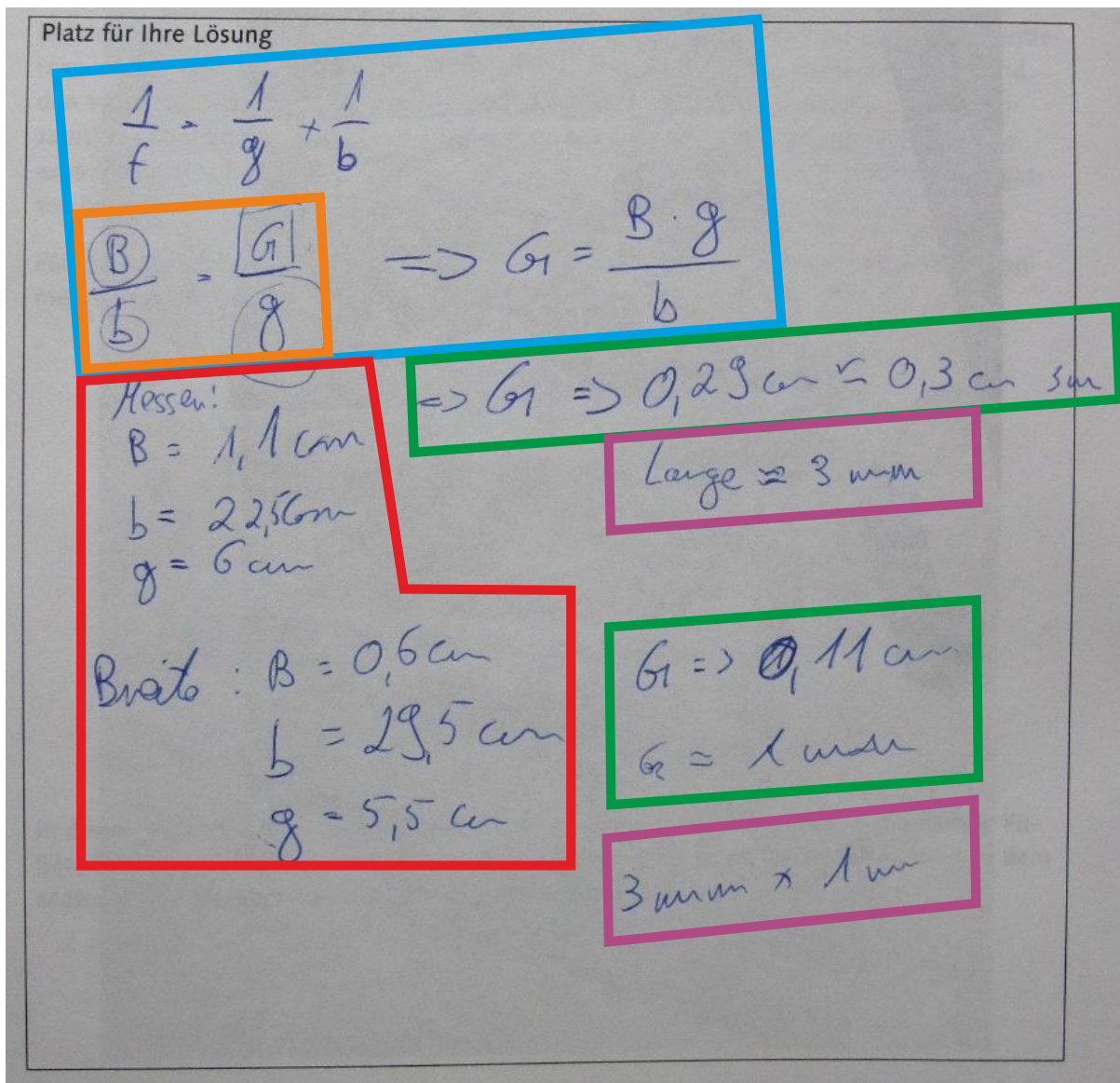
Um die Aufzeichnungen den Fähigkeiten bzw. Fertigkeiten zuzuordnen und sie in eine Reihenfolge zu bringen, wurde das folgende 3-schrittige Verfahren entwickelt.

- (I) Rater ordnen die Aufzeichnungen jeweils den Fähigkeiten bzw. Fertigkeiten nach dem Modell der Experimentierkompetenz zu. Als Hilfestellung und zur Sicherung der Objektivität wurde ein Kodiermanual entwickelt (siehe Anhang A.4.6). Das Ergebnis dieses Kodierschrittes findet sich in Abbildung 16.



**Abbildung 16: Identifikation von Fähigkeiten und Fertigkeiten: Theorie klären (blau), Experiment planen (orange), Messen (rot), Daten analysieren (grün), Schlussfolgerung (lila).**

- (II) Anschließend werden Aufzeichnungen, die bezogen auf die Leserichtung aufeinander folgen und derselben Fähigkeit bzw. Fertigkeit zugeordnet wurden, zusammengefasst. Der Schritt ist in Abbildung 17 dargestellt.



**Abbildung 17: Zusammenfassung von aufeinander folgenden Aufzeichnungen zu gleichen Fähigkeiten bzw. Fertigkeiten.**

- (III) Im dritten Schritt werden die Nummern der Handlungen, in Reihenfolge der Lese- richtung notiert. Die so entstehende Tabelle ist in Tabelle 5 dargestellt.

Die Übereinstimmung bei der Identifikation von Aufzeichnungen und der Zuteilung bzgl. der Fähigkeiten und Fertigkeiten lag zwischen den beiden Ratern bei  $\kappa = .904$ . Damit kann das produktorientierte Verfahren zur Generierung der Bewertungseinheiten als objektiv angesehen werden.

**Tabelle 5: Tabelle mit der Abfolge der Fähigkeiten bzw. Fertigkeiten basierend auf den Aufzeichnungen.**

<b>Handlung in Leserichtung</b>	<b>Fähigkeit bzw. Fertigkeit</b>	<b>Nummer der Fähigkeit bzw. Fertigkeit</b>
<b>1</b>	<b>Versuch planen</b>	<b>3</b>
<b>2</b>	<b>Beobachten und Messen</b>	<b>5</b>
<b>3</b>	<b>Daten Analysieren</b>	<b>6</b>
<b>4</b>	<b>Beobachten und Messen</b>	<b>5</b>

### **7.3.1.2 Prozessorientiertes Auswerteverfahren**

Bei der prozessorientierten Auswertung werden zusätzlich zu den Aufzeichnungen der Probanden auch die Lösungsprozesse mit bewertet (vgl. Kapitel 3). Die Bewertung des Lösungsprozesses lässt sich wie im theoretischen Teil dargestellt auf zwei Arten realisieren. Zum einen könnten Rater bei der Bearbeitung der Aufgaben in Echtzeit mithilfe eines Protokollbogens bewertet werden oder die Handlungen der Probanden werden auf Video aufgezeichnet und nachträglich bewertet. Beide Verfahren sind zeitintensiv. Ein Vorteil von Videodaten ist, dass die Informationen dauerhaft vorliegen. So kann unter anderem das Auswerteverfahren überarbeitet und die Daten rekodiert werden. Auch die Übereinstimmung zwischen Ratern lässt sich leichter untersuchen. Gleichzeitig, und dies ist bezogen auf die vorliegende Arbeit der Hauptvorteil, können Videos angehalten und gespult werden. Dies ermöglicht eine genauere Kodierung durch mehrfaches Ansehen bzw. einem ausführlichen Abgleich mit dem Kodiermanual. Gleichzeitig können verschiedene Aspekte nacheinander an demselben Video kodiert werden (vgl. Fischer & Neumann, 2012).

Entsprechend wurde die Analyse von Videodaten als prozessorientiertes Verfahren gewählt. Dabei werden Probanden bei der Bearbeitung des Experimentiertests gefilmt (siehe Abb. 18). Um die Aufzeichnungen des Probanden mit dem Video zu synchronisieren ist eine Kamera vornehmlich auf das Laborheft gerichtet. Eine zweite Kamera ist auf den gesamten Tisch gerichtet und fokussiert auf die Handlungen des Probanden.



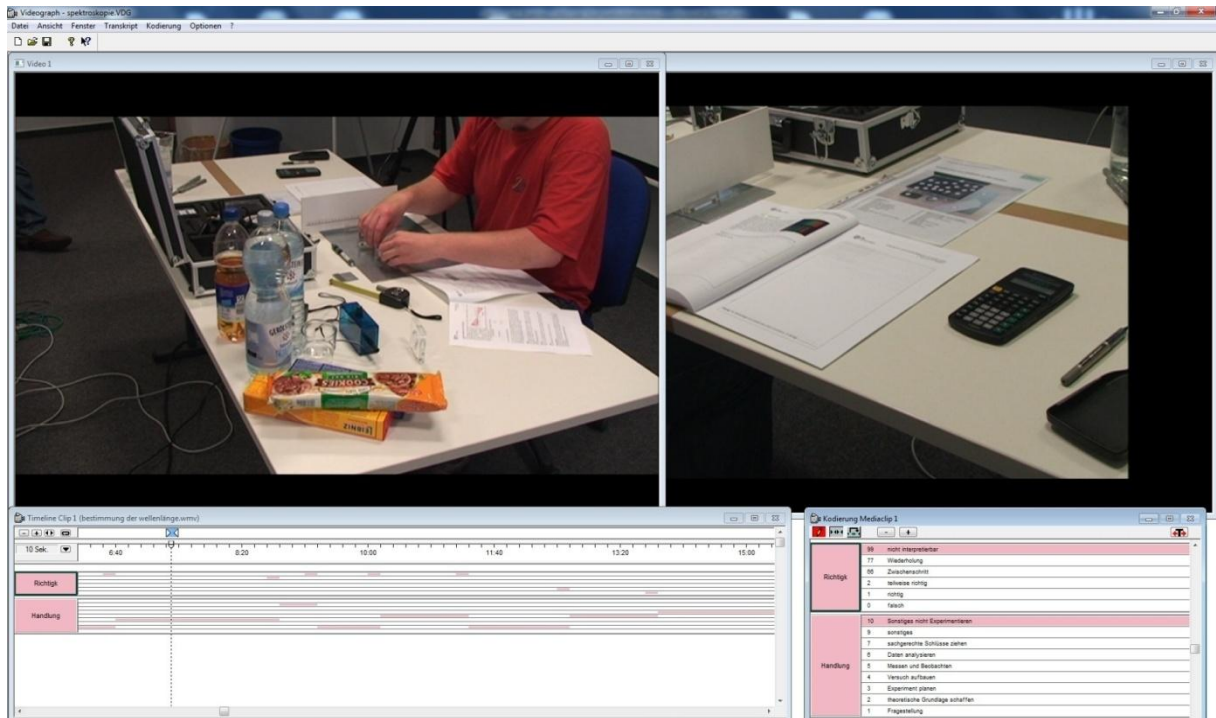
**Abbildung 18 - Testsituation bei der prozessorientierten Auswertung.**

Zur Auswertung werden die Videos mithilfe der Software *Videograph* wie folgt analysiert.

- (I) Die Videoaufzeichnungen werden in 10 Sekundenintervallen einer Fähigkeit bzw. Fertigkeit zugeordnet (siehe Abb. 19). Die Aufzeichnungen im Laborheft stehen als Hilfestellung zur Verfügung. Für diese Kodierung wurde analog zu dem Manual für die produktorientierte Auswertung ein zweites Kodiermanual für die prozessorientierte Auswertung entwickelt (siehe Anhang A.4.7).
- (II) Die Abfolge der Handlungen wird von der Software *Videograph* ausgegeben.
- (III) Mittels eines R Skripts (siehe Anhang A.4.9) werden Intervalle, die aufeinander folgen und derselben Fähigkeit bzw. Fertigkeit zugeordnet wurden, analog zur produktorientierten Auswertung zusammengefasst.

### **7.3.1.3 Referenzauswertung basierend auf dem “lauten Denken“**

Als Referenz wurde, wie im vorherigen Kapitel beschrieben, die Methode des “lauten Denkens“ gewählt. Dabei wurde äquivalent zur prozessorientierten Auswertung vorgegangen. Zusätzlich wurde den Ratern, die mit dem Videomaterial synchronisierte Tonspur des “lauten Denkens“ zur Verfügung gestellt. Wie bei der prozessorientierten Auswertung wurden Handlungen in 10-Sekunden-Intervallen Fähigkeiten bzw. Fertigkeiten zugeordnet. Dazu stand dasselbe Kodiermanual wie bei den Prozessdaten zur Verfügung (siehe Anhang A.4.7).



**Abbildung 19 - Videokodierung in 10 Sekunden Intervallen zur Identifikation der Fähigkeiten bzw. Fertigkeiten.**

### 7.3.2 Maße der Experimentierkompetenz

Jedes der vorgestellten Auswerteverfahren kann als objektiv angesehen werden und liefert eine Abfolge von Fähigkeiten bzw. Fertigkeiten. Im Weiteren wird beschrieben, wie - basierend auf diesen Daten - Maße für die Aspekte der Richtigkeit, Strukturiertheit und Zielorientiertheit sowie das Gesamtmaß der Experimentierkompetenz gebildet wurden. Da die Maße identisch für alle Auswerteverfahren sind, wird im Weiteren nicht auf die unterschiedlichen Auswerteverfahren verwiesen.

#### 7.3.2.1 Maß Richtigkeit

Der Aspekt der Richtigkeit fokussiert darauf, ob Probanden eine Handlung zu einer Fähigkeit bzw. Fertigkeit richtig durchführen und ggf. richtig dokumentieren (siehe Kapitel 2). Entsprechend werden alle identifizierten Fähigkeiten bzw. Fertigkeiten bzgl. der richtigen Durchführung im Falle des Prozesses bzw. der richtigen Aufzeichnung im Falle des Produkts bewertet. Zur Bewertung steht jeweils ein Kodiermanual zur Verfügung (siehe Produkt: Anhang A.4.6, Prozess und Referenz: Anhang A.4.7). Zur Kodierung werden die Aufzeichnungen bzw. Handlungen mit aufgabenspezifischen Musterlösungen zu den unterschiedlichen Lösungswegen verglichen. Bei Messwerten und Ergebnissen wurden Toleranzbereiche angegeben. Für jede Handlung wird dann die Richtigkeit anhand der folgenden Kriterien bewertet:

- Richtig: Das Produkt/die Handlung entspricht der Musterlösung oder ist äquivalent dazu.
- Teilweise Richtig: Das Produkt/die Handlung entspricht in Teilen der Musterlösung bzw. ist äquivalent dazu UND es fehlen Teile der Musterlösung oder es sind Fehler im Produkt/der Handlung.
- Falsch: Das Produkt/die Handlung entspricht nicht der Musterlösung oder ist durchgehend falsch.
- Wiederholung: Das Produkt/die Handlung kommt mindestens zum zweiten Mal in identischer Weise vor.
- Zwischenschritt: Das Produkt/die Handlung ist noch nicht vollständig und wird in einem späteren Intervall vervollständigt.
- Nicht Interpretierbar: Das Produkt/die Handlung sind nicht interpretierbar oder es liegt kein Produkt vor.

In die Testleistung der Richtigkeit gingen die Kategorien Richtig (1 Punkt), Teilweise Richtig (0,5 Punkte) und Falsch (0 Punkte) ein. Die anderen Kategorien dienen der Vollständigkeit, lassen aber keine Bewertung bzgl. der Richtigkeit zu. Für die Testleistung der Richtigkeit wurden sie ignoriert.

Nachdem alle Aufzeichnungen bzw. Handlungen bzgl. der Richtigkeit bewertet wurden, wird der Mittelwert aller Kodierungen zu jeder Fähigkeit und jeder Fertigkeit berechnet. Entsprechend gibt es zu jeder der sieben Fähigkeiten und Fertigkeiten des Modells der Experimentierkompetenz für jede Aufgabe genau einen Wert. Dieser liegt bedingt durch die Kodierung zwischen "0" (falsch) und "1" (richtig). Für den Fall, dass es keine Aufzeichnung bzw. Handlung zu einer Fähigkeit oder Fertigkeit gibt, wird der Score zu dieser Handlung mit "0" (falsch) kodiert. Um einen einzelnen Wert für die Richtigkeit zu jeder Aufgabe zu bekommen, wird der Mittelwert aus den sieben Scores zu den einzelnen Fähigkeiten und Fertigkeiten gebildet. Damit haben alle Fähigkeiten bzw. Fertigkeiten einen gleichwertigen Einfluss, auch wenn beispielsweise nur einmal die Theorie geklärt wurde und neun voneinander unabhängige Messungen durchgeführt wurden.

An dieser Stelle wäre es denkbar gewesen, die Gewichtung anhand der Anzahl der gewählten Handlungen durchzuführen. Aufgrund der Aufgabenentwicklung, die eine gleichgewichtete Ausprägung der Handlungen vorsieht, wurde jedoch das oben skizzierte Verfahren gewählt.

### 7.3.2.2 Maß Strukturiertheit

Bei dem Aspekt der Strukturiertheit soll gemäß Kapitel 2 bewertet werden, ob der Übergang von einer Handlung zur nächsten, bezogen auf die aktuelle Experimentiersituation, logisch sinnvoll strukturiert ist. Die Bewertung erfolgt für jeden Übergang von einer Handlung zur nächsten in “0“ für *kein logisch sinnvoller Übergang* und “1“ für *logisch sinnvoller Übergang*. In Kapitel 2.4.2 wurden verschiedene logisch richtige Handlungsübergänge identifiziert. Dabei handelte es sich um eine idealtypische Abfolge; eine Reaktion auf Fehler um diese zu beheben und Kontrollmechanismen um die Qualität der Arbeit zu überprüfen. Im Folgenden werden die einzelnen Aspekte einer logisch richtigen Abfolge der Handlungen mathematisch beschrieben. Darüber hinaus werden zwei weitere “richtige“ Handlungsübergangsarten beschrieben, die sich nicht aus der Theorie herleiten lassen. Diese Handlungsübergänge werden als notwendig angesehen, um die Qualität des Experimentierprozesses adäquat abzubilden. Bei den Übergängen handelt es sich um Sprünge und Vervollständigungen.

#### Idealtypische Abfolge

Nach dem Modell der Experimentierkompetenz ist die idealtypische Abfolge der Handlungen vorgegeben. Demnach wird zuerst die Fragestellung geklärt (1), anschließend eine theoretische Grundlage geschaffen (2). Danach wird ein Experiment geplant (3), um diesen Versuch anschließend aufzubauen (4). Dann werden Messungen bzw. Beobachtungen gemacht (5), diese Werte z.B. durch Rechnungen analysiert (6), um abschließend eine Schlussfolgerung zu ziehen (7). Da Experimentieren ein iterativer Prozess ist, kann der Schlussfolgerung (7) die erneute Klärung der Fragestellung (1) folgen (vgl. Kapitel 2). Der Übergang einer Handlung zur nächsten würde demnach logisch richtig sein, wenn die Nummer der Handlung um genau 1 zunimmt. Dies würde z.B. wie folgt aussehen:

**Tabelle 6: Handlungsübergänge bei der idealtypischen Abfolge des Experimentierens.**

Handlung	Bewertung bzgl. der Strukturiertheit
1	
2	1
3	1
4	1
5	1
6	1
7	1
1	1



## Reaktion auf Fehler

Eine Alternative für logisch richtige Handlungsübergängen ist die korrekte Reaktion auf Fehler. Erkennt der Proband einen von ihm gemachten Fehler und „springt“ an die entsprechende Stelle im Experimentierprozess zurück, um seinen Fehler zu korrigieren, so ist dies als logisch sinnvoller Handlungsübergang zu verstehen. Dies gilt aber nur, wenn wirklich ein Fehler in dieser Handlung vorliegt. Im Folgenden wird ein Beispiel gegeben:

**Tabelle 7: Handlungsübergänge bei der Reaktion auf Fehler.**

Handlung	Richtigkeit	Bewertung bzgl. der Strukturiertheit
2	1	
3	1	1
4	0	1
5	1	1
6	1	1
4	1	1
5	1	1
6	1	1
4	1	0

Der erste Übergang von Handlung 6 zu Handlung 4 wird als richtig bewertet, weil zuvor die Handlung 4 falsch ausgeführt wurde (siehe rote Markierung). Der zweite Übergang von Handlung 6 zu Handlung 4 wird bezüglich der Strukturiertheit als falsch bewertet, da Handlung 4 zuletzt als richtig bewertet wurde.

## Kontrolle

Beim Experimentieren ist es durchaus legitim, vor längeren oder wichtigen Handlungen zu kontrollieren, ob alles richtig durchdacht, geplant oder aufgebaut ist. Dies zeigt sich, z.B. wenn ein Proband vor der Messung die Theorie kontrolliert und dann direkt wieder zum Messen springt. Allgemein gilt, dass ein Proband zu einer Handlung mit niedrigerer Nummer und wieder zur ursprünglichen Handlung zurückspringt. Für die Bewertung der Strukturiertheit wird die Kontrolle weder als richtig, noch als falsch bewertet. Sie wird für die Bewertung einfach ignoriert. Das Beispiel einer Kontrolle sieht wie folgt aus:

**Tabelle 8: Handlungsübergänge bei Kontrollhandlungen.**

Handlung	Bewertung bzgl. der Strukturiertheit
2	
3	1
4	1
2	-
4	-

Der Proband generiert eine Hypothese oder klärt die theoretischen Grundlagen. Anschließend plant er ein Experiment- dieser Übergang ist bzgl. der Strukturiertheit logisch richtig. Im nächsten Schritt baut der Proband wie geplant einen Versuch auf. Auch dieser Schritt ist logisch sinnvoll. Bevor der Proband im Experimentierprozess voran schreitet, prüft er nun, ob die Hypothese durch seinen Aufbau geprüft werden kann. Er springt entsprechend zum *Hypothesen generieren/Klären der theoretischen Grundlage (II)* zurück. Nachdem er sichergestellt hat, dass sein Experiment geeignet ist, vollendet er den Aufbau (Handlung *IV Versuch fehlerfrei aufbauen*). Die Übergänge 4-2 und 2-4 werden entsprechend als Kontrolle eingestuft und nicht bewertet.

### **Sprünge**

Es ist möglich, dass ein Proband den Kreisprozess beim Experimentieren mehrfach durchläuft. Dabei ist es durchaus wahrscheinlich, dass nach der idealtypischen Abfolge Handlungen erwartet werden, die aus einem vorherigen Durchgang des Prozesses bereits vollständig und richtig vorliegen. Dies könnte z.B. der Fall sein, wenn im ersten Durchlauf ein Versuch aufgebaut wird. In einem zweiten Durchlauf soll am selben Versuchsaufbau etwas anders oder erneut gemessen werden (z.B. durch eine neue Deklaration von abhängigen- und unabhängigen Variablen). Nach der idealtypischen Abfolge wird erwartet, dass der Proband einen Versuch aufbaut. Der Versuch steht aber vollständig richtig aufgebaut bereit. Entsprechend wäre es sinnvoll, wenn der Proband diese Handlung überspringt und direkt mit der Messung beginnt. Das bedeutet, der Übergang von einer Handlung zur nächsten gilt als sinnvoll, wenn bzgl. der Reihenfolge der Handlungen diejenigen Handlungen übersprungen werden, welche in der aktuellen Experimentiersituation bereits richtig vorliegen.

Genauso kann es vorkommen, dass Probanden dem idealtypischen Verlauf folgen, einige triviale Handlungen aber im Kopf vollführen. Diese Handlungen können nur schwer beobachtet werden. Entsprechend werden Sprünge als logisch sinnvolle Handlungsübergänge eingeführt. In der folgenden Tabelle sind Handlungsübergänge und Bedingungen zu sehen, die als richtig bewertet werden:

**Tabelle 9: Handlungsübergänge bedingt durch vorherige Handlungen.**

Handlung	Vorherige Handlung	Bedingung
3	1	
4	1, 2	
5	1, 2, 3	VR <sub>4</sub> = 1 (richtiger Versuch ist aufgebaut)
6	1, 2, 3, 4	VR <sub>5</sub> = 1 (richtige Messwerte liegen vor)
7	1, 2, 3, 4, 5	VR <sub>6</sub> = 1 (richtige Analyse von Messdaten liegt vor)

In der ersten Zeile überspringt der Proband die Klärung der theoretischen Grundlagen, wenn ihm alle Grundlagen bekannt sind. In der zweiten Zeile wird die Planung des Experiments bzw. zusätzlich die Klärung der theoretischen Grundlage übersprungen. Dies ergibt z.B. im weiteren Verlauf eines Experimentierprozesses Sinn. In beiden Fällen lässt sich aus einer korrekten Planung bzw. einem korrekten Versuchsaufbau schließen, dass der Proband die nicht detektierte Handlung vollzogen haben muss. Die letzten drei Zeilen betreffen jeweils den Fall, dass ein Proband in einem späteren Experimentierzyklus genau eine Handlung überspringt, wenn diese bereits richtig durchgeführt wurde.

Ein Beispiel für die Kodierung von Sprüngen findet sich in der folgenden Tabelle. Rot markiert ist der logisch korrekte Übergang eines Sprunges.

**Tabelle 10: Handlungsübergänge bei Sprüngen.**

Handlung	Bewertung bzgl. der logischen Abfolge
1	
4	1
5	1
6	1
7	1

In dem Beispiel überspringt der Proband die Klärung der theoretischen Grundlage und die Planung des Experiments. Da er jedoch ein korrektes Experiment aufbaut, werden ihm die richtige Klärung der Theorie und die Planung des Experiments attestiert, obwohl sie nicht dokumentiert wurden.

### **Vervollständigung**

Es kann sein, dass ein Proband beim Experimentieren einen Aspekt vergessen und/oder (bewusst) ausgelassen hat. Dies lässt sich z.B. beobachten, wenn verschiedene Eigenschaften einer Komponente oder mehrere Komponenten untersucht werden sollen. Wenn ein Proband die Wellenlänge von zwei Lasern bestimmen soll, so werden einige Probanden erst die Messungen und Rechnungen zu einem Laser durchführen und anschließend äquivalent dazu die Messungen und Rechnungen für den zweiten Laser anschließen<sup>12</sup>. Dabei ändert sich nichts an den Handlungen der Planungsphase oder dem Aufbau- ganz unabhängig ob diese richtig sind. Entsprechend zählt ein Handlungsübergang als sinnvoll, wenn der Proband zu einer Handlung geht, die noch nicht vollständig abgeschlossen ist. Zum Beispiel, wenn er eine Messung abschließen will. Um diese Prozesse adäquat zu bewerten, wurde die Vervollständigung eingeführt.

Damit die Strukturiertheit problemlos berechnet werden kann, muss ein Umgang mit “sonstigen Handlungen“ einbezogen werden. Entsprechend wurde festgelegt, dass Handlungsübergänge zu “sonstigen Handlungen“ oder von “sonstigen Handlungen“ nicht bewertet werden können. Das Gesamtmaß für die Strukturiertheit wird als Mittelwert der Bewertungen über alle Handlungsübergänge gebildet. Eine spezifische Gewichtung einzelner Übergänge kann nicht durch vorliegende Erkenntnisse oder Hypothesen begründet werden.

### **Multiple Bewertungsmöglichkeiten**

Durch die vorgestellten Handlungsübergänge, ergeben sich für einen Experimentator in spezifischen Experimentiersituationen teilweise verschiedene sinnvolle Handlungsübergänge. Um das Gütemaß der Strukturiertheit verständlicher zu machen und die verschiedenen möglichen Handlungen in spezifischen Experimentiersituationen nachvollziehen zu können, wird in Tabelle 11 ein fiktives Beispiel eines Experimentierprozesses gegeben. Neben den jeweiligen

---

<sup>12</sup> Die Qualität der Strukturiertheit hängt ausdrücklich nicht davon ab, ob ein Proband vorausschauend alle notwendigen Messungen durchgeführt hat, bevor er mit der Berechnung beginnt oder den beschriebenen Aspekt zunächst ausgelassen hat und ihn nachfolgend vervollständigen muss. Bewertet wird ausschließlich, ob in der gegebenen Situation die gewählte nächste Handlung sinnvoll ist. Dass eine unterschiedliche Qualität beim Experimentieren durch die beiden Lösungswege gegeben ist, wird durch das dritte Gütemaß, die Zielorientiertheit, berücksichtigt.

Handlungen und der Bewertung der Handlung bzgl. der Richtigkeit werden zu jeder spezifischen Experimentiersituation mögliche, als sinnvoll bewertete, Folgehandlungen beschrieben. Wählt der Proband eine der sinnvoll bewerteten Folgehandlungen, so wird ihm dieser Übergang mit einer "1" als richtig bewertet (vgl. 7.3.2.2 die Tabellen zur idealen Abfolge, Kontrolle, Fehlerkorrektur, Sprüngen und Vervollständigungen).

**Tabelle 11: Beispiel eines Experimentierprozesses mit möglichen, sinnvollen Folgehandlungen**

#	Handlung	Mögliche sinnvolle Folgehandlung
1	Frage- bzw. Problemstellung entwickeln oder klären (richtig)	-Hypothese generieren (Idealtypisch)
2	Hypothese generieren/ Klären der theoretischen Grundlage (richtig)	-Experiment planen (Idealtypisch)
3	Experiment planen (falsch)	-Versuch aufbauen (Idealtypisch)
4	Versuch fehlerfrei aufbauen (richtig)	-Messen, Beobachten und Dokumentieren (Idealtypisch) -Experiment planen (Fehlerkorrektur)
5	Messen, Beobachten und Dokumentieren (falsch)	-Daten analysieren (Idealtypisch) -Experiment planen (Fehlerkorrektur)
6	Daten analysieren und Verallgemeinerungen diskutieren (falsch)	-Sachgerechte Schlüsse ziehen (Idealtypisch) -Experiment planen (Fehlerkorrektur) -Messen, Beobachten und Dokumentieren (Fehlerkorrektur)
7	Sachgerechte Schlüsse ziehen und diese Kommunizieren (richtig)	-Fragestellung klären (Idealtypisch) -Hypothese bilden (Sprung) -Experiment planen (Sprung, Fehlerkorrektur) -Messen, Beobachten und Dokumentieren (Fehlerkorrektur) -Daten analysieren (Fehlerkorrektur)
8	Experiment planen (richtig)	-Versuch aufbauen (Idealtypisch) -Messen, Beobachten und Dokumentieren (Sprung, Fehlerkorrektur) -Daten analysieren (Fehlerkorrektur)
9	Messen, Beobachten und Dokumentieren (richtig)	-Daten analysieren (Idealtypisch, Fehlerkorrektur)
10	Daten analysieren und Verallgemeinerungen diskutieren (richtig)	-Sachgerechte Schlüsse ziehen (Idealtypisch)
11	Sachgerechte Schlüsse ziehen und diese Kommunizieren (richtig)	-Fragestellung entwickeln (Idealtypisch) -Experiment abschließen

Für die Bewertung ist unwichtig, ob die sinnvoll zulässige Folgehandlung durch die *idealtypische Abfolge*, eine *Kontrollhandlung*, eine *Fehlerkorrektur*, einen *Sprung* oder zur *Vervollständigung* durchgeführt wurde. Ist die Folgehandlung bzgl. aller dargestellten sinnvollen Übergangsarten nicht sinnvoll, so wird der Übergang mit "0" nicht sinnvoll bewertet.

Obwohl die Intention eines Handlungsübergangs nicht eindeutig geklärt werden kann, lässt sich das Gütemaß eindeutig kodieren. Denn zur Bewertung eines logischen Überganges ist es nicht relevant, nach welcher Intention der Proband einen passenden Übergang gewählt hat, sondern nur ob er einen richtigen Übergang gewählt hat. Zur Bewertung der Strukturiertheit wird entsprechend die Kodierung "1" für *logisch sinnvoller Übergang* gewählt, wenn der Handlungsübergang nach mindestens einer der beschriebenen Übergangsarten (Logische Abfolge, Fehlerkorrektur, Kontrolle, Sprünge oder Vervollständigung) logisch richtig ist. Die Kodierung "0" für *kein logisch sinnvoller Übergang* wird vergeben, wenn der Handlungsübergang nach allen Übergangsarten falsch ist.

### **7.3.2.3 Maß Zielorientiertheit**

In Kapitel 2 wurden Strategien als Element der Zielorientiertheit beim Experimentieren identifiziert. Ausgehend von qualitativ einfachen Strategien wie dem Ausprobieren (Trial and Error) über Bergsteigen (Hillclimbing), Start-Ziel-Analysen (Means-End-Analysen) manifestiert sich die höchste Qualität in planungs- und analogiebasierten Lösungsstrategien. Es wurde beschrieben, dass Strategien geringerer Qualität eine höhere Anzahl an Zyklen beim Kreisprozess des Experimentierens voraussetzen als höherwertige Strategien. Die höchste Qualität ist bei genau einem Zyklus gegeben. Dieses Vorgehen wird beim planungs- bzw. analogiebasierten Experimentieren erreicht.

Zur Bewertung der Zielorientiertheit wird entsprechend die Anzahl der Zyklen betrachtet. Eine mögliche Alternative dazu wäre der Einsatz von Mustererkennungen beim Experimentieren, indem von spezifischen Mustern im Prozess auf angewendete Strategien geschlossen werden kann. Allerdings fehlen bisher Erkenntnisse über spezifische Muster zu den verschiedenen Strategien, weshalb dieses Vorgehen verworfen werden musste. Bei der Anzahl an Zyklen als Bewertungsgegenstand ergibt sich allerdings das Problem, dass Probanden beim Ausprobieren mit Glück direkt den richtigen Lösungsweg wählen. Die Wahrscheinlichkeit für diesen Fall ist durch die offene Experimentierumgebung allerdings stark reduziert. Eine nicht ausreichende interne Konsistenz könnte möglicherweise durch diese "Glückstreffer" begründet werden.

Um die Anzahl der Zyklen einer Güte zuzuordnen, wurde bei einer ersten Erprobung qualitativ ermittelt, wie viele Zyklen die verschiedenen Probanden gebraucht haben. Daraus wurde normativ die in Tabelle 12 vorgestellte Stufung festgelegt und später für die verschiedenen Auswerteverfahren erweitert. Die Unterschiede zwischen der prozessorientierten und produktorientierten Auswertung ergeben sich aus der Gesamtanzahl an identifizierten Handlungen. Im Laborheft werden nur einige der durchgeführten Handlungen aufgeführt. Bei der Videoanalyse zeigen sich jedoch weit mehr Handlungen, die ein Proband nicht durch Aufzeichnungen im Laborheft kenntlich macht.

**Tabelle 12: Normative Setzung zur Zielorientiertheit durch die Anzahl an Zyklen.**

<b>Zyklen</b>	<b>Videoanalyse (mit und ohne Lautem Denken)</b>	<b>Laborheftanalyse (mit und ohne Protokollierung)</b>
1	16/16	6/6
2	15/16	5/6
3	14/16	4/6
4	13/16	3/6
5	12/16	2/6
6	11/16	1/6
7	10/16	0/6
8	9/16	0/6
9	8/16	0/6
10	7/16	0/6
11	6/16	0/6
12	5/16	0/6
13	4/16	0/6
14	3/16	0/6
15	2/16	0/6
16	1/16	0/6
17+	0/16	0/6

## **8 Studie II – Qualitative Untersuchung mittels Lautem Denken**

In Kapitel 7 wurde nach den Anforderungen aus Kapitel 5 und einer Fachinhaltsanalyse (Kapitel 6) ein Testinstrument entwickelt. Dieses soll nach den formulierten Zielen dieser Arbeit (Kapitel 4) das Konstrukt Experimentierkompetenz im universitären Kontext für den Fachinhalt Optik mit einer ausreichenden Güte erfassen. Um herauszufinden, ob dieses Ziel erreicht werden konnte, wurde als zweites Ziel der Arbeit festgelegt, dass die tatsächliche Testgüte detailliert untersucht werden soll. In diesem Kapitel werden die qualitativen Aspekte analysiert.

### **8.1 Forschungsfragen zu den qualitativen Aspekten der Testgüte**

Die Forschungsfragen orientieren sich an den Gütemaßen von Experimentiertests. Nachfolgenden werden die Fragen zur Objektivität, Reliabilität und den verschiedenen Aspekten der Validität entwickelt.

#### **8.1.1 Objektivität**

Die Objektivität wurde als relativ problemloses Gütekriterium beschrieben (3.3.1; 5.8). Dennoch stellt eine ausreichende Objektivität eine Grundlage zur Analyse der generierten Daten dar. Zur Kontrolle der Objektivität ergibt sich:

*Forschungsfrage 2: Lässt sich mit den entwickelten Testmanualen eine ausreichende Objektivität erzielen?*

Hypothese 2a: Für die Übereinstimmung der Rater bei der Kodierung der Daten gilt Cohens  $\kappa > .6$ , was bei einer hochinferenten Kodierung als ausreichende Güte der Objektivität angesehen wird.

#### **8.1.2 Reliabilität**

Bei der Reliabilität ergeben sich in vorliegenden Arbeiten vor allem bei authentischen Testinstrumenten Schwierigkeiten. Diese lassen sich, wie in Kapitel 3.3.2 und 3.3.6 beschrieben, auf ein Problem bei der Validität zurückführen. Dies wurde damit begründet, dass nicht ausreichend Wissen vorliegt, was einzelne Aufgaben eines Testinstruments erfassen. Dadurch ist es nicht möglich, Aufgaben zu entwickeln, die dasselbe erfassen. Dies wiederum führt zwangsläufig zu Problemen bei der Reliabilität. Diese werden dadurch verstärkt, dass die Anzahl bearbeiteter Aufgaben in authentischen Testinstrumenten gering ist. Durch die mathematische Operationalisierung der Reliabilität durch Cronbachs Alpha wird mit kleiner Aufgabenanzahl



auch der Kennwert für die Reliabilität geringer. Als Grundlage für eine valide Datengenerierung ergibt sich:

*Forschungsfrage 3: Inwieweit können die verschiedenen Aspekte der Experimentierkompetenz durch die sechs Aufgaben reliabel abgebildet werden?*

Hypothese 3a: Die Inter-Item-Übereinstimmung liegt bei einem Wert von Cronbachs  $\alpha > .6$ , was als Schwellenwert für die Güte der Reliabilität gilt (vgl. 3.3.2).

### **8.1.3 Inhaltsvalidität**

Bei der Inhaltsvalidität wurde durch die Fachinhaltsanalyse sichergestellt, dass die Fachinhalte repräsentativ abgedeckt werden (5.1; 5.9; 5.10). Bei der Aufgabenentwicklung wurde angenommen, dass alle Fähigkeiten und Fertigkeiten bei allen Lösungswegen notwendig sind (5.1). Diese Hypothese muss geprüft werden, damit ergibt sich:

*Forschungsfrage 4: Nutzen Probanden alle Fähigkeiten und Fertigkeiten der Experimentierkompetenz bei der Bearbeitung jeder Aufgabe?*

### **8.1.4 Kognitive Validität**

In Kapitel 3.3.3.6 wurde beschrieben, dass die Probleme bei der kognitiven Validität darin bestehen, inwieweit das Auswerteverfahren die Handlungen adäquat abbilden kann und inwieweit die Testleistung von der Experimentierkompetenz abhängt. Zur Inhaltsvalidität wurde formuliert, dass die Probanden alle Fähigkeiten und Fertigkeiten nutzen müssen. Genauso relevant ist, dass die Probanden keine anderen, kontext-fremden Fähigkeiten und Fertigkeiten nutzen. Damit ergibt sich:

*Forschungsfrage 5: Inwieweit nutzen Probanden Fähigkeiten bzw. Fertigkeiten, die sich nicht der Experimentierkompetenz zuordnen lassen?*

Hypothese 5a: Die Anzahl der beobachteten Handlungen, die sich keiner Fähigkeit bzw. Fertigkeit der Experimentierkompetenz zuordnen lassen, sind vernachlässigbar.

In Kapitel 3.1.1 und Kapitel 3.2.3.2 wurde außerdem das Spannungsfeld zwischen ökonomischen produktorientierten Auswerteverfahren als Grundlage zur Einsetzbarkeit von Experimentiertests und aufwendigen prozessorientierten Auswerteverfahren, denen eine höhere Validität attestiert wird, beschrieben. Ideal wäre ein möglichst ökonomisches Verfahren, allerdings ist das notwendige Kriterium eine ausreichende Validität. Um zu untersuchen, welche der verschiedenen in Kapitel 3 beschriebenen Auswerteverfahren geeignet sind und um ggf.

Probleme von eher ökonomischen Verfahren zu identifizieren und zu beheben, ergeben sich zwei Herausforderungen. Zum einen muss das Auswerteverfahren geeignet sein die Handlungen der Probanden adäquat abzubilden. Wenn das nicht gelingt, dann werden entweder Konstrukt-relevante Aspekte nicht berücksichtigt und/oder es gibt Einflüsse Konstrukt-fremder Konstrukte, die nicht vernachlässigt werden können. Damit ergibt sich:

*Forschungsfrage 6: Inwieweit ist ein produkt- bzw. prozessorientiertes Auswerteverfahren in der Lage die Fähigkeiten und Fertigkeiten der Probanden bei der Bearbeitung des Tests abzubilden?*

Hypothese 6a: Ein produkt- bzw. prozessorientiertes Auswerteverfahren ist in der Lage alle Fähigkeiten und Fertigkeiten von allen Probanden bei der erfolgreichen Bearbeitung aller Aufgaben abzubilden.

Die zweite Herausforderung bezieht sich darauf, inwieweit die Schlussfolgerung von der Probandenfähigkeit durch das Auswerteverfahren zulässig ist. Also ob die diagnostizierte Leistung durch das Testinstrument mit der Experimentierkompetenz des Probanden übereinstimmt. Damit ergibt sich:

*Forschungsfrage 7: Inwieweit können mit einer produkt- bzw. prozessorientierten Auswertung die kognitiven Prozesse der Probanden adäquat abgebildet werden?*

Hypothese 7a: Die Testleistung basierend auf einer produkt- bzw. prozessorientierten Auswertung korreliert hoch mit einer Referenzleistung der Experimentierkompetenz.

## **8.2 Studiendesign**

Für die Studie wird die qualitative Methode des Lauten Denkens eingesetzt (3.3.3.6; Thelk & Hoole, 2006). Das heißt, in dieser Studie sollen die Probanden bei der Bearbeitung des Experimentiertests ihre Überlegungen laut formulieren. Durch die Formulierungen der Probanden können ihre tatsächlichen kognitiven Prozesse bestmöglich identifiziert werden. Der in Kapitel 3 dargestellte kognitive Overload durch die Bearbeitung und gleichzeitig durch die Beschreibung der Bearbeitung kann bei einer Population aus Studierenden vernachlässigt werden. Studierende sollten durch ihre bisherige Lernentwicklung in der Lage sein die beiden Handlungen problemlos simultan durchzuführen. Bevor der Experimentiertest bearbeitet wird, erhalten die Probanden eine kurze Einführung in das Experimentiermaterial. Dazu gehören Sicherheitshinweise beim Umgang mit Lasern der Klasse 1, die Bestandteil des Experimentiermaterials sind. Diese Einführung dauert etwa 5 Minuten. Anschließend wird die Methode des Lauten Denkens vorgestellt und die Probanden werden darin trainiert. Dies entspricht dem

typischen Vorgehen bei Lautem Denken-Studien (vgl. Thek 2006 #116}). Diese Phase dauert etwa 10 Minuten. Die Objektivität in die Einführung des Testmaterials und das Laute Denken sowie die Interaktionsmöglichkeiten des Testleiters werden durch die Entwicklung und Nutzung eines Manuals gewährleistet (vgl. Kapitel 7 und Anhang A.4.10). Der zeitliche Verlauf der Erhebung wird in Tabelle 13 dargestellt.

**Tabelle 13: Durchführung Studie I**

<b>Testelement</b>	<b>Dauer (in Minuten)</b>
Begrüßung und Sicherheitshinweise	5
Vorstellung und Einübung der Methode des Lauten Denkens	10
Durchführung Experimentiertest mit Lautem Denken	180
Gesamt	195

Bei der Bearbeitung des Tests werden die Probanden gefilmt.

Zur Analyse werden drei voneinander unabhängige Auswertungen durchgeführt (vgl. V.6; V.10). Als produktorientierte Auswertung werden die Aufzeichnungen der Probanden im Laborheft bewertet. Bei der prozessorientierten Auswertung werden die Aufzeichnungen im Laborheft und zusätzlich die Videoaufnahmen ohne Tonspur des Lauten Denkens bewertet. Als Referenzauswertung dient eine Analyse basierend auf Laborheft, Videodaten und Tonspur des Lauten Denkens (5.10). Die Bewertung erfolgt jeweils durch zwei Rater.

Die Hypothese zu Forschungsfrage 2 besagt, dass alle Probanden bei der Bearbeitung jeder Aufgabe alle Fähigkeiten und Fertigkeiten nach dem Modell der Experimentierkompetenz nutzen. Um Forschungsfrage 2 zu beantworten, werden die Anzahlen der jeweils benötigten Fähigkeiten und Fertigkeiten für jeden Probanden bei der Bearbeitung jeder Aufgabe gezählt. Damit die Handlungen der Probanden bestmöglich abgebildet werden, wird die Referenzauswertung für diese Analyse genutzt. Forschungsfrage 3 fokussiert auf den Einfluss von Konstrukt-fremden Einflussgrößen. Zur Untersuchung der Forschungsfrage 3 wird äquivalent zu Forschungsfrage 2 vorgegangen. Im Gegensatz zur Untersuchung von Forschungsfrage 2 werden nicht die einzelnen Handlungen zu den jeweiligen Fähigkeiten bzw. Fertigkeiten ausgezählt, sondern die Handlungen, welche keiner Fähigkeit bzw. Fertigkeit zugeordnet werden können. Diese Handlungen lassen sich Konstrukt fremden Einflussgrößen zuordnen.

Ein Vergleich der durch die verschiedenen Auswerteverfahren abgebildeten Handlungen führt zu einer Aussage von Forschungsfrage 5 zur kognitiven Validität. Aus einem Vergleich der

diagnostizierten Testleistungen zwischen den Auswerteverfahren mittels Korrelation, folgen Erkenntnisse zu Forschungsfrage 6 für die kognitive Validität.

### **8.3 Stichprobe**

Die Auswahl einer Stichprobe sollte ein repräsentatives und erschöpfendes Bild bzgl. der Zielpopulation ergeben. Bei quantitativen Studien ergibt sich die Stichprobengröße zudem aus angestrebten Rechenverfahren und den erwarteten Effekten. Bei einer qualitativen Studie ist der Aufwand sehr hoch. In dieser Arbeit beträgt die Testdurchführung für einen einzelnen Probanden 195 Minuten. Beim lauten Denken, kann bei einem Testleiter jeweils nur ein Proband gleichzeitig den Test durchführen. Andernfalls beeinflussen sich Probanden gegenseitig durch die laut geäußerten Ideen, Gedanken und Lösungswege. Die Testleistung soll basierend auf dem produkt- und dem prozessorientierten Auswerteverfahren sowie dem Referenzverfahren berechnet werden. Während die produktorientierte Auswertung ökonomisch ist, müssen zur Analyse der anderen beiden Verfahren jeweils 3 Stunden Videomaterial analysiert werden (vgl. Kapitel 5.12). Bei einem zeitlichen Faktor von 2 als Kodierdauer von Videodaten, ergibt sich ein zeitlicher Aufwand der reinen Videokodierung von 12 Stunden für jeden Probanden. Damit muss ein zeitlicher Faktor von etwa 15 Stunden für Testdurchführung und Kodierung für einen einzelnen Probanden gerechnet werden. Dies bezieht eine Schulung und Güteprüfung von Ratern, eine Entwicklung und Erprobung der Auswerteverfahren und eine Datenaufbereitung nicht mit ein. Durch diesen Aufwand und die begrenzten Ressourcen ist die Größe der Stichprobe limitiert. Im Rahmen der vorliegenden Dissertation wurde ein Aufwand von 200 Stunden für reine Testdurchführung und Kodierdauer als durchführbar eingeschätzt. Dies ergibt eine Stichprobengröße von 13-14 Probanden. Damit der Anspruch eines repräsentativen und erschöpfenden Bildes bzgl. der Zielpopulation trotzdem gegeben ist, muss die Stichprobe geschickt ausgewählt werden.

Die Zielpopulation des Experimentiertests sind Studierende, die das physikalische Anfängerpraktikum besuchen. Entsprechend werden die Probanden aus dieser Population gewählt. Bedingt durch die logistischen Gegebenheiten konnten dabei nur Probanden getestet werden, die sich freiwillig melden. Insgesamt konnten  $n=9$  Studierende gewonnen werden. Diese  $n=9$  Studierenden setzten sich aus zwei Frauen und 7 Männern zusammen. Drei Probanden waren im vierten Fachsemester, die anderen sechs Probanden im sechsten Fachsemester. Das Alter betrug Durchschnittlich 22,9 Jahre. Die Probanden hatten einen durchschnittlichen Notenschnitt von 2,9 in Klausuren zur Experimentalphysik (1,43 beim besten Probanden und 3,6 beim schlechtesten). Als Vergleich zu der Zielpopulation bearbeiteten zudem  $n=2$  wissen-

schaftliche Mitarbeiter das Testinstrument. Diese weisen eine höhere Erfahrung beim Experimentieren auf. Als zusätzliche Stichprobe wurden die  $n=5$  deutschen Finalisten der Internationalen Physik Olympiade ausgewählt. Diese befinden sich noch in der Schule, zählen aber zu den begabtesten Schülern Deutschlands im Bereich der Physik.

Insgesamt ergibt sich eine Stichprobe von  $n=16$  Probanden, wobei die deutschen Finalisten der Internationalen Physik Olympiade aus logistischen Gründen nur ein Drittel des Experimentiertests bearbeiten konnten. Der zeitliche Aufwand von Testdurchführung und Kodierung der Daten wird damit zu etwa 190 Stunden, zzgl. Schulung und Güteprüfung von Ratern, Entwicklung und Erprobung der Auswerteverfahren und einer Datenaufbereitung geschätzt.

## 8.4 Ergebnisse Studie II

### 8.4.1 Objektivität

Bei der Durchführung der Untersuchung wurden die Daten, wie in Kapitel 7.3 beschrieben, aus Laborheften und Videos extrahiert. Dazu wurden zwei Rater eingesetzt. In Tabelle 14 ist die Übereinstimmung der Rater bzgl. der Kodierungen aufgetragen.

**Tabelle 14: Beurteilerübereinstimmung zu den drei Auswerteverfahren.**

	Identifikation von Handlungen	Bewertung bzgl. der Richtigkeit
Produktorientiert	$\kappa = .904$	$\kappa = .873$
Prozessorientiert	$\kappa = .662$	$\kappa = .682$
Referenz	$\kappa = .711$	$\kappa = .662$

Die Übereinstimmung bei der produktorientierten Auswertung ist sowohl für die Identifikation der Handlungen als auch für die Bewertung bzgl. der Richtigkeit gut. Bei der prozessorientierten Auswertung und der Referenzauswertung handelt es sich um hochinferente Kodierungen. Die erzielten Übereinstimmungen sind im akzeptablen Bereich. Diese Kenngrößen bestätigen eine ausreichende Objektivität. Damit kann Hypothese 1 angenommen werden und die Daten sind für die folgenden Analysen geeignet.

### 8.4.2 Reliabilität

In Hypothese 2a geht es um die interne Konsistenz der verschiedenen Auswerteverfahren bzgl. der Aspekte Richtigkeit, Strukturiertheit und Zielorientiertheit sowie einem Gesamtmaß für Experimentierkompetenz. Die interne Konsistenz resultiert aus den Testleistungen zu den

verschiedenen Aufgaben und Qualitätsmaßen. Bei der Auswertung wird jedem Probanden für jede Aufgabe ein Wert für die Richtigkeit, die Strukturiertheit und die Zielorientiertheit zugeordnet. Aus diesen drei Werten berechnet sich die Gesamtleistung für die einzelnen Aufgaben. Jede Bewertung liegt, bedingt durch die Modellierung (vgl. Kapitel 7.3) zwischen 0 (keine Leistung) und 1 (maximale Leistung). In der nachfolgenden Tabelle sind der Mittelwert und die Standardabweichung der Testleistung über alle Probanden eingetragen. Die aufgegliederte Leistung nach Probanden findet sich im Anhang unter A.6.

**Tabelle 15: Mittelwert und Standardabweichung bzgl. der drei Qualitätsmaße des Experimentierens und der drei Auswerteverfahren.**

	Richtigkeit	Strukturiertheit	Zielorientiertheit	Gesamt
Referenz	$\bar{X} = 0,64$ $\sigma = 0,10$	$\bar{X} = 0,66$ $\sigma = 0,05$	$\bar{X} = 0,49$ $\sigma = 0,13$	$\bar{X} = 0,60$ $\sigma = 0,07$
Produkt	$\bar{X} = 0,30$ $\sigma = 0,10$	$\bar{X} = 0,75$ $\sigma = 0,08$	$\bar{X} = 0,56$ $\sigma = 0,16$	$\bar{X} = 0,58$ $\sigma = 0,09$
Prozess	$\bar{X} = 0,42$ $\sigma = 0,11$	$\bar{X} = 0,38$ $\sigma = 0,06$	$\bar{X} = 0,66$ $\sigma = 0,08$	$\bar{X} = 0,43$ $\sigma = 0,04$

Die Leistung gegliedert nach Aufgaben findet sich in Tabelle 16. Es zeigt sich, dass die ersten drei Aufgaben einen ähnlichen Schwierigkeitsgrad aufweisen. Die Aufgaben 4-5 sind bezogen auf die Referenzauswertung auch ähnlich schwer einzustufen, jedoch schwerer als die ersten drei Aufgaben. Gerade Aufgabe 5 weist einen sehr hohen fachinhaltlichen Schwierigkeitsgrad auf. Dies zeigt sich an der Richtigkeit der durchgeführten Handlungen. Diese liegen mit durchschnittlich 0,38 Punkten deutlich unter den anderen Aufgaben (0,55 – 0,81 Punkten).

**Tabelle 16: Mittlere Testleistung und Standardabweichung für die Aufgaben 1-6 (R: Richtigkeit; S: Strukturiertheit; Z: Zielorientiertheit; G: Gesamtmaß).**

	Aufgabe 1					Aufgabe 2					Aufgabe 3			
	R	S	Z	G		R	S	Z	G		R	S	Z	G
Referenz	$\bar{X} = ,76$ $\sigma = ,24$	$\bar{X} = ,66$ $\sigma = ,08$	$\bar{X} = ,56$ $\sigma = ,14$	$\bar{X} = ,66$ $\sigma = ,12$		$\bar{X} = ,78$ $\sigma = ,13$	$\bar{X} = ,73$ $\sigma = ,09$	$\bar{X} = ,51$ $\sigma = ,14$	$\bar{X} = ,67$ $\sigma = ,07$		$\bar{X} = ,81$ $\sigma = ,10$	$\bar{X} = ,70$ $\sigma = ,10$	$\bar{X} = ,55$ $\sigma = ,18$	$\bar{X} = ,69$ $\sigma = ,08$
Produkt	$\bar{X} = ,43$ $\sigma = ,21$	$\bar{X} = ,50$ $\sigma = ,17$	$\bar{X} = ,84$ $\sigma = ,05$	$\bar{X} = ,59$ $\sigma = ,13$		$\bar{X} = ,51$ $\sigma = ,13$	$\bar{X} = ,41$ $\sigma = ,11$	$\bar{X} = ,82$ $\sigma = ,06$	$\bar{X} = ,58$ $\sigma = ,06$		$\bar{X} = ,42$ $\sigma = ,17$	$\bar{X} = ,54$ $\sigma = ,14$	$\bar{X} = ,80$ $\sigma = ,07$	$\bar{X} = ,59$ $\sigma = ,07$
Prozess	$\bar{X} = ,58$ $\sigma = ,26$	$\bar{X} = ,78$ $\sigma = ,06$	$\bar{X} = ,61$ $\sigma = ,11$	$\bar{X} = ,66$ $\sigma = ,09$		$\bar{X} = ,55$ $\sigma = ,22$	$\bar{X} = ,77$ $\sigma = ,10$	$\bar{X} = ,41$ $\sigma = ,20$	$\bar{X} = ,57$ $\sigma = ,13$		$\bar{X} = ,55$ $\sigma = ,17$	$\bar{X} = ,77$ $\sigma = ,14$	$\bar{X} = ,57$ $\sigma = ,19$	$\bar{X} = ,63$ $\sigma = ,13$
	Aufgabe 4					Aufgabe 5					Aufgabe 6			
	R	S	Z	G		R	S	Z	G		R	S	Z	G
Referenz	$\bar{X} = ,55$ $\sigma = ,22$	$\bar{X} = ,70$ $\sigma = ,08$	$\bar{X} = ,35$ $\sigma = ,20$	$\bar{X} = ,53$ $\sigma = ,12$		$\bar{X} = ,38$ $\sigma = ,13$	$\bar{X} = ,63$ $\sigma = ,12$	$\bar{X} = ,53$ $\sigma = ,21$	$\bar{X} = ,51$ $\sigma = ,11$		$\bar{X} = ,58$ $\sigma = ,11$	$\bar{X} = ,57$ $\sigma = ,11$	$\bar{X} = ,47$ $\sigma = ,23$	$\bar{X} = ,54$ $\sigma = ,07$
Produkt	$\bar{X} = ,31$ $\sigma = ,26$	$\bar{X} = ,41$ $\sigma = ,25$	$\bar{X} = ,68$ $\sigma = ,33$	$\bar{X} = ,46$ $\sigma = ,22$		$\bar{X} = ,01$ $\sigma = ,04$	$\bar{X} = ,01$ $\sigma = ,29$	$\bar{X} = ,25$ $\sigma = ,41$	$\bar{X} = ,12$ $\sigma = ,21$		$\bar{X} = ,15$ $\sigma = ,14$	$\bar{X} = ,08$ $\sigma = ,15$	$\bar{X} = ,56$ $\sigma = ,42$	$\bar{X} = ,26$ $\sigma = ,21$
Prozess	$\bar{X} = ,38$ $\sigma = ,24$	$\bar{X} = ,81$ $\sigma = ,09$	$\bar{X} = ,42$ $\sigma = ,21$	$\bar{X} = ,54$ $\sigma = ,11$		$\bar{X} = ,14$ $\sigma = ,08$	$\bar{X} = ,79$ $\sigma = ,15$	$\bar{X} = ,65$ $\sigma = ,16$	$\bar{X} = ,53$ $\sigma = ,07$		$\bar{X} = ,36$ $\sigma = ,17$	$\bar{X} = ,60$ $\sigma = ,14$	$\bar{X} = ,71$ $\sigma = ,10$	$\bar{X} = ,55$ $\sigma = ,06$

Auf den dargestellten Testleistungen wurde, wie in Kapitel 3 beschrieben, die interne Konsistenz mittels Cronbachs  $\alpha$  bestimmt. Die Ergebnisse sind in Tabelle 17 dargestellt.

**Tabelle 17: Interne Konsistenz und mittlere Trennschärfe der Aspekte der Experimentierkompetenz bzgl. der verschiedenen Auswerteverfahren.**

	Richtigkeit	Strukturiertheit	Zielorientiertheit	Experimentierkompetenz
Produktorientiert	$\alpha = .61$ $r_{it} = .39$	$\alpha = .09$ $r_{it} = .06$	$\alpha = .43$ $r_{it} = .19$	$\alpha = .55$ $r_{it} = .28$
Prozessorientiert	$\alpha = .58$ $r_{it} = .35$	$\alpha = .44$ $r_{it} = .28$	$\alpha = .28$ $r_{it} = .17$	$\alpha = .02$ $r_{it} = .12$
Referenz	$\alpha = .61$ $r_{it} = .36$	$\alpha = .41$ $r_{it} = .21$	$\alpha = .76$ $r_{it} = .50$	$\alpha = .81$ $r_{it} = .58$

Die Ergebnisse offenbaren verschiedene Probleme: die Strukturiertheit kann offensichtlich von keinem Auswerteverfahren befriedigend abgebildet werden; beim Aspekt der Zielorientiertheit liegen Probleme sowohl beim produkt- als auch beim prozessorientierten Auswerteverfahren vor. Ein Gesamtmaß von Experimentierkompetenz kann mit einer guten internen Konsistenz durch das Referenzauswerteverfahren abgebildet werden. Das produktorientierte Auswerteverfahren befindet sich bzgl. der internen Konsistenz in einem Grenzbereich bei der Bestimmung von Experimentierkompetenz.

Zur differenzierteren Analyse der internen Konsistenz wurde die Trennschärfe der einzelnen Aufgaben untersucht (vgl. Tab. 18 – 21).

**Tabelle 18: Trennschärfe für den Aspekt der Richtigkeit.**

	A1	A2	A3	A4	A5	A6
<b>Produkt</b>	.62	.06	.10	.52	.41	.64
<b>Prozess</b>	.24	.11	.16	.63	.28	.67
<b>Referenz</b>	.46	.00	.53	.63	.38	.17



**Tabelle 19: Trennschärfe für den Aspekt der Strukturiertheit.**

	A1	A2	A3	A4	A5	A6
Produkt	-.23	-.01	.06	-.15	.20	.50
Prozess	.74	.04	.54	.22	.03	.09
Referenz	.08	-.22	.28	.59	.31	.22

**Tabelle 20: Trennschärfe für die Zielorientiertheit.**

	A1	A2	A3	A4	A5	A6
Produkt	-.34	.23	.02	.60	.23	.38
Prozess	.73	-.05	-.06	.18	.41	-.19
Referenz	.57	.18	.25	.68	.63	.68

**Tabelle 21: Trennschärfe für die Experimentierkompetenz.**

	A1	A2	A3	A4	A5	A6
Produkt	.21	.01	-.08	.52	.40	.81
Prozess	.10	-.29	-.30	.44	.09	.71
Referenz	.56	.22	.32	.77	.79	.70

Aus Tabelle 18 lässt sich entnehmen, dass bzgl. der Richtigkeit Aufgabe 2 das Konstrukt nicht im gleichen Maße erfasst wie die anderen Aufgaben. Aufgabe 3 trägt bei dem produkt- und prozessorientierten Verfahren zu einer unzureichenden internen Konsistenz bei. Allerdings ist die interne Konsistenz bei Aufgaben 3 in der Referenzauswertung ausreichend. Dieses Ergebnis legt nahe, dass die beiden Auswerteverfahren den Aspekt der Richtigkeit bei dieser Aufgabe nicht adäquat abbilden können.

Die Ergebnisse in Tabelle 19 zeigen, dass bzgl. der Strukturiertheit vor allem Aufgabe 2 nicht das gleiche Konstrukt misst. Allerdings legen die Ergebnisse den Schluss nahe, dass der Aspekt der Strukturiertheit überarbeitet werden muss. Das produkt- und prozessorientierte Auswerteverfahren ist nicht in der Lage, den Aspekt der Strukturiertheit adäquat abzubilden.

Zum Aspekt der Zielorientiertheit zeigen sich in Tabelle 20 Schwierigkeiten mit Aufgabe 2. Gleichzeitig sind die produkt- und prozessorientierten Auswerteverfahren nicht in der Lage diesen Aspekt adäquat abzubilden. Bezogen auf ein Gesamtmaß für Experimentierkompetenz

lässt sich aus Tabelle 21 wiederum Aufgabe 2 identifizieren, die nicht das gleiche Konstrukt misst.

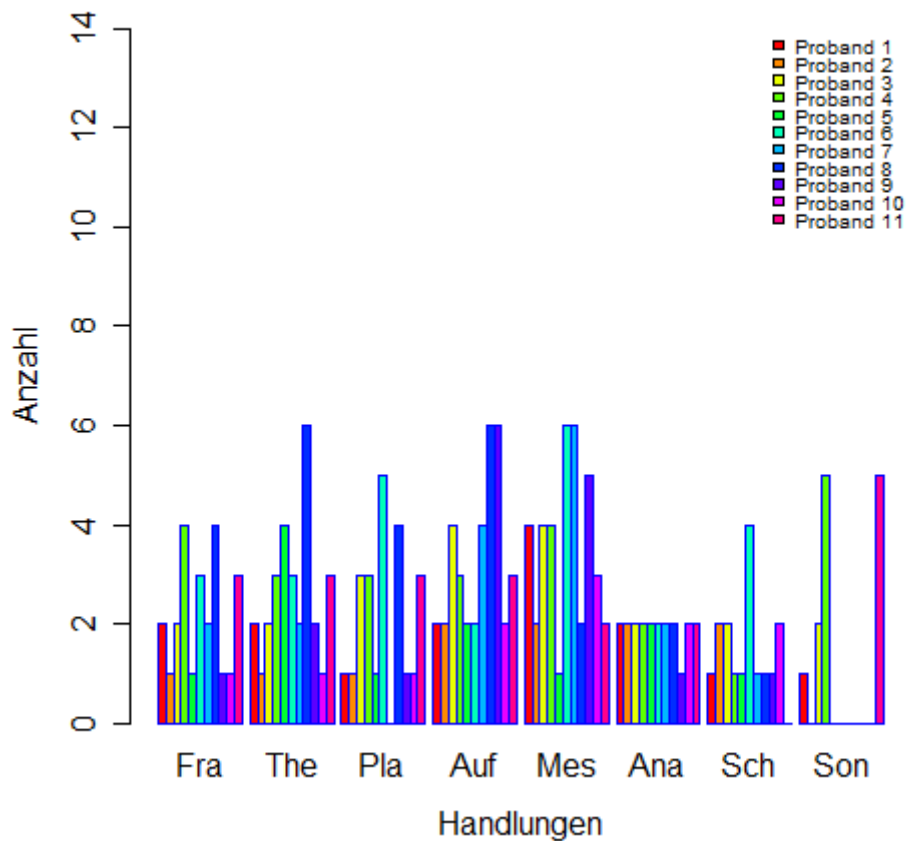
Bezogen auf die Hypothese 2a zur Reliabilität lässt sich festhalten, dass die interne Konsistenz für die Richtigkeit für alle Auswerteverfahren bestätigt werden kann. Bei der Zielorientiertheit kann nur durch das Referenzverfahren eine ausreichende interne Konsistenz erzielt werden und bei der Strukturiertheit kann keines der Auswerteverfahren eine zufriedenstellende interne Konsistenz vorweisen. Ein Gesamtmaß der Experimentierkompetenz kann durch das Referenzverfahren mit einer guten internen Konsistenz abgebildet werden.

### **8.4.3 Inhaltsvalidität**

Forschungsfrage 3 fokussiert auf die Passung der relevanten Fähigkeiten und Fertigkeiten zwischen dem Modell der Experimentierkompetenz und den entwickelten Experimentieraufgaben. Untersucht wird, ob die Probanden zur Bearbeitung der Experimentieraufgaben alle Fähigkeiten bzw. Fertigkeiten nach dem Modell der Experimentierkompetenz nutzen. Dazu wurde ausgezählt, wie oft Handlungen zu den einzelnen Fähigkeiten bzw. Fertigkeiten bei den einzelnen Probanden in den unterschiedlichen Aufgaben identifiziert werden konnten. Dafür wurde das Referenzauswerteverfahren genutzt, da dieses die Handlungen der Probanden bestmöglich abbilden kann (vgl. Kapitel 3). Die Ergebnisse der Handlungen gegliedert nach den sechs Aufgaben finden sich in den Abbildungen 19-24.

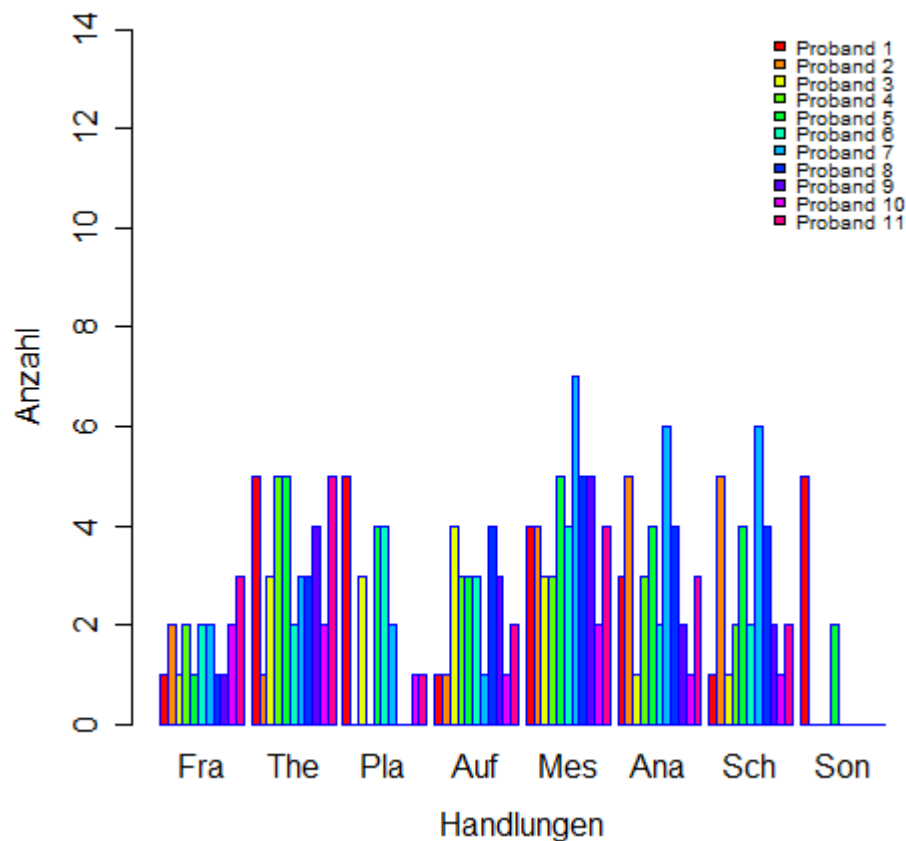
Aus Abbildung 19 lässt sich ablesen, dass die verschiedenen Probanden bei der Bearbeitung von Aufgabe 1 alle Fähigkeiten bzw. Fertigkeiten nach der Modellierung von Experimentierkompetenz genutzt haben. Nur bei zwei Probanden konnten nicht alle Fähigkeiten bzw. Fertigkeiten identifiziert werden (Proband 7 bei *Experiment planen* und Proband 11 bei der *Schlussfolgerung*). Die Anzahl der Handlungen ist relativ ähnlich verteilt, wobei ein leichter Schwerpunkt auf den Handlungen der Durchführung liegt. Handlungen, die sich keiner der Fähigkeiten bzw. Fertigkeiten nach dem Modell der Experimentierkompetenz zuordnen lassen, wurden nur bei 4 Probanden identifiziert.

## Referenz Auswertung - Aufgabe 1



**Abbildung 19: Handlungen bzgl. der verschiedenen Fähigkeiten und Fertigkeiten bei Aufgabe 1 bzgl. der Referenzauswertung; Fra (Fragestellung entwickeln/ klären); The (Theoretische Grundlagen klären/ Hypothese bilden); Pla (Experiment planen); Auf (Versuch fehlerfrei aufbauen); Mes (Messen und Beobachten); Ana (Daten analysieren); Sch (Schlussfolgerungen ziehen und diese kommunizieren); Son (Sonstiges).**

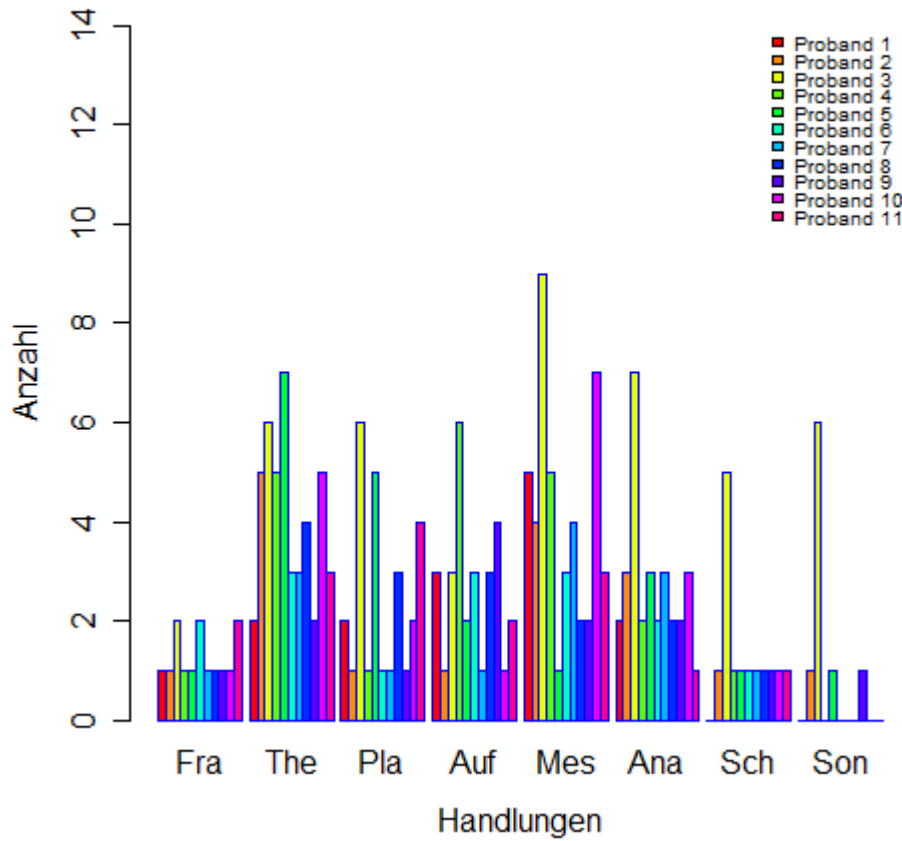
## Referenz Auswertung - Aufgabe 2



**Abbildung 20: Handlungen bzgl. der verschiedenen Fähigkeiten und Fertigkeiten bei Aufgabe 1 bzgl. der Referenzauswertung; Fra (Fragestellung entwickeln/ klären); The (Theoretische Grundlagen klären/ Hypothese bilden); Pla (Experiment planen); Auf (Versuch fehlerfrei aufbauen); Mes (Messen und Beobachten); Ana (Daten analysieren); Sch (Schlussfolgerungen ziehen und diese kommunizieren); Son (Sonstiges).**

Für Aufgabe 2 ergibt sich ein ähnliches Bild (vgl. Abb. 20). Hier konnten bei vier Probanden keine Handlungen zum *Experiment planen* identifiziert werden (Proband 2, 4, 8 und 9), alle anderen Fähigkeiten bzw. Fertigkeiten wurden von allen Probanden bei der Bearbeitung von Aufgabe 2 genutzt. Ähnlich wie in Aufgabe 1 sind die Handlungen relativ ähnlich verteilt. Es findet sich ein leichter Schwerpunkt beim *Messen und Beobachten*, nicht jedoch beim *Versuch fehlerfrei aufbauen*. Handlungen, die sich nicht dem Modell der Experimentierkompetenz zuordnen lassen, konnten bei zwei der Probanden beobachtet werden.

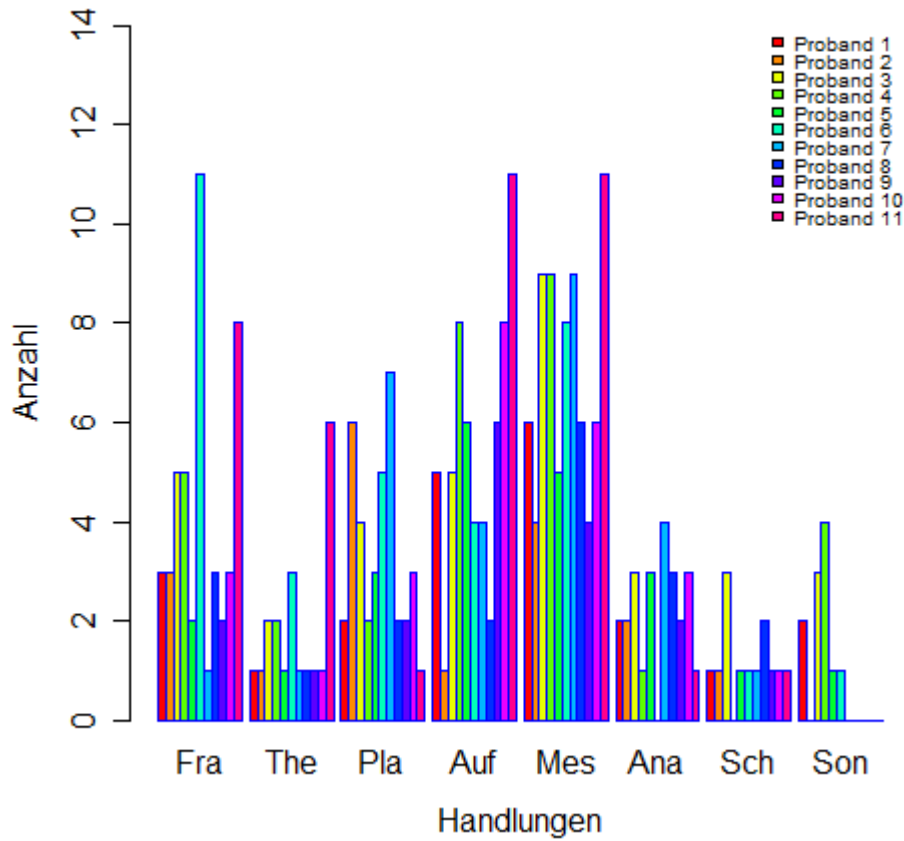
### Referenz Auswertung - Aufgabe 3



**Abbildung 21: Handlungen bzgl. der verschiedenen Fähigkeiten und Fertigkeiten bei Aufgabe 1 bzgl. der Referenzauswertung; Fra (Fragestellung entwickeln/ klären); The (Theoretische Grundlagen klären/ Hypothese bilden); Pla (Experiment planen); Auf (Versuch fehlerfrei aufbauen); Mes (Messen und Beobachten); Ana (Daten analysieren); Sch (Schlussfolgerungen ziehen und diese kommunizieren); Son (Sonstiges).**

Vergleichbar sind auch die Ergebnisse zu Aufgabe 3 (vgl. Abb. 21). Nur ein Proband hat keine *sachgerechten Schlüsse gezogen* (Proband 1), alle anderen Fähigkeiten bzw. Fertigkeiten wurden von allen Probanden bei der Bearbeitung von Aufgabe 3 genutzt. Es gibt einen leichten Schwerpunkt auf dem *Messen und Beobachten* sowie der *Klärung der theoretischen Grundlagen*. Insgesamt ist die Anzahl der identifizierten Handlungen weniger homogen. Sonstige Handlungen konnten nur bei 4 Probanden identifiziert werden.

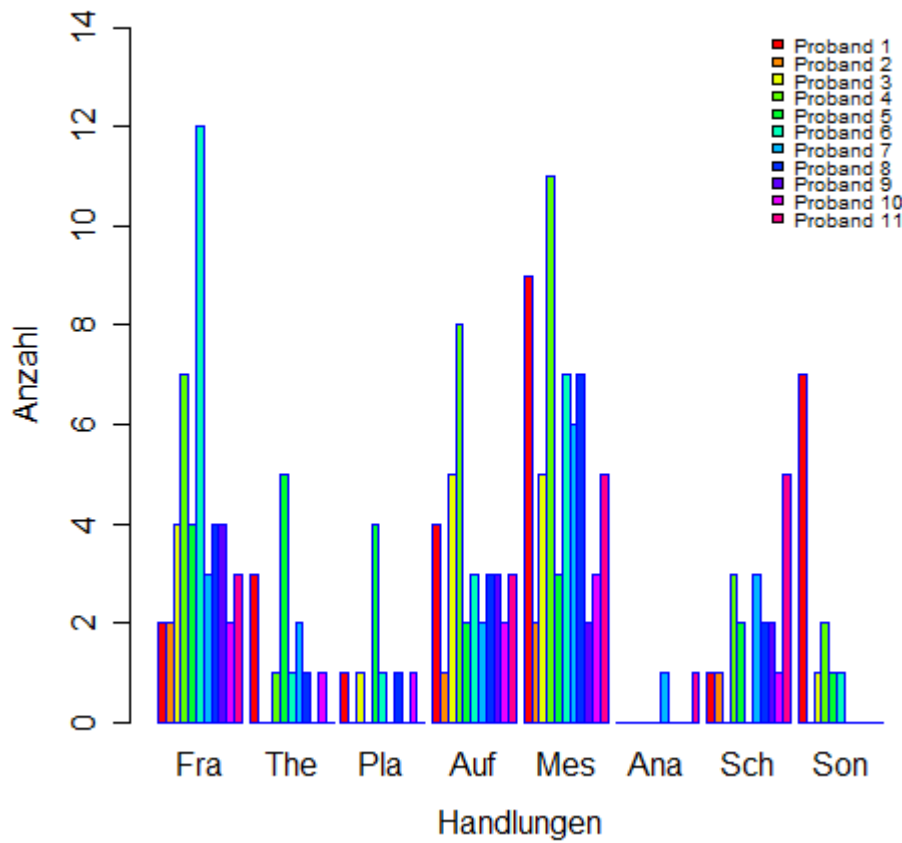
## Referenz Auswertung - Aufgabe 4



**Abbildung 22: Handlungen bzgl. der verschiedenen Fähigkeiten und Fertigkeiten bei Aufgabe 1 bzgl. der Referenzauswertung; Fra (Fragestellung entwickeln/ klären); The (Theoretische Grundlagen klären/ Hypothese bilden); Pla (Experiment planen); Auf (Versuch fehlerfrei aufbauen); Mes (Messen und Beobachten); Ana (Daten analysieren); Sch (Schlussfolgerungen ziehen und diese kommunizieren); Son (Sonstiges).**

Aus Abbildung 22 geht hervor, dass nur zwei Probanden in Aufgabe 4 nicht alle Handlungen vollzogen haben (Proband 6 *Daten analysieren* und Proband 4 *sachgerechte Schlüsse ziehen*). Der Schwerpunkt liegt deutlich auf der Durchführung, vor allem dem *Messen und Beobachten*. Fünf Probanden zeigen sonstige Handlungen.

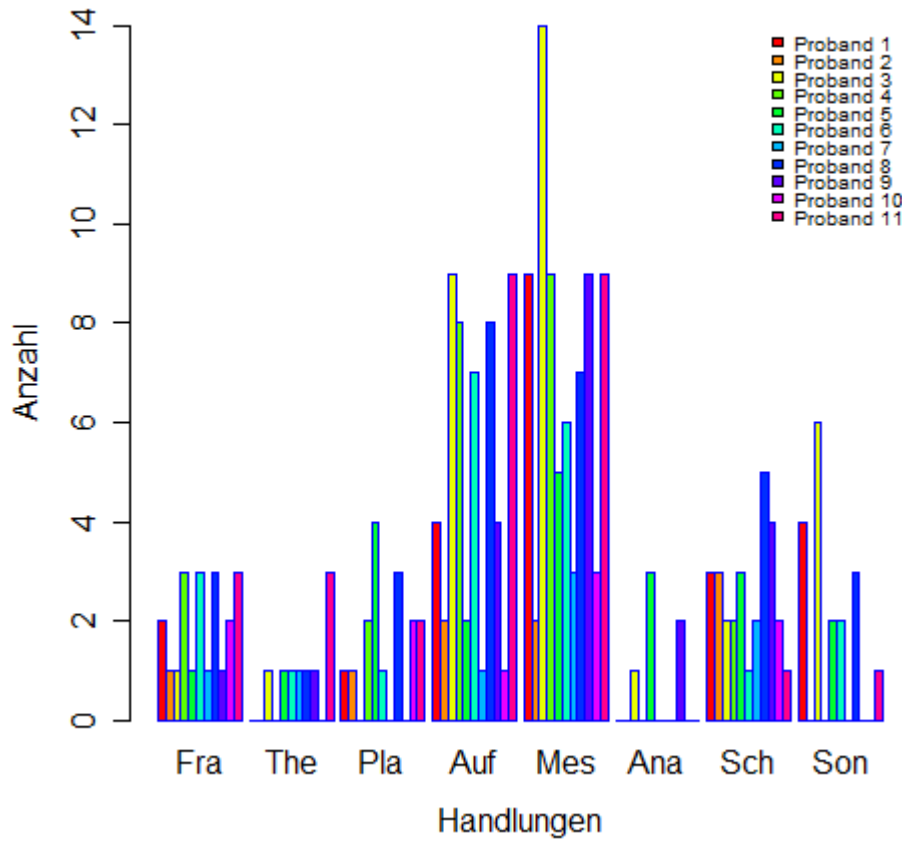
## Referenz Auswertung - Aufgabe 5



**Abbildung 23: Handlungen bzgl. der verschiedenen Fähigkeiten und Fertigkeiten bei Aufgabe 1 bzgl. der Referenzauswertung; Fra (Fragestellung entwickeln/ klären); The (Theoretische Grundlagen klären/ Hypothese bilden); Pla (Experiment planen); Auf (Versuch fehlerfrei aufbauen); Mes (Messen und Beobachten); Ana (Daten analysieren); Sch (Schlussfolgerungen ziehen und diese kommunizieren); Son (Sonstiges).**

Die Ergebnisse zu Aufgabe 5 unterscheiden sich deutlich von den vorherigen Aufgaben (siehe Abb. 23). Konnten bisher bei maximal vier Probanden nur sechs Handlungen identifiziert werden, so nutzten alle anderen Probanden alle sieben Handlungen. Bei Aufgabe 5 zeigte kein Proband alle Handlungen. Fünf Probanden zeigten 6 Handlungen, drei Probanden zeigten 5 Handlungen und drei Probanden zeigten sogar nur 4 Handlungen. Nur bei den Handlungen *Fragestellung klären* und bei der Durchführung wurden von allen Probanden Handlungen identifiziert. Auf diesen Handlungen liegt auch ein deutlicher Schwerpunkt. Bei fünf Probanden konnten sonstige Handlungen identifiziert werden.

## Referenz Auswertung - Aufgabe 6



**Abbildung 24: Handlungen bzgl. der verschiedenen Fähigkeiten und Fertigkeiten bei Aufgabe 1 bzgl. der Referenzauswertung; Fra (Fragestellung entwickeln/ klären); The (Theoretische Grundlagen klären/ Hypothese bilden); Pla (Experiment planen); Auf (Versuch fehlerfrei aufbauen); Mes (Messen und Beobachten); Ana (Daten analysieren); Sch (Schlussfolgerungen ziehen und diese kommunizieren); Son (Sonstiges).**

Bei Aufgabe 6 nutzt nur ein Proband alle Handlungen, während jeweils fünf Probanden 6 bzw. 5 Handlungen zeigen. Vor allem bei *Daten analysieren* konnten nur bei drei Probanden Handlungen identifiziert werden. Sechs Probanden nutzen sonstige Handlungen.

Die Ergebnisse zeigen, dass die Probanden zur Bearbeitung der ersten vier Aufgaben des Experimentiertests alle Fähigkeiten bzw. Fertigkeiten nach dem Modell der Experimentierkompetenz nutzen. Bezogen auf die 7 Handlungen bei 11 Probanden sind die 1 bis 4 nicht identifizierten Handlungen bei den jeweiligen Aufgaben zu vernachlässigen. Dies entspricht einem prozentualen Anteil von 1,3 bis 5,2 % der gesamten Handlungen. Nur ein Proband nutzte in den vier Aufgaben zwei Handlungen nicht (Proband 4). Bei 7 der Probanden konnte insge-



samt eine Handlung nicht identifiziert werden und 3 Probanden zeigten alle Handlungen. Für die ersten vier Aufgaben kann die Hypothese 3a entsprechend bestätigt werden.

Bei Aufgabe 5 und 6 zeigten sich Schwierigkeiten bzgl. der identifizierten Handlungen. Bei Aufgabe 5 zeigte kein Proband alle Handlungen und bei Aufgabe 6 zeigten nur 3 Probanden Handlungen zum Daten analysieren. Hypothese 3a kann daher für diese Aufgaben nicht bestätigt werden. Um die Probleme zu identifizieren, wurden die gesammelten Daten zu diesen Aufgaben qualitativ analysiert.

Bei einer rein explorativen Analyse der Videodaten zu Aufgabe 5 zeigte sich, dass die Probanden Schwierigkeiten hatten, die Fragestellung zu verstehen. Daraus ergibt sich ein Schwerpunkt beim *Fragestellungsnachvollziehen*. Die Probanden arbeiteten nach der Strategie Trial and Error. Dies erklärt den zweiten Schwerpunkt bei der Versuchsdurchführung. Aufgrund der Verständnisprobleme wurden die anderen Handlungen nur teilweise durchgeführt, da grundlegendes Wissen nicht vorhanden war. Wenn jemand nicht weiß, worum es geht und was er tun soll, so lässt sich auch kein Experiment planen oder wenn ein Proband nicht weiß was er messen soll, so kann er keine Daten analysieren. Auch die Testleistung bzgl. der Richtigkeit bekräftigt diese Einschätzung. Diese ist für Aufgabe 5 im Mittel am niedrigsten. Das bedeutet, dass die Probanden im Vergleich zu den anderen Aufgaben mehr Handlungen fehlerhaft durchgeführt haben.

Die fehlenden Handlungen zur Analyse der Daten aus Aufgabe 6 lassen sich mit der Charakteristik der Aufgabe erklären. Die Aufgabe zielt auf die Identifikation der optischen Eigenschaften einer Blackbox (vgl. Anhang A.4.3). Durch die Aufgabentexte wurde der Eindruck erweckt, dass nur die optischen Komponenten bestimmt werden müssen. Entsprechend haben nur wenig Probanden die optischen Eigenschaften der Komponenten, beispielsweise die Gitterkonstante, berechnet.

#### **8.4.4 Kognitive Validität**

Forschungsfrage 4 fokussiert auf den Einfluss von Konstrukt-fremden Fähigkeiten bzw. Fertigkeiten auf die Testleistung. Untersucht wird wie viele der Handlungen der verschiedenen Probanden sich nicht dem Modell der Experimentierkompetenz zuordnen lassen.

Dazu werden die Handlungen betrachtet, welche als “sonstige Handlung“ kodiert wurden. Dies sind wie in Kapitel 5.9 dargestellt sowohl *nicht eindeutig zuordnungsbar* als auch Konstrukt-fremde Handlungen. Wenn alle diese Handlungen als Konstrukt-fremd interpretiert werden, wird der Einfluss von Konstrukt-fremden Fähigkeiten bzw. Fertigkeiten entsprechend

überschätzt. Sollte die Gesamtanzahl an "sonstigen Handlungen" vernachlässigbar sein, so ist erst recht die Anzahl an Konstrukt-fremden Fähigkeiten bzw. Fertigkeiten zu vernachlässigen. Zur Untersuchung werden basierend auf der Referenzauswertung die Handlungen von jedem Probanden bzgl. der verschiedenen Fähigkeiten bzw. Fertigkeiten ausgezählt. Betrachtet wird der Anteil der sonstigen Handlungen (siehe Inhaltsvalidität Abb. 19-24, [Son]).

Aus den Abbildungen 19-24 lässt sich ablesen, dass 2-6 Probanden pro Aufgabe Handlungen aufweisen, die als sonstige Handlung kodiert werden. Für alle Probanden wurde der Anteil sonstiger Handlungen für jede Aufgabe bestimmt. Durchschnittlich wurden je nach Aufgabe 2,48 – 6,15 % der Handlungen als "sonstige Handlung" kodiert. Das bedeutet, der Einfluss von konstrukt-fremden Fähigkeiten bzw. Fertigkeiten ist gering.

Allerdings finden sich Fälle, in denen einzelne Probanden in spezifischen Aufgaben einen deutlich höheren Anteil sonstiger Handlungen aufweisen. Das bedeutet, diese Probanden könnten eine nicht vernachlässigbare Menge an Konstrukt-fremden Fähigkeiten bzw. Fertigkeiten nutzen. Die Videodaten der Bearbeitung dieser Probanden bei den jeweiligen Aufgaben wurden erneut analysiert, um zu überprüfen, ob die "sonstigen Handlungen" als Teile von Experimentierkompetenz oder Konstrukt-fremden Fähigkeiten bzw. Fertigkeiten verstanden werden müssen. Zur Analyse wurden alle Bearbeitungen einer Aufgabe durch einen Probanden analysiert, in denen mehr als 10% "sonstige Handlungen" kodiert wurden. Diese Bearbeitungen (Proband und Aufgabe) sind in Tabelle 22 notiert. Zusätzlich enthält die Tabelle den jeweiligen kategorisierten Anteil an "sonstigen Handlungen". In der vierten Spalte der Tabelle sind die qualitativen Beschreibungen der jeweiligen als "sonstige Handlung" kodierten Handlungen. Insgesamt wurden auf diese Weise 38 Handlungen beschrieben. Dabei zeigt sich, dass bei 12 Handlungen trotz lautem Denken und Videoaufzeichnungen nicht beschrieben werden konnte, was die Probanden machen. Basierend auf insgesamt 1348 identifizierten Handlungen sind diese 12 Handlungen zu vernachlässigen, so dass die Methode des lauten Denkens trotzdem als erfolgreich bei der Identifikation der Fähigkeiten und Fertigkeiten angesehen werden kann. Von den 38 "sonstigen" Handlungen wurden fünf als Verständnisfragen identifiziert. Auch diese sind im Vergleich zu der Gesamtanzahl an identifizierten Handlungen zu vernachlässigen. Von den verbleibenden 21 "sonstigen" Handlungen waren 19 zwar der Experimentierkompetenz zuzuordnen, allerdings nicht eindeutig einer Fähigkeit bzw. Fertigkeit. Beispielsweise schaut ein Proband sich einen aufgebauten Versuch an, plant dabei ein Experiment und diskutiert zeitgleich die Theorie.

**Tabelle 22: Einzelfälle mit hohem Einfluss sonstiger Handlungen.**

Proband	Aufgabe	Anteil sonstiger Handlungen	Qualitative Beschreibung der als "Sonstiges" kodierten Handlungen, basierend auf den Video- und Lautes Denken-Daten
Proband 1	Aufgabe 2	20 %	<p>-Der Proband benennt ein Problem, dass ihn am Weiterarbeiten hindert. Er kann dieses Problem lösen und nimmt den Experimentierprozess wieder auf.</p> <p>-Der Proband beschreibt, dass er etwas Messen möchte. Anschließend sucht er nach dem Maßband. Nach dem Kodiermanual ist diese Handlung sowohl der Fertigkeit „Messen und Beobachten“ als auch der Fertigkeit „Aufbauen“ zuzuordnen. Da keine eindeutige Kodierung möglich ist, wird eine sonstige Handlung kodiert.</p> <p>-Der Proband führt im gesamten Bewertungsintervall weder eine Handlung aus, noch teilt er seine Gedanken laut mit.</p>
Proband 1	Aufgabe 5	25,93 %	<p>-Der Proband wechselt sehr schnell zwischen verschiedenen Handlungen, die sich der Experimentierkompetenz zuordnen lassen. Entsprechend ist keine eindeutige Zuordnung möglich.</p> <p>-Der Proband äußert ein Problem, dass verschiedenen Fähigkeiten bzw. Fertigkeiten nach der Experimentierkompetenz zugeordnet werden kann. Eine eindeutige Kodierung ist nicht möglich.</p> <p>-Der Proband äußert eine Verständnisfrage, die der Testleiter beantwortet.</p> <p>-Der Proband führt im gesamten Bewertungsintervall weder eine Handlung aus, noch teilt er seine Gedanken laut mit.</p> <p>-Der Proband äußert eine Verständnisfrage, die der Testleiter beantwortet.</p>

<b>Proband</b>	<b>Aufgabe</b>	<b>Anteil sonstiger Handlungen</b>	<b>Qualitative Beschreibung der als "Sonstiges" kodierten Handlungen, basierend auf den Video- und Lautes Denken-Daten</b>
Proband 1	Aufgabe 6	17,39 %	<p>-Der Proband führt im gesamten Bewertungsintervall weder eine Handlung aus, noch teilt er seine Gedanken laut mit.</p> <p>-Der Proband äußert eine Verständnisfrage, die der Testleiter beantwortet.</p> <p>-Der Proband befolgt einen Sicherheitshinweis vom Testleiter.</p>
Proband 3	Aufgabe 3	13,64 %	<p>-Der Proband notiert ein Ergebnis und zieht im Kopf eine Schlussfolgerung, um das Ergebnis zu verifizieren. Eine eindeutige Zuordnung ist nicht möglich.</p> <p>-Der Proband zeigt Handlungen die verschiedenen Fähigkeiten bzw. Fertigkeiten der Experimentierkompetenz zugeordnet werden können. Keine eindeutige Zuordnung möglich.</p> <p>-Der Proband führt im gesamten Bewertungsintervall weder eine Handlung aus, noch teilt er seine Gedanken laut mit.</p> <p>-Es ist nicht ersichtlich, ob der Proband in den theoretischen Grundlagen oder in seinen Aufzeichnungen liest. Entsprechend ist keine eindeutige Zuordnung möglich.</p> <p>-Der Proband führt im gesamten Bewertungsintervall weder eine Handlung aus, noch teilt er seine Gedanken laut mit.</p> <p>-Der Proband geht das Experiment noch einmal im Kopf durch, um einen Fehler zu finden. Das Vorgehen ist nicht eindeutig zuzuordnen.</p>

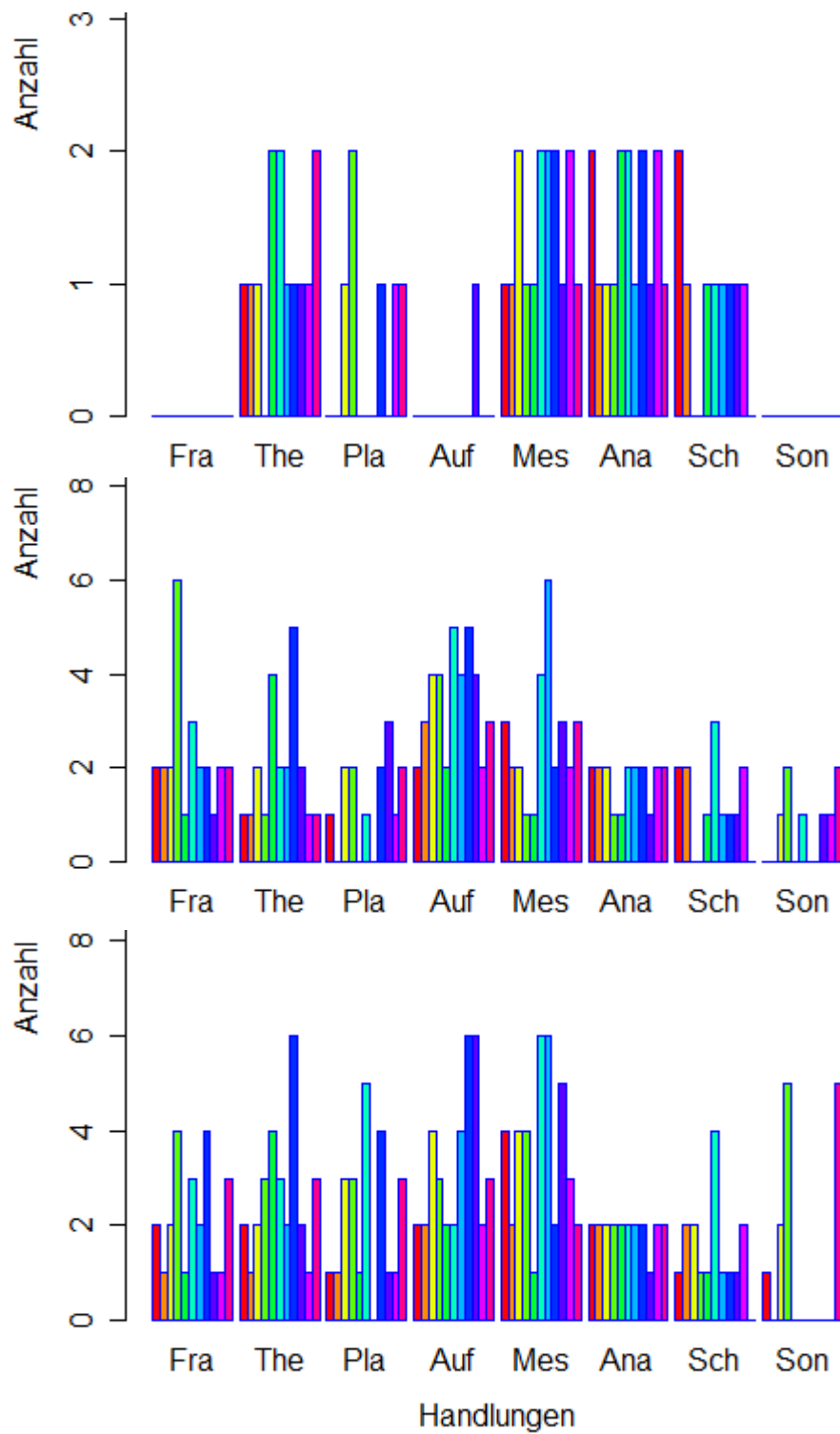
Proband	Aufgabe	Anteil sonstiger Handlungen	Qualitative Beschreibung der als "Sonstiges" kodierten Handlungen, basierend auf den Video- und Lautes Denken-Daten
Proband 3	Aufgabe 6	17,65 %	<p>-Der Proband äußert eine Verständnisfrage.</p> <p>-Der Proband führt im gesamten Bewertungsintervall weder eine Handlung aus, noch teilt er seine Gedanken verständlich mit.</p> <p>-Der Proband führt im gesamten Bewertungsintervall weder eine Handlung aus, noch teilt er seine Gedanken laut mit.</p> <p><b>-Der Proband schaut sich die Aufzeichnungen zu vorherigen Aufgaben an.</b></p> <p>-Der Proband führt im gesamten Bewertungsintervall weder eine Handlung aus, noch teilt er seine Gedanken verständlich mit.</p> <p>-Der Proband führt im gesamten Bewertungsintervall weder eine Handlung aus, noch teilt er seine Gedanken laut mit.</p>
Proband 4	Aufgabe 1	20 %	<p>-Der Proband macht eine Pause um ein Glas Wasser zu trinken.</p> <p>-Der Proband schaut sich die nachfolgenden Seiten im Testheft an, um anschließend die theoretischen Grundlagen zu suchen.</p> <p>-Der Proband reflektiert das bisherige Vorgehen. Eine eindeutige Zuordnung ist nicht möglich.</p> <p>-Der Proband führt im gesamten Bewertungsintervall weder eine Handlung aus, noch teilt er seine Gedanken verständlich mit.</p> <p>-Der Proband äußert seinen Gefühlszustand.</p>

Proband	Aufgabe	Anteil sonstiger Handlungen	Qualitative Beschreibung der als „Sonstiges“ kodierten Handlungen, basierend auf den Video- und Lautes Denken-Daten
Proband 4	Aufgabe 4	12,90 %	<p>-Der Proband baut einen Versuchsaufbau von einer vorherigen Aufgabe ab.</p> <p>-Der Proband äußert eine Verständnisfrage, die der Testleiter beantwortet.</p> <p><b>-Der Proband blättert im Testheft.</b></p> <p>-Der Proband führt im gesamten Bewertungsintervall weder eine Handlung aus, noch teilt er seine Gedanken laut mit.</p>
Proband 11	Aufgabe 1	23,81 %	<p>-Der Proband schaut sich einen aufgebauten Versuch an, plant dabei ein Experiment und diskutiert die Theorie. Dies lässt keine eindeutige Zuordnung zu.</p> <p>-Der Proband überlegt, wo er einen Fehler gemacht hat.-Der Proband benennt einen Fehler, den er aufgrund fehlender Zeit nicht finden kann und beendet die Aufgabe.</p>

Es verbleiben zwei Handlungen, die als Konstrukt-fremde Fähigkeiten bzw. Fertigkeiten oder alternative Strategien angesehen werden können: Einmal blättert ein Proband durch das Testheft um möglicherweise einen Hinweis zur Bearbeitung der aktuellen Aufgaben zu finden und ein anderes Mal schaut ein anderer Proband sich die Bearbeitung vorheriger Aufgaben an, um möglicherweise Hinweise zur Bearbeitung der aktuellen Aufgaben zu erlangen. Dies lässt sich als experimentierunabhängige Strategie werten. Das bedeutet, von insgesamt 1348 Handlungsintervallen wurden 38 Intervalle als „sonstige Handlungen“ bewertet. Von diesen 38 „sonstigen Handlungen“ lassen sich 2 Handlungen als Konstrukt-fremd beschreiben. Dies ist ein zu vernachlässigender Anteil. Allerdings wurden 10-Sekunden Intervalle in denen sowohl „sonstige Handlungen“ als auch Handlungen die einer Fähigkeit bzw. Fertigkeit nach dem Modell der Experimentierkompetenz zugeordnet werden können, den jeweiligen Handlungen im Experimentierprozess zugeordnet. Die 38 beschriebenen „sonstigen Handlungen“ enthalten ausschließlich „sonstige Handlungen“ über das gesamte 10-Sekunden Intervall. Die Ergebnisse der Untersuchung bestätigen Hypothese 4a, dass die Testleistung auf die Experimentierkompetenz der Probanden und keine weiteren, konstrukt-fremden Konstrukte zurückzuführen ist.

In Forschungsfrage 5 wird auf die Frage fokussiert, inwieweit ein produkt- bzw. prozessorientiertes Auswerteverfahren dieselben Handlungen wie das Referenzverfahren abbilden. Dazu wurde äquivalent zu Hypothese 2 die Anzahl der jeweiligen Handlungen für jeden Probanden bei jeder Aufgabe ausgezählt. Im Vergleich zu Hypothese 2 wurde dies für alle drei Auswerteverfahren gemacht. Die Daten finden sich zum Vergleich am Beispiel von Aufgabe 1 in Abbildung 25. Die Daten zu den anderen Aufgaben finden sich im Anhang unter A.7.

Nachfolgend werden die Ergebnisse des Vergleichs zwischen der produkt- bzw. prozessorientierten Auswertung und der Referenzauswertung mittels der Methode des lauten Denkens (vgl. Ergebnisse der Referenzauswertung in 8.4.3) dargestellt.



**Abbildung 25: Handlungen der einzelnen Probanden (Proband 1 jeweils in rot auf der linken Seite) über die drei Auswerteverfahren Produkt (oben), Prozess (Mitte) und Referenz (unten) für Aufgabe 1.**



### **Produktorientiert**

Beim produktorientierten Auswerteverfahren wurden in Aufgabe 1 deutlich weniger Handlungen identifiziert. Während die maximale Ausprägung bei prozessorientierter Auswertung und der Referenz bei sechs Handlungen lagen, können durch die produktorientierte Auswertung in Aufgabe 1 maximal 2 Handlungen zu einer Handlung abgebildet werden. Zusätzlich konnten keine Handlungen zur *Klärung der Fragestellung* und nur eine Handlung zum *Aufbau* abgebildet werden. Auch die *Planung* konnte nur bei 5 von 11 Probanden identifiziert werden. Die Handlungen zu den anderen Fähigkeiten bzw. Fertigkeiten sind abgesehen von der geringeren Anzahl ähnlich verteilt wie bei der Referenzauswertung. Ein vergleichbares Bild ergibt sich für Aufgabe 3. Aufgabe 2 unterscheidet sich nur bei der Handlung *Experiment planen*, mit identifizierten Handlungen von 9 Probanden. Die Verteilung der Handlungen in Aufgabe 4 ist ähnlich zur ersten Aufgabe. Es fehlen jedoch mehr identifizierte Handlungen im Bereich *Messen und Beobachten* sowie *Daten analysieren*. Dafür zeigen 4 Probanden Handlungen zur *Klärung der Fragestellung*. Die Aufgaben 5 und 6 unterscheiden bzgl. der identifizierten Handlungen deutlich von den anderen vier Aufgaben. Insgesamt wurden deutlich weniger Handlungen identifiziert. Vor allem die *Klärung der theoretischen Grundlagen* konnte fast komplett nicht identifiziert werden.

### **Prozessorientiert**

Ein Vergleich der identifizierten Handlungen zu Aufgabe 1 (vgl. Abb. 25) zeigt, dass durch das prozessorientierte Auswerteverfahren vergleichbare Datengeneriert werden konnten. Einzig zu der Fähigkeit *Planung* lassen sich leichte Unterschiede zwischen den Auswerteverfahren erkennen. Beim prozessorientierten Verfahren konnten weniger Handlungen zu dieser Fähigkeit identifiziert werden als beim Referenzauswertungsverfahren. Auch bei den anderen Aufgaben ergibt sich eine gute Übereinstimmung zur Verteilung der identifizierten Handlungen. In Aufgabe 2 werden etwas mehr Handlungen abgebildet. Bei den Aufgaben 3, 4 und 6 kann durch das prozessorientierte Auswerteverfahren die *Planung* nur schlecht abgebildet werden. Insgesamt zeigt sich, dass beim prozessorientierten Auswerteverfahren häufiger die Handlung *Beobachten und Messen* und seltener die Handlung *Versuch fehlerfrei aufbauen* identifiziert wurde.

Es lässt sich zusammenfassen, dass beim produktorientierten Auswerteverfahren deutlich weniger Handlungen identifiziert werden. Dies ist erwartungskonform, da nur Handlungen identifiziert werden können, zu denen Aufzeichnungen gemacht werden, während bei den anderen

Auswerteverfahren auch Handlungen ohne Aufzeichnungen identifiziert werden können. Die Probanden machen Aufzeichnungen, wenn sie sich einigermaßen sicher sind, dass sie das Richtige tun. Entsprechend gibt es zu der schwierigen Aufgabe 5 fast keine Aufzeichnungen, obwohl die Probanden wie aus der Referenz deutlich hervorgeht trotzdem viele Handlungen durchgeführt haben. Allgemein zeigen sich bei der produktorientierten Auswertung Probleme die Fähigkeiten bzw. Fertigkeiten *Fragestellung*, *Planung* und *Versuch fehlerfrei aufbauen* adäquat abzubilden. Hypothese 5a zur Abbildung der Handlungen kann für das produktorientierte Auswerteverfahren nicht bestätigt werden.

Das prozessorientierte Auswerteverfahren ist in der Lage, eine ähnliche Anzahl und eine ähnliche Verteilung von Handlungen abzubilden wie das Referenzverfahren. Dies legt den Schluss nahe, dass sich dieselben Handlungen abbilden lassen. Allerdings zeigen sich teilweise Probleme die Fähigkeiten bzw. Fertigkeiten *Planung*, *Versuch fehlerfrei aufbauen* und *Messen und Beobachten* adäquat voneinander zu trennen. Entsprechend gibt es bzgl. der Handlungsverteilungen im Vergleich zur Referenzauswertung an diesen Stellen Unterschiede. Hypothese 5a kann bzgl. des prozessorientierten Auswerteverfahrens mit der Einschränkung bei der Trennbarkeit von *Planung*, *Versuch fehlerfrei aufbauen* und *Messen und Beobachten* bestätigt werden.

Zu Hypothese 6a sollte die Übereinstimmung der Testleistung des produkt- bzw. prozessorientierten Auswerteverfahrens zur Referenz überprüft werden. Die Ergebnisse zu Reliabilität zeigen, dass die verschiedenen Aufgaben bei den Aspekten der Zielorientiertheit und der Strukturiertheit basierend auf dem produkt- und prozessorientierten Auswerteverfahren keine ausreichende interne Konsistenz aufweisen. Daraus folgt, dass die einzelnen Aufgaben jeweils verschiedene Konstrukte abbilden. Geplant war ein Vergleich des einheitlichen Maßes Experimentierkompetenz basierend auf der produkt- bzw. prozessorientierten Auswertung mit der Referenzauswertung. Diese einheitlichen Maße konnten, wie die interne Konsistenz zeigt, nicht generiert werden. Ein Vergleich zur Referenz ist daher wenig aussagekräftig. Entsprechend wurde nur der reliable Aspekt der Richtigkeit und das Gesamtmaß der Experimentierkompetenz mit der Referenz verglichen. Die Korrelation findet sich in Tabelle 23 bzw. 24.

**Tabelle 23: Zusammenhänge bzgl. des Aspekts der Richtigkeit.**

	Produktorientiert	Prozessorientiert	Referenz
Produktorientiert	1.00	.92***	.60
Prozessorientiert		1.00	.42
Referenz			1.00

**Tabelle 24: Zusammenhänge bzgl. des Gesamtmaßes der Experimentierkompetenz.**

	Produktorientiert	Prozessorientiert	Referenz
Produktorientiert	1.00	.47	.44
Prozessorientiert		1.00	.16
Referenz			1.00

Die Zusammenhänge bzgl. der Referenz sind wegen der geringen Stichprobe nicht signifikant. Bei 11 Probanden ist nach G-Power erst ab Effekten von .733 mit einem signifikanten Effekt zu rechnen. Damit liefern die Daten Hinweise darauf, dass die Hypothese bzgl. des Aspekts der Richtigkeit sinnvoll erscheint. Für die anderen beiden Aspekte kann keine Aussage getroffen werden. Bemerkenswert ist der große Zusammenhang der Richtigkeit zwischen produkt- und prozessorientierter Auswertung. Dieser lässt sich durch die Bewertung anhand der Aufzeichnungen erklären. In der Referenzauswertung können darüber hinaus viele Handlungen ohne Aufzeichnungen bzgl. der Richtigkeit bewertet werden. Diese, in den anderen Auswertungen fehlenden Bewertungseinheiten, führen zu einer schlechteren Passung. Eine Aussage bzgl. des Gesamtmaßes von Experimentierkompetenz lässt sich mit der vorliegenden Stichprobe nicht treffen.

## **8.5 Diskussion**

Mit dieser Studie sollten die qualitativen Aspekte der Testgüte für den in Kapitel 7 entwickelten Test untersucht werden. Nachfolgend wird basierend auf den Erkenntnissen von Studie II diskutiert, inwieweit die entwickelten Aufgaben des Testinstruments (vgl. Kapitel 7.2) und die entwickelten Auswerteverfahren (vgl. Kapitel 7.3) für eine Datenerhebung mit ausreichender Güte geeignet sind.

## Qualität der Aufgaben

Die Erkenntnisse der Studie belegen, dass durch die entwickelten Aufgaben sämtliche Handlungen des Modells der Experimentierkompetenz von den Probanden genutzt werden. Bis auf wenige Ausnahmen konnten mit der Referenzauswertung alle Handlungen beim Bearbeiten jeder Aufgabe bei jedem einzelnen Probanden beobachtet werden. Damit kann Forschungsfrage 3 (*Nutzen Probanden alle Fähigkeiten und Fertigkeiten der Experimentierkompetenz bei der Bearbeitung jeder Aufgabe?*) im Sinne der Testentwicklung positiv beantwortet werden. Gleichzeitig nutzen Probanden nur in einem sehr geringen Maß Konstrukt-fremde Handlungen. Weniger als 0,15% der Handlungen wurden anderen Konstrukten zugeordnet. Damit kann Hypothese 4a (*Die Anzahl der beobachteten Handlungen, die sich keiner Fähigkeit bzw. Fertigkeit der Experimentierkompetenz zuordnen lassen, sind vernachlässigbar.*) bestätigt werden.

Die Analyse der Aufgaben offenbarte dennoch Probleme bzgl. einzelner Aufgaben. Es zeigten sich Probleme bei der Reliabilität (Forschungsfrage 2: *Inwieweit können die verschiedenen Aspekte der Experimentierkompetenz durch die sechs Aufgaben reliabel abgebildet werden?*). Aufgabe 1 hängt bzgl. der Strukturiertheit kaum mit den anderen Aufgaben zusammen. Die Aufgabe wurde von den Probanden zuerst bearbeitet und es liegt der Schluss nahe, dass eine Gewöhnungsphase an den Experimentierkoffer und die Testsituation zu einer schlechteren Güte bei dieser Aufgabe geführt haben könnten.

Aufgabe 2 erfasst ein anderes Konstrukt als die übrigen Aufgaben. Die Aufgabe wurde qualitativ analysiert. Die Bearbeitung der Aufgabe erfolgt sehr gradlinig. Aus dem Experimentierprozess der Probanden ließ sich ableiten, dass ihr Vorgehen durchweg klar war. Eine Schwierigkeit der Aufgabe wird durch die Genauigkeit in der Messung bedingt. Kleine Messungenauigkeiten führen zu größeren Abweichungen und damit zu Problemen im Vergleich mit den Referenzgraphen. Dies bereitet nicht nur einzelnen z.B. leistungsschwachen Probanden, sondern allen Probanden Probleme. Sie unterschätzen die Bedeutung der Genauigkeit der Messung. Bedingt durch das klare Vorgehen und die fehlende Differenzierung ist die diagnostizierte Testleistung der Probanden sehr ähnlich. Hypothese 2a (*Die Inter-Item-Übereinstimmung liegt bei einem Wert von Cronbachs  $\alpha > .6$* ) kann für die Aspekte der Richtigkeit, der Zielorientiertheit und für das Gesamtmaß der Experimentierkompetenz bestätigt werden. Für das Qualitätsmaß der Strukturiertheit kann die Hypothese nicht bestätigt werden. Zur Bewertung der Ergebnisse bzgl. dieses Aspekts wurden Experten aus dem Bereich des Experimentierens aus der Fachdidaktik eingeladen und die Modellierung und Operationalisierung wurde anhand der Daten aus Studie II analysiert und diskutiert. Nach Einschätzung der

Experten sind sowohl Modellierung als auch Operationalisierung sinnvoll, logisch und geeignet. Für die Verbesserung dieses Aspekts bietet sich eine Überarbeitung von Aufgabe 2 an.

Neben den konkreten Erkenntnissen zu den spezifischen Hypothesen konnten aus den Bemerkungen der Probanden im Rahmen des lauten Denkens weitere Hinweise gewonnen werden. Es zeigte sich, dass die Aufgabenformulierung von Aufgabe 5 einigen Probanden Schwierigkeiten bereitete. Zudem war das benötigte Fachwissen den meisten Probanden nicht bekannt. Auch mithilfe der theoretischen Hinweise, gelang es ihnen im Rahmen einer halben Stunde nicht ein ausreichendes Verständnis zur Polarisierung zu entwickeln. Daraus resultierte, dass die geplanten bzw. aufgebauten Versuche zu großen Teilen nicht zur Beantwortung der Fragestellung geeignet waren. Auch in Aufgabe 6 ergaben sich Unklarheiten zur Fragestellung, was die Bestimmung konkreter Größen des unbekanntes Bauteils angeht.

### **Qualität der Auswerteverfahren**

Als Grundlage zur Auswertung zeigte sich bzgl. aller drei Auswerteverfahren eine ausreichende Übereinstimmung zwischen den Ratern. Hypothese 1a zur Auswerteeobjektivität kann bestätigt werden. Im Rahmen der Diskussion zu den Aufgaben wurde bereits dargestellt, dass die Reliabilität (Forschungsfrage 2) in Verbindung mit dem Referenz-Auswerteverfahren für alle Aspekte außer der Strukturiertheit gegeben ist. Damit liefern die Ergebnisse der Studie Evidenz dafür, dass mit dem Testinstrument (in Verbindung mit dem Referenz-Auswerteverfahren) Experimentierkompetenz valide erfasst werden kann. Die Bewertung berücksichtigt sämtliche Fähigkeiten bzw. Fertigkeiten nach dem Modell der Experimentierkompetenz und der Einfluss Konstrukt-fremder Fähigkeiten bzw. Fertigkeiten ist zu vernachlässigen (vgl. Forschungsfragen 3 und 4).

Das vorliegende prozessorientierte Auswerteverfahren ist in dieser Form nur bedingt geeignet Experimentierkompetenz adäquat abzubilden. Zwar konnten bis auf eine Trennung von den Fähigkeiten bzw. Fertigkeiten der *Planung* und des *Aufbaus* bzw. des *Aufbaus* und des *Messens* vergleichbare Handlungen identifiziert werden (Hypothese 5a: *Ein prozessorientiertes Auswerteverfahren ist in der Lage alle Fähigkeiten und Fertigkeiten von allen Probanden bei der erfolgreichen Bearbeitung aller Aufgaben abzubilden.*). Allerdings finden sich Probleme bei der internen Konsistenz des Auswerteverfahrens bzw. der Aspekte Strukturiertheit und Zielorientiertheit (Hypothese 2a). Aus einer qualitativen Analyse der Daten lässt sich vermuten, dass Probanden gerade an den Stellen, wo sie ein Problem haben und nicht direkt weiter-

wissen, keine oder nur unklare Handlungen zeigen. Sie springen stark zwischen verschiedenen Handlungen hin und her oder versuchen das Problem im Kopf zu lösen. Bei der Referenzauswertung lassen sich diese schnellen Übergänge identifizieren und die Sequenzen ohne sichtbare Handlungen lassen sich durch die Aussagen des Probanden den jeweiligen Fähigkeiten zuordnen. Dies könnte man als die "relevanten" Passagen zur Ausdifferenzierung bezeichnen. Fehlen diese "relevanten" Passagen mit vielen Handlungsübergängen, so lassen sich in diesem Abschnitt auch die Strukturiertheit und die Zielorientiertheit nicht rekonstruieren. Entsprechend konnte nur für den Aspekt der Richtigkeit ein Zusammenhang zum Referenzauswerteverfahren gezeigt werden (Hypothese 6a: *Die Testleistung basierend auf einer prozessorientierten Auswertung korreliert hoch mit einer Referenzleistung der Experimentierkompetenz.*).

Das hier genutzte produktorientierte Auswerteverfahren ist noch weniger gut geeignet Experimentierkompetenz adäquat abzubilden als das prozessorientierte Auswerteverfahren. Beim produktorientierten Auswerteverfahren können die Handlungen *Fragestellung*, *Planung* und *Aufbau* gar nicht abgebildet werden. Die Probleme zur Fragestellung lassen sich erklären, weil Probanden eine Fragestellung lesen, diese jedoch nicht im Laborheft notieren. Entsprechend gibt es keine Bewertungsgrundlage, um diese Handlung identifizieren zu können. Ähnlich ist es mit dem *Aufbau*. Haben Probanden ein Experiment aufgebaut, fehlt ihnen schlicht die Möglichkeit dieses adäquat und schnell dokumentieren zu können. Zudem denken die Probanden ohne Aufforderung nicht daran ihren Aufbau mit einer Skizze (zeitaufwendig) zu dokumentieren. Die fehlende Dokumentation *Planung* ist möglicherweise auf die zu wenig konkrete Arbeitsanweisung zurück zu führen. Die Planung sollte vor allem durch Skizzen und kurze Beschreibungen dokumentiert werden. Diese haben die Probanden selten angefertigt um Zeit zu sparen. Zur Aufgabenlösung sahen sie wahrscheinlich keine Notwendigkeit ihr Vorgehen nachvollziehbar zu dokumentieren. Zusätzlich zu den fehlenden Aufzeichnungen zu diesen drei Handlungen können ganz allgemein nur Fähigkeiten bzw. Fertigkeiten identifiziert werden, zu denen Aufzeichnungen angefertigt werden (Hypothese 5a: *Ein produktorientiertes Auswerteverfahren ist in der Lage alle Fähigkeiten und Fertigkeiten von allen Probanden bei der erfolgreichen Bearbeitung aller Aufgaben abzubilden.*). Entsprechend werden einige Fähigkeiten bzw. Fertigkeiten nicht abgebildet und es ist problematisch die Strukturiertheit und die Zielorientiertheit aus den lückenhaften Daten adäquat zu rekonstruieren. Dies zeigt sich auch in einer unzureichenden internen Konsistenz bzw. der Aspekte Strukturiertheit

und Zielorientiertheit (Hypothese 2a). Die Passung des Aspekts der Richtigkeit zur Referenz konnte bestätigt werden (Hypothese 6a).

## **8.6 Überarbeitung des Testinstruments**

Das Referenzauswerteverfahren scheint eine valide Erfassung zu ermöglichen. Allerdings zeigten sich noch kleinere Schwierigkeiten bzgl. einiger Aufgaben. Diese sollen im weiteren Verlauf überarbeitet werden. Durch den ökonomischen Aufwand des Referenzauswerteverfahrens ist dieses nicht für den Einsatz im Feld geeignet. Das produkt- bzw. prozessorientierte Auswerteverfahren wurde zum Einsatz im Feld entwickelt. Beide Verfahren konnten die Ansprüche jedoch nicht in ausreichender Weise erfüllen. Aus den in Studie II gewonnen Erkenntnissen konnten sich jedoch Hinweise für die Überarbeitung der Aufgaben ableiten lassen. Diese Erkenntnisse wurden zur Entwicklung eines neuen, ökonomisch effizienten Auswerteverfahrens genutzt. Die dafür notwendigen Schritte werden im Folgenden vorgestellt.

### **Überarbeitung der Aufgaben**

Bei Aufgabe 2 schien den Probanden nicht klar zu sein, dass die Messgenauigkeit eine zentrale Rolle spielt. Deshalb wurden im Aufgabentext Hinweise zur Messgenauigkeit aufgenommen und der Referenzgraph zur Identifikation in den theoretischen Hinweisen aufgearbeitet. Der komplexe Arbeitsauftrag von Aufgabe 5 wurde in zwei Teile zerlegt und zum besseren Verständnis neu formuliert. Zusätzlich wurden zusätzliche Hinweise zur Polarisierung und dem Einsatz von Polarisationsfiltern in die theoretischen Hinweise aufgenommen. Aufgabe 6 wurde bzgl. des Arbeitsauftrages angepasst und somit konkretisiert.

### **Entwicklung eines ökonomisch effizienten Auswerteverfahrens**

Zum Einsatz des Testinstruments in einem größeren Rahmen, wird ein Auswerteverfahren benötigt, welches ökonomisch effizient ist. Von den eingesetzten Auswerteverfahren kann nur das produktorientierte Auswerteverfahren als ökonomisch effizient bezeichnet werden. Allerdings konnte gezeigt werden, dass ein produktorientiertes Auswerteverfahren in der vorliegenden Form nicht geeignet ist um Experimentierkompetenz adäquat abzubilden. Entsprechend wird das Auswerteverfahren bzgl. der gefundenen Probleme überarbeitet.

Die Ergebnisse von Studie II zeigen, dass einige Handlungen durch das produktorientierte Auswerteverfahren nicht identifiziert werden können. Entsprechend muss das Verfahren erweitert werden, um diese Handlungen abbilden zu können. Dazu wurde das Laborheft um einen Protokollierungsbogen erweitert (siehe Abb. 26). Im Bereich der Aufzeichnungen wer-

den äquivalent wie bei der produktorientierten Auswertung alle Aufzeichnungen wie in einem Laborbuch niedergeschrieben. Zusätzlich werden die Probanden aufgefordert Handlungen zu notieren, indem sie diese ankreuzen, Aufzeichnungen machen oder einen Kommentar schreiben. Aus einer qualitativen Analyse der Videodaten ging hervor, dass es sich dabei vor allem um das Lesen oder Verstehen der Fragestellung; die Beschäftigung mit theoretischen Grundlagen; Verifizierungen von Rechnungen oder Abschätzungen sowie die Interaktion mit dem Experimentiermaterial handelt. Entsprechend wurden diese Kategorien aufgenommen. Damit Probanden bei der Bearbeitung die Dokumentation nicht vergessen, werden sie alle 60 Sekunden daran erinnert, sich zu fragen, ob alle Handlungen der letzten Minute eingetragen sind. Mit diesem Protokoll sollte die Identifikation der Handlungen *Klärung der Fragestellung* ermöglicht werden. Zudem sollten auch Handlungen abgebildet werden können, die zu keiner Aufzeichnung führen. Allerdings wird eine Bewertung der Richtigkeit in diesem Falle nicht möglich sein.

Aus den Ergebnissen zur Kohärenz der Auswerteverfahren zeigt sich aber auch, dass die Fähigkeiten bzw. Fertigkeiten *Experiment planen* und *Versuch fehlerfrei aufbauen* nicht adäquat abgebildet werden können. Entsprechend werden die Probanden beim neuen Auswerteverfahren aufgefordert ihre Aufzeichnungen und Dokumentationen so anzulegen, dass jemand anders mit ihren Aufzeichnungen denselben Versuch anfertigen könnte und zu denselben Ergebnissen kommen würde. Dies sollte zu Aufzeichnung zur Fähigkeit *Experiment planen* führen. Damit die Fertigkeit *Versuch fehlerfrei aufbauen* abgebildet werden kann, werden die Probanden aufgefordert ihre Auf- und Umbauten mit einem Foto festzuhalten. Dazu erhalten die Probanden Zahlenkarten, um die Fotos ihren Aufzeichnungen zuzuordnen (siehe Abb. 26). Die Protokollierung und die Aufforderung zur Nachvollziehbarkeit der Dokumentation gibt den Probanden die Möglichkeit "relevante" Passagen wie im Abschnitt zur Qualität der Auswerteverfahren beschrieben zu dokumentieren. Zusätzlich werden Probanden daher dazu aufgefordert, Handlungen falls notwendig mit Stichworten zu dokumentieren.



Aufgabe: 2.1				Seite:
Fragestellung	Theorie	Rechnen	Experimentieren	Aufzeichnungen
X	X		X	Bestimmung der Länge und Breite des Glühwändels
			X	Material für Versuch heraus suchen (Lichtquelle, Linse, Schirm)
			X	Versuchsaufbau
			X	Abbildungsgleichung $\frac{B}{b} = \frac{G}{g}$
			X	Foto (1)
			X	Aufbau: Feinjustierung (Abstände g und b verändert, um ein scharfes Bild zu erzeugen)
			X	Abstände ausmessen mit Hilfe des Maßbandes g = 25cm b = 19,5cm
			X	Abbildung des Glühwändels ausmessen Länge G = 0,8cm Breite d = 0,1cm
		X		G ist gemacht: $\frac{B}{b} = \frac{G}{g} \Rightarrow G = \frac{B}{b} \cdot g$ Länge bestimmen: $G = \frac{0,8}{19,5} \cdot 25 = 1,0256 \text{ cm} = 1,03 \text{ cm}$

Aufgabe: 2.1				Seite:
Fragestellung	Theorie	Rechnen	Experimentieren	Aufzeichnungen
		X		<p>Breite d bestimmen: <math>B = g \cdot d \Rightarrow d = \frac{B}{g} = \frac{0,1 \text{ cm}}{19,5 \text{ cm}} = 0,0051 \text{ cm} = 0,0051 \text{ cm} = 0,051 \text{ mm}</math></p> <p>Der Glühwandel hat eine Länge von 1,03cm und eine Breite von 0,051mm.</p>

Experimentieraufgabe: 2.1				Lösungsseite: 1
Fragestellung	Theorie	Rechnen	Experimentieren	Aufzeichnungen
X	X		X	<p>Foto 1 Versuchsaufbau</p> <p>Skizze für B, b, g und G im Theoriefall</p> <p>Längenberechnung</p> $\frac{B}{b} = \frac{G}{g} \quad B = (11,9 - 10,4) \text{ cm} = 1,5 \text{ cm}$ $L = 32,7 \text{ cm}$ $g = 6,4 \text{ cm}$ $\Rightarrow b = \frac{B \cdot g}{G} \quad b = \left( \frac{1,5 \cdot 6,4}{0,1} \right) \text{ cm} = 96 \text{ cm} = 960 \text{ mm}$ <p>Breiteberechnung</p> $B = 0,2 \text{ cm}$ $b = 32,7 \text{ cm}$ $g = 6,4 \text{ cm}$ $B = \frac{b \cdot g}{G} \quad B = \left( \frac{0,2 \cdot 6,4}{32,7} \right) \text{ cm} = 0,039 \text{ cm} = 0,39 \text{ mm}$

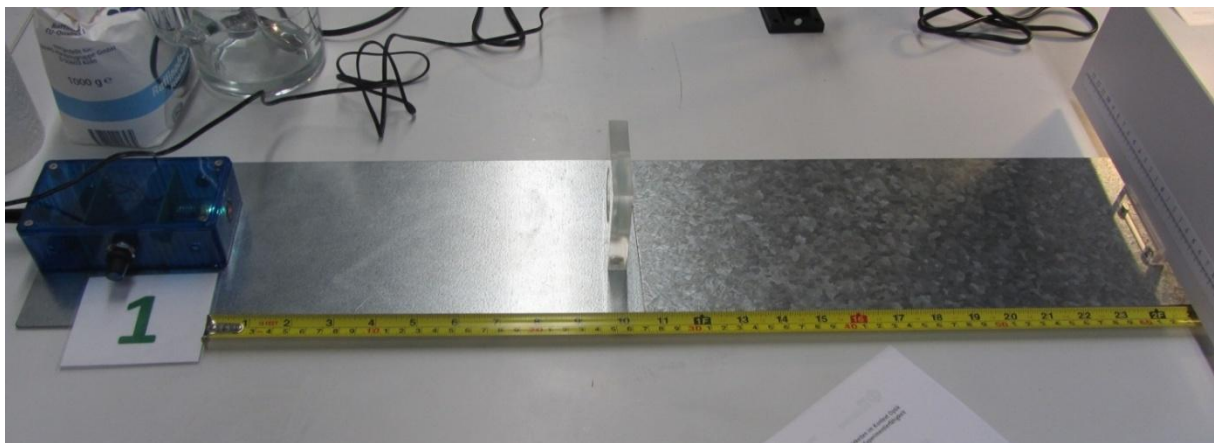


Abbildung 26: Aufzeichnungen im neu entwickelten Auswerteverfahren (oben) mit einem Fotoverweis auf einen Versuchsaufbau (unten).

## **9 Studie III – Quantitative Untersuchung mittels Feldtest**

In Kapitel 7 wurde nach den Anforderungen aus Kapitel 5 und einer Fachinhaltsanalyse (Kapitel 6) ein Testinstrument entwickelt. Dieses wurde aufgrund der Erkenntnisse von qualitativen Untersuchungen in Kapitel 7 überarbeitet. Das Testinstrument soll nach den formulierten Zielen dieser Arbeit (Kapitel 4) das Konstrukt Experimentierkompetenz im universitären Kontext für den Fachinhalt Optik mit einer ausreichenden Güte erfassen. Um herauszufinden, ob dieses Ziel erreicht werden konnte, wurde als zweites Ziel der Arbeit festgelegt, dass die tatsächliche Testgüte detailliert untersucht werden soll. In diesem Kapitel werden die quantitativen Aspekte analysiert.

### **9.1 Forschungsfragen zu den quantitativen Aspekten der Testgüte**

Die Forschungsfragen orientieren sich an den Gütemaßen von Experimentiertests. Nachfolgenden werden die Fragen zur Objektivität, Reliabilität und den verschiedenen Aspekten der Validität entwickelt.

#### **9.1.1 Objektivität**

Die Objektivität wurde als relativ problemloses Gütekriterium beschrieben (Kapitel 3.3.1; Kapitel 5.8). Dennoch stellt eine ausreichende Objektivität eine Grundlage zur Analyse der generierten Daten dar. Zur Kontrolle der Objektivität ergibt sich:

*Forschungsfrage 2: Lässt sich mit den entwickelten Testmanualen eine ausreichende Objektivität erzielen?*

Hypothese 2a: Für die Übereinstimmung der Rater bei der Kodierung der Daten gilt Cohens  $\kappa > .6$ , was bei einer hochinferenten Kodierung als ausreichende Güte der Objektivität angesehen wird.

#### **9.1.2 Reliabilität und Strukturelle Validität**

Die Aspekte der Reliabilität und der strukturellen Validität sind eng miteinander verknüpft (vgl. Kapitel 3). In diesem Teilkapitel werden die beiden Aspekte zusammen thematisiert.

In Kapitel 3.3.3.3 wurde herausgearbeitet, dass in jede Testentwicklung bestimmte theoretische Annahmen über die Struktur des Konstrukts einfließen. Ein Problem bei der strukturellen Validität liegt vor, wenn diese strukturellen theoretischen Annahmen nicht durch die beobachteten Daten bestätigt werden können. Bei der Beschreibung der Anforderungen (Kapitel 5.3) wurde festgestellt, dass sich Experimentierkompetenz durch die drei Indikatoren Richtigkeit,

Strukturiertheit und Zielorientiertheit *formativ operationalisieren* lässt. Dieses formative Messmodell muss sich in den Testdaten wiederfinden. In diesem Fall sollten die Indikatoren Richtigkeit, Strukturiertheit und Zielorientiertheit voneinander disjunkte Teilkonstrukte der Experimentierkompetenz sein.

Bezogen auf die einzelnen Skalen (der Richtigkeit, der Strukturiertheit, der Zielorientiertheit und der Experimentierkompetenz als Gesamtmaß) sollen dasselbe Konstrukt messen (vgl. auch Inter-Item-Korrelation; Reliabilität). Entsprechend muss die interne Konsistenz der einzelnen Skalen hoch sein. In diesem Fall liegt bzgl. der einzelnen Aufgaben *reflektives Messmodell* für die einzelnen Subskalen vor. Das heißt, die Aufgaben bilden zu jedem Teilkonstrukt eine streng eindimensionale Skala.

Für die strukturelle Validität ergibt sich damit:

*Forschungsfrage 8: Inwieweit liegt eine Passung der strukturellen theoretischen Annahmen bzgl. der Modellierung von Experimentierkompetenz und der mit dem Testinstrument generierten Struktur der Daten vor?*

Hypothese 8a: Die interne Konsistenz der Teilkonstrukte Richtigkeit, Strukturiertheit und Zielorientiertheit sowie das Gesamtmaß der Experimentierkompetenz ist ausreichend hoch ( $\alpha > .60$ ) (Reliabilität).

Hypothese 8b: Die testinternen Korrelationen zwischen den Aspekten Richtigkeit, Strukturiertheit und Zielorientiertheit sind maximal gering, was ein formatives Messmodell bestätigt.

### **9.1.3 Externe Validität**

Bei der Prüfung der strukturellen Validität wird auf die Passung der theoretischen Annahmen über die Strukturen des Konstrukts und die in den Daten gefundenen Strukturen fokussiert. In Kapitel 3.3.3.4 wurde erörtert, dass auch theoretisch erwartete Zusammenhänge mit anderen Konstrukten zur Prüfung der Validität herangezogen werden können. Äquivalent zu der strukturellen Validität muss auch bei der externen Validität eine Passung zwischen den externen Strukturen erklärt durch theoretische Annahmen und den Strukturen in den tatsächlich generierten Daten vorliegen. Speziell sollten Konstrukte die ähnlich bzw. teilweise deckungsgleich modelliert wurden, mit Experimentierkompetenz korrelieren. In Kapitel 3 und 4 wurden verschiedene ähnliche Modellierungen von Experimentierkompetenz dargestellt und darauf aufbauend Testinstrumente operationalisiert. Die meisten Auswerteverfahren beruhen vor allem auf dem Aspekt der Richtigkeit (vgl. Kapitel 3.3.3.3). Durch kleinere Unterschiede in der Modellierung und Operationalisierung wird keine vollkommene Übereinstimmung erwar-

tet. Dennoch sollte die Testleistung bzgl. alternativer Bewertungsverfahren der Experimentierkompetenz durch die Testergebnisse des in dieser Arbeit entwickelten Testinstruments vorhergesagt werden können (konvergente Validität). Relevant ist vor allem die Bewertungsgrundlage im Praktikum, also die Praktikumsnote im Bereich Optik. Diese setzt sich üblicherweise aus Prüfungsgesprächen zu bisher durchgeführten Optik-Versuchen zusammen. Es wird vor allem bewertet, ob Studierende die Versuchsdurchführung fachlich richtig wiedergeben können. Diese Note sollte durch die Teilnahme am Experimentiertest vorhersagbar sein.

Im Gegensatz zu verschiedenen, teilweise deckungsgleichen Konstrukten, gibt es auch andere Konstrukte, die begründeten Einfluss beim Experimentieren haben könnten. Als potentielle Konstrukte wurden Fachwissen zur Optik, Selbstwirksamkeit und Strategiewissen identifiziert. Um zu zeigen, dass die Testleistung auf Experimentierkompetenz und keinem der anderen Konstrukte basiert (vgl. Konstrukt-Irrelevante-Varianz, Kapitel 3), sollte aber gerade keine oder höchstens eine geringe Korrelation zwischen diesen drei Konstrukten und der Experimentierkompetenz vorliegen (divergente Validität).

Es ergibt sich:

*Forschungsfrage 9: Inwieweit liegt eine Passung der erwarteten Strukturen durch externe theoretische Annahmen und den tatsächlichen Strukturen in den generierten Testdaten vor?*

Hypothese 9a: Die Testleistung basierend auf gängigen Auswerteverfahren von Experimentiertests lassen sich durch die drei Aspekte der Experimentierkompetenz, vorrangig der Richtigkeit aufklären(konvergente Validität).

Hypothese 9b: Zwischen der Experimentierkompetenz bzw. vorrangig beim Aspekt der Richtigkeit und der Praktikumsnote in Optik findet sich mittlerer bis hoher Zusammenhang(konvergente Validität).

Hypothese 9c: Es findet sich höchstens ein geringer Zusammenhang zwischen dem Fachwissen, der Selbstwirksamkeit bzw. des Strategiewissens und der Experimentierkompetenz(divergente Validität).

## **9.2 Studiendesign**

Die Untersuchung wurde als Feldstudie im physikalischen Anfängerpraktikum für Hauptfachstudierende an der CAU zu Kiel im Sommersemester 2014 durchgeführt. Dazu wurde einer der sechs Versuche zum Thema Optik, wie sie im physikalischen Anfängerpraktikum für Hauptfachstudierende durchzuführen sind, durch den Experimentiertest ersetzt. Die erfolgrei-

che Teilnahme an der Studie war damit elementarer Bestandteil des Anfängerpraktikums und die Inhalte prüfungsrelevant.

Die Testerhebung umfasste eine Sicherheitseinweisung zum Umgang mit Lasern und eine Einführung in den Testablauf (10 Minuten). Anschließend wurden persönliche Daten erhoben (5 Minuten). Als drittes wurde der Experimentiertest durchgeführt (180 Minuten). Als Begleiterhebung zur Beantwortung von Hypothese 9c (*Es findet sich höchstens ein geringer Zusammenhang zwischen dem Fachwissen, der Selbstwirksamkeit bzw. des Strategiewissens und der Experimentierkompetenz (divergente Validität).*) werden ein Fachwissenstest (Selbstentwicklung, 15 Minuten), ein Strategiewissenstest (Thillmann, 15 Minuten; vgl. Thillmann, 2007) und ein Selbstwirksamkeitstest (adaptiert nach Rabe; Rabe, Meinhardt, & Krey, 2012, Rabe, Krey, & Meinhardt, 2013) durchgeführt. Die Probanden führen bei der Bearbeitung ein Laborheft und protokollieren ihre Handlungen (vgl. Kapitel 8.6). Ein Manual zur Testeinführung findet sich im Anhang (Anhang A.4.11). Die Daten werden von zwei Ratern ausgewertet.

Als alternative Auswerteverfahren zur Beantwortung von Hypothese 9a (*Die Testleistung basierend auf gängigen Auswerteverfahren von Experimentiertests lassen sich durch die drei Aspekte der Experimentierkompetenz, vorrangig der Richtigkeit aufklären (konvergente Validität).*) wird das Endergebnis bewertet, es wird eine subjektive Expertenbewertung vorgenommen, es wird eine Checkliste eingesetzt und es wird jeweils ein Globalscore für die Aspekte Richtigkeit, Strukturiertheit und Zielorientiertheit bestimmt (vgl. Anhang A.4.12). Zusätzlich wird bzgl. Hypothese 9b (*Zwischen der Experimentierkompetenz bzw. vorrangig beim Aspekt der Richtigkeit und der Praktikumsnote in Optik findet sich mittlerer bis hoher Zusammenhang (konvergente Validität).*) die Praktikumsnote für die Optik aufgenommen.

Die Untersuchung der aufgeklärten Varianz der Testleistungen basierend auf den alternativen Auswerteverfahren durch die (Teilkonstrukte) der Experimentierkompetenz (Hypothese 9a und 9b) lässt sich üblicherweise mithilfe einer SEM realisieren. Dabei werden die (Teil)Aufgaben im Sinne eines *reflektiven Messmodells* als streng eindimensionale Skala interpretiert. Eine SEM setzt jedoch eine große Stichprobengröße voraus. Diese ist im Rahmen des physikalischen Anfängerpraktikums in Kiel innerhalb eines Semesters nicht zu realisieren. Deshalb wurde als Analyseverfahren eine PLS-SEM gewählt (Hair, 2014; Stender, 2014). Eine PLS-SEM nutzt ein alternatives Schätzverfahren, welches bereits bei kleineren Stichproben zu einer Lösung konvergiert. Für die konvergente Validität bzgl. der Praktikumsnote (Hypothese 9b) und der divergenten Validität (Hypothese 9c) werden die Zusammenhänge zwischen den Begleiterhebungen Fachwissen, Selbstwirksamkeit und Strategiewissen sowie der Praktikumsnote zur Experimentierkompetenz bestimmt.

Zur Beantwortung von Hypothese 8a (*Die interne Konsistenz der Teilkonstrukte Richtigkeit, Strukturiertheit und Zielorientiertheit sowie dem Gesamtmaß der Experimentierkompetenz ist ausreichend hoch ( $\alpha > .60$ ).*) werden für die drei Aspekte der Experimentierkompetenz und das Gesamtmaß jeweils das Cronbachs  $\alpha$  für die interne Konsistenz berechnet. Für Hypothese 8b (*Die testinternen Korrelationen zwischen den Aspekten Richtigkeit, Strukturiertheit und Zielorientiertheit sind maximal gering, was ein formatives Messmodell bestätigt.*) werden die Korrelationen zwischen den Aspekten Richtigkeit, Strukturiertheit und Zielorientiertheit berechnet.

### **9.3 Stichprobe**

An der Erhebung nahmen alle Teilnehmer des Physikalischen Anfängerpraktikums für Hauptfachstudierende der CAU Universität zu Kiel teil, die im Sommersemester 2014 den Themenbereich Optik bearbeiteten. Dies umfasst etwa die Hälfte der Studierenden. Insgesamt nahmen  $N=58$  Studierende teil. Davon waren  $n=35$  männlich und  $n=23$  weiblich. Sie verteilten sich auf die Studiengänge 1-Fach-Bachelor Physik ( $n=6$ ), 2-Fach-Bachelor mit Physik ( $n=8$ ), Physik des Erdsystems ( $n=23$ ) und Materialwissenschaften ( $n=21$ ). Die Probanden waren im Mittel 23,28 Jahre alt und besuchten das 4. Fachsemester ( $n=21$ ), das 6. Fachsemester ( $n=34$ ) bzw. ein höheres Fachsemester ( $n=3$ ).

### **9.4 Instrumente**

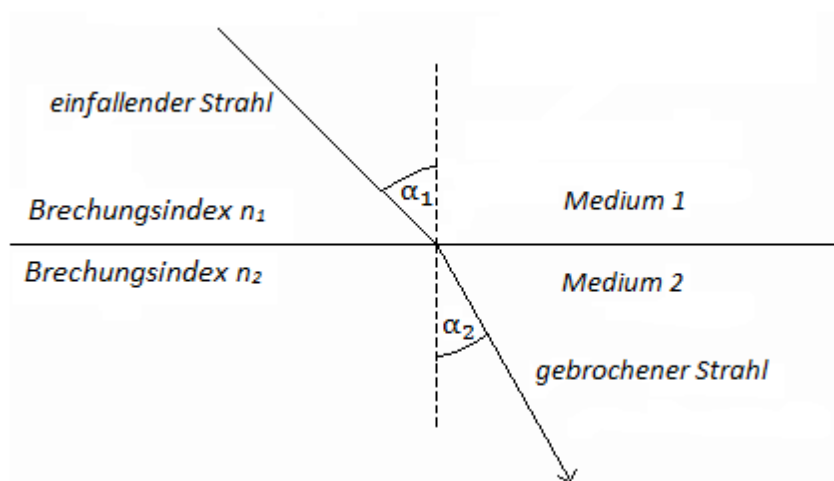
Für die Erfassung von Experimentierkompetenz wurde das in Kapitel 7 entwickelte und in Kapitel 8 überarbeitete Testinstrument genutzt. Die Begleiterhebungen orientieren sich an denen von Emden (Emden, 2011, vgl. auch Klos et al., 2008). Da die Erhebungen im universitären Kontext durchgeführt werden, wurde auf die Unterrichtswahrnehmung verzichtet. Da als Probanden Hauptfachstudierende eines physikalischen Studiengangs gewählt wurden, wird auch das Interesse nicht erhoben. Anders als in einer schulischen Umgebung ist davon auszugehen, dass Studierende die sich für ein Studium der Physik entschieden haben, auch ein relatives hohes Interesse an Physik besitzen.

Um das Fachwissen im Bereich der Optik zu erfassen, wurde in Zusammenarbeit mit dem Projekt Messung professioneller Kompetenzen in mathematischen und naturwissenschaftlichen Lehramtsstudiengängen (KiL) ein Fachwissenstest entwickelt und erprobt (Kleickmann et al., 2014). Dieser basiert auf den identifizierten relevanten Fachinhalten, die auf den Aufgaben des Experimentiertests zugrunde liegen (vgl. Kapitel 6). In der Hauptstudie konnten bei

einer Aufgabenanzahl von acht Aufgaben, eine interne Konsistenz von  $\alpha = .59$  erreicht werden. Zur Erhebung des Fachwissens wurden im Rahmen der vorliegenden Studie 15 Aufgaben im Papier-und-Bleistift-Format eingesetzt. Als Antwortoptionen wurden Multiple-Choice- und offene Aufgaben gewählt. Eine dieser Aufgaben ist in Abbildung 27 dargestellt. Der vollständige Test findet sich im Anhang (A.5).

### Aufgabe 1

Welche der folgenden Formeln gibt das Brechungsgesetz für den in der Abbildung gezeigten Strahlenverlauf korrekt wieder?



$$\frac{\sin \alpha_1}{\sin \alpha_2} = \frac{n_1}{n_2} \quad \square$$

$$\frac{\sin \alpha_1}{\sin \alpha_2} = \frac{n_2}{n_1} \quad \square$$

$$\frac{\cos \alpha_1}{\cos \alpha_2} = \frac{n_1}{n_2} \quad \square$$

$$\frac{\cos \alpha_1}{\cos \alpha_2} = \frac{n_2}{n_1} \quad \square$$

**Abbildung 27: Beispielaufgabe aus dem Fachwissenstest zur Optik.**

Zur Einschätzung der Leistung der Probanden wurde außerdem die Selbstwirksamkeit erhoben. Im Vergleich zum Selbstkonzept wird dabei nicht gefragt, wie gut die Probanden experimentieren können, sondern wie gut sie Aufgaben beim Experimentieren lösen können (Wild & Möller, 2009). Zwischen den beiden Konstrukten wird ein starker Zusammenhang erwartet (Dickhäuser, Schöne, Spinath, & Stiensmeier-Pelster, 2002; Frank, 2014). Zu diesem Zweck

wurde das von Rabe (Rabe et al., 2012; Rabe et al., 2013) entwickelte Testinstrument adaptiert und auf den Kontext des Experimentierens in Laboren angepasst. Das entstandene Testinstrument umfasst 13 Papier-und-Bleistift-Aufgaben. Die Probanden mussten ihre Zustimmung zu Aussagen wie z.B. *Auch Zeitdruck während des Experimentierens hindert mich nicht daran, das Experiment korrekt durchzuführen.* oder *Es bereitet mir keine Probleme zu experimentieren, auch wenn die zugehörige Theorie schwierig ist.* auf einer 4-stufigen Likert-Skala angeben.

Damit bezogen auf die divergente Validität sichergestellt werden kann, dass die Aspekte der Strukturiertheit und vor allem der Zielorientiertheit nicht dem Konstrukt des Wissens über Strategien entspricht, wurde auch zu diesem Thema eine Begleiterhebung durchgeführt. Dazu wurden der Strategiewissenstest von Thillmann (Thillmann, 2007) eingesetzt. Er umfasst sechs Aufgaben mit insgesamt 30 Teilaufgaben. In dem Test werden den Probanden Fragestellungen und mögliche Vorgehensweisen vorgeschlagen. Auf einer 6-stufigen Likert-Skala sollen die Probanden dann bewerten, wie geeignet das jeweilige Vorgehen ist. Das Beispiel einer Aufgabe findet sich in Tabelle 25, der gesamte Test findet sich bei Thillmann (Thillmann, 2007).

**Tabelle 25: Beispielaufgabe zum Strategiewissen.**

<b>1. Sie haben die Aufgabe herauszufinden, ob sich die Raumtemperatur verändert, wenn die Kühlschranktür offen stehen bleibt.</b>						
<i>Ihnen fallen folgende Vorgehensweisen ein, um die Frage zu beantworten. Bewerte die Vorgehensweisen mit Noten von 1 bis 6:</i>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>
a) Ich vergleiche die Temperatur innerhalb des Kühlschranks mit der Temperatur außerhalb des Kühlschranks.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b) Ich messe die Raumtemperatur bei offen stehender Kühlschranktür.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c) Ich messe die Raumtemperatur vor und nach dem Öffnen der Kühlschranktür.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d) Ich messe die Raumtemperatur mehrmals vor dem Öffnen und mehrmals nach dem Öffnen der Kühlschranktür.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e) Ich messe die Raumtemperatur erst dann, wenn Fenster und Zimmertür geschlossen sind.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



## 9.5 Ergebnisse Studie III

### 9.5.1 Objektivität

Bei der Durchführung der Untersuchung wurden die Daten, wie in Kapitel 7.3 und Kapitel 8.6 beschrieben, aus den Protokollbögen mithilfe von Fotos extrahiert. Dazu wurden zwei Rater eingesetzt. In Tabelle 26 ist die Übereinstimmung der Rater bzgl. der Kodierungen aufgetragen.

**Tabelle 26: Beurteilerübereinstimmungen zu Studie III.**

	Identifikation von Handlungen	Bewertung bzgl. der Richtigkeit
Protokollbogen	$\kappa = .691$	$\kappa = .708$

Die Übereinstimmung bei der Auswertung ist sowohl für die Identifikation der Handlungen als auch für die Bewertung bzgl. der Richtigkeit gut. Aufgrund eines sicherheitsbedingten Austausches der verwendeten Laserpointer, konnten die Probanden eines Labortages ( $n=4$  Probanden) keinen identischen Test durchführen und wurden für die Auswertung nicht berücksichtigt. Hypothese 2a (*Für die Übereinstimmung der Rater bei der Kodierung der Daten gilt Cohens  $\kappa > .6$ , was bei einer hochinferenten Kodierung als ausreichende Güte der Objektivität angesehen wird.*) zur Objektivität kann als erfüllt angesehen werden.

### 9.5.2 Reliabilität und Strukturelle Validität

Bei Hypothese 3a und 8a ging es um die interne Konsistenz der Subskalen für Richtigkeit, Strukturiertheit und Zielorientiertheit sowie einem daraus resultierenden Gesamtmaß der Experimentierkompetenz. Zur Analyse wurde wie in Kapitel 3.3.3.2 bzw. Kapitel 3.3.3.6 beschrieben analog zu Kapitel 8.4 Cronbachs  $\alpha$  als Maß für die interne Konsistenz berechnet. Die Daten finden sich in Tabelle 27.

**Tabelle 27: Interne Konsistenz und mittlere Trennschärfe der Aspekte der Experimentierkompetenz.**

	Richtigkeit	Strukturiertheit	Zielorientiertheit	Experimentierkompetenz
Protokollbogen	$\alpha = .60$ $r_{it} = .34$	$\alpha = .54$ $r_{it} = .29$	$\alpha = .73$ $r_{it} = .46$	$\alpha = .67$ $r_{it} = .41$

Die Güte für die interne Konsistenz der Experimentierkompetenz ist etwas schlechter als die Güte von eher authentischen Testinstrumenten (vgl. Kapitel 3.1.4.2,  $\alpha = .73 - .86$ ). Allerdings wurden dabei auch die 4-6 fache Menge an Aufgaben verwendet. Für authentische Testinstrumente liegen bisher keine Werte zur internen Konsistenz von Gesamtskalen vor. Im Vergleich zu Studie II konnte nicht die Qualität der aufwendigen Referenzauswertung erreicht werden. Allerdings liegt die Qualität bzgl. der internen Konsistenz deutlich über der von produkt- und prozessorientierter Auswertung aus Studie II.

Auf Subskalenebene zeigen sich etwas bessere (Klos, 2008; Klos et al., 2008) bzw. deutlich bessere (Emden, 2011) Kennwerte für die Güte bei eher authentischen Testinstrumenten (vgl. Kapitel 3). Im Vergleich zu authentischen Testinstrumenten ist die interne Konsistenz deutlich besser (Emden, 2011). Die Güte der Richtigkeit konnte bezogen auf Studie II beibehalten werden. Die Strukturiertheit konnte deutlich verbessert werden. Allerdings ist die interne Konsistenz mit  $\alpha = .54$  weiterhin kritisch. Bezogen auf die Zielorientiertheit konnte eine vergleichbare interne Konsistenz wie bei der Referenzauswertung aus Studie II erzielt werden. Zusammenfassend lässt sich sagen, dass die interne Konsistenz bzw. Reliabilität des Testinstruments besser als in Studie II ist. Im Vergleich mit bisherigen Arbeiten konnten insgesamt bessere Kennwerte als bei eher authentischen und authentischen Testinstrumenten erreicht werden.

Hypothese 8b war, dass die drei Aspekte der Experimentierkompetenz ein *formatives Messmodell* bilden. Das bedeutet, sie messen verschiedene Aspekte von Experimentierkompetenz. Um diese Hypothese zu untersuchen, wurden die Korrelationen zwischen den drei Aspekten berechnet. Diese finden sich in Tabelle 28.

**Tabelle 28: Korrelationen zwischen den Aspekten der Experimentierkompetenz**

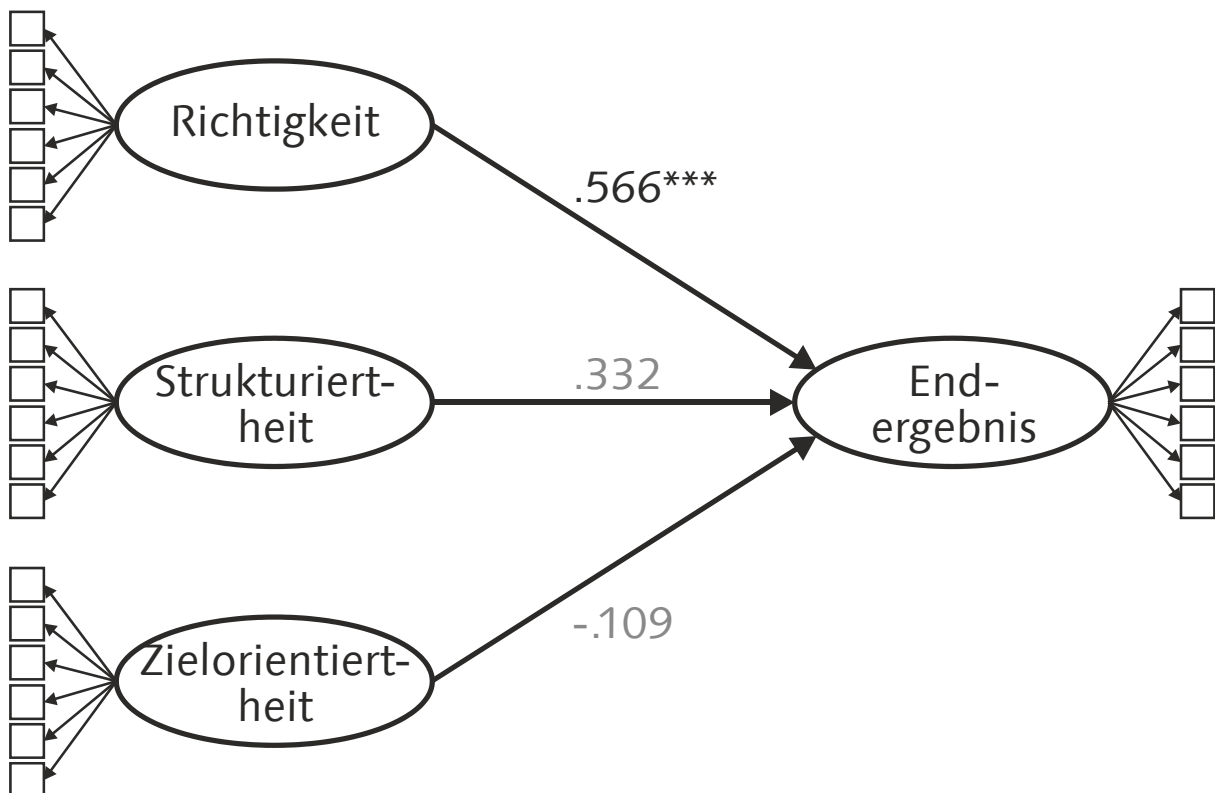
	<b>Richtigkeit</b>	<b>Strukturiertheit</b>	<b>Zielorientiertheit</b>
<b>Richtigkeit</b>		<b>.33*</b>	<b>-.22</b>
<b>Strukturiertheit</b>			<b>.29*</b>
<b>Zielorientiertheit</b>			

Der Aspekt der Strukturiertheit korreliert gering mit den anderen beiden Aspekten. Zwischen den Aspekten der Richtigkeit und der Zielorientiertheit konnte keine signifikante Korrelation gefunden werden. Daraus lässt sich schließen, dass die drei Aspekte des Experimentierens

verschiedene Indikatoren von Experimentierkompetenz sind. Entsprechend kann das formative Messmodell bestätigt werden.

### 9.5.3 Externe Validität

In Hypothese 9a wird darauf fokussiert, inwieweit Testleistungen basierend auf gängigen, alternativen Auswerteverfahren durch Experimentierkompetenz vorhergesagt werden können. Zur Verifizierung dieser Hypothese wird untersucht, ob und in welchem Umfang die Testleistung verschiedener Auswerteverfahren durch die Aspekte der Experimentierkompetenz aufgeklärt werden können. Ein gängiges Strukturgleichungsmodell (SEM) lässt sich wegen der geringen Stichprobengröße nicht nutzen. Entsprechend wurde mit dem Partial Least Square - SEM (PLS-SEM) ein Verfahren genutzt, das auch bei kleineren Stichproben geeignet ist (Hair, 2014; Stender, 2014). Die Modellierung für das alternative Auswerteverfahren, das nur das Endergebnis als Bewertungsgegenstand nutzt ist in Abbildung 28 dargestellt.



**Abbildung 28: Modellierung eines PLS-SEM zur Varianzaufklärung des Endergebnisses aus den Aspekten der Experimentierkompetenz.**

Dargestellt sind die jeweils sechs Aufgaben zu den latenten Variablen Richtigkeit, Strukturiert-heit und Zielorientiert-heit sowie zur latenten Variable des Endergebnisses. Genau wie bei

gängigen SEM werden die Beziehungen zwischen den latenten Variablen durch Pfeile gekennzeichnet. In der Darstellung mit angegeben sind die jeweiligen Signifikanzniveaus. Aus dem Beispiel lässt sich entsprechend ablesen, dass die Testleistung basierend nur auf dem Endergebnis als alternativem Auswerteverfahren signifikant durch den Aspekt der Richtigkeit aufgeklärt werden kann. In dieser Darstellung nicht enthalten ist die aufgeklärte Varianz. Diese liegt für das Endergebnis bei .518.

Auf dieselbe Weise wurden durch Experteneinschätzungen der Aufzeichnungen qualitative Globalmaße für Richtigkeit, Strukturiertheit und Zielorientiertheit sowie der Experimentierkompetenz analysiert. Als letztes Verfahren wurde die Testleistung basierend auf einer Checkliste ermittelt und untersucht. Äquivalent zum Endergebnis wurden die aufgeklärte Varianz und die Vorhersage durch die drei Aspekte Richtigkeit, Strukturiertheit und Zielorientiertheit analysiert. Die jeweiligen Ergebnisse finden sich in Tabelle 29.

**Tabelle 29: Varianzaufklärung von alternativen Auswerteverfahren durch die Aspekte der Experimentierkompetenz.**

	Aufgeklärte Varianz	Vorhergesagt durch
Endergebnis	.518	Richtigkeit b=.566***
Globalscore R	.468	Richtigkeit b=.527***
Globalscore S	.527	Richtigkeit b=.482*** Zielorientiertheit b=.413***
Globalscore Z	.506	Richtigkeit b=.403* Zielorientiertheit b=.503**
Globalmaß Experimentierkompetenz	.701	Richtigkeit b=.816***
Checkliste	.622	Richtigkeit b=.646*** Strukturiertheit b=.249*

Es zeigt sich, dass das Endergebnis zu 52% durch die Richtigkeit aufgeklärt werden kann. Ähnliches gilt auch für das Globalmaß der Experimentierkompetenz (70%) und die Checkliste (62%). Bei der Checkliste wird die Varianz nicht nur von der Richtigkeit sondern zu einem geringen Teil auch von der Strukturiertheit erklärt. Der Globalscore zur Richtigkeit lässt sich zu 47% durch die Richtigkeit aufklären. Das Globalmaß zur Richtigkeit wird erwartungsgemäß durch die Richtigkeit aufgeklärt. Ähnliches gilt für das Globalmaß der Zielorientiertheit,

wobei auch hier die Richtigkeit einen signifikanten Einfluss hat. Das Globalmaß für die Strukturiertheit wird zwar zu 53% aufgeklärt, allerdings nicht durch die Strukturiertheit sondern durch die anderen beiden Aspekte.

Während das Globalmaß für die Strukturiertheit problematisch ist, kann die Hypothese für die anderen alternativen Auswerteverfahren bestätigt werden. Das bedeutet, die Leistung basierend auf anderen Testverfahren bzw. ähnlicher Modellierungen kann durch Experimentierkompetenz bzw. die verschiedenen Aspekte von Experimentierkompetenz vorhergesagt werden. Dieses Erkenntnis liefert Evidenz dafür, dass durch den Experimentiertest tatsächlich Experimentierkompetenz erfasst wurde (konvergente Validität).

Gemäß Hypothese 9b wird ein positiver Zusammenhang zwischen Experimentierkompetenz und der Praktikumsnote zur Optik erwartet. Um diese Hypothese zu untersuchen wurde die Korrelation zwischen den Aspekten der Experimentierkompetenz und der Praktikumsnote zur Optik berechnet. Alternativ hätte eine Regression gerechnet werden können. Diese folgt demselben mathematischen Vorgehen. Einzig die Interpretation unterscheidet sich. Für die Hypothese 9b ist es jedoch nicht relevant, ob die Praktikumsnote durch die Leistung beim Experimentiertest bzw. die Leistung beim Experimentiertest durch die Praktikumsnote vorhergesagt wird oder ob die beiden Konstrukte gemeinsame Varianz teilen. Ein Zusammenhang von .44\*\*\* zwischen Richtigkeit und der Note zeigt, dass, im Sinne einer konvergenten Validität, ein ähnliches Konstrukt vorliegt. Dieses Erkenntnis liefert Evidenz dafür, dass durch das entwickelte Testinstrument tatsächlich Experimentierkompetenz erfasst wird.

Zwischen Note und den anderen beiden Aspekten konnte kein signifikanter Zusammenhang gefunden werden. Dies entspricht den Erwartungen und bestätigt die Hypothese.

Zur divergenten Validität wurde Hypothese 9c formuliert. In dieser wird postuliert, dass die Konstrukte Fachwissen, Strategiewissen und Selbstwirksamkeit ein anderes Konstrukt darstellen als im Experimentiertest gemessen wird. Entsprechend sollten höchstens geringe Korrelationen zwischen den Aspekten der Experimentierkompetenz sowie dem Gesamtwert der Experimentierkompetenz und den Testleistungen für Fachwissen, Strategiewissen und Selbstwirksamkeit vorliegen.

Für den Test zum Fachwissen konnte keine ausreichende interne Konsistenz ( $\alpha = .403$ ) erreicht werden. Dieses Ergebnis kann nicht erklärt werden, da bei dem Einsatz des Instruments in der KiL-Studie bei einer vergleichbaren Population deutlich bessere Kennwerte erreicht

wurden. Bei den Testinstrumenten zum Strategiewissen ( $\alpha = .802$ ) und der Selbstwirksamkeit ( $\alpha = .806$ ) konnte eine gute interne Konsistenz erreicht werden.

Zur Untersuchung der Hypothese wurden die Korrelationen zwischen den drei Konstrukten und der Experimentierkompetenz berechnet. Es gab keine signifikanten Zusammenhänge. Dies gilt sowohl für die drei Aspekte der Experimentierkompetenz als auch für das Gesamtmaß. Daraus lässt sich schließen, dass der Einfluss von Fachwissen, Strategiewissen und dem Selbstkonzept keinen signifikanten Einfluss auf die Testleistung zur Experimentierkompetenz hat.

## 9.6 Diskussion

Die Überarbeitung des Auswerteverfahrens und die Anpassung der Aufgaben (vgl. Kapitel 6.2.5) haben bei einer gleichbleibend ausreichenden Objektivität dazu geführt, dass die interne Konsistenz trotz eines ökonomischen Auswerteverfahrens ausreichend ist. Dies ist bisher bei (eher) authentischen, auch zeitaufwendigen Testinstrumenten zur Experimentierkompetenz nicht gelungen. Dort waren einige oder sogar der Großteil der Subskalen sehr problematisch.

Der Aspekt der Strukturiertheit konnte im Vergleich zur Referenzauswertung verbessert werden, allerdings zeigen sich weiter Probleme mit diesem Aspekt. Dieses Bild ergibt sich bei allen Analysen. Bei der internen Konsistenz sind die Kennwerte an der Grenze zum ausreichenden. Bei Verwendung eines formativen Messmodells finden sich noch Zusammenhänge zu den anderen Konstrukten. Auch wenn diese gering sind, so konnte der Aspekt der Strukturiertheit bisher noch nicht von den anderen beiden entkoppelt werden und auch bei den alternativen Auswerteformaten, klärt die Strukturiertheit keine Varianz im Globalmaß der Strukturiertheit auf. Diese Ergebnisse liefern zwar Hinweise, dass die Strukturiertheit ein Bestandteil der Experimentierkompetenz ist- allerdings sollte die Erfassung der Strukturiertheit beim Experimentieren weiter ausgearbeitet werden.

Bei der Analyse der externen Validität konnten die Hypothesen zur konvergenten Validität (Hypothese 9a: *Die Testleistung basierend auf gängigen Auswerteverfahren von Experimentiertests lassen sich durch die drei Aspekte der Experimentierkompetenz, vorrangig der Richtigkeit aufklären* (konvergente Validität).; Hypothese 9b: *Zwischen der Experimentierkompetenz bzw. vorrangig beim Aspekt der Richtigkeit und der Praktikumsnote in Optik findet sich mittlerer bis hoher Zusammenhang* (konvergente Validität).) und zur divergenten Validität (Hypothese 9c: *Es findet sich höchstens ein geringer Zusammenhang zwischen dem Fachwis-*

sen, der Selbstwirksamkeit bzw. des Strategiewissens und der Experimentierkompetenz (divergente Validität.) bestätigt werden.

Bei der konvergenten Validität konnte die Leistung basierend auf verschiedenen alternativen Auswerteverfahren (vgl. 9.5.3) durch die Aspekte der Experimentierkompetenz zu durchschnittlich 56% aufgeklärt werden. Zur Vorhersage diente vor allem der Aspekt der Richtigkeit. Dies entspricht den Erwartungen, da die Bewertung zumeist auf die Korrektheit von Ergebnissen bzw. Teilschritten zurückzuführen ist. Bei der Checkliste hat aber auch die Strukturiertheit einen Einfluss, der gerade signifikant wird. Dieser ist unerwartet und kann nicht interpretiert werden. Möglicherweise ist er durch die geringe Korrelation zwischen den Gütemaßen entstanden. Als konkreter Indikator galt die Praktikumsnote, auch diese korreliert mit den Maßen der Experimentierkompetenz. Dadurch, dass verschiedene Auswerteverfahren, die verschiedene Aspekte von Experimentierkompetenz erfassen, mit den Ergebnissen des hier entwickelten Testinstruments zusammenhängen, ist es sehr wahrscheinlich, dass auch mit dem vorliegenden Testinstrument das Konstrukt Experimentierkompetenz erfasst wird. Im Sinne einer konvergenten Validierung konnte daher viel Evidenz für eine ausreichende Testgüte generiert werden.

Bei der divergenten Validität ist die Erfassung des Fachwissens (unzureichende interne Konsistenz) und der Selbstwirksamkeit (Deckeneffekt) nicht adäquat gelungen. Dennoch liefern alle Erkenntnisse Hinweise darauf, dass mit dem Testinstrument Experimentierkompetenz und keines der anderen Konstrukte erfasst wurde. Durch diese divergente Validierung wurde Evidenz für die Qualität des Testinstruments generiert.

Insgesamt sind die Erkenntnisse bei der externen Validierung vor allem auf die Richtigkeit bezogen. Dies entspricht zwar den Erwartungen, allerdings fehlen bisher sowohl theoretische Erkenntnisse als auch empirische Daten zur Strukturiertheit und Zielorientiertheit. Um Evidenz zu diesen beiden Aspekten der Experimentierkompetenz generieren zu können, müssten Testinstrumente zum Vergleich der Konstrukte (konvergente Validität) entwickelt werden oder die Struktur innerhalb der Aspekte weiter ausdifferenziert werden.

Zusätzlich zu den eigentlichen Analysen zeigte sich, dass sich die Bewertungseinheiten für die Kodierung durch das eigenständige protokollieren der Probanden stark unterscheiden. Bei vergleichbaren Experimentierprozessen notieren und protokollieren einige Probanden deutlich mehr als andere. Da die Aufzeichnungen und Protokolle die Bewertungseinheiten als Grundlage der Kodierung bilden, stellt dies ein Problem dar. Das bedeutet nämlich, dass die Anzahl

der Protokolleinträge der Probanden einen Einfluss auf die Bewertung hat. Bei der Richtigkeit ist keine Änderung zu erwarten, da die Bewertung abhängig ist von der Richtigkeit der Aufzeichnungen. Wenn es mehr Aufzeichnungen gibt, das Verhältnis von Richtigem und Falschem aber unverändert bleibt, dann ändert sich auch an der Bewertung nichts. Allerdings könnten die Probleme der Strukturiertheit ggf. durch die Unterschiede in den Bewertungseinheiten begründet werden. Denn wenn Probanden Handlungen nicht notieren, so könnte der Prozess möglicherweise nicht adäquat abgebildet werden. Entsprechend können die Handlungsübergänge, die der Bewertung der Strukturiertheit zugrunde liegen, teilweise verfälscht sein. Auch die Zielorientiertheit könnte durch fehlende Protokolleinträge beeinflusst werden. Die Bewertung basiert auf der Anzahl an Forschungszyklen. Diese werden aus den Aufzeichnungen extrahiert. Entsprechend wird einem Probanden, der mehr Handlungen notiert, eher eine höhere Anzahl an Forschungszyklen attestiert. Im Vergleich zur Strukturiertheit ist der Einfluss der Anzahl von Protokolleinträgen für alle Aufgaben konstant, so dass sich keine Hinweise in der internen Konsistenz finden.

Um das skizzierte Problem zu lösen, könnte der Einfluss der Anzahl von Protokolleinträgen auf die Testleistung untersucht oder das Auswerteverfahren weiter optimiert werden. Bei einer Optimierung müssten die Bewertungseinheiten extern vorgegeben werden.



## 10 Fazit

In Kapitel 3 wurde dargestellt, dass sich die Erfassung von Experimentierkompetenz problematisch gestaltet. Ökonomische Testinstrumente können keine inhaltlich validen Daten, nach der Modellierung von Experimentierkompetenz aus Kapitel 2 generieren. Bei authentischen Testinstrumenten fehlt eine ausreichende Passung von theoretisch erwarteten Strukturen und den tatsächlichen Strukturen in generierten Testdaten. Die Diagnostik zur Entwicklung von Experimentierkompetenz z.B. in physikalischen Anfängerpraktika ist mit den vorliegenden Testinstrumenten nicht möglich. Das Ziel dieser Arbeit war es, ein Testinstrument zur Diagnostik von Experimentierkompetenz im universitären Kontext zu entwickeln, mit dem sich Daten generieren lassen, die valide Schlussfolgerungen ermöglichen.

Zur Bearbeitung des Forschungsziels wurden vorliegende Testinstrumente bzgl. möglicher Probleme bei der Validität analysiert. Auf diese Weise konnten diverse Probleme gegliedert nach den Aspekten der Validität nach Messick (Messick, 1995) identifiziert werden (vgl. Kapitel 3). Zu diesen Problemen wurden Lösungsansätze aus vorliegenden Arbeiten (z.B. Stecher et al., 2000) vorgestellt bzw. eigene Ansätze entwickelt. Basierend auf diesen Ansätzen wurden Anforderungen an ein Experimentiertest entwickelt (Kapitel 5), eine Fachinhaltsanalyse durchgeführt (Kapitel 6) und ein Testinstrument entwickelt (Kapitel 7). Mithilfe einer qualitativen und einer quantitativen Studie wurde die Testgüte detailliert analysiert. Im Folgenden werden die Erkenntnisse zur Objektivität, Reliabilität und den sechs Aspekten der Validität aus den Erkenntnissen der drei Studien zusammengefügt und abschließend diskutiert.

### Objektivität

Die Objektivität wurde in Kapitel 3 als *leicht zu realisierendes* Testgütekriterium beschrieben. Diese Einschätzung konnte in allen Studien bestätigt werden. In allen Studien wurde eine ausreichende Objektivität erzielt (vgl. 8.4.1; 9.5.1; Forschungsfrage 2). Damit kann Hypothese 2a (*Für die Übereinstimmung der Rater bei der Kodierung der Daten gilt Cohens  $\kappa > .6$ , was bei einer hochinferenten Kodierung als ausreichende Güte der Objektivität angesehen wird.*) bestätigt werden.

### Reliabilität

Die Reliabilität hängt direkt mit dem Aspekt der internen Konsistenz der Validität zusammen (vgl. Kapitel 3). Für die Referenzauswertung von Studie II (vgl. Kapitel 8.1.2) und für Studie III (vgl. Kapitel 9.1.2) wurde eine ausreichende Reliabilität erzielt. Die interne Konsistenz ist

im Bereich der Strukturiertheit jedoch problematisch. Die Hypothese 3a (*Die Inter-Item-Übereinstimmung liegt bei einem Wert von Cronbachs  $\alpha > .6$ , was als Schwellenwert für die Güte der Reliabilität gilt (vgl. Kapitel 3.3.2.)*) kann nur bzgl. der Richtigkeit, der Zielorientiertheit und des Gesamtmaßes der Experimentierkompetenz bestätigt werden. Eine ausführliche Diskussion findet beim Aspekt der internen Konsistenz statt.

### **Inhaltliche Validität**

Zur Sicherung der inhaltlichen Validität musste geprüft werden, inwieweit notwendiges Wissen und notwendige Fähigkeiten sowie Fertigkeiten zur Testbearbeitung dem Konstrukt der Experimentierkompetenz entsprechen. Das (Fach)Wissen bezieht sich durch die Wahl des Kontextes auf die Optik. Die relevanten Fachinhalte sind die Themen in den physikalischen Anfängerpraktika. Durch die Ergebnisse einer Fachinhaltsanalyse (vgl. Kapitel 6) konnte bestätigt werden, dass die Fachinhalte des Experimentiertests denen von physikalischen Praktika im Bereich Optik entsprechen (vgl. Kapitel 7.2). Damit lässt sich der Schluss ziehen, dass die grundlegenden Fachinhalte eine ausreichende Passung aufweisen. Die relevanten Fähigkeiten und Fertigkeiten werden durch die Modellierung der Experimentierkompetenz (Kapitel 2) vorgegeben. Mit Studie II (Kapitel 7.2) wurde bestätigt, dass die in den Aufgaben genutzten Fähigkeiten und Fertigkeiten mit denen des Modells übereinstimmen.

Durch diese beiden Schritte konnten die Anforderungen bzgl. der inhaltlichen Validität (vgl. 5.1) erreicht werden. Die Erkenntnisse lassen den Schluss zu, dass die generierten Daten inhaltlich valide Schlussfolgerungen zulassen. Forschungsfrage 4 kann damit bestätigt werden.

### **Generalisierbarkeit**

Die Testentwicklung erfolgt gerade so, dass die Ergebnisse möglichst verallgemeinerbar sind. Entsprechend wurde eine Testsituation geschaffen, die authentisch für physikalische Anfängerpraktika ist. Dazu wurde das Format eines Realexperiments mit einem authentischen Versuchsablauf gewählt. Die Passung zum physikalischen Anfängerpraktikum ist insofern gelungen, dass der Test als regulärer Versuch eingesetzt wurde. Dies zeigt eine Akzeptanz des örtlichen Praktikumsleiters und eine Passung zu den Ausbildungsrichtlinien. Da sich der Ablauf der Praktika in den verschiedenen Studiengängen im deutschsprachigen Bereich kaum unterscheidet, ist davon auszugehen, dass diese Passung auch auf den deutschsprachigen Bereich erweitert werden kann. Das entsprechende gilt für die Zielpopulation. Bei der Erprobung des Testinstruments nahmen Hauptfachstudierende der Physik aus dem physikalischen Anfängerpraktikum der CAU zu Kiel statt. Diese sind die Zielpopulation des Testinstruments und we-

gen der Standardisierung wird bzgl. der Population eine Verallgemeinerbarkeit für Hauptfachstudierende der Physik auf den gesamten deutschsprachigen Bereich angenommen.

Bei der Analyse der notwendigen Fachinhalte zeigte sich eine unerwartete Divergenz zwischen den verschiedenen Praktika (vgl. Kapitel 6.5 und 6.6). Daraus folgt der Schluss, dass unterschiedliche fachinhaltliche Schwerpunkte in den verschiedenen Praktika vorliegen. Allerdings konnten relevante Inhaltsbereiche der Optik identifiziert werden. Diese Inhalte besitzen in allen Praktika einen übergeordneten Stellenwert. Darauf aufbauend konnte eine Rangabfolge von relevanten zu weniger relevanten Fachinhalten gebildet werden. Der Test konnte so entwickelt werden, dass die Fachinhalte des Testinstruments diese Rangabfolge der Fachinhalte von Praktika abbilden (vgl. Kapitel 7.2).

Basierend auf der Testentwicklung und den Erkenntnissen kann davon ausgegangen werden, dass eine Generalisierung auf die deutschsprachigen physikalischen Anfängerpraktika zulässig ist.

### **Strukturelle Validität**

In Kapitel 3 wurde als ein zentrales Problem der Validität die interne Konsistenz identifiziert. Dies gilt vor allem für Testinstrumente, die authentisch sind. Das hier entwickelte Testinstrument lässt sich im Spannungsfeld zwischen Ökonomie und Authentizität im Bereich der authentischen Tasteninstrumente einordnen. Dennoch konnten erstmals ausreichende Kennwerte für die interne Konsistenz von authentischen Testinstrumenten, auch für die einzelnen Subskalen nachgewiesen werden (vgl. Kapitel 8.1.2 für die Referenzauswertung; Kapitel 9.1.2). Die Kennwerte reichen jedoch weiterhin nicht an die Güte von typischen Papier- und Bleistift-Tests heran. Zudem ist der Aspekt der Strukturiertheit noch kritisch zu betrachten. Hypothese 8a (*Die interne Konsistenz der Teilkonstrukte Richtigkeit, Strukturiertheit und Zielorientiertheit sowie dem Gesamtmaß der Experimentierkompetenz ist ausreichend hoch ( $\alpha > .60$ ).*) kann mit diesen Befunden teilweise bestätigt werden.

Weitere Strukturen, die in den Testdaten gefunden werden konnten, waren die geringen bzw. nicht vorhandenen Zusammenhänge zwischen den einzelnen Aspekten Richtigkeit, Strukturiertheit und Zielorientiertheit der Experimentierkompetenz. Es kann davon ausgegangen werden, dass die Aspekte, wie modelliert (vgl. Kapitel 2), voneinander unabhängige Teilkonstrukte sind. Das Konstrukt der Experimentierkompetenz wird durch diese drei Indikatoren abgebildet. Hypothese 9a (*Die testinternen Korrelationen zwischen den Aspekten Richtigkeit, Strukturiertheit und Zielorientiertheit sind maximal gering, was ein formatives Messmodell*

*bestätigt.*) kann mit diesen Daten bestätigt werden. Eine Faktorenanalyse zur Bestätigung dieser Erkenntnis konnte bedingt durch die kleine Stichprobe nicht durchgeführt werden.

### **Externe Validität**

Um Evidenz zur externen Validität zu generieren, wurden gängige alternative Auswerteverfahren auf die Aufgaben angewendet. Die dadurch generierten Testleistungen konnten durch die Aspekte der Experimentierkompetenz aufgeklärt werden. Die meisten Auswerteverfahren basieren vor allem auf der richtigen Durchführung. Dies konnte durch eine hohe Varianzaufklärung der Richtigkeit verifiziert werden. Hypothese 9a (*Die Testleistung basierend auf gängigen Auswerteverfahren von Experimentiertests lässt sich durch die drei Aspekte der Experimentierkompetenz, vorrangig der Richtigkeit aufklären (konvergente Validität).*) kann bestätigt werden.

Um die Passung des Konstrukts (*Konstrukt-Unterrepräsentation*) zu bestätigen, wurde die Testleistung mit der Praktikumsnote im Bereich Optik verglichen. Die Note ist kein optimaler Indikator, denn die Notengebung basiert auf der fachlich korrekten Beschreibung der Durchführung einzelner Versuche. Entsprechend sind die Aspekte der Strukturiertheit und Zielorientiertheit nicht notenrelevant. Dies zeigt sich auch in den Ergebnissen. Nur die Richtigkeit korreliert mit der Praktikumsnote im Bereich Optik. Entsprechend dieses Aspekts ist die Passung zum Konstrukt bestätigt, für die anderen beiden Aspekte steht die Passung aus. Die Schwierigkeit ist, dass keine Testinstrumente zur Verfügung stehen, die diese Aspekte abbilden können. Entsprechend kann derzeit kein Vergleich dieser Aspekte durchgeführt werden. Hypothese 9b (*Zwischen der Experimentierkompetenz bzw. vorrangig beim Aspekt der Richtigkeit und der Praktikumsnote in Optik findet sich mittlerer bis hoher Zusammenhang (konvergente Validität).*) kann durch die Ergebnisse von Studie III bestätigt werden.

Die Abgrenzung des Konstrukts der Experimentierkompetenz (*Konstrukt-irrelevante Varianz*) wurde mit einer Untersuchung zur divergenten Validität sichergestellt. Nahe Konstrukte, bei denen ein Einfluss auf die Testleistung erwartet bzw. beobachtet wurde sind Fachwissen, Selbstwirksamkeit und bezogen auf die Aspekte der Strukturiertheit und Zielorientiertheit das Strategiewissen. Die Ergebnisse der Untersuchung konnten zeigen, dass diese Konstrukte keinen signifikanten Einfluss auf die Testleistung haben. Unter der Annahme, dass die beschriebenen Instrumente die verschiedenen Konstrukte adäquat abbilden können, kann die Abgrenzung des Konstrukts bestätigt werden. Allerdings sind sowohl die Erhebung des Fachwissens als auch die Erhebung des Strategiewissens nicht aussagekräftig gelungen. Ent-

sprechend kann hier nur von Tendenzen gesprochen werden und eine Verifizierung der Ergebnisse steht aus. Die Bestätigung von Hypothese 9c (*Es findet sich höchstens ein geringer Zusammenhang zwischen dem Fachwissen, der Selbstwirksamkeit bzw. des Strategiewissens und der Experimentierkompetenz (divergente Validität).*) kann daher nur angenommen werden. Es wurde keine Evidenz erzeugt, die der Hypothese widerspricht und alle Erkenntnisse bzw. Hinweise bestätigen die Hypothese.

### **Prädiktive Validität**

Die prädiktive Validität konnte nicht analysiert werden. Zur Sicherstellung der Vorhersagekraft von späterem Erfolg beim Experimentieren, wurde das Konstrukt basierend auf normativen Setzungen, fachdidaktischen- und berufsspezifischen Anforderungen modelliert. Das Modell umfasst entsprechend genau die Fähigkeiten, Fertigkeiten und Wissensaspekte, die konsensfähig die Qualität beim Experimentieren repräsentieren- und die relevant für spätere Berufe sind. Da die Passung des Modells mit dem Test verifiziert wurde, kann davon ausgegangen werden, dass späterer Erfolg beim Experimentieren durch das Testinstrument vorhergesagt werden kann. Um diese Hypothese zu testen, müsste eine Langzeitstudie durchgeführt werden. Dies war im Rahmen der hier vorliegenden Qualifikationsarbeit nicht möglich.

### **Kognitive Validität**

Im Rahmen der kognitiven Validität wurde geprüft, dass keine Konstrukt-fremden Fähigkeiten oder Fertigkeiten testrelevant sind (vgl. Kapitel 8.4.4). Dies gilt für alle Probanden. Gleichzeitig nutzen auch alle Probanden wirklich alle Fähigkeiten und Fertigkeiten bei der Testbearbeitung (vgl. VI.2). Das bedeutet, dass *Konstrukt-irrelevante Varianz* und *Konstrukt Unterrepräsentation* für alle Probanden der qualitativen Studie ausgeschlossen werden kann. Hypothese 5a (*Die Anzahl der beobachteten Handlungen, die sich keiner Fähigkeit bzw. Fertigkeit der Experimentierkompetenz zuordnen lassen, ist vernachlässigbar.*) kann mit den Erkenntnissen bestätigt werden.

Der Aspekt der kognitiven Validität fokussiert weiterhin darauf, wie adäquat die Auswerteverfahren die kognitiven und auch motorischen Prozesse der Probanden abbilden. Also wie zutreffend die Testleistung die wahre Leistung der Probanden abbilden kann. Um diesen Aspekt zu untersuchen, wurden zunächst drei Auswerteverfahren entwickelt. Bei einem produktorientierten Verfahren, dass die Testleistung aus den Aufzeichnungen der Probanden abbildet, zeigten sich diverse Probleme. Erstens konnten nicht alle Fähigkeiten und Fertigkeiten beim Experimentieren identifiziert werden, obwohl diese stattgefunden haben. Zweitens gibt

es keine Korrelation zwischen den Maßen und einer Referenzauswertung. Entsprechend wird durch das produktorientierte Auswerteverfahren ein anderes Konstrukt abgebildet. Dieselben Ergebnisse zeigen sich für das entwickelte prozessorientierte Auswerteverfahren. Bei der Referenzauswertung, die auf Videoanalyse sowie lautem Denken beruht, konnte die kognitive Validität bestätigt werden. Da dieses Verfahren allerdings zu zeitaufwendig für den Einsatz im Feld ist, wurde basierend auf den Erkenntnissen der qualitativen Studie ein neues Auswerteverfahren entwickelt und basierend auf den Erkenntnissen der quantitativen Studie Konzepte zur Weiterentwicklung vorgeschlagen.

Es fehlt eine qualitative Vergleichsstudie des neu entwickelten Auswerteverfahrens zu einer Referenz, um diese Annahme zu bestätigen.

### **Gesamtgüte**

In Kapitel 3 wurde dargestellt, dass die Testgüte bei der Erfassung von Experimentierkompetenz kritisch ist. Bei ökonomischen Testinstrumenten können einige Aspekte nach der Modellierung von Experimentierkompetenz aus Kapitel 2 nicht abgebildet werden. Bei authentischen Testinstrumenten zeigen sich diverse Schwierigkeiten bei der Passung von theoretisch erwarteten und tatsächlich beobachteten Strukturen.

Bei dem hier entwickelten Testinstrument konnten Objektivität, inhaltliche Validität, kognitive Validität, Generalisierbarkeit und die externe Validität zufriedenstellend nachgewiesen werden. Die prädiktive Validität konnte nicht analysiert werden, da zur Testfairness keine Erkenntnisse zu einer Untersuchung vorlagen und eine Langzeitstudie zum späteren Erfolg im Rahmen der Arbeit nicht realisierbar war. Kritisch sind vor allem die Aspekte der Reliabilität und eng damit verwoben der strukturellen Validität. Die Kennwerte für die Maße der Richtigkeit, Zielorientiertheit und dem Gesamtmaß der Experimentierkompetenz waren ausreichend. Die Subskalen konnten bessere Kennwerte aufweisen, als bei authentischen bzw. eher authentischen Testinstrumenten bisher erreicht wurden. Beim Qualitätsmaß der Strukturiertheit konnte eine ausreichende interne Konsistenz jedoch noch nicht erreicht werden. Dies ist zwar bedingt durch die geringe Aufgabenanzahl, doch im Rahmen eines Experimentiertests wird sich die Anzahl der Aufgaben nicht weiter steigern lassen, um eine bessere Inter-Aufgaben-Korrelation zu erreichen. Daraus folgt, dass der Aspekt der Strukturiertheit in der Modellierung weiter ausdifferenziert werden muss, oder dass die Operationalisierung überarbeitet werden muss. Eine Faktorenanalyse zur Prüfung der 3-faktoriellen Lösung zur Bestätigung der unabhängigen Qualitätsmaße Richtigkeit, Strukturiertheit und Zielorientiertheit konnte aufgrund der geringen Stichprobengrößen nicht durchgeführt werden. Mit dieser Untersuchung

könnten ggf. weitere Erkenntnisse zu den Problemen mit der Strukturiertheit generiert werden.

Ein zweites, nicht zufriedenstellendes Ergebnis, ist die fehlende Prüfung der kognitiven Validität des Auswerteverfahrens basierend auf dem Protokollbogen. Zwar wurden aus den Erkenntnissen von Studie II spezifische Veränderungen vorgenommen. Eine zweite Lautes-Denken-Studie zum Vergleich der Referenzauswertung mit der Protokollierungs-Auswertung zur Prüfung der kognitiven Validität konnte aus Zeitgründen nicht realisiert werden. Zweifel zur Qualität der Protokollierungs-Auswertung ergeben sich auch durch den Einfluss der Anzahl an Protokollierungen auf die Kodierung. Probanden, die mehr Einträge bei einem vergleichbaren Experimentierprozess anfertigen, erhalten eine andere Leistungsdiagnose.

### **Vorschläge zur Weiterentwicklung des Testinstruments (Laborheft-APP)**

Zur weiteren Verbesserung des Testinstruments bzgl. der skizzierten Probleme, wurde ein neues Konzept entwickelt. Dieses fokussiert vor allem auf die Probleme mit der Individualität der Bewertungseinheiten. Dieses kann möglicherweise auch die vorhandenen Probleme der Strukturiertheit erklären. Um dieses Problem zu beheben, wurde eine Überarbeitung des in Kapitel 7.3 bzw. Kapitel 8.6 vorgestellten Auswerteverfahrens vorgeschlagen. Das Ziel dieses Auswerteverfahrens ist es, die Bewertungseinheiten unabhängig von der Anzahl der Protokolleinträge der Probanden zu identifizieren. Ohne starke Strukturierungshilfen, die zu anderen Problemen bei der Erfassung führen würden, lässt sich dies nur durch ein Liverating erreichen. Eine Bewertung in Echtzeit führt wiederum zu einem stark erhöhten Aufwand und wird keine ökonomische Auswertung ermöglichen.

Zur Realisierung eines Liveratings auf der einen Seite und der Reduktion des Testaufwands bietet sich die Implementierung eines computergestützten Liveratings an. Eine Möglichkeit der Realisierung wäre eine App. Die App fungiert als Labor- bzw. Protokollheft und loggt die Handlungen der Probanden automatisch mit. Alles was die Probanden vorher in das Laborheft geschrieben haben, notieren sie nun in der Laborheft-App. Zudem werden der Taschenrechner und der Fotoapparat in die App integriert. Auch die Fragestellungen der einzelnen Aufgaben und theoretischen Hinweise werden ausschließlich über die App verfügbar. Dieses Vorgehen entlastet den Probanden und den Testleiter. Gleichzeitig wird durch den computergestützten Zugriff jede Handlung automatisch festgehalten. Eine vollständige Konzeption der Laborheft-App findet sich im Anhang (A.4.5).

## **Erkenntnisse zur Experimentierkompetenz**

Neben Erkenntnissen zur Validität des Testinstruments wurden einige Erkenntnisse zur Experimentierkompetenz gewonnen. So konnte Evidenz dazu generiert werden, dass Experimentieren ein eigenständiges Konstrukt ist, das sich nicht aus anderen Konstrukten wie z.B. *Fachwissen* und *Problemlösen* zusammensetzen lässt. Das Konstrukt konnte z.B. gegen Fachwissen abgegrenzt werden. Auch die Gliederung von Experimentierkompetenz in die drei Aspekte Richtigkeit, Strukturiertheit und Zielorientiertheit konnte bestätigt werden. Das bedeutet, es gibt nun Evidenz dazu, dass Experimentieren wirklich mehr als nur die Summe einzelner Fähigkeiten und Fertigkeiten ist. Sowohl eine Qualitätssicherung beim Experimentieren als auch eine Steuerung des Prozesses sind unabhängige aber gleichzeitig relevante Aspekte der Experimentierkompetenz.

Aus diesen Erkenntnissen lassen sich Schlüsse für die Lehre von Experimentierkompetenz ziehen. Es ist nicht ausreichend die einzelnen Fähigkeiten und Fertigkeiten zu trainieren, auch die Aspekte der Strukturiertheit (Fehlerkorrektor, Kontrolle, Einschätzung des Experimentierprozesses) müssen Gegenstand der Lehre sein. Gleiches gilt für die Zielorientiertheit. Die bewusste Wahl von Strategien, deren Voraussetzungen und Vorteile müssen thematisiert werden, damit Lernende Experimentierkompetenz erwerben können.



## 11 Literaturverzeichnis

- Baxter, G. P., & Shavelson, R. J. (1994). Science Performance Assessments: Benchmarks and Surrogates. *International journal of educational research*, (21), 279–298.
- Beaton, A. E. (1996). *Science achievement in the middle school years: IEA's third international mathematics and science study (TIMSS)*. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.
- Bergmann, L., Schäfer, C., Kassing, R., Blügel, S., Niedrig, H., & Eichler, H.-J. (2004). *Lehrbuch der Experimentalphysik: Zum Gebrauch bei akademischen Vorlesungen und zum Selbststudium* (10. Aufl.). Berlin [u.a.]: De Gruyter.
- Bortz, J., Bortz-Döring, & Döring, N. (;2010). *Forschungsmethoden und Evaluation: Für Human- und Sozialwissenschaftler ; mit 87 Tabellen* (4th ed.). *Springer-Lehrbuch Bachelor, Master*. Heidelberg: Springer-Medizin-Verl.
- Bybee, R., McCrae, B., & Laurie, R. (2009). PISA 2006: An assessment of scientific literacy. *Journal of Research in Science Teaching*, 46(8), 865–883.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, pp. 37–46.
- Comber, L., & Keeves, J. P. (1973). *Science education in 19 countries: An empirical study. International studies in evaluation: Vol. 1*. New York: Wiley.
- Cronbach, L. J. (1951). Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*, pp. 297–334.
- David Wetzel. *What is Scientific Inquiry?* Retrieved from <http://ezinearticles.com/?What-is-Scientific-Inquiry?&id=1223526>
- Demtröder, W. (2009). *Experimentalphysik 2: Elektrizität und Optik ; mit 19 Tabellen, zahlreichen durchgerechneten Beispielen und 144 Übungsaufgaben mit ausführlichen Lösungen* (5., überarb. und erw. Aufl.). *Springer-Lehrbuch*. Berlin, Heidelberg: Springer.
- Dickhäuser, O., Schöne, C., Spinath, B., & Stiensmeier-Pelster, J. (2002). Die Skalen zum akademischen Selbstkonzept. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 23(4), 393–405. doi:10.1024//0170-1789.23.4.393
- Emden, M., Hübinger, R., & Sumfleth, E. (2010). Erkenntnisgewinnung im naturwissenschaftlichen Anfangsunterricht - Unterrichtsmaterial zur Unterstützung der Kompetenzförderung. *Der mathematische und naturwissenschaftliche Unterricht*, (5/10), 279–286.
- Emden, M. (2011). *Prozessorientierte Leistungsmessung des naturwissenschaftlich-experimentellen Arbeitens: Eine vergleichende Studie zu Diagnoseinstrumenten zu Beginn der Sekundarstufe I*. Univ., Diss.--Duisburg-Essen, 2011. *Studien zum Physik- und Chemielernen: Vol. 118*. Berlin: Logos. Retrieved from [http://deposit.d-nb.de/cgi-bin/dokserv?id=3842951&prov=M&dok\\_var=1&dok\\_ext=htm](http://deposit.d-nb.de/cgi-bin/dokserv?id=3842951&prov=M&dok_var=1&dok_ext=htm)

- Erpenbeck, J. (Ed.). (2007). *Handbuch Kompetenzmessung: Erkennen, verstehen und bewerten von Kompetenzen in der betrieblichen, pädagogischen und psychologischen Praxis* (2., überarb. und erw. Aufl.). Stuttgart: Schäffer-Poeschel.
- Fischer, H. E., & Neumann, K. (2012). Video Analysis as a Tool for Understanding Science Instruction. In J. Dillon & D. Jorde (Eds.), *The World of Science Education. Handbook of Research in Europe* (pp. 115–139). Rotterdam: Sense Publishers.
- Frank, C. (2014). *Arbeitswelt als Kontext: Empirische Grundlagen der Gestaltung berufsorientierender Lehr- und Lernprozesse für Naturwissenschaft und Technik*. Retrieved from [http://www.qucosa.de/recherche/frontdoor/?tx\\_slubopus4frontend\[id\]=urn:nbn:de:bsz:14-qucosa-147216](http://www.qucosa.de/recherche/frontdoor/?tx_slubopus4frontend[id]=urn:nbn:de:bsz:14-qucosa-147216)
- Fraser, B. (1980). Development and validation of a test of enquiry skills. *Journal of Research in Science*, 17(1), 7–16.
- Gesellschaft für Didaktik des Sachunterrichts. (2002). *Perspektivrahmen Sachunterricht*. Bad Heilbrunn/Obb.: Klinkhardt.
- Gott, R., & Duggan, S. (2002). Problems with the Assessment of Performance in Practical Science: Which way now? *Cambridge Journal of Education*, 32(2), 183–201. doi:10.1080/03057640220147540
- Grube & Christiane. (2010). *Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung*. Retrieved from <https://kobra.bibliothek.uni-kassel.de/handle/urn:nbn:de:hebis:34-2011041537247>
- Gut, C., Labudde, P., & Ramseier, E. (2010). Large-scale Experimentiertests: Ansätze zur Analyse von Itemschwierigkeit. In D. Höttecke (Ed.), *Entwicklung naturwissenschaftlichen Denkens zwischen Phänomenen und Systematik. Gesellschaft für Didaktik der Chemie und Physik, [36.] Jahrestagung in Dresden 2009 ; [Gesellschaft für Didaktik der Chemie und Physik, Band 30]* (pp. 245–247). Münster, Westf: LIT-Verl.
- Gut-Glanzmann, C. (2012). *Modellierung und Messung experimenteller Kompetenz: Analyse eines large-scale Experimentiertests*. Univ., Diss.--Basel, 2012. *Studien zum Physik- und Chemielernen: Vol. 134*. Berlin: Logos-Verl.
- Hacker, W. (1986). *Arbeitspsychologie: Psychische Regulation von Arbeitstätigkeit*. Berlin: Deutscher Verlag der Wissenschaft.
- Hair, J. F. (2014). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Los Angeles: Sage Publ.
- Halliday, D. (2009). *Halliday Physik* (2., überarb. und erg. Aufl.). Weinheim: Wiley-VCH.
- Hammann, M., & Hoi Phan, T. B. H. (2007). Experimentieren als Problemlösen: Lässt sich das SDDS-Modell nutzen, um unterschiedliche Dimensionen beim Experimentieren zu messen? *Zeitschrift für Erziehungswissenschaft*, (Sonderheft 8), S33-.
- Hammann, M., Phan, T. T. H., Ehmer, M., & Grimm, T. (2008). Assessing pupils' skills in experimentation. *Journal of Biological Education*, 42(2), 66–72. doi:10.1080/00219266.2008.9656113

- Hanauer, D. I., Hatfull, G. F., & Jacobs-Sera, D. (2009). *Active assessment: Assessing scientific inquiry. Mentoring in academia and industry: Vol. 2*. New York, London: Springer.
- Harmon, M. (1997). *Performance assessment in IEA's third international mathematics and science study (TIMSS)*. Chestnut Hill, MA: TIMSS International Study Center.
- Heinicke, S., Rieß, F., & Heering, P. (2012). *Aus Fehlern wird man klug: Eine genetisch-didaktische Rekonstruktion des "Messfehlers"*. Univ., Diss.--Oldenburg, 2011. Berlin: Logos-Verl.
- Hodson, D. (1996). Laboratory work as scientific method: three decades of confusion and distortion. *Journal of Curriculum Studies*, 28(2), 115–135.  
doi:10.1080/0022027980280201
- Hofstein, A., & Lunetta, V. N.
- Hofstein, A., & Lunetta, V. N. (2004). The laboratory in science education: Foundations for the twenty-first century. *Science Education*, 88(1), 28–54. doi:10.1002/sc.10106
- IEA. (1988). *Science achievement in seventeen countries: A preliminary report* (1st ed.). Oxford, England, New York: Pergamon Press.
- Janík, T. (2009). *The Power of Video Studies in Investigating Teaching and Learning in the Classroom*. (Seidel, T., Ed.). s.l.: Waxmann Verlag GmbH. Retrieved from [http://ebooks.ciando.com/book/index.cfm/bok\\_id/40151](http://ebooks.ciando.com/book/index.cfm/bok_id/40151)
- Kane, M. T. (2001). Current Concerns in Validity Theory. *Journal of Educational Measurement*, 38(4), 319–342. doi:10.1111/j.1745-3984.2001.tb01130.x
- Kempa, R. (1986). *Assessment in science* (Reprint). *Cambridge science education series*. Cambridge: Cambridge University Press.
- Kipnis, M., & Hofstein, A. (2008). The Inquiry Laboratory as a Source for Development of Metacognitive Skills. *International Journal of Science and Mathematics Education*, 6(3), 601–627. doi:10.1007/s10763-007-9066-y
- Klahr, D., & Dunbar, K. (2000). *Exploring science: The cognition and development of discovery processes*. Cambridge, Mass.: MIT Press. Retrieved from <http://www.gbv.de/dms/bs/toc/269299742.pdf>
- Klahr, D., Triona, L. M., & Williams, C. Hands on what? The relative effectiveness of physical versus virtual materials in an engineering design project by middle school children. *Journal of Research in Science Teaching*, pp. 183–203.
- Kleickmann, T., Großschedl, J., Harms, U., Heinze, A., Herzog, S., Hohenstein, F., . . . Zimmermann, F. (2014). Professionswissen von Lehramtsstudierenden der mathematisch-naturwissenschaftlichen Fächer - Testentwicklung im Rahmen des Projekts KiL. *Unterrichtswissenschaft*, pp. 280–288.
- Klieme, E., Funke, J., Leutner, D., Reimann, P., & Wirth, J. (2001). Problemlösen als fächerübergreifende Kompetenz: Konzeption und erste Resultate aus einer Schulleistungsstudie. *Zeitschrift für Pädagogik*, 47(2), 179–200.

- Klos, S., Henke, C., Kieren, C., Walpuski, M., & Sumfleth, E. (2008). Naturwissenschaftliches Experimentieren und chemisches Fachwissen - zwei verschiedene Kompetenzen. *Zeitschrift für Pädagogik*, 304–321.
- Klos, S. (2008). *Kompetenzförderung im naturwissenschaftlichen Anfangsunterricht - der Einfluss eines integrierten Unterrichtskonzepts. Studien zum Physik- und Chemielernen: Vol. 89*. Berlin: Logos Verl.
- Labudde, P., Metzger, S., & Gut, C. (2009). Bildungsstandards: Validierung des Schweizer Kompetenzmodells. In D. Höttecke (Ed.), *Gesellschaft für Didaktik der Chemie und Physik: Bd. 29. Chemie- und Physikdidaktik für die Lehramtsausbildung*. In Schwäbisch Gmünd 2008 (pp. 307–317). Berlin, Münster: Lit.
- Lunetta, V. N. (1998). The School Science Laboratory: Historical Perspectives and Contexts for Contemporary Teaching. *International Handbook of Science Education*, pp. 249–262.
- Lunetta, V. N., Hofstein, A., & Clough, M. (2007). Learning and teaching in the school science laboratory: an analysis of research, theory, and practice. In N. Lederman & S. Abel (Eds.), *Handbook of research on science education*. New York: Lawrence Erlbaum.
- Maiseyenka, V., Nawrath, D., & Schecker, H. (2011). Modellbasierte Diagnose und Förderung experimenteller Kompetenz. In D. Höttecke (Ed.), *Naturwissenschaftliche Bildung als Beitrag zur Gestaltung partizipativer Demokratie. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Potsdam 2010 ; [Gesellschaft für Didaktik der Chemie und Physik, Band 31]*. Münster, Westf: LIT-Verl.
- Martin, M. O. (2000). *TIMSS 1999: International science report : findings from IEA's repeat of the Third International Mathematics and Science Study at the eighth grade*. Chestnut Hill, MA: International Study Center, Boston College, Lynch School of Education.
- Martin, M. O. (2004). *TIMSS 2003: International science report : findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- Martin, M. O. (2008). *TIMSS 2007 international science report: Findings from IEA's trends in international mathematics and science study at the fourth and eighth grades*. Chestnut Hill, Mass.: TIMSS & PIRLS International Study Center Lynch School of Education Boston College.
- Martin, M. O. (2012). *TIMSS 2011 international results in science*. Boston: TIMSS & PIRLS International Study Center.
- Mayer, J. Erkenntnisgewinnung als wissenschaftliches Problemlösen.
- Mayer, J., Bauer, A., & Kattmann, U. (2003). *Erkenntnisgewinnung als wissenschaftliches Problemlösen /// Entwicklung von Wissen und Kompetenzen im Biologieunterricht: Internationale Tagung der Sektion Biologiedidaktik im VDBiol, berlin, 14. bis 19. September 2003*. Kiel: IPN.

- Mayer, J., Grube, C., & Möller, A. (2008). Kompetenzmodell naturwissenschaftlicher Erkenntnisgewinnung. In U. Harms, A. Sandmann, R. Klee, & U. Harms (Eds.), *Forschungen zur Fachdidaktik: Vol. 10. Ausbildung und Professionalisierung von Lehrkräften. Internationale Tagung der Fachsektion Didaktik der Biologie im VBiO, Essen 2007* (pp. 63–79). Innsbruck: StudienVerl.
- Mayring, P. (2008). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (10th ed.). Pädagogik. Weinheim, Basel: Beltz.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement (3rd edition)* (pp. 13–103). New York: Macmillan.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performance as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749.
- Ministeriums für Bildung und Wissenschaft. (2014). schulgesetz2014. Retrieved from <http://www.schulrecht-sh.de/schulgesetz/schulgesetz2014.pdf>
- Mullis, I. V. S. (2007). *TIMSS 2007 assessment frameworks*. Chestnut Hill, Mass.: TIMSS & PIRLS International Study Center Lynch School of Education Boston College.
- Mullis, I. V. S., & Martin, M. O. (2013). *TIMSS 2015 assessment frameworks* (1st ed.). Chestnut Hill, MA: International Study Center, Lynch School of Education, Boston College.
- Murphy, P. (1996). The IEA Assessment of Science Achievement. *Assessment in Education: Principles, Policy & Practice*, 3(2), 213–232. doi:10.1080/0969594960030207
- National Assessment Governing board U.S. Department of Education. (2008). *Science Framework for the 2009 National Assessment of Educational Progress*. Washington, DC: U.S. Government Printing Office.
- National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas* (Prepublication copy.). Washington, D.C: National Academies Press.
- National Science Education Standards: Observe, interact, change, learn* (4th ed.). (1996). Washington, DC: National Academy Press.
- Nawrath, D., Maiseyenko, V., & Schecker, H. (2011). Experimentierkompetenz - modellbasierte Förderung und Diagnose. *Praxis der Naturwissenschaften - Physik in der Schule*, pp. 42–49.
- Neumann, K. (2004). *Didaktische Rekonstruktion eines physikalischen Praktikums für Physiker*. Pädag. Hochsch., Diss.--Heidelberg, 2004. *Studien zum Physiklernen: Vol. 38*. Berlin: Logos-Verl.
- Nickolaus, R., Retelsdorf, J., Winther, E., & Köller, O. (2013). *Mathematisch-naturwissenschaftliche Kompetenzen in der beruflichen Erstausbildung: Stand der Forschung und Desiderata*. *Zeitschrift für Berufs- und Wirtschaftspädagogik: Vol. 26*. Stuttgart: Franz Steiner Verlag. Retrieved from <http://gbv.ebib.com/patron/FullRecord.aspx?p=1660153>

- Oesterreich, R. (1981). *Handlungsregulation und Kontrolle. U&S-Psychologie Forschung*. München: Urban & Schwarzenberg.
- Olson, J. F., Martin, M. O., Mullis, I. V., & Arora, A. (op. 2008). *TIMSS 2007 technical report*. Boston, MA: IEA TIMSS & PIRLS.
- Organisation for Economic Co-operation and Development. (2007). *Analysis. PISA 2006, science competencies for tomorrow's world: Vol. 1*. Paris: OECD. Retrieved from <http://www.nbbmuseum.be/doc/seminar2010/nl/bibliografie/opleiding/analysis.pdf>
- Prenzel, M., Artelt, C., Baumert, J., Blum, W., Hammann, M., Klieme, E., & Pekrun, R. (2007). *PISA 2006: Die Ergebnisse der dritten internationalen Vergleichsstudie*. Zusammenfassung. Retrieved from [http://www.ipn.uni-kiel.de/pisa/zusammenfassung\\_PISA2006.pdf](http://www.ipn.uni-kiel.de/pisa/zusammenfassung_PISA2006.pdf)
- Rabe, T., Krey, O., & Meinhardt, C. (2013). Physikdidaktische Selbstwirksamkeitserwartungen zukünftiger Physiklehrkräfte I. In S. Bernholt (Ed.), *Gesellschaft für Didaktik der Chemie und Physik: Vol. 33. Inquiry-based Learning - Forschendes Lernen. Jahrestagung in Hannover 2012* (pp. 635–637). Kiel: IPN.
- Rabe, T., Meinhardt, C., & Krey, O. (2012). Entwicklung eines Instruments zur Erhebung von Selbstwirksamkeitserwartungen in physikdidaktischen Handlungsfeldern. *Zeitschrift für Didaktik der Naturwissenschaften*, pp. 293–315.
- Rauner, F., & Spöttl, G. (2002). *Der Kfz-Mechatroniker: Vom Neuling zum Experten. Berufsbildung, Arbeit und Innovation: Vol. 12*. Bielefeld: Bertelsmann.
- Schecker, H. (2001). TIMSS - Konsequenzen für den Physikunterricht. In BMBF (Ed.), *TIMSS - Impulse für Schule und Unterricht* (pp. 85–97). München: Mediahaus Biering Grafischer Betrieb.
- Schenk, W., & Kremer, F. (2011). *Physikalisches Praktikum: Mit 100 Versuchen* (13., überarb. Aufl). *Studium*. Wiesbaden: Vieweg + Teubner.
- Schreiber, N., Theyßen, H., & Schecker, H. (2009). Experimentelle Kompetenz messen?! *Physik und Didaktik in Schule und Hochschule*, 8(3), 92–101. Retrieved from [http://www.idn.uni-bremen.de/pubs/DD\\_56\\_NicoSchreiber.pdf](http://www.idn.uni-bremen.de/pubs/DD_56_NicoSchreiber.pdf)
- Schreiber, N., Theyßen, H., & Schecker, H. (2010). Vom Versuch experimentelle Kompetenz zu messen... In D. Höttecke (Ed.), *Entwicklung naturwissenschaftlichen Denkens zwischen Phänomenen und Systematik. Gesellschaft für Didaktik der Chemie und Physik, [36.] Jahrestagung in Dresden 2009 ; [Gesellschaft für Didaktik der Chemie und Physik, Band 30]* (pp. 203–205). Münster, Westf: LIT-Verl.
- Schreiber, N., Theyßen, H., & Schecker, H. (2011a). Auswertungsmethodik im Projekt Diagnostik experimenteller Kompetenz. *PhyDid B - Didaktik der Physik - Beiträge zur DPG Frühjahrstagung* (2),

- Schreiber, N., Theyßen, H., & Schecker, H. (2011b). Diagnostik experimenteller Kompetenz: Ergebnisse von Pilotstudien. In D. Höttecke (Ed.), *Naturwissenschaftliche Bildung als Beitrag zur Gestaltung partizipativer Demokratie. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Potsdam 2010 ; [Gesellschaft für Didaktik der Chemie und Physik, Band 31]* (pp. 224–246). Münster, Westf: LIT-Verl.
- Schreiber, N. (2012). *Diagnostik experimenteller Kompetenz: Validierung technologiegestützter Testverfahren im Rahmen eines Kompetenzstrukturmodells*. Essen, Univ., Diss.--Duisburg, 2012. *Studien zum Physik- und Chemielernen: Vol. 139*. Berlin: Logos-Verl.
- Schreiber, N., Theyßen, H., & Schecker, H. Diagnostik experimenteller Kompetenz: Kann man Realexperimente durch Simulationen ersetzen? *Zeitschrift für Didaktik der Naturwissenschaften*,
- Schumacher, D., & Planinšič, G. (2007). Student undergraduate laboratory and project work. *European Journal of Physics*, 28(3). doi:10.1088/0143-0807/28/3/E01
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK]. (2005a). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss*. München: Luchterhand.
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK]. (2005b). *Bildungsstandards im Fach Chemie für den Mittleren Schulabschluss*. München: Luchterhand.
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK]. (2005c). *Bildungsstandards im Fach Physik für den Mittleren Schulabschluss*. München: Luchterhand.
- Shavelson, R. J., Baxter, G. P., & Pine, J. (1991). Performance Assessment in Science. *Applied measurement in education*, (4), 347–362.
- Shavelson, R. J. & Ruiz-Primo, M. A. (1998). *On the assessment of science achievement: CSE Technical Report 491*. Retrieved from <http://www.cse.ucla.edu/products/Reports/TECH491.pdf>
- Shavelson, R. J., Solano-Flores, G., & Ruiz-Primo, M. A. (1998). Toward a science performance assessment technology. *Evaluation and program planning : an international journal*, (21), 171–184.
- Shavelson, R. J., Ruiz-Primo, M. A., & Wiley, E. W. (1999). Note on Sources of Sampling Variability in Science Performance Assessments. *Journal of Educational Measurement*, 36(1), 61–71. doi:10.1111/j.1745-3984.1999.tb00546.x
- Shavelson R. J., Gao X., & Baxter G. P. (1993). *Sampling Variability of Performance Assessments: CSE Technical Report 361*. Retrieved from <http://www.cse.ucla.edu/products/Reports/TECH361.pdf>
- Stecher, B. M., Klein, S. P., Solano-Flores, G., McCaffrey, D., Robyn, A., Shavelson, R. J., & Haertel, E. (2000). The Effects of Content, Format, and Inquiry Level on Science Performance Assessment Scores. *Applied measurement in education*, 13(2), 139–160. doi:10.1207/S15324818AME1302\_2

- Stender, A. (2014). *Unterrichtsplanung: Vom Wissen zum Handeln : theoretische Entwicklung und empirische Überprüfung des Transformationsmodells der Unterrichtsplanung*. Univ., Diss.--Kiel, 2014. *Studien zum Physik- und Chemielernen: Vol. 170*. Berlin: Logos Verl.
- Tamir, P., Doran, R. L., & Chye, Y. O. (1992). Practical skills testing in science. *Studies in Educational Evaluation*, pp. 263–275. Retrieved from [http://ac.els-cdn.com/0191491X9290001T/1-s2.0-0191491X9290001T-main.pdf?\\_tid=71b36e94-5679-11e3-ad32-00000aab0f27&acdnat=1385456715\\_4800e938225952720e83c8c47c447688](http://ac.els-cdn.com/0191491X9290001T/1-s2.0-0191491X9290001T-main.pdf?_tid=71b36e94-5679-11e3-ad32-00000aab0f27&acdnat=1385456715_4800e938225952720e83c8c47c447688)
- Tamir, P., Doran, R. L., Kojima, S., & Bathory, Z. (1992). Procedures used in practical skills testing in science. *Studies in Educational Evaluation*, 18(3), 277–290. doi:10.1016/0191-491X(92)90002-U
- Tamir, P., & Lunetta, V. N. (1981). Inquiry-Related Tasks in High School Science Laboratory Handbooks. *Science Education*, 65(5), 477–484. doi:10.1002/sce.3730650503
- Tannenbaum, R. S. (1971). The Development of the Test of Science Process. *Journal of Research in Science*, 8(2), 123–136.
- Tesch, M. (2005). *Das Experiment im Physikunterricht: Didaktische Konzepte und Ergebnisse einer Videostudie*. *Studien zum Physik- und Chemielernen: Vol. 42*. Berlin: Logos-Verl.
- Thelk, A. D., & Hoole, E. R. (2006). What Are You Thinking? Postsecondary Student Think-Alouds of Scientific and Quantitative Reasoning Items. *The Journal of General Education*, 55(1), 17–39. doi:10.1353/jge.2006.0019
- Theyßen, H. (1999). *Ein Physikpraktikum für Studierende der Medizin: Darstellung der Entwicklung und Evaluation eines adressatenspezifischen Praktikums nach dem Modell der Didaktischen Rekonstruktion*. Univ., Diss.--Bremen, 1999. *Studien zum Physiklernen: Vol. 9*. Berlin: Logos-Verl.
- Thillmann, M. (2007). *Selbstreguliertes Lernen durch Experimentieren: Von der Erfassung zur Förderung*. Retrieved from [http://duepublico.uni-duisburg-essen.de/servlets/DerivateServlet/Derivate-18970/Dissertation\\_Thillmann\\_online-Version.pdf](http://duepublico.uni-duisburg-essen.de/servlets/DerivateServlet/Derivate-18970/Dissertation_Thillmann_online-Version.pdf)
- Tipler, P. A., Baumgartner, M., Gerlich, D., & Jerke, G. (1994). *Physik. Spektrum-Lehrbuch*. Heidelberg: Spektrum.
- Walpuski, M., & Sumfleth, E. (2009). The use of video data to evaluate inquiry situations in chemistry educations. In T. Seidel (Ed.), *The Power of Video Studies in Investigating Teaching and Learning in the Classroom*. 1. Aufl. s.l.: Waxmann Verlag GmbH.
- Walpuski, M. (2006). *Optimierung von experimenteller Kleingruppenarbeit durch Strukturierungshilfen und Feedback: Eine empirische Studie*. *Studien zum Physik- und Chemielernen: Vol. 49*. Berlin: Logos-Verl.
- Weinert, F. E. (2001). *Leistungsmessungen in Schulen* (Dr. nach Typoskript.). *Beltz-Pädagogik*. Weinheim: Beltz.



- Welzel, M., Haller, K., Bandiera, M., Hammelew, D., Koumaras, P., Niedderer, H., . . . Aufschnaiter, S. von. (1998). Ziele, die Lehrende mit dem Experimentieren in der naturwissenschaftlichen Ausbildung verbinden: Ergebnisse einer europäischen Umfrage. *Zeitschrift für Didaktik der Naturwissenschaften*, 4(1), 29–44.
- Wild, E., & Möller, J. (2009). *Pädagogische Psychologie: Mit 27 Tabellen*. Springer-Lehrbuch. Heidelberg: Springer. Retrieved from <http://www.socialnet.de/rezensionen/isbn.php?isbn=9783540885726>
- Wirth, J. (2004). *Selbstregulation von Lernprozessen*. Humboldt-Univ., Diss.--Berlin, 2003. *Pädagogische Psychologie und Entwicklungspsychologie: Vol. 39*. Münster: Waxmann. Retrieved from <http://www.gbv.de/dms/bsz/toc/bsz109501101inh.pdf>
- Wirtz, M., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen: Hogrefe Verl. für Psychologie.
- Wirtz, M. A., & Strohmer, J. (2013). *Dorsch: Lexikon der Psychologie* (16. vollst. überarb. Aufl.). Bern: H. Huber.
- Wissenschaftliches Konsortium HarmoS Naturwissenschaften+. (2008). *HarmoS Naturwissenschaften+: Kompetenzmodell und Vorschläge für Bildungsstandards*. Wissenschaftlicher Schlussbericht (1st ed.). Ostermündingen: suterprint.
- Woolnough, B. E. (1983). Exercises, investigations and experiences. *Physics Education*, pp. 60–63.

## 12 Abbildungsverzeichnis

Abbildung 1: Der Prozess des Experimentierens.....	16
Abbildung 2: Handlungen des potentiell iterativen Kreisprozess des Experimentierens. ....	23
Abbildung 3: Modell der Experimentierkompetenz mit Handlungen und Gütemaßen. ....	31
Abbildung 4: Beispielaufgabe einer produktorientierten Auswertung (aus Mullis & Martin, 2013). ...	35
Abbildung 5: Beispiel des zeitlichen Verlaufs von Handlungen bei Schreiber (Schreiber, 2012, Seite 71). ....	43
Abbildung 6: Beispielaufgabe des Experimentiertests aus dem HarmoS Projekt (aus Gut-Glanzmann, 2012, Seite 216). ....	44
Abbildung 7: Beispiel einer Aufgabe im Realexperiment Format von TIMSS (aus: Shavelson, Baxter, & Pine, 1991, Seite 354). ....	45
Abbildung 8: Beispiel einer Bewertung der Experimentierkompetenz durch ein Produkt (aus: Hammann & Hoi Phan, 2007, Seite 38). ....	48
Abbildung 9: Fachinhaltsverteilung deutschsprachiger physikalischer Praktika im Bereich Optik. ....	73
Abbildung 10 - Beispielaufgabe des Experimentiertests. ....	82
Abbildung 11 – Versuchsaufbau zu einer Beispielaufgabe a) mit einem Prisma (links) oder b) einem optischen Gitter (rechts) des Experimentiertests. ....	82
Abbildung 12: Theoretische Grundlage zu ähnlichen Dreiecken bei der Spektroskopie. ....	83
Abbildung 13: Rutschfeste, magnetische Arbeitsfläche und optische Komponenten vom Experimentiertest. ....	84
Abbildung 14: Experimentierkoffer mit Material. ....	85
Abbildung 15 - Beispiel eines ausgefüllten Laborhefts. ....	88
Abbildung 16: Identifikation von Fähigkeiten und Fertigkeiten: Theorie klären (blau), Experiment planen (orange), Messen (rot), Daten analysieren (grün), Schlussfolgerung (lila). ....	89
Abbildung 17: Zusammenfassung von aufeinander folgenden Aufzeichnungen zu gleichen Fähigkeiten bzw. Fertigkeiten. ....	90
Abbildung 18 - Testsituation bei der prozessorientierten Auswertung. ....	92
Abbildung 19: Handlungen bzgl. der verschiedenen Fähigkeiten und Fertigkeiten bei Aufgabe 1 bzgl. der Referenzauswertung; Fra (Fragestellung entwickeln/ klären); The (Theoretische Grundlagen klären/ Hypothese bilden); Pla (Experiment planen); Auf (Versuch fehlerfrei aufbauen); Mes (Messen und Beobachten); Ana (Daten analysieren); Sch (Schlussfolgerungen ziehen und diese kommunizieren); Son (Sonstiges). ....	114
Abbildung 20: Handlungen bzgl. der verschiedenen Fähigkeiten und Fertigkeiten bei Aufgabe 1 bzgl. der Referenzauswertung; Fra (Fragestellung entwickeln/ klären); The (Theoretische Grundlagen klären/ Hypothese bilden); Pla (Experiment planen); Auf (Versuch fehlerfrei aufbauen); Mes (Messen und Beobachten); Ana (Daten analysieren); Sch (Schlussfolgerungen ziehen und diese kommunizieren); Son (Sonstiges). ....	115
Abbildung 21: Handlungen bzgl. der verschiedenen Fähigkeiten und Fertigkeiten bei Aufgabe 1 bzgl. der Referenzauswertung; Fra (Fragestellung entwickeln/ klären); The (Theoretische Grundlagen klären/ Hypothese bilden); Pla (Experiment planen); Auf (Versuch fehlerfrei aufbauen); Mes (Messen und Beobachten); Ana (Daten analysieren); Sch (Schlussfolgerungen ziehen und diese kommunizieren); Son (Sonstiges). ....	116
Abbildung 22: Handlungen bzgl. der verschiedenen Fähigkeiten und Fertigkeiten bei Aufgabe 1 bzgl. der Referenzauswertung; Fra (Fragestellung entwickeln/ klären); The (Theoretische Grundlagen klären/ Hypothese bilden); Pla (Experiment planen); Auf (Versuch fehlerfrei aufbauen); Mes (Messen	

und Beobachten); Ana (Daten analysieren); Sch (Schlussfolgerungen ziehen und diese kommunizieren); Son (Sonstiges). .....	117
Abbildung 23: Handlungen bzgl. der verschiedenen Fähigkeiten und Fertigkeiten bei Aufgabe 1 bzgl. der Referenzauswertung; Fra (Fragestellung entwickeln/ klären); The (Theoretische Grundlagen klären/ Hypothese bilden); Pla (Experiment planen); Auf (Versuch fehlerfrei aufbauen); Mes (Messen und Beobachten); Ana (Daten analysieren); Sch (Schlussfolgerungen ziehen und diese kommunizieren); Son (Sonstiges). .....	118
Abbildung 24: Handlungen bzgl. der verschiedenen Fähigkeiten und Fertigkeiten bei Aufgabe 1 bzgl. der Referenzauswertung; Fra (Fragestellung entwickeln/ klären); The (Theoretische Grundlagen klären/ Hypothese bilden); Pla (Experiment planen); Auf (Versuch fehlerfrei aufbauen); Mes (Messen und Beobachten); Ana (Daten analysieren); Sch (Schlussfolgerungen ziehen und diese kommunizieren); Son (Sonstiges). .....	119
Abbildung 25: Handlungen der einzelnen Probanden (Proband 1 jeweils in rot auf der linken Seite) über die drei Auswerteverfahren Produkt (oben), Prozess (Mitte) und Referenz (unten) für Aufgabe 1. ....	127
Abbildung 26: Aufzeichnungen im neu entwickelten Auswerteverfahren (oben) mit einem Fotoverweis auf einen Versuchsaufbau (unten). .....	136
Abbildung 27: Beispielaufgabe aus dem Fachwissenstest zur Optik. ....	142
Abbildung 28: Modellierung eines PLS-SEM zur Varianzaufklärung des Endergebnisses aus den Aspekten der Experimentierkompetenz. ....	146
Abbildung 29: Beispiel der Aufzeichnungen in einem Testheft .....	239
Abbildung 30: Markierung der Aufzeichnungen zu "Wissen aktivieren (2´)" .....	240
Abbildung 31: Markierung der Aufzeichnungen zu "5. Messen und Beobachten" .....	240
Abbildung 32: Markierung der Aufzeichnungen zu allen Handlungen .....	241
Abbildung 33: Verschiedene, mögliche Kodierungen zu "5. Messen und Beobachten" .....	241
Abbildung 34: Eindeutige Kodierung.....	242
Abbildung 35: Beispiel der Aufzeichnungen in einem Testheft .....	276
Abbildung 36: Markierung der ersten Aufzeichnungen zu "Versuch fehlerfrei aufbauen (4´)" .....	277

## 13 Tabellenverzeichnis

Tabelle 1: Klassifizierung von Testinstrumenten im Spannungsfeld von Authentizität und Ökonomie. .....	41
Tabelle 2: Rangordnung der Fachinhalte in deutschsprachigen physikalischen Anfängerpraktika.....	74
Tabelle 3: Durchschnittliche Rangordnung der physikalischen Anfängerpraktika und des entwickelten Experimentiertests.....	79
Tabelle 4: Entwickelte Aufgaben des Experimentiertests.....	81
Tabelle 5: Tabelle mit der Abfolge der Fähigkeiten bzw. Fertigkeiten basierend auf den Aufzeichnungen. ....	91
Tabelle 6: Handlungsübergänge bei der idealtypischen Abfolge des Experimentierens. ....	95
Tabelle 7: Handlungsübergänge bei der Reaktion auf Fehler. ....	96
Tabelle 8: Handlungsübergänge bei Kontrollhandlungen.....	97
Tabelle 9: Handlungsübergänge bedingt durch vorherige Handlungen. ....	98
Tabelle 10: Handlungsübergänge bei Sprüngen. ....	98
Tabelle 11: Beispiel eines Experimentierprozesses mit möglichen, sinnvollen Folgehandlungen.....	100
Tabelle 12: Normative Setzung zur Zielorientiertheit durch die Anzahl an Zyklen.....	102
Tabelle 13: Durchführung Studie I.....	106
Tabelle 14: Beurteilerübereinstimmung zu den drei Auswerteverfahren. ....	108
Tabelle 15: Mittelwert und Standardabweichung bzgl. der drei Qualitätsmaße des Experimentierens und der drei Auswerteverfahren. ....	109
Tabelle 16: Mittlere Testleistung und Standardabweichung für die Aufgaben 1-6 (R: Richtigkeit; S: Strukturiertheit; Z: Zielorientiertheit; G: Gesamtmaß). ....	110
Tabelle 17: Interne Konsistenz und mittlere Trennschärfe der Aspekte der Experimentierkompetenz bzgl. der verschiedenen Auswerteverfahren. ....	111
Tabelle 18: Trennschärfe für den Aspekt der Richtigkeit. ....	111
Tabelle 19: Trennschärfe für den Aspekt der Strukturiertheit. ....	112
Tabelle 20: Trennschärfe für die Zielorientiertheit. ....	112
Tabelle 21: Trennschärfe für die Experimentierkompetenz.....	112
Tabelle 22: Einzelfälle mit hohem Einfluss sonstiger Handlungen. ....	122
Tabelle 23: Zusammenhänge bzgl. des Aspekts der Richtigkeit.....	130
Tabelle 24: Zusammenhänge bzgl. des Gesamtmaßes der Experimentierkompetenz. ....	130
Tabelle 25: Beispielaufgabe zum Strategiewissen.....	143
Tabelle 26: Beurteilerübereinstimmungen zu Studie III. ....	144
Tabelle 27: Interne Konsistenz und mittlere Trennschärfe der Aspekte der Experimentierkompetenz. .....	144
Tabelle 28: Korrelationen zwischen den Aspekten der Experimentierkompetenz.....	145
Tabelle 29: Varianzaufklärung von alternativen Auswerteverfahren durch die Aspekte der Experimentierkompetenz.....	147
Tabelle 1: Verkürzte Notation und Doppelkodierung.....	243
Tabelle 2: Kategorien zur Kodierung der Aufzeichnungen im Laborheft.....	244
Tabelle 2 Richtigkeit bezüglich der Teilkategorien der Kategorie "Messen und Beobachten" .....	264
Tabelle 1: Verkürzte Notation und Doppelkodierung.....	280
Tabelle 2: Kategorien zur Kodierung der Aufzeichnungen im Laborheft.....	281
Tabelle 8: Gesamtleistung über alle Auswertungsverfahren.....	313

A.1 Experimentierkompetenz: Fähigkeiten und Fertigkeiten nach vorliegenden Modellierungen

	Fragestellung bzw. Problemstellung entwickeln oder klären			
	Stellen bzw. Formulieren	Identifizieren	Klären	Entwickeln
(David Wetzel)		x		
(Emden, 2011); (Emden et al., Juli 2010)				
(Gut et al., 2010)	x			
(Hammann & Hoi Phan, 2007); (Hammann et al., 2008)				
(Hanauer et al., 2009)	x			
(Hodson, 1996)				
(Hofstein & Lunetta, 2004)	x			
(Kempa, 1986)	x			
(Kipnis & Hofstein, 2008)		x		
(Klahr & Dunbar, 2000)				
(Klieme et al., 2001)				
(Labudde et al., 2009)	x			
(Lunetta, 1998)				
(Maiseyenko et al., 2011)				x
(Mayer et al., 2003); (Mayer et al., 2008); (Grube & Christiane, 2010)	x			
(Murphy, 1996)		x		
(National Research Council, 2012)	x			
("National Science Education Standards," 1996)	x			
(Prenzel et al., 04.12.2007)		x		
(Schreiber, 2012)			x	
(Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005c); (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005a); (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005b)				
(Tamir, Doran, Kojima et al., 1992); (Tamir, Doran, & Chye, 1992); (Tamir & Lunetta, 1981)	x			
(Walpuski, 2006)				
(Welzel et al., 1998)				
(Wissenschaftliches Konsortium HarmoS Naturwissenschaften+, 2008)	x			

	<b>Hypothesen generieren/ Klären der theoretischen Grundlage</b>			
	<b>Hypothesen bilden</b>	<b>Formulierung von Erwartungen oder möglichen Lösungen</b>	<b>Herausfinden was bereits bekannt ist oder sich notwendiges Fachwissen aneignen</b>	<b>Analogien und Modelle nutzen sowie Mathematisierungen anwenden</b>
(David Wetzel)		x		
(Emden, 2011); (Emden et al., Juli 2010)				
(Gut et al., 2010)	x			
(Hammann & Hoi Phan, 2007); (Hammann et al., 2008)		x		
(Hanauer et al., 2009)			x	
(Hodson, 1996)				
(Hofstein & Lunetta, 2004)	x			
(Kempa, 1986)				
(Kipnis & Hofstein, 2008)	x			
(Klahr & Dunbar, 2000)		x		
(Klieme et al., 2001)				
(Labudde et al., 2009)				
(Lunetta, 1998)				
(Maiseyenka et al., 2011)	x			
(Mayer et al., 2003); (Mayer et al., 2008); (Grube & Christiane, 2010)	x			
(Murphy, 1996)				
(National Research Council, 2012)	x		x	
("National Science Education Standards," 1996)	x		x	
(Prenzel et al., 04.12.2007)			x	
(Schreiber, 2012)		x		

	Hypothesen bilden	Formulierung von Erwartungen oder möglichen Lösungen	Herausfinden was bereits bekannt ist oder sich notwendiges Fachwissen aneignen	Analogien und Modelle nutzen sowie Mathematisierungen anwenden
(Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005c); (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005a); (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005b)	x		x	x
(Tamir, Doran, Kojima et al., 1992); (Tamir, Doran, & Chye, 1992); (Tamir & Lunetta, 1981)	x			
(Walpuski, 2006)	x			
(Welzel et al., 1998)				
(Wissenschaftliches Konsortium HarmoS Naturwissenschaften+, 2008)	x		x	

	Experiment planen	
	Experiment planen	Variablen -Kontroll-Strategie planen
(David Wetzel)	x	
(Emden, 2011); (Emden et al., Juli 2010)	x	
(Gut et al., 2010)	x	
(Hammann & Hoi Phan, 2007); (Hammann et al., 2008)		
(Hanauer et al., 2009)	x	
(Hodson, 1996)	x	
(Hofstein & Lunetta, 2004)	x	
(Kempa, 1986)	x	
(Kipnis & Hofstein, 2008)	x	
(Klahr & Dunbar, 2000)		
(Klieme et al., 2001)	x	
(Labudde et al., 2009)		
(Lunetta, 1998)	x	
(Maiseyenko et al., 2011)	x	
(Mayer et al., 2003); (Mayer et al., 2008); (Grube & Christiane, 2010)	x	
(Murphy, 1996)	x	
(National Research Council, 2012)	x	x
("National Science Education Standards," 1996)	x	x
(Prenzel et al., 04.12.2007)		
(Schreiber, 2012)	x	
(Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005c); (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005a); (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005b)	x	
(Tamir, Doran, Kojima et al., 1992); (Tamir, Doran, & Chye, 1992); (Tamir & Lunetta, 1981)	x	x
(Walpuski, 2006)		
(Welzel et al., 1998)	x	
(Wissenschaftliches Konsortium HarmoS Naturwissenschaften+, 2008)	x	



	Versuch fehlerfrei aufbauen	
	Versuch wie geplant aufbauen	Fehler erkennen und beheben
(David Wetzel)		
(Emden, 2011); (Emden et al., Juli 2010)		
(Gut et al., 2010)	x	
(Hamann & Hoi Phan, 2007); (Hamann et al., 2008)		
(Hanauer et al., 2009)		
(Hodson, 1996)		
(Hofstein & Lunetta, 2004)	x	
(Kempa, 1986)	x	
(Kipnis & Hofstein, 2008)		
(Klahr & Dunbar, 2000)		
(Klieme et al., 2001)		
(Labudde et al., 2009)		
(Lunetta, 1998)		
(Maiseyenko et al., 2011)	x	x
(Mayer et al., 2003); (Mayer et al., 2008); (Grube & Christiane, 2010)		
(Murphy, 1996)		
(National Research Council, 2012)	x	
("National Science Education Standards," 1996)	x	
(Prenzel et al., 04.12.2007)		
(Schreiber, 2012)	x	
(Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005c); (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005a); (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005b)		
(Tamir, Doran, Kojima et al., 1992); (Tamir, Doran, & Chye, 1992); (Tamir & Lunetta, 1981)		
(Walpuski, 2006)		
(Welzel et al., 1998)		x
(Wissenschaftliches Konsortium HarmoS Naturwissenschaften+, 2008)	x	

	<b>Messen, Beobachten und Dokumentieren</b>					
	<b>Experiment durchführen</b>	<b>Sicherheitsmaßnahmen beobachten</b>	<b>Variablen-Kontroll-Strategie durchführen</b>	<b>Qualität der Messung beurteilen bzw. sicherstellen</b>	<b>Beobachtungen anstellen</b>	<b>Dokumentieren</b>
(David Wetzel)	x					
(Emden, 2011); (Emden et al., Juli 2010)	x					
(Gut et al., 2010)	x					
(Hammann & Hoi Phan, 2007); (Hammann et al., 2008)	x					
(Hanauer et al., 2009)					x	
(Hodson, 1996)	x					
(Hofstein & Lunetta, 2004)	x			x	x	x
(Kempa, 1986)			x		x	x
(Kipnis & Hofstein, 2008)			x			
(Klahr & Dunbar, 2000)	x					
(Klieme et al., 2001)						
(Labudde et al., 2009)						
(Lunetta, 1998)						
(Maiseyenka et al., 2011)			x		x	
(Mayer et al., 2003); (Mayer et al., 2008); (Grube & Christiane, 2010)						
(Murphy, 1996)	x					x
(National Research Council, 2012)	x	x	x	x	x	x
("National Science Education Standards," 1996)	x				x	
(Prenzel et al., 04.12.2007)						
(Schreiber, 2012)			x			x
(Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005c); (Sekretariat der Ständigen Konferenz der Kultusminister der Län-	x					x

der in der Bundesrepublik Deutschland [KMK], 2005a); (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005b)						
(Tamir, Doran, Kojima et al., 1992); (Tamir, Doran, & Chye, 1992); (Tamir & Lunetta, 1981)	x		x		x	x
(Walpuski, 2006)	x					
(Welzel et al., 1998)	x				x	
(Wissenschaftliches Konsortium HarmoS Naturwissenschaften+, 2008)	x					

	<b>Daten analysieren und Verallgemeinerbarkeit diskutieren</b>				
	<b>Daten aufbereiten</b>	<b>Berechnungen anstellen</b>	<b>Zeichnungen und Graphen anfertigen oder auswerten</b>	<b>Analyse der Daten</b>	<b>Verallgemeinerbarkeit z.B. durch Fehlerrechnung</b>
(David Wetzel)					
(Emden, 2011); (Emden et al., Juli 2010)				x	
(Gut et al., 2010)	x			x	
(Hammann & Hoi Phan, 2007); (Hammann et al., 2008)				x	
(Hanauer et al., 2009)				x	
(Hodson, 1996)	x			x	
(Hofstein & Lunetta, 2004)					x
(Kempa, 1986)				x	
(Kipnis & Hofstein, 2008)					
(Klahr & Dunbar, 2000)				x	
(Klieme et al., 2001)					
(Labudde et al., 2009)					
(Lunetta, 1998)				x	
(Maiseyenko et al., 2011)				x	
(Mayer et al., 2003); (Mayer et al., 2008); (Grube & Christiane, 2010)				x	
(Murphy, 1996)				x	
(National Research Council, 2012)		x	x	x	x
("National Science Education Standards," 1996)				x	
(Prenzel et al., 04.12.2007)					
(Schreiber, 2012)		x	x		

	Daten aufbereiten	Berechnungen anstellen	Zeichnungen und Graphen anfertigen oder auswerten	Analyse der Daten	Verallgemeinerbarkeit z.B. durch Fehlerrechnung
(Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005c); (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005a); (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005b)	x	x	x	x	x
(Tamir, Doran, Kojima et al., 1992); (Tamir, Doran, & Chye, 1992); (Tamir & Lunetta, 1981)		x			x
(Walpuski, 2006)					
(Welzel et al., 1998)					
(Wissenschaftliches Konsortium HarmoS Naturwissenschaften+, 2008)	x			x	

	<b>Sachgerechte Schlüsse ziehen und diese Kommunizieren</b>			
	<b>Daten bzgl. der Frage- bzw. Problemstellung diskutieren und interpretieren</b>	<b>Bezug zu den Hypothesen/ Erwartungen/Grundlagen</b>	<b>Schlussfolgerung/ Ergebnis formulieren</b>	<b>Kommunizieren</b>
(David Wetzel)				
(Emden, 2011); (Emden et al., Juli 2010)				
(Gut et al., 2010)	x	x		x
(Hammann & Hoi Phan, 2007); (Hammann et al., 2008)				
(Hanauer et al., 2009)	x	x		x
(Hodson, 1996)				
(Hofstein & Lunetta, 2004)	x	x		x
(Kempa, 1986)	x			
(Kipnis & Hofstein, 2008)			x	
(Klahr & Dunbar, 2000)				
(Klieme et al., 2001)				
(Labudde et al., 2009)				
(Lunetta, 1998)		x		
(Maiseykenka et al., 2011)			x	
(Mayer et al., 2003); (Mayer et al., 2008); (Grube & Christiane, 2010)			x	
(Murphy, 1996)	x		x	
(National Research Council, 2012)	x	x	x	x
("National Science Education Standards," 1996)	x	x	x	x
(Prenzel et al., 04.12.2007)				
(Schreiber, 2012)	x			

	Daten bzgl. der Frage- bzw. Problemstellung diskutieren und interpretieren	Bezug zu den Hypothesen/ Erwartungen/Grundlagen	Schlussfolgerung/ Ergebnis formulieren	Kommunizieren
(Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005c); (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005a); (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005b)		x		
(Tamir, Doran, Kojima et al., 1992); (Tamir, Doran, & Chye, 1992); (Tamir & Lunetta, 1981)		x		
(Walpuski, 2006)			x	
(Welzel et al., 1998)	x			
(Wissenschaftliches Konsortium HarmoS Naturwissenschaften+, 2008)	x	x		x

## A.2 Klassifizierung von Experimentiertests

Test	Datum	Testgegenstand	Format	Auswertungs-gegenstand	Experimentieren (ja/nein)	Interaktion zwischen F&F (ja/nein)
FISS (Comber & Keeves, 1973)	1968-1973	Beobachten; Messen; Experimentieren	P&P	Produkt	nein	nein
Test of Science Processes (Tannenbaum, 1971)	1971	<i>Observaing; Comparing; Classifying; Quantifying; Measuring; Experimenting; Inferring und Predicting</i>	P&P	Produkt	nein	nein
TOES (Fraser, 1980)	1980	<i>Libarry usage; Index &amp; table of contents; Sclaes; Averages, percentages &amp; propotions; Charts &amp; tables; Graphs; Comprehension of science reading; Design of experimental procedures; Conclusions &amp; generalizations</i>	P&P	Produkt	nein	nein
SISS P&P (IEA, 1988)	1988	<i>terms of simple information, the understanding of a principle, and the application of information and understanding to solve a practical problem</i>	P&P	Produkt	nein	nein
SISS RE (Tamir, Doran, & Chye, 1992)	1992	<i>Ziel of the practical skills testing in SISS was to test student's abilities and knowledge connected with experimental work in the field of science education [...] the testing program sought to examine the cooperation between head and hand</i>	RE	Produkt	ja	ja
TIMSS 95 (Beaton, 1996)	1995	<i>Recall/Recognize; Define; Describe; Use Tools and Procedures; Illustrate with Examples; Compare/Contrast/Classify; Represent/Model; Relate; Extract/Apply Information (Interpret Information); Find Solutions; Explain; Analyze/Interpret/Solve Problems; Integrate/Synthesize; Hypothesize/Predict; Design/Plan; Collect/Analyze/Interpret Data; Draw Conclusions; Generalize; Evaluate und Justity</i>	P&P	Produkt	nein	nein



Test	Datum	Testgegenstand	Format	Auswertungs-gegenstand	Experimentieren (ja/nein)	Interaktion zwischen F&F (ja/nein)
TIMSS 99 (Martin, 2000)	1999	<i>Recall/Recognize; Define; Describe; Use Tools and Procedures; Illustrate with Examples; Compare/Contrast/Classify; Represent/Model; Relate; Extract/Apply Information (Interpret Information); Find Solutions; Explain; Analyze/Interpret/Solve Problems; Integrate/Synthesize; Hypothesize/Predict; Design/Plan; Collect/Analyze/Interpret Data; Draw Conclusions; Generalize; Evaluate und Justify</i>	P&P	Produkt	nein	nein
TIMSS 03 (Martin, 2004)	2003	<i>Recall/Recognize; Define; Describe; Use Tools and Procedures; Illustrate with Examples; Compare/Contrast/Classify; Represent/Model; Relate; Extract/Apply Information (Interpret Information); Find Solutions; Explain; Analyze/Interpret/Solve Problems; Integrate/Synthesize; Hypothesize/Predict; Design/Plan; Collect/Analyze/Interpret Data; Draw Conclusions; Generalize; Evaluate und Justify</i>	P&P	Produkt	nein	nein
TIMSS 07 (Martin, 2008)	2007	<i>Recall/Recognize; Define; Describe; Use Tools and Procedures; Illustrate with Examples; Compare/Contrast/Classify; Represent/Model; Relate; Extract/Apply Information (Interpret Information); Find Solutions; Explain; Analyze/Interpret/Solve Problems; Integrate/Synthesize; Hypothesize/Predict; Design/Plan; Collect/Analyze/Interpret Data; Draw Conclusions; Generalize; Evaluate und Justify</i>	P&P	Produkt	nein	nein
TIMSS 11 (Martin, 2012)	2011	<i>Recall/Recognize; Define; Describe; Use Tools and Procedures; Illustrate with Examples; Compare/Contrast/Classify; Represent/Model; Relate; Extract/Apply Information (Interpret Information); Find Solutions; Explain; Analyze/Interpret/Solve Problems; Integrate/Synthesize; Hypothesize/Predict; Design/Plan; Collect/Analyze/Interpret Data; Draw Conclusions; Generalize; Evaluate und Justify</i>	P&P	Produkt	nein	nein

Test	Datum	Testgegenstand	Format	Auswertungs-gegenstand	Experimentieren (ja/nein)	Interaktion zwischen F&F (ja/nein)
TIMSS-R (Harmon, 1997)	1995	<i>recalling and using simple or complex information; using equipment, routine procedures, and experimental processes; problem solving; designing and conducting an investigation; analyzing and interpreting findings; formulating and justifying conclusions; and communicating scientific or mathematical information (..)</i>	RE	Produkt	ja	ja
NAW (auch Protokollbögen) (Walpuski, 2006); (Klos, 2008)	2004+	<i>Suche im Hypothesenraum; Testen von Hypothesen; und Analyse von Evidenzen</i>	P&P;RE	Produkt	nein	nein / ja
PISA (Organisation for Economic Co-operation and Development, 2007)	2006	<i>: Identify scientific issues; Explain phenomena scientifically; and Use scientific evidence</i>	P&P	Produkt	nein	nein
Hammann (Hammann & Hoi Phan, 2007)	2007	<i>Suche im Hypothesenraum; Testen von Hypothesen; und Analyse von Evidenzen</i>	P&P	Produkt	nein	nein
Hammann (Hammann et al., 2008)	2008	<i>Suche im Hypothesenraum; Testen von Hypothesen; und Analyse von Evidenzen</i>	RE	Produkt	ja	ja
NEAPs (National Assessment Governing board U.S. Department of Education, 2008)	2008	<i>Identifying Science Principles; Using Science Principles; Using Scientific Inquiry; und Using Technological Design</i>	P&P	Produkt	nein	ja

Test	Datum	Testgegenstand	Format	Auswertungsgegenstand	Experimentieren (ja/nein)	Interaktion zwischen F&F (ja/nein)
NEAPs (National Assessment Governing board U.S. Department of Education, 2008)	2008	<i>the quality of students: reasoning while framing their research questions; planning for data collection and the execution of the plan; abilities to meet unpredictable challenges that arise during an actual, ongoing scientific investigation; lines of argument in deciding how to alter their experimental approach in the light of new evidence; engagement with fellow students and/or the teacher in interpreting an observation or result and deciding what to do about it; and deliberations and reasoning when settling on the defensible conclusions that might be drawn from their work</i>	RE	Produkt	ja	ja
HarmoS (Wissenschaftliches Konsortium HarmoS Naturwissenschaften+, 2008)	2008	Fragen, Probleme und Hypothesen aufwerfen; geeignete Werkzeuge, Instrumente und Materialien auswählen; Informationen recherchieren; Informationen und Informationsquellen einordnen; Planen; Vorhaben planen; Bewusst wahrnehmen; geeignete Werkzeuge, Instrumente und Materialien verwenden; Erkdungen, Untersuchungen oder Experimente durchführen; Vorhaben umsetzen; Über Ergebnisse und Untersuchungsmethoden nachdenken; Analysieren und Strukturieren; Ergebnisse aufbereiten und präsentieren; Beschreiben, Präsentieren und Begründen	P&P	Produkt	nein	nein
HarmoS (Wissenschaftliches Konsortium HarmoS Naturwissenschaften+, 2008)	2008	Fragen, Probleme und Hypothesen aufwerfen; geeignete Werkzeuge, Instrumente und Materialien auswählen; Informationen recherchieren; Informationen und Informationsquellen einordnen; Planen; Vorhaben planen; Bewusst wahrnehmen; geeignete Werkzeuge, Instrumente und Materialien verwenden; Erkdungen, Untersuchungen oder Experimente durchführen; Vorhaben umsetzen; Über Ergebnisse und Untersuchungsmethoden nachdenken; Analysieren und Strukturieren; Ergebnisse aufbereiten und präsentieren; Beschreiben, Präsentieren und Begründen	RE	Produkt	ja	ja

Test	Datum	Testgegenstand	Format	Auswertungsgegenstand	Experimentieren (ja/nein)	Interaktion zwischen F&F (ja/nein)
AG Bremen (Maiseyenka et al., 2011)	2011+	Fragestellung entwickeln; Hypothese generieren; Experiment planen; Versuch fehlerfrei aufbauen; Beobachten und Messen; Daten analysieren; Sachgerechte Schlüsse ziehen	P&P	Produkt	ja	teilweise
Emden (Emden, 2011)	2011	Vorbereitung, Durchführung, Auswertung	P&P	Produkt	nein	ja
Emden (Emden, 2011)	2011	Vorbereitung, Durchführung, Auswertung	RE	Prozess	ja	ja
Schreiber (Schreiber, 2012)	2012	Fragestellung klären; Versuchsplan entwerfen; Geräte zusammenstellen; Versuchsanordnung aufbauen; Messungen durchführen; Messungen dokumentieren; mit Daten Berechnungen anstellen; Ergebnisse interpretieren	CS; RE	Prozess	ja	ja

### A.3 Kategorienbildung Fachinhaltsanalyse

Tipler	Bergmann Schäfer	Halliday	Demtröder	Kategorie
Die Lichtgeschwindigkeit	Lichtgeschwindigkeit		Messung der Lichtgeschwindigkeit	1. Lichtgeschwindigkeit
Die Ausbreitung des Lichts	geradlinige Ausbreitung			
das Huygenssche Prinzip	Schatten			
Reflexion	Reflexion	Reflexion	Reflexionsgesetz	2. Reflexion
			Brewsterwinkel	
	Totalreflexion		totale Reflexion	
			Phasenänderung bei der Reflexion	
ebene- und sphärische Spiegel	ebene- und gekrümmte Spiegel	ebene- und sphärische Spiegel		
Brechung	Brechung	Brechung	Brechungsindex	3. Brechung
	Brechung an Kugeln		Brechung an gekrümmten Flächen	
	Prismen		Prismen	
	Dispersion		Dispersion	
		sphärische brechende Flächen		
Dünne Linsen	Linsen	dünne Linsen	dünne-, dicke und sphärische Linsen	
			Linsensysteme	
Polarisation	Polarisation	Polarisation	Polarisation	4. Polarisation
	Herstellung von polarisiertem Licht	Herstellung von polarisiertem Licht	Herstellung von polarisiertem Licht	
			Arten von polarisiertem Licht	

Tipler	Bergmann Schäfer	Halliday	Demtröder	Kategorie
	elektromagnetische Welle	elektromagnetische Welle	elektromagnetische Wellen	5. Licht als Welle
	Helligkeit	Intensität		
		Licht und Welle		
		Wellenoptik	Wellengleichung	
			ebene- und periodische Wellen	
			Wellen in Leitern	
			Lichtmodulation	
	Lichtelektrische Effekte			6. Teilcheneigenschaften
	Absorption		Absorption	
	Infrarot- und Ultraviolette Strahlen			7. elektromagnetisches Spektrum
	Spektren		elektromagnetisches Frequenzspektrum	
	Spektroskopie		Spektroskopie	
Geometrische Optik	Geometrische Optik	Abbildungen	Geometrische Optik	8. Bildentstehung
	Bildentstehung		optische Abbildungen	
	Strahlenbegrenzung; Blenden			
Bilder durch Brechung	Abbildungen durch Brechung			
Abbildungsfehler	Abbildungsfehler		Auflösungsvermögen	

Tipler	Bergmann Schäfer	Halliday	Demtröder	Kategorie
Optische Instrumente	optische Instrumente	optische Instrumente	optische Instrumente und Auflösungsvermögen optischer Instrumente	9. optische Instrumente
Auge	Auge		Auge	
Lupe			Lupe	
Kamera				
Mikroskop			Mikroskop	
Teleskop				
			Fernrohr	
Interferenz	Interferenz	Interferenz	Interferenz	10. Interferenz
Phasendifferenz und Kohärenz	Phasengeschwindigkeit; Kohärenz	Kohärenz	Kohärenz; Erzeugung und Überlagerung von kohärenten Wellen	
Interferometer		Interferometer	Interferometer	
Interferenz beim Doppelspalt				
vektoraddition von harmonischen Wellen				
Interferenz bei mehreren äquidistanten Quellen				
			Holographie	

<b>Tipler</b>	<b>Bergmann Schäfer</b>	<b>Halliday</b>	<b>Demtröder</b>	<b>Kategorie</b>
Beugung	Beugung	Beugung	Beugung und Streuung	11. Beugung
Beugungsmuster am Einfach- und Doppelspalt		Beugung am Einfach- und Doppelspalt	Beugung an Einfach- und Doppelspalt	
Fraunhofersche- und Fresnelsche Beugung	Fresnelsche Beugung		Fraunhofersche- und Fresnelsche Beugung	
Beugung und Auflösung	Auflösungsvermögen			
Gitter			Gitter	
			Lichtstreuung	
			Fourieroptik	
			Diffraktive Optik.	

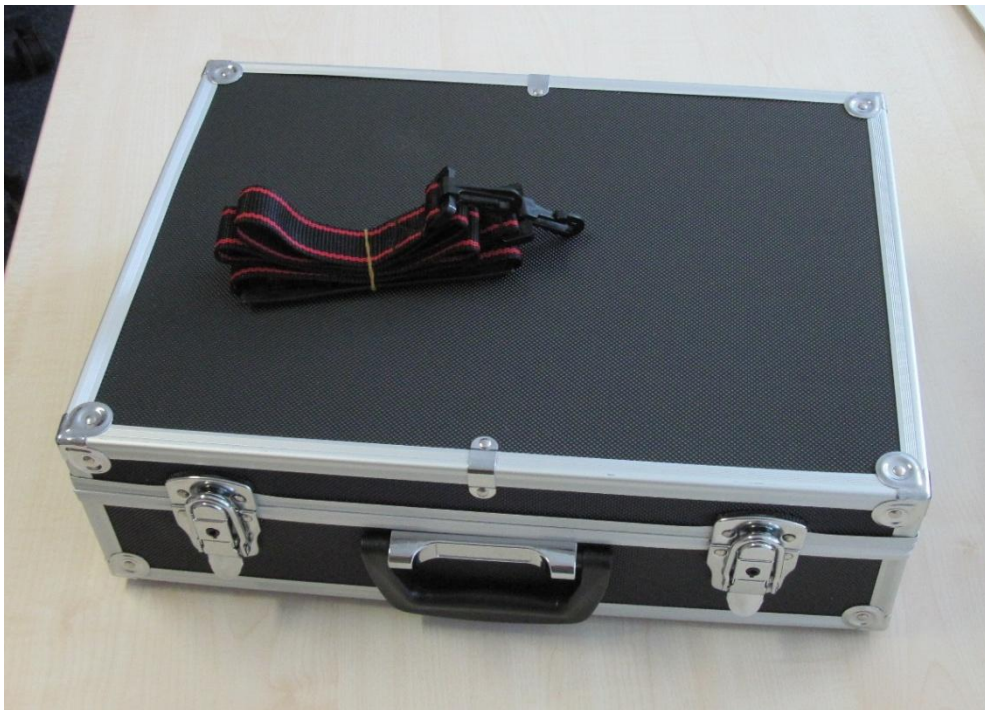


## **A.4 Experimentiertes**

### **A.4.1 Theoretische Hinweise**

# **Erfassung von Experimentierfähigkeiten im Kontext Optik**

## **Theoretische Hinweise und Formeln**



# Inhalt

## Eigenschaften von Licht

- Elektromagnetische Welle und Teilchen
- Intensität
- Polarisierung
  - o Filter

## Wechselwirkung von Licht mit Materie

- Beugung
- Interferenz
- Optische Komponenten und ihre Anwendungen
  - o Linsen
  - o Spiegel
  - o Brennweite
  - o Blenden
- Auflösungsvermögen
- Brechung
  - o Prismen
  - o Dispersion
  - o Dispersionsgraph
- Verschiedenes

## Eigenschaften von Licht

### Elektromagnetische Welle und Teilchen

Licht lässt sich als elektromagnetische Welle und als Teilchen beschreiben. Für eine elektromagnetische Welle charakteristisch sind die Größen der Wellenlänge ( $\lambda$ ), der Frequenz ( $f$ ) und der Amplitude ( $A$ ). Bei der Beschreibung des Lichts als Teilchen sind Impuls ( $p$ ) und Energie ( $E$ ) die charakteristischen Größen. Frequenz einer Welle und Energie eines Teilchens lassen sich über die Beziehung  $E=h \cdot f$  ineinander überführen.

### Intensität

Als Intensität  $I$  bezeichnet man die Energie, welche in einer bestimmten Zeit auf eine Fläche trifft. Für elektromagnetische Wellen ist die Intensität näherungsweise das Quadrat der Amplitude  $I=|A|^2$ . Bewegt sich Licht durch ein Medium (beispielsweise Luft, Glas oder Wasser) so nimmt die Intensität ab. Dies liegt in der Wechselwirkung der Welle bzw. des Teilchens mit dem Medium begründet. Die Intensität  $I$  nimmt in Abhängigkeit zum Absorptionskoeffizienten  $\alpha$  des Mediums gegenüber ihrer Ausgangsintensität  $I_0$  ab, je weiter die Strecke  $x$  ist, welche das Licht durch das Medium zurücklegt  $I(x)=I_0 \cdot e^{-\alpha x}$ . Entsprechend bezeichnet man die Strecke  $x_0$ , nach welcher sich die Intensität bzgl. der Ausgangsintensität halbiert hat, als Halbwertsdicke eines Mediums. Messen lässt sich die Intensität mit einem Luxmeter.

### Polarisation

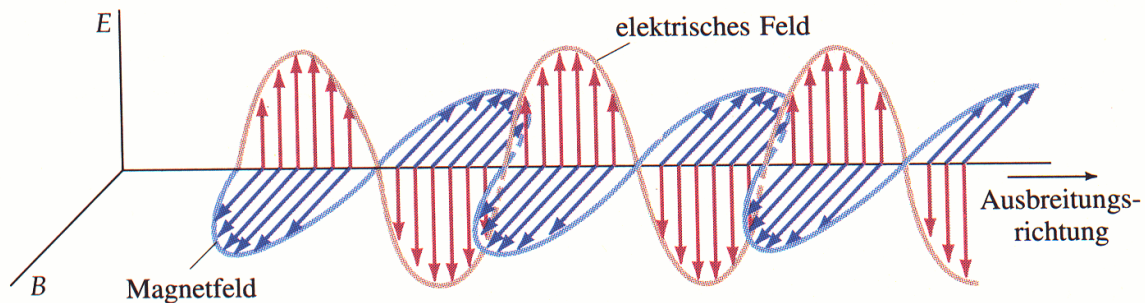
Licht ist eine elektromagnetische Welle. Die Welle besteht darin, dass ein zeitlich veränderliches elektrisches Feld ein Magnetfeld hervorruft, und aus einem zeitlich veränderlichen Magnetfeld entsteht wiederum ein elektrisches Feld.

Elektrisches und magnetisches Feld sind durch ihren Betrag und ihre Richtung gekennzeichnet, sind also Vektoren. Sie stehen senkrecht zueinander. Im Allgemeinen ist Licht unpolarisiert. Bleibt bei der

ständigen Umwandlung entweder die Richtung des elektrischen Vektors (außer dem Vorzeichen) oder der Betrag erhalten, spricht man von polarisiertem Licht.

Betrachten wir zunächst den Fall, dass die Richtung des elektrischen Vektors, abgesehen vom Vorzeichenwechsel, erhalten bleibt. Zusammen mit der Ausbreitung des Lichtes wird eine Ebene aufgespannt, man sagt vereinfachend oft, das Licht schwingt in einer Ebene. Der Physiker spricht von Linearer Polarisation. Dies ist in der Abbildung dargestellt.

(Quelle: <http://www.aip.de/de/forschung/forschungsschwerpunkt-kmf/cosmic-magnetic-fields/sonnenphysik/optische-sonnenphysik/forschungsprogramm/polarisiertes-licht>)



(Quelle: [http://web.physik.rwth-aachen.de/~hebbeker/lectures/ph2\\_02/tipl293.gif](http://web.physik.rwth-aachen.de/~hebbeker/lectures/ph2_02/tipl293.gif))

Linear polarisiertes Licht lässt sich durch den Einsatz eines Polarisationsfilters erzeugen, dieser wird Polarisator genannt. Der Polarisator lässt nur Licht durch, das in einer Ebene schwingt. Um zu prüfen, ob Licht linear polarisiert ist, werden Polarisationsfilter genutzt, in diesem Fall wird dieser Analysator genannt. Ist die Intensität unabhängig von den Einstellungen des Analysators, so ist das Licht nicht polarisiert. Finden sich hingegen Maxima und Minima, so ist das Licht linear polarisiert.

Optisch aktive Medien drehen die Schwingungsebene des elektrischen- bzw. magnetischen Feldes von Licht, das sich durch das optisch aktive Medium bewegt. Der Drehwinkel hängt dabei von der durchlaufenden Strecke im optisch aktiven Medium und einer spezifischen Konstante abhängig von dem optisch aktiven Medium ab.

## Wechselwirkung von Licht mit Materie

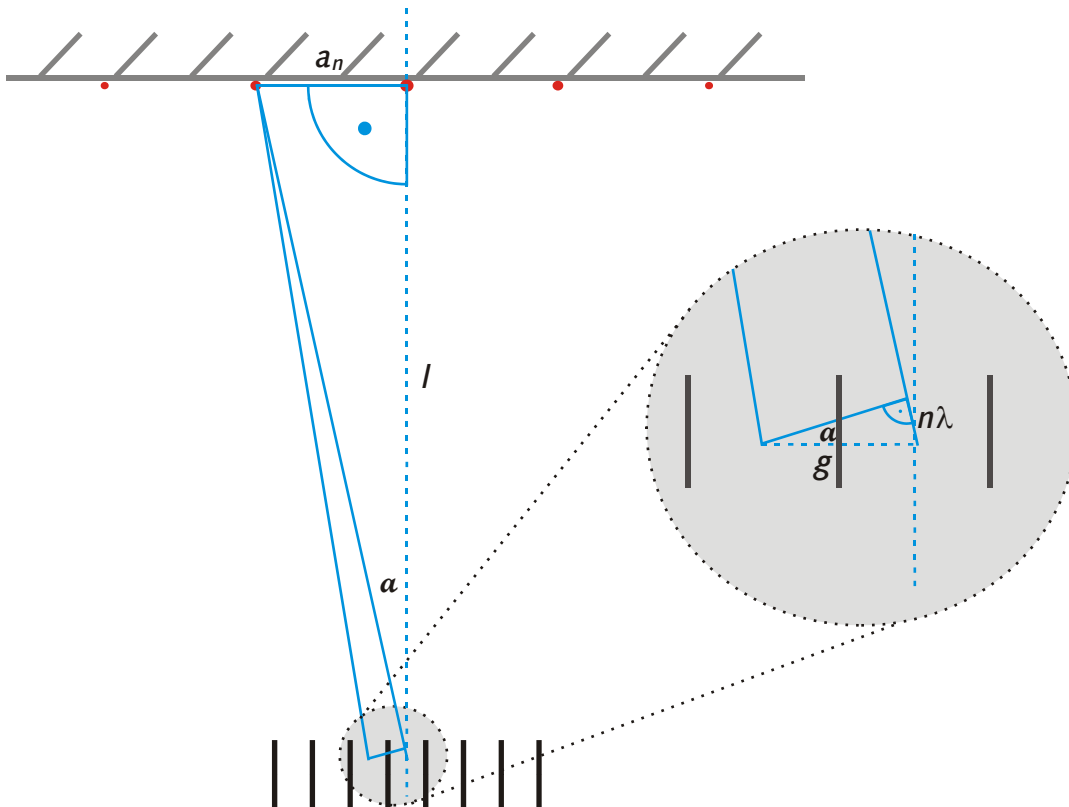
### Beugung

Beugung ist die Ablenkung einer elektromagnetischen Welle an einem Hindernis. Gemäß des Huygens-Fresnelschen Prinzips lässt sich jeder Punkt entlang einer Wellenfront als Elementarerreger verstehen. Damit kann erklärt werden, wie sich eine Welle in den geometrischen Schattenraum eines Hindernisses ausbreiten kann.

### Interferenz

Betrachtet man verschiedene Wellen, so können diese miteinander interagieren. Nach dem Superpositionsprinzip werden die Auslenkungen aller Wellen an einem Punkt des Raumes zusammen addiert und bilden eine resultierende Welle.

Zur Berechnung von konstruktiver Interferenz am Gitter hilft die folgende Abbildung.



Darin finden sich zwei ähnliche Dreiecke, ein kleines und ein großes. Mit  $l \gg a_n$  gilt für diese Dreiecke gilt:

$$n\lambda = g \sin \alpha, \quad a_n = l \sin \alpha$$

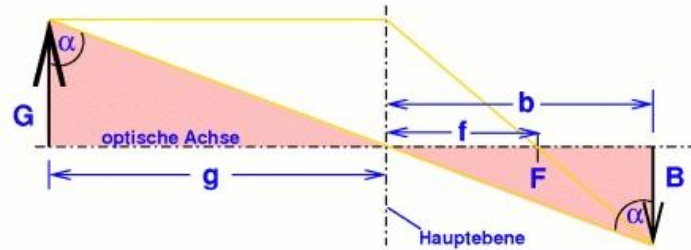
Dabei bezeichnet:

- $\lambda$  die Wellenlänge,
- $g$  den Abstand zwischen zwei Spalten,
- $\alpha$  den Winkel zwischen eingehender und ausgehender Wellenfront bzw. zwischen der optischen Achse und dem n-ten Maximum,
- $l$  den Abstand von Gitter zum Schirm und
- $a_n$  den Abstand von der optischen Achse zum n-ten Maximum.

## Optische Komponenten und ihre Anwendungen

### Linse

Eine Sammellinse ist eine sphärische Linse mit positiver, vergrößernder Brechkraft. Sie sammelt parallel einfallendes Licht in der Brennebene. Speziell das parallel zur optischen Achse eingestrahle Licht wird im Brennpunkt fokussiert. Die Sammellinse macht aus einem parallelen Lichtstrahlenbündel konvergente Strahlen, wodurch sie ein reelles Abbild der Umgebung erlaubt. Linsen stellen ein Hauptbauelement optischer Systeme, wie sie beispielsweise für Foto-Objektive verwendet werden, dar.



Linsen stellen ein Hauptbauelement optischer Systeme, wie sie beispielsweise für Foto-Objektive verwendet werden, dar.

- Für die Bildentstehung gelten die folgenden Gleichungen:
  - Abbildungsgleichung:

$$\frac{B}{b} = \frac{G}{g}$$

- Linsengleichung:

$$\frac{1}{f} = \frac{1}{g} + \frac{1}{b}$$

(Quelle: <http://de.wikipedia.org/wiki/Bikonvexlinse>)

Eine Streulinse ist eine sphärische Linse mit negativer Brechkraft. Sie bricht parallel zur optischen Achse auf eine dünne Zerstreuungslinse fallende Strahlen so, dass die gebrochenen Strahlen alle von einem Punkt zu kommen scheinen, dem virtuellen Bildbrennpunkt. Zum Brennpunkt hinzielende Strahlen verlassen die Linse parallel zur optischen Achse (Umkehrung des Strahlenganges). Abbildungsgleichung ist die gleiche wie für die Sammellinsen, die Bildbrennweite ist aber negativ.

(Quelle: [http://www.sprott.net/science/physik/taschenbuch/daten/kap\\_11/node54.htm](http://www.sprott.net/science/physik/taschenbuch/daten/kap_11/node54.htm))

### Spiegel

Konvexe Spiegel oder Wölbspiegel oder erhabene Spiegel ermöglichen „um die Ecke zu schauen“ und durch einen Weitwinkelleffekt den Blickwinkel zu vergrößern und werden deshalb an Verkehrsstraßen zum Einblick in sonst unüberseh- oder schlecht einsehbare Abzweigungen eingesetzt

(Quelle: <http://de.wikipedia.org/wiki/Konvexspiegel>)

### Brennweite

Spiegel und Linsen besitzen eine sogenannte Brennweite  $f$ . Die Brennweite beschreibt in welchem Abstand von der Komponente parallel zur optischen Achse einfallende Lichtstrahlen sich treffen.

### Blenden

Durch Lochblenden lassen sich Objekte wie mit Linsen abbilden.

### Auflösungsvermögen:

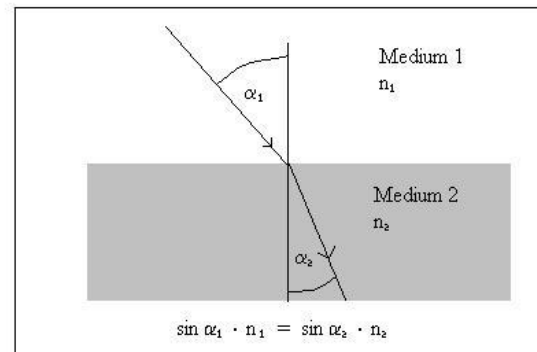
Der Begriff Auflösungsvermögen bezeichnet die Unterscheidbarkeit feiner Strukturen, also z. B. den kleinsten noch wahrnehmbaren Abstand zweier punktförmiger Objekte. Durch die Angabe eines Winkelabstandes oder durch die Angabe des Abstandes gerade noch trennbarer Strukturen lässt er sich quantifizieren.

(Quelle: <http://de.wikipedia.org/wiki/Aufl%C3%B6sungsverm%C3%B6gen>)

## Brechung

Das Snelliussche Brechungsgesetz besagt, dass eine Welle (z. B. ein Lichtstrahl) ihre Richtung ändert – man sagt gebrochen wird – wenn sie von einem transparenten Medium in ein anderes transparentes Medium auf Grund unterschiedlicher Materialeigenschaften (wie der Dichte) mit einer anderen Phasengeschwindigkeit übergeht. Das Gesetz gilt für alle Wellenarten. Es besagt nur, in welche Richtung die Welle abgelenkt wird, nicht aber, wie viel von der Welle an dem Übergang zwischen den beiden Medien transmittiert bzw. reflektiert wird. Im Fall der Totalreflexion ist das reelle Brechungsgesetz ungültig.

(Quelle: [http://de.wikipedia.org/wiki/Snelliussches\\_Brechungsgesetz](http://de.wikipedia.org/wiki/Snelliussches_Brechungsgesetz))



Daraus ergibt sich das Brechungsgesetz:  $\frac{\sin \hat{\alpha}_1}{\sin \hat{\alpha}_2} = \frac{n_2}{n_1}$ .

Dabei bezeichnen  $n_1; n_2$  die Brechungsindizes, auch die Brechzahlen genannt. Sie sind eine optische Materialeigenschaft. Diese physikalische dimensionslose Größe gibt an, um welchen Faktor die Wellenlänge und die Phasengeschwindigkeit des Lichts kleiner sind als im Vakuum. In Luft ist der Brechungsindex ziemlich genau 1.

(Quelle: <http://de.wikipedia.org/wiki/Brechungsindex>)

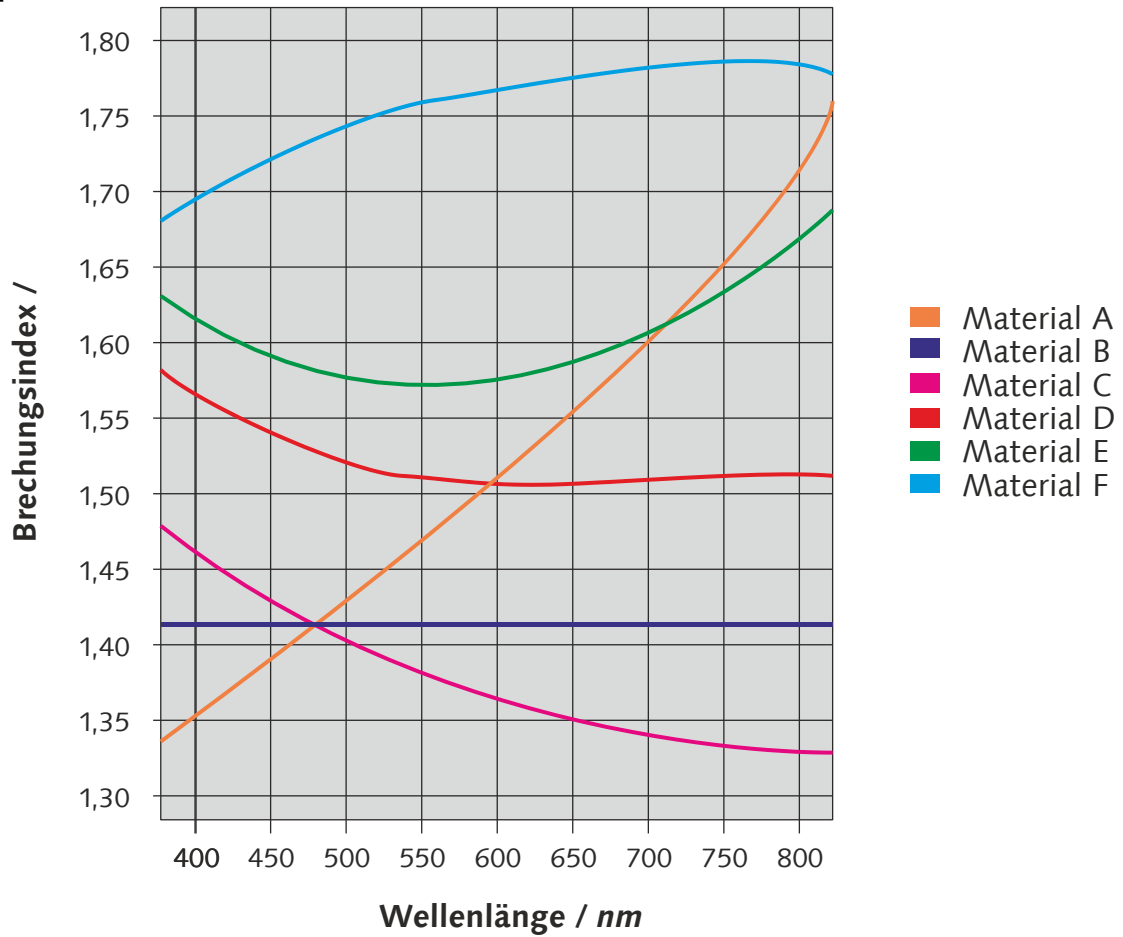
## Prisma

Als Prismen bezeichnet man in der Optik Körper die einem geometrischen Prisma entsprechen und für Licht des sichtbaren Spektrums durchlässig sind. Mit ihnen lässt sich Licht bzgl. seiner Frequenz aufspalten oder linear polarisieren.

## Dispersion

Die Aufspaltung des Lichts bzgl. seiner Frequenz basiert auf der Abhängigkeit der Ausbreitungsgeschwindigkeit von der Frequenz. Dieses Phänomen wird Dispersion genannt. Entsprechend ist der Brechungsindex eines Prismas Abhängig von der Frequenz der elektromagnetischen Welle.

### Dispersionskurve verschiedener Materialien



### Verschiedenes

1 l = 0,001 m<sup>3</sup>

1 m = 1000 mm = 1000000000 nm

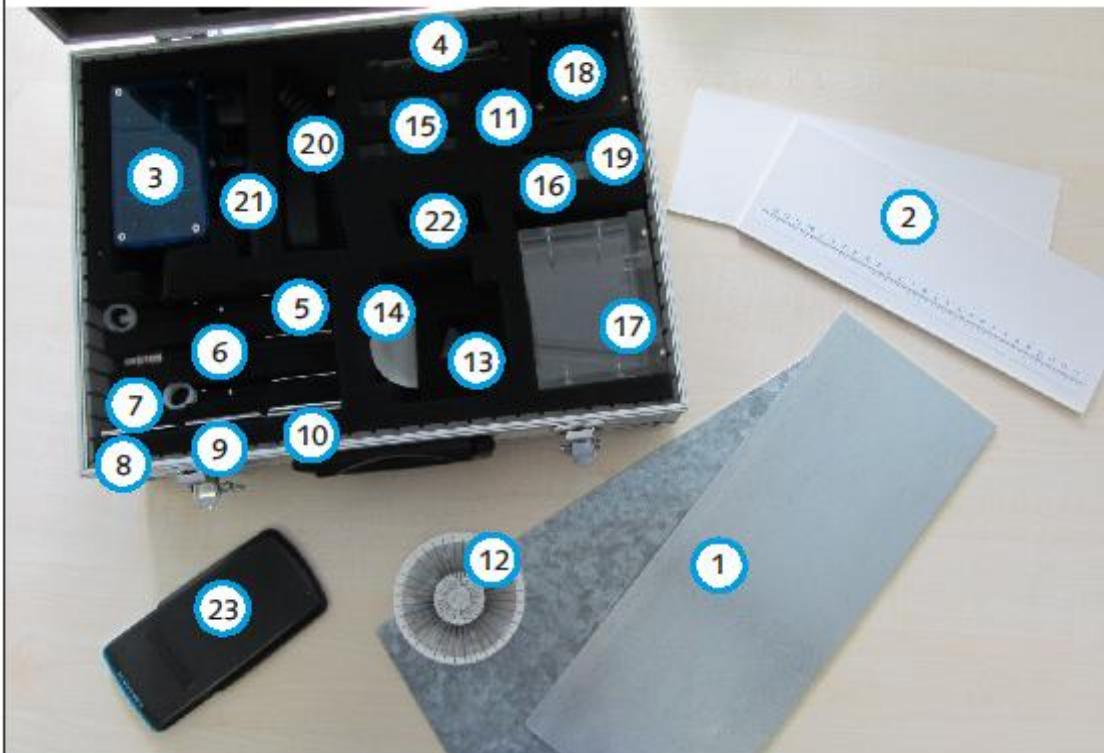
1m = 10<sup>3</sup> mm = 10<sup>9</sup>nm

## A.4.2 Inventarliste

Erfassung von Experimentierkompetenz

Inventar

### Inventarliste und Hinweise zu den Geräten




- |     |   |     |                                       |
|-----|---|-----|---------------------------------------|
| 1.  | Arbeitsfläche                             | 13. | gleichschenkliges Prisma (Material B) |
| 2.  | Schirm                                    | 14. | halbkreisförmiges Prisma (unbekannt)  |
| 3.  | thermische Lichtquelle                    | 15. | Polarisationsfilter                   |
| 4.  | Halterung Laserpointer                    | 16. | bikonvexe Linse ( $f=50\text{mm}$ )   |
| 5.  | Laserpointer (unbekannt)                  | 17. | Kuvette                               |
| 6.  | Laserpointer (532 nm)                     | 18. | Blackbox (Aufgabe 6)                  |
| 7.  | Laserpointer (408 nm)                     | 19. | Blackbox (Nachbau Bodenplatte)        |
| 8.  | optisches Gitter                          | 20. | Luxmeter                              |
| 9.  | Pinhole                                   | 21. | Halterung Luxmeter                    |
| 10. | Farbfilter                                | 22. | Maßband                               |
| 11. | universal Halterung (für Schirm und Dias) | 23. | Taschenrechner                        |
| 12. | Prismatisch                               |     |                                       |



## Inventar

## Erfassung von Experimentierkompetenz


1. Arbeitsfläche  
Metallplatten mit rutschfester Unterlage.




2. Schirm  
Mit Millimeter Skala.  
Nutzbar mit Halterung (11).



3. thermische Lichtquelle  
Glühlampe mit 6 Volt,  
frei regelbar.  
Integrierte Magnethalterung.



4. Halterung Laserpointer  
Verstellbare Höhe und Winkel  
durch Schrauben. Integrierte  
Magnethalterung.




5-7. Laserpointer  
Dauerbetrieb-Stellung  
Wellenlänge: 408nm;  
532nm;  
unbekannt



8-10. Dia  
Gitter  
Pinhole  
Farbfilter  
nutzbar mit Halterung (11).




11. universal Halterung  
integrierte Magnete.  
Für Schirm und Dia.




12. Prismatisch  
integrierte Magnete  
und Winkelskala.  
13-14. Prisma  
Dreieck (Material B)  
Halbkreis (unbekannt)




15. Polarisationsfilter  
integrierte Magnethalterung  
und Winkelskala.



16. Linse  
integrierte Magnethalterung.  
Bikonvex,  $f=50\text{mm}$ .




17. Küvette  
integrierte Magnethalterung.




18. Blackbox (Aufgabe 6)  
integrierte Magnete.  
19. Blackbox  
nachbau Bodenplatte  
mit integrierten Magneten.



20. Luxmeter  
21. Halterung Luxmeter  
integrierte Magnete.



22. Maßband  
23. Taschenrechner



## Erfassung von experimenteller Kompetenz im Bereich Optik

Bitte mache die folgenden Angaben:

--	--	--	--	--	--	--	--	--

Erster Buchstabe des Vornamens

Letzter Buchstabe des Nachnamens

Dein Geburtsdatum  
TTMMYY

Beispiel:

J	H		2	9	0	4	8	3
---	---	--	---	---	---	---	---	---

**LiebeStudentin, lieber Student,**

in den nächsten 210 Minuten werden Sie einTestheft zum Thema Optik bearbeiten. Es besteht aus drei Teilen. Mit dem ersten Teil werden wir allgemeine Daten zu Ihrer Person und Ihrem Fortschritt im Studium erfassen. Im zweiten Teil werden Ihre Fähigkeiten beim Experimentieren getestet und im dritten Teil finden sich Fragestellungen zur Erfassung Ihres Fachwissens im Bereich Optik.

**Vielen Dank für Ihre Mitarbeit!**

Jan Heidrich

# 1. Allgemeine Daten zur Person

In diesem Abschnitt möchten wir einige allgemeine Daten zu Ihrer Person und Ihrem Fortgang im Studium erfragen. Dazu sollen Sie an entsprechenden Stellen Kreuze setzen oder in Kästchen entsprechende Angaben machen.

Beispiel:  
Wie alt sind Sie?

Haben Sie sich verschrieben, so streichen Sie ihre Antwort und notieren rechts neben dem Kästchen die korrekte Antwort.

Beispiel:  
Bearbeiten Sie gerade ein Testheft?  Ja  Nein

Sollten Sie versehentlich eine falsche Antwort gewählt haben, so malen Sie das Kästchen mit der falschen Antwort aus und Kreuzen die richtige Antwort an.

Beispiel:  
Bearbeiten Sie gerade ein Testheft?  Ja  Nein

**1.1) Sind Sie weiblich oder männlich?**

weiblich       männlich

**1.2) Wie lautet die genaue Bezeichnung Ihres Studienganges?**

**1.3) In welchem Fachsemester sind sie?**

**1.4) Welche Noten hatten Sie in bisherigen Experimentalphysik Vorlesungen?**  
(Markieren Sie bisher nicht besuchte Experimentalphysik Vorlesungen mit einem "--".)

Experimentalphysik I

Experimentalphysik II

Experimentalphysik III

Experimentalphysik IV

**1.5) Wie alt sind sie?**

**1.6) Haben Sie eine angefangene oder abgeschlossene technische Ausbildung neben dem Abitur und dem Studium?**

ja       nein

**1.6.1) Wenn Sie eine angefangene oder abgeschlossene technische Ausbildung haben, geben Sie bitte die Bezeichnung dieser Ausbildung an.**

## 2. Experimentierfähigkeiten im Bereich Optik

In diesem Abschnitt möchten wir Ihre Fähigkeiten beim Experimentieren erfassen. Dazu werden offene Aufgaben verwendet.

Bei **offenen Aufgaben** werden keine Antwortalternativen vorgegeben und Sie werden aufgefordert etwas zu erklären bzw. zu skizzieren. Ihre Antwort sollte in den dafür vorgesehenen Kästen notiert werden.

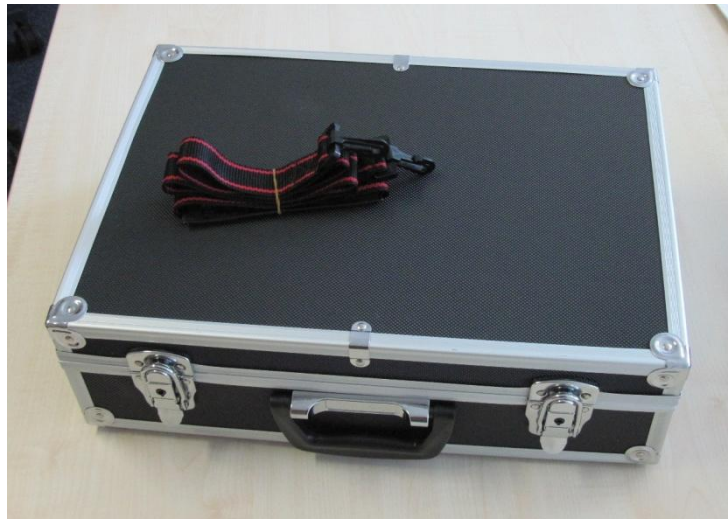
Beispiel:

Beschreiben Sie das Wetter am heutigen Testtag.

*Heute Morgen hat die Sonne geschienen, jetzt beginnt es zu regnen.*

Bei der Bearbeitung dieser Aufgaben werden sie gefilmt, um den Prozess des Experimentierens erfassen zu können. Dazu werden zwei Kameras genutzt: eine Kamera filmt Sie und den Versuchsplatz, die andere Kamera ist auf ihre Schreibunterlage gerichtet. Damit wir zuordnen können wann Sie welche Aufzeichnungen machen ist es erforderlich, dass Sie ausschließlich auf der Schreibunterlage zeichnen bzw. schreiben. Um dies nicht zu vergessen belassen Sie das Testheft bitte immer auf der Schreibunterlage.

Für die Bearbeitung aller Experimentieraufgaben steht Ihnen die bereits bekannte Experimentierbox zur Verfügung.



In dieser finden Sie diverse optische Materialien sowie Halterungen mit magnetischen Füßen, eine Inventarliste mit technischen Daten zu den Komponenten, kurze theoretische Grundlagen zur Optik mit Formeln und Metallplatten als feste Experimentierunterlage.



Mit dem Material aus der Experimentierbox lassen sich alle Aufgaben auf verschiedene Weisen bearbeiten und lösen. Es ist ausreichend, wenn Sie einen möglichen Lösungsweg notieren.

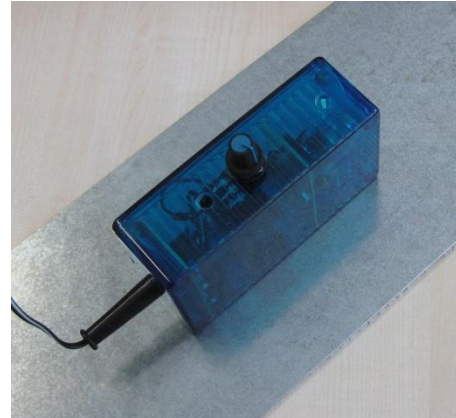
Am Ende jedes Lösungskastens finden Sie eine erwartete Lösung. Wenn Sie nicht genau wissen was Sie machen sollen oder wenn Sie nicht weiter kommen, so kann Blick auf diese erwartete Lösung hilfreich sein.

**Wichtiger Hinweis:**

Sie werden mit Lasern der Laserklasse 1 arbeiten. Achten Sie darauf nicht in den Laserstrahl zu schauen. Legen Sie Schmuck und Uhren ab, um den Laserstrahl nicht ungewollten zu reflektieren.

### 2.1) Vermessung einer Glühwendel

Mit Hilfe von Linsen, Spiegeln oder Lochblenden lassen sich Objekte abbilden. Dabei bestimmt die Anordnung der Komponenten das Verhältnis der Größen von Gegenstand und Bild. Damit lassen sich Gegenstände vermessen.

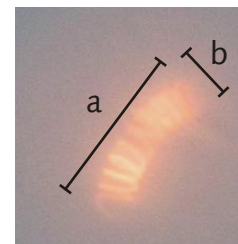


**Bestimmen Sie experimentell die Abmessung der Glühwendel der Lampe (Inventarnr. 3) ohne sich Gefahr auszusetzen oder Versuchsmaterial zu beschädigen.**

Platz für Ihre Lösung



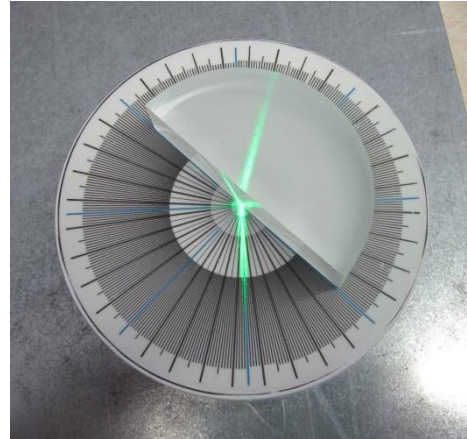
**Lösung:** Die Glühwendel hat eine Länge von  $a =$  \_\_\_\_\_ und eine Breite von  $b =$  \_\_\_\_\_.



## 2.2) Optische Dichte

Trifft Licht auf eine optische Grenzschicht (zum Beispiel Luft zu Wasser) so wird es teilweise reflektiert und teilweise transmittiert. Für den transmittierten Anteil lässt sich eine Veränderung der Ausbreitungsrichtung feststellen. Diese Veränderung hängt von der Wellenlänge des Lichts und der optischen Dichte der beiden Medien ab.

**Identifizieren Sie das Material (siehe Dispersionsgraphen) aus dem das Prisma (Inventarnr. 14) besteht.**

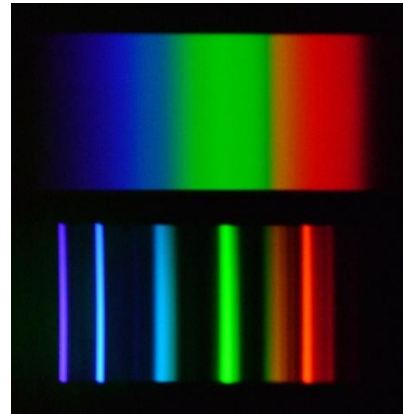


Platz für Ihre Lösung

**Lösung:** Das Prisma (Inventarnr. 14) besteht aus folgendem Material: \_\_\_\_\_.

### 2.3) Spektroskopie

Jede elektromagnetische Welle hat eine Frequenz die charakteristisch für die Welle ist. Liegt die Frequenz im vom menschlichen Auge wahrnehmbaren Bereich, so kann dem Licht eine Farbe zugeordnet werden. Überlagern sich elektromagnetische Wellen verschiedener Frequenzen so lässt sich Licht in einer Farben beobachten. Mit Hilfe von Gittern oder Prismen lassen sich überlagerte elektromagnetische Wellen bzgl. ihrer Frequenz aufspalten. Dabei entstehen Spektren die sich nutzen lassen um Stoffe oder einzelne Wellenlängen zu identifizieren.



**Bestimmen Sie experimentell die Wellenlänge des unbekanntes Lasers (Inventarnr. 5).**

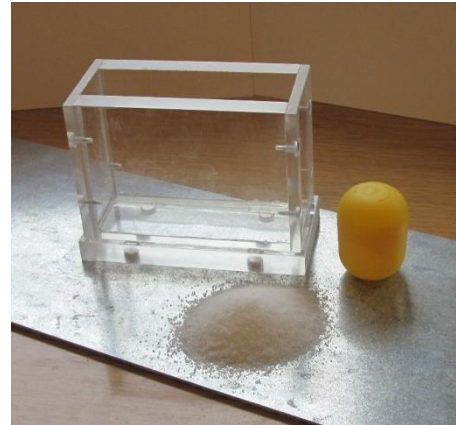
Platz für Ihre Lösung

**Lösung:** Die Wellenlänge des unbekanntes Lasers (Inventarnr. 5) beträgt: \_\_\_\_\_.

#### 2.4) Optische Aktivität von Zuckerwasser

Licht ist eine elektromagnetische Welle. Wenn die Schwingungsebenen der elektrischen Felder aller elektromagnetischen Wellen eines Lichtbündels zu jeder Zeit parallel zueinander liegen, so ist das Lichtbündel polarisiert. Sind die Schwingungsebenen zeitlich und räumlich konstant wenn es zu keiner Wechselwirkung mit Materie kommt, spricht man von linearer Polarisation.

Durch optisch aktive Medien (beispielsweise zuckerhaltiges Wasser) lässt sich die Schwingungsebene linear polarisierten Lichts ändern.



**Bestimmen Sie experimentell bei welcher Zuckerkonzentration (gestrichene Löffel pro Liter) des Wassers die Schwingungsebene des elektrischen Feldes um  $0,75^\circ/\text{cm}$  gedreht wird.**

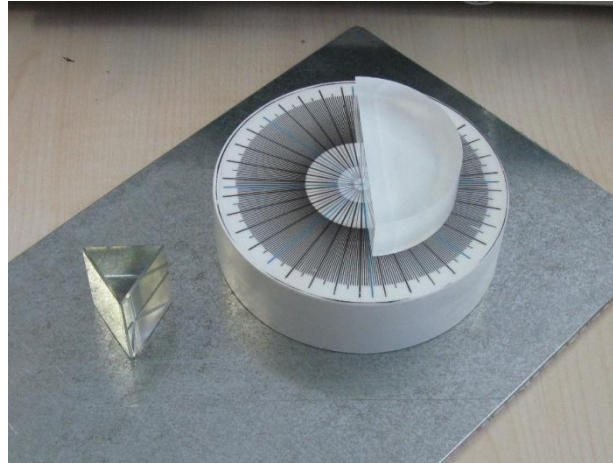
Platz für Ihre Lösung

**Lösung:** Bei einer Zuckerkonzentration von \_\_\_\_\_ gestrichenen Löffeln Zucker pro Liter Wasser wird die Schwingungsebene der elektrischen Felder um  $0,75^\circ/\text{cm}$  gedreht.

### 2.5) Polarisations-eigenschaften eines Prismas

Licht ist eine elektromagnetische Welle. Wenn die Schwingungsebenen der elektrischen Felder aller elektromagnetischen Wellen eines Lichtbündels zu jeder Zeit parallel zueinander liegen, so ist das Lichtbündel polarisiert. Sind die Schwingungsebenen zeitlich und räumlich konstant wenn es zu keiner Wechselwirkung mit Materie kommt, spricht man von linearer Polarisation.

Es gibt verschiedene Möglichkeiten linear polarisiertes Licht zu erzeugen.



**Trifft Licht in einem bestimmten Winkel auf ein Prisma, so lässt sich das Licht linear polarisieren. Bestimmen Sie experimentell diesen Winkel und geben Sie die Schwingungsebene des transmittierten und des reflektierten Lichts an.**

Platz für Ihre Lösung

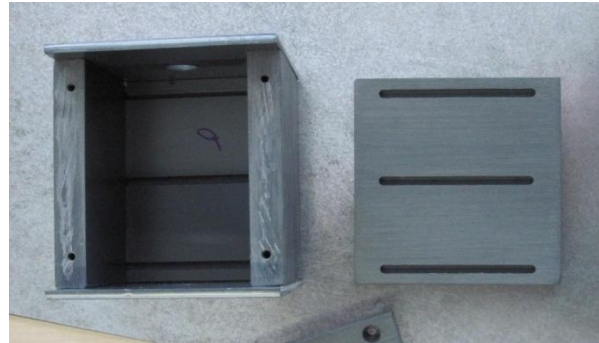


**Lösung:** Der Winkel unter dem Licht am stärksten polarisiert wird beträgt \_\_\_\_\_.  
Die Lage des elektrischen Feldes eines reflektierten Strahls hat bzgl. der Polarisationsfilter einen Winkel von \_\_\_\_\_.  
Die Lage des elektrischen Feldes eines reflektierten Strahls hat bzgl. der Polarisationsfilter einen Winkel von \_\_\_\_\_.

### 2.6) Identifikation optischer Eigenschaften von unbekannten Bauteilen

Nicht immer sind die optischen Eigenschaften von Objekten bekannt oder direkt zu erkennen.

Für die Nutzung dieser Objekte kann es aber notwendig sein diese Eigenschaften zu kennen. Deshalb werden in der Materialkunde die optischen Eigenschaften gezielt untersucht, teilweise werden die Objekte dazu nachgebaut.



**Bestimmen Sie experimentell die optischen Eigenschaften des unbekannten Bauteils.**

Platz für Ihre Lösung

**Lösung:** Im ersten Slot der Blackbox befindet sich \_\_\_\_\_, im zweiten Slot befindet sich \_\_\_\_\_ und im dritten Slot befindet sich \_\_\_\_\_.

**Vielen Dank**  
**für Ihre Mitarbeit.**

## **Erfassung von Experimentierkompetenz im Bereich Optik**

**Bitte mache die folgenden Angaben:**

**Nachname**

**Vorname**

**Liebe Studentin, lieber Student,**

in den nächsten 240 Minuten werden Sie einen Test zum Thema Optik bearbeiten. Dieser besteht aus drei Teilen. Mit dem ersten Teil werden wir allgemeine Daten zu Ihrer Person und Ihrem Fortschritt im Studium erfassen. Im zweiten Teil werden Ihre Fähigkeiten beim Experimentieren getestet und im dritten Teil finden sich Fragestellungen zu ihrem Fachwissen, ihrer Selbstwirksamkeitserwartung und ihrem Strategiewissen.

**Vielen Dank für Ihre Mitarbeit!**

Jan Heidrich

## 2. Allgemeine Daten zur Person

In diesem Abschnitt möchten wir einige allgemeine Daten zu Ihrer Person und Ihrem Fortgang im Studium erfragen. Dazu sollen Sie an entsprechenden Stellen Kreuze setzen oder in Kästchen entsprechende Angaben machen.

Beispiel:  
Wie alt sind Sie?

Haben Sie sich verschrieben, so streichen Sie ihre Antwort und notieren rechts neben dem Kästchen die korrekte Antwort.

Beispiel:  
Bearbeiten Sie gerade ein Testheft?  Ja  Nein

Sollten Sie versehentlich eine falsche Antwort gewählt haben, so malen Sie das Kästchen mit der falschen Antwort aus und Kreuzen die richtige Antwort an.

Beispiel:  
Bearbeiten Sie gerade ein Testheft?  Ja  Nein

**1.7) Sind Sie weiblich oder männlich?**

weiblich       männlich

**1.8) Wie lautet die genaue Bezeichnung Ihres Studienganges?**

**1.9) In welchem Fachsemester sind sie?**

**1.10) Welche Noten hatten Sie in bisherigen Experimentalphysik Vorlesungen?**  
(Markieren Sie bisher nicht besuchte Experimentalphysik Vorlesungen mit einem "--".)

Experimentalphysik I

Experimentalphysik II

Experimentalphysik III

Experimentalphysik IV

**1.11) Wie alt sind sie?**

**1.12) Haben Sie eine angefangene oder abgeschlossene technische Ausbildung neben dem Abitur und dem Studium?**

ja       nein

**2.6.1) Wenn Sie eine angefangene oder abgeschlossene technische Ausbildung haben, geben Sie bitte die Bezeichnung dieser Ausbildung an.**



### 3. Experimentierfähigkeiten im Bereich Optik

In diesem Abschnitt möchten wir Ihre Fähigkeiten beim Experimentieren erfassen. Dazu werden offene Aufgaben verwendet.

Bei **offenen Aufgaben** werden keine Antwortalternativen vorgegeben und Sie werden aufgefordert etwas zu erklären bzw. zu skizzieren. Ihre Antwort sollte in den dafür vorgesehenen Kästen notiert werden.

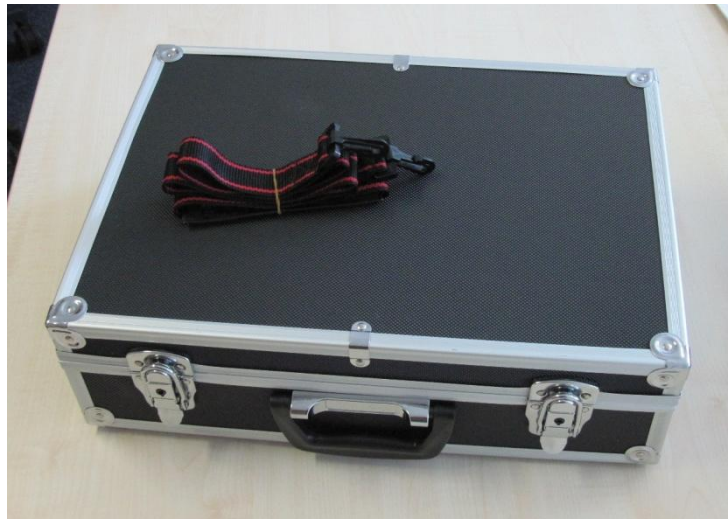
Beispiel:

Beschreiben Sie das Wetter am heutigen Testtag.

*Heute Morgen hat die Sonne geschienen, jetzt beginnt es zu regnen.*

Bei der Bearbeitung dieser Aufgaben werden sie gefilmt, um den Prozess des Experimentierens erfassen zu können. Dazu werden zwei Kameras genutzt: eine Kamera filmt Sie und den Versuchsplatz, die andere Kamera ist auf ihre Schreibunterlage gerichtet. Damit wir zuordnen können wann Sie welche Aufzeichnungen machen ist es erforderlich, dass Sie ausschließlich auf der Schreibunterlage zeichnen bzw. schreiben. Um dies nicht zu vergessen belassen Sie das Testheft bitte immer auf der Schreibunterlage.

Für die Bearbeitung aller Experimentieraufgaben steht Ihnen die bereits bekannte Experimentierbox zur Verfügung.



In dieser finden Sie diverse optische Materialien sowie Halterungen mit magnetischen Füßen, eine Inventarliste mit technischen Daten zu den Komponenten, kurze theoretische Grundlagen zur Optik mit Formeln und Metallplatten als feste Experimentierunterlage.



Mit dem Material aus der Experimentierbox lassen sich alle Aufgaben auf verschiedene Weisen bearbeiten und lösen. Es ist ausreichend, wenn Sie einen möglichen Lösungsweg notieren.

**Zur Bearbeitung der Aufgaben erhalten Sie beliebig viele Protokollbögen. Dokumentieren Sie ihren Lösungsprozess so, dass andere mit ihren Aufzeichnungen den Lösungsweg nachvollziehen können und zu denselben Ergebnissen kommen würden.**

**Wichtiger Hinweis:**

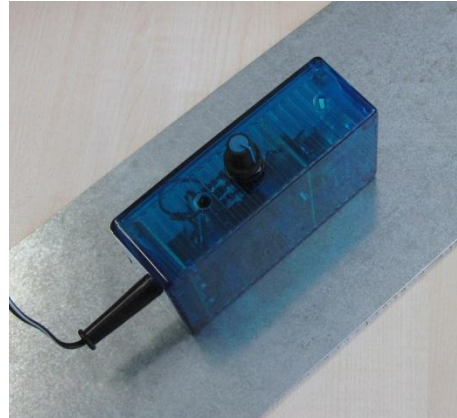
Sie werden mit Lasern der Laserklasse 1 arbeiten. Achten Sie darauf nicht in den Laserstrahl zu schauen. Legen Sie Schmuck und Uhren ab, um den Laserstrahl nicht ungewollten zu reflektieren.

## 2.1) Vermessung einer Glühwendel

Mit Hilfe von Linsen, Spiegeln oder Lochblenden lassen sich Objekte abbilden. Dabei bestimmt die Anordnung der Komponenten das Verhältnis der Größen von Gegenstand und Bild.

Damit lassen sich Gegenstände vermessen.

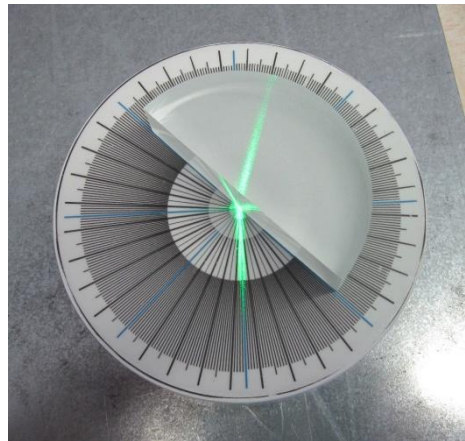
**Bestimmen Sie experimentell die Abmessung (Länge und Breite) der Glühwendel der Lampe (Inventarnr. 3) ohne sich Gefahr auszusetzen oder Versuchsmaterial zu beschädigen.**



## 2.2) Optische Dichte

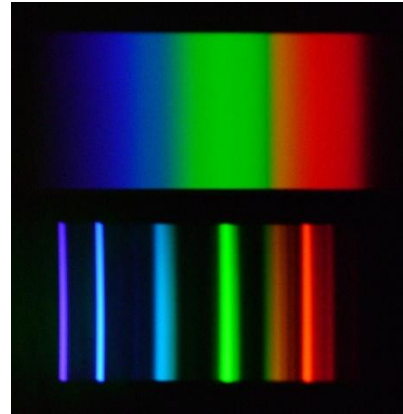
Trifft Licht auf eine optische Grenzschicht (zum Beispiel Luft zu Wasser) so wird es teilweise reflektiert und teilweise transmittiert. Für den transmittierten Anteil lässt sich eine Veränderung der Ausbreitungsrichtung feststellen. Diese Veränderung hängt von der Wellenlänge des Lichts und der optischen Dichte der beiden Medien ab.

Identifizieren Sie das Material (siehe Dispersionsgraphen) aus dem das Prisma (Inventarnr. 14) besteht und diskutieren Sie den Fehler der Messung.



### 2.3) Spektroskopie

Jede elektromagnetische Welle hat eine Frequenz die charakteristisch für die Welle ist. Liegt die Frequenz im vom menschlichen Auge wahrnehmbaren Bereich, so kann dem Licht eine Farbe zugeordnet werden. Überlagern sich elektromagnetische Wellen verschiedener Frequenzen so lässt sich Licht in einer Farben beobachten. Mit Hilfe von Gittern oder Prismen lassen sich überlagerte elektromagnetische Wellen bzgl. ihrer Frequenz aufspalten. Dabei entstehen Spektren die sich nutzen lassen um Stoffe oder einzelne Wellenlängen zu identifizieren.

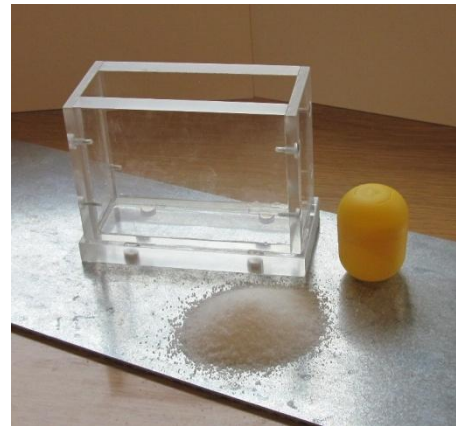


**Bestimmen Sie experimentell die Wellenlänge des unbekanntens Lasers (Inventarnr. 5).**

## 2.4) Optische Aktivität von Zuckerwasser

Licht ist eine elektromagnetische Welle. Wenn die Schwingungsebenen der elektrischen Felder aller elektromagnetischen Wellen eines Lichtbündels zu jeder Zeit parallel zueinander liegen, so ist das Lichtbündel polarisiert. Sind die Schwingungsebenen zeitlich und räumlich konstant wenn es zu keiner Wechselwirkung mit Materie kommt, spricht man von linearer Polarisation.

Durch optisch aktive Medien (beispielsweise zuckerhaltiges Wasser) lässt sich die Schwingungsebene linear polarisierten Lichts ändern.

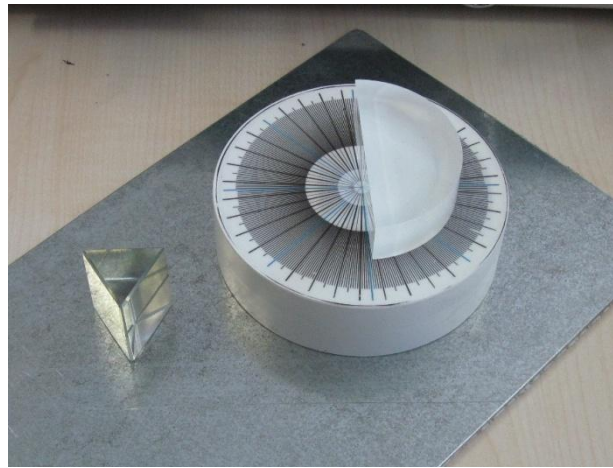


**Bestimmen Sie experimentell bei welcher Zuckerkonzentration (in der Einheit: gestrichene Löffel pro Liter) des Wassers die Schwingungsebene des elektrischen Feldes um  $0,75^\circ/\text{cm}$  gedreht wird.**

## 2.5) Polarisations-eigenschaften eines Prismas

Licht ist eine elektromagnetische Welle. Wenn die Schwingungsebenen der elektrischen Felder aller elektromagnetischen Wellen eines Lichtbündels zu jeder Zeit parallel zueinander liegen, so ist das Lichtbündel polarisiert. Sind die Schwingungsebenen zeitlich und räumlich konstant wenn es zu keiner Wechselwirkung mit Materie kommt, spricht man von linearer Polarisation.

Es gibt verschiedene Möglichkeiten linear polarisiertes Licht zu erzeugen.

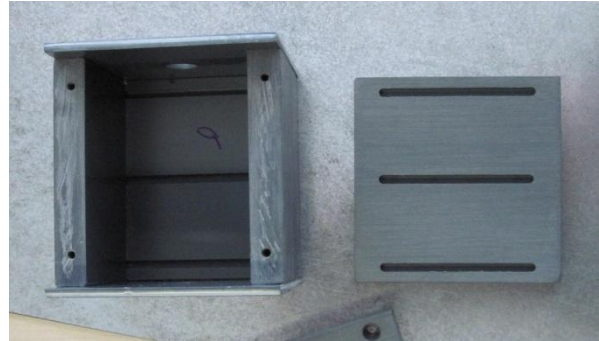


**Trifft Licht in einem bestimmten Winkel (dem Brewsterwinkel) auf ein Prisma, so lässt sich das Licht vollständig linear polarisieren. Bestimmen Sie experimentell wie dieser Winkel ist und geben Sie an in welchem Verhältnis der transmittierte- und reflektierte Anteil des Lichts zueinander polarisiert sind.**

## 2.6) Identifikation optischer Eigenschaften von unbekanntem Bauteilen

Nicht immer sind die optischen Eigenschaften von Objekten bekannt oder direkt zu erkennen.

Für die Nutzung dieser Objekte kann es aber notwendig sein diese Eigenschaften zu kennen. Deshalb werden in der Materialkunde die optischen Eigenschaften gezielt untersucht, teilweise werden die Objekte dazu nachgebaut.



**Bestimmen Sie experimentell welche optischen Komponenten in dem unbekanntem Bauteil sind.**

**Sammeln Sie ausreichend Informationen, um das optische Bauteil nachbauen zu können.**



**Vielen Dank**  
**für Ihre *Mitarbeit.***

## **A.4.5 Laborheft-App Konzept**

### **1) Ziele der Erfassung exp. Kompetenz mittels Labor-App (Jan Heidrich)**

Die Labor-App soll eine Erweiterung der Auswertung basierend auf dem Laborheft sein. Bei der Think Aloud Studie konnte gezeigt werden, dass die Handlungen „Fragestellung klären“ und „Experiment aufbauen“ mit einem Laborheft nicht abgebildet werden können und der Score für alle Aspekte der Experimentierkompetenz entsprechend verfälscht wird. Durch die Labor-App sollen diese Handlungen abgebildet werden können ohne den Aufwand einer Videostudie zu haben (Fragestellung über die App. abrufbar / Fotofunktion). Außerdem konnte mit der Think Aloud Studie gezeigt werden, dass zur Erfassung des Aspekts Zielorientiertheit Informationen fehlen. Dies gilt sowohl für die Analyse von Videos als auch vom Laborheft. Dies zeigt sich vor allem, wenn der Proband nicht weiß, wie er weiter vorgehen soll. Die Handlungen wechseln dann mitunter schnell und sind nicht identifizierbar. Mit der Labor-App sollen diese Informationen identifizierbar gemacht werden (automatische Speicherung aller Handlungen / Möglichkeit Probleme zu benennen und Lösungswege zu skizzieren). Zielgruppe sind Studierende im Anfängerpraktikum.

### **2) Ziel der Erfassung exp. Kompetenz mittels Labor-App (Martin Schwichow)**

Die Labor-App soll im Rahmen meines Promotionsprojektes als ein Testverfahren zur Überprüfung der Wirkung einer Intervention bezüglich der VKS genutzt werden. Zielpopulation sind SuS der 7. Jahrgangsstufe an Gemeinschaftsschulen des Landes Schleswig-Holstein. Wie die Ergebnisse meiner Meta-Analyse gezeigt haben, hat das Testformat einen nicht zu vernachlässigenden Einfluss auf den gemessenen Interventionserfolg. Daher sind zur Überprüfung der Wirkung von Interventionen mehrere Testinstrumente unterschiedlichen Formats zu wählen um eine Über- bzw. Unterschätzung zu vermeiden. Neben dem Einsatz klassischer Papier- Bleistift-Tests sollen die Probanden mit praktischen experimentellen Aufgaben getestet werden. Der Einsatz der Labor-App eignet sich besonders, da sie nicht nur die Fähigkeit der SuS zur Kontrolle von Variablen, sondern weitere Teilkompetenzen experimenteller Kompetenz erfasst. Es kann somit die Interventionswirkung breiter erfasst werden. Von besonderem Interesse ist, ob sich eine Intervention bezüglich der VKS positive auf die Fähigkeit der SuS auswirkt Hypothesen zu bilden, Experimente zu Planen und Schlussfolgerungen aus diesen Experimenten zu ziehen.

### **3) Aufbau der Labor-App**

Die Labor-App soll SuS einerseits bei der Planung, Durchführung und Auswertung von Experimenten unterstützen und andererseits ihr experimentelles Vorgehen aufzeichnen und für spätere fachdidaktische Analysen zur Verfügung stellen. Ziel ist es die SuS minimal anzuleiten und gleichzeitig möglichst viele Informationen über ihr Vorgehen bei der Bearbeitung von experimentellen Aufgaben zu bekommen. Die App soll für unterschiedliche Aufgaben durch minimale Anpassungen (z.B. der Theorihilfen) nutzbar sein.

Die SuS können die unterschiedlichen Handlungsoptionen auswählen und nutzen um ein Laborprotokoll zu erstellen. Das Protokoll ist vollständig mittels der App zu erstellen. Die Auswahl der Handlungsoptionen sollte soweit wie möglich keinen Restriktionen unterliegen. Der Bildschirm gliedert sich in die zwei Teilaspekte Handlungsoptionen und Laborprotokoll. Die Aufgabensteller können Anpassungen über die Bearbeitung einer „Master-Ebene“ vornehmen.

#### **a. Handlungsoptionen**

Die folgende Liste enthält alle Handlungen, die den Probanden zur Bearbeitung der experimentellen Aufgabenstellung zur Verfügung stehen. Die Darstellung der Handlungen sollte ein möglichst intuitives Erkennen ihrer Funktion ermöglichen. Die einzelnen Handlungen sollen ohne Vorbedingungen auszuwählen sein.

<b>Handlung</b>	<b>Beschreibung</b>	
Textfeld	Eingabe von Texten. Texte können entweder Fragestellungen, Beschreibungen des experimentellen Vorgehens oder Schlussfolgerungen beinhalten. Einzugebende Texte werden wahrscheinlich nicht von „ausufernder“ Länge sein. Eingabe soll über ein Tastenfeld in englischer und deutscher Sprache möglich sein. Es besteht die Möglichkeit Handschriftliche Aufzeichnungen zu machen.	
Foto	Aufnehmen von Fotos mit der im Endgerät integrierten Kamera. Fotos sollen entweder als Dokumentation des Versuchsaufbaus oder als Ergebnisdokumentation nutzbar sein. Fotos sollen in dem Laborheft frei verschiebbar sein.	
Film	Aufnehmen von kurzen Filmen mit der im Endgerät integrierten Kamera. Audiospur sollte ebenfalls aufgenommen werden.	
Tabelle	Tabellen sollen die Aufnahme von Messwerten ermöglichen. Bezeichnungen der einzelnen Spalten sollten durch die Probanden erfolgen. Erst einmal keine Tabellenkalkulation, da langen Messreihen nicht zu erwarten sind. Probanden können jedoch die Taschenrechnerfunktion nutzen um neue Werte aus den Messwerten „per Hand“ zu berechnen. Die Tabellen sollen als Grundlage für die Graphikfunktion dienen.	
Graphen	Zeichnen von einfachen Abhängigkeiten. Probanden können Spalten, die als X-, bzw. Y-Achse dienen sollen aus bisher aufgenommenen Tabellen wählen. Wünschenswerte Zusatzfunktion: Einzeichnen von Ausgleichsgeraden (linearen Zusammenhänge), sowie der zugehörigen Gradengleichung.	
Taschenrechner	Taschenrechner, der neben Grundrechenarten, Potenzen (Wurzeln) auch trigonometrische exponentielle Funktionen beherrscht. Der Rechner soll jederzeit und unabhängig von der gewählten Handlungsoption genutzt werden können. Probanden sollen Rechenergebnisse in Textfelder oder Tabellen kopieren können.	
Skizze	Zeichnen von Skizzen, die wie Fotos zum Festhalten des Versuchsaufbaus oder schematischen Darstellung von Erkenntnissen dienen kann. Skizzen sollen in jedes Feld des Laborbuches kopiert werden können. Als Zeichenhilfe ist der Einsatz eines Stiftes denkbar	
Aufbauen	Dieses Feld soll keine weiteren Funktionen ermöglichen. Es erfolgt nur eine Eintrag in das Auswertungsskript. Diese Funktion soll helfen die nicht App gebundene Handlungen der Probanden festzuhalten. Um diese Handlung abzuschließen und die App wieder normal nutzen zu können, muss der Proband ein Foto aufnehmen.	
Theorie	Zugriff auf pdf-Dokumente die vom Aufgabensteller zur Verfügung gestellt werden und die den theoretischen Hintergrund bzw. praktische Hilfestellungen anbieten.	
Hilfe	Allgemeine Hilfefunktion, welche das Ziel von Experimenten und den Aufbau von Laborheften erklärt.	
Aufgabenstellung	Textfeld mit der Aufgabenstellung für die jeweilige Aufgabe.	





## **A.4.6 Kodiermanual produktorientierte Auswertung**

### **Produktorientierte Auswertung von Experimentierkompetenz**

#### **1. Einleitung**

Das vorliegende Manual dient der Auswertung der Aufzeichnungen, die Probanden bei der Bearbeitung von sechs Experimentieraufgaben angefertigt haben. Die Aufzeichnungen sind vergleichbar mit solchen, wie sie z.B. im Physikalischen Praktikum für Physiker entstehen. Anhand der Aufzeichnungen sollen mit Hilfe des vorliegenden Manuals die Handlungen, die beim Experimentieren durchgeführt werden, identifiziert und hinsichtlich ihrer Richtigkeit bewertet werden.

#### **2. Grundlagen**

Experimentieren ist ein komplexer Prozess, beim dem unterschiedliche Fähigkeiten und Fertigkeiten eine Rolle spielen. In physik- bzw. naturwissenschaftsdidaktischer Literatur hat sich eine Gliederung entsprechend der Handlungen beim idealtypischen Experimentieren als potentiell iterativer Prozess etabliert. Experimentieren beginnt idealtypisch mit der Planung eines Experiments. Das heißt, aus theoretischen Überlegungen oder empirischen Beobachtungen wird eine Fragestellung entwickelt (1) und es werden dazu passende fachliche Hypothesen gebildet (2). Anschließend wird ein geeignetes Experiment geplant (3), mit dem sich die Frage beantworten lässt. Auf die Planungsphase folgt die Durchführung. Im Rahmen der Durchführung wird zunächst ein Versuch aufgebaut (4) und anschließend wird gezielt gemessen und beobachtet (5). Den Abschluss des idealtypischen Experimentierens bildet die Auswertung. Dabei werden Daten analysiert (6) indem Berechnungen auf Basis der Messwerte durchgeführt oder Graphen angefertigt werden. Anschließend werden sachgerechte Schlüsse (7) gezogen, indem die Beobachtungen und Ergebnisse der Analyse von Daten auf die Fragestellung bezogen werden.

Diese Handlungen bilden die Grundlage für die Beschreibung von Experimentierkompetenz. Experimentierkompetenz umfasst demnach die Fähigkeiten und Fertigkeiten die entsprechenden Handlungen durchführen zu können. Die Ausprägung von Experimentierkompetenz lässt sich dann anhand von drei Kriterien beschreiben: 1) Richtigkeit: Erfolgreiches Experimentieren setzt die richtige Durchführung der einzelnen Handlungen voraus. 2) Strukturiertheit: Die Handlungen müssen in einer sinnvollen Reihenfolge durchgeführt werden. 3) Zielorientiertheit: Die Handlungen müssen zu Beantwortung der Fragestellung führen. Damit ergeben sich drei Dimensionen der Experimentierkompetenz auf der Grundlage der Handlungen des idealtypischen Experimentierens.

In der Realität lässt sich der Prozess idealtypischen Experimentierens jedoch nur selten vollständig beobachten. Insbesondere im physikalischen Praktikum der universitären Ausbildung wird häufig auf die Entwicklung eigenständiger Fragestellungen und die selbständige Formulierung von Hypothesen verzichtet. Stattdessen wird von den Studierenden erwartet, dass sie sich nach einem Studium der theoretischen Grundlagen eines Versuchs mit einer entsprechenden experimentellen Aufgabenstellung auseinandersetzen und diese experimentell bearbeiten. Statt der Entwicklung einer Fragestellung (1) muss nun also die Aufgabenstellung gelesen und verstanden werden (1'); statt der Formulierung entsprechender Hypothesen (2) muss Wissen aktiviert werden, z.B. indem Literatur gelesen wird, um die theoretischen Grundlagen für die Bearbeitung der Aufgabenstellung zu schaffen (2').

Die Experimentieraufgaben im Rahmen deren Bearbeitung die zu kodierenden Aufzeichnungen entstanden sind, sind entsprechend konstruiert. Im Rahmen der Bearbeitung der Aufgaben sollten also prinzipiell die folgenden Handlungen auftreten (können): Fragestellung lesen und verstehen (1'),

Aktivierung des Wissens zur Schaffung einer theoretischen Grundlagen (2'), Experiment planen (3), Versuch aufbauen (4), Messen und Beobachten (5), Daten auswerten (6) und Schlussfolgerungen ziehen (7).

### 3. Kodieranweisungen

Die folgenden Anweisungen dienen der Kodierung der Handlungen die bei der Bearbeitung der Experimentieraufgaben entstanden sind. Die Kodierung gliedert sich in zwei Schritte: die Identifikation der Handlungen in den Aufzeichnungen und die Bewertung der Korrektheit der durchgeführten Handlungen. Die Maße für Richtigkeit, Strukturiertheit und Zielorientiertheit werden anschließend aus den gewonnenen Daten berechnet.

#### 3.1 Identifikationen von Handlungen

Die Identifikation der Handlungen in den Aufzeichnungen erfolgt in mehreren Schritten. Zunächst werden nacheinander die Aufzeichnungen oder Teile von Aufzeichnungen, die sich dieser Handlung zuordnen lassen, markiert. Für jede Handlungen werden am Ende des Abschnittes in Tabelle 2 entsprechende Kategorien definiert, Indikatoren aufgelistet und Beispiele genannt. Zusätzlich wird eine Kategorie „Sonstiges“ definiert, der alle nicht identifizierbaren Aufzeichnungen zugeordnet werden. Im Anschluss an die Markierung der Handlungen in den Aufzeichnungen werden aneinander grenzende Aufzeichnungen zu Blöcken zusammengefasst. Diese werden dann abschließend entsprechend der Leserichtung in eine Reihenfolge gebracht. Die Vorgehensweise wird im Folgenden anhand einer Beispielaufzeichnung (Abbildung 1) beschrieben.

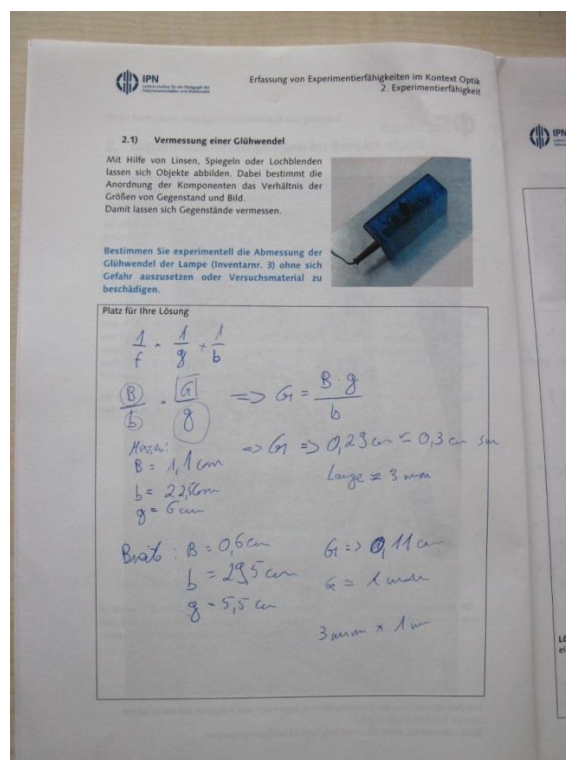


Abbildung 29: Beispiel der Aufzeichnungen in einem Testheft

Die ersten Aufzeichnungen lassen sich der Kategorie „Aktivierung von Wissen (2')“ zuordnen und werden entsprechend markiert. Das Ergebnis zeigt Abbildung 2.



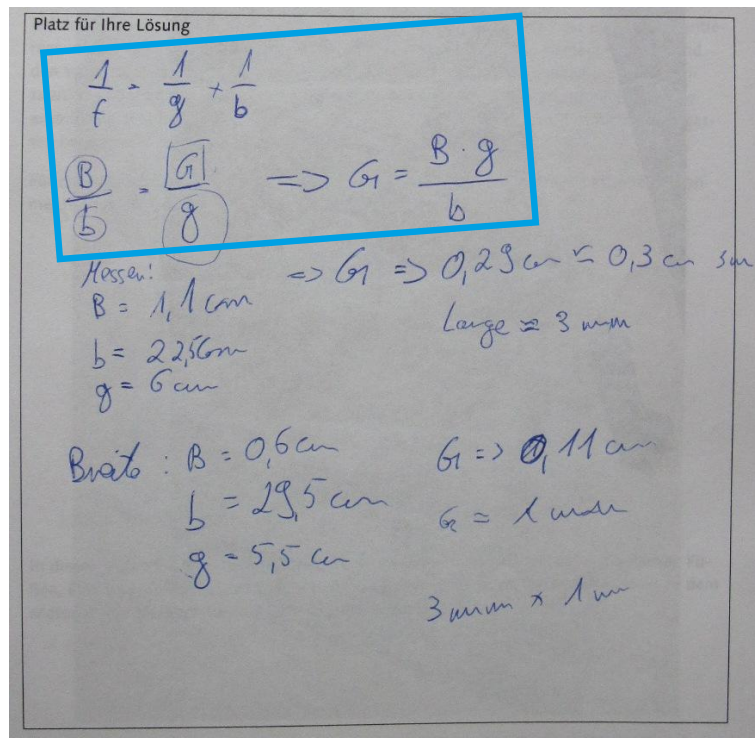


Abbildung 30: Markierung der Aufzeichnungen zu "Wissen aktivieren (2)'"

Teile dieser Aufzeichnungen müssen auch der Kategorie „Experiment planen (3)“ zuordnen und entsprechend kodiert werden. Im Beispiel sind die Messgrößen  $B$ ,  $b$ ,  $g$  durch einen Kreis und der gesuchten Größe  $G$  durch ein Rechteck markiert (vgl. Abbildung 3).

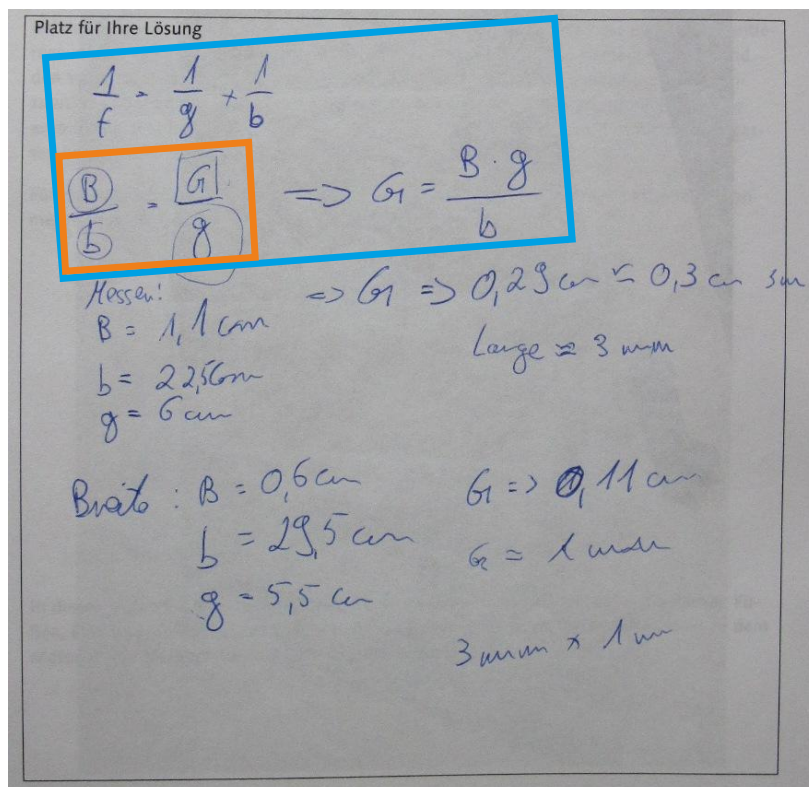


Abbildung 31: Markierung der Aufzeichnungen zu "5. Messen und Beobachten"

Es folgen Markierungen der weiteren Aufzeichnungen zu den weiteren Kategorien: „Messen und Beobachten (5)“ (rot), „Daten analysieren (6)“ (grün) und „Sachgerechte Schlüsse ziehen (7)“ (lila) (vgl. Abbildung 4).

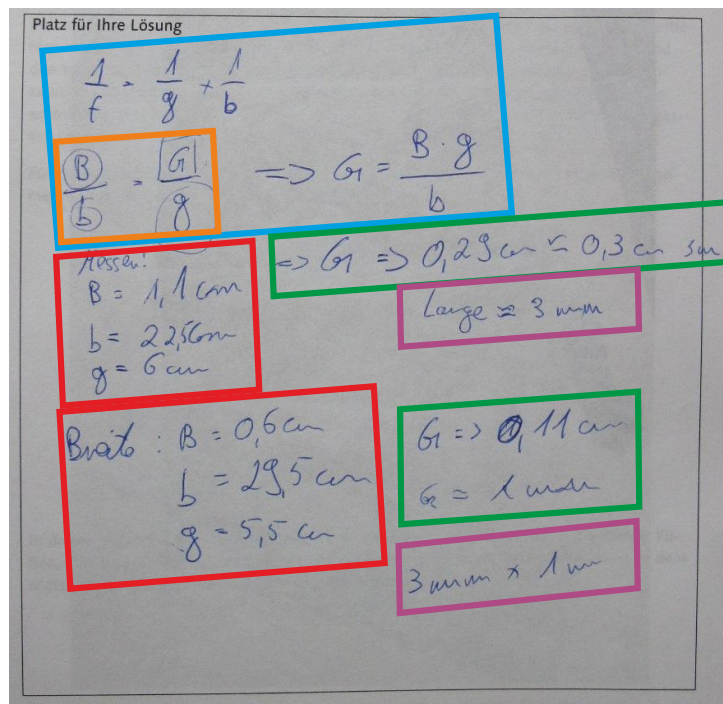


Abbildung 32: Markierung der Aufzeichnungen zu allen Handlungen

Zuletzt werden alle bisher nicht markierten Aufzeichnungen als „Sonstiges (9)“ markiert. Da es im vorliegenden Beispiel keine Aufzeichnungen dieser Art gibt, werden hier keine Markierungen gemacht.

Nun sind alle Aufzeichnungen mindestens einmal markiert. Allerdings können viele kleine oder wenige große Markierungen gemacht worden sein. (vgl. Abbildung 5).

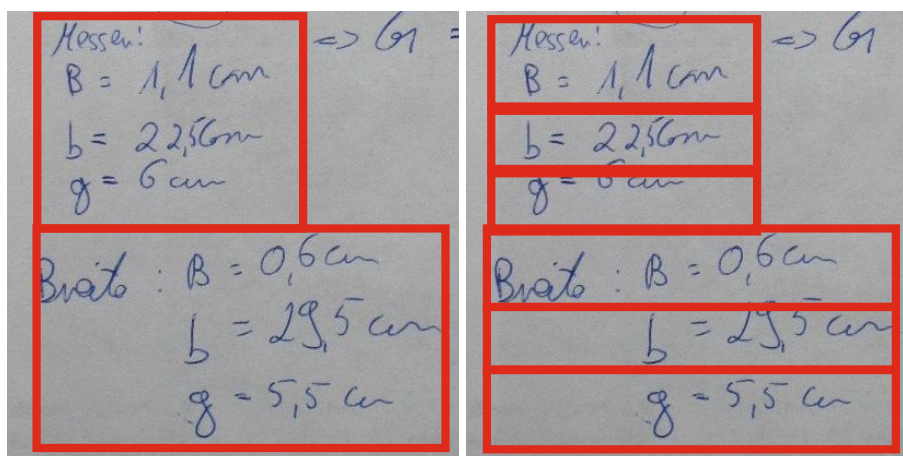


Abbildung 33: Verschiedene, mögliche Kodierungen zu "5. Messen und Beobachten"

Um zu einer eindeutigen Kodierung zu kommen werden alle Aufzeichnungen zu derselben Handlung die räumlich aneinander Grenzen zu einem Block zusammengefasst. Für das Beispiel aus Abbildung 1-4 ergibt sich damit die folgende, eindeutige Kodierung (siehe Abbildung 6).

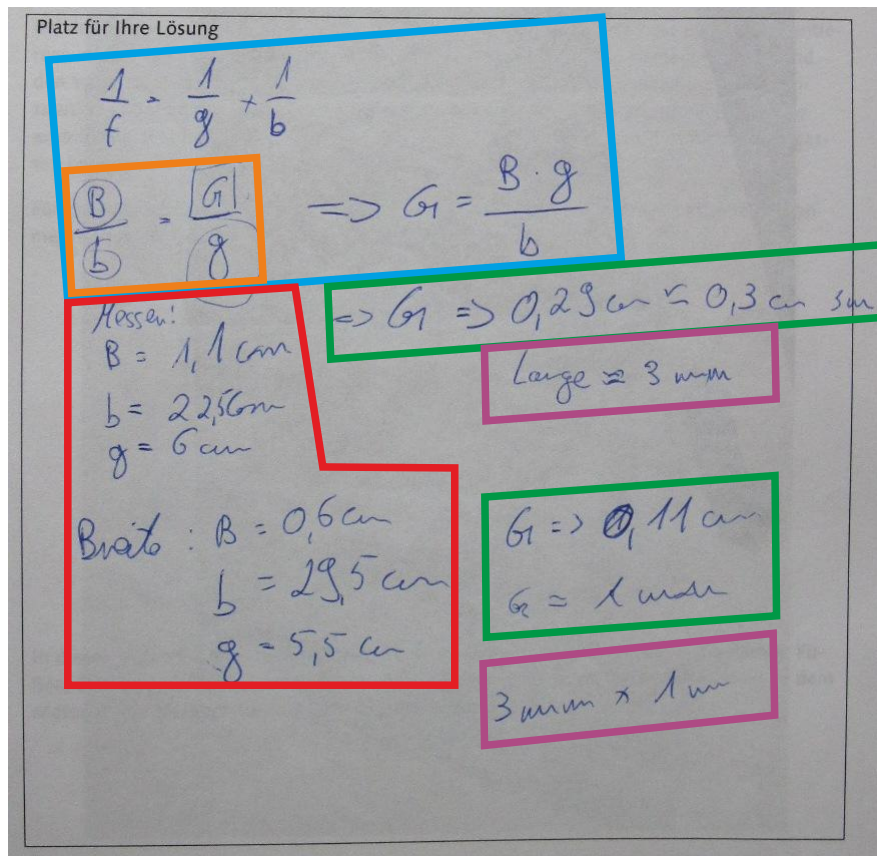


Abbildung 34: Eindeutige Kodierung

### Reihenfolge und Notation der Kodierung

Nachdem die Kodierung der Aufzeichnungen durch Bildung von Blöcken eindeutig ist, wird die Kodierung in eine Tabelle eingetragen. Eine solche Tabelle sieht wie folgt aus:

Aufgabe 1 - Glühwendel							
Handlungsnummer							
Richtigkeit							

Nun wird in Leserichtung jeweils die Nummer der Handlung des ersten Blocks eingetragen. In dem Beispiel ist dies der blaue Kasten zur Handlung „Wissen aktivieren (2)“. Kam es, so wie im Beispiel, zu einer Doppelkodierung so wird, die Handlung zuerst in die Tabelle eingetragen, dessen Block zuerst beginnt (hier der blaue Kasten). Für den Fall das beide Blöcke an derselben Stelle beginnen wird die Handlung mit der niedrigeren Nummer zuerst eingetragen.

Aufgabe 1 - Glühwendel							
Handlungsnummer	2						
Richtigkeit							

Anschließend folgen die nächsten Blöcke: (orange) „Experiment planen (3)“, (rot) „Messen und Beobachten (5)“, (grün 1) „Daten analysieren (6)“, (lila 1) „Sachgerechte Schlüsse ziehen (7)“, (grün 2) „Daten analysieren (6)“ und (lila 2) „Sachgerechte Schlüsse ziehen (7)“. Die fertige Kodiertabelle sieht wie folgt aus:

Aufgabe 1 - Glühwendel							
Handlungsnummer	2	3	5	6	7	6	7
Richtigkeit							

Sollten in der Tabelle zwei aufeinanderfolgende Blöcke dieselbe Handlungsnummer zugewiesen bekommen, so werden diese beiden Blöcke zu einem Block zusammen gezogen. Dies kommt etwa dann vor, wenn Aufzeichnungen durch den Seitenumbruch getrennt werden.

Um die Aufzeichnungen bzgl. der Handlungen zu identifizieren, gibt es aufgabenspezifischen expliziten Erwartungen was in den jeweiligen Aufzeichnungen stehen soll. Die Erwartungen finden sich im Anhang (siehe A.1).

## Hinweise

Es ist möglich, dass Sonderfälle auftreten, die zu Problemen bei der Identifikation von Handlungen führen. Welche Sonderfälle das sind und wie in diesem Fall bei der Kodierung vorzugehen ist, wird im Folgenden beschrieben.

### Durchgestrichene Aufzeichnungen

Sollten Aufzeichnungen durchgestrichen sein, so werden diese für die Kodierung ignoriert und wie leeres Papier behandelt.

### Doppelte Zuordnungen

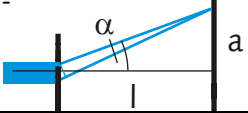
Es ist durchaus möglich, dass einer Aufzeichnung mehrere Kategorien zugeordnet werden können. Dies ist zum Beispiel der Fall, wenn mehrere Produkte zu verschiedenen Handlungen beim Experimentieren gleichzeitig notiert werden (siehe Tabelle 1).

**Tabelle 30: Verkürzte Notation und Doppelkodierung**

Ideale Aufzeichnungen	verkürzte Aufzeichnungen
$B/b = G/g; \Rightarrow G=B \cdot g/G$  $B=1,1 \text{ cm}$ $b=22,5\text{cm}$ $g=6\text{cm}$  $G=1,1\text{cm} \cdot 22,5\text{cm}/6\text{cm}=0,3\text{cm}$	$1,1\text{cm} \cdot 22,5\text{cm}/6\text{cm}=0,3\text{cm}$

Während die idealen Aufzeichnungen wie in Abbildung 1-4 beschrieben leicht zu kodieren ist, hat man bei der verkürzten Aufzeichnung eine Mehrfachkodierung. In der verkürzten Aufzeichnung findet sich eine richtige Formel die vorher nicht eingeführt wurde [Wissen aktivieren (2)], es finden sich Messwerte die vorher nicht aufgeführt werden [Messen und Beobachten (5)] und es findet sich ein Ergebnis einer Rechnung [Daten analysieren (6)]. Entsprechend werden auch in den verkürzten Aufzeichnungen alle drei Handlungen kodiert.

Tabelle 31: Kategorien zur Kodierung der Aufzeichnungen im Laborheft

Handlung	Beschreibung	Indikatoren	Beispiel	Abgrenzung
Fragestellung klären (1)	Die Fragestellung wird gelesen und verstanden.	Die Fragestellung wird wörtlich oder in eigenen Worten aufgeschrieben.		
Wissen aktivieren und theoretische Grundlage schaffen (2)	Das für die Bearbeitung der Fragestellung notwendige Wissen wird aktiviert.	Es finden sich Erklärungen zur Theorie der Fragestellung, dazu gehören auch Formeln, Umformungen ODER das Notieren von bekannten Größen.	<p>Formel zur Berechnung der Gitterkonstante: <math>n \cdot \lambda = g \cdot \sin(\alpha_n)</math>; <math>\sin(\alpha_n) = a_i / l</math>.</p> <p>Variablen: <math>n_1=1</math>; <math>\lambda_{\text{grün}}=532 \text{ nm}</math></p>	
Experiment planen (3)	Die Planung eines Experiments zur Beantwortung der Fragestellung.	Die Planung des Experiments wird beschrieben, dazu gehören zu messende Werte bzw. zu machende Beobachtungen und die Erklärung des Versuchsaufbaus inklusive Skizze.	<p>Skizze eines Versuchsaufbaus (siehe Abbildung) und/oder Beschreibung des Versuchs.</p> 	Abgrenzung zu „Versuch aufbauen (4)“: Die Skizze enthält keine konkreten Werte und Abmessungen.
Versuch aufbauen (4)	Ein Versuch wird aufgebaut.	Es findet sich eine Zeichnung des Versuchsaufbaus die sich 1:1 nachbauen lässt.		Abgrenzung zu „Versuch aufbauen (4)“: Die Zeichnung enthält konkrete Werte und Abmessungen wodurch der Aufbau 1:1 nachgebaut werden kann.
Messen und Beobachten (5)	Gezielt Messwerte aufnehmen oder Beobachtungen anstellen.	Handlungen zur Fertigkeit „Messen und Beobachten“ zeigen sich durch Aufzeichnungen von Messwerten, Messtabellen und Beobachtungen bei der Versuchsdurchführung.	$n=1$ ; $\lambda=532 \text{ nm}$ ; $a_i=2,2 \text{ cm}$ ; $l=30 \text{ cm}$ .	
Daten analysieren (6)	Auf Basis von Messwerten werden Berechnungen durchgeführt oder Graphen gezeichnet.	Handlungen zur Fähigkeit „Daten analysieren“ sind erkennbar durch Rechnungen die auf den Messwerten basieren, sowie dem Anfertigen und Auswerten von Graphen. Dazu kann es auch gehören Fehlerrechnung anfertigen oder Fehlerbalken zu	$g=n\lambda/a_i=7,25 \cdot 10^{-6} \text{ m}$	

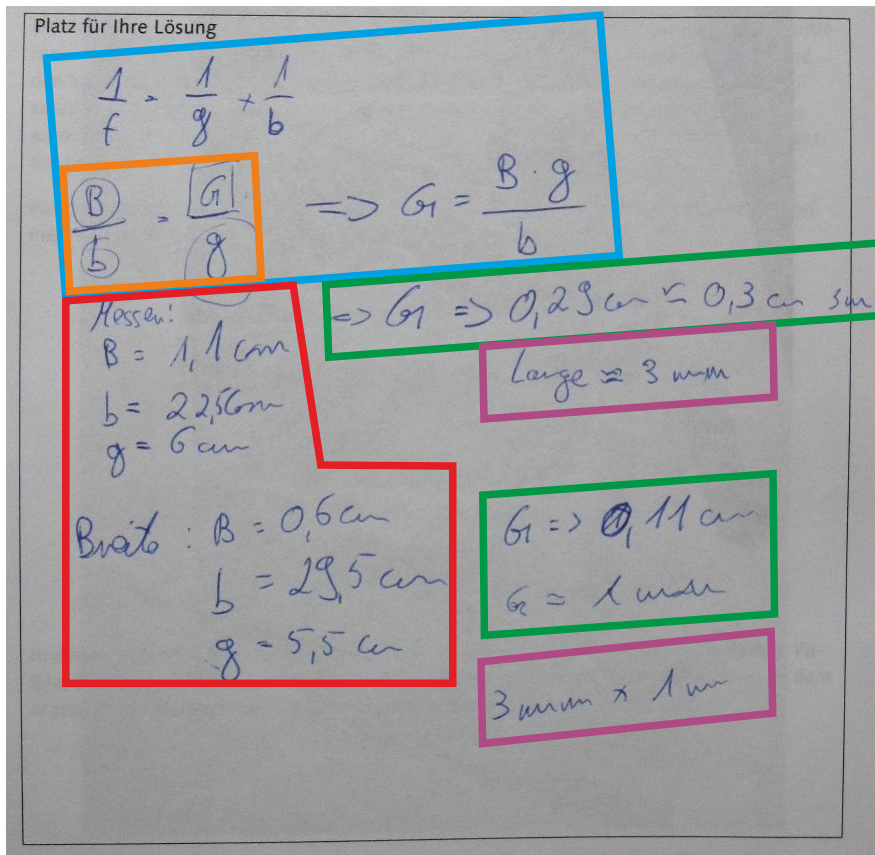
		zeichnen.		
Sachgerechte Schlüsse ziehen (7)	Berechnungen, Messgrößen und Beobachtungen werden auf die Fragestellung bezogen um zu prüfen, ob die Fragestellung beantwortet wurde.	Handlungen zur Fähigkeit „Sachgerechte Schlüsse ziehen“ zeigen sich in den Aufzeichnungen durch Antwortsätze und Schlussfolgerungen zu den Messwerten. Messfehler und Ungenauigkeiten können zur Erklärung genutzt werden.	<i>Die Gitterkonstante liegt im erwarteten Bereich. Da keine Fehler bei der Messung aufgetreten sind, wird angenommen, dass die Gitterkonstante <math>g \approx 7,25 \cdot 10^6 \text{ m}</math> ist.</i>	
Sonstiges (9)	Sonstige Aufzeichnungen.	Produkte die keiner Handlung zugeteilt werden können.		

### 3.2 Richtigkeit

Nachdem die Handlungen identifiziert wurden, erfolgt nun deren Kodierung bezüglich der Richtigkeit. Jedem Block an Aufzeichnungen kann dabei einer von vier Codes bzgl. der korrekten Durchführung zugeordnet werden:

- (1) Die Aufzeichnungen entsprechen der Musterlösung oder sind äquivalent dazu (richtig).
- (0,5) Die Aufzeichnungen enthalten Teile der Musterlösung bzw. äquivalente Aufzeichnungen UND es fehlen Teile der Musterlösung oder es sind Fehler in den Aufzeichnungen (teilweise richtig)
- (0) Die Aufzeichnungen entsprechend nicht der Musterlösung oder sind durchgehend falsch (falsch)
- (99) Die Aufzeichnungen sind nicht interpretierbar (nicht interpretierbar)

Um diese Codes den jeweiligen Aufzeichnungen zuordnen zu können, finden sich im Anhang aufgabenspezifische Tabellen (siehe A.2). In diesen wird beschrieben, was das Kriterium in den Aufzeichnungen ist anhand dessen die Richtigkeit bewertet wird. Zudem findet sich die Musterlösung, die zur Bewertung als Referenz genutzt wird. Im Folgenden wird die Bewertung der Richtigkeit an dem Beispiel aus 3.1 dokumentiert.



Aufgabe 1 – Glühwendel							
Handlungsnummer	2	3	5	6	7	6	7
Richtigkeit							

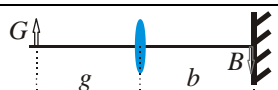
Der erste Block der Tabelle ist „Wissen aktivieren und Theorie schaffen (2)“. Aus der Tabelle im Anhang findet sich dafür bei der Aufgabe 1 – Glühwendel:

Handlung	Kriterium für Richtigkeit	Musterlösung
Wissen aktivieren (2)	Beschreibungen und Erklärungen zur Theorie und Formeln	Formel: $G=B \cdot g/b$

Die Formel der Musterlösung findet sich genauso in den Aufzeichnungen. Entsprechend wird hier „1“ (richtig) kodiert. Für die Kodiertabelle ergibt sich damit:

Aufgabe 1 – Glühwendel							
Handlungsnummer	2	3	5	6	7	6	7
Richtigkeit	1						

Die nächste Handlung ist „Experiment planen (3)“, die aufgabenspezifische Tabelle dafür sieht wie folgt aus:

Handlung	Kriterium für Richtigkeit	Musterlösung
Experiment planen (3)	Skizzen und Erklärungen zum Versuchsaufbau	Zu Messen: $B, g, b$ Skizze: 

In den Aufzeichnungen finden sich die zu messenden Größen. Die entsprechende Kodierung wäre hier „0,5“ (teilweise richtig), weil die Messgrößen mit der Musterlösung übereinstimmen, der zweite Teil der Musterlösung aber fehlt und diese deshalb unvollständig ist. Die Kodiertabelle wird ergänzt zu:

Aufgabe 1 – Glühwendel							
Handlungsnummer	2	3	5	6	7	6	7
Richtigkeit	1	0,5					

Entsprechend wird für die weiteren Handlungen verfahren. Die Handlung „Messen und Beobachten (5)“ wird mit „1“ (richtig) kodiert, weil sich aus den Messwerten  $B=1,1\text{cm}$ ,  $b=22,5\text{cm}$  und  $g=6\text{cm}$  bzw.  $B'=0,6\text{cm}$ ,  $b'=29,5\text{cm}$  und  $g'=5,5\text{cm}$  ein Wert  $G=0,29\text{cm}$  bzw.  $G'=0,11\text{cm}$  berechnen lässt und diese Werte innerhalb der gegebenen Toleranz von  $2,5\text{mm} - 3,5\text{mm}$  bzw.  $0,5\text{mm} - 1,5\text{mm}$  liegen. Die Handlung „Daten analysieren (6)“ wird mit „1“ (richtig) kodiert, da die Rechnung des Probanden ein Ergebnis von  $G=0,29\text{cm}$  ergibt, dass sich durch nachrechnen aus den Werten reproduzieren lässt. Die Handlung „Sachgerechte Schlüsse ziehen (7)“ wird ebenfalls mit „1“ (richtig) kodiert, da der Wert als von  $0,3\text{cm}$  als Länge der Glühwendel beschrieben wird und noch keine Daten zur Erklärung der Breite vorliegen. Der zweite Block zu der Kategorie „Daten analysieren (6)“ wird mit „1“ (richtig) kodiert, da die Rechnung des Probanden ein Ergebnis von  $G'=0,11\text{cm}$  ergibt, dass sich durch nachrechnen aus den Werten reproduzieren lässt. Der zweite Block zu der Kategorie „Sachgerechte Schlüsse ziehen (7)“ wird ebenfalls mit „1“ (richtig) kodiert, da der Wert als von  $0,1\text{cm}$  als Breite der Glühwendel beschrieben wird und die Daten zur Erklärung der Länge bereits ausgewertet sind. Die fertige Kodiertabelle sieht wie folgt aus:

Aufgabe 1 - Glühwendel							
Handlungsnummer	2	3	5	6	7	6	7
Richtigkeit	1	0,5	1	1	1	1	1

### Wiederholungen

Kommt eine Aufzeichnung in identischer Form an mehreren Stellen vor (z.B. eine Formel, eine Vari-



able  $n_1=1, \dots$ ) dann wird nur für die erste Stelle eine Richtigkeit kodiert. Die Wiederholungen werden mit „99“ (nicht interpretierbar) kodiert.

## A.1 Aufgabenspezifische Kodieranweisungen zur Identifikation der Handlungen

### Vermessung einer Glühwendel

Handlung	Produkt
Fragestellung klären (1)	Notierte Fragestellung
Wissen aktivieren (2)	Beschreibungen und Erklärungen zur Theorie und Formeln
Experiment planen (3)	Beschreibung und Skizze
Versuch aufbauen (4)	Zeichnung zum aufgebauten Versuch
Messen / Beobachten (5)	Messwerte für $g, b, B_i$
Daten analysieren (6)	Berechnen von $\hat{G}_i$
Sachgerechte Schlüsse ziehen (7)	Lösungssatz
Sonstiges (9)	Sonstige Aufzeichnung

### Optische Dichte

Handlung	Produkt
Fragestellung klären (1)	Notierte Fragestellung
Wissen aktivieren (2)	Beschreibungen und Erklärungen zur Theorie und Formeln
Experiment planen (3)	Beschreibung und Skizze
Versuch aufbauen (4)	Zeichnung zum aufgebauten Versuch
Messen / Beobachten (5)	Einfallswinkel $\alpha_1$ und Ausfallswinkel $\alpha_2$ für Laser verschiedener Wellenlänge
Daten analysieren (6)	Berechnung von $n_2$
Sachgerechte Schlüsse ziehen (7)	Lösungssatz
Sonstiges (9)	Sonstige Aufzeichnung

### Spektroskopie

Handlung	Produkt
Fragestellung klären (1)	Notierte Fragestellung
Wissen aktivieren (2)	Beschreibungen und Erklärungen zur Theorie und Formeln
Experiment planen (3)	Beschreibung und Skizze
Versuch aufbauen (4)	Zeichnung zum aufgebauten Versuch
Messen / Beobachten (5)	Messwerte $a_i$ und $l$
Daten analysieren (6)	Berechnung von $g$ oder $\ddot{e}$
Sachgerechte Schlüsse ziehen (7)	Lösungssatz
Sonstiges (9)	Sonstige Aufzeichnung

## Zuckerwasser

Handlung	Produkt
Fragestellung klären (1)	Notierte Fragestellung
Wissen aktivieren (2)	Beschreibungen und Erklärungen zur Theorie und Formeln
Experiment planen (3)	Beschreibung und Skizze
Versuch aufbauen (4)	Zeichnung zum aufgebauten Versuch
Messen / Beobachten (5)	Drehung der Polarisations Ebene $\alpha$
Daten analysieren (6)	Berechnung der Zuckerkonzentration und der Zuckerkonzentration pro Liter
Sachgerechte Schlüsse ziehen (7)	Lösungssatz
Sonstiges (9)	Sonstige Aufzeichnung

## Polarisationseigenschaften

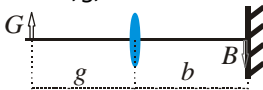
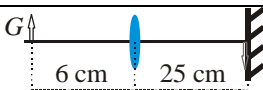
Handlung	Produkt
Fragestellung klären (1)	Notierte Fragestellung
Wissen aktivieren (2)	Beschreibungen und Erklärungen zur Theorie und Formeln
Experiment planen (3)	Beschreibung und Skizze
Versuch aufbauen (4)	Zeichnung zum aufgebauten Versuch
Messen / Beobachten (5)	Messen der Intensität des transmittierten- und reflektierten Strahls in Abhängigkeit zum Einfallswinkel UND Messen der Intensität von transmittierten- und reflektierten Strahl in Abhängigkeit zum Polarisationsfilter bei festem „Winkel der stärksten Polarisation“
Daten analysieren (6)	Ablesen des „am stärksten polarisierenden Winkels“
Sachgerechte Schlüsse ziehen (7)	Lösungssatz
Sonstiges (9)	Sonstige Aufzeichnung

## Blackbox

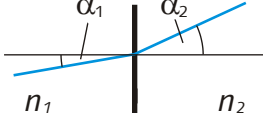
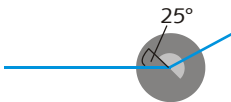
Handlung	Produkt
Fragestellung klären (1)	Notierte Fragestellung
Wissen aktivieren (2)	Beschreibungen und Erklärungen zur Theorie und Formeln
Experiment planen (3)	Beschreibung und Skizze
Versuch aufbauen (4)	Zeichnung zum aufgebauten Versuch
Messen / Beobachten (5)	Beobachten und Messe
Daten analysieren (6)	Berechnung der Gitterposition
Sachgerechte Schlüsse ziehen (7)	Lösungssatz
Sonstiges (9)	Sonstige Aufzeichnung

## A.2 Aufgabenspezifische Kodieranweisungen zur Richtigkeit

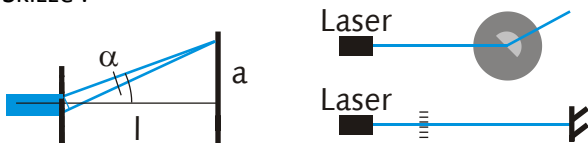
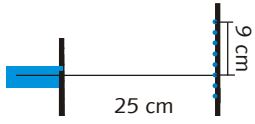
### Vermessung einer Glühwendel

Handlung	Kriterium für Richtigkeit	Musterlösung
Fragestellung klären (1)	Notierte Fragestellung	Wie ist die Abmessung der Glühwendel von der Lampe (Inventarnr. 3)?
Wissen aktivieren (2)	Beschreibungen und Erklärungen zur Theorie und Formeln	Formel: $G=B \cdot g/b$
Experiment planen (3)	Skizzen, Erklärungen zum Versuchsaufbau ODER Messgrößen ODER Überschriften zu Teilabschnitten	Zu Messen: $B, g, b$  Skizze: Überschriften zu Abschnitten sind richtig, wenn sie ein Teil der Musterlösung sind.
Versuch aufbauen (4)	Zeichnung zum aufgebauten Versuch	
Messen und Beobachten (5)	Messwerte für $g, b, B_i$	Zur Kodierung muss der Kodierer $G_i$ aus den Messwerten $g, b, B_i$ anhand der Formel $G_i=B_i \cdot g/b$ berechnen. Die Phase gilt als richtig wenn $G_i$ zwischen 2,5mm und 3,5mm oder zwischen 0,5mm und 1,5mm liegt.
Daten analysieren (6)	Berechnen von $G_i$	Richtig, wenn das vom Probanden aus den Messwerten $g, b, B_i$ errechnete $G_i$ mit dem vom Kodierer errechneten $G_i$ übereinstimmt.
Sachgerechte Schlüsse ziehen (7)	Lösungssatz oder Erklärung der Messwerte.	Abhängig von bisher errechneten Daten muss entweder die Länge $a$ im Bereich von 2,5mm und 3,5mm liegen oder die Breite $b$ im Bereich von 0,5mm und 1,5mm liegen. Wird beides angegeben, müssen beide Werte innerhalb der Messgenauigkeit liegen. ODER Erklärung, dass die Ergebnisse nicht stimmen können.
Sonstiges (9)	Sonstige Aufzeichnung	Richtigkeit wird nicht kodiert „-“

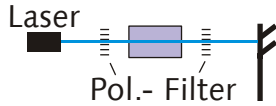
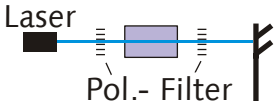
## Optische Dichte

Handlung	Kriterium für Richtigkeit	Musterlösung
Fragestellung klären (1)	Notierte Fragestellung	Aus welchem Material ist das Prisma (Inventarnr. 14)?
Wissen aktivieren (2)	Beschreibungen und Erklärungen zur Theorie und Formeln	Formel: $n_1 \cdot \sin(\alpha_1) = n_2 \cdot \sin(\alpha_2)$ Variable: $n_1=1$
Experiment planen (3)	Skizzen, Erklärungen zum Versuchsaufbau ODER Messgrößen ODER Überschriften zu Teilabschnitten	Zu Messen: $\alpha_1, \alpha_2$ Skizze:  Überschriften zu Abschnitten sind richtig, wenn sie ein Teil der Musterlösung sind.
Versuch aufbauen (4)	Zeichnung zum aufgebauten Versuch	
Messen und Beobachten (5)	Einfalls- $\alpha_1$ und Ausfallswinkel $\alpha_2$ für Laser verschiedener Wellenlänge	Zur Kodierung muss der Kodierer $n_2$ aus den Messwerten $\alpha_1, \alpha_2$ für $n_1 = 1$ berechnen. Die Phase gilt als richtig wenn $n_2$ zwischen 1,40 und 1,55 liegt.
Daten analysieren (6)	Berechnung von $n_2$	Richtig, wenn das vom Probanden aus den Messwerten $\alpha_1$ und $\alpha_2$ errechneten $n_2$ mit dem vom Kodierer errechneten $n_2$ übereinstimmt.
Sachgerechte Schlüsse ziehen (7)	Lösungssatz	Es wurde Graph „F“ gewählt ODER Erklärung das ein Messfehler vorliegen muss
Sonstiges (9)	Sonstige Aufzeichnung	Richtigkeit wird nicht kodiert „-“

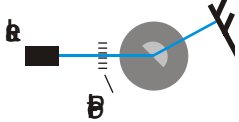
## Spektroskopie

Handlung	Kriterium für Richtigkeit	Musterlösung
Fragestellung klären (1)	Notierte Fragestellung	Wie ist die Wellenlänge des unbekanntes Lasers (Inventarnr. 5)?
Wissen aktivieren (2)	Beschreibungen und Erklärungen zur Theorie und Formeln	Formel: $n \cdot \ddot{e} = g \cdot \sin(\alpha_n)$ ; $\sin(\alpha_n) = a_i / l$ ; $\tan(\alpha_n) = a_i / l'$ Variable : $n_1=1$ ; $\ddot{e}_{\text{blau}}=430\text{nm}$ ; $\ddot{e}_{\text{grün}}=532\text{nm}$
Experiment planen (3)	Skizzen, Erklärungen zum Versuchsaufbau ODER Messgrößen ODER Überschriften zu Teilabschnitten	Zu Messen: $a_i$ , $l$ bzw. $a_i$ , $l'$ bzw. $\alpha_i$ Skizze :  Überschriften zu Abschnitten sind richtig, wenn sie ein Teil der Musterlösung sind.
Versuch aufbauen (4)	Zeichnung zum aufgebauten Versuch	
Messen und Beobachten (5)	Messwerte $a_i$ und $l$	Zur Kodierung muss der Kodierer $g$ aus den Messwerten $a_i$ und $l$ für $\ddot{e}_{\text{rot}}= 642 \text{ nm}$ ; $\ddot{e}_{\text{grün}}= 532 \text{ nm}$ ; $\ddot{e}_{\text{blau}}= 430 \text{ nm}$ berechnen: $g=n \cdot \ddot{e} \cdot l/a$ Die Phase gilt als richtig wenn $g$ zwischen $7,00 \cdot 10^{-6} \text{ m}$ und $7,25 \cdot 10^{-6} \text{ m}$ liegt.
Daten analysieren (6)	Berechnung von $g$ oder $\ddot{e}$	Zur Kodierung muss der Kodierer $g$ bzw. $\ddot{e}$ aus den Messwerten nachrechnen. Die Phase gilt als richtig wenn ein entsprechendes Ergebnis von Kodierer und Proband erzielt wird.
Sachgerechte Schlüsse ziehen (7)	Lösungssatz	Richtig wenn $\lambda_{\text{rot}}$ zwischen 600 nm und 680 nm ODER Erklärung, dass ein Messfehler vorliegen muss
Sonstiges (9)	Sonstige Aufzeichnung	Richtigkeit wird nicht kodiert „-“

## Zuckerwasser

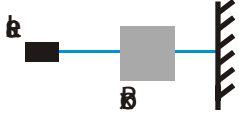
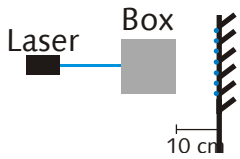
Handlung	Kriterium für Richtigkeit	Musterlösung
Fragestellung klären (1)	Notierte Fragestellung	Bei welcher Zuckerkonzentration wird die Polarisationssebene um $0,75^\circ/\text{cm}$ gedreht?
Wissen aktivieren (2)	Beschreibungen und Erklärungen zur Theorie und Formeln	Formel: $V=h \cdot l \cdot b$ ; $\rho=m/V$
Experiment planen (3)	Skizzen, Erklärungen zum Versuchsaufbau ODER Messgrößen ODER Überschriften zu Teilabschnitten	Skizze:  oder Skizze der Küvette Überschriften zu Abschnitten sind richtig, wenn sie ein Teil der Musterlösung sind.
Versuch aufbauen (4)	Zeichnung zum aufgebauten Versuch	 Mit Polarisationsfilter $90^\circ$ versetzt.
Messen und Beobachten (5)	Drehung der Polarisationssebene $\alpha$	Richtig wenn die minimale und die maximale Auslöschung des Lichts bei einem Winkel von $85^\circ - 95^\circ$ zwischen den Polarisationsfiltern auftritt. ODER Richtige Messung, wenn sich Anzahl an Löffeln Zucker für die Änderung die Polarisationssebene um $0,75^\circ/\text{cm}$ zwischen $9 \cdot 1000 \text{cm}^3/\text{Volumen}(\text{in cm}^3)$ und $15 \cdot 1000 \text{cm}^3/\text{Volumen}(\text{in cm}^3)$ liegt. ODER Volumen= $10 \text{cm} \cdot 1,5 \text{cm} \cdot b$ , mit $b$ zwischen $1 \text{cm}$ und $6,5 \text{cm}$ .
Daten analysieren (6)	Berechnung der Zuckerkonzentration und der Zuckerkonzentration pro Liter	Richtig, wenn die Konzentration zu einer Drehung von $7,5^\circ$ oder $1,9^\circ$ aus den Daten „Löffel“ und „Drehwinkel“ berechnet werden UND Die Konzentration auf Liter umgerechnet werden
Sachgerechte Schlüsse ziehen (7)	Lösungssatz	Richtiges Ergebnis, wenn 9-15 Löffel angegeben werden ODER Erklärung eines Messfehlers
Sonstiges (9)	Sonstige Aufzeichnung	Richtigkeit wird nicht kodiert „-“

## Polarisationseigenschaften

Handlung	Kriterium für Richtigkeit	Musterlösung
Fragestellung klären (1)	Notierte Fragestellung	Wie ist der Winkel unter dem das Licht linear polarisiert wird? Wie ist die Schwingungsebene des transmittierten- und des reflektierten Lichts?
Wissen aktivieren (2)	Beschreibungen und Erklärungen zur Theorie und Formeln	-
Experiment planen (3)	Skizzen, Erklärungen zum Versuchsaufbau ODER Messgrößen ODER Überschriften zu Teilabschnitten	Zu Messen: $\alpha_1, \alpha_2$ Skizze:  Überschriften zu Abschnitten sind richtig, wenn sie ein Teil der Musterlösung sind.
Versuch aufbauen (4)	Zeichnung zum aufgebauten Versuch	-
Messen und Beobachten (5)	Messen der Intensität des transmittierten- und reflektierten Strahls in Abhängigkeit zum Einfallswinkel UND Messen der Intensität von transmittierten- und reflektierten Strahl in Abhängigkeit zum Polarisationsfilter bei festem „Winkel der stärksten Polarisation“	Richtig, wenn das Maximum des transmittierten/reflektierten und das Minimum des reflektierten/transmittierten Strahls bei einem Einfallswinkel von $\alpha = 50^\circ - 62^\circ$ auftreten UND $\beta, \gamma$ stehen senkrecht aufeinander ( $90^\circ$ Differenz im Winkel)
Daten analysieren (6)	Ablesen des „am stärksten polarisierenden Winkels“	Richtig, wenn der „Winkel der stärksten Polarisation“ gewählt wird, wenn der transmittierte/reflektierte Strahl sein Maximum/Minimum erreicht
Sachgerechte Schlüsse ziehen (7)	Lösungssatz	Der „am stärksten polarisierende Winkel“ $\alpha = 50^\circ - 62^\circ$ UND $\beta, \gamma$ stehen senkrecht aufeinander ( $90^\circ$ Differenz im Winkel) ODER Erklärung eines Messfehlers
Sonstiges (9)	Sonstige Aufzeichnung	Richtigkeit wird nicht kodiert „-“



## Blackbox

Handlung	Kriterium für Richtigkeit	Musterlösung
Fragestellung klären (1)	Notierte Fragestellung	Wie sind die optischen Eigenschaften des unbekanntes Bauteils?
Wissen aktivieren (2)	Beschreibungen und Erklärungen zur Theorie und Formeln	-
Experiment planen (3)	Skizzen, Erklärungen zum Versuchsaufbau ODER Messgrößen ODER Überschriften zu Teilabschnitten	<p>Skizze:</p>  <p>oder Skizze der Blackbox</p> <p>Überschriften zu Abschnitten sind richtig, wenn sie ein Teil der Musterlösung sind. Erklärung das ein bestimmter Laser genutzt wird und um ein Bild der Komponenten in der Blackbox zu erzeugen.</p>
Versuch aufbauen (4)	Zeichnung zum aufgebauten Versuch	
Messen und Beobachten (5)	Beobachten und Messen	<p>Richtig, wenn:</p> <p>Bemerkt wird, dass Licht bestimmter Wellenlängen gefiltert wird  UND  Ein Beugungsbild entsteht  UND  <math>a_i</math> für die Abstände der <math>i</math>-ten Beugungsordnung zur optischen Achse gemessen werden</p>
Daten analysieren (6)	Berechnung der Gitterposition	<p>Richtig wenn berechnet wird, dass sich das Gitter an Position 1 befinden muss  ODER  Wenn die Gitterkonstante korrekt bestimmt wird (<math>g</math> liegt zwischen <math>7,00 \cdot 10^{-6}</math> m und <math>7,25 \cdot 10^{-6}</math> m)</p>
Sachgerechte Schlüsse ziehen (7)	Lösungssatz	<p>Richtig wenn:</p> <p>Ein Farbfilter wird diagnostiziert  UND  Ein Gitter in einem der Randpositionen diagnostiziert wird  UND  Der letzte Slot leer ist oder nicht gesagt werden kann, was sich im letzten Slot befindet.  Teilweise richtig (0,5 Punkte) wenn zwei der Slots richtig sind.</p>
Sonstiges (9)	Sonstige Aufzeichnung	Richtigkeit wird nicht kodiert „-“

## **A.4.7 Kodiermanual prozessorientierte- und Referenzauswertung**

### **Prozessorientierte Auswertung von Experimentierkompetenz**

#### **1. Einleitung**

Das vorliegende Manual dient der Auswertung von Videos, die Probanden bei der Bearbeitung von sechs Experimentieraufgaben zeigen. Anhand der Videos sollen mit Hilfe des vorliegenden Manuals die Handlungen, die beim Experimentieren durchgeführt werden, identifiziert und hinsichtlich ihrer Richtigkeit bewertet werden.

#### **2. Grundlagen**

Experimentieren ist ein komplexer Prozess, beim dem unterschiedliche Fähigkeiten und Fertigkeiten eine Rolle spielen. In physik- bzw. naturwissenschaftsdidaktischer Literatur hat sich eine Gliederung entsprechend der Handlungen beim idealtypischen Experimentieren als potentiell iterativer Prozess etabliert. Experimentieren beginnt idealtypisch mit der Planung eines Experiments. Das heißt, aus theoretischen Überlegungen oder empirischen Beobachtungen wird eine Fragestellung entwickelt (1) und es werden dazu passende fachliche Hypothesen gebildet (2). Anschließend wird ein geeignetes Experiment geplant (3), mit dem sich die Frage beantworten lässt. Auf die Planungsphase folgt die Durchführung. Im Rahmen der Durchführung wird zunächst ein Versuch aufgebaut (4) und anschließend wird gezielt gemessen und beobachtet (5). Den Abschluss des idealtypischen Experimentierens bildet die Auswertung. Dabei werden Daten analysiert (6) indem Berechnungen auf Basis der Messwerte durchgeführt oder Graphen angefertigt werden. Anschließend werden sachgerechte Schlüsse (7) gezogen, indem die Beobachtungen und Ergebnisse der Analyse von Daten auf die Fragestellung bezogen werden.

Diese Handlungen bilden die Grundlage für die Beschreibung von Experimentierkompetenz. Experimentierkompetenz umfasst demnach die Fähigkeiten und Fertigkeiten die entsprechenden Handlungen durchführen zu können. Die Ausprägung von Experimentierkompetenz lässt sich dann anhand von drei Kriterien beschreiben: 1) Richtigkeit: Erfolgreiches Experimentieren setzt die richtige Durchführung der einzelnen Handlungen voraus. 2) Strukturiertheit: Die Handlungen müssen in einer sinnvollen Reihenfolge durchgeführt werden. 3) Zielorientiertheit: Die Handlungen müssen zu Beantwortung der Fragestellung führen. Damit ergeben sich drei Dimensionen der Experimentierkompetenz auf der Grundlage der Handlungen des idealtypischen Experimentierens.

In der Realität lässt sich der Prozess idealtypischen Experimentierens jedoch nur selten vollständig beobachten. Insbesondere im physikalischen Praktikum der universitären Ausbildung wird häufig auf die Entwicklung eigenständiger Fragestellungen und die selbständige Formulierung von Hypothesen verzichtet. Stattdessen wird von den Studierenden erwartet, dass sie sich nach einem Studium der theoretischen Grundlagen eines Versuchs mit einer entsprechenden experimentellen Aufgabenstellung auseinandersetzen und diese experimentell bearbeiten. Statt der Entwicklung einer Fragestellung (1) muss nun also die Aufgabenstellung gelesen und verstanden werden (1'); statt der Formulierung entsprechender Hypothesen (2) muss Wissen aktiviert werden, z.B. indem Literatur gelesen wird, um die theoretischen Grundlagen für die Bearbeitung der Aufgabenstellung zu schaffen (2').

Die Experimentieraufgaben im Rahmen deren Bearbeitung die zu kodierenden Aufzeichnungen entstanden sind, sind entsprechend konstruiert. Im Rahmen der Bearbeitung der Aufgaben sollten also prinzipiell die folgenden Handlungen auftreten (können): Fragestellung lesen und verstehen (1'), Aktivierung des Wissens zur Schaffung einer theoretischen Grundlagen (2'), Experiment planen (3),

Versuch aufbauen (4), Messen und Beobachten (5), Daten auswerten (6) und Schlussfolgerungen ziehen (7).

### 3. Kodieranweisungen

Die folgenden Anweisungen dienen der Kodierung der Handlungen die bei der Bearbeitung der Experimentieraufgaben durchgeführt wurden und auf Video festgehalten sind. Die Kodierung gliedert sich in zwei Schritte: die Identifikation der Handlungen in den Videos und die Bewertung der Korrektheit der durchgeführten Handlungen. Zur besseren Identifikation der Handlungen der Probanden und der Korrektheit wurde ein Video angefertigt, was den Experimentierbereich zeigt und eines das die Produkte im Testheft zeigt. Die Maße für Richtigkeit, Strukturiertheit und Zielorientiertheit werden anschließend aus den gewonnenen Daten berechnet.

#### 3.1 Identifikationen von Handlungen

Die Identifikation der Handlungen in dem Video erfolgt mithilfe des Programms „Videograph“. Zu der Kodierung werden Abschnitte des Videos abgespielt und jeweils Handlungen zugeordnet. Für jede Handlung werden am Ende des Abschnittes in Tabelle 2 entsprechende Kategorien definiert, Indikatoren aufgelistet und Beispiele genannt. Zusätzlich werden zwei Kategorie „Sonstiges“ und „Sonstiges nicht Experimentieren“ definiert, der alle nicht identifizierbaren Handlungen zugeordnet werden.

Die Vorgehensweise wird im Folgenden anhand eines Beispielvideos (externe Videodatei) beschrieben. Die Länge der zu bewertenden Videoabschnitte beträgt 10 Sekunden. Jedem dieser Intervalle wird genau eine Handlung in Videograph zugeordnet. Sollten mehrere Handlungen vorkommen, so wird die Kategorie gewählt, welche den größten Teil des Intervalls ausfüllt.

Intervall	Zeitspanne	Handlung	Kodierung
1	0-10	Probandin liest im Testheft	Klärung der Fragestellung (1)
2-4	20-40	Probandin liest im Testheft	Klärung der Fragestellung (2)
5	40-50	Probandin liest im Testheft, in der Inventarliste und nimmt Komponenten aus dem Koffer.	

Während das Lesen im Testheft zur Kategorie „Klärung der Fragestellung (1)“ gehört, lassen sich die anderen beiden Handlungen der Kategorie „Versuch aufbauen (4)“ zuordnen. Da nur eine Kategorie gewählt werden kann, wird die Kategorie zugeordnet die den Großteil des Intervalls ausfüllt. In diesem Fall ist das Lesen im Testheft etwas länger als die restlichen Handlungen. Entsprechend wird auch für dieses Intervall „Klärung der Fragestellung (1)“ zugewiesen.

Intervall	Zeitspanne	Handlung	Kodierung
5	40-50	Probandin liest im Testheft, in der Inventarliste und nimmt Komponenten aus dem Koffer. Schwerpunkt: Probandin liest im Testheft	Klärung der Fragestellung (2)
6	50-60	Probandin agiert mit dem Experimentiermaterial.	Versuch aufbauen (4)
7-8	60-80	Probanden agiert mit dem Experimentiermaterial.	Versuch aufbauen (4)

Nach diesem Verfahren wird das ganze Video angeschaut und die Intervalle werden bzgl. der Kategorien eingeordnet.

Um die Handlungen zu identifizieren, gibt es aufgabenspezifischen expliziten Erwartungen was in den jeweiligen Handlungen passieren soll und welche Aufzeichnungen dabei angefertigt werden. Die Erwartungen finden sich im Anhang (siehe A.1).

### **Hinweise**

Es ist möglich, dass Sonderfälle auftreten, die zu Problemen bei der Identifikation von Handlungen führen. Welche Sonderfälle das sind und wie in diesem Fall bei der Kodierung vorzugehen ist, wird im Folgenden beschrieben.

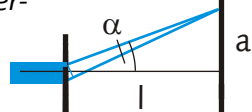
#### **Handlungen die erst später identifizierbar werden**

Es ist möglich, dass eine Handlung nicht im ersten Intervall identifiziert werden kann, sondern erst durch die folgenden Intervalle erkennbar wird. Bevor also ein Intervall mit einer der „Sonstigen“ Kategorien markiert wird, werden die folgenden Intervalle angeschaut.

#### **Aufzeichnungen aus dem Testheft zuordnen**

Zwar gibt es zwei Videos, eines für das Experiment und eines für das Testheft. Dennoch lässt sich auf den Videos nicht erkennen was genau die Probanden schreiben. Entsprechend soll das Testheft genutzt werden, um eindeutig identifizieren zu können, welche Aufzeichnungen der Proband zu bestimmten Zeiten anfertigt.

**Tabelle 1: Kategorien zur Kodierung der Handlungen im Laborheft**

Handlung	Beschreibung	Indikatoren	Beispiel	Abgrenzung
Fragestellung klären (1)	Die Fragestellung wird gelesen und verstanden.	Die Fragestellung wird wörtlich oder in eigenen Worten aufgeschrieben. ODER Es wird im Testheft die Einführung zur Aufgabe bzw. die Fragestellung gelesen.		
Wissen aktivieren und theoretische Grundlage schaffen (2)	Das für die Bearbeitung der Fragestellung notwendige Wissen wird aktiviert.	Es werden Formeln oder bekannte Größen notiert, Formeln umgestellt oder im Theorieblatt gelesen.	<i>Formel zur Berechnung der Gitterkonstante: <math>n \cdot \ddot{e} = g \cdot \sin(\alpha_n)</math>; <math>\sin(\alpha_n) = a_i / l</math>.</i> <i>Variablen: <math>n_1=1</math>; <math>\ddot{e}_{\text{grün}}=532 \text{ nm}</math></i>	
Experiment planen (3)	Die Planung eines Experiments zur Beantwortung der Fragestellung.	Eine Versuchsbeschreibung wird notiert, dazu gehört auch eine Skizze. ODER Die zu messenden Werte werden notiert. ODER Es werden Überschriften eingefügt.	<i>Skizze eines Versuchsaufbaus (siehe Abbildung) und/oder Beschreibung des Versuchs.</i> 	Abgrenzung zu „Versuch aufbauen (4)“: Die Skizze wird vor dem Aufbau des Versuchs angefertigt und enthält in der Regel keine konkreten Werte und Abmessungen.
Versuch aufbauen (4)	Ein Versuch wird aufgebaut.	Eine Zeichnung des Ausbaus wird angefertigt in der alle relevanten Größen enthalten sind um den Versuchsaufbau 1:1 nachzubauen. ODER Nimmt Komponenten, positioniert Komponenten auf der Arbeitsfläche, verbindet Halterungen mit Objekten oder prüft den Strahlengang. ODER Liest in der Inventarliste. Ein Versuch gilt spätestens als fertig aufgebaut, wenn damit gemessen wird.		Abgrenzung zu „Experiment planen (3)“: Die Zeichnung wird nach dem Aufbau des Versuchs angefertigt und enthält in der Regel konkrete Werte und Abmessungen wodurch der Aufbau 1:1 nachgebaut werden kann.

Messen und Beobachten (5)	Gezielt Messwerte aufnehmen oder Beobachtungen anstellen.	Verändert den Aufbau um eine Messung durchzuführen. Verändert Einstellungen an Messgeräten (Luxmeter, Polarisationsfilter) oder Veränderung des zu messenden Gegenstands (z.B. Zucker ins Wasser geben) Notiert Messwerte und Beobachtungen.	$n=1; \lambda=532nm; a_i=2,2cm; l=30cm.$	Abgrenzung zu „Versuch aufbauen (4)“: Der Aufbau wird nicht verändert außer um eine Messung durchzuführen.
Daten analysieren (6)	Auf Basis von Messwerten werden Berechnungen durchgeführt oder Graphen gezeichnet.	Wertet die Daten aus indem er Rechnungen durchführt, Graphen zeichnet oder Werte aus Graphen abliest. Dazu kann es auch gehören eine Fehlerrechnung anzufertigen oder Fehlerbalken zu zeichnen.	$g=n\lambda/a_i=7,25*10^{-6}m$	
Sachgerechte Schlüsse ziehen (7)	Berechnungen, Messgrößen und Beobachtungen werden auf die Fragestellung bezogen um zu prüfen, ob die Fragestellung beantwortet wurde. Es werden Schlussfolgerungen aus den Messergebnissen/ Beobachtungen gezogen	Es wird eine Antwort formuliert. Messfehler und Ungenauigkeiten können zur Erklärung genutzt werden. ODER Messwerte werden mit Tabellen aus der Theorie verglichen. ODER Fehler werden bemerkt und notiert bzw. fehlerhafte Aufzeichnungen (z.B. Messwerte) durchgestrichen.	<i>Die Gitterkonstante liegt im erwarteten Bereich. Da keine Fehler bei der Messung aufgetreten sind, wird angenommen, dass die Gitterkonstante <math>g \approx 7,25 * 10^{-6} m</math> ist.</i>	Abgrenzung zu „Wissen aktivieren (2)“: Nach Berechnung von Größen wird in Tabellen im Theorieteil gelesen.
Sonstiges (9)	Sonstige experimentierbezogenen Handlungen	Der Proband experimentiert, die Handlungen können aber keiner der Kategorien zugeteilt werden.	<i>Der Proband denkt nach, ohne das ersichtlich ist worüber er nachdenkt oder betrachtet Aufzeichnungen bzw. den Versuchsaufbau.</i>	
Sonstiges nicht Experimentieren (10)	Sonstige nicht experimentelle Handlung	Der Proband macht Handlungen die nichts mit der Aufgabenstellung zu tun haben.	<i>Der Proband macht eine Pause und trinkt etwas.</i>	

### 3.2 Richtigkeit

Nachdem die Handlungen identifiziert wurden, erfolgt nun deren Kodierung bezüglich der Richtigkeit. Dafür werden die Produkte der Handlungen genutzt. Jeder Handlung kann dabei einer von sechs Codes bzgl. der korrekten Durchführung zugeordnet werden:

- (1) Das Produkt der Handlung entspricht der Musterlösung oder ist äquivalent dazu (richtig).
- (0,5) Das Produkt der Handlung entspricht in Teilen der Musterlösung bzw. ist äquivalent dazu UND es fehlen Teile der Musterlösung oder es sind Fehler im Produkt (teilweise richtig)
- (0) Das Produkt der Handlung entspricht nicht der Musterlösung oder ist durchgehend falsch (falsch)
- (99) Die Handlung und das Produkt sind nicht interpretierbar oder es liegt kein Produkt vor (nicht interpretierbar)
- (77) Das Produkt zu der Handlung kommt mindestens zum zweiten Mal in identischer Weise vor (Wiederholung)
- (66) Das Produkt zu einer Handlung ist noch nicht vollständig (z.B. wird nur die Hälfte eines Versuches aufgebaut oder eine Messreihe enthält nur die ersten Werte) und wird in einem späteren Intervall zu dieser Handlung vervollständigt (Zwischenschritt, keine Wertung)

Um die Handlungen zu bewerten wird nicht jedes einzelne 10-Sekunden Intervall bewertet, sondern das gesamte Intervall in dem nur eine Handlung kodiert wurde. Dazu wird die Richtigkeit nur im letzten 10-Sekunden Intervall durchgeführt und in den anderen Fällen einfach offen gelassen. Jedem dieser Handlungs-Intervalle können die beschriebenen Codes zugeordnet werden. Als Hilfestellung finden sich im Anhang aufgabenspezifische Tabellen (siehe A.2). In diesen wird beschrieben, was das Kriterium in den Aufzeichnungen ist, anhand dessen die Richtigkeit bewertet wird. Zudem findet sich die Musterlösung, die zur Bewertung als Referenz genutzt wird. Im Folgenden wird die Bewertung der Richtigkeit an dem Beispiel aus 3.1 dokumentiert.

Intervall	Zeitspanne	Handlung	Kodierung	Richtigkeit
1-5	0-50	Probandin liest im Testheft	Klärung der Fragestellung (1)	
6-8	50-80	Probandin agiert mit dem Experimentiermaterial.	Versuch aufbauen (4)	
9-10	80-100	Probandin liest im Testheft.	Klärung der Fragestellung (2)	

Im ersten Intervall des Videos liest die Probandin im Testheft. Entsprechend wurde kodiert, dass sie die Handlung „Klärung der Fragestellung (1)“ durchführt. Da in diesem Abschnitt aber kein Produkt entsteht, kann keine Wertung der Richtigkeit vorgenommen werden. Die Handlung wird entsprechend mit „99“ (nicht interpretierbar) kodiert.

Im nächsten Intervall wählt die Probandin eine Komponente und beginnt mit dem Aufbau eines Versuchs, wie er in der Musterlösung beschrieben ist. Der Versuchsaufbau wird aber nicht abgeschlossen. Aus den nachfolgenden Intervallen zum Aufbau des Versuchs wird aber eindeutig sichtbar, dass der Versuch komplett und richtig aufgebaut wird. Das Intervall zur Handlung „Versuch aufbauen (4)“ würde entsprechend mit einer „66“ (Zwischenschritt, keine Wertung) kodiert werden.

Im dritten Abschnitt liest die Probandin wieder im Testheft. Da kein Produkt entsteht kann diese Handlung nicht bzgl. der Richtigkeit bewertet werden. Entsprechend wird „99“ (nicht interpretierbar) kodiert.

Intervall	Zeitspanne	Handlung	Kodierung	Richtigkeit
1-5	0-50	Probandin liest im Testheft	Klärung der Fragestellung (1)	„99“
6-8	50-80	Probandin agiert mit dem Experimentiermaterial.	Versuch aufbauen (4)	„66“
9-10	80-100	Probandin liest im Testheft.	Klärung der Fragestellung (2)	„99“

Entsprechend wird für alle Handlungen verfahren.

### **Wiederholungen**

Kommt ein Produkt in identischer Form an mehreren Stellen vor (z.B. eine Formel, eine Variable  $n_1=1$ , der erneute Aufbau eines identischen Versuchs, ...) dann wird nur für die erste Stelle eine Richtigkeit kodiert. Die Wiederholungen werden mit „77“ (Wiederholung) kodiert.

### **Kein vorliegendes Bewertungskriterium**

Liegt zu einer Handlung kein Kriterium der Bewertung von Richtigkeit vor (z.B. wenn keine Messwerte notiert wurden) dann wird „99“ (nicht interpretierbar) bewertet.

### **Bewertung der Kategorie: „Messen und Beobachten“**

Zur Bewertung der Richtigkeit in der Kategorie „Messen und Beobachten“ werden zwei Aspekte beurteilt: (1) die eigentliche Handlung, also ob der Proband an der richtigen Stelle und mit dem richtigen Messgerät arbeitet; und (2) der Messwert, also ob die aufgenommenen Werte des Probanden richtig sind. Zur Bewertung der Richtigkeit von Messwerten ist unter den aufgabenspezifischen Kodieranweisungen für jeden Messwert ein Intervall angegeben, in dem der Messwert als „richtig“ bezeichnet werden kann. Um aus den beiden Aspekten einen Wert für die Richtigkeit des Probanden für ein Handlungsintervall zu erhalten hilft die nachfolgende Tabelle 2.



Tabelle 32 Richtigkeit bezüglich der Teilkategorien der Kategorie "Messen und Beobachten"

		Handlung des Probanden ist ...		
		richtig	falsch	nicht interpretierbar
Messwerte ist ...	richtig	Kategorie Messen und Beobachten ist als <b>richtig</b> zu bewerten	Kategorie Messen und Beobachten ist als <b>falsch</b> zu bewerten	Kategorie Messen und Beobachten ist als <b>richtig</b> zu bewerten
	falsch	Kategorie Messen und Beobachten ist als <b>falsch</b> zu bewerten	Kategorie Messen und Beobachten ist als <b>falsch</b> zu bewerten	Kategorie Messen und Beobachten ist als <b>falsch</b> zu bewerten
	nicht interpretierbar	Kategorie Messen und Beobachten ist als <b>nicht interpretierbar</b> zu bewerten	Kategorie Messen und Beobachten ist als <b>falsch</b> zu bewerten	Kategorie Messen und Beobachten ist als <b>nicht interpretierbar</b> zu bewerten

## Anhang

### A.1 Aufgabenspezifische Kodieranweisungen zur Identifikation der Handlungen

#### Vermessung einer Glühwendel

Handlung	Handlung
Fragestellung klären (1)	Lese oder Notiere eine Fragestellung
Wissen aktivieren (2)	Notiere oder Beschreibe Formeln bekannte Größen
Experiment planen (3)	Versuchsbeschreibung inklusive Skizze wird notiert und die Messgrößen werden festgelegt
Versuch aufbauen (4)	Versuch wird aufgebaut, Zeichnung wird angefertigt
Messen / Beobachten (5)	Messen der Werte für $g$ , $b$ , $B_i$
Daten analysieren (6)	Berechnen von $G_i$
Sachgerechte Schlüsse ziehen (7)	Schreiben eines Lösungssatzes
Sonstiges (9)	Sonstige experimentierbezogene Handlungen
Sonstiges nicht Experiment (10)	Sonstige nicht experimentelle Handlung

#### Optische Dichte

Handlung	Produkt
Fragestellung klären (1)	Lese oder Notiere eine Fragestellung
Wissen aktivieren (2)	Notiere oder Beschreibe Formeln bekannte Größen
Experiment planen (3)	Versuchsbeschreibung inklusive Skizze wird notiert und die Messgrößen werden festgelegt
Versuch aufbauen (4)	Versuch wird aufgebaut, Zeichnung wird angefertigt
Messen / Beobachten (5)	Messen der Werte für Einfallswinkel $\alpha_1$ und Ausfallswinkel $\alpha_2$ für Laser verschiedener Wellenlänge
Daten analysieren (6)	Berechnung von $n_2$
Sachgerechte Schlüsse ziehen (7)	Schreiben eines Lösungssatzes
Sonstiges (9)	Sonstige experimentierbezogene Handlungen
Sonstiges nicht Experiment (10)	Sonstige nicht experimentelle Handlung

#### Spektroskopie

Handlung	Produkt
Fragestellung klären (1)	Lese oder Notiere eine Fragestellung
Wissen aktivieren (2)	Notiere oder Beschreibe Formeln bekannte Größen
Experiment planen (3)	Versuchsbeschreibung inklusive Skizze wird notiert und die Messgrößen werden festgelegt
Versuch aufbauen (4)	Versuch wird aufgebaut, Zeichnung wird angefertigt
Messen / Beobachten (5)	Messen der Werte für $a_i$ und $l$
Daten analysieren (6)	Berechnung von $g$ oder $\ddot{e}$
Sachgerechte Schlüsse ziehen (7)	Schreiben eines Lösungssatzes
Sonstiges (9)	Sonstige experimentierbezogene Handlungen
Sonstiges nicht Experiment (10)	Sonstige nicht experimentelle Handlung

### Zuckerwasser

Handlung	Produkt
Fragestellung klären (1)	Lese oder Notiere eine Fragestellung
Wissen aktivieren (2)	Notiere oder Beschreibe Formeln bekannte Größen
Experiment planen (3)	Versuchsbeschreibung inklusive Skizze wird notiert und die Messgrößen werden festgelegt
Versuch aufbauen (4)	Versuch wird aufgebaut, Zeichnung wird angefertigt
Messen / Beobachten (5)	Messen der Minima bzw. Maxima und des Drehwinkels zwischen den Polarisationsfiltern $\alpha$ Zucker wird in die Küvette gegeben
Daten analysieren (6)	Berechnung der Zuckerkonzentration und der Zuckerkonzentration pro Liter
Sachgerechte Schlüsse ziehen (7)	Schreiben eines Lösungssatzes
Sonstiges (9)	Sonstige experimentierbezogene Handlungen
Sonstiges nicht Experiment (10)	Sonstige nicht experimentelle Handlung

### Polarisationseigenschaften

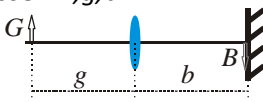
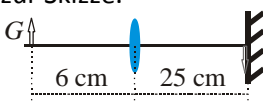
Handlung	Produkt
Fragestellung klären (1)	Lese oder Notiere eine Fragestellung
Wissen aktivieren (2)	Notiere oder Beschreibe Formeln bekannte Größen
Experiment planen (3)	Versuchsbeschreibung inklusive Skizze wird notiert und die Messgrößen werden festgelegt
Versuch aufbauen (4)	Versuch wird aufgebaut, Zeichnung wird angefertigt
Messen / Beobachten (5)	Messen der Intensität des transmittierten- und reflektierten Strahls in Abhängigkeit zum Einfallswinkel UND Messen der Intensität von transmittierten- und reflektierten Strahl in Abhängigkeit zum Polarisationsfilter bei festem „Winkel der stärksten Polarisation“
Daten analysieren (6)	Ablesen des „am stärksten polarisierenden Winkels“
Sachgerechte Schlüsse ziehen (7)	Schreiben eines Lösungssatzes
Sonstiges (9)	Sonstige experimentierbezogene Handlungen
Sonstiges nicht Experiment (10)	Sonstige nicht experimentelle Handlung

**Blackbox**

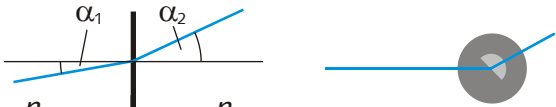
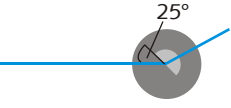
<b>Handlung</b>	<b>Produkt</b>
Fragestellung klären (1)	Lese oder Notiere eine Fragestellung
Wissen aktivieren (2)	Notiere oder Beschreibe Formeln bekannte Größen
Experiment planen (3)	Versuchsbeschreibung inklusive Skizze wird notiert und die Messgrößen werden festgelegt
Versuch aufbauen (4)	Versuch wird aufgebaut, Zeichnung wird angefertigt
Messen / Beobachten (5)	Beobachten und Messen
Daten analysieren (6)	Berechnung der Gitterposition
Sachgerechte Schlüsse ziehen (7)	Schreiben eines Lösungssatzes
Sonstiges (9)	Sonstige experimentierbezogene Handlungen
Sonstiges nicht Experiment (10)	Sonstige nicht experimentelle Handlung

## A.2 Aufgabenspezifische Kodieranweisungen zur Richtigkeit

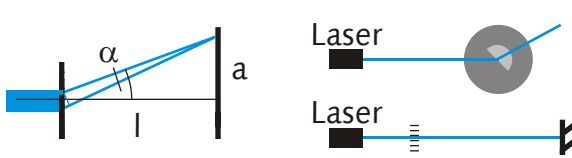
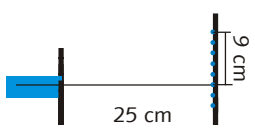
### Vermessung einer Glühwendel

Handlung	Kriterium für Richtigkeit	Musterlösung
Fragestellung klären (1)	Notierte Fragestellung	Wie ist die Abmessung der Glühwendel von der Lampe (Inventarnr. 3)?
Wissen aktivieren (2)	Beschreibungen und Erklärungen zur Theorie und Formeln	Formel: $G=B \cdot g/b$ ODER Proband liest die folgenden Abschnitte im Theorieheft und identifiziert diese als Aufgabenrelevant: - Optische Komponenten und ihre Anwendungen inklusive Abbildungsgleichung
Experiment planen (3)	Skizzen, Erklärungen zum Versuchsaufbau ODER Messgrößen	Zu Messen: $B, g, b$ Skizze: 
Versuch aufbauen (4)	Zeichnung zum aufgebauten Versuch ODER Aufbau eines Versuchs wie in der Musterlösung	Material, Aufbau und Versuchszeichnung äquivalent zur Skizze. 
Messen und Beobachten (5)	Messwerte für $g, b, B_i$	Zur Kodierung muss der Kodierer $G_i$ aus den Messwerten $g, b, B_i$ anhand der Formel $G_i = B_i \cdot g/b$ berechnen. Die Phase gilt als richtig wenn $G_i$ zwischen 2,5mm und 3,5mm oder zwischen 0,5mm und 1,5mm liegt.
Daten analysieren (6)	Berechnen von $G_i$	Richtig, wenn das vom Probanden aus den Messwerten $g, b, B_i$ errechnete $G_i$ mit dem vom Kodierer errechneten $G_i$ übereinstimmt.
Sachgerechte Schlüsse ziehen (7)	Lösungssatz oder Erklärung der Messwerte.	Abhängig von bisher errechneten Daten muss entweder die Länge $a$ im Bereich von 2,5mm und 3,5mm liegen oder die Breite $b$ im Bereich von 0,5mm und 1,5mm liegen. ODER Erklärung, dass die Ergebnisse nicht stimmen können.
Sonstiges (9)	Sonstige Handlung	Richtigkeit wird nicht kodiert „-“
Sonstiges nicht Experiment (10)	Sonstige nicht experimentelle Handlung	Richtigkeit wird nicht kodiert „-“

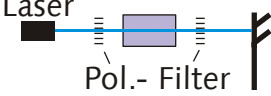
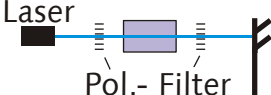
## Optische Dichte

Handlung	Kriterium für Richtigkeit	Musterlösung
Fragestellung klären (1)	Notierte Fragestellung	Aus welchem Material ist das Prisma (Inventarnr. 14)?
Wissen aktivieren (2)	Beschreibungen und Erklärungen zur Theorie und Formeln	Formel: $n_1 \cdot \sin(\alpha_1) = n_2 \cdot \sin(\alpha_2)$ Variable: $n_1=1$ ODER Proband liest die folgenden Abschnitte im Theorieheft und identifiziert diese als Aufgabenrelevant: - Brechung inklusive Brechungsgesetz
Experiment planen (3)	Skizzen, Erklärungen zum Versuchsaufbau ODER Messgrößen	Zu Messen: $\alpha_1, \alpha_2$ Skizze:  Überschriften bzgl. der einzelnen Laser
Versuch aufbauen (4)	Zeichnung zum aufgebauten Versuch ODER Aufbau eines Versuchs wie in der Musterlösung	Material, Aufbau und Versuchszeichnung äquivalent zur Skizze. 
Messen und Beobachten (5)	Einfallswinkel $\alpha_1$ und Ausfallswinkel $\alpha_2$ für Laser verschiedener Wellenlänge ODER Austausch des Laserpointers	Zur Kodierung muss der Kodierer $n_2$ aus den Messwerten $\alpha_1, \alpha_2$ für $n_1 = 1$ berechnen. Die Phase gilt als richtig wenn $n_2$ zwischen 1,40 und 1,55 liegt.
Daten analysieren (6)	Berechnung von $n_2$	Richtig, wenn das vom Probanden aus den Messwerten $\alpha_1$ und $\alpha_2$ errechnete $n_2$ mit dem vom Kodierer errechneten $n_2$ übereinstimmt.
Sachgerechte Schlüsse ziehen (7)	Lösungssatz ODER Es wird der Graph in den Theorieangaben gelesen.	Es wurde Graph „F“ gewählt ODER Erklärung das ein Messfehler vorliegen muss
Sonstiges (9)	Sonstige Handlung	Richtigkeit wird nicht kodiert „-“
Sonstiges nicht Experiment (10)	Sonstige nicht experimentelle Handlung	Richtigkeit wird nicht kodiert „-“

## Spektroskopie

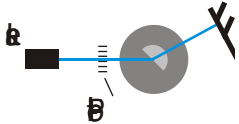
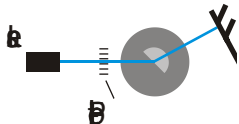
Handlung	Kriterium für Richtigkeit	Musterlösung
Fragestellung klären (1)	Notierte Fragestellung	Wie ist die Wellenlänge des unbekanntes Lasers (Inventarnr. 5)?
Wissen aktivieren (2)	Beschreibungen und Erklärungen zur Theorie und Formeln	Formel: $n \cdot \ddot{e} = g \cdot \sin(\alpha_n)$ ; $\sin(\alpha_n) = a_i / l$ ; $\tan(\alpha_n) = a_i / l'$ Variable : $n_1=1$ ; $\ddot{e}_{blau}=430nm$ ; $\ddot{e}_{grün}=532nm$ ODER Proband liest die folgenden Abschnitte im Theorieheft und identifiziert diese als Aufgabenrelevant: - Wechselwirkung von Licht mit Materie inklusive der Gleichungen: $n = g \sin \alpha, \quad a_n = l \sin \alpha$ - Dispersionsgraph
Experiment planen (3)	Skizzen, Erklärungen zum Versuchsaufbau ODER Messgrößen	Zu Messen: $a_i$ , $l$ bzw. $a_i$ , $l'$ bzw. $\alpha_i$ Skizze : 
Versuch aufbauen (4)	Zeichnung zum aufgebauten Versuch ODER Aufbau eines Versuchs wie in der Musterlösung	Material, Aufbau und Versuchszeichnung äquivalent zur Skizze. 
Messen und Beobachten (5)	Messwerte $a_i$ und $l$ ODER Messwerte $\alpha_{\text{Einfallswinkel}}$ und $\alpha_{\text{Ausfallswinkel}}$	Zur Kodierung muss der Kodierer $g$ aus den Messwerten $a_i$ und $l$ für $\ddot{e}_{\text{rot}} = 642 \text{ nm}$ ; $\ddot{e}_{\text{grün}} = 532 \text{ nm}$ ; $\ddot{e}_{\text{blau}} = 430 \text{ nm}$ berechnen: $g = n \cdot \ddot{e} / a$ Die Phase gilt als richtig wenn $g$ zwischen $7,00 \cdot 10^{-6} \text{ m}$ und $7,25 \cdot 10^{-6} \text{ m}$ liegt.
Daten analysieren (6)	Berechnung von $g$ oder $\ddot{e}$	Zur Kodierung muss der Kodierer $g$ bzw. $\ddot{e}$ aus den Messwerten nachrechnen. Die Phase gilt als richtig wenn ein entsprechendes Ergebnis von Kodierer und Proband erzielt wird.
Sachgerechte Schlüsse ziehen (7)	Lösungssatz	Richtig wenn $\lambda_{\text{rot}}$ zwischen 600 nm und 680 nm ODER Erklärung, dass ein Messfehler vorliegen muss
Sonstiges (9)	Sonstige Handlung	Richtigkeit wird nicht kodiert „-“
Sonstiges nicht Experiment (10)	Sonstige nicht experimentelle Handlung	Richtigkeit wird nicht kodiert „-“

## Zuckerwasser

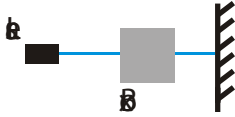
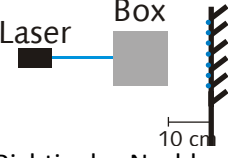
Handlung	Kriterium für Richtigkeit	Musterlösung
Fragestellung klären (1)	Notierte Fragestellung	Bei welcher Zuckerkonzentration wird die Polarisationssebene um $0,75^\circ/\text{cm}$ gedreht?
Wissen aktivieren (2)	Beschreibungen und Erklärungen zur Theorie und Formeln	Formel: $V=h \cdot l \cdot b$ ; $\rho=m/V$ ODER Proband liest die folgenden Abschnitte im Theorieheft und identifiziert diese als Aufgabenrelevant: - Polarisation
Experiment planen (3)	Skizzen, Erklärungen zum Versuchsaufbau ODER Messgrößen	Skizze:   Durchführung: - Vermessung der Küvette - Erhöhung der Zuckerkonzentration um eine Einheit - Ablesen des Winkel bei dem die Lichtintensität minimal/maximal wird - Bestimmung der Konzentration für den Drehwinkel von $0,75^\circ/\text{cm}$
Versuch aufbauen (4)	Zeichnung zum aufgebauten Versuch ODER Aufbau eines Versuchs wie in der Musterlösung	Material, Aufbau und Versuchszeichnung äquivalent zur Skizze.  
Messen und Beobachten (5)	Drehung der Polarisationssebene $\alpha$	Richtig wenn die minimale und die maximale Auslöschung des Lichts bei einem Winkel von $85^\circ - 95^\circ$ zwischen den Polarisationsfiltern auftritt. ODER Richtige Messung, wenn sich Anzahl an Löffeln Zucker für die Änderung die Polarisationssebene um $0,75^\circ/\text{cm}$ zwischen $9 \cdot 1000 \text{cm}^3/\text{Volumen}(\text{in cm}^3)$ und $15 \cdot 1000 \text{cm}^3/\text{Volumen}(\text{in cm}^3)$ liegt. ODER Volumen= $10 \text{cm} \cdot 2,8 \text{cm} \cdot b$ , mit $b$ zwischen 1cm und 6,5 cm.
Daten analysieren (6)	Berechnung der Zuckerkonzentration und der Zuckerkonzentration pro Liter	Richtig, wenn die Zuckerkonzentration bei einer Drehung der Polarisationssebene um $0,75^\circ/\text{cm}$ aus den im Wasser gelösten Zucker (Einheit in Löffeln) und dem Drehwinkel berechnet wird. UND Die Konzentration wird wenn nötig auf Liter umgerechnet werden
Sachgerechte Schlüsse ziehen (7)	Lösungssatz	Richtiges Ergebnis, wenn 9-15 Löffel angegeben werden ODER Erklärung eines Messfehlers
Sonstiges (9)	Sonstige Handlung	Richtigkeit wird nicht kodiert „-“
Sonstiges nicht Experiment (10)	Sonstige nicht experimentelle Handlung	Richtigkeit wird nicht kodiert „-“



## Polarisationseigenschaften

Handlung	Kriterium für Richtigkeit	Musterlösung
Fragestellung klären (1)	Notierte Fragestellung	Wie ist der Winkel unter dem das Licht linear polarisiert wird? Wie ist die Schwingungsebene des transmittierten- und des reflektierten Lichts?
Wissen aktivieren (2)	Beschreibungen und Erklärungen zur Theorie und Formeln	Proband liest die folgenden Abschnitte im Theorieheft und identifiziert diese als Aufgabenrelevant: - Polarisation - Brechung
Experiment planen (3)	Skizzen, Erklärungen zum Versuchsaufbau ODER Messgrößen	Zu Messen: $\alpha_1, \alpha_2$ Skizze: 
Versuch aufbauen (4)	Zeichnung zum aufgebauten Versuch ODER Aufbau eines Versuchs wie in der Musterlösung	Material, Aufbau und Versuchszeichnung äquivalent zur Skizze. (Polarisationsfilter hinter dem Prisma anstelle davor ist ein äquivalenter Aufbau) 
Messen und Beobachten (5)	Messen der Intensität des transmittierten- und reflektierten Strahls in Abhängigkeit zum Einfallswinkel UND Messen der Intensität von transmittierten- und reflektierten Strahl in Abhängigkeit zum Polarisationsfilter bei festem „Winkel der stärksten Polarisation“	Richtig, wenn das Maximum des transmittierten/reflektierten und das Minimum des reflektierten/transmittierten Strahls bei einem Einfallswinkel von $\alpha = 50^\circ$ - $62^\circ$ auftreten UND $\beta, \gamma$ stehen senkrecht aufeinander ( $90^\circ$ Differenz im Winkel)
Daten analysieren (6)	Ablesen des „am stärksten polarisierenden Winkels“	Richtig, wenn der „Winkel der stärksten Polarisation“ gewählt wird, wenn der transmittierte/reflektierte Strahl sein Maximum/Minimum erreicht
Sachgerechte Schlüsse ziehen (7)	Lösungssatz	Der „am stärksten polarisierende Winkel“ $\alpha = 50^\circ$ - $62^\circ$ UND $\beta, \gamma$ stehen senkrecht aufeinander ( $90^\circ$ Differenz im Winkel) ODER Erklärung eines Messfehlers
Sonstiges (9)	Sonstige Handlung	Richtigkeit wird nicht kodiert „-“
Sonstiges nicht Experiment (10)	Sonstige nicht experimentelle Handlung	Richtigkeit wird nicht kodiert „-“

## Blackbox

Handlung	Kriterium für Richtigkeit	Musterlösung
Fragestellung klären (1)	Notierte Fragestellung	Wie sind die optischen Eigenschaften des unbekanntes Bauteils?
Wissen aktivieren (2)	Beschreibungen und Erklärungen zur Theorie und Formeln	Proband liest die folgenden Abschnitte im Theorieheft und identifiziert diese als Aufgabenrelevant: - Intensität - Polarisation - Wechselwirkung von Licht mit Materie - Optische Komponenten und ihre Anwendungen
Experiment planen (3)	Skizzen, Erklärungen zum Versuchsaufbau ODER Messgrößen	Skizze: oder Skizze der Blackbox   Zu untersuchende optische Eigenschaften.
Versuch aufbauen (4)	Zeichnung zum aufgebauten Versuch ODER Aufbau eines Versuchs wie in der Musterlösung ODER Nachbau der Blackbox	Material, Aufbau und Versuchszeichnung äquivalent zur Skizze. (Einführung verschiedener optischer Komponenten wie z.B. Polarisationsfilter in den Strahlengang sind äquivalent)   Richtig des Nachbaus: Wenn mindestens eine Komponente im Nachbau enthalten ist und mit der Blackbox verglichen werden kann.
Messen und Beobachten (5)	Beobachten und Messen	Richtig, wenn: Bemerk wird, dass Licht bestimmter Wellenlängen gefiltert wird UND Ein Beugungsbild entsteht UND $a_i$ für die Abstände der $i$ -ten Beugungsordnung zur optischen Achse gemessen ODER Ausschluss von Eigenschaften die nicht Bestandteil der box sind ODER beim Nachbau: Wenn Bilder von Nachbau und Blackbox verglichen werden ODER beim Nachbau: Wenn Komponenten im nachbau getauscht werden oder in einem anderen Slot platziert werden
Daten analysieren	Berechnung der Gitterposition	Richtig wenn berechnet oder erklärt wird,

(6)		dass sich das Gitter an Position 1 befinden muss UND Das die Position des Farbfilters nicht eindeutig bestimmt werden kann
Sachgerechte Schlüsse ziehen (7)	Lösungssatz	Richtig wenn: Ein Farbfilter diagnostiziert wird UND Ein Gitter in einem der Randpositionen diagnostiziert wird UND Der letzte Slot leer ist oder nicht gesagt werden kann, was sich im letzten Slot befindet. ODER Optische Bauteile ausgeschlossen werden Teilweise richtig (0,5 Punkte) wenn zwei der Slots richtig sind.
Sonstiges (9)	Sonstige Handlung	Richtigkeit wird nicht kodiert „-“
Sonstiges nicht Experiment (10)	Sonstige nicht experimentelle Handlung	Richtigkeit wird nicht kodiert „-“

## **A.4.8 Kodiermanual Protokollierung**

### **Protokollbasierte Auswertung von Experimentierkompetenz**

#### **1. Einleitung**

Das vorliegende Manual dient der Auswertung der Aufzeichnungen, die Probanden bei der Bearbeitung von sechs Experimentieraufgaben angefertigt haben. Die Aufzeichnungen sind vergleichbar mit solchen, wie sie z.B. im Physikalischen Praktikum für Physiker entstehen. Anhand der Aufzeichnungen sollen mit Hilfe des vorliegenden Manuals die Handlungen, die beim Experimentieren durchgeführt werden, identifiziert und hinsichtlich ihrer Richtigkeit bewertet werden.

#### **2. Grundlagen**

Experimentieren ist ein komplexer Prozess, beim dem unterschiedliche Fähigkeiten und Fertigkeiten eine Rolle spielen. In physik- bzw. naturwissenschaftsdidaktischer Literatur hat sich eine Gliederung entsprechend der Handlungen beim idealtypischen Experimentieren als potentiell iterativer Prozess etabliert. Experimentieren beginnt idealtypisch mit der Planung eines Experiments. Das heißt, aus theoretischen Überlegungen oder empirischen Beobachtungen wird eine Fragestellung entwickelt (1) und es werden dazu passende fachliche Hypothesen gebildet (2). Anschließend wird ein geeignetes Experiment geplant (3), mit dem sich die Frage beantworten lässt. Auf die Planungsphase folgt die Durchführung. Im Rahmen der Durchführung wird zunächst ein Versuch aufgebaut (4) und anschließend wird gezielt gemessen und beobachtet (5). Den Abschluss des idealtypischen Experimentierens bildet die Auswertung. Dabei werden Daten analysiert (6) indem Berechnungen auf Basis der Messwerte durchgeführt oder Graphen angefertigt werden. Anschließend werden sachgerechte Schlüsse (7) gezogen, indem die Beobachtungen und Ergebnisse der Analyse von Daten auf die Fragestellung bezogen werden.

Diese Handlungen bilden die Grundlage für die Beschreibung von Experimentierkompetenz. Experimentierkompetenz umfasst demnach die Fähigkeiten und Fertigkeiten die entsprechenden Handlungen durchführen zu können. Die Ausprägung von Experimentierkompetenz lässt sich dann anhand von drei Kriterien beschreiben: 1) Richtigkeit: Erfolgreiches Experimentieren setzt die richtige Durchführung der einzelnen Handlungen voraus. 2) Strukturiertheit: Die Handlungen müssen in einer sinnvollen Reihenfolge durchgeführt werden. 3) Zielorientiertheit: Die Handlungen müssen zu Beantwortung der Fragestellung führen. Damit ergeben sich drei Dimensionen der Experimentierkompetenz auf der Grundlage der Handlungen des idealtypischen Experimentierens.

In der Realität lässt sich der Prozess idealtypischen Experimentierens jedoch nur selten vollständig beobachten. Insbesondere im physikalischen Praktikum der universitären Ausbildung wird häufig auf die Entwicklung eigenständiger Fragestellungen und die selbständige Formulierung von Hypothesen verzichtet. Stattdessen wird von den Studierenden erwartet, dass sie sich nach einem Studium der theoretischen Grundlagen eines Versuchs mit einer entsprechenden experimentellen Aufgabenstellung auseinandersetzen und diese experimentell bearbeiten. Statt der Entwicklung einer Fragestellung (1) muss nun also die Aufgabenstellung gelesen und verstanden werden (1'); statt der Formulierung entsprechender Hypothesen (2) muss Wissen aktiviert werden, z.B. indem Literatur gelesen wird, um die theoretischen Grundlagen für die Bearbeitung der Aufgabenstellung zu schaffen (2').

Die Experimentieraufgaben im Rahmen deren Bearbeitung die zu kodierenden Aufzeichnungen entstanden sind, sind entsprechend konstruiert. Im Rahmen der Bearbeitung der Aufgaben sollten also prinzipiell die folgenden Handlungen auftreten (können): Fragestellung lesen und verstehen (1'),

Aktivierung des Wissens zur Schaffung einer theoretischen Grundlagen (2'), Experiment planen (3), Versuch aufbauen (4), Messen und Beobachten (5), Daten auswerten (6) und Schlussfolgerungen ziehen (7).

### 3. Kodieranweisungen

Die folgenden Anweisungen dienen der Kodierung der Handlungen die bei der Bearbeitung der Experimentieraufgaben entstanden sind. Die Kodierung gliedert sich in zwei Schritte: die Identifikation der Handlungen in den Aufzeichnungen und die Bewertung der Korrektheit der durchgeführten Handlungen. Die Maße für Richtigkeit, Strukturiertheit und Zielorientiertheit werden anschließend aus den gewonnenen Daten berechnet.

#### 3.1 Identifikationen von Handlungen

Die Identifikation der Handlungen in den Aufzeichnungen erfolgt in mehreren Schritten. Zunächst werden nacheinander die Aufzeichnungen oder Teile von Aufzeichnungen, die sich dieser Handlung zuordnen lassen, markiert. Für jede Handlungen werden am Ende des Abschnittes in Tabelle 2 entsprechende Kategorien definiert, Indikatoren aufgelistet und Beispiele genannt. Zusätzlich wird eine Kategorie „Sonstiges“ definiert, der alle nicht identifizierbaren Aufzeichnungen zugeordnet werden. Im Anschluss an die Markierung der Handlungen in den Aufzeichnungen werden aneinander grenzende Aufzeichnungen zu Blöcken zusammengefasst. Diese werden dann abschließend entsprechend der Leserichtung in eine Reihenfolge gebracht. Die Vorgehensweise wird im Folgenden anhand einer Beispielaufzeichnung (Abbildung 1) beschrieben.

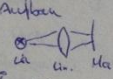
Experimentieraufgabe: <i>Gühwendel</i>				Lösungsseite: <i>1</i>
Fragestellung	Theorie	Rechnen	Experimentieren	Aufzeichnungen
<i>x</i>	<i>x</i> <i>x</i>	<i>x</i>		<i>Koffer ausspacken</i>  <i>Materialien zusammensuchen</i> <i>- 4 - (13), (16.)</i> <i>Aufbau</i>  <i>Aufbau</i>  <i>Bild rot 1 - Überlegung: Abbildungsgleichung <math>\frac{B}{b} = \frac{G}{g}</math> siehe Theorie</i> <i>messen <math>g = 11\text{cm}</math></i> <i><math>b = 3\text{cm}</math></i> <i><math>B = 0,3\text{cm}</math></i>  $\frac{B}{b} = \frac{G}{g} \rightarrow G = \frac{B \cdot g}{b} = \frac{0,3\text{cm} \cdot 11\text{cm}}{3\text{cm}} \approx 0,33\text{cm}$  <i>Voränderung der Abstände <math>\rightarrow</math> wech. messung Aufbau wie rot 1</i> <i><math>g = 13,5\text{cm}</math> <math>b = 7,5\text{cm}</math> <math>B = 0,2\text{cm}</math></i> $G = \frac{B \cdot g}{b} = \frac{0,2\text{cm} \cdot 13,5\text{cm}}{7,5\text{cm}} \approx 0,36\text{cm}$

Abbildung 35: Beispiel der Aufzeichnungen in einem Testheft

Die ersten Aufzeichnungen lassen sich der Kategorie „Versuch fehlerfrei aufbauen (4)“ zuordnen und werden entsprechend markiert. Das Ergebnis zeigt Abbildung 2.

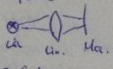
Experimentieraufgabe: <u>Gluhwendel</u>				Lösungsseite: <u>1</u>
Fragestellung	Theorie	Rechnen	Experimentieren	Aufzeichnungen
				Koffer auspacken
x				
	x			Materialien zusammensuchen -x- (13), (16)
				Aufbau
	x			Aufbau 
				Bild rot 1. Überlegung: Abbildungsgleichung $\frac{B}{b} = \frac{G}{g}$ siehe Theorie
			x	messen $g = 11 \text{ cm}$
			x	$b = 9 \text{ cm}$
			x	$B = 0,3 \text{ cm}$
			x	
		x		$\frac{B}{b} = \frac{G}{g} \rightarrow G = \frac{B \cdot g}{b} = \frac{0,3 \text{ cm} \cdot 11 \text{ cm}}{9 \text{ cm}} \approx 0,37 \text{ cm}$
		x		Voränderung der Abstände $\rightarrow$ Wdh. messung. Aufbau wie rot 1 $g = 13,5 \text{ cm}$ $b = 7,5 \text{ cm}$ $B = 0,2 \text{ cm}$ $G = \frac{B \cdot g}{b} = \frac{0,2 \text{ cm} \cdot 13,5 \text{ cm}}{7,5 \text{ cm}} \approx 0,36 \text{ cm}$

Abbildung 36: Markierung der ersten Aufzeichnungen zu "Versuch fehlerfrei aufbauen (4)"

Es folgen Markierungen der weiteren Aufzeichnungen zu den weiteren Kategorien: „Klärung der Fragestellung (1)“ (grün), „Klärung der theoretischen Grundlagen (2)“ (orange), „Experiment planen (3)“ (blau), „Messen und Beobachten (5)“ (braun) und „Daten analysieren (6)“ (schwarz) (vgl. Abbildung 3).

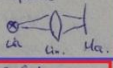
Experimentieraufgabe: <u>Gluhwendel</u>				Lösungsseite: <u>1</u>
Fragestellung	Theorie	Rechnen	Experimentieren	Aufzeichnungen
				Koffer auspacken
x				
	x			Materialien zusammensuchen -x- (13), (16)
				Aufbau
	x			Aufbau 
				Bild rot 1. Überlegung: Abbildungsgleichung $\frac{B}{b} = \frac{G}{g}$ siehe Theorie
			x	messen $g = 11 \text{ cm}$
			x	$b = 9 \text{ cm}$
			x	$B = 0,3 \text{ cm}$
			x	
		x		$\frac{B}{b} = \frac{G}{g} \rightarrow G = \frac{B \cdot g}{b} = \frac{0,3 \text{ cm} \cdot 11 \text{ cm}}{9 \text{ cm}} \approx 0,37 \text{ cm}$
		x		Voränderung der Abstände $\rightarrow$ Wdh. messung. Aufbau wie rot 1 $g = 13,5 \text{ cm}$ $b = 7,5 \text{ cm}$ $B = 0,2 \text{ cm}$ $G = \frac{B \cdot g}{b} = \frac{0,2 \text{ cm} \cdot 13,5 \text{ cm}}{7,5 \text{ cm}} \approx 0,36 \text{ cm}$

Abbildung 3: Markierung der Aufzeichnungen zu allen Handlungen

Zuletzt werden alle bisher nicht markierten Aufzeichnungen als „Sonstiges (9)“ markiert. Da es im vorliegenden Beispiel keine Aufzeichnungen dieser Art gibt, werden hier keine Markierungen gemacht.

Nun sind alle Aufzeichnungen mindestens einmal markiert. Allerdings können viele kleine oder wenige große Markierungen gemacht worden sein. (vgl. Abbildung 4).

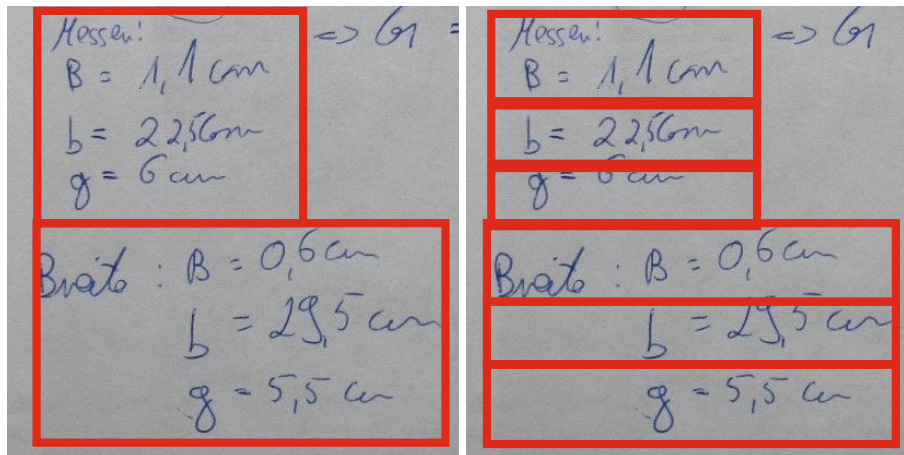


Abbildung 4: Verschiedene, mögliche Kodierungen zu "5. Messen und Beobachten"

Um zu einer eindeutigen Kodierung zu kommen werden alle Aufzeichnungen zu derselben Handlung die räumlich aneinander Grenzen zu einem Block zusammengefasst. Für das Beispiel aus Abbildung 1-3 ergibt sich damit die folgende, eindeutige Kodierung (siehe Abbildung 5).

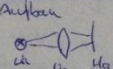
Experimentieraufgabe: Glühwendel				Lösungsseite: 1
Fragestellung	Theorie	Rechnen	Experimentieren	Aufzeichnungen
				Koffer auspacken
x				
	x			Wahlrad zusammensuchen - 4 - (13), (16)
				Aufbau
	x			
				Aufbau  u. Ma.
				Bild rot 1
			x	Überlegung: Abbildungsgleichung $\frac{3}{b} = \frac{g}{f}$ siehe Theorie (f zunächst die Breite)
			x	messen $g = 11 \text{ cm}$
			x	$b = 9 \text{ cm}$
			x	$f = 0,3 \text{ cm}$
			x	
x				$\frac{3}{b} = \frac{g}{f} \rightarrow G = \frac{3 \cdot g}{b} = \frac{0,3 \text{ cm} \cdot 11 \text{ cm}}{9 \text{ cm}} \approx 0,37 \text{ cm}$
x				Veränderung der Abstände $\rightarrow$ Wdh. messung. Aufbau wie rot 1 $g = 13,5 \text{ cm}$ $b = 7,5 \text{ cm}$ $f = 0,3 \text{ cm}$
				$G = \frac{3 \cdot g}{b} = \frac{0,3 \text{ cm} \cdot 13,5 \text{ cm}}{7,5 \text{ cm}} \approx 0,36 \text{ cm}$

Abbildung 5: Eindeutige Kodierung

Sollten Teile der Aufzeichnungen auch anderen Kategorien zugeordnet werden, so wird wie folgt vorgegangen. Im Beispiel wurde der blaue Kasten der „Klärung der Theorie“. Die Messgrößen  $B$ ,  $b$ ,  $g$  sind durch einen Kreis und die gesuchte Größe  $G$  durch ein Rechteck markiert (vgl. Abbildung 3, orange Kasten). Diese Aufzeichnungen lassen sich der Kategorie „Experiment planen (3)“ zuordnen. Für die Kodierung wird zunächst der als erstes beginnende Kasten (also der blaue) und anschließend der zweite Kasten kodiert werden.

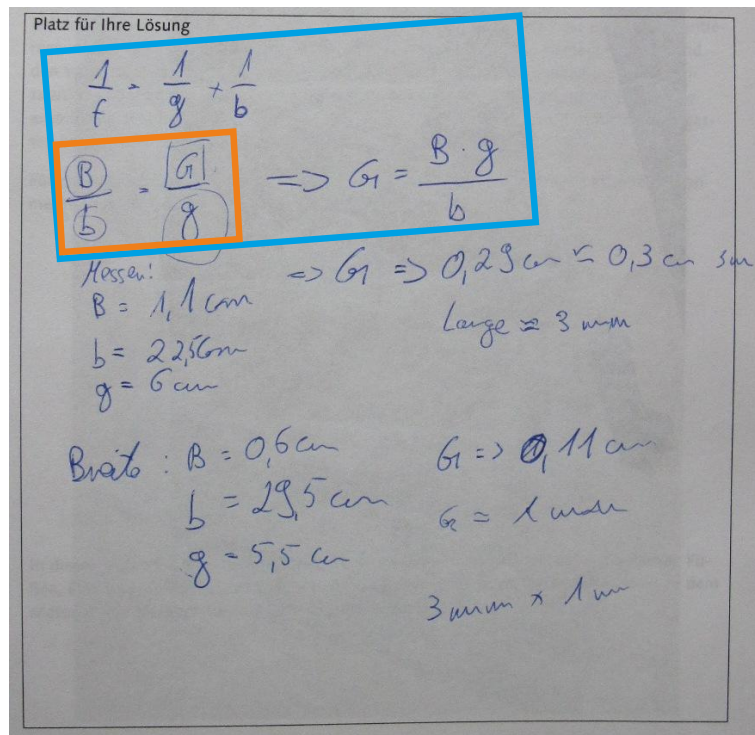


Abbildung 6: Markierung der Aufzeichnungen zu "5. Messen und Beobachten"

### Reihenfolge und Notation der Kodierung

Nachdem die Kodierung der Aufzeichnungen durch Bildung von Blöcken eindeutig ist, wird die Kodierung in eine Tabelle eingetragen. Eine solche Tabelle sieht wie folgt aus:

Aufgabe 1 - Glühwendel							
Handlungsnummer							
Richtigkeit							

Nun wird in Leserichtung jeweils die Nummer der Handlung des ersten Blocks eingetragen. In dem Beispiel ist dies der blaue Kasten zur Handlung „Versuch fehlerfrei aufbauen (4)“. Kam es, so wie im Beispiel, zu einer Doppelkodierung so wird, die Handlung zuerst in die Tabelle eingetragen, dessen Block zuerst beginnt (hier der blaue Kasten). Für den Fall das beide Blöcke an derselben Stelle beginnen wird die Handlung mit der niedrigeren Nummer zuerst eingetragen.

Aufgabe 1 - Glühwendel							
Handlungsnummer	4						
Richtigkeit							

Anschließend folgen die nächsten Blöcke: (Grün) „Fragestellung klären (1)“, (Orange) „Theoretische Grundlagen klären (2)“, usw. Die fertige Kodiertabelle sieht wie folgt aus:



Aufgabe 1 - Glühwendel												
Handlungsnummer	4	1	2	4	2	3	4	2	5	6	5	6
Richtigkeit												

Sollten in der Tabelle zwei aufeinanderfolgende Blöcke dieselbe Handlungsnummer zugewiesen bekommen, so werden diese beiden Blöcke zu einem Block zusammen gezogen. Dies kommt etwa dann vor, wenn Aufzeichnungen durch den Seitenumbruch getrennt werden.

Um die Aufzeichnungen bzgl. der Handlungen zu identifizieren, gibt es aufgabenspezifischen expliziten Erwartungen was in den jeweiligen Aufzeichnungen stehen soll. Die Erwartungen finden sich im Anhang (siehe A.1).

## Hinweise

Es ist möglich, dass Sonderfälle auftreten, die zu Problemen bei der Identifikation von Handlungen führen. Welche Sonderfälle das sind und wie in diesem Fall bei der Kodierung vorzugehen ist, wird im Folgenden beschrieben.

### Durchgestrichene Aufzeichnungen

Sollten Aufzeichnungen durchgestrichen sein, so werden diese für die Kodierung ignoriert und wie leeres Papier behandelt.

### Doppelte Zuordnungen

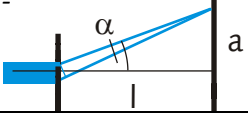
Es ist durchaus möglich, dass einer Aufzeichnung mehrere Kategorien zugeordnet werden können. Dies ist zum Beispiel der Fall, wenn mehrere Produkte zu verschiedenen Handlungen beim Experimentieren gleichzeitig notiert werden (siehe Tabelle 1).

**Tabelle 33: Verkürzte Notation und Doppelkodierung**

Ideale Aufzeichnungen	verkürzte Aufzeichnungen
$B/b = G/g; \Rightarrow G=B \cdot g/G$  B=1,1 cm b=22,5cm g=6cm  $G=1,1\text{cm} \cdot 22,5\text{cm}/6\text{cm}=0,3\text{cm}$	$1,1\text{cm} \cdot 22,5\text{cm}/6\text{cm}=0,3\text{cm}$

Während die idealen Aufzeichnungen wie in Abbildung 1-3 beschrieben leicht zu kodieren ist, hat man bei der verkürzten Aufzeichnung eine Mehrfachkodierung. In der verkürzten Aufzeichnung findet sich eine richtige Formel die vorher nicht eingeführt wurde [Wissen aktivieren (2)], es finden sich Messwerte die vorher nicht aufgeführt werden [Messen und Beobachten (5)] und es findet sich ein Ergebnis einer Rechnung [Daten analysieren (6)]. Entsprechend werden auch in den verkürzten Aufzeichnungen alle drei Handlungen kodiert.

**Tabelle 34: Kategorien zur Kodierung der Aufzeichnungen im Laborheft**

Handlung	Beschreibung	Indikatoren	Beispiel	Abgrenzung
Fragestellung klären (1)	Die Fragestellung wird gelesen und verstanden.	Die Fragestellung wird wörtlich oder in eigenen Worten aufgeschrieben.		
Wissen aktivieren und theoretische Grundlage schaffen (2)	Das für die Bearbeitung der Fragestellung notwendige Wissen wird aktiviert.	Es finden sich Erklärungen zur Theorie der Fragestellung, dazu gehören auch Formeln, Umformungen ODER das Notieren von bekannten Größen.	<p>Formel zur Berechnung der Gitterkonstante: <math>n \cdot \lambda = g \cdot \sin(\alpha_n)</math>; <math>\sin(\alpha_n) = a_i / l</math>.</p> <p>Variablen: <math>n_1=1</math>; <math>\lambda_{\text{grün}}=532 \text{ nm}</math></p>	
Experiment planen (3)	Die Planung eines Experiments zur Beantwortung der Fragestellung.	Die Planung des Experiments wird beschrieben, dazu gehören zu messende Werte bzw. zu machende Beobachtungen und die Erklärung des Versuchsaufbaus inklusive Skizze.	<p>Skizze eines Versuchsaufbaus (siehe Abbildung) und/oder Beschreibung des Versuchs.</p> 	Abgrenzung zu „Versuch aufbauen (4)“: Die Skizze enthält keine konkreten Werte und Abmessungen.
Versuch aufbauen (4)	Ein Versuch wird aufgebaut.	Es findet sich eine Zeichnung des Versuchsaufbaus die sich 1:1 nachbauen lässt.		Abgrenzung zu „Versuch aufbauen (4)“: Die Zeichnung enthält konkrete Werte und Abmessungen wodurch der Aufbau 1:1 nachgebaut werden kann.
Messen und Beobachten (5)	Gezielt Messwerte aufnehmen oder Beobachtungen anstellen.	Handlungen zur Fertigkeit „Messen und Beobachten“ zeigen sich durch Aufzeichnungen von Messwerten, Messtabellen und Beobachtungen bei der Versuchsdurchführung.	$n=1$ ; $\lambda=532 \text{ nm}$ ; $a_i=2,2 \text{ cm}$ ; $l=30 \text{ cm}$ .	
Daten analysieren (6)	Auf Basis von Messwerten werden Berechnungen durchgeführt oder Graphen gezeichnet.	Handlungen zur Fähigkeit „Daten analysieren“ sind erkennbar durch Rechnungen die auf den Messwerten basieren, sowie dem Anfertigen und Auswerten von Graphen. Dazu kann es auch gehören Fehlerrechnung anfertigen oder Fehlerbalken zu	$g=n\lambda/a_i=7,25 \cdot 10^{-6} \text{ m}$	

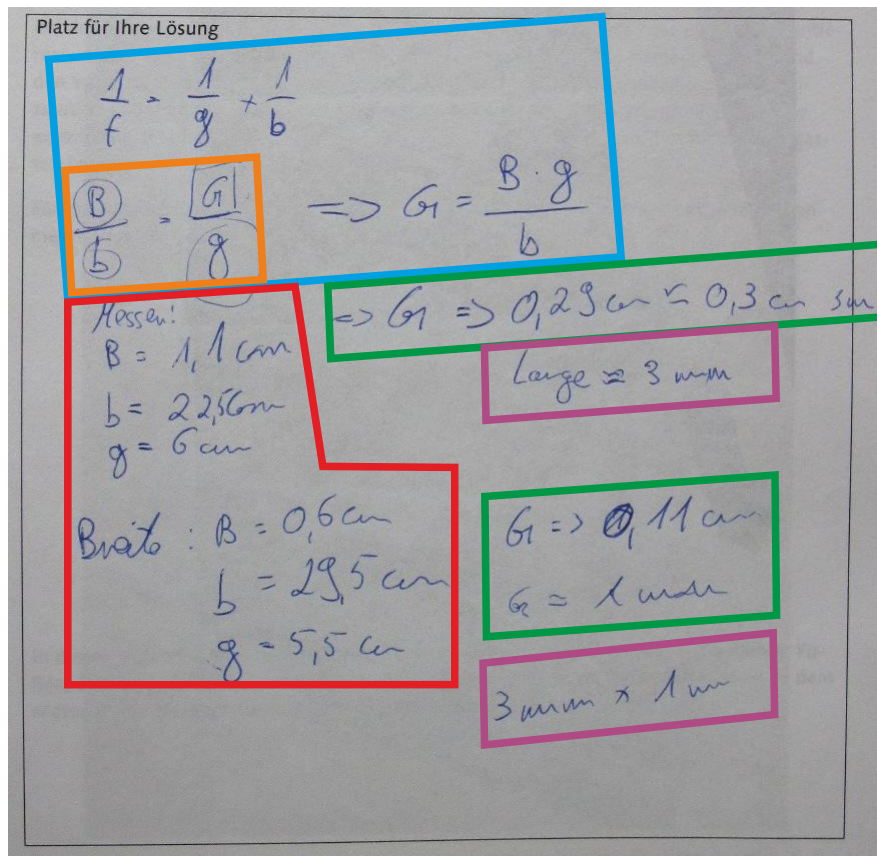
		zeichnen.		
Sachgerechte Schlüsse ziehen (7)	Berechnungen, Messgrößen und Beobachtungen werden auf die Fragestellung bezogen um zu prüfen, ob die Fragestellung beantwortet wurde.	Handlungen zur Fähigkeit „Sachgerechte Schlüsse ziehen“ zeigen sich in den Aufzeichnungen durch Antwortsätze und Schlussfolgerungen zu den Messwerten. Messfehler und Ungenauigkeiten können zur Erklärung genutzt werden.	<i>Die Gitterkonstante liegt im erwarteten Bereich. Da keine Fehler bei der Messung aufgetreten sind, wird angenommen, dass die Gitterkonstante <math>g \approx 7,25 \cdot 10^{-6} \text{ m}</math> ist.</i>	
Sonstiges (9)	Sonstige Aufzeichnungen.	Produkte die keiner Handlung zugeteilt werden können.		

## 3.2 Richtigkeit

Nachdem die Handlungen identifiziert wurden, erfolgt nun deren Kodierung bezüglich der Richtigkeit. Jedem Block an Aufzeichnungen kann dabei einer von vier Codes bzgl. der korrekten Durchführung zugeordnet werden:

- (1) Die Aufzeichnungen entsprechen der Musterlösung oder sind äquivalent dazu (richtig).
- (0,5) Die Aufzeichnungen enthalten Teile der Musterlösung bzw. äquivalente Aufzeichnungen UND es fehlen Teile der Musterlösung oder es sind Fehler in den Aufzeichnungen (teilweise richtig)
- (0) Die Aufzeichnungen entsprechend nicht der Musterlösung oder sind durchgehend falsch (falsch)
- (99) Die Aufzeichnungen sind nicht interpretierbar (nicht interpretierbar)
- (77) Die Aufzeichnung kommt mindestens zum zweiten Mal in identischer Weise vor (Wiederholung)
- (66) Die Aufzeichnung ist noch nicht vollständig (z.B. wird nur die Hälfte eines Versuches aufgebaut oder eine Messreihe enthält nur die ersten Werte) und wird in einem späteren Intervall zu dieser Handlung vervollständigt (Zwischenschritt, keine Wertung)

Um diese Codes den jeweiligen Aufzeichnungen zuordnen zu können, finden sich im Anhang aufgabenspezifische Tabellen (siehe A.2). In diesen wird beschrieben, was das Kriterium in den Aufzeichnungen ist anhand dessen die Richtigkeit bewertet wird. Zudem findet sich die Musterlösung, die zur Bewertung als Referenz genutzt wird. Im Folgenden wird die Bewertung der Richtigkeit an dem Beispiel aus 3.1 dokumentiert.



Aufgabe 1 – Glühwendel							
Handlungsnummer	2	3	5	6	7	6	7
Richtigkeit							

Der erste Block der Tabelle ist „Wissen aktivieren und Theorie schaffen (2)“. Aus der Tabelle im Anhang findet sich dafür bei der Aufgabe 1 – Glühwendel:

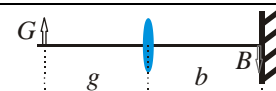
Handlung	Kriterium für Richtigkeit	Musterlösung
Wissen aktivieren (2)	Beschreibungen und Erklärungen zur Theorie und Formeln	Formel: $G=B \cdot g/b$

Die Formel der Musterlösung findet sich genauso in den Aufzeichnungen. Entsprechend wird hier „1“ (richtig) kodiert. Für die Kodiertabelle ergibt sich damit:

Aufgabe 1 – Glühwendel							
Handlungsnummer	2	3	5	6	7	6	7
Richtigkeit	1						

Die nächste Handlung ist „Experiment planen (3)“, die aufgabenspezifische Tabelle dafür sieht wie folgt aus:

Handlung	Kriterium für Richtigkeit	Musterlösung
Experiment planen (3)	Skizzen und Erklärungen zum Versuchsaufbau	Zu Messen: $B, g, b$ Skizze:



In den Aufzeichnungen finden sich die zu messenden Größen. Die entsprechende Kodierung wäre hier „0,5“ (teilweise richtig), weil die Messgrößen mit der Musterlösung übereinstimmen, der zweite

Teil der Musterlösung aber fehlt und diese deshalb unvollständig ist. Die Kodiertabelle wird ergänzt zu:

Aufgabe 1 – Glühwendel							
Handlungsnummer	2	3	5	6	7	6	7
Richtigkeit	1	0,5					

Entsprechend wird für die weiteren Handlungen verfahren. Die Handlung „Messen und Beobachten (5)“ wird mit „1“ (richtig) kodiert, weil sich aus den Messwerten  $B=1,1\text{cm}$ ,  $b=22,5\text{cm}$  und  $g=6\text{cm}$  bzw.  $B'=0,6\text{cm}$ ,  $b'=29,5\text{cm}$  und  $g'=5,5\text{cm}$  ein Wert  $G=0,29\text{cm}$  bzw.  $G'=0,11\text{cm}$  berechnen lässt und diese Werte innerhalb der gegebenen Toleranz von  $2,5\text{mm} - 3,5\text{mm}$  bzw.  $0,5\text{mm} - 1,5\text{mm}$  liegen. Die Handlung „Daten analysieren (6)“ wird mit „1“ (richtig) kodiert, da die Rechnung des Probanden ein Ergebnis von  $G=0,29\text{cm}$  ergibt, dass sich durch nachrechnen aus den Werten reproduzieren lässt. Die Handlung „Sachgerechte Schlüsse ziehen (7)“ wird ebenfalls mit „1“ (richtig) kodiert, da der Wert als von  $0,3\text{cm}$  als Länge der Glühwendel beschrieben wird und noch keine Daten zur Erklärung der Breite vorliegen. Der zweite Block zu der Kategorie „Daten analysieren (6)“ wird mit „1“ (richtig) kodiert, da die Rechnung des Probanden ein Ergebnis von  $G'=0,11\text{cm}$  ergibt, dass sich durch nachrechnen aus den Werten reproduzieren lässt. Der zweite Block zu der Kategorie „Sachgerechte Schlüsse ziehen (7)“ wird ebenfalls mit „1“ (richtig) kodiert, da der Wert als von  $0,1\text{cm}$  als Breite der Glühwendel beschrieben wird und die Daten zur Erklärung der Länge bereits ausgewertet sind. Die fertige Kodiertabelle sieht wie folgt aus:

Aufgabe 1 - Glühwendel							
Handlungsnummer	2	3	5	6	7	6	7
Richtigkeit	1	0,5	1	1	1	1	1

### Wiederholungen

Kommt eine Aufzeichnung in identischer Form an mehreren Stellen vor (z.B. eine Formel, eine Variable  $n_1=1, \dots$ ) dann wird nur für die erste Stelle eine Richtigkeit kodiert. Die Wiederholungen werden mit „99“ (nicht interpretierbar) kodiert.

## Anhang

### A.1 Aufgabenspezifische Kodieranweisungen zur Identifikation der Handlungen

#### Vermessung einer Glühwendel

Handlung	Produkt
Fragestellung klären (1)	Notierte Fragestellung
Wissen aktivieren (2)	Beschreibungen und Erklärungen zur Theorie und Formeln
Experiment planen (3)	Beschreibung und Skizze
Versuch aufbauen (4)	Zeichnung zum aufgebauten Versuch; Foto eines Versuchsaufbaus
Messen / Beobachten (5)	Messwerte für $g$ , $b$ , $B_i$
Daten analysieren (6)	Berechnen von $G_i$
Sachgerechte Schlüsse ziehen (7)	Lösungssatz
Sonstiges (9)	Sonstige Aufzeichnung

#### Optische Dichte

Handlung	Produkt
Fragestellung klären (1)	Notierte Fragestellung
Wissen aktivieren (2)	Beschreibungen und Erklärungen zur Theorie und Formeln
Experiment planen (3)	Beschreibung und Skizze
Versuch aufbauen (4)	Zeichnung zum aufgebauten Versuch; Foto eines Versuchsaufbaus
Messen / Beobachten (5)	Einfalls- $\alpha_1$ und Ausfallswinkel $\alpha_2$ für Laser verschiedener Wellenlänge
Daten analysieren (6)	Berechnung von $n_2$
Sachgerechte Schlüsse ziehen (7)	Lösungssatz
Sonstiges (9)	Sonstige Aufzeichnung

#### Spektroskopie

Handlung	Produkt
Fragestellung klären (1)	Notierte Fragestellung
Wissen aktivieren (2)	Beschreibungen und Erklärungen zur Theorie und Formeln
Experiment planen (3)	Beschreibung und Skizze
Versuch aufbauen (4)	Zeichnung zum aufgebauten Versuch; Foto eines Versuchsaufbaus
Messen / Beobachten (5)	Messwerte $a_i$ und $l$
Daten analysieren (6)	Berechnung von $g$ oder $\ddot{e}$
Sachgerechte Schlüsse ziehen (7)	Lösungssatz
Sonstiges (9)	Sonstige Aufzeichnung

### Zuckerwasser

Handlung	Produkt
Fragestellung klären (1)	Notierte Fragestellung
Wissen aktivieren (2)	Beschreibungen und Erklärungen zur Theorie und Formeln
Experiment planen (3)	Beschreibung und Skizze
Versuch aufbauen (4)	Zeichnung zum aufgebauten Versuch; Foto eines Versuchsaufbaus
Messen / Beobachten (5)	Drehung der Polarisations Ebene $\alpha$
Daten analysieren (6)	Berechnung der Zuckerkonzentration und der Zuckerkonzentration pro Liter
Sachgerechte Schlüsse ziehen (7)	Lösungssatz
Sonstiges (9)	Sonstige Aufzeichnung

### Polarisationseigenschaften

Handlung	Produkt
Fragestellung klären (1)	Notierte Fragestellung
Wissen aktivieren (2)	Beschreibungen und Erklärungen zur Theorie und Formeln
Experiment planen (3)	Beschreibung und Skizze
Versuch aufbauen (4)	Zeichnung zum aufgebauten Versuch; Foto eines Versuchsaufbaus
Messen / Beobachten (5)	Messen der Intensität des transmittierten- und reflektierten Strahls in Abhängigkeit zum Einfallswinkel UND Messen der Intensität von transmittierten- und reflektierten Strahl in Abhängigkeit zum Polarisationsfilter bei festem „Winkel der stärksten Polarisation“
Daten analysieren (6)	Ablesen des „am stärksten polarisierenden Winkels“
Sachgerechte Schlüsse ziehen (7)	Lösungssatz
Sonstiges (9)	Sonstige Aufzeichnung

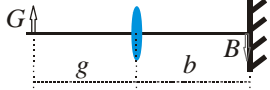
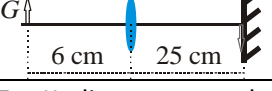
### Blackbox

Handlung	Produkt
Fragestellung klären (1)	Notierte Fragestellung
Wissen aktivieren (2)	Beschreibungen und Erklärungen zur Theorie und Formeln
Experiment planen (3)	Beschreibung und Skizze
Versuch aufbauen (4)	Zeichnung zum aufgebauten Versuch; Foto eines Versuchsaufbaus
Messen / Beobachten (5)	Beobachten und Messe
Daten analysieren (6)	Berechnung der Gitterposition
Sachgerechte Schlüsse ziehen (7)	Lösungssatz
Sonstiges (9)	Sonstige Aufzeichnung


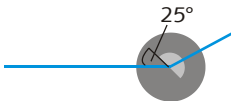


## A.2 Aufgabenspezifische Kodieranweisungen zur Richtigkeit

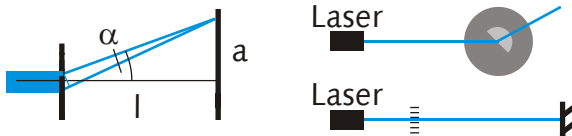
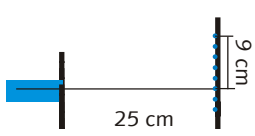
### Vermessung einer Glühwendel

Handlung	Kriterium für Richtigkeit	Musterlösung
Fragestellung klären (1)	Notierte Fragestellung	Wie ist die Abmessung der Glühwendel von der Lampe (Inventarnr. 3)?
Wissen aktivieren (2)	Beschreibungen und Erklärungen zur Theorie und Formeln	Formel: $G=B \cdot g/b$
Experiment planen (3)	Skizzen, Erklärungen zum Versuchsaufbau ODER Messgrößen ODER Überschriften zu Teilabschnitten	Zu Messen: $B, g, b$  Skizze: Überschriften zu Abschnitten sind richtig, wenn sie ein Teil der Musterlösung sind.
Versuch aufbauen (4)	Zeichnung zum aufgebauten Versuch	
Messen und Beobachten (5)	Messwerte für $g, b, B_i$	Zur Kodierung muss der Kodierer $G_i$ aus den Messwerten $g, b, B_i$ anhand der Formel $G_i=B_i \cdot g/b$ berechnen. Die Phase gilt als richtig wenn $G_i$ zwischen 2,5mm und 3,5mm oder zwischen 0,1mm und 1mm liegt.
Daten analysieren (6)	Berechnen von $G_i$	Richtig, wenn das vom Probanden aus den Messwerten $g, b, B_i$ errechnete $G_i$ mit dem vom Kodierer errechneten $G_i$ übereinstimmt.
Sachgerechte Schlüsse ziehen (7)	Lösungssatz oder Erklärung der Messwerte.	Abhängig von bisher errechneten Daten muss entweder die Länge $a$ im Bereich von 2,5mm und 3,5mm liegen oder die Breite $b$ im Bereich von 0,1mm und 1mm liegen. ODER Erklärung, dass die Ergebnisse nicht stimmen können.
Sonstiges (9)	Sonstige Aufzeichnung	Richtigkeit wird nicht kodiert „-“

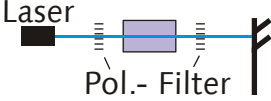
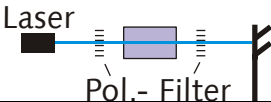
## Optische Dichte

Handlung	Kriterium für Richtigkeit	Musterlösung
Fragestellung klären (1)	Notierte Fragestellung	Aus welchem Material ist das Prisma (Inventarnr. 14)?
Wissen aktivieren (2)	Beschreibungen und Erklärungen zur Theorie und Formeln	Formel: $n_1 \cdot \sin(\alpha_1) = n_2 \cdot \sin(\alpha_2)$ Variable: $n_1=1$
Experiment planen (3)	Skizzen, Erklärungen zum Versuchsaufbau ODER Messgrößen ODER Überschriften zu Teilabschnitten	Zu Messen: $\alpha_1, \alpha_2$ Skizze:  Überschriften bzgl. der einzelnen Laser
Versuch aufbauen (4)	Zeichnung zum aufgebauten Versuch	Material, Aufbau und Versuchszeichnung äquivalent zur Skizze. 
Messen und Beobachten (5)	Einfalls- $\alpha_1$ und Ausfallswinkel $\alpha_2$ für Laser verschiedener Wellenlänge	Zur Kodierung muss der Kodierer $n_2$ aus den Messwerten $\alpha_1, \alpha_2$ für $n_1 = 1$ berechnen. Die Phase gilt als richtig wenn $n_2$ zwischen 1,45 und 1,60 liegt.
Daten analysieren (6)	Berechnung von $n_2$	Richtig, wenn das vom Probanden aus den Messwerten $\alpha_1$ und $\alpha_2$ errechneten $n_2$ mit dem vom Kodierer errechneten $n_2$ übereinstimmt.
Sachgerechte Schlüsse ziehen (7)	Lösungssatz	Es wurde Graph „D“ gewählt ODER Erklärung das ein Messfehler vorliegen muss
Sonstiges (9)	Sonstige Aufzeichnung	Richtigkeit wird nicht kodiert „-“

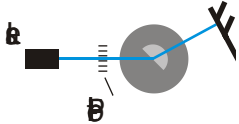
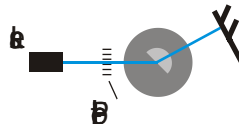
## Spektroskopie

Handlung	Kriterium für Richtigkeit	Musterlösung
Fragestellung klären (1)	Notierte Fragestellung	Wie ist die Wellenlänge des unbekannten Lasers (Inventarnr. 5)?
Wissen aktivieren (2)	Beschreibungen und Erklärungen zur Theorie und Formeln	Formel: $n \cdot \ddot{e} = g \cdot \sin(\alpha_n)$ ; $\sin(\alpha_n) = a_i / l$ ; $\tan(\alpha_n) = a_i / l'$ Variable : $n_1=1$ ; $\ddot{e}_{\text{blau}}=430\text{nm}$ ; $\ddot{e}_{\text{grün}}=532\text{nm}$
Experiment planen (3)	Skizzen, Erklärungen zum Versuchsaufbau ODER Messgrößen ODER Überschriften zu Teilabschnitten	Zu Messen: $a_i$ , $l$ bzw. $a_i$ , $l'$ bzw. $\alpha_i$ Skizze : 
Versuch aufbauen (4)	Zeichnung zum aufgebauten Versuch	Material, Aufbau und Versuchszeichnung äquivalent zur Skizze. 
Messen und Beobachten (5)	Messwerte $a_i$ und $l$	Zur Kodierung muss der Kodierer $g$ aus den Messwerten $a_i$ und $l$ für $\ddot{e}_{\text{rot}} = 650 \text{ nm}$ ; $\ddot{e}_{\text{grün}} = 532 \text{ nm}$ ; $\ddot{e}_{\text{blau}} = 405 \text{ nm}$ berechnen: $g = n \cdot \ddot{e} / a$ Die Phase gilt als richtig wenn $g$ zwischen $7,00 \cdot 10^{-6} \text{ m}$ und $7,25 \cdot 10^{-6} \text{ m}$ liegt.
Daten analysieren (6)	Berechnung von $g$ oder $\ddot{e}$	Zur Kodierung muss der Kodiererg bzw. $\ddot{e}$ aus den Messwerten nachrechnen. Die Phase gilt als richtig wenn ein entsprechendes Ergebnis von Kodierer und Proband erzielt wird.
Sachgerechte Schlüsse ziehen (7)	Lösungssatz	Richtig wenn $\lambda_{\text{rot}}$ zwischen 600 nm und 700 nm ODER Erklärung, dass ein Messfehler vorliegen muss
Sonstiges (9)	Sonstige Aufzeichnung	Richtigkeit wird nicht kodiert „-“

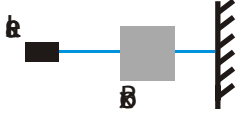
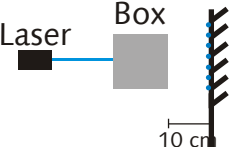
## Zuckerwasser

Handlung	Kriterium für Richtigkeit	Musterlösung
Fragestellung klären (1)	Notierte Fragestellung	Bei welcher Zuckerkonzentration wird die Polarisations Ebene um $0,75^\circ/\text{cm}$ gedreht?
Wissen aktivieren (2)	Beschreibungen und Erklärungen zur Theorie und Formeln	Formel: $V=h \cdot l \cdot b$ ; $\rho=m/V$
Experiment planen (3)	Skizzen, Erklärungen zum Versuchsaufbau ODER Messgrößen ODER Überschriften zu Teilabschnitten	Skizze:  <p>Durchführung:</p> <ul style="list-style-type: none"> <li>- Vermessung der Küvette</li> <li>- Erhöhung der Zuckerkonzentration um eine Einheit</li> <li>- Ablesen des Winkel bei dem die Lichtintensität minimal/maximal wird</li> <li>- Bestimmung der Konzentration für den Drehwinkel von <math>0,75^\circ/\text{cm}</math></li> </ul>
Versuch aufbauen (4)	Zeichnung zum aufgebauten Versuch	Material, Aufbau und Versuchszeichnung äquivalent zur Skizze. 
Messen und Beobachten (5)	Drehung der Polarisations Ebene $\alpha$	Richtig wenn die minimale und die maximale Auslöschung des Lichts bei einem Winkel von $85^\circ - 95^\circ$ zwischen den Polarisationsfiltern auftritt. ODER Richtige Messung, wenn sich Anzahl an Löffeln Zucker für die Änderung die Polarisations Ebene um $0,75^\circ/\text{cm}$ zwischen $9 \cdot 1000\text{cm}^3/\text{Volumen}(\text{in cm}^3)$ und $15 \cdot 1000\text{cm}^3/\text{Volumen}(\text{in cm}^3)$ liegt. ODER Volumen= $10\text{cm} \cdot 2,8\text{cm} \cdot b$ , mit $b$ zwischen 1cm und 6,5 cm.
Daten analysieren (6)	Berechnung der Zuckerkonzentration und der Zuckerkonzentration pro Liter	Richtig, wenn die Zuckerkonzentration bei einer Drehung der Polarisations Ebene um $0,75^\circ/\text{cm}$ aus den im Wasser gelösten Zucker (Einheit in Löffeln) und dem Drehwinkel berechnet wird. UND Die Konzentration wird wenn nötig auf Liter umgerechnet werden
Sachgerechte Schlüsse ziehen (7)	Lösungssatz	Richtiges Ergebnis, wenn 9-15 Löffel angegeben werden ODER Erklärung eines Messfehlers
Sonstiges (9)	Sonstige Aufzeichnung	Richtigkeit wird nicht kodiert „-“

## Polarisationseigenschaften

Handlung	Kriterium für Richtigkeit	Musterlösung
Fragestellung klären (1)	Notierte Fragestellung	Wie ist der Winkel unter dem das Licht linear polarisiert wird? Wie ist die Schwingungsebene des transmittierten- und des reflektierten Lichts?
Wissen aktivieren (2)	Beschreibungen und Erklärungen zur Theorie und Formeln	Proband identifiziert als Aufgabenrelevant: - Polarisation - Brechung
Experiment planen (3)	Skizzen, Erklärungen zum Versuchsaufbau ODER Messgrößen ODER Überschriften zu Teilabschnitten	Zu Messen: $\alpha_1, \alpha_2$ Skizze: 
Versuch aufbauen (4)	Zeichnung zum aufgebauten Versuch	Material, Aufbau und Versuchszeichnung äquivalent zur Skizze. (Polarisationsfilter hinter dem Prisma anstelle davor ist ein äquivalenter Aufbau) 
Messen und Beobachten (5)	Messen der Intensität des transmittierten- und reflektierten Strahls in Abhängigkeit zum Einfallswinkel UND Messen der Intensität von transmittierten- und reflektierten Strahl in Abhängigkeit zum Polarisationsfilter bei festem „Winkel der stärksten Polarisation“	Richtig, wenn das Maximum des transmittierten/reflektierten und das Minimum des reflektierten/transmittierten Strahls bei einem Einfallswinkel von $\alpha = 50^\circ$ - $62^\circ$ auftreten UND $\beta, \gamma$ stehen senkrecht aufeinander ( $90^\circ$ Differenz im Winkel)
Daten analysieren (6)	Ablesen des „am stärksten polarisierenden Winkels“	Richtig, wenn der „Winkel der stärksten Polarisation“ gewählt wird, wenn der transmittierte/reflektierte Strahl sein Maximum/Minimum erreicht
Sachgerechte Schlüsse ziehen (7)	Lösungssatz	Der „am stärksten polarisierende Winkel“ $\alpha = 50^\circ$ - $65^\circ$ UND $\beta, \gamma$ stehen senkrecht aufeinander ( $90^\circ$ Differenz im Winkel) ODER Erklärung eines Messfehlers
Sonstiges (9)	Sonstige Aufzeichnung	Richtigkeit wird nicht kodiert „-“

## Blackbox

Handlung	Kriterium für Richtigkeit	Musterlösung
Fragestellung klären (1)	Notierte Fragestellung	Wie sind die optischen Eigenschaften des unbekanntes Bauteils?
Wissen aktivieren (2)	Beschreibungen und Erklärungen zur Theorie und Formeln	Proband identifiziert als Aufgabenrelevant: - Intensität - Polarisation - Wechselwirkung von Licht mit Materie - Optische Komponenten und ihre Anwendungen
Experiment planen (3)	Skizzen, Erklärungen zum Versuchsaufbau ODER Messgrößen ODER Überschriften zu Teilabschnitten	Skizze: oder Skizze der Blackbox  Zu untersuchende optische Eigenschaften.
Versuch aufbauen (4)	Zeichnung zum aufgebauten Versuch	Material, Aufbau und Versuchszeichnung äquivalent zur Skizze. (Einführung verschiedener optischer Komponenten wie z.B. Polarisationsfilter in den Strahlengang sind äquivalent)  Richtig des Nachbaus: Wenn mindestens eine Komponente im Nachbau enthalten ist und mit der Blackbox verglichen werden kann.
Messen und Beobachten (5)	Beobachten und Messe	Richtig, wenn: Bemerkt wird, dass Licht bestimmter Wellenlängen gefiltert wird UND Ein Beugungsbild entsteht UND $a_i$ für die Abstände der $i$ -ten Beugungsordnung zur optischen Achse gemessen ODER Ausschluss von Eigenschaften die nicht Bestandteil der box sind ODER beim Nachbau: Wenn Bilder von Nachbau und Blackbox verglichen werden ODER beim Nachbau: Wenn Komponenten im nachbaugetauscht werden oder in einem anderen Slot platziert werden
Daten analysieren (6)	Berechnung der Gitterposition	Richtig wenn berechnet oder erklärt wird, dass sich das Gitter an Position 1 befinden muss

		<p>UND</p> <p>Das die Position des Farbfilters nicht eindeutig bestimmt werden kann</p>
<p>Sachgerechte Schlüsse ziehen (7)</p>	<p>Lösungssatz</p>	<p>Richtig wenn:</p> <p>Ein Farbfilter diagnostiziert wird</p> <p>UND</p> <p>Ein Gitter in einem der Randpositionen diagnostiziert wird</p> <p>UND</p> <p>Der letzte Slot leer ist oder nicht gesagt werden kann, was sich im letzten Slot befindet.</p> <p>ODER</p> <p>Optische Bauteile ausgeschlossen werden</p> <p>Teilweise richtig (0,5 Punkte) wenn zwei der Slots richtig sind.</p>
<p>Sonstiges (9)</p>	<p>Sonstige Aufzeichnung</p>	<p>Richtigkeit wird nicht kodiert „-“</p>

## A.4.9 R-Skript zur Auswertung

```
## Auswertung Experimentierkompetenz Studie III
##Verzeichnis einstellen
pfad<- "F:\\Promotion\\Studie 3\\R_Daten"
setwd (pfad)

## Alle Probanden einfügen
Probanden_Prod<-read.delim("Stichprobe_Studie_III.txt", header=FALSE)
names(Probanden_Prod) <- c("Probandennamen")
max<- length(Probanden_Prod$Probandennamen)
for (i in 1:max) {

  ##Einlesen eines Probanden
  for (j in 1:6) {

    ## Daten einlesen und benennen
    inh<- read.delim(paste(paste("Studielll_",Probanden_Prod$Probandennamen[i],sep=""), paste(j,
    ".txt", sep=""), sep="_"), header=FALSE)
    names(inh) <- c("H", "R")

    ## Datei richtig benennen und speichern
    nam<- paste(paste("",Probanden_Prod$Probandenname[i], sep=""), paste(j, "", sep=""), sep="_")
    assign(nam, inh)

    save (list = paste(paste("",Probanden_Prod$Probandenname[i], sep=""), "", sep="_1"),
    file=paste(paste("",Probanden_Prod$Probandenname[i], sep=""), paste(j, ".rda", sep=""), sep="_"))
  }
}

#### Testleistung bestimmen

##Einladen der Auswertungsstichprobe
Probanden_AllOver<- Probanden_Prod
names(Probanden_AllOver) <- c("Probandennamen")

##Leere Übersicht für jeden Probanden erzeugen
max<- length(Probanden_AllOver$Probandennamen); Auswertung <- c(0) ##TODO: Das mit der Aus-
wertung ist noch nicht sauber!
for (i in 1:max) {
  Name <- c(paste(Probanden_AllOver$Probandennamen[i], "", sep="")); Name <- factor(Name)

  aspekt<- c("R","LA","S", "H", "G")); j<- 1; k<- 1; max_as<- length(aspekt)
  for (j in 1:max_as) {
    for (k in 1:6) {
      nam<- paste("Aufgabe_", paste(k, paste(aspekt[j]), sep="_"), sep="")

```



```

inh<- c(0)
assign(nam, inh)
}
}
Richtigkeit <- c(0); Logische_Abfolge<- c(0); Strategie <- c(0); Testleistung <- c(0)
Prob_Dat<- data.frame(Name, Aufgabe_1_R, Aufgabe_1_LA, Aufgabe_1_S, Aufgabe_1_H, Aufga-
be_1_G, Aufgabe_2_R, Aufgabe_2_LA, Aufgabe_2_S, Aufgabe_2_H, Aufgabe_2_G, Aufgabe_3_R,
Aufgabe_3_LA, Aufgabe_3_S, Aufgabe_3_H, Aufgabe_3_G, Aufgabe_4_R, Aufgabe_4_LA, Aufga-
be_4_S, Aufgabe_4_H, Aufgabe_4_G, Aufgabe_5_R, Aufgabe_5_LA, Aufgabe_5_S, Aufgabe_5_H,
Aufgabe_5_G, Aufgabe_6_R, Aufgabe_6_LA, Aufgabe_6_S, Aufgabe_6_H, Aufgabe_6_G, Richtigkeit,
Logische_Abfolge, Strategie, Testleistung)
Prob_Nam<- paste(Probanden_AllOver$Probandennamen[i], "", sep="")
R1 <- c(0); R2 <- c(0); R3 <- c(0); R4 <- c(0); R5 <- c(0); R6 <- c(0); R7 <- c(0)
A1 <- c(0); A2 <- c(0); A3 <- c(0); A4 <- c(0); A5 <- c(0); A6 <- c(0); A7 <- c(0)

##Testleistung - Richtigkeit
j<-1
for (j in 1:6) {
max<- length(get(paste(Probanden_AllOver$Probandenname[i], j, sep="_")) [,1])
nam_auf<- paste("Aufgabe_", paste(j, "_H", sep=""), sep="") ##TODO: Brauche ich diese Zeile für
irgendwas??
k <- 0; Richtigkeit <- 0; anzahl<- 0; R1 <- 0; R2 <- 0; R3 <- 0; R4 <- 0; R5 <- 0; R6 <- 0; R7 <- 0;
A1 <- 0; A2 <- 0; A3 <- 0; A4 <- 0; A5 <- 0; A6 <- 0; A7 <- 0
use_data<- get(paste(Probanden_AllOver$Probandenname[i], j, sep="_"))
h<-0
if (length(use_data[,1]) <= 1) { Prob_Dat[1,j*5-3] <- 0} else {
repeat{h=h+1;
if (use_data [h,1] == 1) {if (use_data [h,2] == 1) {R1=R1+1; A1=A1+1} else {if (use_data [h,2] == 0)
{A1=A1+1} else {if (use_data [h,2] == 2) {R1=R1 + 0.5; A1=A1+1} else {}}}} else {}
if (use_data [h,1] == 2) {if (use_data [h,2] == 1) {R2=R2+1; A2=A2+1} else {if (use_data [h,2] == 0)
{A2=A2+1} else {if (use_data [h,2] == 2) {R2=R2 + 0.5; A2=A2+1} else {}}}} else {}
if (use_data [h,1] == 3) {if (use_data [h,2] == 1) {R3=R3+1; A3=A3+1} else {if (use_data [h,2] == 0)
{A3=A3+1} else {if (use_data [h,2] == 2) {R3=R3 + 0.5; A3=A3+1} else {}}}} else {}
if (use_data [h,1] == 4) {if (use_data [h,2] == 1) {R4=R4+1; A4=A4+1} else {if (use_data [h,2] == 0)
{A4=A4+1} else {if (use_data [h,2] == 2) {R4=R4 + 0.5; A4=A4+1} else {}}}} else {}
if (use_data [h,1] == 5) {if (use_data [h,2] == 1) {R5=R5+1; A5=A5+1} else {if (use_data [h,2] == 0)
{A5=A5+1} else {if (use_data [h,2] == 2) {R5=R5 + 0.5; A5=A5+1} else {}}}} else {}
if (use_data [h,1] == 6) {if (use_data [h,2] == 1) {R6=R6+1; A6=A6+1} else {if (use_data [h,2] == 0)
{A6=A6+1} else {if (use_data [h,2] == 2) {R6=R6 + 0.5; A6=A6+1} else {}}}} else {}
if (use_data [h,1] == 7) {if (use_data [h,2] == 1) {R7=R7+1; A7=A7+1} else {if (use_data [h,2] == 0)
{A7=A7+1} else {if (use_data [h,2] == 2) {R7=R7 + 0.5; A7=A7+1} else {}}}} else {};
if(h == max) break}
}
if (A1 == 0) { R1 <- 0} else{ R1 <- R1/A1}
if (A2 == 0) { R2 <- 0} else{ R2 <- R2/A2}
if (A3 == 0) { R3 <- 0} else{ R3 <- R3/A3}

```

```

if (A4 == 0) { R4 <- 0 } else{ R4 <- R4/A4}
if (A5 == 0) { R5 <- 0 } else{ R5 <- R5/A5}
if (A6 == 0) { R6 <- 0 } else{ R6 <- R6/A6}
if (A7 == 0) { R7 <- 0 } else{ R7 <- R7/A7}
Prob_Dat[1,j*5-3] <- (R1+R2+R3+R4+R5+R6+R7)/7
}

```

## Testleistung - Strukturiertheit

```

j<-1
for (j in 1:6) {
max_sl<- length(get(paste(Probanden_AllOver$Probandenname[i], j, sep="_")) [,1]); h <- 0;
sachlogik<- 0; anzahl<- 0
  R1 <- 88; R2 <- 88; R3 <- 88; R4 <- 88; R5 <- 88; R6 <- 88; R7 <- 88
  use_data<- get(paste(Probanden_AllOver$Probandenname[i], j, sep="_"))
  if (length(use_data[,1]) <= 1) { Prob_Dat[1,j*5-2] <- 0 } else {repeat{h=h+1; if (use_data [h,1] == 1
&&use_data [h,2] != 77) {R1 <- use_data [h,2]} else {}; if (use_data [h,1] == 2 &&use_data [h,2] != 77)
{R2 <- use_data [h,2]} else {}; if (use_data [h,1] == 3 &&use_data [h,2] != 77) {R3 <- use_data [h,2]}
else {}; if (use_data [h,1] == 4 &&use_data [h,2] != 77) {R4 <- use_data [h,2]} else {}; if (use_data [h,1]
== 5 &&use_data [h,2] != 77) {R5 <- use_data [h,2]} else {}; if (use_data [h,1] == 6 &&use_data [h,2]
!= 77) {R6 <- use_data [h,2]} else {}; if (use_data [h,1] == 7 &&use_data [h,2] != 77) {R7 <- use_data
[h,2]} else {}; if (use_data [h+1,1] == 9) {h=h+1} else{if ((use_data [h,1] == use_data [h+1,1]-1) ||
((use_data [h,1] == 7) && (use_data [h+1,1] == 1))) {sachlogik=sachlogik+1; anzahl=anzahl+1} else {if
(((use_data [h+1,1] == 1) && (R1 == 99)) || ((use_data [h+1,1] == 2) && (R2 == 99)) || ((use_data
[h+1,1] == 3) && (R3 == 99)) || ((use_data [h+1,1] == 4) && (R4 == 99)) || ((use_data [h+1,1] == 5)
&& (R5 == 99)) || ((use_data [h+1,1] == 6) && (R6 == 99)) || ((use_data [h+1,1] == 7) && (R7 == 99)))
} else {if (((use_data [h+1,1] == 1) && (R1 == 0 || R1 == 2 || R1 == 66)) || ((use_data [h+1,1] == 2)
&& (R2 == 0 || R2 == 2 || R2 == 66)) || ((use_data [h+1,1] == 3) && (R3 == 0 || R3 == 2 || R3 == 66))
|| ((use_data [h+1,1] == 4) && (R4 == 0 || R4 == 2 || R4 == 66)) || ((use_data [h+1,1] == 5) && (R5
== 0 || R5 == 2 || R5 == 66)) || ((use_data [h+1,1] == 6) && (R6 == 0 || R6 == 2 || R6 == 66)) ||
((use_data [h+1,1] == 7) && (R7 == 0 || R7 == 2 || R7 == 66))) {sachlogik=sachlogik+1;
anzahl=anzahl+1} else{if((use_data [h,1] == use_data [h+2,1]) && (is.logical(use_data [h+2,1])))
{h=h+1} else{if (((use_data [h,1] == 1) && (use_data [h+1,1] == 3)) || ((use_data [h,1] < 3) &&
(use_data [h+1,1] == 4)) || ((use_data [h,1] < 4) && (use_data [h+1,1] == 5) && (R4 == 1)) ||
((use_data [h,1] < 5) && (use_data [h+1,1] == 6) && (R5 == 1)) || ((use_data [h,1] < 6) && (use_data
[h+1,1] == 7) && (R6 == 1))) {sachlogik=sachlogik+1;anzahl=anzahl+1} else{anzahl=anzahl+1}}}} }; if(h
>= max_sl-1) break}; Prob_Dat[1,j*5-2] <- sachlogik/anzahl}
}

```

## Testleistung - Zielorientiertheit

```

for (j in 1:6) {
  use_data<- get(paste(Probanden_AllOver$Probandenname[i], j, sep="_"))
  max_stra<- length(get(paste(Probanden_AllOver$Probandenname[i], j, sep="_")) [,1])
  if (length(get(paste(Probanden_AllOver$Probandenname[i], j, sep="_")) [,1]) <= 1) {
  Prob_Dat[1,j*5-1] <- 0 } else {
  repeat{max_stra=max_stra-1; if (use_data [max_stra+1,1] == 9) { use_data<- use_data [-
c(max_stra+1,)] } else {}; if(max_stra == 0) break};

```

```

max_stra<- length(use_data [,1]); h <- 0; Prod_Zielorientiertheit<- 0; anzahl<- 1;
if (length(use_data [,1]) <= 1) { Prob_Dat[1,j*5-1] <- 0} else {
repeat{h=h+1; if (use_data [h,1] <use_data [h+1,1]) {} else {anzahl=anzahl+1}; if(h == max_stra-1)
break};
if (anzahl> 6) {Prob_Dat[1,j*5-1] <- 0} else {Prob_Dat[1,j*5-1] <- (1-1/6*anzahl)}
}
}
}

```

```

##Übersichtsdatei für den einzelnen Probanden und Gesamtübersicht erzeugen
assign(Prob_Nam, Prob_Dat)
save (list = paste(Probanden_AllOver$Probandenname[i], "", sep=""),
file=paste(Probanden_AllOver$Probandennamen[i], ".rda", sep=""))
Auswertung <- rbind(Auswertung, Prob_Dat)
}

```

```

## Gesamtübersicht
Auswertung <- Auswertung [-c(1),]

```

```

## Gesamtscores
max<- length(Auswertung[,1])
for (i in 1:max) {
Auswertung[i,6] <- (Auswertung[i,2]+Auswertung[i,3]+Auswertung[i,4])/3
Auswertung[i,11] <- (Auswertung[i,7]+Auswertung[i,8]+Auswertung[i,9])/3
Auswertung[i,16] <- (Auswertung[i,12]+Auswertung[i,13]+Auswertung[i,14])/3
Auswertung[i,21] <- (Auswertung[i,17]+Auswertung[i,18]+Auswertung[i,19])/3
Auswertung[i,26] <- (Auswertung[i,22]+Auswertung[i,23]+Auswertung[i,24])/3
Auswertung[i,31] <- (Auswertung[i,27]+Auswertung[i,28]+Auswertung[i,29])/3
Auswertung[i,35] <- (Auswertung[i,6]+Auswertung[i,11]+Auswertung[i,16]+Auswertung[i,21]+Auswertung[i,26]+Auswertung[i,31])/6
Auswertung[i,32] <- (Auswertung[i,2]+Auswertung[i,7]+Auswertung[i,12]+Auswertung[i,17]+Auswertung[i,22]+Auswertung[i,27])/6
Auswertung[i,33] <- (Auswertung[i,3]+Auswertung[i,8]+Auswertung[i,13]+Auswertung[i,18]+Auswertung[i,23]+Auswertung[i,28])/6
Auswertung[i,34] <- (Auswertung[i,4]+Auswertung[i,9]+Auswertung[i,14]+Auswertung[i,19]+Auswertung[i,24]+Auswertung[i,29])/6
}
}

```

```

Auswertung

```

## A.4.10 Manual zur Testeinführung Think Aloud

### 1. Erklärung warum es in den nächsten 4 Stunden gehen wird

*Liebe(r) Studierende(r), in den nächsten etwa 4 Stunden sollen Sie einen Experimentiertest und einen Fachwissenstest zum Thema Optik bearbeiten. Dies dient nicht dazu ihr Wissen und ihre Fähigkeiten zu testen, sondern um den Test zu erproben. Wir möchten herausfinden, wie gut sich Ihre Lösungsprozesse beim Experimentieren mit dem Test erfassen lassen.*

### 2. Think Aloud Demonstrieren

*Dazu werden wir eine Technik benutzen die „Lautes Denken“ heißt. Dabei sollen Sie Ihre Gedanken laut äußern. Das heißt:*

- *Wenn Sie etwas lesen, lesen Sie es bitte laut vor.*
- *Wenn Sie etwas schreiben, sagen Sie bitte laut was Sie schreiben.*
- *Wenn Sie über etwas nachdenken, teilen Sie ihre Gedanken laut mit.*
- *Wenn Sie etwas tun, zum Beispiel Material nehmen oder einen Versuch aufbauen, sagen Sie laut, welche Überlegungen dazu geführt haben.*

*Am besten wäre, wenn Sie die ganzen 4 Stunden ohne Pause reden. Als Testleiter werde ich Sie daran erinnern, wenn Sie aufhören laut zu denken. Zudem werde ich Ihnen nach dem Bearbeiten jeder Experimentieraufgabe einige Fragen stellen. Abgesehen von diesen Fragen und der Erinnerung ans laute Denken werde ich keine Tipps oder Anregungen geben.*

*Um zu konkretisieren wie die Methode „Lautes Denken“ funktioniert, gebe ich Ihnen nun ein kurzes Beispiel:*

*<Beispiel: Anschließen und Testen der Kameras>*

### 3. Warm Up

*Nachdem ich nun erklärt habe warum es geht, was lautes Denken ist und es an einem Beispiel demonstriert habesind Sie nun an der Reihe. Bevor wir mit dem eigenlichtenTest beginnen gibt es noch eine Übungsphase. Zweck dieser Übungsphase ist, dass Sie die Experimentierbox zunächst kennen lernen und sich dabei an die Methode „Lautes Denken“ gewöhnen.*

*Nutzen Sie die Inventarliste um die Experimentierbox kennen zu lernen und sich mit dem Material vertraut zu machen.*

*Haben Sie noch Fragen?*

*Wenn nicht, dann beginnen wir nun mit dem Experimentiertest.*

#### 4. Testdurchführung (jeweils 6x mit kurzer Pause zwischen den Aufgaben)

*Wir beginnen jetzt mit Aufgabe X. Bitte bearbeiten Sie die Aufgabe. Denken Sie daran alle Ihre Gedanken laut mitzuteilen*

<Proband bearbeitet eine Experimentieraufgabe für maximal 25 Minuten oder so lange bis er angibt die Aufgabe gelöst zu haben>

##### Intervention:

Handlung des Probanden	Intervention
Wenn der Proband aufhört zu reden und ...	
liest.	<i>Bitte lesen Sie laut.</i>
schreibt, etwas notiert oder zeichnet.	<i>Bitte lesen Sie laut vor was Sie schreiben, notieren oder zeichnen.</i>
mit Material interagiert bzw. experimentiert.	<i>Erzählen Sie bitte was Sie gerade tun.</i>
keine der drei Handlungen vollzieht.	<i>Können Sie bitte erzählen worüber Sie nachdenken.</i>
Wenn noch 10 Minuten Bearbeitungszeit zur Verfügung stehen.	<i>Sie haben noch 10 Minuten um diese Aufgabe zu bearbeiten.</i>
Wenn noch 5 Minuten Bearbeitungszeit zur Verfügung stehen.	<i>Sie haben noch 5 Minuten um diese Aufgabe zu bearbeiten.</i>

Interview nach Abschluss der Experimentieraufgabe:

*Vielen Dank. Bevor wir mit der nächsten Aufgabe weiter machen / Bevor wir den Test beenden, habe ich einige kurze Fragen zu der bearbeiteten Aufgabe.*

- Wie haben Sie die Problemstellung bearbeitet? Bitte fassen Sie Ihre zentralen Überlegungen kurz zusammen!*
- Bitte schätzen Sie ein, wie wichtig es war „Hypothesen zu bilden“ und „Versuchsaufbauten zu planen“ um die Problemstellung zu lösen. War es „nicht wichtig“, „wenig wichtig“, „eher wichtig“ oder „sehr wichtig“?*
- Bitte schätzen Sie ein, wie wichtig es war „Versuche aufzubauen“ und „zu messen bzw. zu beobachten“ um die Problemstellung zu lösen. War es „nicht wichtig“, „wenig wichtig“, „eher wichtig“ oder „sehr wichtig“?*
- Bitte schätzen Sie ein, wie wichtig es war „Daten zu analysieren“ und „Sachgerechte Schlüsse zu ziehen“ um die Problemstellung zu lösen. War es „nicht wichtig“, „wenig wichtig“, „eher wichtig“ oder „sehr wichtig“?*

*Bevor wir fortfahren können wir eine kurze Pause von 5 Minuten machen, in der Sie sich entspannen oder eine Kleinigkeit essen bzw. trinken können.*

## 5. Fachwissenstest

*Vielen Dank für die Bearbeitung der Aufgaben. Zum Abschluss möchte ich Sie bitten, noch einen kurzen Wissenstest zu den fachlichen Inhalten des Experimentiertests zu beantworten. Bearbeiten Sie dazu alle 15 Aufgaben und sagen Sie dabei laut was Ihnen in den Kopf kommt.*

Intervention:

<b>Handlung des Probanden</b>	<b>Intervention</b>
Wenn der Proband aufhört zu reden und ...	
liest.	<i>Bitte lesen Sie laut.</i>
schreibt, etwas notiert oder zeichnet.	<i>Bitte lesen Sie laut vor was Sie schreiben, notieren oder zeichnen.</i>
keine der beiden Handlungen vollzieht.	<i>Können Sie bitte erzählen worüber Sie nachdenken.</i>

Vielen Dank das Sie an der Erprobung teilgenommen haben

## A.4.11 Manual zur Testeinführung Feldstudie

### 1. Erklärung worum es in den nächsten 4 Stunden gehen wird

*Liebe(r) Studierende(r), in den nächsten etwa 4 Stunden sollen Sie einen Experimentiertest und ein schriftlichen Test zum Thema Optik bearbeiten. Dies dient dazu ihr Wissen und ihre Fähigkeiten zu testen, und um den Test zu erproben. Wir möchten herausfinden, wie gut sich Ihre Lösungsprozesse beim Experimentieren mit dem Test erfassen lassen.*

*Um diesen Praktikumsversuch zu bestehen, benötigen Sie die Hälfte der möglichen Punkte aus dem Experimentiertest. Alle Aufgaben sind gleich gewichtet. Teilen Sie sich wie in einer normalen Testsituation Ihre Zeit selbstständig ein, um ein möglichst gutes Ergebnis zu erzielen und alle Aufgaben zu bearbeiten. Sie haben bis 11:45 / 16:15 Uhr Zeit für die Bearbeitung des Experimentiertests. Danach müssen Sie mit den schriftlichen Aufgaben beginnen.*

### 2. Prozessdaten

*Damit wir nicht nur die Ergebnisse ihrer Arbeit bewerten können, versuchen wir den Prozess einzubeziehen. Dazu ist es notwendig, dass wir wissen was sie tun. Entsprechend fordern wir sie auf, während der Bearbeitung ihre Handlungen zu notieren. Damit sie keine Zeit verlieren, steht ihnen ein Raster zur Verfügung in dem sie Kreuze machen können. <Austeilen der Aufzeichnungszettel>*

*Auf diesen Blättern sollen Sie alle Ergebnisse festhalten und wie im Folgenden beschrieben Handlungen angekreuzen. Es ist ausreichend wenn jede Handlung entweder mit einer Aufzeichnung, einem Stichwort oder einem Kreuz dokumentiert wird. Damit wir die Aufzeichnungen und Kreuze einer Reihenfolge zuordnen können, schreiben sie bitte von oben nach unten und starten Sie nach jedem Kreuz bzw. jeder Aufzeichnung in der jeweiligen nächsten Zeile. Das bedeutet in jeder Zeile sollte entweder ein Kreuz oder eine Aufzeichnung stehen. Haben Sie fragen dazu?*

*Damit Sie nicht vergessen ihre Handlungen zu notieren, werden wir sie alle 60 Sekunden mit diesem Ton <Glocke> daran erinnern. Das bedeutet nicht, dass sie bei jedem Ton ein Kreuz machen müssen. Es bedeutet aber, dass sie kurz nachdenken sollten, ob alle ihre Handlungen der letzten Minute notiert wurden.*

### 3. Handlungen

*Auf dem Blatt für ihre Lösung sind vier Handlungen zum ankreuzen angegeben: Fragestellung, Theorie, Rechnen und Experimentieren. Im Folgenden möchten wir ihnen kurz vorstellen, welche Handlungen entsprechend zugeordnet werden sollten.*

*Fragestellung:*

*Immer wenn Sie sich mit der Fragestellung beschäftigen, würden Sie hier ein Kreuz machen. Das bedeutet wenn Sie die Fragestellung lesen, darüber nachdenken oder versuchen sie zu verstehen.*

*Theorie:*

*Sie erhalten theoretische Hinweise und haben die Möglichkeit sich selbst Gedanken zu der benötigten Theorie zum Experiment zu machen. Sollten Sie dies tun oder in den theoretischen Hinweisen lesen, so machen Sie ihr Kreuz hier.*

*Rechnen:*

*Sollten Sie Rechnungen durchführen, auch kurze Rechnungen um Ergebnisse abschätzen zu können, so machen sie hier ein Kreuz.*

*Experimentieren:*

*Jede Interaktion mit dem Versuchsmaterial fällt unter diese Kategorie. Wenn Sie sich also Material anschauen, einen Versuch aufbauen oder Messungen durchführen, kurz immer wenn sie irgendetwas mit dem Experimentiermaterial machen, so notieren Sie dies mit einem Kreuz in dieser Kategorie.*

*Sonstiges:*

*Sollten Sie Handlungen durchführen, die zu keiner der Kategorien passen und zu denen Sie keine Aufzeichnungen machen, so schreiben Sie KURZ 1-3 Stichworte auf.*

#### **4. Fotos**

*Damit wir ihre Versuchsaufbauten bewerten können, die Sie für ihre Messungen bzw. Beobachtungen nutzen, müssen wir diese sehen. Deshalb erhalten Sie einen Fotoapparat und Karten mit Nummern. Bitte legen Sie nach jedem Auf- und Umbau eines Versuchs eine Karte mit einer Zahl neben den Versuch und machen Sie 1-3 Fotos. In den Aufzeichnungen notieren Sie die Zahl auf der Karte damit wir die Fotos zuordnen können. Es ist durchaus möglich, dass pro Aufgabe mehrere Aufbauten und entsprechend Fotos gemacht werden. Sie können den Fotoapparat und die Zahlenkarten auch nutzen, um ihre Aufzeichnungen zusätzlich zu dokumentieren, beispielsweise bei Beobachtungen. Haben Sie Fragen?*

*Wenn nicht, dann beginnen wir nun mit dem Experimentiertest.*



## A.4.12 Globalmaße zur Experimentierkompetenz

### Globalmaß Richtigkeit

- **Richtige Planung** (Der Probanden hat die Theorie verstanden, weiß was er messen muss und wie er messen muss. – z.B. Richtige Formel, Zu messende Variablen, Beschreibung der Messung, Skizze des Versuchsaufbaus - 1 Punkt)
- **Richtige Umsetzung** (Das Ergebnis ist richtig und ergibt sich aus den Messwerten. – z.B. richtige Messwerte, richtige Rechnungen, richtiges Ergebnis - 1 Punkt)

### Globalmaß Strukturiertheit

- **Strukturierte Aufzeichnungen** (Klare Struktur der Aufzeichnungen erkennbar. – z.B. Gliederung nach 1. Planung, 2. Durchführung und 3. Auswertung - 1 Punkt)
- **Qualitätssicherung** (Der Proband reflektiert die Qualität der Arbeit – z.B. werden Fehler erkannt und gelöst - 1 Punkt)

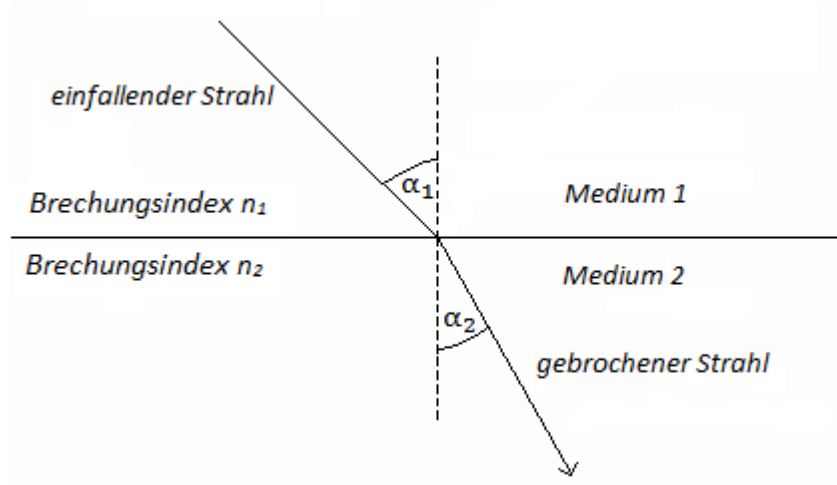
### Globalmaß Zielorientiertheit

- **Zielorientiertes Vorgehen** (Es ist ein Fortschritt zur Lösung des Problems zu erkennen – z.B. indem ein Erkenntnisgewinn deutlich wird - 1 Punkt; ist dieser Fortschritt sogar stringent - 2 Punkte)

## A.5 Fachwissenstest

### Aufgabe 1

Welche der folgenden Formeln gibt das Brechungsgesetz für den in der Abbildung gezeigten Strahlenverlauf korrekt wieder?



$$\frac{\sin \alpha_1}{\sin \alpha_2} = \frac{n_1}{n_2} \quad \input{checkbox}$$

$$\frac{\sin \alpha_1}{\sin \alpha_2} = \frac{n_2}{n_1} \quad \input{checkbox}$$

$$\frac{\cos \alpha_1}{\cos \alpha_2} = \frac{n_1}{n_2} \quad \input{checkbox}$$

$$\frac{\cos \alpha_1}{\cos \alpha_2} = \frac{n_2}{n_1} \quad \input{checkbox}$$

### Aufgabe 2

Wie groß ist die Bildweite  $b$  bei der Abbildung eines Gegenstands mit einer Linse der Brennweite  $f = 40$  mm und einer Gegenstandsweite  $g = 90$  mm?

$$b = 50 \text{ mm} \quad \input{checkbox}$$

$$b = 72 \text{ mm} \quad \input{checkbox}$$

$$b = 225 \text{ mm} \quad \input{checkbox}$$

Das hängt von Bild- und Gegenstandsgröße ab.

### Aufgabe 3

Wodurch wird das Auflösungsvermögen optischer Mikroskope fundamental limitiert?

Durch die Vergrößerung des Mikroskops.

Durch die Reflexion an den Innenflächen.

Durch die Beugung des Lichts.

Durch die Streuung des Lichts.

### Aufgabe 4

Welche der folgenden Vorgehensweisen ist geeignet, um herausfinden, ob es sich bei einem konkaven Spiegel um einen Parabolspiegel handelt?

Man vergleicht die Lage der Brennpunkte für Strahlen verschiedener Wellenlänge.

Man vergleicht die Lage der Brennpunkte für achsennahe- und achsenferne Strahlen die parallel zur optischen Achse auf den Spiegel treffen.

Man vergleicht die Lage der Brennpunkte für achsennahe Strahlen die parallel oder in einem Winkel von  $45^\circ$  auf den Spiegel treffen.

Man vergleicht die Lage der Brennpunkte bei der Abbildung von nahen und weit entfernten Objekten.

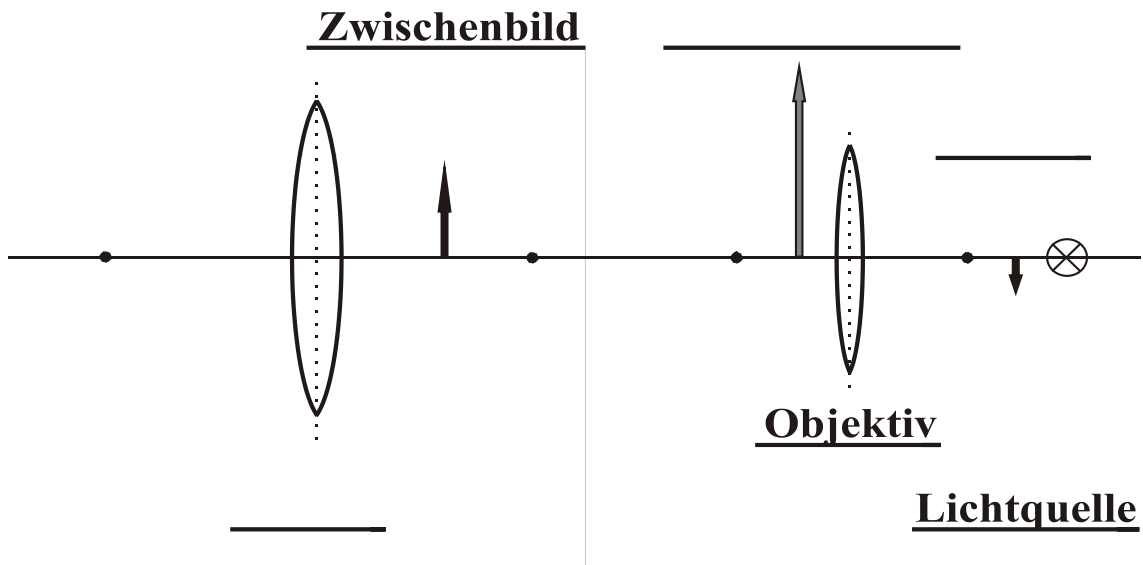
### Aufgabe 5

Welche der folgenden Aussagen zur Polarisation eines Lichtbündels sind richtig, welche falsch?

	Richtig	Falsch
Wenn das magnetische Feld aller Lichtwellen in einer Ebene schwingt, so ist das Lichtbündel polarisiert.	<input type="checkbox"/>	<input type="checkbox"/>
Wenn die Amplitude aller Lichtwellen gleich ist, so ist das Lichtbündel polarisiert.	<input type="checkbox"/>	<input type="checkbox"/>
Wenn die Ausbreitungsgeschwindigkeit aller Lichtwellen gleich ist, so ist das Lichtbündel polarisiert.	<input type="checkbox"/>	<input type="checkbox"/>
Wenn das elektrische Feld aller Lichtwellen in einer Ebene schwingt, so ist das Lichtbündel polarisiert.	<input type="checkbox"/>	<input type="checkbox"/>

### Aufgabe 6

Die folgende Abbildung zeigt den schematischen Aufbau eines Mikroskops. Die Punkte auf der optischen Achse bezeichnen die Brennpunkte der einzelnen Linsen. Vervollständigen Sie die Beschriftung der abgebildeten Bestandteile des Mikroskops.



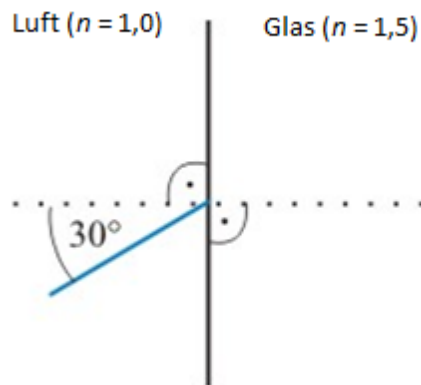
### Aufgabe 7

Woran kann man erkennen, ob ein Farbspektrum durch ein Gitter oder ein Prisma erzeugt wurde?

- Beim Gitter wird Licht kleiner Wellenlängen weniger stark abgelenkt, beim Prisma ist es umgekehrt.
- Im Gegensatz zum Prisma lässt sich durch das Gitter der gesamte Wellenlängenbereich abbilden.
- Beim Gitter entsteht ein Spektrum aus einzelnen Linien, wohingegen beim Prisma ein kontinuierliches Bandspektrum entsteht.
- Gitter und Prismen erzeugen immer gleiche Farbspektren. Es gibt keinen Unterschied.

### Aufgabe 8

Ein Lichtstrahl trifft – wie abgebildet – unter einem Winkel von  $30^\circ$  auf eine Glasscheibe.



In welchem Winkel, bezüglich der Normalen zur Oberfläche, tritt der Lichtstrahl in das Glas ein?  
[Hinweis:  $\sin(30^\circ) = 0,50$ ]

$\arcsin\left(\frac{1}{3}\right) \approx 19^\circ$

$\arcsin\left(\frac{3}{8}\right) \approx 22^\circ$

$\arcsin\left(\frac{2}{3}\right) \approx 42^\circ$

$\arcsin\left(\frac{3}{4}\right) \approx 49^\circ$

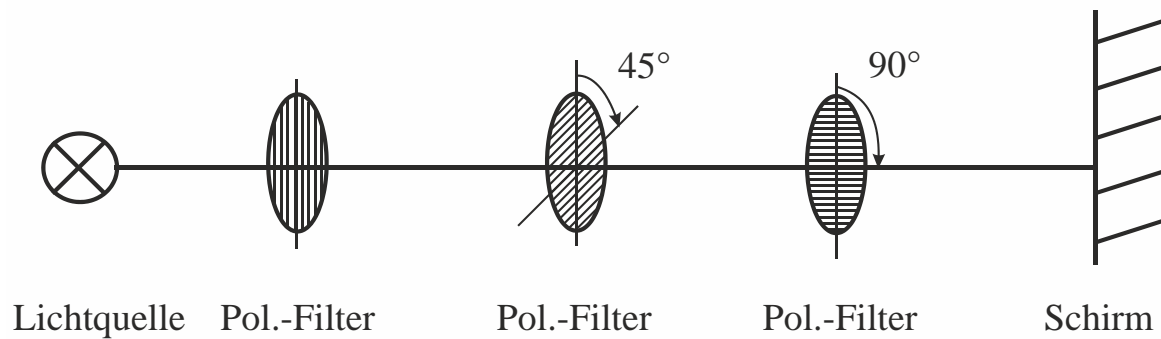
### Aufgabe 9

Welche der folgenden Eigenschaften einer elektromagnetischen Welle ändert sich beim Übergang zwischen verschiedenen optischen Medien.

	Ändert sich	Ändert sich nicht
Wellenlänge	<input type="checkbox"/>	<input type="checkbox"/>
Frequenz	<input type="checkbox"/>	<input type="checkbox"/>
Ausbreitungsgeschwindigkeit	<input type="checkbox"/>	<input type="checkbox"/>
Amplitude	<input type="checkbox"/>	<input type="checkbox"/>

### Aufgabe 10

Eine Lichtquelle emittiert unpolarisiertes Licht, welches auf einen Schirm trifft. Zwischen die Lichtquelle und den Schirm werden, wie in der Abbildung gezeigt, drei Polarisationsfilter gebracht. Der Winkel der Polarisations Ebene des ersten Filters beträgt  $0^\circ$ , der des Zweiten  $45^\circ$  und der des Dritten  $90^\circ$ .



Welcher Anteil der ursprünglichen Lichtintensität erreicht den Schirm?

0 %

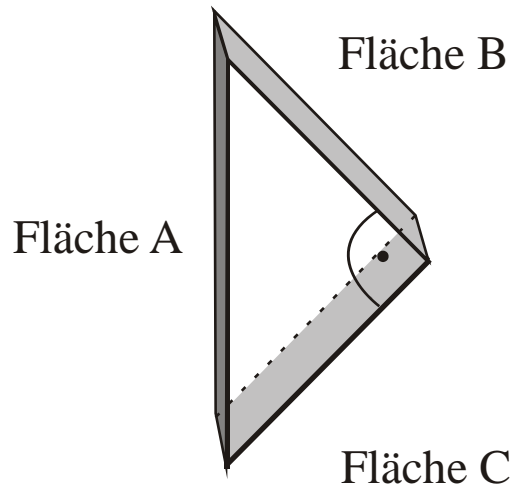
12,5 %

25 %

50 %

### Aufgabe 11

Ein Lichtstrahl tritt durch die Fläche A in das abgebildete gleichschenklige Prisma ein, wird an den zueinander senkrechten Flächen B und C reflektiert und tritt durch die Fläche A wieder aus.



Welche der folgenden Aussagen über den Ausfallswinkel des Strahls ist richtig?

	Richtig	Falsch
Der Ausfallswinkel ist kleiner als der Einfallswinkel.	<input type="checkbox"/>	<input type="checkbox"/>
Der Ausfallswinkel ist gleich dem Einfallswinkel.	<input type="checkbox"/>	<input type="checkbox"/>
Der Ausfallswinkel ist größer als der Einfallswinkel.	<input type="checkbox"/>	<input type="checkbox"/>
Das Verhältnis von Aus- und Einfallswinkel hängt vom Brechungsindex des Prismas ab.	<input type="checkbox"/>	<input type="checkbox"/>

### Aufgabe 12

Welche Art von Spektrum entsteht durch Sonnenlicht, welches durch eine Schicht Natriumgas auf einen Schirm trifft?

monochromatisches Spektrum

Linienpektrum

Bandspektrum

kontinuierliches Spektrum

### Aufgabe 13

An einem Strichgitter wird Licht gebeugt. Auf einem Schirm lässt sich ein Interferenzmuster beobachten.

Wodurch entsteht dieses Interferenzmuster?

Durch die Abbildung des Schattenmusters von dem Strichgitter.

Durch die Wellenlängenabhängigkeit der Brechung am Strichgitter.

Durch konstruktive und destruktive Überlagerung von Elementarwellen.

Durch Verringerung der Intensitäten bei bestimmten Streuwinkeln.



### Aufgabe 14

Der Durchmesser des Strahlenbündels, welcher das Okular eines Fernrohrs verlässt, wird als *Austrittspupille* bezeichnet. Welche der folgenden Formeln gibt die Austrittspupille  $AP$  eines Fernrohrs mit Objektivdurchmesser  $D$  und Vergrößerung  $V$  an?

$$AP = D \cdot V \quad \square$$

$$AP = D \cdot V^2 \quad \square$$

$$AP = \frac{D}{V} \quad \square$$

$$AP = \frac{V}{D} \quad \square$$

### Aufgabe 15

Bei welchen der folgenden Phänomene wird polarisiertes Licht erzeugt?

	Richtig	Falsch
Beugung	<input type="checkbox"/>	<input type="checkbox"/>
Brechung	<input type="checkbox"/>	<input type="checkbox"/>
Reflexion	<input type="checkbox"/>	<input type="checkbox"/>
Streuung	<input type="checkbox"/>	<input type="checkbox"/>

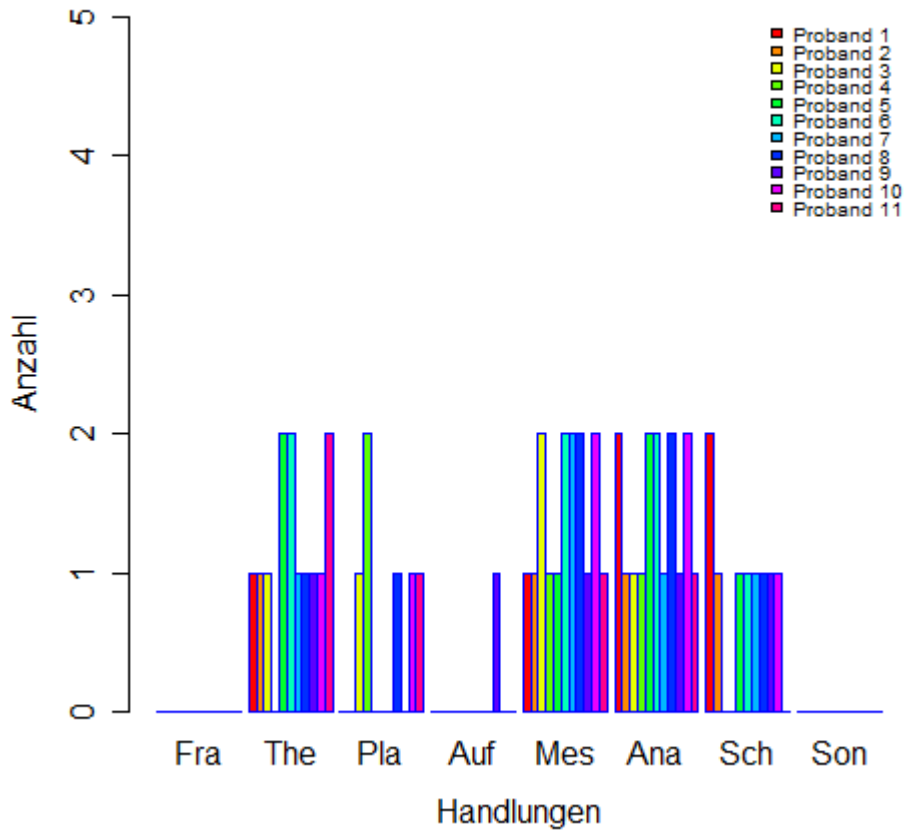
## A.6 Lautes Denken: Gesamtleistung der Probanden über alle Auswertungsverfahren

Tabelle 35: Gesamtleistung über alle Auswertungsverfahren.

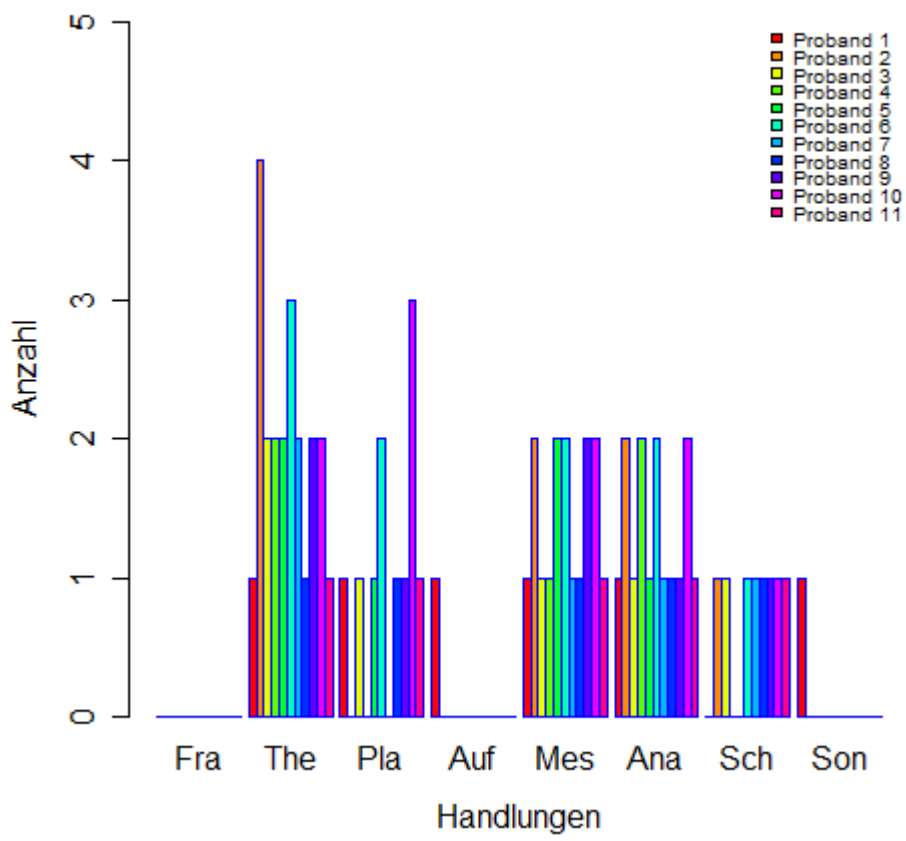
		Richtigkeit	Strukturiertheit	Zielorientiertheit	Gesamt
P1	Referenz	.68	.58	.46	.57
	Prozess	.49	.71	.67	.62
	Produkt	.32	.34	.54	.40
P2	Referenz	.71	.60	.71	.67
	Prozess	.44	.77	.58	.60
	Produkt	.36	.32	.71	.46
P3	Referenz	.67	.72	.33	.57
	Prozess	.48	.74	.62	.61
	Produkt	.35	.36	.71	.47
P4	Referenz	.44	.66	.37	.49
	Prozess	.21	.83	.58	.54
	Produkt	.095	.34	.58	.34
P5	Referenz	.50	.70	.51	.57
	Prozess	.47	.81	.65	.64
	Produkt	.33	.36	.69	.46
P6	Referenz	.55	.62	.35	.51
	Prozess	.37	.72	.38	.49
	Produkt	.25	.21	.37	.28
P7	Referenz	.75	.74	.57	.69
	Prozess	.22	.85	.55	.54
	Produkt	.20	.25	.43	.29
P8	Referenz	.72	.68	.47	.62
	Prozess	.52	.75	.51	.60
	Produkt	.42	.28	.70	.47
P9	Referenz	.70	.69	.58	.66
	Prozess	.54	.66	.49	.56
	Produkt	.45	.44	.83	.57
P10	Referenz	.73	.69	.69	.70
	Prozess	.55	.66	.60	.60
	Produkt	.37	.53	.82	.57
P11	Referenz	.61	.62	.40	.54
	Prozess	.37	.78	.55	.57
	Produkt	.21	.30	.87	.46

# 7 Lauter Denken Studie: Handlungsverteilung über alle Aufgaben

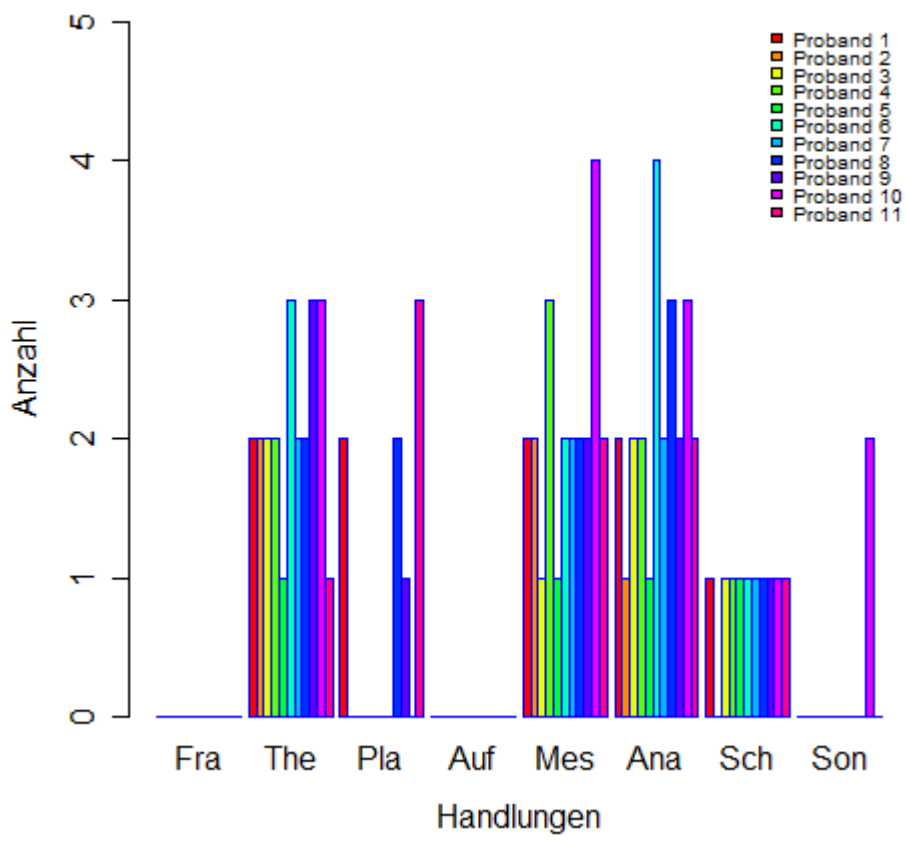
## Produktorientierte Auswertung - Aufgabe 1



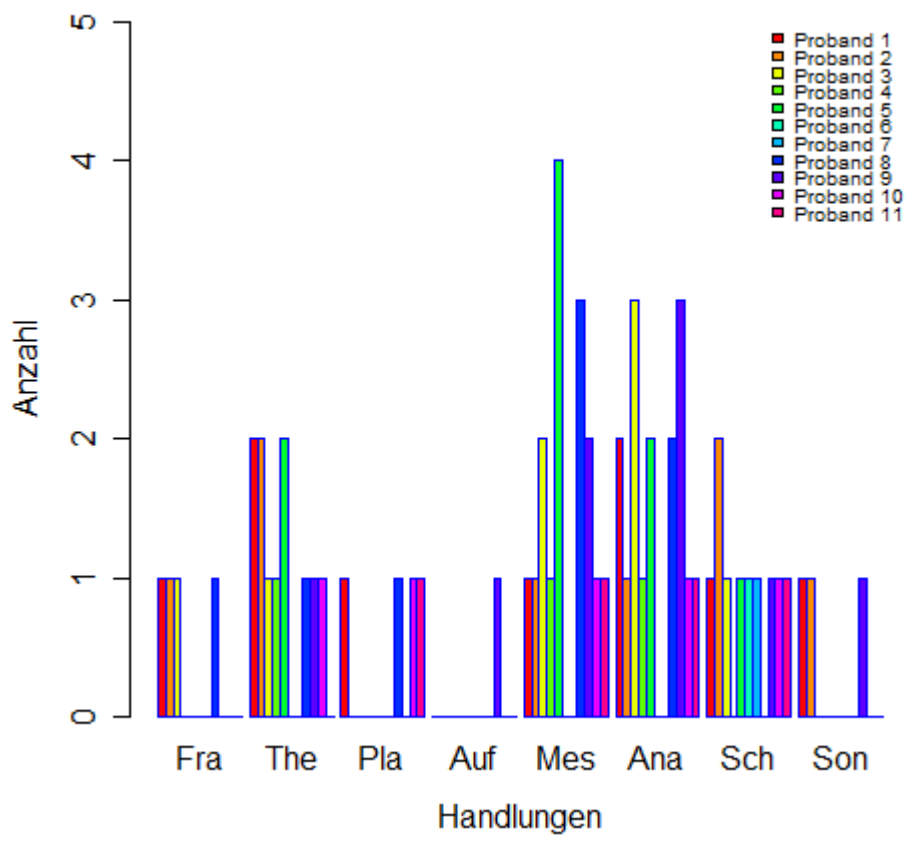
## Produktorientierte Auswertung - Aufgabe 2



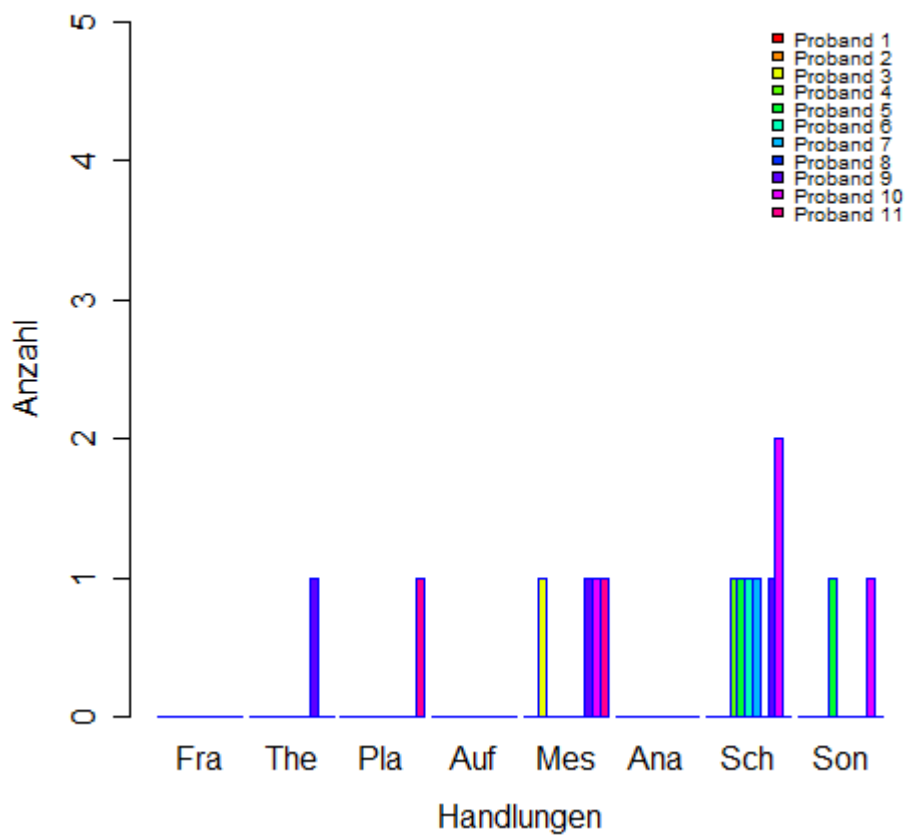
### Produktorientierte Auswertung - Aufgabe 3



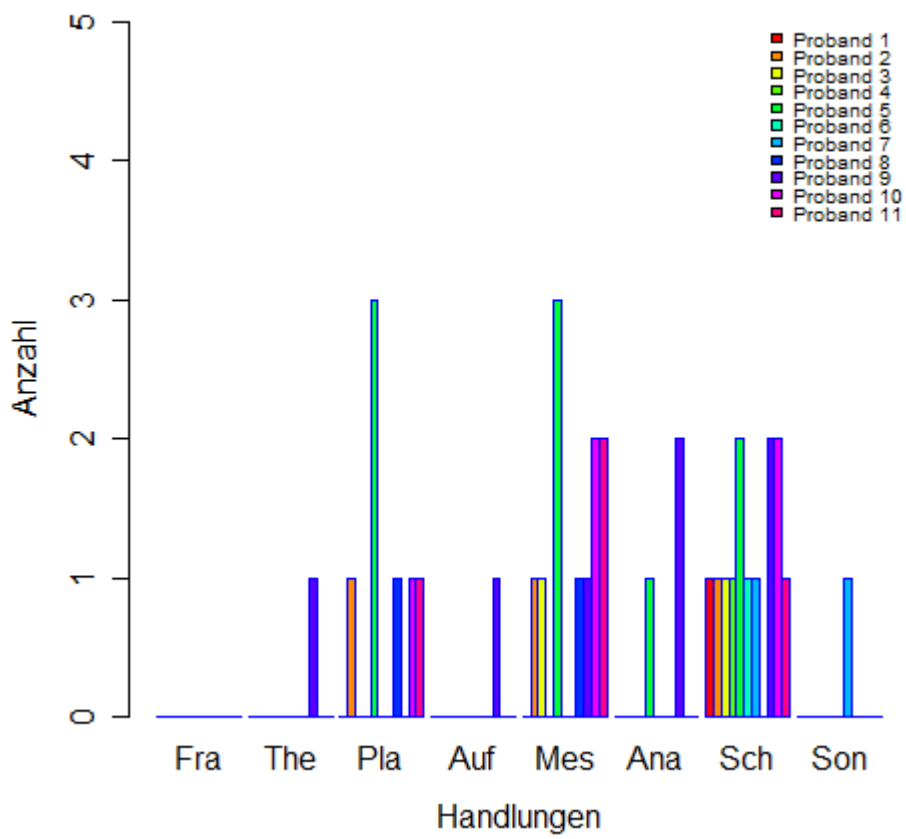
## Produktorientierte Auswertung - Aufgabe 4



## Produktorientierte Auswertung - Aufgabe 5

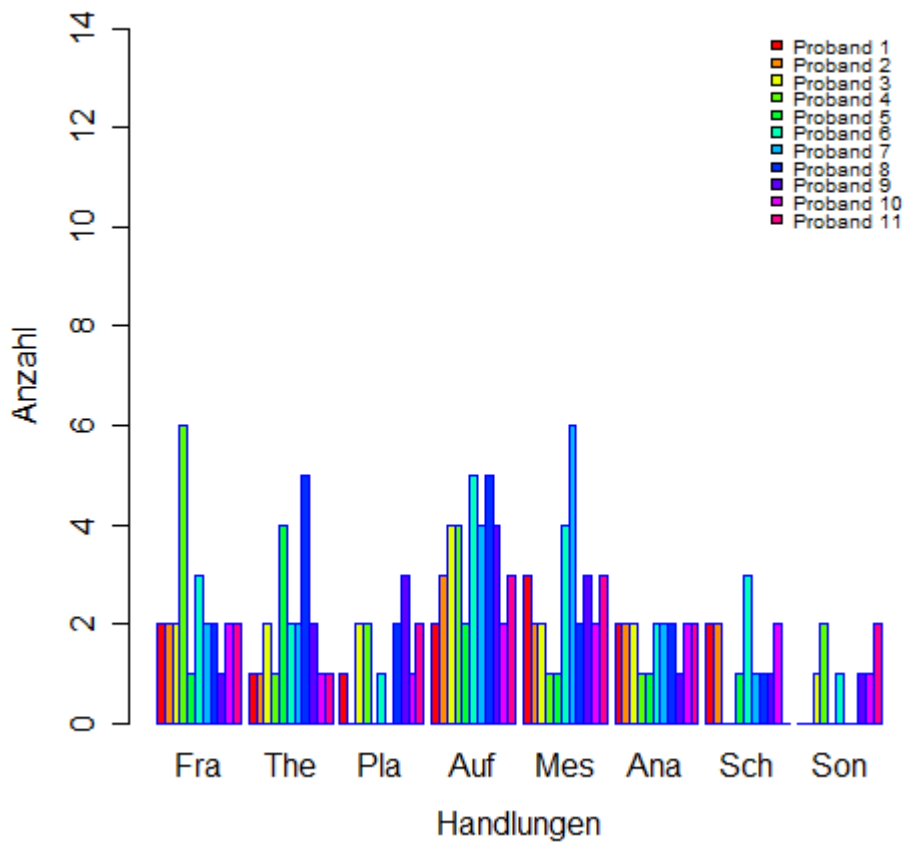


## Produktorientierte Auswertung - Aufgabe 6

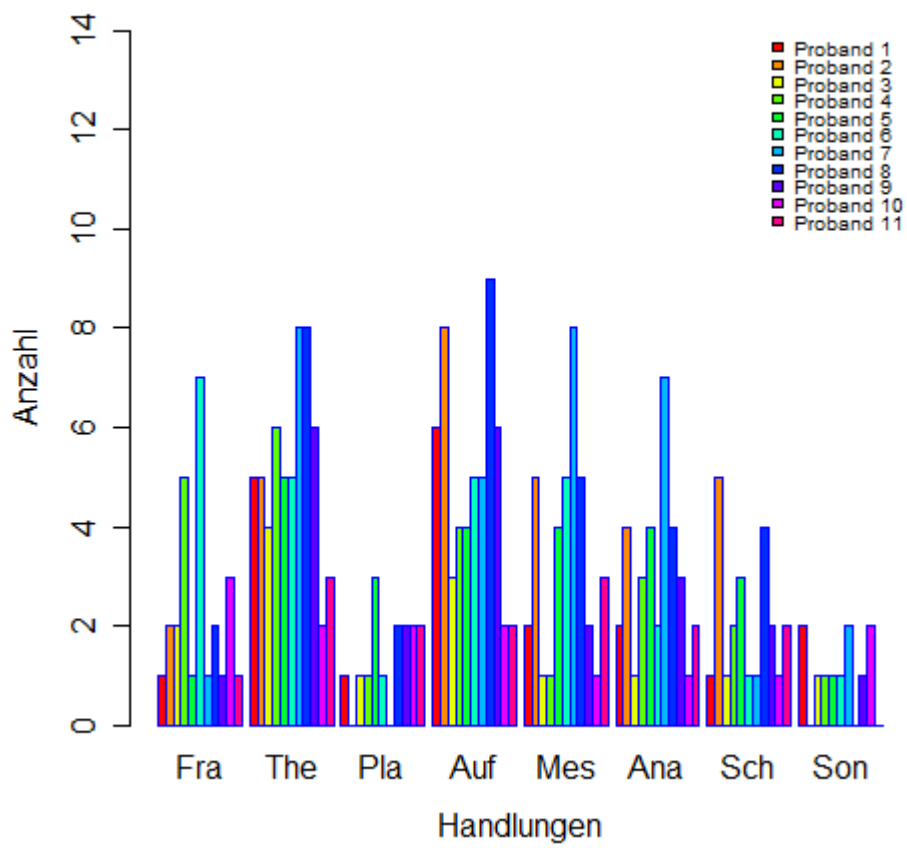




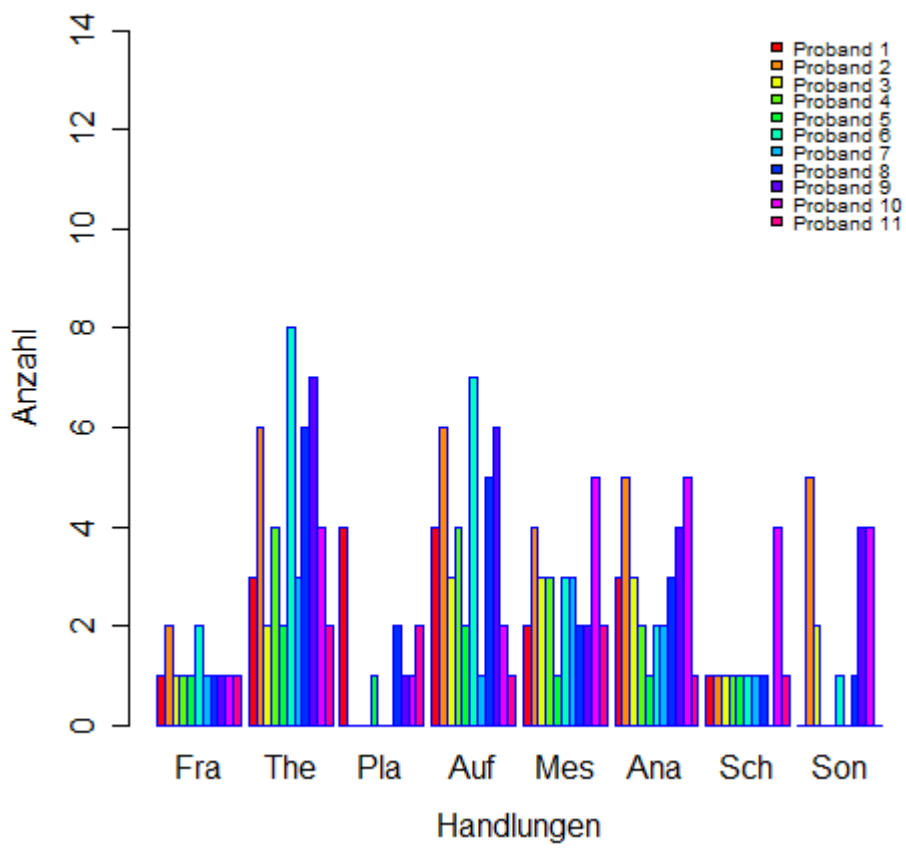
## Prozessorientierte Auswertung - Aufgabe 1



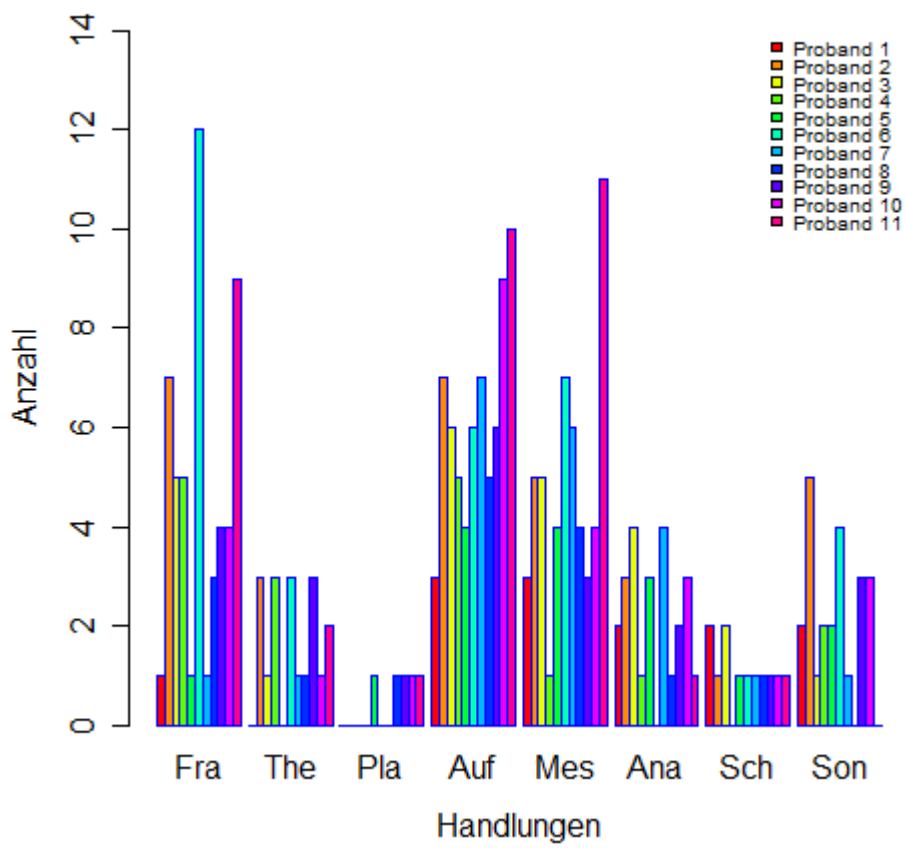
## Prozessorientierte Auswertung - Aufgabe 2



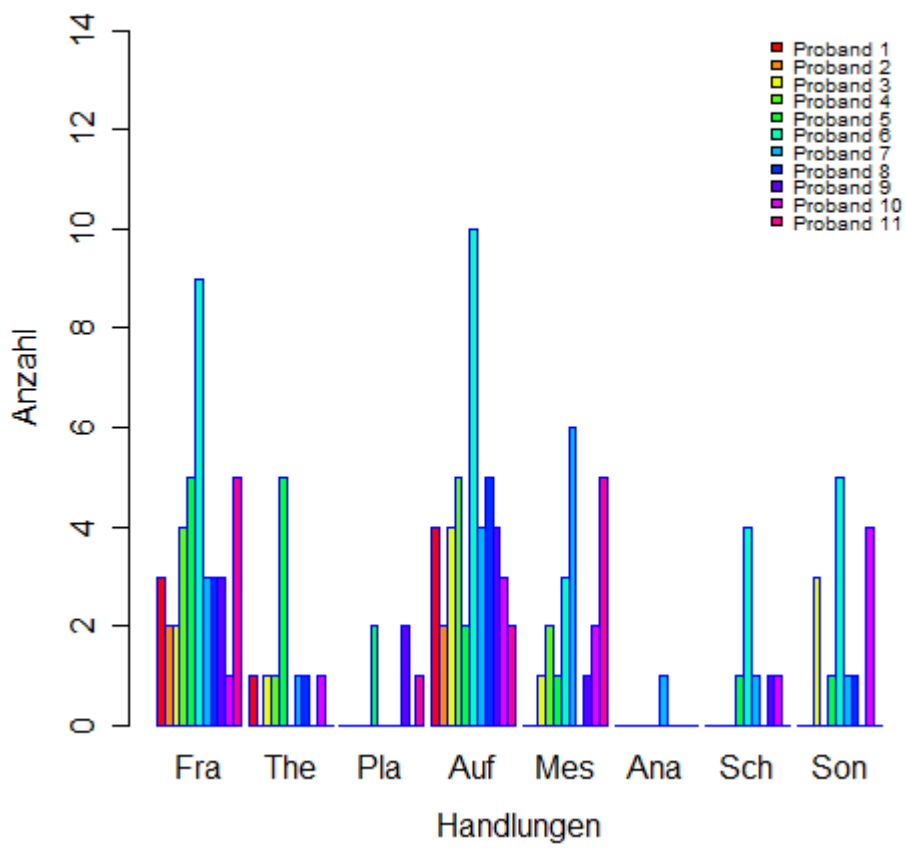
### Prozessorientierte Auswertung - Aufgabe 3



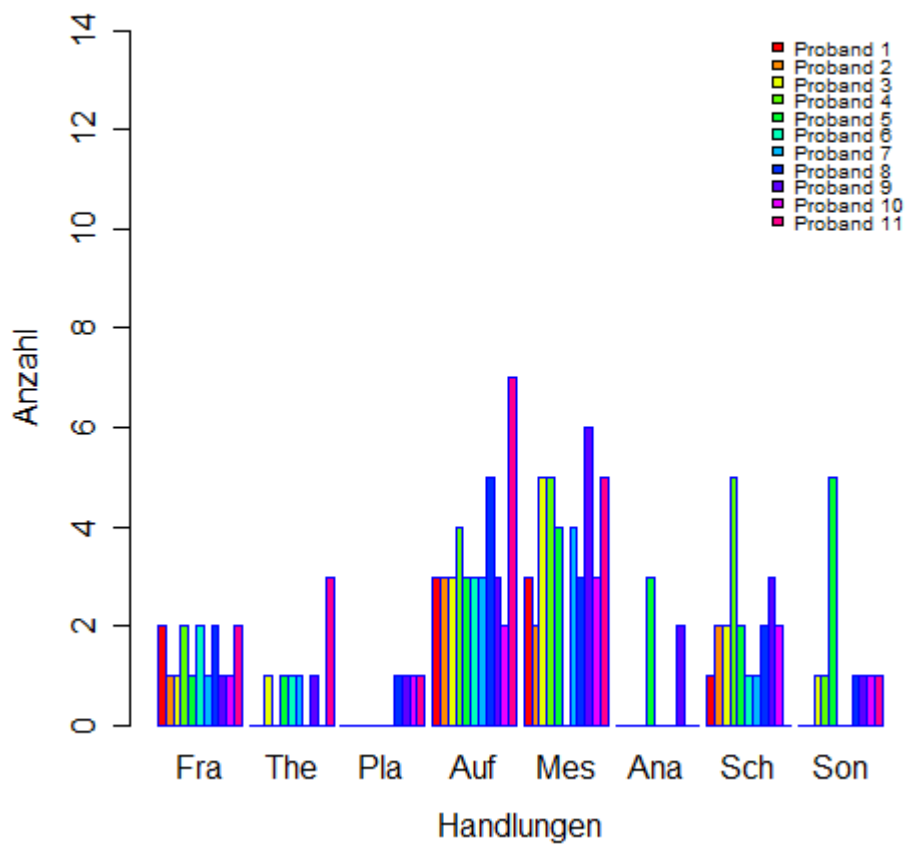
## Prozessorientierte Auswertung - Aufgabe 4



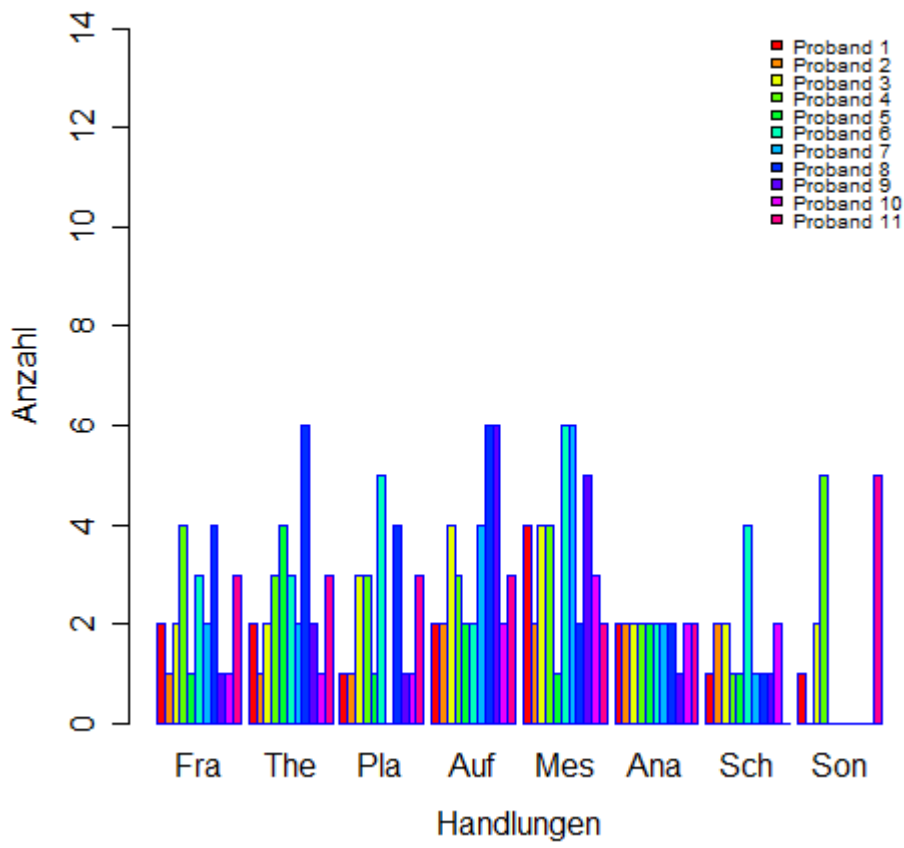
## Prozessorientierte Auswertung - Aufgabe 5



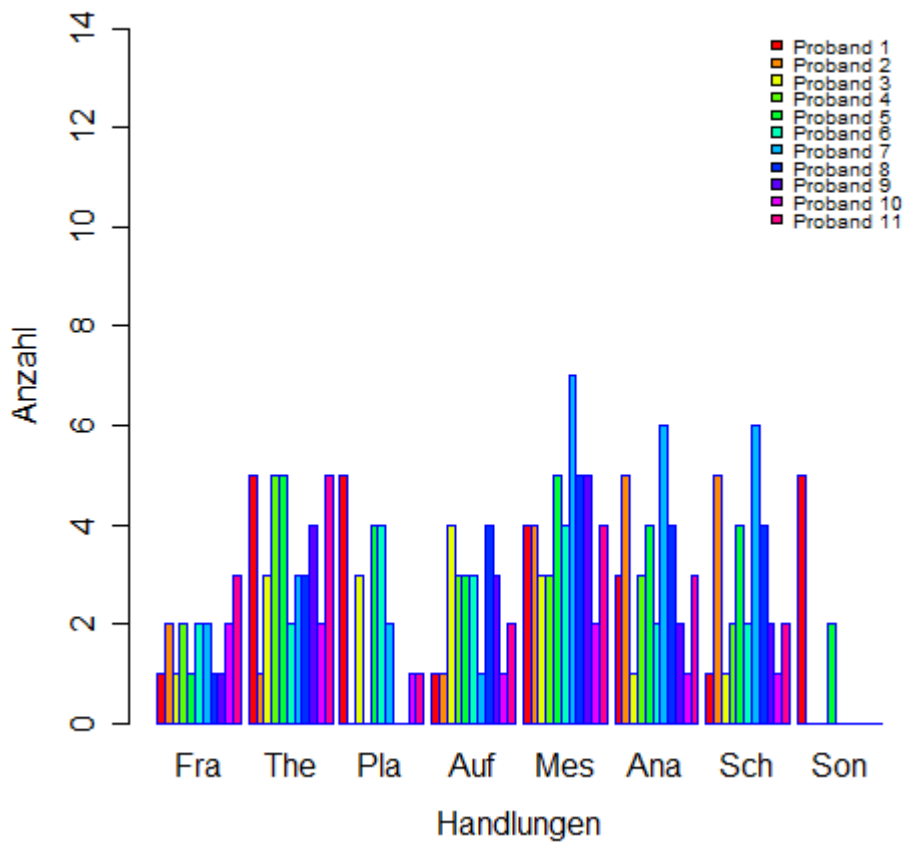
## Prozessorientierte Auswertung - Aufgabe 6



## Referenz Auswertung - Aufgabe 1

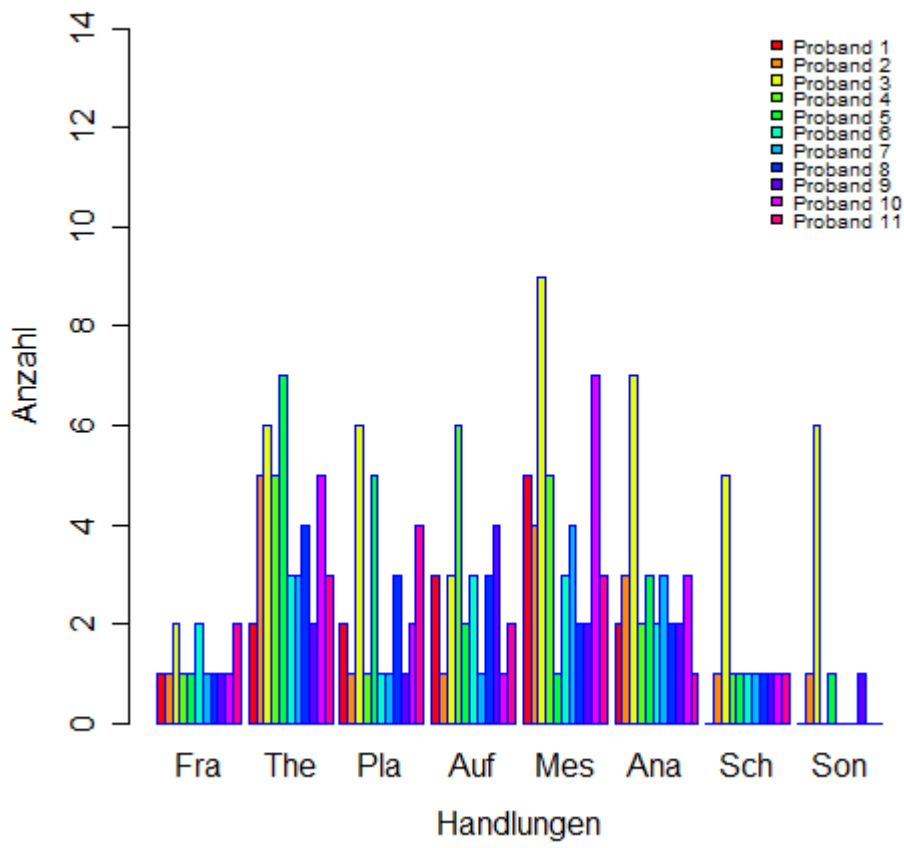


## Referenz Auswertung - Aufgabe 2

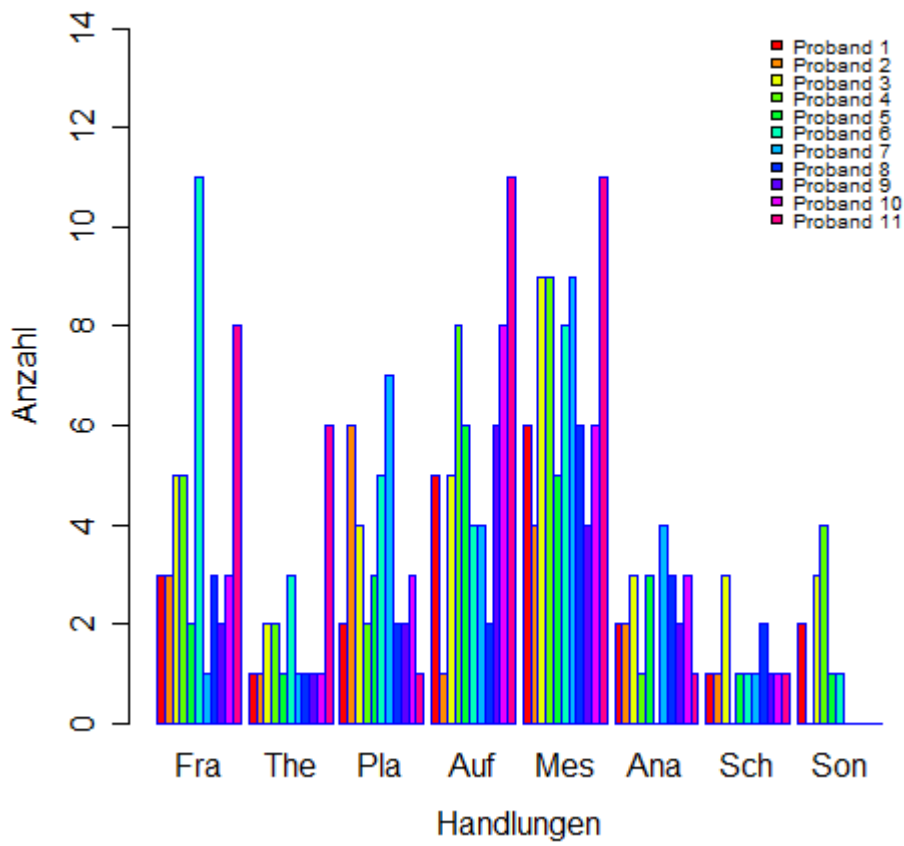




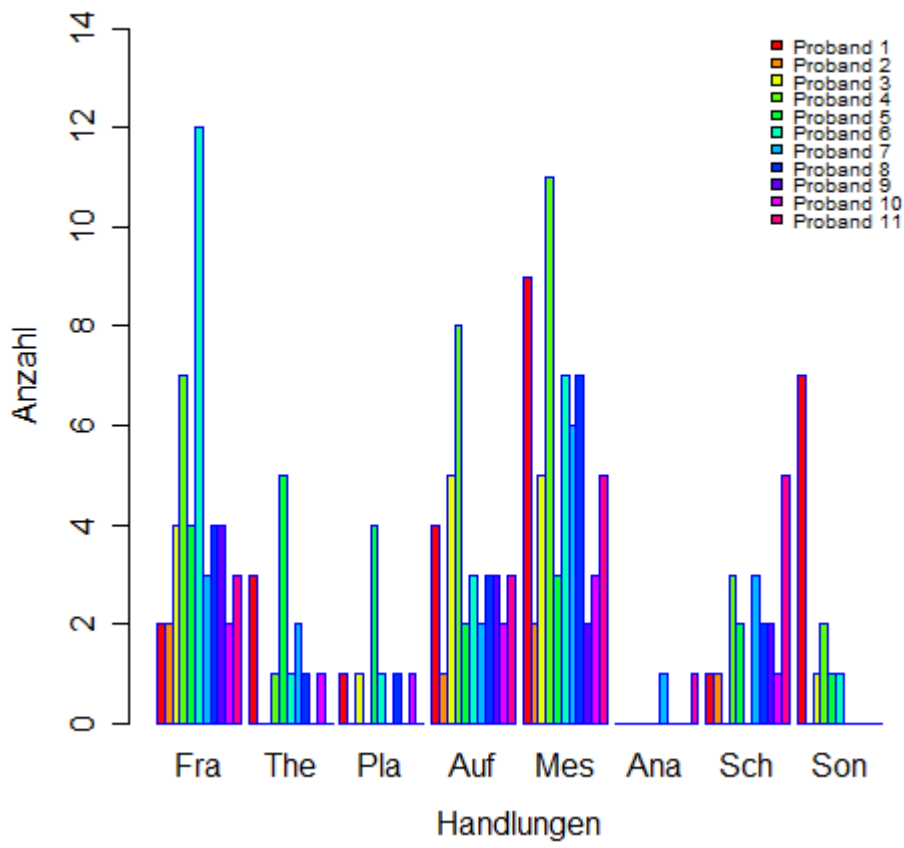
### Referenz Auswertung - Aufgabe 3



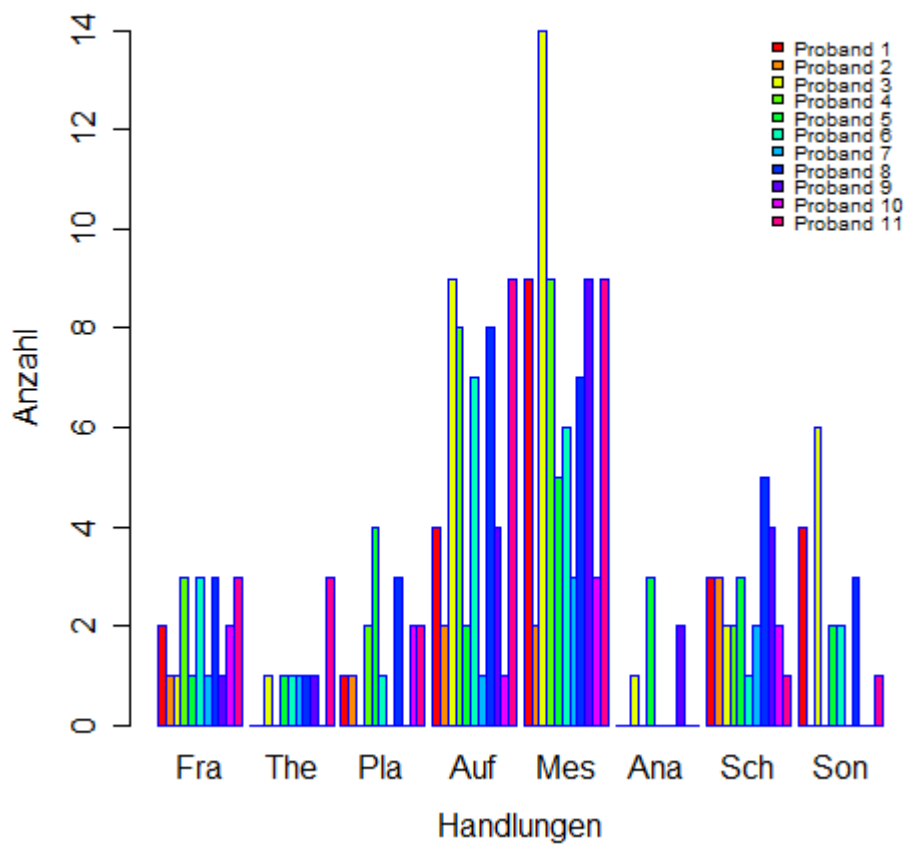
## Referenz Auswertung - Aufgabe 4



## Referenz Auswertung - Aufgabe 5



## Referenz Auswertung - Aufgabe 6



## **A.8 Forschungsdefizite**

### **Verifizierung der externen Validierung**

Bei der divergenten Validierung gab es Probleme mit den Testinstrumenten zum Fachwissen und zum Strategiewissen. Entsprechend ist es notwendig weitere Evidenz zur Abgrenzung dieser Konstrukte zu sammeln. Gleichzeitig wurde bei der konvergenten Validierung nur ein Verfahren zur Bewertung im Praktikum angewendet. Es gibt aber eine Reihe verschiedener Ansätze um die Leistung von Studierenden im Praktikum zu bewerten. Entsprechend wäre eine konvergente Validierung zu den verschiedenen Ansätzen der Bewertung in Praktika wünschenswert, um die gesammelten Ergebnisse zu verifizieren und weiter generalisieren zu können.

### **Weiterentwicklung der Strukturiertheit**

Es gibt eine Menge Hinweise darauf, dass der Aspekt der Strukturiertheit ein von den anderen Aspekten unabhängiger Bestandteil der Experimentierkompetenz ist. Im theoretischen Hintergrund wurde festgestellt:

*Reflektiertes Handeln, Selbstreflektion und konzeptuelles Wissen lassen sich als nach diesen Beschreibungen als Gütekriterium zur Experimentierkompetenz bezeichnen. Aber woran macht man fest, ob ein Proband sein Handeln reflektiert und evaluiert oder woran lässt sich ein mögliches konzeptuelles Wissen erkennen?*(vgl. Kapitel II)

Allerdings gab es kaum konkrete Ansätze um diese Frage ausreichend zu klären. Zwar wurde der Versuch unternommen, Kriterien zu finden (vgl. II.4.2.1 – II.4.2.5) und zu operationalisieren (vgl. V), allerdings zeigen die Ergebnisse von Studie I (vgl. VI.2) und Studie II (vgl. VI.4) dass dies nicht ausreichend gelungen ist. Um dieses Problem zu lösen, wurden drei mögliche Ansätze gefunden. Zum einen könnte eine Weiterentwicklung des Auswerteverfahrens wie in Kapitel VI.4.5 vorgeschlagen, die Probleme der Strukturiertheit möglicherweise lösen. Zweitens kann eine qualitative Analyse der Intention die zu Handlungsübergängen führt zu einem Erkenntnisgewinn bei der Strukturiertheit führen. Die dritte Alternative wäre der Einsatz von Testverfahren zum Umgang mit Fehlern und zu Kontrollverfahren, um Zusammenhänge oder bisherige Probleme mit der Strukturiertheit genauer zu identifizieren.

### **Weiterentwicklung und Verifizierung des Auswerteverfahrens**

In Kapitel VI.4 wurde das Problem des Auswerteverfahrens durch individuelle Bewertungseinheiten beschrieben. Als ein möglicher Lösungsansatz wurde eine konzeptuelle Weiterentwicklung des Auswerteverfahrens durchgeführt. Auf diese Weise entstand das Konzept einer Laborheft-App (siehe Kapitel VI.4.5). Neben der Forschung könnte diese App auch in Praktika zur Datenaufnahme, Gliederung und Protokollierung genutzt werden.

Da die Entwicklung des Auswerteverfahrens die identifizierten Probleme beheben soll, wird vermutet, dass sich kognitiv valide Daten damit erfassen lassen. Allerdings müsste eine Vergleichsstudie mit einer Referenzauswertung (vgl. Studie II) durchgeführt werden, um diese Annahme zu verifizieren.

Eine Umsetzung der Laborheft-App inklusiver einer Analyse zur Beseitigung der beschriebenen Probleme (vgl. Kapitel VII.4) wäre eine schlüssige Möglichkeit zur Verbesserung des Testinstruments. Eine Untersuchung zum Einsatz der Laborheft-App in Praktika könnte zu ersten Verbesserungen der Ausbildung im Bereich des Experimentierens führen.

### **Übertragbarkeit von Experimentierkompetenz auf Fachinhaltsebene**

Im Kapitel III.2.3.5 wurde beschrieben, dass bisher keine Evidenz vorliegt ob oder inwieweit Experimentierkompetenz auf andere Inhaltsbereiche oder Naturwissenschaften übertragbar ist. Mithilfe der Beschreibungen zur Testentwicklung könnte ein äquivalentes Testinstrument für einen anderen Inhaltsbereich z.B. Elektrizitätslehre oder Mechanik bzw. für die Biologie oder Chemie entwickelt werden. In einer Studie, in der Probanden Aufgaben aus verschiedenen Inhaltsbereichen bzw. Naturwissenschaften bearbeiten, könnte Evidenz zur Übertragbarkeit generiert werden. Dies würde einen wichtigen Beitrag zur Entwicklung von Lehrangeboten liefern.

### **Übertragbarkeit von Experimentierkompetenz auf Populationsebene**

Bisher wurden nur Erkenntnisse von Studierenden an der CAU zu Kiel gesammelt. Es wurde zwar die Vermutung formuliert, dass die Ergebnisse wegen der Vergleichbarkeit der Physik Studiengänge auch auf Studierende anderer Universitäten übertragbar ist, Evidenz gibt es dafür allerdings nicht. Entsprechend steht eine Studie zur Übertragbarkeit auf andere Populationen (vgl. III.2.3.5) noch aus.