

**DIE MODELLIERUNG
DES WISSENSCHAFTLICHEN DENKENS
IM GRUNDSCHULALTER**

**ZUSAMMENHÄNGE ZU KOGNITIVEN FÄHIGKEITEN
UND MOTIVATIONALEN ORIENTIERUNGEN**

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Philosophie an der Ludwig-Maximilians-Universität
München

vorgelegt von
Daniela Mayer

München, 2011

Erstgutachterin: Prof. Dr. Beate Sodian
Zweitgutachter: Prof. Dr. Frank Fischer
Tag der mündlichen Prüfung: 24.01.2012

Danksagung

An dieser Stelle möchte ich die Gelegenheit nutzen, all denen zu danken, die mich während meiner Promotionszeit unterstützt haben.

Besonders danken möchte ich meiner Doktormutter Prof. Dr. Beate Sodian sowie Prof. Dr. Susanne Koerber und Prof. Dr. Knut Schwippert für die fachliche Begleitung und Betreuung der Arbeit sowie die wertvollen Ratschläge, die zum Gelingen der Arbeit beigetragen haben.

Mein weiterer Dank gilt dem Doctoral Training Program des Munich Center of the Learning Sciences, meinen Kolleginnen und Kollegen an der Ludwig-Maximilians-Universität München, besonders Dr. Hannah Perst und Christopher Osterhaus, sowie meiner Projektkollegin Judith Pollmeier an der Universität Münster für die konstruktive Zusammenarbeit und die Zeit, die sie sich für auftauchende Fragen oder Probleme immer wieder genommen haben. Außerdem danke ich den studentischen Mitarbeitern, vor allem Susanne Mikschl und Verena Pehl, für ihre engagierte Unterstützung im Forschungsprojekt Science-P.

Des Weiteren möchte ich mich bei den Lehrkräften und Kindern der Grundschulen Camerloherstraße, Fritz-Lutz-Straße und Wolfratshausen bedanken, die sich zur Teilnahme an der Untersuchung bereit erklärt haben, und ohne deren Mitarbeit die vorliegende Arbeit nicht möglich gewesen wäre.

Schließlich gilt großer Dank meiner Familie und meinen Freunden, die mich durch die Höhen und Tiefen meiner Promotionszeit begleitet haben und immer unterstützend an meiner Seite waren.

Zusammenfassung

Ausgehend von neueren entwicklungspsychologischen Forschungsergebnissen zu Basisfähigkeiten im wissenschaftlichen Denken bereits im Grundschulalter wurde ein Kompetenzmodell mit den Komponenten Wissenschaftsverständnis (Verständnis wissenschaftlicher Konzepte und der Ziele von Wissenschaften, Theorieverständnis) und Methodenkompetenz (Experimentierstrategien, Dateninterpretation) aufgestellt sowie ein gruppentestfähiges Testinstrument im schriftlichen Antwortformat entwickelt. Das Ziel dieser Arbeit war die Modellierung der latenten Struktur der Kompetenz wissenschaftliches Denken, die Beschreibung von Entwicklungsveränderungen im Grundschulalter sowie die Untersuchung von Zusammenhängen zwischen wissenschaftlichem Denken und kognitiven Fähigkeiten und motivationalen Orientierungen. In der Studie wurde an 285 Grundschulkindern (72 Zweitklässler, 58 Drittklässler und 155 Viertklässler) wissenschaftliches Denken, Leseverständnis, Intelligenz, Problemlösen, Inhibition, räumliches Denken, formal-operatorische Fähigkeiten, Einstellungen zu Naturwissenschaften, naturwissenschaftsbezogenes Fähigkeitsselbstkonzept sowie Interesse an forschenden Aktivitäten erfasst. Der Vergleich von ein- und mehrdimensionalen Rasch-Modellen ergab eine eindimensionale Struktur der Kompetenz wissenschaftliches Denken mit einer zufrieden stellenden Reliabilität. Zwischen den Teilkompetenzen des wissenschaftlichen Denkens wurden hohe latente Interkorrelationen und bidirektionale Beziehungen ermittelt. Es zeigten sich signifikante Leistungsverbesserungen von der zweiten zur vierten Klassenstufe. In der zweiten Klasse bestanden mehrheitlich Fehlvorstellungen, während in der vierten Klasse schon wissenschaftlich adäquate Vorstellungen vorherrschten. In der dritten Jahrgangsstufe traten markante Entwicklungsveränderungen auf. Innerhalb der Teilkompetenzen des wissenschaftlichen Denkens wurde ein sequenzieller Entwicklungsverlauf von einem Verständnis der Ziele von Wissenschaften über ein Verständnis von Experimentierstrategien und Dateninterpretation hin zu einem Theorieverständnis festgestellt. Darüber hinaus bestanden substantielle Zusammenhänge zu den kognitiven Fähigkeiten, jedoch konnte die Kompetenz wissenschaftliches Denken jeweils als separates Konstrukt abgegrenzt werden. In latenten hierarchischen Regressionsanalysen hatten Leseverständnis und Intelligenz einen signifikanten Einfluss auf die Kompetenz im wissenschaftlichen Denken. Daneben waren Problemlösefähigkeiten von herausragender prädiktiver Bedeutung, in geringerem Ausmaß auch räumliche Fähigkeiten und das naturwissenschaftsbezogene Fähigkeitsselbstkonzept. Besonders im frühen Grundschulalter zeigten sich positive Zusammenhänge zwischen der Kompetenz im wissenschaftlichen Denken und den motivationalen Variablen.

INHALTSVERZEICHNIS

0	Einleitung	8
1	Theoretischer Hintergrund.....	12
1.1	Begriffliche Klärung der Kompetenz wissenschaftliches Denken.....	12
1.2	Entwicklung des wissenschaftlichen Denkens.....	13
1.2.1	Studien zur Methodenkompetenz	14
1.2.1.1	Experimentierstrategien.....	15
1.2.1.2	Dateninterpretation.....	17
1.2.2	Studien zum Wissenschaftsverständnis	19
1.2.2.1	Verständnis wissenschaftlicher Konzepte und der Ziele von Wissenschaften.....	20
1.2.2.2	Theorieverständnis	21
1.2.3	Studien zur selbstgesteuerten Exploration komplexer Mikrowelten	22
1.2.4	LOGIC-Studie zur Entwicklung des wissenschaftlichen Denkens von der Kindheit bis ins Erwachsenenalter.....	24
1.2.5	Studien zur Bedeutung des wissenschaftlichen Denkens für den Erwerb naturwissenschaftlichen Wissens.....	25
1.3	Modellierung der Kompetenz wissenschaftliches Denken im Grundschulalter	27
1.3.1	Kompetenzmodelle	27
1.3.2	Kompetenzmessung.....	31
1.3.3	Abgrenzung und Zusammenhang der Kompetenz wissenschaftliches Denken zu kognitiven Fähigkeiten und motivationalen Orientierungen.....	34
1.3.3.1	Intelligenz	35
1.3.3.2	Leseverständnis.....	36
1.3.3.3	Problemlösen.....	37
1.3.3.4	Inhibition.....	39
1.3.3.5	Räumliches Denken.....	40
1.3.3.6	Formal-operatorische Fähigkeiten.....	41
1.3.3.7	Motivationale Orientierungen.....	42
1.3.4	Rahmenmodell	43
2	Ziele, Fragestellungen und Hypothesen	45
3	Methode.....	48
3.1	Stichprobe	48

3.2	Material.....	49
3.2.1	Fragebogen zur Erfassung des wissenschaftlichen Denkens.....	49
3.2.1.1	Strukturelle Grundlagen der Itementwicklung.....	50
3.2.1.1.1	Kompetenzstruktur.....	50
3.2.1.1.2	Kompetenzniveaus.....	51
3.2.1.1.3	Antwortformate.....	55
3.2.1.2	Aufgaben.....	56
3.2.1.3	Durchführung.....	62
3.2.1.4	Kodierung, fehlende Werte und Skalierung.....	62
3.2.2	Fragebögen zur Erfassung ausgewählter kognitiver Fähigkeiten.....	64
3.2.2.1	Leseverständnis.....	64
3.2.2.2	Intelligenz.....	64
3.2.2.3	Problemlösen.....	65
3.2.2.4	Inhibition.....	66
3.2.2.5	Räumliches Denken.....	67
3.2.2.6	Formal-operatorische Fähigkeiten.....	67
3.2.3	Fragebögen zur Erfassung ausgewählter motivationaler Orientierungen.....	68
3.2.3.1	Naturwissenschaftsbezogene Einstellungen und Fähigkeitsselbstkonzept.....	68
3.2.3.2	Interesse an forschenden Aktivitäten.....	68
3.3	Ablauf der Studie.....	69
3.4	Statistische Verfahren.....	71
3.4.1	Item-Response-Theorie.....	71
3.4.2	Prüfung von Gruppenunterschieden.....	79
4	Ergebnisse.....	82
4.1	Skalierung der Kompetenz wissenschaftliches Denken.....	82
4.1.1	Modellprüfung.....	82
4.1.1.1	Modellvergleiche.....	82
4.1.1.2	Latente Korrelationen.....	83
4.1.1.3	Reliabilität.....	86
4.1.2	Statistische Kennwerte des eindimensionalen Rasch-Modells.....	87
4.1.2.1	Wright map.....	87
4.1.2.2	Itemkennwerte.....	89
4.1.3	Latente Klassenanalysen.....	90
4.2	Entwicklung der Kompetenz wissenschaftliches Denken im Grundschulalter.....	93
4.2.1	Deskriptive Statistiken.....	93
4.2.2	Kompetenzwerte.....	98

4.2.2.1 Kompetenz wissenschaftliches Denken – Gesamttest.....	98
4.2.2.2 Kompetenz wissenschaftliches Denken – Subskalen Komponenten	100
4.2.2.3 Kompetenz wissenschaftliches Denken – Subskalen Teilkompetenzen....	102
4.2.3 Sequenzieller Entwicklungsverlauf.....	107
4.2.4 Wechselseitige Beziehungen zwischen den Teilkompetenzen des wissenschaftlichen Denkens	111
4.3 Abgrenzung und Zusammenhang des wissenschaftlichen Denkens zu ausgewählten kognitiven Fähigkeiten.....	113
4.3.1 Leseverständnis.....	113
4.3.2 Intelligenz	117
4.3.3 Problemlösen.....	121
4.3.4 Inhibition	125
4.3.5 Räumliches Denken.....	129
4.3.6 Formal-operatorische Fähigkeiten.....	133
4.4 Zusammenhang zwischen wissenschaftlichem Denken und ausgewählten motivationalen Orientierungen	138
4.4.1 Naturwissenschaftsbezogene Einstellungen	138
4.4.2 Naturwissenschaftsbezogenes Fähigkeitsselbstkonzept.....	141
4.4.3 Interesse an forschenden Aktivitäten	144
4.5 Latente hierarchische Regressionsanalysen zur Vorhersage der Kompetenz im wissenschaftlichen Denken	147
4.5.1 Gesamtstichprobe	148
4.5.2 Frühes Grundschulalter.....	151
4.5.3 Spätes Grundschulalter	154
5 Diskussion	157
6 Ausblick.....	180
7 Literaturverzeichnis.....	182
8 Tabellen- und Abbildungsverzeichnis.....	196
9 Anhang	201

0 Einleitung

Naturwissenschaften und Technik begegnen uns in zahlreichen Alltagssituationen und sind ein wichtiger Teil unserer Kultur und Gesellschaft. Naturwissenschaftliche Grundbildung gilt als wichtige Voraussetzung, um an einer von Naturwissenschaften geprägten Gesellschaft teilhaben und sich mit einer sich ständig verändernden Welt auseinandersetzen zu können (Prenzel, Rost, Senkbeil, Häußler & Klopp, 2001). Sie ist im Zuge der Ergebnisse der internationalen Schulleistungsstudien wie PISA¹, TIMSS² und IGLU-E³ in den Blickpunkt der Bildungsdiskussion in Deutschland gerückt.

Werden die Übereinstimmungen unterschiedlicher nationaler (z. B. Prenzel et al., 2001) und internationaler Konzeptionen (z. B. American Association for the Advancement of Science (AAAS), 1993; Bybee, 1997, 2002) zusammengefasst, beinhaltet naturwissenschaftliche Grundbildung bzw. scientific literacy folgende Wissensbereiche (Duit, Häußler & Prenzel, 2001, S. 172):

- „Naturwissenschaftliche Begriffe und Prinzipien (Konzepte)
- Naturwissenschaftliche Untersuchungsmethoden und Denkweisen (Prozesse)
- Vorstellungen zur Natur der Naturwissenschaften [Wissenschaftsverständnis]
- Vorstellungen und Einstellungen zur Relevanz der Naturwissenschaften in Gesellschaft und Technik“.

Entsprechend dieser Definition ist wissenschaftliches Denken, das naturwissenschaftliche Arbeitsweisen und Untersuchungsmethoden sowie ein Verständnis der Natur von Wissenschaften umfasst, ein wichtiger Aspekt der naturwissenschaftlichen Grundbildung. Die aktuelle Bildungsdiskussion über scientific literacy hat Auswirkungen sowohl auf die Gestaltung von Lehrplänen und Unterricht als auch auf die Erfassung der Kompetenzen von Schülern⁴ (Duit et al., 2001). Nicht nur im Sekundarschul- sondern auch im Primarschulbereich sind konzeptuelle und prozedurale scientific literacy zum Bestandteil von

¹ Programm for International Student Assessment (Baumert et al., 2001)

² Third International Mathematics and Science Study (Baumert et al., 1997; TIMSS 2007 / Bos et al., 2008)

³ Erweiterungsstudie zu IGLU: Erhebung des mathematischen und naturwissenschaftlichen Verständnisses von Grundschulern (Bos et al., 2003)

⁴ In der vorliegenden Arbeit wird hier und im Folgenden die männliche Form verwendet, gemeint sind sowohl Schülerinnen als auch Schüler.

Bildungsstandards geworden (z. B. Bos et al., 2003, Gesellschaft für die Didaktik des Sachunterrichts [GDSU], 2002; Bybee, 1997). Während im naturwissenschaftlichen Unterricht in Deutschland bislang fast ausschließlich die Vermittlung naturwissenschaftlichen Inhaltswissens in den Domänen Physik, Chemie und Biologie vorherrscht, gibt es für domänenübergreifende Kompetenzen, wie das Verständnis der experimentellen Methode, die Evaluation von Daten und das metabegriffliche Verständnis von Wissenschaften, in der Grundschule keine etablierten Curricula. Gerade diese epistemologischen, metakognitiven Aspekte, d. h. die explizite Reflexion über die Natur der Naturwissenschaften, die Rolle von Theorien und Experimenten im wissenschaftlichen Erkenntnisprozess sowie die Differenzierung und Koordination von Theorie und Evidenz, gelten als wesentlich sowohl für den Prozess der Erkenntnisgewinnung als auch für den Erwerb naturwissenschaftlichen Wissens (Kuhn, 2002; Sodian, Jonen, Thoermer & Kircher, 2006, Grygier, 2008).

Die Forderung nach einem frühen Lernen im Bereich der Naturwissenschaften wird gestützt durch die Befunde aktueller Schulleistungsstudien (z. B. Bos et al., 2003, 2008), durch die Ergebnisse neuerer entwicklungspsychologischer Forschung zu Basisfähigkeiten im wissenschaftlichen Denken bereits im Grundschulalter (siehe Zimmerman, 2007, für einen Überblick) sowie durch die Belege fachdidaktischer Forschung, dass auch schon Grundschulkindern durch konstruktivistisch orientierten Unterricht ein elaboriertes Verständnis naturwissenschaftlicher Konzepte und Vorgehensweisen erreichen konnten (Möller, Hardy, Jonen, Kleickmann & Blumberg, 2006). In der Grundschule sollen Alltagsvorstellungen über Naturwissenschaften aufgegriffen und gegebenenfalls korrigiert werden mit dem Ziel, eine anschlussfähige, naturwissenschaftliche Grundbildung aufzubauen (Grygier, 2008).

Bisher fehlen allerdings noch breit angelegte Studien, die die frühe Entwicklung des wissenschaftlichen Denkens sowohl im Wissenschaftsverständnis als auch in den Methodenkompetenzen sowie die Beziehung zu naturwissenschaftlichem Inhaltswissen untersuchen. Es ist ein effektives, reliables und valides Testinstrument notwendig, um die Kompetenz von Grundschulern im wissenschaftlichen Denken breit erfassen zu können und darauf aufbauend die Kompetenzentwicklung und deren Bedingungen zu beschreiben sowie Handlungsanweisungen für die pädagogische Praxis geben zu können. Diese Ziele verfolgt das Forschungsprojekt „Science-P Entwicklung naturwissenschaftlicher Kompetenz in der Grundschule“⁵ sowie die vorliegende Arbeit, die im Rahmen dieses Projektes entstand.

⁵ Die vorliegende Studie entstand im Rahmen des Projekts „Science-P – Entwicklung naturwissenschaftlicher Kompetenz in der Grundschule“ und wurde durch die DFG im Schwerpunktprogramm „Kompetenzmodelle zur Erfassung individueller Lernergebnisse und Bilanzierung von Bildungsprozessen“ (SPP 1293) gefördert.

Zielstellung der Arbeit

Die vorliegende Studie leistet einen Beitrag zur Erfassung und Modellierung des wissenschaftlichen Denkens im Grundschulalter sowie zur Untersuchung der frühen Kompetenzentwicklung und deren individuellen Bedingungsfaktoren.

Auf dem Hintergrund des im Rahmen des Forschungsprojektes Science-P entwickelten Testinstrumentes im geschlossenen Antwortformat wurden die psychometrischen Eigenschaften der Aufgaben anhand des Rasch-Modells überprüft. Bei der Testentwicklung wurde auf zwei zentrale Kompetenzkomponenten fokussiert, die in je zwei Teilkompetenzen unterschieden wurden: Zum einen das Wissenschaftsverständnis mit dem Verständnis der Konzepte Hypothese, Theorie, Experiment und der Ziele von Wissenschaften sowie dem Verständnis von Theorien als alternative Interpretationsrahmen; zum anderen die Methodenkompetenz mit dem Verständnis von Experimentierstrategien und der Interpretation von Daten.

Das Ziel der vorliegenden Arbeit war es, das Kompetenzstrukturmodell empirisch zu überprüfen und die latente Struktur der Kompetenz wissenschaftliches Denken im Grundschulalter zu modellieren. Darüber hinaus stand im Fokus des Interesses, individuelle Unterschiede im wissenschaftlichen Denken für das Alter von acht bis zehn Jahren sowie frühe Entwicklungsveränderungen und den Entwicklungsverlauf in den einzelnen Teilkompetenzen zu beschreiben. Eine weitere Zielstellung war es, den Zusammenhang zwischen wissenschaftlichem Denken und potentiell zugrunde liegenden, kognitiven Fähigkeiten zu untersuchen. Hier wurden basierend auf einschlägiger Forschungsliteratur und Ergebnissen aus dem Sekundarschulalter die kognitiven Fähigkeiten Leseverständnis, Intelligenz, formal-operatorische Fähigkeiten, Problemlösen, inhibitorische Kontrolle und räumliches Denken ausgewählt, um den Zusammenhang zum wissenschaftlichen Denken zu beschreiben und die einzelnen Kompetenzen zu differenzieren. Neben kognitiven Bedingungen wurde die Bedeutung wissenschaftsbezogener, motivationaler Orientierungen für die Kompetenzentwicklung im wissenschaftlichen Denken betrachtet und die Beziehungen zu Einstellungen zu Naturwissenschaften, dem naturwissenschaftsbezogenem Fähigkeits-selbstkonzept sowie dem Interesse an forschenden Aktivitäten untersucht.

Gliederung der Arbeit

Die vorliegende Arbeit ist folgendermaßen gegliedert. Der theoretische Hintergrund gibt zunächst einen allgemeinen Überblick über die Definition des Konstrukts wissenschaftliches Denken und den Prozess des wissenschaftlichen Erkenntnisgewinns. Auf dem Hintergrund neuerer entwicklungspsychologischer Forschung wird die Entwicklung des wissenschaftlichen Denkens in den einzelnen Teilkompetenzen beschrieben. Dies beinhaltet eine Darstellung früher Kompetenzen als auch von Defiziten im Grundschulalter und der weiteren Kompetenzentwicklung bis ins Jugend- und Erwachsenenalter. Es folgt die Betrachtung der Kompetenz wissenschaftliches Denken unter dem Aspekt der Kompetenzmodellierung, der Abgrenzung und Beziehung zu kognitiven Fähigkeiten sowie des Zusammenhangs zu motivationalen Variablen.

An den theoretischen Überblick schließt sich die Beschreibung der empirischen Studie an. Das entwickelte Testinstrument zur Erfassung wissenschaftlichen Denkens von Grundschulkindern und der Einsatz in Verbindung mit weiteren kognitiven und motivationalen Maßen werden dargestellt, um die Fragestellungen der Kompetenzmodellierung und -entwicklung zu untersuchen. Schließlich folgt die Darstellung der empirischen Ergebnisse im Hinblick auf die psychometrischen Eigenschaften des Testinstruments, die Kompetenzmodellierung, die Kompetenzentwicklung und die Zusammenhänge zu den kognitiven und motivationalen Variablen. Die Befunde werden hinsichtlich der aufgestellten Hypothesen und aktueller Forschung diskutiert, und es werden Limitationen der vorliegenden Studie aufgezeigt. Abschließend werden im Ausblick, basierend auf den berichteten Fähigkeiten der Grundschulkinde, Möglichkeiten der Förderung wissenschaftlichen Denkens und Implikationen für die pädagogische Praxis in der Grundschule erörtert.

1 Theoretischer Hintergrund

1.1 Begriffliche Klärung der Kompetenz wissenschaftliches Denken

Wissenschaftliches Denken kann nach Deanna Kuhn (2002) definiert werden als die intentionale Suche nach Erkenntnis und erfordert die bewusste Unterscheidung und Koordination von Theorie und Evidenz. Der Prozess der wissenschaftlichen Erkenntnissuche beinhaltet Fähigkeiten zur Bildung, Prüfung und Revision von Theorien und Hypothesen über beliebige Phänomenbereiche sowie die Reflexion über diesen Prozess der Wissenskonstruktion (Wilkening & Sodian, 2005; Zimmerman, 2007). Die Anwendung von Methoden und Prinzipien des wissenschaftlichen Erkenntnisprozesses auf Denk- und Problemlösesituationen ist generell von großer Bedeutung für den Erkenntnisgewinn und Wissenserwerb (Zimmerman, 2007). Daher stehen wissenschaftliches Denken und andere Formen des Denkens, wie Problemlösen oder Schlussfolgern, in Zusammenhang (Kuhn 2002; Kuhn & Franklin, 2006).

Der wissenschaftliche Erkenntnisprozess umfasst vier grundlegende Phasen bzw. Komponenten (Kuhn, 2002): Basierend auf Theorien und aufgestellten Hypothesen über das interessierende Phänomen werden Experimente durchgeführt, gewonnene Daten interpretiert und Schlussfolgerungen gezogen mit dem Ziel, Theorien zu bilden und/oder zu revidieren (vgl. Abbildung 1.1).

Zusätzlich ist das metakonzeptuelle Verständnis von Wissenschaften ein weiterer entscheidender Aspekt, der sich im wissenschaftlichen Denken entwickeln muss (Kuhn, 2002). Das Wissenschaftsverständnis, im Sinne eines epistemologischen Verständnisses über die Natur von Wissenschaften und über die Natur wissenschaftlichen Wissens, umfasst den gesamten Zyklus des wissenschaftlichen Erkenntnisprozesses und ermöglicht die Reflexion über diesen Prozess. Wissenschaftsverständnis schließt Einsicht in erkenntnistheoretische, wissenschaftstheoretische und wissenschaftsethische Aspekte von Naturwissenschaften ein (Sodian et al., 2006).

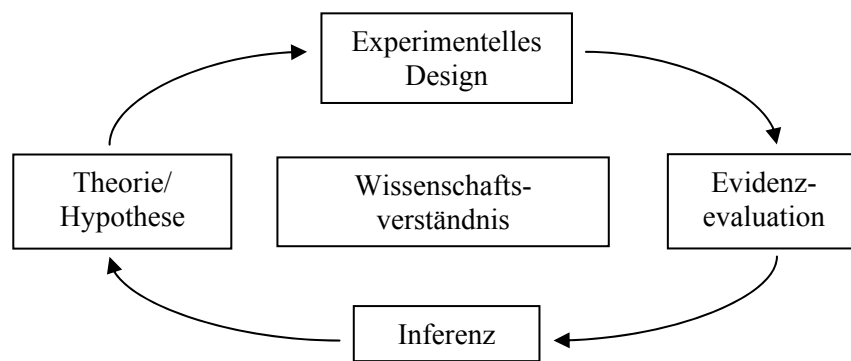


Abbildung 1.1. *Zyklus des wissenschaftlichen Erkenntnisprozesses*

1.2 Entwicklung des wissenschaftlichen Denkens

Die traditionelle auf der Arbeit von Inhelder und Piaget (1958) basierende entwicklungspsychologische Forschung sah wissenschaftliches Denken an die Entwicklung formaler Operationen gebunden und damit als Entwicklungsphänomen des Jugendalters an. Piagets Theorie der Denkentwicklung stellt die kognitive Entwicklung als bereichsübergreifend und diskontinuierlich im Sinne eines Durchlaufens von qualitativ unterschiedlichen, universell vorkommenden Stadien dar. Inhelder und Piaget beschrieben das konkret-operatorische Denken des Grundschulkindes defizitär beispielsweise im kausalen, perspektivischen und mehrdimensionalen Denken. Sie prüften in einer Serie von Aufgaben mit verschiedenen Inhaltsbereichen die Fähigkeit von Kindern, Jugendlichen und Erwachsenen, Hypothesen über kausale Zusammenhänge zu entwickeln und zu prüfen sowie die Ergebnisse der Tests zu interpretieren. Im Gegensatz zu Jugendlichen stellten Kinder nicht systematisch Versuchspläne auf. Sie manipulierten die relevanten Variablen unsystematisch und zogen voreilige, confirmatorische Schlussfolgerungen aus inkonklusiven Tests. Die Kinder schienen eher an der Produktion von Effekten interessiert zu sein, als an der Aufklärung der Ursachen für das Zustandekommen dieser Effekte. Piagets Theorie ging daher von der Annahme aus, dass sich in einer universellen Entwicklung erst mit dem Erreichen des Stadiums der formalen Operationen (12 Jahre) eine kognitive Struktur bzw. ein kohärentes Set an kognitiven Fähigkeiten ausbildet, das ab dem Jugendalter ermöglicht, Problemstellungen systematisch und analytisch zu erkunden und so rationale Erkenntnis zu gewinnen.

Neuere entwicklungspsychologische Ansätze betrachten die Denkentwicklung als variabler, kontinuierlicher und bereichsspezifischer: Kognitive Entwicklung wird als Aufbau bereichsspezifischer Wissenssysteme angesehen, aber es werden auch bereichsübergreifender Veränderungen im Denken angenommen (Sodian, 1998; Sodian, 2005). Die neuere Forschung zum wissenschaftlichen Denken führt die Schwierigkeiten von Kindern im wissenschaftlichen Denken weniger auf strukturelle Einschränkungen sensu Piaget zurück, sondern auf Defizite im (meta)begrifflichem Verständnis im Sinne der Unterscheidung der Konzepte Theorie und Hypothese einerseits und empirische Evidenz andererseits (z. B. Kuhn, Amsel & O'Loughlin, 1988; Kuhn & Pearsall, 2000) oder auf eine Unfähigkeit, die Problemlösesituation vollständig repräsentieren zu können (Dunbar & Klahr, 1989; Klahr, 2000).

Zimmerman (2007) folgerte aus einer Zusammenschau neuerer, entwicklungspsychologischer Forschung zum wissenschaftlichen Denken, dass Kinder im wissenschaftlichen Denken kompetenter sind als zunächst angenommen, und ihre Fähigkeiten unterschätzt wurden, wohingegen die Fähigkeiten von Erwachsenen überschätzt wurden. Bereits im Vorschulalter beginnt die Entwicklung einer Kompetenz im wissenschaftlichen Denken in Form eines impliziten Verständnisses und dauert bis ins Jugend- und Erwachsenenalter an (Sodian & Bullock, 2008).

Kompetenzen von Kindern im wissenschaftlichen Denken konnten vor allen dann gezeigt werden, wenn statt multivariaten Aufgaben vereinfachte, kindgerechte (kontextualisierte) Aufgabenstellungen verwendet wurden, die weniger Anforderungen an die Informationsverarbeitungsprozesse der Kinder stellten (z. B. Sodian, Zaitchik & Carey, 1991). Auch die Art der Aufgabenstellung bzw. das Antwortformat spielte eine entscheidende Rolle: In Aufgaben, in denen beispielsweise Experimente zur Auswahl vorgegeben wurden, schnitten Kinder besser ab, als wenn es die Aufgabe war, ein Experiment selbst zu produzieren (Bullock & Ziegler, 1999).

1.2.1 Studien zur Methodenkompetenz

Zentrale Aspekte der wissenschaftlichen Methodenkompetenz sind Methoden der Datengewinnung (Verständnis von Experimentierstrategien) und Methoden der Datenanalyse (Fähigkeit zur Interpretation von Daten) mit dem Ziel eine Hypothese zu prüfen. Ausgewählte Studien zu Fähigkeiten von Grundschulkindern im Bereich von Experimentierstra-

tegien und der Dateninterpretation werden in den beiden nachfolgenden Abschnitten dargestellt.

1.2.1.1 Experimentierstrategien

Die im Folgenden zusammengefassten Studien zu Experimentierstrategien untersuchten Fähigkeiten zur Wahl und Produktion eines konklusiven Tests oder eines geeigneten, experimentellen Designs sowie zur Isolation, Manipulation und Kontrolle von Variablen (siehe Zimmerman, 2007, für einen Überblick).

Sodian und Kollegen (1991) untersuchten, ob Erst- und Zweitklässler zwischen Hypothesen und Evidenz unterscheiden können. Es wurde eine Geschichte von zwei Kindern präsentiert, die entweder durch ein Experiment herausfinden wollten, ob eine Maus in ihrem Haus groß oder klein war (find-out-condition: Hypothesenprüfung), oder die Maus füttern wollten (feed-condition: Effektproduktion). Die Kinder sollten in den beiden Versuchsbedingungen jeweils aus zwei Schachteln mit einer großen und einer kleinen Öffnung, in denen jeweils Käse gelegt worden war, wählen und ihre Wahl begründen. Wenn Kinder in der Lage sind, zwischen Hypothesenprüfung und Effektproduktion zu unterscheiden, sollten sie in den beiden Bedingungen unterschiedliche Schachteln wählen. Mehr als die Hälfte der Erstklässler und 86 % der Zweitklässler konnten die beiden Aufgabestellungen unterscheiden und in der find-out-Bedingung zwischen einem inkonklusiven Test (große Öffnung) bzw. einem konklusiven Test (kleine Öffnung) differenzieren sowie ihre Wahl der Schachtel korrekt begründen. In einer zweiten Variante ging es darum, herauszufinden, ob ein Haustier einen guten oder schlechten Geruchssinn hat. Ungefähr ein Viertel der Kinder beider Klassenstufen konnte einen konklusiven Test spontan produzieren. Die Ergebnisse wurden dahingehend interpretiert, dass Kinder schon im Alter von sechs Jahren zwischen einem konklusiven und inkonklusiven Test und dadurch die Prüfung einer einfachen Hypothese durch ein Experiment von der Produktion positiver Effekte unterscheiden können. Somit zeigten schon Grundschulkinder ein beginnendes Verständnis der experimentellen Methode mit dem Ziel des Testens von Hypothesen. Die Ergebnisse widersprechen der Annahme eines Defizits in der Unterscheidung zwischen Hypothesen und Theorien einerseits und empirischer Evidenz andererseits (Kuhn et al., 1988; Kuhn & Pearsall, 2000).

Bullock und Ziegler (1999) überprüften in der LOGIC-Studie (Munich Longitudinal Study on the Genesis of Individual Competencies; Weinert & Schneider, 1999) bei Kin-

dern der dritten bis sechsten Klassenstufe das Verständnis und den Einsatz der Variablenkontrollstrategie als Test zur Prüfung einer Ursache-Wirkung-Beziehung zwischen Variablen. Die sogenannte Variablenkontrollstrategie als domänenübergreifende Fähigkeit besteht darin, die fokale Variablendimension zu manipulieren und gleichzeitig alle anderen Variablendimensionen konstant zu halten. Dadurch können aus dem resultierenden unkonfundierten Experiment valide Schlussfolgerungen abgeleitet werden. In verschiedenen Kontexten präsentierten Bullock und Ziegler (1999) den Kindern eine Geschichte über einen Protagonisten, der ein Produkt herstellen und dabei testen wollte, ob eine bestimmte Variablendimension einen positiven Effekt hat. Beispielsweise ging es um einen Flugzeugkonstrukteur, der Flugzeuge bauen wollte, die möglichst wenig Treibstoff verbrauchen (sogenannte Flugzeugaufgabe). Drei Variablen wurden für den Treibstoffverbrauch als relevant erachtet und präsentiert: die Form der Nase (spitz oder rund), die Art der Flügel (einfach oder Doppeldecker) und die Stellung der Höhenruder (oben oder unten). Der Flugzeugingenieur wollte dann seine Hypothese, dass die Stellung der Höhenruder (fokale Variable) einen Effekt auf den Treibstoffverbrauch habe, testen. Das Verständnis der Logik des Experimentierens wurde durch zwei Maße erfasst: Zuerst sollten die Kinder spontan einen Test selbst produzieren. Im zweiten Leistungsmaß war es die Aufgabe, aus den acht möglichen Variablenkombinationen, präsentiert in Form von Bildkarten, einen kontrollierten Test, in dem die fokale Variable (Position der Höhenruder) manipuliert und die anderen Variablendimensionen (Form der Nase, Art der Flügel) konstant gehalten wurden, auszuwählen. Die Dritt- und Viertklässler produzierten mehrheitlich einen kontrastiven Test, indem sie die fokale Variable isolierten und manipulierten, allerdings nicht darauf achteten, die anderen Variablen zu kontrollieren. Dieses Ergebnis zeigte, dass schon Kinder ab der dritten Klassenstufe das Testen von Hypothesen durch einen kritischen Bedingungsvergleich verstanden und nicht nur positive Effekte produzierten. Allerdings zeigte sich erst ab der fünften Klassenstufe eine spontane Produktion eines kontrollierten Tests. Im Gegensatz dazu erkannten in der Wahlaufgabe 30 % der Drittklässler, 60 % der Viert- und Fünftklässler und 80 % der Sechstklässler ein kontrolliertes Experiment und konnten mehrheitlich diese Wahl auch korrekt begründen. Dieser Befund wurde dahingehend interpretiert, dass schon Achtjährige in Wahlaufgaben ein beginnendes (implizites) Verständnis der experimentellen Methode zeigten: Sie setzten zwar die Variablenkontrollstrategie zur Produktion eines kontrollierten Tests noch nicht spontan ein, aber konnten ein kontrolliertes Experiment wählen und die Wahl meist auch korrekt begründen (beginnendes explizites Verständnis). Nur wenige Grundschul Kinder produzierten einen kontrollierten Test. Der

spontane Einsatz der Variablenkontrollstrategie verbesserte sich erst ab der fünften Klasse und die Leistung der Sechstklässler entsprach der von Erwachsenen (Bullock, Sodian & Koerber, 2009).

1.2.1.2 Dateninterpretation

Neben der Forschung zur Fähigkeit zum systematischen Experimentieren, in der es darum geht, ein experimentelles Design aufzustellen und Daten zu gewinnen, befassen sich viele Studien zum wissenschaftlichen Denken mit dem Aspekt der Evidenzevaluation. Dabei sollen Daten interpretiert werden und aus der Evidenz Schlussfolgerungen gezogen werden. Dafür ist es notwendig, zwischen Theorien/Hypothesen und Evidenz zu unterscheiden und Theorie und Evidenz zu koordinieren und systematisch aufeinander zu beziehen (Kuhn & Pearsall, 2000; Kuhn, 2002).

Kuhn und Kollegen (1988) untersuchten in einer Reihe von Studien zur Evaluation von Kovariationsevidenz die Fähigkeit zur Koordination von Theorie und Evidenz. Den Kindern, Jugendlichen und Erwachsenen wurden multivariate Kovariationsdaten präsentiert, die mit den intuitiven Theorien der Probanden über Kausalzusammenhänge konsistent oder inkonsistent waren. Die Probanden sollten die Datenmuster interpretieren: beispielsweise hingen zwei Faktoren (Fruchtsorte, Müslisorte) mit Erkältungen zusammen und zwei Variablen (Kartoffelsorte, Gewürz) nicht. Die Antworten der Probanden wurden als evidenzbasiert (d. h. die Antworten bezogen sich auf die Kovariationsmuster) oder theoriebasiert (d. h. die Antworten bezogen sich auf die Theorien, Überzeugungen und das Vorwissen der Probanden) bewertet. Die Grundschulkinder antworteten eher theoriegeleitet. Sie ließen sich von ihren eigenen Überzeugungen über Zusammenhänge bei der Dateninterpretation leiten statt evidenzbasiert zu antworten. Sie beachteten die Daten nur selektiv. Insbesondere dann, wenn die Daten mit ihren eigenen Theorien inkonsistent waren, wurde die Evidenz ignoriert oder verzerrt. Häufig wurde die eigene Theorie unbewusst an die Evidenz angepasst. Die Fähigkeiten zur Differenzierung und Koordination von Theorie und Evidenz und zur Inhibition der eigenen Überzeugungen bei der Interpretation von Daten zeigten Entwicklungsfortschritte von der Kindheit bis ins Jugend- und Erwachsenenalter. Kuhn und Kollegen diskutierten die Befunde dahingehend, dass die Probanden ihre Theorien nicht zum Gegenstand der bewussten Reflexion machten und nicht zwischen Theorien

und Evidenz als distinkte, epistemologische Repräsentationen unterscheiden konnten (Kuhn, 2002; Kuhn & Franklin, 2006).

Ruffman, Perner, Olson und Doherty (1993) verwendeten bei Vier- bis Siebenjährigen statt multivariaten Aufgaben weniger komplexe Aufgabenstellungen mit einer geringeren Anzahl an Faktoren. Bereits Sechsjährige interpretierten einfache Kovariationsdaten korrekt, wenn nur ein Kausalfaktor (rotes oder grünes Essen) mit einem Ergebnis (gute oder schlechte Zähne) perfekt oder teilweise kovarierte. Die meisten Kinder konnten basierend auf der Evidenz die Kausalannahme einer Geschichtenfigur vorhersagen.

Koerber, Sodian, Thoermer und Nett (2005) replizierten die Ergebnisse von Ruffman et al. (1993) für Vierjährige. Die Kinder konnten auch verstehen, dass eine Geschichtenfigur eine ursprüngliche Kausalhypothese aufgrund von Gegenevidenz ändern kann. Allerdings hatten auch Sechsjährige Schwierigkeiten bei Interpretation von Daten, die keinen Zusammenhang zwischen der angenommenen Ursache und der Ergebnisvariable zeigten (Non-Kovariationsdaten). Die Fähigkeit zur Evaluation von Kovariationsdaten war von vorherrschenden Kausalannahmen beeinflusst: In einem artifiziellen Kontext (z. B. Marsmenschen), in dem die Kinder keine eigenen Überzeugungen über einen Kausalzusammenhang hatten, erzielten die Kinder bessere Leistungen, als in einem Kontext, in dem die Daten mit den eigenen subjektiven Theorien inkonsistent waren.

Ein weiterer Aspekt der Evidenzevaluation ist die Fähigkeit zur Interpretation von Kovariationsdaten aus Graphen und Diagrammen (Koerber & Sodian, 2009). In der Wissenschaft sind diese graphischen Repräsentationsformen ein wichtiges Werkzeug, um die Art und das Ausmaß der Beziehungen zwischen zwei oder mehreren Variablen darzustellen. Beispielsweise werden Balkendiagramme dafür verwendet, um Quantitäten und deren Beziehung zueinander abzutragen, und Liniendiagramme, um den zeitlichen Verlauf einer oder mehrerer Variablen (z. B. Wachstumsrate) abzubilden. Die Fähigkeit, nicht räumliche Beziehungen (z. B. Quantität und Zeit) auf räumliche Beziehungen und somit auf den Abbildungsmodus von Graphen zu übertragen, zeigten bereits Vierjährige (Koerber & Sodian, 2008). Sechsjährige konnten in Liniendiagrammen die Wachstumsrate auf die Steigung des Graphen und die Quantität auf die Höhe (y-Achse) abbilden (Gattis, 2002). Koerber und Sodian (2009) konnten zeigen, dass Kinder im Alter von vier Jahren Kausalbeziehungen zwischen perfekt kovariierenden Variablen aus Balkendiagrammen ablesen konnten. Fünf- und Sechsjährige konnten auch Daten mit imperfekter Kovariation interpretieren. Koerber (2011) berichtete ebenfalls von einem Grundverständnis der Interpretation von Säulen- und Liniendiagrammen und der Fähigkeit, zwischen relevanten und irrelevanten

ten Merkmalen von Graphen zu differenzieren, mit einem Entwicklungspotential von der zweiten zur vierten Grundschulklasse. Diese Fähigkeit gilt für ein elaboriertes Verständnis graphischer Repräsentationsformen als entscheidend.

Fasst man die dargestellten Studien zusammen, so wird deutlich, dass bereits Vorschulkinder einfache Kovariationsmuster evaluieren und aus Graphen interpretieren konnten (Ruffman et al., 1993; Koerber et al., 2005; Koerber & Sodian, 2009; Koerber 2006). Wenn allerdings die Effekte von mehr als zwei Variablen auf ein Ergebnis berücksichtigt werden mussten, hatten Grundschul Kinder Schwierigkeiten in Aufgaben zur Dateninterpretation (Kuhn et al., 1988). Außerdem traten im Vor- und Grundschulalter bei der Interpretation von Kovariationsdaten mehr Verzerrungen in Abhängigkeit von Vorwissen bzw. eigenen Überzeugungen zu Kausalzusammenhängen auf als bei Jugendlichen und Erwachsenen (Amsel & Brock, 1996). Kinder hatten vor allem dann Schwierigkeiten bei der Interpretation von Kovariationsdaten, wenn die Daten nicht perfekt kovariierten (imperfekte Kovariationsmuster) und wenn die Daten zusätzlich noch ihren eigenen Überzeugungen widersprachen (Koerber et al., 2005). In derartigen Aufgaben antworteten Kinder häufig eher theoriegeleitet. Sie ließen sich von ihren eigenen Überzeugungen über Zusammenhänge bei der Dateninterpretation leiten statt evidenzbasiert zu antworten. Häufig wurde die eigene Hypothese unbewusst an die Evidenz angepasst.

Kritisch sei hier erwähnt, dass bei Kausalattributionen nicht nur Kovariationsevidenz, sondern insbesondere auch Erklärungen über einen Kausalmechanismus oder Plausibilitätsannahmen eine entscheidende Rolle spielen (Koslowski, 1996).

1.2.2 Studien zum Wissenschaftsverständnis

In den beiden nachfolgenden Abschnitten werden Studien zum Wissenschaftsverständnis beschrieben, wobei wissenschaftstheoretische Aspekte im Fokus des Interesses stehen. Wissenschaftsverständnis beinhaltet demnach das Verständnis der Konzepte Theorie, Hypothese, Experiment und der Ziele von Wissenschaften sowie ein Theorieverständnis auf einer metakognitiven Ebene. Das metabegriffliche Wissenschaftsverständnis ist bisher im Grundschulalter noch nicht systematisch untersucht worden. Der Fokus der Forschung lag fast ausschließlich im Sekundarschulalter.

1.2.2.1 Verständnis wissenschaftlicher Konzepte und der Ziele von Wissenschaften

Carey, Evans, Honda, Jay und Ungerer (1989) untersuchten in einer Interviewstudie das Verständnis epistemologischer Konzepte (Understanding the Nature of Science) von Siebtklässlern. Carey und Kollegen stellten Fragen zu wissenschaftstheoretischen Aspekten (sogenanntes Nature-of-Science-Interview), wie z. B. Worum geht es in Wissenschaften? Was sind Ziele von Wissenschaften? Wie kommen Wissenschaftler zu neuen Erkenntnissen? Was ist eine Hypothese? Was ist eine Theorie? Was ist ein Experiment? Nach dem mehrstufigen Kodierschema befand sich die Mehrheit der Schüler auf dem untersten Verständnisniveau (Level 1a-1b): Sie verstanden Wissenschaft als Aktivität zur Produktion positiver Effekte oder als Sammlung faktischer Information und stellten keine Verbindung zwischen Hypothesen, Experimenten und Evidenzen her. Nur wenige Kinder zeigten ein metakognitives Verständnis von Wissenschaft als Suche nach überprüfbaren Erklärungen (Evidenz) und ein Verständnis der Hypothesenprüfung (Level 2).

Folgestudien (z. B. Thoermer & Sodian, 2002; Carey & Smith, 1993) stellten fest, dass auch naturwissenschaftlich gebildete Erwachsene selten das höchste, epistemologische Niveau (Level 3), im Sinne eines Verständnisses des Hypothese-Evidenz-Bezugs, der forschungsleitenden Funktion von Theorien sowie des zyklischen und kumulativen Charakters von Wissenschaften bei der Suche nach überprüfbaren Erklärungen, erreichten. Lederman (1992) fasste in seinem Reviewartikel über Studien zu Vorstellungen von Schülern und Lehrern über die Natur von Wissenschaften zusammen, dass Schüler der Sekundarstufe und auch Naturwissenschaftslehrer kein adäquates Wissenschaftsverständnis besitzen. Driver, Leach, Millar und Scott (1996) kamen in ihrer Studie an neun-, zwölf- und sechszehnjährigen Schülern zu dem Schluss, dass Schüler mit einem Naturwissenschaftsunterricht, in dem die Natur der Naturwissenschaften und naturwissenschaftliche Denk- und Arbeitsweisen nicht explizit aufgegriffen werden, in der Regel naive Vorstellungen von der Natur von Wissenschaften besitzen und kein Verständnis des zyklischen und tentativen Charakters wissenschaftlichen Wissens erwerben.

Carey et al. (1989) erreichten anhand eines expliziten, wissenschaftstheoretischen Curriculums positive Effekte auf das Wissenschaftsverständnis von Schülern. Auch Smith Maclin, Houghton und Hennessey (2000) berichteten anhand eines Vergleichs von konstruktivistischem versus eher traditionell, induktivistischem Unterrichts über den Zeitraum

von der ersten bis zur sechsten Klasse positive Effekte des langfristigen, wissenschaftstheoretisch reflektierten Curriculums auf das Wissenschaftsverständnis der Schüler.

In Studien zum deklarativen Wissenschaftsverständnis von Viertklässlern fanden Sodian und Kollegen (Sodian, Thoermer, Kircher, Grygier & Günter, 2002; Sodian et al., 2006; Grygier, 2008) mit einer adaptierten Version des Nature-of-Science-Interviews nach Carey et al. (1989), dass die Kinder überwiegend nur das niedrigste Level 1 erreichten und damit ein Verständnis von Wissenschaft als Sammlung faktischer Information oder als Aktivität (Experimente als Prozedur zur Produktion positiver Effekte) in der vierten Klassenstufe vorherrschte. Nach kurzzeitiger Instruktion anhand eines wissenschaftstheoretischen Unterrichts konnte allerdings bereits am Ende des Grundschulalters das Verständnisniveau angehoben werden. Es konnte ein Grundverständnis von Wissenschaft als Suche nach Erklärungen und ein Verständnis von wissenschaftlichem Wissen als Ergebnis der Prüfung von Hypothesen und Theorien vermittelt werden.

1.2.2.2 Theorieverständnis

Ein elaboriertes Theorieverständnis beinhaltet ein Verständnis von Theorien als alternative Interpretationsrahmen bzw. Überzeugungssysteme, die die Bildung von Hypothesen und die Evaluation von Daten leiten und die Denkperspektive über einen Phänomenbereich bestimmen. Es besteht ein Verständnis des Hypothese-Evidenz-Bezugs. Theorien und Evidenzen können differenziert werden.

Bullock und Ziegler (1994) untersuchten in der LOGIC-Studie in einem kindgerechten kontextualisierten Interview (sogenanntes Hexerei-Interview) nach Sodian, Carey, Grosslight und Smith (1992) das Verständnis der Theorieabhängigkeit des wissenschaftlichen Erkenntnisprozesses am Beispiel einer mittelalterlichen Hexerei-Theorie über die Entstehung von Krankheiten. 29 % der Viert- und Fünftklässler verfügten über ein Verständnis von kohärenten Interpretationsrahmen und ihrer leitenden Rolle bei der Interpretation von Daten. Sie zeigten damit ein intuitives Verständnis des Konzeptes Überzeugungssysteme. Die Kinder konnten die Interpretationsperspektive des mittelalterlichen Wissenschaftlers annehmen (Hexen als Ursachen für Krankheiten) und theoriekonsistente Schlussfolgerungen über die Interpretation von Daten ziehen (beispielsweise wurden die von dem modernen Wissenschaftler präsentierten Krankheitserreger Bakterien als Helfer von Hexen uminterpretiert). Die Analysen der Längsschnittdaten der LOGIC-Studie ergaben, dass nur etwa

Zweidrittel der Erwachsenen ein explizites Verständnis alternativer Überzeugungssysteme zeigten (Bullock et al., 2009).

In einer weiteren Studie zum Theorieverständnis bei Erwachsenen mit einem kontextualisierten Interview (sogenanntes Nature-Nurture-Interview) nach Sodian et al. (1992) über einen Interpretationskonflikt eines Anlagetheoretikers und eines Umwelttheoretikers über die Entwicklung von Nervosität bzw. Aggressivität bei Hunden stellten Thoermer und Sodian (2002) fest, dass auch bei Studenten und Doktoranden keine Deckeneffekte im Theorieverständnis auftraten, sondern vielmehr kein ausgeprägtes Verständnis der Theorieabhängigkeit des wissenschaftlichen Erkenntnisprozesses bestand.

1.2.3 Studien zur selbstgesteuerten Exploration komplexer Mikrowelten

In der Forschung zum wissenschaftlichen Denken fokussierten die meisten Studien hauptsächlich nur auf einen Teil des wissenschaftlichen Erkenntnisprozesses. Nur wenige Studien umfassten mehrere oder alle Komponenten des wissenschaftlichen Denkens. In Studien zur selbstgesteuerten Exploration komplexer Mikrowelten durchliefen die Probanden in einstündigen Sitzungen oder über die Dauer von mehreren Wochen hinweg den gesamten Zyklus des wissenschaftlichen Erkenntnisprozesses und erforschten Kausalbeziehungen in multivariablen Spielwelten (z. B. Kuhn, Schauble, & Garcia-Mila, 1992; Kuhn, Garcia-Mila, Zohar & Andersen, 1995; Schauble, 1990, 1996; Dunbar & Klahr, 1989; Klahr, Fay & Dunbar, 1993; s. Zimmerman, 2000 und 2007, für einen Überblick). Beispielsweise konnten in der sogenannten Kanalaufgabe die Tiefe des Kanals und die Eigenschaften von Booten (z. B. Gewicht, Größe etc.) verändert werden, um zu untersuchen, welche Faktoren die Geschwindigkeit der Boote beeinflussen. Dadurch dass die Probanden in diesen mikrogenetischen Studien mehrmals, oft über mehrere Wochen hinweg, an den Aufgaben teilnahmen, konnten die Forscher sowohl die individuelle Entwicklung der Veränderung des Strategieeinsatzes als auch die Entwicklung domänenspezifischen Wissens beobachten (Kuhn, 1995, 2002).

Zimmermann (2000, 2007) fasste die Ergebnisse zu Entwicklungsmustern und -veränderungen aus mikrogenetischen Studien mit Kindern, Jugendlichen und Erwachsenen wie folgt zusammen: Die selbstgesteuerte Exploration von Kindern war stärker von Vorwissen und eigenen Überzeugungen beeinflusst. Statt eigene Überzeugungen zu falsifizieren, versuchten sie meist, diese zu bestätigen und gewünschte Effekte zu produzieren.

Kinder führten eher uninformative, inkonklusive Experimente durch und zogen ihre Schlussfolgerungen bereits aus einem einzigen Experiment. Dabei ignorierten sie häufig Daten, die nicht mit ihren eigenen Theorien und Hypothesen konsistent waren. Wenn sie Notizen von Experimenten, Daten und Ergebnissen machten, gingen sie unsystematisch vor. Im Alter der fünften und sechsten Klassenstufe zeigten sich dann Verbesserungen dahingehend, dass vermehrt unkonfundierte und konklusive Tests durch die Kontrolle anderer Variablen durchgeführt, valide Schlussfolgerungen gezogen und evidenzbasierte Urteile getroffen wurden. Erwachsene begannen ihre selbstgesteuerte Exploration auf einem höheren Niveau und erreichten größere Fortschritte, aber erschlossen sich auch nicht immer vollständig das Kausalsystem der Mikrowelt. Sie generierten durch den Einsatz der Variablenkontrollstrategie häufiger kontrollierte und valide Tests und waren dabei weniger durch ihre eigenen Überzeugungen beeinflusst. Als Grundlage für ihre Schlussfolgerungen verwendeten sie die Ergebnisse von nicht nur einem, sondern von mehreren Experimenten und versuchten sich Inkonsistenzen zu erklären statt zu ignorieren. Erwachsene berücksichtigten und testeten bei der selbstgesteuerten Exploration häufiger mehrere, alternative Hypothesen und waren dadurch erfolgreicher. Auf allen Altersstufen zeigten die Probanden eine gleichzeitige Anwendung einer Vielzahl von weniger und mehr effektiven Strategien. Mit dem Alter und der vermehrten Beschäftigung mit der selbstgesteuerten Exploration nahm allmählich die Verwendung von weniger effektiven Strategien ab und der Einsatz von fortgeschrittenen, effektiveren Strategien zu. Es zeigten sich wechselseitige Beziehungen zwischen dem Strategie- und Wissenserwerb: Ausgeprägtes Inhaltswissen unterstützte den Einsatz von effektiven Experimentierstrategien, was wiederum zur Entwicklung von elaborierterem Inhaltswissen führte.

Ein weiteres wichtiges Ergebnis von Studien zur selbstgesteuerten Exploration ist die Rolle des metakognitiven Verständnisses im wissenschaftlichen Erkenntnisprozess (Kuhn et al., 1995; Kuhn, 1995, 2001, 2002): Unter prozeduralen, metastrategischen Kompetenzen wird das Verständnis der Ziele und das Wissen über Strategien des Erkenntnisgewinns sowie die Fähigkeit, diese Strategien im Hinblick auf das Ziel der Aufgabe auszuwählen und die Anwendung der Strategien zu überwachen und zu kontrollieren, verstanden. Deklaratives, metakonzeptuelles Verständnis im Sinne eines epistemologischen Verständnisses beinhaltet das Verständnis der Natur von Wissenschaften und wissenschaftlichen Wissens und die Unterscheidung zwischen Hypothesen/Theorien und Evidenzen. Kuhn und Kollegen (Kuhn et al., 1995; Kuhn & Pearsall, 1998) fanden bidirektionale Beziehungen zwischen metastrategischem Verständnis und dem erfolgreichen Einsatz der Strategien des

Erkenntnisprozesses: “metastrategic understanding may both guide (in implicit form) and follow (in more explicit form) strategy development” (Kuhn et al., 1995, S. 114).

Metakognitive Reflexionsfähigkeiten umfassen den gesamten wissenschaftlichen Erkenntnisprozess und spielen daher eine bedeutsame Rolle bei der Generierung von Hypothesen, dem Durchführen von Experimenten, der Evaluation von Daten und dem Ziehen von Schlussfolgerungen sowie dem Wissenserwerb. Die Fähigkeiten, zwischen Theorie und Evidenz zu unterscheiden sowie eigene Überzeugungen und deren Einfluss auf die Interpretation von Daten zu inhibieren, werden als Erfolgsfaktoren im wissenschaftlichen Denken angesehen (Kuhn & Franklin, 2006).

1.2.4 LOGIC-Studie zur Entwicklung des wissenschaftlichen Denkens von der Kindheit bis ins Erwachsenenalter

In der Munich Longitudinal Study on the Genesis of Individual Competencies (LOGIC-Studie; Weinert & Schneider, 1999) wurde das wissenschaftliche Denken von ca. 200 Kindern längsschnittlich vom Kindesalter bis ins Erwachsenenalter anhand einer Reihe von Maßen erfasst, um die Entwicklung und den Zusammenhang von verschiedenen Komponenten des wissenschaftlichen Denkens nachzuzeichnen (Bullock et al., 2009; Sodian, Bullock & Koerber, 2008). Die in der LOGIC-Studie verwendeten Instrumente und gefundenen Ergebnisse zu den einzelnen Teilkompetenzen des wissenschaftlichen Denkens wurden zum Teil schon separat in den Abschnitten 1.2.1.1 (sogenannte Flugzeugaufgabe: Variablenkontrollstrategie) und 1.2.1.4 (sogenanntes Hexerei-Interview: Theorieverständnis) dargestellt und sollen nun sowohl längsschnittlich betrachtet als auch in Zusammenhang gesetzt werden. Ein weiteres Maß der Batterie zum wissenschaftlichen Denken war das sogenannte Design-Interview: In verschiedenen Aufgaben mussten Aspekte bzw. Fehler des experimentellen Designs, beispielsweise die Variation der fokalen Variablendimension oder die Berücksichtigung der Basisrate im Prä-Post-Design, erkannt werden.

Die Entwicklung der verschiedenen Teilkompetenzen des wissenschaftlichen Denkens gestaltete sich unterschiedlich schnell. Kompetenz in der Wahl eines kontrollierten Tests zeigte sich bereits im Grundschulalter, aber erst später im Jugendalter konnten Fehler im experimentellen Design erkannt werden und spontan ein kontrollierter Test durch den Einsatz der Variablenkontrollstrategie produziert werden. Entwicklungsfortschritte im Theorieverständnis traten erst im Jugend- und Erwachsenenalter auf. Darüber hinaus zeig-

ten die Ergebnisse, dass bereits im Grundschulalter drastische individuelle Unterschiede vorhanden waren und diese im weiteren Entwicklungsverlauf relativ stabil blieben. In keinem der Maße zum wissenschaftlichen Denken traten Geschlechterdifferenzen auf.

Bullock und Kollegen (2009) fanden mittlere Zusammenhänge zwischen den Aufgaben zum Theorieverständnis und zu Experimentierstrategien. Zu den einzelnen Messzeitpunkten und über die Messzeitpunkte hinweg zeigten sich Korrelationen zwischen dem Wissen über Theorien und Experimente und der Fähigkeit, Experimente zu planen und zu produzieren. Pfadanalysen stellten bidirektionale Beziehungen zwischen Wissenschaftsverständnis und Experimentierstrategien heraus. Dieses Ergebnis wurde von Sodian und Kollegen (2008, S. 78-79) wie folgt interpretiert:

„Wer schon in der Kindheit überdurchschnittlich viel Wissen über Experimente und Theorien hat, wird wahrscheinlich im Jugend- und Erwachsenenalter bessere Strategien einsetzen, als jemand, der weniger Wissen hat. Umgekehrt ist die Strategienutzung auch eine Wissensquelle: Wer gute Strategien zur Lösung komplexer Probleme einsetzen kann, lernt auch, über den Prozess der Erkenntnisgewinnung zu reflektieren, und Wissen über den Wissenserwerb zu abstrahieren.“

Des Weiteren war ein früh ausgeprägtes Wissenschaftsverständnis ein guter Prädiktor für die spätere Produktion eines kontrollierten Experiments. Diese Befunde stützen die Annahme, dass das metakognitive Verständnis wesentlich für die Entwicklung des wissenschaftlichen Denkens ist (z. B. Sodian et al., 2006; Kuhn, 2002).

Bullock und Kollegen (2009) folgerten aus den dargestellten Befunden, dass sich die Entwicklung des wissenschaftlichen Denkens eher als ein gradueller Aufbau verschiedener Teilkompetenzen beschreiben lässt, weniger als qualitativer, stadienähnlicher Wandel des Denkens im Sinne von Piaget. Daher bilden die Komponenten des wissenschaftlichen Denkens vermutlich ein Bündel miteinander verknüpfter Teilfähigkeiten und keine einheitliche Kompetenz, die in einem bestimmten Entwicklungsstadium erreicht wird.

1.2.5 Studien zur Bedeutung des wissenschaftlichen Denkens für den Erwerb naturwissenschaftlichen Wissens

Allgemein werden wechselseitige Beziehungen zwischen wissenschaftlichem Denken und domänenspezifischem Inhaltswissen angenommen (z. B. Zimmermann, 2007). Dabei wird von funktionalen Zusammenhängen zwischen dem Wissenschaftsverständnis sowie den Methodenkompetenzen und dem Erwerb naturwissenschaftlichen Wissens ausgegangen, in

dem Sinne dass ein elaborierteres Wissenschaftsverständnis und bessere Methodenkompetenzen für das erfolgreiche Lernen inhaltlichen Wissens hilfreich sind.

Carey und Smith (1993) stellten allerdings in ihrem Überblicksartikel heraus, dass es bis dato nur wenige empirische Untersuchungen darüber gab, wie wissenschaftliches Denken und der Konzeptwechsel in naturwissenschaftlichen Domänen zusammenhängen. Einige wenige Studien fanden im Sekundarschulalter positive Zusammenhänge zwischen einem konstruktivistischem Wissenschaftsverständnis und dem Erwerb naturwissenschaftlichen Inhaltswissens: Schüler mit einer dynamischen bzw. konstruktivistischen Vorstellung von Wissenschaften erreichten ein besser integriertes naturwissenschaftliches Wissen als Schüler mit einer statischen, positivistischen Vorstellung (z. B. Songer & Linn, 1991; Wallace, Tsoi, Calkin & Darley, 2003). Eine dynamische bzw. konstruktivistische Vorstellung von Wissenschaft beinhaltete ein Verständnis der aktiven Wissenskonstruktion und des tentativen und kumulativen Charakters naturwissenschaftlicher Ideen und naturwissenschaftlichen Wissens, wohingegen eine statische, positivistische Vorstellung Wissenschaft als eine Sammlung von Fakten verstand. Grygier (2008) fand bereits für das Grundschulalter positive Effekte eines wissenschaftstheoretischen Curriculums auf den Lernzuwachs im Sachwissen. Es zeigten sich Zusammenhänge zwischen einem reflektiertem Wissenschaftsverständnis und dem Erwerb fachinhaltlichen Wissens. Die Befunde wurden dahingehend interpretiert, dass bereits im Grundschulalter ein elaboriertes epistemologisches Verständnis von der Natur der Naturwissenschaften die Umstrukturierung naturwissenschaftlicher Konzepte positiv beeinflussen kann.

Wenige Studien untersuchten die wechselseitigen Beziehungen zwischen domänenspezifischen Inhaltswissen und domänenübergreifenden Methodenkompetenzen: Schauble (1996) fasste den gefundenen Zusammenhang in ihrer Studie zum selbstgesteuerten Experimentieren in physikalischen Mikrowelten von Fünft-, Sechst- und Zehntklässlern wie folgt zusammen: „appropriate knowledge supports the selection of appropriate experimentation strategies, and systematic and valid experimentation strategies support the development of more accurate and complete knowledge“ (Schauble, 1996, S. 188). Die Befunde von Kwon und Lawson (2000) belegten ebenfalls diesen Zusammenhang, dadurch dass in multiplen Regressionsanalysen die gemessene Kompetenz im wissenschaftlichen Denken signifikant die Leistung im naturwissenschaftlichen Wissenstests vorhersagte. Auch wirkte sich ein gezieltes Training der domänenübergreifenden Variablenkontrollstrategie bei amerikanischen Grundschulkindern positiv auf den Erwerb domänenspezifischen, naturwissenschaftlichen Inhaltswissens aus (Chen & Klahr, 1999). Chen und Klahr schlossen

aus ihren Ergebnissen, dass der Erwerb domänenübergreifender Methodenkompetenzen den Erwerb domänenspezifischen Wissens erleichtern kann, indem durch den Einsatz effektiver Experimentierstrategien valide Schlussfolgerungen aus unkonfundierten Experimenten gezogen werden können. Dadurch können Fehlvorstellungen erkannt und ein fortgeschrittenes naturwissenschaftliches Verständnis erlangt werden.

1.3 Modellierung der Kompetenz wissenschaftliches Denken im Grundschulalter

1.3.1 Kompetenzmodelle

Weinert (2001) definiert Kompetenzen als kontextspezifische kognitive Leistungsdispositionen, die durch Lernen und Erfahrung erworben werden und daher durch Bildungsprozesse beeinflussbar und trainierbar sind. Nach Weinert soll das Kompetenzkonstrukt aus pragmatischen Gründen allgemeine kognitive Grundfunktionen und Intelligenz ausschließen. Allgemeine intellektuelle Fähigkeiten tragen zwar durchaus zur erfolgreichen Bearbeitung bereichsspezifischer Anforderungen bei, allerdings müssen diese nicht erworben werden. Sie gehören zu den allgemeinen menschlichen Grundfunktionen und generellen Dispositionen und können nur eingeschränkt durch Trainings oder Interventionen beeinflusst werden (Hartig & Klieme, 2006). Eine weitere Restriktion des Kompetenzbegriffs betrifft die Einschränkung auf den kognitiven Bereich und den Ausschluss von motivationalen oder affektiven Leistungsvoraussetzungen. Weinert schlägt eine getrennte Messung von kognitiven und motivationalen Kompetenzaspekten vor, um deren Interaktion empirisch untersuchen zu können. Kompetenzen werden als bereichsspezifische Leistungsdispositionen als notwendig für die erfolgreiche Bewältigung von Anforderungen in alltagsbezogenen Situationen oder Aufgaben bestimmter Domänen erachtet (Koeppen, Hartig, Klieme & Leutner, 2008) und können in einem bestimmten Ausmaß auch über ähnliche Situationen generalisiert werden (Hartig & Klieme, 2006).

Zur empirischen Erfassung und Modellierung von Kompetenzen ist es notwendig die interessierende Kompetenz theoretisch und systematisch in einem Kompetenzmodell zu definieren. Kompetenzmodelle ermöglichen es, die Struktur und die Entwicklung von Kompetenzen sowie individuelle Unterschiede in Form von Kompetenzniveaus kriteriumsorientiert zu beschreiben (Schrader, Helmke & Hosenfeld, 2008).

Kompetenzstrukturmodelle befassen sich mit der Dimensionalität von Kompetenzen, das heißt, welche und wie viele verschiedene Kompetenzdimensionen unterschieden werden können, und bilden die Struktur der Kompetenz ab (Schecker & Parchmann, 2006; Koeppen et al., 2008). Kompetenzniveauemodelle unterteilen eine kontinuierliche Kompetenzdimension in diskrete, ordinale Kategorien der Ausprägung: die sogenannten Kompetenzniveaus (Hartig & Klieme, 2006). Durch diese Kompetenzniveaus werden qualitative, kriteriumsorientierte Beschreibungen der erfassten Kompetenz möglich. Je Kompetenzniveau werden die Anforderungen bzw. die Leistungen der Personen inhaltlich beschrieben. Kompetenzentwicklung wird darin verstanden, dass die Kompetenzniveaus sukzessiv zum jeweils nächst höheren Niveau durchlaufen werden.

Im Bereich der Naturwissenschaften differenzierte Bybee (1997, 2002) die Entwicklung der scientific literacy in folgende vier Stufen eines Kontinuums (Schecker & Parchmann, 2006; Prenzel et al., 2001):

- Nominale naturwissenschaftliche Grundbildung (zusammenhangsloses Verständnis von Begriffen, Fakten und Formeln, die jedoch auf einem Niveau von naiven Theorien und Fehlvorstellungen aus wissenschaftlicher Sicht falsch sind)
- Funktionale naturwissenschaftliche Grundbildung (in einem eingegrenzten Bereich von Situationen und Tätigkeiten kann naturwissenschaftliches Vokabular und Faktenwissen korrekt verwendet werden, aber kein tieferes Verständnis der Begriffe und Zusammenhänge)
- Konzeptuelle und prozedurale naturwissenschaftliche Grundbildung (Verständnis und passende Anwendung zentraler naturwissenschaftlicher Konzepte, Prinzipien sowie Denk- und Arbeitsweisen, Herstellen von Zusammenhängen zwischen Fakten, Prinzipien und Begriffen)
- Multidimensionale naturwissenschaftliche Grundbildung (Verständnis der Besonderheiten naturwissenschaftlichen Wissens und Denkens, Herstellen von Verbindungen zwischen verschiedenen naturwissenschaftlichen Disziplinen, Einordnung der Naturwissenschaften in geschichtliche, soziale und kulturelle Zusammenhänge).

Das Stufenmodell von Bybee beschreibt naturwissenschaftliche Kompetenz allerdings vorwiegend normativ. In den in Anlehnung daran entwickelten Kompetenzmodellen inter-

nationaler Schulleistungsstudien wie PISA und IGLU-E wurde die Abgrenzung und Beschreibung der Kompetenzniveaus erst post hoc auf Basis der ermittelten Schwierigkeitsparameter der Aufgaben und anhand von Anforderungsanalysen vorgenommen. Daher wurden diese Kompetenzmodelle aus theoretischer Sicht als unzureichend kritisiert (vgl. Prenzel et al., 2001; Bos et al., 2003).

Für das Sekundarschulalter hat die empirische Überprüfung der Struktur naturwissenschaftlicher Kompetenz sowie die Erarbeitung von Kompetenzentwicklungsmodellen bereits begonnen (vgl. Parchmann, 2010). Die Modellierung der naturwissenschaftlichen Kompetenzentwicklung im Grundschulalter ist ein derzeit noch nicht systematisch untersuchter Forschungsgegenstand. Nur wenige Arbeiten thematisierten das Grundschulalter: Beispielsweise beschrieb das Modell von Hammann (2004) zum Experimentieren die Kompetenz des Grundschulkindes ausschließlich auf dem Niveau der Fehlvorstellungen. Eine Modellierung einer Kompetenzentwicklung für das Grundschulalter wurde nicht berücksichtigt.

Im Gegensatz dazu wurde im Forschungsprojekt Science-P basierend auf den beschriebenen entwicklungspsychologischen Befunden zu Basiskompetenzen von Grundschulkindern im wissenschaftlichen Denken (vgl. Abschnitte 1.2.1 und 1.2.2.), der Forschung zum begrifflichen Wandel im Bereich des naturwissenschaftlichen Lernens (z. B. diSessa, 2001; Vosniadou, Ioannides, Dimitrakopoulou & Papademetriou, 2001) und in Anlehnung an das Stufenmodell zur scientific literacy von Bybee (1997, 2002) ein Struktur-Niveau-Modell naturwissenschaftlicher Kompetenz für das Grundschulalter postuliert (Hardy et al., 2010): Es wurden die beiden Kompetenzdimensionen naturwissenschaftliches Wissen und Wissen über Naturwissenschaften sowie je Dimension a priori drei hierarchisch angeordnete Kompetenzniveaus angenommen. Die Kompetenzdimension naturwissenschaftliches Wissen fokussierte auf die physikalischen Inhaltsbereiche Schwimmen und Sinken sowie Verdunstung und Kondensation. Auf die Modellierung des naturwissenschaftlichen Inhaltswissens wird in der vorliegenden Studie nicht näher eingegangen (für eine detaillierte Beschreibung dieser Kompetenzdimension siehe z. B. Pollmeier et al., 2009). In der Kompetenzdimension Wissen über Naturwissenschaften lag der Fokus auf der Modellierung der beiden Kompetenzkomponenten Wissenschaftsverständnis und Methodenkompetenz. Die Kompetenzkomponente Wissenschaftsverständnis umfasst die Teilkompetenzen Verständnis wissenschaftlicher Konzepte und der Ziele von Wissenschaften sowie das Theorieverständnis. Die zweite Komponente Methodenkompetenz beinhaltet die Teilkompetenzen Experimentierstrategien und Dateninterpretation. In Abbildung 1.2 ist das Kompetenz-

strukturmodell der Kompetenz wissenschaftliches Denken schematisch dargestellt. Eine inhaltliche Beschreibung der Kompetenzkomponenten wurde im Abschnitt 1.2 vorgenommen.

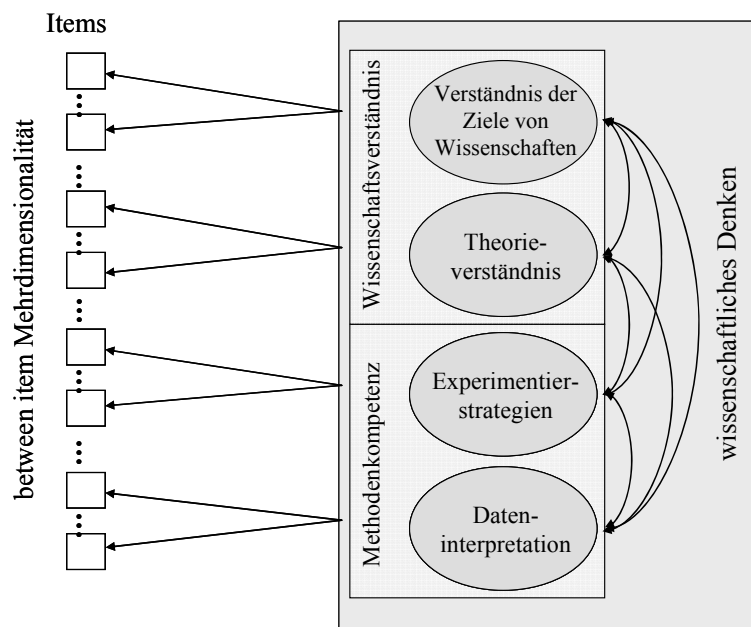


Abbildung 1.2. Schematische Darstellung des Strukturmodells der Kompetenz wissenschaftliches Denken

Entsprechend den Ergebnissen der Konzeptwechselforschung (conceptual change) wurde die Kompetenzentwicklung in den beiden Kompetenzdimensionen als graduelle Umstrukturierung im Sinne eines begrifflichen Wandels angenommen und ein Entwicklungsverlauf von drei hierarchisch angeordneten Kompetenzniveaus postuliert: naive Vorstellungen (Fehlvorstellungen), die einer empirischen Prüfung nicht standhalten; Zwischenvorstellungen, die naturwissenschaftliche Phänomene nur begrenzt erklären können; wissenschaftliche Vorstellungen, die auf wissenschaftlich adäquaten Konzepten beruhen. Aufbauend auf den Verständnisebenen von Carey et al. (1989) können im wissenschaftlichen Denken die drei Kompetenzniveaus wie folgt beschrieben werden (vgl. Tabelle 3.3, S. 52): Auf dem Niveau der naiven Vorstellungen wird Wissenschaft als reine Aktivität oder Prozedur verstanden. Dabei fehlen das Ziel des Erkenntnisgewinns und der Bezug zu Theorien oder Hypothesen. Experimente werden vorwiegend zur Produktion erwünschter Effekte statt als Methode zur Prüfung von Hypothesen gesehen. In einem graduellen Prozess der Umstrukturierung müssen diese nicht belastbaren Konzepte aufgegeben werden. Durch eine Verknüpfung von naiven Vorstellungen und wissenschaftlich belastbaren Erklärungen handelt es sich bei dem Niveau der Zwischenvorstellungen um Konzepte, die sich im Alltag der

Kinder bewähren, und die durch eine Differenzierung in das Niveau der wissenschaftlichen Vorstellungen überführt werden können. Auf dem Niveau der Zwischenvorstellungen besteht ein Verständnis von Wissenschaft als objektivistisches Sammeln von Fakten, Hestellen einfacher Zusammenhänge und Testen von einzelnen Hypothesen. Aber im Gegensatz zu Niveau 3 fehlt hier die Einbettung einzelner Experimente zur Evaluation einer Theorie. Auf dem höchsten Niveau handelt es sich um Vorstellungen, die mit wissenschaftlichen Erklärungen vereinbar sind. Wissenschaft wird als die Suche nach überprüfbareren Erklärungen und wissenschaftliches Wissen als das Ergebnis der Hypothesenprüfung gesehen. Es wird der zyklische und kumulative Prozess der Bildung, Prüfung und Revision von Theorien und Hypothesen verstanden. Theorien werden als forschungsleitend sowohl bei der Hypothesenbildung als auch bei der Evidenzevaluation gesehen.

Die empirische Modellierung der latenten Struktur der Kompetenz wissenschaftliches Denken war Ziel der vorliegenden Forschungsarbeit. Eine psychometrische Überprüfung des Kompetenzniveaumodells steht bisher noch aus. Dieser Fragestellung wird in den weiteren Phasen des Projektes Science-P nachgegangen.

1.3.2 Kompetenzmessung

Theoretische Kompetenzmodelle bilden die Basis für die Konstruktion von Messinstrumenten und damit zur Diagnose von Kompetenzen. Um Leistungen und Kompetenzen erfassen zu können, gibt es verschiedene Messverfahren (Duit et al., 2001): Zur Messung naturwissenschaftlicher Kompetenzen werden neben Interviews oder Aufgaben im offenen Antwortformat, in denen freie Antworten produziert werden müssen, häufig Multiple-Choice-Tests verwendet. In Multiple-Choice-Aufgaben werden ein Aufgabenstamm, in dem die Aufgabenstellung dargestellt ist, sowie mehrere Antwortmöglichkeiten, von denen in der Regel eine richtige Alternative ausgewählt werden soll, vorgegeben. Weitere geschlossene Antwortformate sind: Forced-Choice-Aufgaben, in denen eine von zwei Antwortalternativen gewählt werden muss, Richtig-Falsch-Aufgaben, in denen eine Aussage als richtig oder falsch bewertet werden muss, sowie Multiple-Select-Aufgaben, bei denen unter einem Aufgabenstamm mehrere Items im Richtig-Falsch-Format beantwortet werden müssen. In Experimental- bzw. Performanztests ist es die Aufgabe des Probanden, zur Lösung eines vorgegebenen Problems bzw. zur Hypothesenprüfung selbständig eine Untersuchung zu planen, durchzuführen, zu protokollieren, die Ergebnisse auszuwerten und

zu interpretieren. Tests im geschlossenen Antwortformat können sowohl effizient, standardisiert und ökonomisch durchgeführt als auch ökonomisch und objektiv ausgewertet und interpretiert werden. Dagegen sind offene Formate in der Anwendung und Auswertung unökonomischer, werden aber im Hinblick auf ihre Validität höher eingestuft. Ein sorgfältig konstruierter und validierter Papier- und Bleistift-Test kann ein Konstrukt breit und zeitökonomisch erfassen und bietet daher große Vorteile für die Forschung.

Zur Messung des wissenschaftlichen Denkens wurden überwiegend Interviews durchgeführt (z. B. Carey et al., 1989; Sodian et al., 1991; Bullock & Ziegler, 1999). Für das Sekundarschulalter existieren einige wenige Testverfahren im geschlossenen Antwortformat. Ausgewählte Instrumente sind im Nachfolgenden kurz dargestellt: Klopfer und Cooley (1961) entwickelten den Fragebogen Test on Understanding Science (TOUS), um bei Sekundarschülern und Lehrern ihr Wissenschaftsverständnis und ihr Verständnis von Zielen und Methoden von Wissenschaften anhand von 60 Multiple-Choice-Items (Reliabilität 0.76) zu erfassen und kamen zu dem Ergebnis, dass das Wissenschaftsverständnis von Sekundarschülern als inadäquat einzustufen war. Der Test of Science Processes von Tanenbaum (1971) misst die Kompetenzen von Schülern der Klassenstufen 7-9 in insgesamt 96 Multiple-Choice-Items in den acht Subskalen Beobachten, Vergleichen, Klassifizieren, Quantifizieren, Messen, Experimentieren, Schlussfolgern und Vorhersagen (Reliabilitäten zwischen 0.30 und 0.81). Ein weiteres Beispiel für einen Multiple-Choice-Test ist der von Fraser (1980) für Schüler der 7. bis 10. Klasse entwickelte Test of Enquiry Skills (TOES). Dieser Test beinhaltet 87 inhaltsfreie, vorwissensarme Items in neun Subskalen (Verwendung von Bibliotheken und Inhaltsverzeichnissen, Interpretieren und Verarbeiten von Informationen aus Skalen, aus Mittelwerten/Prozenten/Verhältnissen, aus Tabellen und Graphen, Verständnis wissenschaftlicher Texte, Design experimenteller Verfahren, Schlussfolgern und Verallgemeinern; Reliabilitäten von 0.65 bis 0.82). Zwei neuere, deutschsprachige Instrumente beruhen auf dem Scientific Discovery as Dual Search-Modell (SDDS) von Klahr (2000): Hammann, Phan und Bayrhuber (2008) entwickelten einen Test zur Messung von Kompetenzen beim Experimentieren für Schüler der fünften und sechsten Klassenstufe. Die drei Subskalen Suche im Hypothesenraum, Testen von Hypothesen und Analyse von Evidenzen (Reliabilitäten zwischen 0.69 und 0.78) waren moderat bis hoch korreliert und erreichten als Gesamtskala aus 24 Items mit 0.88 eine zufrieden stellende Reliabilität. Die Überprüfung der Dimensionalität gab Hinweise auf die Existenz einer bis zwei Dimensionen. Glug (2009) beschrieb in ihrer Dissertation die Entwicklung eines Tests zur Erfassung der prozessbezogenen naturwissenschaftlichen

Grundbildung in der neunten Klassenstufe. Der Test besteht aus 30 Items (Reliabilität 0.77) in den drei Subskalen Identifizieren wissenschaftlicher Hypothesen, Planen einer wissenschaftlichen Untersuchung und Nutzen wissenschaftlicher Ergebnisse (Reliabilitäten von 0.48 bis 0.65). Die Vergleiche von ein- und mehrdimensionalen Rasch-Modellen sowie die hohen Korrelationen der Teildimensionen ergaben eine Eindimensionalität der Kompetenz.

Für den Primarbereich fehlt es an validierten Testverfahren im geschlossenen Antwortformat. Das Forschungsprojekt Science-P entwickelte, basierend auf dem postulierten Kompetenzmodell (vgl. Abschnitt 1.3.1) und in Anlehnung an in der Forschung etablierten Interviewinstrumenten, gruppentestfähige Aufgaben im geschlossenen Antwortformat für das Grundschulalter, um mit einer Skala einen breiteren Kompetenzbereich abdecken sowie effizient und ökonomisch erfassen zu können. Um dem Aspekt der Kompetenzentwicklung nachzugehen, wurden in den drei Antwortalternativen der Multiple-Choice-Items die theoretisch begründeten Kompetenzniveaus naive Vorstellungen, Zwischenvorstellungen und wissenschaftliche Vorstellungen abgebildet. Eine genaue Beschreibung der Aufgabenentwicklung zu den Subskalen Verständnis der Ziele von Wissenschaften, Theorieverständnis, Experimentierstrategien und Dateninterpretation nach dem von Wilson (2005) vorgeschlagenen Ansatz des „Construct Modeling“ findet sich in Abschnitt 3.2.1.1 im Methodenteil der vorliegenden Arbeit. Die Testkonstruktion erfolgte in einem gestuften Prozess: In der Präpiloting- und Itempilotingphase wurden die psychometrischen Eigenschaften der Testitems anhand der klassischen und probabilistischen Testtheorie überprüft (Koerber, Kropf, Mayer, Sodian & Schwippert, 2009; Koerber, Sodian, Kropf, Mayer & Schwippert, 2011): Die Analysen zeigten, dass ein reliables Testinstrument zur Erfassung des wissenschaftlichen Denkens im Grundschulalter entwickelt werden konnte (Cronbachs Alpha 0.76; Koerber et al., 2011). Es fanden sich signifikante Verbesserungen im wissenschaftlichen Denken von der zweiten zur vierten Klassenstufe. Die Mehrheit der Zweitklässler konnte dem Niveau der naiven Vorstellung und ein Großteil der Viertklässler dem Niveau der Zwischenvorstellung zugeordnet werden. Des Weiteren zeigten sich bereits im Grundschulalter substantielle Korrelationen zwischen den Teilkompetenzen Experimentierstrategien und Dateninterpretation.

In der Science-P Validierungsstudie, die die Übereinstimmung der schriftlichen Aufgaben im geschlossenen Antwortformat mit Einzelinterviews in der dritten Klassenstufe untersuchte, wurde belegt, dass das wissenschaftliche Denken anhand des entwickelten Gruppentests valide erfasst werden konnte (Kropf, 2010). Allerdings erzielten die Kinder

in den Multiple-Choice-Items generell bessere Leistungen als in den entsprechenden Interviewfragen. Wurden die Items, die in den Raschskalierungen der Itempiloting zufriedenstellende statistische Kennwerte aufwiesen, in den Vergleich der erreichten mittleren Verständnisniveaus aufgenommen, waren die Aufgaben der Methodenkompetenz, mit Experimentierstrategien gefolgt von Dateninterpretation, sowohl in den schriftlichen Aufgaben als auch im Interview am einfachsten zu beantworten. Die nächst besseren Leistungen erzielten die Kinder in den Testitems der Teilkompetenz Verständnis wissenschaftlicher Konzepte und der Ziele von Wissenschaften, wohingegen im Interview konsistent mit früheren Studien lediglich Bodeneffekte gefunden wurden (Sodian et al., 2006; Grygier, 2008). Vergleichbar mit dem Entwicklungstrend der LOGIC-Studie (Bullock et al., 2009) hatten die Aufgaben, die das Theorieverständnis erfassten, den höchsten Schwierigkeitsgrad.

1.3.3 Abgrenzung und Zusammenhang der Kompetenz wissenschaftliches Denken zu kognitiven Fähigkeiten und motivationalen Orientierungen

Im Rahmen der Kompetenzmodellierung ist es darüber hinaus interessant und bedeutsam die spezifische interessierende Kompetenz, hier wissenschaftliches Denken, von Intelligenz und anderen kognitiven Fähigkeiten, die bei der Bearbeitung der Aufgaben eine Rolle spielen könnten, zu differenzieren sowie den Zusammenhang zu individuellen Bedingungsfaktoren, die der Kompetenzentwicklung im wissenschaftlichen Denken zugrunde liegen könnten, zu untersuchen.

Allgemein wird angenommen, dass Leistungs- und Kompetenzentwicklung multifaktoriell bedingt sind (Weinert & Helmke, 1997): Dabei haben kognitive Fähigkeiten gegenüber motivationalen und sozialen Bedingungsfaktoren einen größeren Einfluss. Jedoch weisen Studien darauf hin, dass kognitive Fähigkeiten eine notwendige, aber durchaus nicht hinreichende Voraussetzung für den Kompetenzerwerb sind. Vielmehr wird die Bedeutung der wechselseitigen Beziehungen zwischen kognitiven Fähigkeiten und motivationalen Faktoren hervorgehoben.

In der vorliegenden Studie wurden nach dem derzeitigen Forschungsstand relevante und potentielle individuelle Bedingungsfaktoren für die Kompetenz im wissenschaftlichen Denken ausgewählt, deren Einfluss im Grundschulalter bisher noch nicht systematisch untersucht wurde. Bei den kognitiven Variablen handelt es sich dabei um Intelligenz,

Leseverständnis, Problemlösen, Inhibition, räumliches Denken und formal-operatorische Fähigkeiten. Für die motivationalen Orientierungen wurden die Einstellungen zu Naturwissenschaften, das naturwissenschaftsbezogene Fähigkeitsselbstkonzept sowie das Interesse an forschenden Aktivitäten betrachtet.

1.3.3.1 Intelligenz

Intelligenz als grundlegendes Denkvermögen wird allgemein definiert als die Fähigkeit, „sich in neuen Situationen auf Grund von Einsichten zurechtzufinden oder Aufgaben mit Hilfe des Denkens zu lösen, ohne dass hierfür die Erfahrung, sondern vielmehr die Erfassung von Beziehungen das Wesentliche ist“ (Häcker & Stapf, 2004, S. 447). Damit kann Intelligenz umschrieben werden als die allgemeine geistige Fähigkeit, sich an neuartige Aufgaben, Anforderungen und Situationen anzupassen.

Allgemeine intellektuelle Fähigkeiten werden zwar aus der Definition von Kompetenzen ausgeschlossen, allerdings spielen sie bei der erfolgreichen Bearbeitung bereichsspezifischer Anforderungen und Aufgaben eine wichtige Rolle (Hartig & Klieme, 2006). Steht der Aspekt kognitive Leistungsdisposition im Vordergrund der Kompetenzdefinition bestehen große inhaltliche Überschneidungen mit dem Intelligenzkonzept.

Zwar stellen Aufgaben zur Messung von Kompetenzen meist Anforderungen, die entsprechend der Kompetenzdefinition realen Situationen entsprechen, wohingegen Intelligenztests eher abstrakt, kontextunabhängig und möglichst vorwissensarm konstruiert sind. Allerdings ähneln sich die Verfahren in Form von standardisierten Leistungstests zur Erfassung von Kompetenzen und zur Diagnose von Intelligenzaspekten (Hartig & Klieme, 2006). Daher sind Zusammenhänge zwischen Kompetenz- und Intelligenzmaßen anzunehmen. Typischerweise zeigen sich recht hohe empirische Zusammenhänge zwischen Intelligenz und Kompetenz: Beispielsweise wurden in PISA 2003 messfehlerbereinigte latente Korrelationen von 0.64 bis 0.74 zwischen Maßen von Intelligenz und Kompetenz berichtet (Hartig & Klieme, 2006).

Auch in der LOGIC-Studie wurden die Kompetenzen im wissenschaftlichen Denken hinsichtlich der Zusammenhänge zu Intelligenzmaßen untersucht. Zwischen allen Maßen des wissenschaftlichen Denkens und den erfassten Intelligenzaspekten fanden sich kleine bis mittlere Zusammenhänge (Bullock et al., 2009).

Die Untersuchung des Zusammenhangs zwischen Intelligenzaspekten und der Kompetenz im wissenschaftlichen Denken anhand des schriftlichen Tests des Forschungsprojekts Science-P im Rahmen der Validierungsstudie ergab kein einheitliches Bild (Kropf, 2010): vereinzelt wurden signifikante Korrelationen zu den Begabungsaspekten Verarbeitungsgeschwindigkeit, fluides Schlussfolgern und räumliches Vorstellungsvermögen gefunden. Es bedarf einer weiteren Überprüfung der Zusammenhänge zwischen dem Test zum wissenschaftlichem Denken und Intelligenzmaßen an einer größeren Stichprobe, um die beiden Konstrukte differenziert betrachten sowie den Einfluss der Intelligenz auf die Kompetenz im wissenschaftlichen Denken untersuchen zu können.

1.3.3.2 Leseverständnis

Lesekompetenz kann beschrieben werden als „die Fähigkeit, geschriebene Texte unterschiedlicher Art in ihren Aussagen, ihren Absichten und ihrer formalen Struktur zu verstehen und in einen größeren sinnstiftenden Zusammenhang einordnen zu können, sowie in der Lage zu sein, Texte für verschiedene Zwecke sachgerecht zu nutzen“ (Baumert, Artelt, Klieme & Stanat, 2001, S. 290).

Generell spielen Sprach- und Lesekompetenzen eine wichtige Rolle in der Kompetenz- und Leistungsentwicklung (Schrader et al., 2008). Sprache und sprachgebundene Kompetenzen gelten als wichtige Voraussetzung für Denkvorgänge und Erfolg in verschiedenen Wissensbereichen. Ein ausgeprägtes Leseverständnis wird als eine Grundlage erachtet, um Aufgaben in schriftlicher Form bearbeiten zu können. Die noch nicht ausgereifte Lesekompetenz der Grundschul Kinder könnte die Bearbeitung der Aufgaben einschränken und daher die Validität der Kompetenzmessung beeinflussen. Denn besonders im Verlauf des Grundschulalters finden noch markante Verbesserungen in Aufmerksamkeits-, Gedächtnis- und metakognitiven Strategien sowie dem sinnerfassenden und flüssigen Lesen statt (Paris, Morrison & Miller, 2006). Allerdings zeigten Lesestudien wie IGLU (Internationale Grundschul-Lese-Untersuchung) und VERA (Vergleichsarbeiten in der Grundschule), dass bereits Grundschul Kinder sinn- und textverstehendes Lesen beherrschen (Schrader et al., 2008).

Zur Erfassung wissenschaftlichen Denkens im Grundschulalter wurden bisher überwiegend Interviews und freie Antwortformate verwendet (z. B. Carey et al., 1989; Sodian et al., 1991). Generell werden Testinstrumente im geschlossenen Antwortformat als weni-

ger geeignet erachtet (Duit et al., 2001). In der Validierungsstudie des Projektes Science-P wurde durch den Vergleich der Leistungen in den schriftlichen Testitems und den Einzelinterviews belegt, dass eine valide Erfassung wissenschaftlichen Denkens im Grundschulalter mit geschlossenen, schriftlichen Antwortformaten möglich ist (Kropf, 2010). Es fanden sich mittlere Zusammenhänge zwischen den Aufgaben zum wissenschaftlichen Denken und dem Leseverständnis der Kinder. Zur näheren Untersuchung der diskriminanten Validität gilt allerdings weiterhin zu klären, welchen Einfluss das Textverständnis auf die Bearbeitung der schriftlichen Aufgaben zum wissenschaftlichen Denken hat und inwieweit sich die erfasste Kompetenz wissenschaftliches Denken von der Lesekompetenz der Kinder differenzieren lässt.

1.3.3.3 Problemlösen

Allgemein kann die exekutive Funktion Problemlösen definiert werden als „zielorientiertes Denken und Handeln in Situationen, für deren Bewältigung keine routinierten Vorgehensweisen verfügbar sind“ (Klieme, Funke, Leutner, Reimann & Wirth, 2001, S. 185). Problemlösendes Denken besteht darin, dass von einer Ausgangssituation ein mehr oder weniger gut definierter Zielzustand erreicht werden soll, ohne dass der Lösungsweg unmittelbar erkennbar ist. Dieser zielgerichtete Problemlöseprozess beinhaltet vier Hauptphasen von exekutiven Funktionen (Zelazo, Carter, Reznick & Fryne, 1997; Tucha & Lange, 2004; Klahr, 2000): Zunächst muss die Problemstellung erfasst und der Problemraum repräsentiert werden. Der Problemraum besteht aus dem Anfangs- und Zielzustand, verschiedenen Zwischenzuständen sowie Operationen und Einschränkungen. Darauf folgt die Planungsphase: Geeignete Lösungsstrategien müssen entwickelt und ein passender Plan aus den möglichen Alternativen ausgewählt werden. Der nächste Schritt besteht in der Ausführung des gewählten Handlungsplans. Dabei muss die gewählte Lösungsstrategie im Gedächtnis behalten und in die Tat umgesetzt werden. In der letzten Phase des Problemlöseprozesses wird die Handlung dahingehend evaluiert, ob eine Lösung des Problems erzielt wurde. Die Evaluationsphase beinhaltet gegebenenfalls das Finden von Fehlern und die Korrektur dieser durch eine Änderung der Lösungsstrategie. Zusammenfassend kann der Prozess des Problemlösens als die Suche nach einem Lösungsweg, um von dem Ausgangszustand zu dem Zielzustand zu gelangen, beschrieben werden. Entscheidend für den Problemlöseprozess sind das Begreifen und die Repräsentation der Problemsituation sowie deren schritt-

weise Veränderung, basierend auf planendem und schlussfolgerndem Denken (Klieme et al., 2001).

Untersuchungen zur Entwicklung der Problemlösefähigkeit zeigten, dass die Testleistung mit steigendem Alter der Kinder in standardisierten Testverfahren, wie beispielsweise dem Turm von London, zunahm und mit einem Alter von 12 Jahren das Niveau von Erwachsenen erreichte; in einigen Maßen entwickelten sich Problemlösefähigkeiten bis ins Jugend- und Erwachsenenalter weiter (Zelazo & Müller, 2002; Krikorian, Bartok & Gay, 1994).

Klahr und Kollegen (Klahr & Dunbar, 1988; Dunbar & Klahr, 1989; Klahr, 2000) stellen in ihrem Scientific-Discovery-As-Dual-Search (SDDS)-Modell den zyklischen Prozess der wissenschaftlichen Erkenntnissuche als Problemlöseprozess dar. Der Prozess der wissenschaftlichen Erkenntnisgewinnung wird als die Suche in zwei Problemräumen, dem Hypothesenraum und den Experimentraum, beschrieben. Dabei werden drei interdependente Hauptkomponenten des wissenschaftlichen Denkens unterschieden: Suche im Hypothesen-Raum, Suche im Experiment-Raum und Evidenzevaluation.

Mit der Suche im Hypothesenraum beginnt das Lösen eines wissenschaftlichen Problems: Ausgehend von dem Wissen in einer Domäne und den Ergebnissen vorangegangener Experimente muss eine überprüfbare Hypothese zur Erklärung eines interessierenden Phänomens aufgestellt werden. Als nächstes folgt das Testen der Hypothese: Es müssen geeignete Experimente geplant (Suche im Experiment-Raum) und durchgeführt werden, um Daten im Hinblick auf die Überprüfung der Hypothese zu erhalten. Zum Schluss müssen die gewonnenen Evidenzen im Hinblick auf die aufgestellte Hypothese analysiert werden (Evaluation von Evidenzen). Es folgt die Entscheidung, ob die Hypothese akzeptiert, abgelehnt oder in weiteren Experimenten getestet werden muss.

Die Studie von Dunbar und Klahr (1989; Klahr, 2000) zeigte sowohl Ähnlichkeiten als auch Unterschiede in der Art und Weise, wie Kinder und Erwachsene Hypothesen generierten, Experimente aufstellten und durchführten sowie Evidenz evaluierten. Die Ergebnisse fanden einen alterskorrelierten Entwicklungstrend, wie systematisch und effektiv die Probanden im Hypothesen- und Experiment-Raum suchten. Klahr und Kollegen führten die Schwierigkeiten der Kinder im wissenschaftlichen Denken darauf zurück, dass sie die Problemlösesituation (Problemraum) im wissenschaftlichen Erkenntnisprozess nicht vollständig repräsentieren und die Suche im Hypothesen- und Experiment-Raum nicht koordinieren können.

Der Vergleich des wissenschaftlichen Denkens mit einem Problemlöseprozess legt die Schlussfolgerung nahe, dass allgemeine Problemlösekompetenzen mit dem erfolgreichen Lösen von wissenschaftlichen Problemen in Zusammenhang stehen. In einer Studie von Kwon und Lawson (2000) konnte gezeigt werden, dass Problemlösefähigkeiten von Jugendlichen (13 - 17 Jahre) ihre Leistung im schriftlichen Test zum wissenschaftlichen Denken beeinflusste.

Der Zusammenhang zwischen Problemlösefähigkeiten und wissenschaftlichen Denken im Grundschulalter wurde bisher nicht systematisch untersucht. Es gilt zu überprüfen, inwieweit allgemeine Problemlösefähigkeiten und wissenschaftliches Problemlösen differenziert werden können bzw. inwiefern problemlösendes Denken eine kognitive Voraussetzung für wissenschaftliches Denken ist.

1.3.3.4 Inhibition

Inhibitionsfähigkeiten gehören zu den exekutiven Funktionen. Unter Inhibition versteht man „kognitive Prozesse, die der Unterdrückung einer bestimmten Handlungstendenz und damit einer bereits initiierten Reaktion dienen“ (Seiferth, Thienel & Kircher, 2007, S. 272). Somit besteht die Inhibitionsleistung in der Fähigkeit, dominante und automatische Reaktionstendenzen bewusst zu hemmen.

Es zeigen sich Zusammenhänge zwischen der Inhibitionsleistung und dem Lebensalter: Zwischen drei und fünf Jahren finden markante Entwicklungsfortschritte statt und weitere Verbesserungen sind vom Grundschul- bis ins Erwachsenenalter zu beobachten (Zelazo & Müller, 2002).

Die Ergebnisse zu individuellen und alterskorrelierten Unterschieden im wissenschaftlichen Denken aus Studien zur selbstgesteuerten Exploration komplexer Mikrowelten diskutierten Kuhn und Kollegen (Kuhn & Franklin, 2006) dahingehend, dass die Inhibition des Einflusses von eigenen Vorstellungen über Ursachefaktoren auf die Generierung von Hypothesen, Planung von Experimenten und auf die Interpretation von Daten als Erfolgsfaktor für wissenschaftliches Denken gesehen werden kann.

St. Clair-Thompson und Gathercole (2006) fanden signifikante Zusammenhänge zwischen inhibitorischer Kontrolle und der Leistung in standardisierten Tests zur Erfassung naturwissenschaftlicher Kompetenz bei Elf- und Zwölfjährigen. In einer Studie von Kwon und Lawson (2000) konnte gezeigt werden, dass die Fähigkeit von Jugendlichen (13 - 17

Jahre), aufgabenirrelevante Informationen zu inhibieren, wissenschaftliches Denken beeinflusste. Eine Untersuchung des Einflusses von inhibitorischen Fähigkeiten auf das wissenschaftliche Denken steht für das Grundschulalter bisher noch aus.

1.3.3.5 Räumliches Denken

Unter räumlichen Fähigkeiten versteht man „die Begabung, in der Vorstellung räumlich zu sehen und zu denken, d. h. im Gedächtnis gespeicherte mehrdimensionale Vorstellungsbilder zu reproduzieren und mit ihnen mental zu operieren“ (Quaiser-Pohl, 2001, S. 422).

Im räumlichen Denken finden sich alterskorrelierte und geschlechtsspezifische Unterschiede (Liben, 2002). Geschlechtsspezifische Leistungsvorteile von Jungen und Männern im räumlichen Denken sind seit langem ein stabiler Befund der psychologischen Geschlechterunterschiedsforschung, wobei die deutlichsten Geschlechterdifferenzen im Bereich der mentalen Rotation auftreten (Linn & Petersen, 1985; Lohaus, Schuhmann-Hengsteler & Kessler, 1999; Kerns & Berenbaum, 1991).

Studien weisen darauf hin, dass die Fähigkeit zur räumlichen Visualisierung bedeutsam für die Leistungen in vielen wissenschaftlichen und technischen Bereichen ist (Shea, Lubinski, & Benbow, 2001). Visuell-räumliches Denken wird als notwendig erachtet, um im wissenschaftlichen Erkenntnisprozess einerseits Ideen über Zusammenhänge mental und schematisch darzustellen und andererseits Daten in Form von Diagrammen, Karten, etc. zu repräsentieren und zu kommunizieren (Mathewson, 1999). In der Wissenschaft sind strukturierte Repräsentationsformen, wie Graphen und Diagramme, ein wichtiges Werkzeug, um die Art und das Ausmaß der Beziehungen zwischen zwei oder mehreren Variablen zu veranschaulichen (Koerber & Sodian, 2009). Für die Fähigkeit, Informationen in Graphen abzubilden bzw. Informationen aus Graphen abzulesen, ist es notwendig Beziehungen zwischen Variablen auf räumliche Relationen abbilden zu können (Koerber & Sodian, 2008).

Ob ein Zusammenhang zwischen wissenschaftlichem Denken und räumlichen Fähigkeiten im Grundschulalter besteht und inwieweit räumliches Denken eine Voraussetzung zum erfolgreichen wissenschaftlichen Denken ist, wurde bisher noch nicht systematisch untersucht. St. Clair-Thompson und Gathercole (2006) berichteten von signifikanten Zusammenhängen zwischen Maßen zum visuell-räumlichen Arbeitsgedächtnis und der Leistung in standardisierten Tests zur Erfassung naturwissenschaftlicher Kompetenz bei Elf-

und Zwölfjährigen. Die Studie von Kropf (2010) zeigte ebenfalls positive Beziehungen zwischen den Leistungen von Drittklässlern im visuell-räumlichem Vorstellungsvermögen und dem Theorieverständnis sowie dem Verständnis wissenschaftlicher Arbeitsweisen.

1.3.3.6 Formal-operatorische Fähigkeiten

Piaget schlug in seiner strukturalistischen Stufentheorie der geistigen Entwicklung ein höchstes Stadium der formalen Operationen vor, das nach einem Entwicklungsverlauf durch die vorangehenden Stadien, sensumotorisches (0 - 2 Jahre), vor-operatorisches (2 - 7 Jahre) und konkret-operatorisches (7 - 12 Jahre) Stadium, in einer universellen Entwicklung ab dem Alter von 12 Jahren erreicht wird (vgl. Siegler, DeLoache & Eisenberg, 2008; Montada, 1998; Lohaus, Vierhaus & Maass, 2010; Kuhn & Franklin, 2006). In den niedrigeren Stufen ist das Denken an die konkrete Wahrnehmung gebunden und geprägt durch egozentrisches, statisches, wenig prozesshaftes Denken sowie durch die mangelnde Berücksichtigung mehrerer Dimensionen. Im formal-operatorischen Stadium werden diese Defizite überwunden und die Abstraktionsfähigkeit nimmt zu. Dadurch ist das formal-operatorische Denken gekennzeichnet durch abstraktes, systematisches Denken nach formal-logischen Regeln. Variablen können abstrahiert, isoliert und gedanklich kombiniert werden. Diese Möglichkeit zur Variablenkontrolle gilt als Voraussetzung für die systematische Bildung von Hypothesen und planvolles Experimentieren. Weitere Aspekte der formal-operatorischen Struktur sind proportionales und korrelatives Denken. Das formal-operatorische Denken ermöglicht Denkoperationen zweiter Ordnung und das Denken über das eigene Denken. Piaget sah wissenschaftliches Denken an die Entwicklung der formalen Operationen gebunden. Erst durch die universelle, strukturelle Veränderung des Übergangs von der Stufe des konkret-operatorischen Denkens und der Ausbildung einer einheitlichen Struktur der formalen Operationen sei die Möglichkeit zum hypothetischen und systematischen Schlussfolgern gegeben.

Um die These von Inhelder und Piaget (1958), dass die Fähigkeit zu formalen Operationen eine Voraussetzung für das Verständnis der experimentellen Methodik sei, zu überprüfen, wurde in der LOGIC-Studie (Bullock & Ziegler, 1999; Bullock et al., 2009) ab dem Alter von 12 Jahren zusätzlich zu den Maßen des wissenschaftlichen Denkens ein Test zur Erfassung der formal-operatorischen Fähigkeiten durchgeführt. Es zeigte sich ein Leistungsvorteil der Jungen in den formal-operatorischen Fähigkeiten. Außerdem fanden

sich Zusammenhänge des operatorischen Niveaus zu allen Aufgaben des wissenschaftlichen Denkens, auch wenn Intelligenz ausparialisiert wurde.

Eine Studie unter Einschluss des Grundschulalters wurde bisher noch nicht durchgeführt. Daher gilt es zu klären, ob Zusammenhänge zwischen formal-operatorischen Fähigkeiten bzw. dem operatorischen Niveau und wissenschaftlichem Denken im Grundschulalter festgestellt werden können und inwieweit die Fähigkeit zu formalen Operationen die Kompetenzentwicklung im wissenschaftlichen Denken bedingt.

1.3.3.7 Motivationale Orientierungen

Es gibt bisher wenig Forschung über die Rolle von kompetenzbezogener Motivation bei der Entwicklung wissenschaftlichen Denkens im Grundschulalter.

Motivation kann vereinfacht verstanden werden als „jegliche Form von Handlungsveranlassung“ (Hartinger & Fölling-Albers, 2002, S. 17). Einstellungen, Interesse und Fähigkeitsselbstkonzept sind bedeutsame Variablen der Lern- und Bildungsmotivation. Unter dem Begriff Einstellung ist eine mit einer Wertung verbundene Haltung gegenüber einer Sache, Person oder Idee zu verstehen (Six, 2004). Interesse entsteht durch eine Person-Gegenstand-Beziehung und -Auseinandersetzung und kann durch drei Kennzeichen beschrieben werden: Erkenntnisorientierung, positive Emotionen und Freiwilligkeit bzw. Selbstbestimmtheit (Prenzel, 1994; Krapp, 1998). Im Fähigkeitsselbstkonzept bzw. Selbstkonzept der Begabung ist die eigene Einschätzung der Leistungsfähigkeit bzw. der Kompetenz repräsentiert (Faulstich-Wieland, 2004).

Generell wird von wechselseitigen Beziehungen zwischen Kompetenzentwicklung und kompetenzbezogenen Einstellungen, Fähigkeitsselbstkonzept und Interessen ausgegangen (Weinert, 1999; Krapp, Hidi & Renninger, 1992). Das bedeutet, dass sowohl eine Kompetenzsteigerung durch motivationale Variablen erreicht werden kann als auch eine Steigerung in kompetenzbezogenem motivationalen Einstellungen durch Kompetenzerleben. Als bedeutsamer Mediator wird das Fähigkeitsselbstkonzept angesehen (Faulstich-Wieland, 2004; Krapp, 1998).

Hinsichtlich der Zusammenhänge zwischen Motivation und der Kompetenz im wissenschaftlichen Denken wurden in der TIMS-Studie bei Viertklässlern positive Korrelationen zwischen den motivationalen Variablen Einstellungen zu Naturwissenschaften und Fähigkeitsselbstkonzept und naturwissenschaftlicher Kompetenz, zusammengesetzt aus natur-

wissenschaftlichem Inhaltswissen und dem Verständnis wissenschaftlicher Methoden und Arbeitsweisen, gefunden (Wittwer, Saß & Prenzel, 2008).

Die Beziehung zwischen wissenschaftlichem Denken und kompetenzbezogenem Interesse wurde im Grundschulalter bislang noch nicht differenziert betrachtet. Im Sekundar- schulalter konnten kleine bis mittlere Zusammenhänge zwischen den Leistungen im Test zum wissenschaftlichen Denken und naturwissenschaftlichem Fachinteresse sowie dem Interesse an naturwissenschaftsbezogenen Aktivitäten und Tätigkeiten gezeigt werden (Glug, 2009).

1.3.4 Rahmenmodell

In Abbildung 1.3 ist ein Rahmenmodell der Kompetenz wissenschaftliches Denken dargestellt, das die in den vorangegangenen Abschnitten beschriebenen theoretischen Grundlagen der Kompetenz wissenschaftliches Denken sowie die Beziehungen zu ausgewählten kognitiven und motivationalen Kompetenzaspekten zusammenfasst.

Die Kompetenz wissenschaftliches Denken setzt sich zusammen aus den miteinander verknüpften Komponenten Wissenschaftsverständnis und Methodenkompetenz. Die Kompetenzkomponenten können ihrerseits weiter untergliedert werden in ein Verständnis der wissenschaftlichen Konzepte Hypothese, Theorie, Experiment und der Ziele von Wissenschaften sowie Theorieverständnis bzw. in Experimentierstrategien und Dateninterpretation.

Im Hinblick auf kognitive Fähigkeiten, die der Kompetenz im wissenschaftlichen Denken zugrunde liegen könnten, stehen Leseverständnis und Intelligenz sowie formal-operatorische Fähigkeiten, Problemlösen, inhibitorische Kontrolle und räumliches Denken im Fokus des Interesses.

Bezogen auf motivationale Orientierungen werden die als wechselseitig beschriebenen Beziehungen zwischen wissenschaftlichem Denken und Einstellungen zu Naturwissenschaften, naturwissenschaftsbezogenes Fähigkeitsselbstkonzept sowie Interesse an wissenschaftlichen und forschenden Aktivitäten betrachtet.

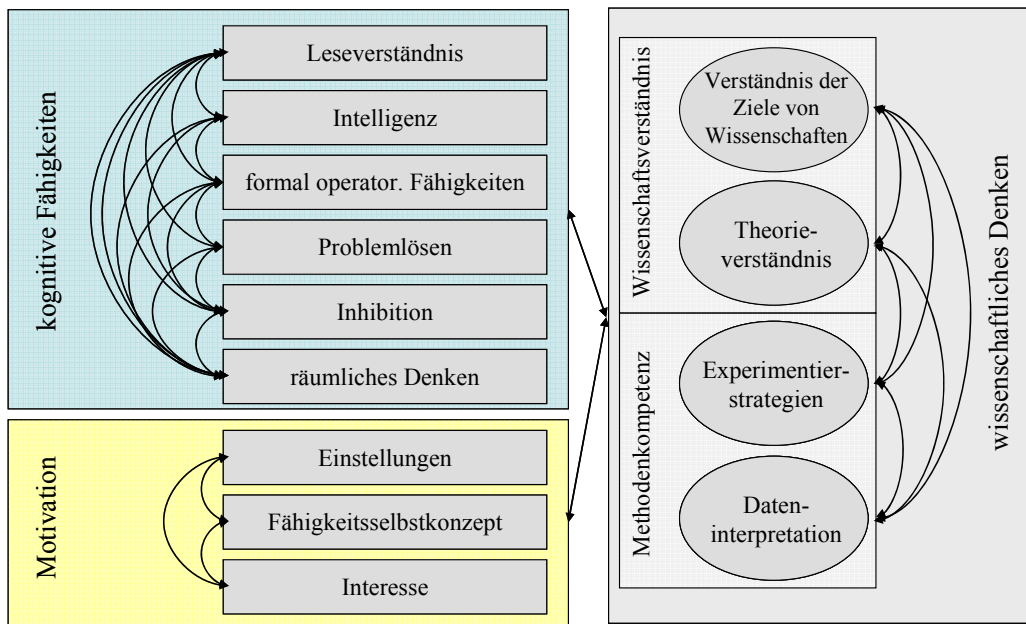


Abbildung 1.3. Rahmenmodell der Kompetenz wissenschaftliches Denken

2 Ziele, Fragestellungen und Hypothesen

Die Ziele der vorliegenden Arbeit waren es a) die latente Struktur der Kompetenz wissenschaftliches Denken im Grundschulalter zu modellieren, b) individuelle Unterschiede im wissenschaftlichen Denken für das Alter von acht bis zehn Jahren sowie frühe Entwicklungsveränderungen zu beschreiben, c) den Zusammenhang zwischen wissenschaftlichem Denken und möglichen zugrunde liegenden kognitiven Fähigkeiten zu untersuchen sowie d) die Beziehung zwischen wissenschaftsbezogenen, motivationalen Orientierungen und der Kompetenz im wissenschaftlichen Denken zu betrachten.

Es wurden folgende Fragestellungen untersucht:

1. Lässt sich die Kompetenz wissenschaftliches Denken als eindimensionale Struktur im Sinne eines einheitlichen, zusammenhängenden Konstrukts oder als mehrdimensionale Struktur von verschiedenen Teilfähigkeiten im Grundschulalter erfassen?
2. Welche Interkorrelationen bestehen im Grundschulalter zwischen den Komponenten bzw. Teilkompetenzen des wissenschaftlichen Denkens und welche Kompetenzprofile lassen sich beschreiben?
3. Wie entwickelt sich die Kompetenz wissenschaftliches Denken im Verlauf der Grundschulzeit von der zweiten zur vierten Jahrgangsstufe? Lassen sich Kompetenzunterschiede zwischen den Jahrgangsstufen und Geschlechterunterschiede im Grundschulalter feststellen? Können Entwicklungssequenzen in den Teilkompetenzen beschrieben werden?
4. Welche Rolle spielen das metakonzeptuelle Theorieverständnis, das Verständnis der experimentellen Methodik und der Dateninterpretation innerhalb der Kompetenzentwicklung im wissenschaftlichen Denken?
5. Welche korrelativen Zusammenhänge bestehen zu anderen kognitiven Fähigkeiten, die einerseits bei der Bearbeitung des schriftlichen Gruppentests zum wissenschaftlichen Denken eine Rolle spielen und andererseits dem wissenschaftlichen Denken zugrunde liegen könnten?

6. Welche korrelativen Zusammenhänge bestehen zu motivationalen Variablen, die einen Einfluss auf die weitere Entwicklung im wissenschaftlichen Denken haben könnten?

Ausgehend von den dargestellten Forschungsergebnissen wurden nachfolgende Hypothesen aufgestellt:

- H1: Wissenschaftliches Denken lässt sich besser als mehrdimensionales Konstrukt von miteinander verknüpften Teilfähigkeiten statt als eindimensionales, einheitliches Konstrukt abbilden.
- H2a: Es bestehen substantielle Zusammenhänge zwischen den Teilkompetenzen des wissenschaftlichen Denkens.
- H2b: Es existieren qualitativ unterschiedliche Kompetenzprofile im Antwortverhalten auf die Teilkompetenzen.
- H3a: Die Kompetenz im wissenschaftlichen Denken steigt von der zweiten zur vierten Klassenstufe an.
- H3b: Es bestehen keine Geschlechterunterschiede.
- H3c: Die einzelnen Teilkompetenzen des wissenschaftlichen Denkens entwickeln sich in folgender Reihenfolge: Experimentierstrategien, Dateninterpretation, Verständnis wissenschaftlicher Konzepte und der Ziele von Wissenschaften, metakonzeptuelles Theorieverständnis.
- H4a: Kinder mit einem ausgeprägten metakonzeptuellen Theorieverständnis erzielen bessere Leistungen in Aufgaben zu Experimentierstrategien und Dateninterpretation.
- H4b: Kinder mit einem ausgeprägten Verständnis der experimentellen Methodik erzielen bessere Leistungen in Aufgaben zu Theorieverständnis und Dateninterpretation.
- H4c: Kinder mit einem ausgeprägten Verständnis der Evidenzevaluation erzielen bessere Leistungen in Aufgaben zu Theorieverständnis und Experimentierstrategien.
- H5a: Leseverständnis und Intelligenz haben einen Einfluss auf die Leistung der im schriftlichen Gruppentest erfassten Kompetenz wissenschaftliches Denken.

H5b: Es besteht ein Zusammenhang zwischen wissenschaftlichem Denken und Problemlösen, Inhibition, räumlichem Denken und formal-operatorischen Fähigkeiten.

H6: Es ergibt sich ein Zusammenhang zwischen wissenschaftlichem Denken und Einstellungen zu Naturwissenschaften, naturwissenschaftsbezogenem Fähigkeitsselbstkonzept sowie dem Interesse an forschenden Aktivitäten.

3. Methode

3.1 Stichprobe

An der Studie nahmen Ende des Schuljahres 2009/2010 insgesamt $N = 285$ Schüler der zweiten, dritten und vierten Jahrgangsstufe der Grundschule aus dem Raum München mit einem durchschnittlichen Alter von 9.55 Jahren ($SD = 0.96$) teil. Es wurden drei Grundschulen nach vergleichbaren Übertrittsquoten an weiterführende Schulen ausgewählt. Davon nahmen drei zweite Klassen mit 72 Schülern (31 Jungen, 41 Mädchen) im durchschnittlichen Alter von 8.21 Jahren ($SD = 0.45$), 58 Drittklässler (35 Jungen, 23 Mädchen) aus drei Klassen mit einem Durchschnittsalter von 9.24 Jahren ($SD = 0.41$) und sieben vierte Klassen mit 155 Schülern (74 Jungen, 78 Mädchen, 3 ohne Angabe) im durchschnittlichen Alter von 10.26 Jahren ($SD = 0.43$) teil. Die Beschreibung der Stichprobe ist in Tabelle 3.1 zusammengefasst.

Tabelle 3.1

Beschreibung der Stichprobe

Jahrgangsstufe	Anzahl Klassen	Anzahl Schüler N	Jungen/Mädchen*	Alter in Jahren $M (SD)$
2	3	72	31/41	8.21 (0.45)
3	3	58	35/23	9.24 (0.41)
4	7	155	74/78	10.26 (0.43)
Gesamt	13	285	140/142	9.55 (0.96)

Anmerkung: N = Anzahl der Schüler, M = Mittelwert, SD = Standardabweichung. *bei drei Kindern fehlte die Angabe des Geschlechts.

3.2 Material

3.2.1 Fragebogen zur Erfassung des wissenschaftlichen Denkens

Das wissenschaftliche Denken wurde anhand einer Auswahl an gruppentestfähigen Fragebogen-Aufgaben aus dem Aufgabenpool des DFG-Projektes Science-P Entwicklung naturwissenschaftlicher Kompetenz in der Grundschule erhoben. Diese Aufgaben, die auf entwicklungspsychologischer Forschung zum wissenschaftlichen Denken aufbauen, waren im Rahmen des Projektes Science-P präpilotiert (Koerber et al., 2009, 2011), mithilfe von Einzelinterviews validiert (Kropf, 2010) und in einer groß angelegten Itempilotierungsstudie auf ihre psychometrischen Eigenschaften überprüft worden. Zur Auswahl der Aufgaben für den in der vorliegenden Studie verwendeten Fragebogen wurden inhaltliche Aspekte sowie statistische Kennwerte (Item-Fit, Aufgabenschwierigkeit, Trennschärfe) der Rasch-Skalierung der Itempilotierung herangezogen. Aus inhaltlichen Gründen sollten in dem zusammengestellten Testheft die beiden postulierten Kompetenzkomponenten Wissenschaftsverständnis und Methodenkompetenz mit ihren untergeordneten Teilkompetenzen Verständnis wissenschaftlicher Konzepte und der Ziele von Wissenschaften sowie Theorieverständnis bzw. Experimentierstrategien und Dateninterpretation repräsentiert sein. Das Testheft bestand aus 20 Items in 18 Aufgaben, wobei jede Seite nur eine einzige Aufgabe enthielt (siehe Anhang A.1). Die Aufgaben waren in einem neutralen, inhaltsfreien Kontext gestellt, um domänenübergreifendes, formal-wissenschaftliches Denken zu erfassen. Die in einem einheitlichen Format dargestellten Aufgaben waren an den Grundschulwortschatz angepasst, durch farbige Bilder illustriert sowie kindgerecht formuliert und aufbereitet.

3.2.1.1 Strukturelle Grundlagen der Itementwicklung

3.2.1.1.1 Kompetenzstruktur

Die Aufgaben des Fragebogens wurden auf Grundlage der in Tabelle 3.2 dargestellten Forschungsstudien zum wissenschaftlichen Denken sowie dem darauf aufbauendem Kompetenzstrukturmodell (vgl. Abbildung 1.3.1, S. 30) entwickelt.

Die Kompetenzkomponente Wissenschaftsverständnis wurde in insgesamt acht Aufgaben erfasst. Fünf Aufgaben erhoben das intuitive Verständnis der Konzepte Hypothese, Theorie und Experiment sowie das Verständnis der Ziele von Wissenschaft bzw. Wissenschaftlern und dessen Arbeitsmethoden (vgl. Nature-of-Science-Interview / Carey et al., 1989; BIQUA / Sodian et al., 2002). Das Theorieverständnis wurde in drei Aufgaben überprüft. In spezifischen Kontexten wurde das metakonzeptuelle Verständnis alternativer Interpretationsrahmen anhand von Interpretationskonflikten erfasst. Beispielsweise stritten zwei Wissenschaftler, die unterschiedlichen Theorien anhängten, über die Genese von Krankheiten (vgl. Hexerei-Interview LOGIC / Bullock et al., 2009). Evidenz sollte aus der Sicht der beiden Wissenschaftler interpretiert werden. Die Kinder sollten die potentiellen Interpretationen eines Wissenschaftlers aufgrund des Zusammenhangs von Theorie und Evidenz beurteilen und Inferenzen ziehen, die mit einem alternativen Theoriegebäude konsistent waren.

Die Kompetenzkomponente Methodenkompetenz erfasste das Verständnis der experimentellen Methodik und der Dateninterpretation in zehn Aufgaben. Sechs Aufgaben untersuchten das Verständnis von Experimentierstrategien. Dazu zählten die Unterscheidung zwischen Hypothesenprüfung und Effektproduktion, die Wahl eines konklusiven Tests (vgl. Sodian et al., 1991), die Wahl eines kontrollierten Experiments (vgl. Flugzeugaufgabe LOGIC / Bullock & Ziegler, 1999) sowie das Verständnis des Aspekts der Basisrate und der Variation der fokalen Dimension beim experimentellen Design (vgl. Design-Interview LOGIC / Bullock, Ziegler & Martin, 1993; Bullock et al., 2009). In vier Aufgaben zur Dateninterpretation wurde das Verständnis von Tabellen, Diagrammen und Kovariationsdaten überprüft. Es wurden Daten in Form von Liniendiagrammen und Tabellen (vgl. Koerber, 2006; Koerber & Sodian, 2008; Koerber & Sodian, 2009) sowie in einfachen Bildern präsentiert (vgl. Kuhn et al., 1988). Anhand dieser Darstellungsformen sollte Evidenz evaluiert und interpretiert werden.

Tabelle 3.2

Übersicht über die Zuordnung der Aufgaben zu Komponenten bzw. Teilkompetenzen des wissenschaftlichen Denkens und über die Grundlagen der Aufgabenentwicklung

Komponente	Teilkompetenz	Aufgaben	Interview-Studien
Wissenschafts- verständnis	Verständnis der Ziele von Wissenschaften	a02 Hypothese a20 Pflanzenmittel a08 Theorie a16 Wissenschaftler a21 Fehler	Nature-of-Science-Interview / Carey, Evans, Honda, Jay, & Unger (1989); BIQUA / Sodian, Thoermer, Kircher, Grygier, & Günther (2002)
	Theorie- verständnis	a23 Wichtel a24 Mittelalter a27 Geburtstagsparty	Hexerei-Interview LOGIC / Bullock, Sodian, & Koerber (2009)
Methoden- kompetenz	Experimentier- strategien	a35 Flugzeug a49 Pizzabacken a51 Maus-Haus a53 Spürnase a44 Ärzte a57 Fußpilz	Flugzeugaufgabe LOGIC / Bullock & Ziegler (1999) Sodian, Zaitchik, & Carey (1991) Design-Interview LOGIC / Bullock, Ziegler & Martin (1993); Bullock et al. (2009)
	Daten- interpretation	a64 Zähne a66 Ravianer a76 Diagramm a82 Tabelle	Kuhn, Amsel & O'Loughlin (1988) Koerber (2006); Koerber & Sodian (2008); Koerber & So- dian (2009)

3.2.1.1.2 Kompetenzniveaus

Auf der Grundlage von entwicklungspsychologischen Befunden (vgl. Abschnitt 1.2 und Tabelle 3.2) und von Kompetenzmodellen zur scientific literacy (z. B. Bybee, 1997), wurde im Projekt Science-P angenommen, dass sich der Entwicklungsverlauf im wissenschaftlichen Denken vereinfachend beschreiben lässt als Übergang von einem Niveau von naiven Vorstellungen über ein Niveau der Zwischenvorstellungen hin zu einem Niveau wissenschaftlich adäquater Vorstellungen. Aufbauend auf den Verständnisebenen von Carey et al. (1989) können die hierarchisch angeordneten Kompetenzniveaus wie folgt beschrieben werden (vgl. Tabelle 3.3):

Auf der untersten Kompetenzstufe handelt es sich um naive Vorstellungen. Diese Konzepte sind nicht mit wissenschaftlichen Vorstellungen und Erklärungsmodellen vereinbar. Auf dem Niveau der naiven Vorstellungen wird Wissenschaft als reine Aktivität oder

Prozedur verstanden. Dabei fehlen das Ziel des Erkenntnisgewinns und der Bezug zu Theorien oder Hypothesen. Experimente werden vorwiegend zur Produktion erwünschter Effekte statt als Methode zur Prüfung von Hypothesen gesehen. In einem graduellen Prozess der Umstrukturierung müssen diese nicht belastbaren Konzepte aufgegeben werden. Wenn Kinder beginnen, ihre vorherrschenden naiven Vorstellungen umzustrukturieren, kann ihr Lernweg sie dabei über sogenannte Zwischenvorstellungen führen. Durch eine Verknüpfung von naiven Vorstellungen und wissenschaftlich belastbaren Erklärungen handelt es sich um Konzepte, die sich im Alltag der Kinder bewähren, und die durch eine Differenzierung in das Niveau der wissenschaftlichen Vorstellungen überführt werden können. Auf dem Niveau der Zwischenvorstellungen besteht ein Verständnis von Wissenschaft als Sammeln von Fakten und Testen von einzelnen Hypothesen. Im Gegensatz zu Niveau 3 fehlt hier die Einbettung einzelner Experimente zur Überprüfung einer Theorie. Das höchste Niveau sind die wissenschaftlichen Vorstellungen. Dabei handelt es sich um Vorstellungen, die mit wissenschaftlichen Erklärungen vereinbar sind. Auf dem Niveau der wissenschaftlichen Vorstellungen wird der zyklische und kumulative Prozess der Bildung, Prüfung und Revision von Theorien und Hypothesen verstanden. Theorien werden als forschungsleitend sowohl bei der Bildung von Hypothesen als auch bei der Interpretation von Daten gesehen.

Tabelle 3.3

Kompetenzniveaus des wissenschaftlichen Denkens (Koerber et al., 2009)

Niveau 1: Naive Vorstellungen	Wissenschaft als reine Aktivität oder Prozedur ohne Ziel des Erkenntnisgewinns und ohne Bezug zu Theorien oder Hypothesen. Experimente werden vorwiegend zur Produktion erwünschter Effekte statt als Methode zur Prüfung von Hypothesen gesehen.
Niveau 2: Zwischenvorstellungen	Wissenschaft als Sammeln von Fakten, Herstellung einfacher Kausalzusammenhänge und Testen von einzelnen Hypothesen. Im Gegensatz zu Niveau 3 fehlt hier die Einbettung einzelner Experimente in den Kontext der Evaluation einer Theorie.
Niveau 3: Wissenschaftliche Vorstellungen	Wissenschaft wird als Suche nach überprüfbaren Erklärungen (Evidenz) verstanden. Es besteht ein Verständnis des zyklischen und kumulativen Prozesses der Bildung, Prüfung und Revision von Theorien und Hypothesen. Theorien werden als forschungsleitend sowohl bei der Bildung von Hypothesen als auch bei der Interpretation von Daten gesehen.

Bei der Aufgabenentwicklung wurde nach dem von Mark Wilson (2005) vorgeschlagenen Ansatz des „Construct Modeling“ vorgegangen. In einer sogenannten Construct Map wurden differenziert für die einzelnen Teilkompetenzen jeweils die drei Kompetenzniveaus

naive Vorstellungen, Zwischenvorstellungen und wissenschaftliche Vorstellungen definiert.

Die Aufgaben und Kompetenzniveaus der Construct Map (siehe Tabelle 3.4) zum Verständnis wissenschaftlicher Konzepte (Hypothese, Theorie, Experiment) und der Ziele von Wissenschaften wurden in Anlehnung an das Nature-of-Science-Interview nach Carey et al. (1989) entwickelt.

Tabelle 3.4

Construct Map der Teilkompetenz Verständnis der Ziele von Wissenschaften

Niveau 1: Naive Vorstellungen	Wissenschaft wird als Beschreibung der Umwelt und als Aktivität, im Sinne von etwas ausprobieren, etwas erfinden, verstanden.
Niveau 2: Zwischenvorstellungen	Wissenschaft wird als Sammeln von Fakten, im Sinne von genau hinschauen und untersuchen, ob es das gibt, verstanden.
Niveau 3: Wissenschaftliche Vorstellungen	Wissenschaft wird als Suche nach überprüfbaren Erklärungen (Evidenz) verstanden. Es werden Hypothesen getestet und kontrollierte Experimente durchgeführt. Es besteht ein Verständnis des zyklischen und kumulativen Prozesses der Bildung, Prüfung und Revision von Theorien und Hypothesen.

Die Aufgaben und Kompetenzniveaus der Construct Map zum Theorieverständnis (vgl. Tabelle 3.5) wurden in Orientierung an das Hexerei-Interview über das Verständnis eines alternativen Interpretationsrahmens der LOGIC-Studie (vgl. Bullock et al., 2009) konstruiert.

Tabelle 3.5

Construct Map der Teilkompetenz Theorieverständnis

Niveau 1: Naive Vorstellungen	Auf diesem Niveau zeigen die Kinder noch kein Verständnis eines alternativen Interpretationsrahmens. Fehlvorstellungen sind einfach, z. B. durch einen Evidenznachweis und Beweisführung, korrigierbar.
Niveau 2: Zwischenvorstellungen	Auf diesem Niveau werden existierende Überzeugungssysteme beibehalten.
Niveau 3: Wissenschaftliche Vorstellungen	Auf diesem Niveau zeigen die Kinder ein Verständnis eines alternativen Interpretationsrahmens. Theorien leiten die Interpretation neuer Information und behindern zum Teil. Die Kinder sind in der Lage, konsistent in einem Überzeugungssystem bzw. Theorierahmen zu argumentieren und neue Informationen werden in das Überzeugungssystem integriert.

Die Aufgaben und die Construct Map zur Teilkompetenz Experimentierstrategien (vgl. Tabelle 3.6) basierten ebenfalls auf Interviewstudien und fokussierten auf das Verständnis

der Hypothesenprüfung und der Unterscheidung zwischen einem inkonklusiven und einem konklusiven Test (vgl. Sodian et al., 1991), sowie auf das Verständnis der Variablenkontrolle und des Aspekts der Basisrate im experimentellen Design (vgl. Flugzeugaufgabe LOGIC / Bullock & Ziegler, 1999; Design-Interview LOGIC / Bullock et al., 1993).

Tabelle 3.6

Construct Map der Teilkompetenz Experimentierstrategien

Niveau 1: Naive Vorstellungen	Die Kinder verstehen die Logik der Hypothesenprüfung und des experimentellen Designs (z. B. Kontrolle von Variablen, Berücksichtigung der Basisrate) noch nicht. Auf diesem Niveau wird zur Hypothesenprüfung kein systematisches Experiment durchgeführt, stattdessen werden Effekte produziert.
Niveau 2: Zwischenvorstellungen	Die Kinder können einen schlüssigen Test nicht von einem un-schlüssigen Test unterscheiden. Zur Überprüfung einer Hypothese wird nur ein kontrastiver Test ausgewählt, indem zwar die fokale Variable variiert wird, die anderen Variablen allerdings nicht konstant gehalten werden. Weitere Aspekte der Logik des experimentellen Designs, wie z. B. die Berücksichtigung der Basisrate, werden im Versuchsaufbau nicht beachtet.
Niveau 3: Wissenschaftliche Vorstellungen	Zur Überprüfung einer Hypothese wird ein schlüssiger Test, ein kontrolliertes Experiment gewählt. Dabei wird die fokale Variable manipuliert, während gleichzeitig alle anderen Variablen konstant gehalten werden. Weitere Aspekte des experimentellen Designs, wie z. B. die Berücksichtigung der Basisrate, werden im Versuchsaufbau beachtet.

Die Aufgaben und die Construct Map zur Teilkompetenz Dateninterpretation (vgl. Tabelle 3.7) wurden entwickelt in Orientierung an Interviewstudien von Kuhn und Kollegen (1988) zur Beachtung der Relevanz und Evaluation kovariierender Daten präsentiert in einfachen Bildern sowie in Anlehnung an experimentelle Aufgaben nach Koerber und Sodian (2008, 2009) zur Interpretation von Daten aus den Repräsentationsformen Tabelle und Diagramm.

Tabelle 3.7

Construct Map der Teilkompetenz Dateninterpretation

Niveau 1: Naive Vorstellungen	Fehlendes Verständnis von Tabellen, Diagrammen und Kovariationsdaten. Die Kinder können nicht zwischen Theorien und Evidenz differenzieren. Es treten Urteilsverzerrungen durch Vorwissen auf, anstatt einfache Datenmuster allein aufgrund vorliegender Daten zu interpretieren.
Niveau 2: Zwischenvorstellungen	Fehlerhaftes Verständnis von Tabellen, Diagrammen und Kovariationsdaten. Es leiten sowohl vorliegende Daten (Evidenz) als auch Vorwissen (Theorie) die Evidenzevaluation.
Niveau 3: Wissenschaftliche Vorstellungen	Es besteht ein Verständnis von Tabellen, Diagrammen und Kovariationsdaten. Auf diesem Niveau kann zwischen Theorie und Evidenz differenziert werden. Bei der Interpretation von Datenmustern treten keine Urteilsverzerrungen durch Vorwissen auf. Allein aufgrund vorliegender Daten werden Schlussfolgerungen gezogen.

3.2.1.1.3 Antwortformate

Die Antwortalternativen der Forced-Choice- und Multiple-Choice-Aufgaben entsprachen den verschiedenen Kompetenzniveaus naive Vorstellungen, Zwischenvorstellungen und wissenschaftliche Vorstellungen. Bei Forced-Choice-Aufgaben, welche zwei Kompetenzniveaus direkt gegenüberstellten, sollte die bessere von zwei vorgegebenen Antwortalternativen gewählt werden. Bei Multiple-Choice-Aufgaben sollte die beste der drei Antwortalternativen angekreuzt werden. Dieses Format deckte simultan die gesamte Spannbreite der drei angenommenen Kompetenzniveaus ab. Bei ja/nein-Items konnte eine naive Vorstellung angenommen oder abgelehnt werden. Bei dem vierten Aufgabentypus war ein ja/nein-Item mit einem Item im offenen Antwortformat kombiniert. Zuerst musste in einer ja/nein-Frage eine naive Vorstellung angenommen oder abgelehnt werden. In einem zweiten Schritt sollten die Kinder schriftlich eine kurze Erklärung ihrer gegebenen Antwort geben.

Eine Übersicht über die Verteilung der Aufgaben bzw. Items auf Komponenten und Aufgabenformate findet sich in Tabelle 3.8.

Tabelle 3.8

Anzahl und Verteilung der Items auf Teilkompetenzen und Aufgabenformate

Komponente	Subkomponente	Aufgabenformat				Anzahl
		Forced-Choice	Multiple-Choice	ja/nein	offen	
Wissenschaftsverständnis	Verständnis der Ziele von Wissenschaften	2	3	--	--	5
	Theorieverständnis	--	3	--	--	3
Methodenkompetenz	Experimentierstrategien	--	4	2	2	8
	Dateninterpretation	1	2	1	--	4
Anzahl		3	12	3	2	20

3.2.1.2 Aufgaben*Verständnis wissenschaftlicher Konzepte und der Ziele von Wissenschaften*

Die Aufgabe *a02_Hypothese* überprüfte das Verständnis von Hypothesen in Form eines Forced-Choice-Items. Als Einleitung wurde eine kurze Erklärung des Begriffes Hypothese gegeben. Als Antwortalternativen wurden eine Zwischenvorstellung, im Sinne von Sammeln von Fakten („Es gibt viele Arten von Walen“), und eine wissenschaftliche Vorstellung entsprechend dem Verständnis von Wissenschaft als Suche nach überprüfbaren Erklärungen („Sonnenblumen, die gedüngt werden, werden größer“) gegenübergestellt.

Die Aufgabe *a08_Theorie* untersuchte das intuitive Verständnis des Konzepts Theorie anhand des Beispiels unterschiedlicher Ideen darüber, warum ein Badesee warm ist. In der Multiple-Choice-Aufgabe wurden drei Antwortmöglichkeiten auf den drei Kompetenzniveaus zur Auswahl angeboten. Die Antwortalternative auf dem Niveau der naiven Vorstellung gab eine Erklärung, die eine reine Beschreibung der Umwelt („weil es Juli und Hochsommer ist“) darstellte. Auf dem Niveau der Zwischenvorstellung wurde eine Sammlung von Fakten („weil es die letzten drei Wochen nicht geregnet hat“) als Erklärung dargeboten. Die wissenschaftliche Vorstellung erklärte den Zusammenhang zwischen Sonne und der Wärme des Badesees durch einen Erklärungsmechanismus („weil die Sonne geschienen hat und der See die Wärme speichert“).

In der Aufgabe *a16_Wissenschaftler* wurde der Frage nach den Zielen von Wissenschaft bzw. Wissenschaftlern nachgegangen. Diese Multiple-Choice-Aufgabe erforderte die Wahl der besten der drei Antwortalternativen auf den beschriebenen Kompetenzniveaus. Auf dem Niveau der naiven Vorstellung wurde Wissenschaft als reine Aktivität und Prozedur verstanden ohne das Ziel des Erkenntnisgewinns (Winterreifen werden getestet, damit sie kein Loch haben). Die Zwischenvorstellung beschrieb die Tätigkeit von Wissenschaftlern als das Sammeln von Fakten (Es wird getestet, wie viel Zeit Autos mit Winterreifen benötigen, bis sie eine Geschwindigkeit von 100 km/h erreichen). Auf der Stufe der wissenschaftlichen Vorstellung wurde erkannt, dass Wissenschaftler auf der Suche nach überprüfbaren Erklärungen sind (Wissenschaftler führen Experimente durch, um beispielsweise herauszufinden, welche Winterreifen die beste Rutschfestigkeit aufweisen).

In einer weiteren Aufgabe *a20_Pflanzmittel* wurden die experimentelle Überprüfung der Hypothese und das Ergebnis des Experiments beschrieben. Im Folgenden wurde nach den weiteren Arbeitsmethoden von Wissenschaftlern gefragt. In einem Forced-Choice-Format wurden zwei Antwortalternativen auf den Niveaus der naiven Vorstellung und der Zwischenvorstellung dargeboten. Die naive Vorstellung entsprach einem Wissenschaftsverständnis als Aktivität, wonach nach einer erfolgreichen Effektproduktion durch das Experiment kein weiterer Handlungsbedarf bestand. Auf dem Niveau der Zwischenvorstellung ging es um die Replikation des Ergebnisses, um Fehler bei der Versuchsdurchführung auszuschließen. Es ging weniger um den zyklischen Charakter von Wissenschaft und Hypothesenprüfung durch ein gezieltes Experiment in dem Kontext der Evaluation einer Theorie, was einer wissenschaftlichen Vorstellung entspräche.

In der letzten Aufgabe zum Verständnis wissenschaftlicher Konzepte und der Ziele von Wissenschaften, Aufgabe *a21_Fehler*, wurde die Frage nach Fehlern von Wissenschaftlern gestellt. Die drei Kompetenzniveaus waren in den drei Antwortalternativen des Multiple-Choice-Items realisiert. In der naiven Vorstellung wurden Wissenschaftler als Autorität dargestellt, die aufgrund ihres großen Wissens im Normalfall keine Fehler im Wissenschaftsprozess begehen. Auf dem Niveau der Zwischenvorstellung konnte es bei Wissenschaftlern zu Fehlern kommen, wenn beim Sammeln von Fakten fehlerhafte Geräte verwendet wurden. Bei der wissenschaftlichen Vorstellung wurde erkannt, dass Theorien forschungsleitend sind, sowohl bei der Hypothesenprüfung als auch bei der Interpretation von Daten, so dass fehlerhafte Theorien zu falschen Schlussfolgerungen führen können.

Theorieverständnis

In der Aufgabe *a23_Wichtel* wurde ein Interpretationskonflikt über die Heilung von Krankheiten aufgebaut, dadurch dass die Kinder aufgrund ihres Vorwissens, in der Aufgabe präsentiert durch das Faktum Kräuter, bereits wussten, dass die in der Aufgabe dargestellte Sichtweise der Heilung durch Wichtel in der heutigen Zeit falsch war. In einem Multiple-Choice-Item sollte die beste der drei Antwortalternativen, die potentielle Sichtweisen der Menschen von damals darstellten, gewählt werden. Die wissenschaftliche Vorstellung entsprach einer mit dem damaligen Theoriegebäude konsistenten Interpretation (Die Wichtel heilen Krankheiten). Auf dem Niveau der Zwischenvorstellung wurde das bestehende Überzeugungssystem einfach beibehalten (Sowohl Wichtel als auch Kräuter sind für die Genesung verantwortlich). Die naive Vorstellung bildete die einfache Übernahme der heutigen Sichtweise (Krankheiten können durch Kräuter geheilt werden).

In der Aufgabe *a24_Mittelalter* wurde ebenfalls ein Interpretationskonflikt zwischen zwei Wissenschaftlern über die Ursache von Krankheiten geschildert. Einerseits wurde die mittelalterliche Denkweise von Hexen als Verursacher von Krankheiten dargestellt, die den Kindern aufgrund ihres Vorwissens bereits als falsch bekannt war. Andererseits wurde durch den Befund eines in der Aufgabe präsentierten Experiments das in der heutigen Zeit vorherrschende und somit dem Wissen der Kinder entsprechende Verständnis von Bakterien als Krankheitserreger beschrieben. Das Multiple-Choice-Item erforderte die Wahl der besten der drei Antwortalternativen, die die potentiellen Interpretationen des mittelalterlichen Wissenschaftlers repräsentierten. Die Alternativen entsprachen der naiven Vorstellung, dass Fehlvorstellungen durch einfachen Evidenznachweis korrigierbar sind, der Zwischenvorstellung, dass Überzeugungssysteme beibehalten werden, sowie der wissenschaftlichen Vorstellung, im Sinne einer mit dem alternativen Theoriegebäude konsistenten Interpretation.

Die Aufgabe *a27_Geburtstagsparty* überprüfte das Verständnis alternativer Interpretationsrahmen in einem sozialen Kontext. Es wurde ein Interpretationskonflikt zweier Schwestern über die Ursache von Ketchupflecken auf ihren Kleidern dargestellt. Die naive Vorstellung sowie die Zwischenvorstellung der Multiple-Choice-Aufgabe stellten eine einseitige Betrachtung des Problems dar. Die wissenschaftliche Vorstellung berücksichtigte beide Sichtweisen des Interpretationskonflikts.

Experimentierstrategien

Die Aufgabe *a35_Flugzeug* untersuchte das Verständnis der Variablenkontrollstrategie. Ein Flugzeugkonstrukteur hatte den Auftrag, Flugzeuge mit möglichst wenig Treibstoffverbrauch zu bauen. Dabei konnten beim Flugzeug die Variablendimensionen Nase, Höhenruder oder Flügel variiert werden. Zur Überprüfung des Effekts der Höhenruder auf den Treibstoffverbrauch wurden in den drei Antwortalternativen der Multiple-Choice-Aufgabe verschiedene Versuchsanordnungen dargestellt: Die Effektproduktion sowie die unsystematische Variation von Variablen entsprachen Fehlvorstellungen. Bei der wissenschaftlichen Vorstellung wurde hingegen ein kontrolliertes Experiment durchgeführt, in dem die fokale Variable manipuliert und gleichzeitig alle anderen Variablen konstant gehalten wurden.

Bei der Aufgabe *a44_Ärzte* im offenen Aufgabentypus ging es um das Verständnis der Variation der fokalen Variablen. Zwei Ärzte untersuchten die Wirkung zweier Tabletten. Allerdings wurde nur der Effekt einer Tablette, welche die Probanden heilte, untersucht. Die Kinder sollten durch die Zustimmung oder Ablehnung dieser naiven Vorstellung entscheiden, ob es sich dabei um ein gutes Experiment handelte. In dem anschließenden Item des offenen Antwortformats sollten die Kinder ihre Antwort begründen. Diese offene Antwort wurde anhand eines Kodiermanuals (vgl. Anhang A.2) danach beurteilt, ob die Kinder erkannt hatten, dass in dem Experimentaufbau die Manipulation der fokalen Variable fehlte, wonach beide Tabletten auf ihre Wirkung hin untersucht hätten werden müssen.

Die Aufgabe *a49_Pizzabacken* betrachtete den Aspekt der Variation der fokalen Dimension. Es sollte die Wirkung der Hefe auf den Pizzateig untersucht werden. In den Antwortalternativen der Multiple-Choice-Aufgabe wurden verschiedene Versuchsanordnungen beschrieben. Auf dem Niveau der naiven Vorstellung wurde ein Effekt produziert, indem ein Pizzateig nach einem Rezept gebacken wurde. Das wiederholte Mischen verschiedener Zutaten, ohne dabei die fokale Variable Hefe zu berücksichtigen, stellte das Niveau der Zwischenvorstellung dar. Die systematische Manipulation der fokalen Variablen bei gleichzeitigem Konstanthalten der anderen Variablen entsprach einer wissenschaftlichen Vorstellung.

Die Aufgaben *a51_Maus Haus* und *a53_Spürnase* untersuchten den Aspekt der Hypothesenprüfung und die Unterscheidung zwischen einem inkonklusiven und konklusiven Test. In der Aufgabe *a51_Maus Haus* wollten zwei Mädchen herausfinden, ob die Maus in ihrem Keller dick oder dünn ist. Dafür stellten sie ein Häuschen mit einem großen Eingang

auf und legten Käse hinein, der in der Nacht von der Maus gefressen wurde. Die Kinder sollten nun beurteilen, ob dies ein guter Versuch war, um die Größe der Maus herauszufinden. Die Zustimmung mit der Begründung, dass die Maus etwas zu essen bekommen hat, entspricht der Effektproduktion der naiven Vorstellung. Die Zwischenvorstellung stellte die Zustimmung und die Wahl eines inkonklusiven Tests dar. Auf dem Niveau der wissenschaftlichen Vorstellung wurde der Versuch aufgrund eines unschlüssigen Tests als schlecht abgelehnt.

In der Aufgabe *a53_Spürnase* wollten zwei Mädchen anhand einer Kiste, schwach riechendem Brot und stark riechendem Käse herausfinden, ob ihr neues, seltenes Haustier eine gute Spürnase hat. Die Kinder sollten von drei Möglichkeiten die beste auswählen. Die Antwortalternative „Käse auf der Kiste“ stellte als Effektproduktion die naive Vorstellung dar. Die Antwort „Käse in der Kiste“ beschrieb als inkonklusiven Test das Niveau der Zwischenvorstellung. Die Möglichkeit „Brot in der Kiste“ stellte als schlüssiger Test die wissenschaftliche Vorstellung dar.

Die Aufgabe *a57_Fußpilz* überprüfte das Verständnis der Basisrate im experimentellen Design. Bei der beschriebenen Untersuchung zur Wirkung einer Tablette gegen Fußpilz wurde die Basisrate nicht berücksichtigt. In einem ja/nein-Item sollten die Kinder durch die Annahme oder Ablehnung dieser naiven Vorstellung des experimentellen Designs beurteilen, ob es sich um eine gute Untersuchung handelte. In einem zweiten Schritt sollten die Kinder ihre Antwort schriftlich begründen. Diese offene Antwort wurde anhand eines Kodiermanuals (vgl. Anhang A.2) danach beurteilt, ob die Kinder erkannt hatten, dass in dem Experimentaufbau die Beachtung der Basisrate fehlte, in diesem Falle die Registrierung des Vorhandenseins oder Fehlens von Fußpilz bei den Probanden vor der Untersuchung.

Dateninterpretation

Die Aufgaben *a64_Zähne* und *a66_Ravianer* überprüften das Verständnis von kovariierenden Daten. In der Aufgabe *a64_Zähne* wurden den Kindern in einer Internatsgeschichte zwei Dimensionen von je zwei Variablen (Gemüse: Maiskekse oder Karotten, Süßigkeiten: Kaugummi oder Schokolade) präsentiert, die einen Einfluss auf den Zahnzustand der Kinder im Internat hatten. Es wurde ein Datenmuster mit konfundierenden Variablen gezeigt. In diesem kovariierenden Datenbeispiel beeinflussten zwei Faktoren, Maiskekse und Schokolade, den Gesundheitszustand der Zähne. In der Aufgabe hatten die Forscher herausgefunden, dass sie anhand der vorliegenden Daten nicht wissen, ob die Schokolade

die Ursache für schlechte Zähne ist. Die Kinder sollten nun ebenfalls anhand der Daten die Aussage der Forscher begründen. Dazu wurden in der Multiple-Choice-Aufgabe drei Antwortalternativen dargeboten. Auf dem naiven Niveau war die Beurteilung durch das Vorwissen der Kinder, dass Schokolade schlecht für Zähne ist, beeinflusst. Bei der Zwischenvorstellung leiteten zwar vorliegende Daten die Evidenzevaluation, allerdings wurde nur ein Teilaspekt der Daten, in diesem Fall der Einfluss der Maiskekse auf den Zahnzustand, betrachtet. Die wissenschaftliche Vorstellung entsprach der korrekten Interpretation, dass aufgrund der Datenkovariation (Maiskekse und Schokolade stehen im Zusammenhang mit schlechten Zähnen) keine eindeutige Schlussfolgerung für die Ursache schlechter Zähne gezogen werden konnte.

In der Aufgabe *a66_Ravianer* wurden Kovariationsdaten in einer fiktiven Geschichte in einem artifiziellen Kontext präsentiert, so dass die vorliegenden Daten nicht mit dem Vorwissen der Kinder interferierten und allein evidenzbasiert evaluiert werden konnten. Auf einem fremden Planeten machten Steine, die in je zwei Ausprägungen der Variablen-dimensionen Farbe, Größe, Form und Oberfläche variierten, die Einwohner des Planeten Ravianer entweder fröhlich oder traurig. Forscher ordneten anhand eines Experiments die Steine auf zwei Tischen danach an, ob sie fröhlich oder traurig machten. Die Forscher kamen in der Aufgabe zu dem Ergebnis, dass sie anhand der vorliegenden Daten nicht wissen, woran es lag, dass die Ravianer fröhlich oder traurig wurden. Die Kinder sollten nun anhand der Daten entscheiden, wie die Forscher ihr Ergebnis begründen würden, und eine der drei Antwortalternativen auswählen. Die Alternative auf dem Niveau der naiven Vorstellungen zielte auf eine Erklärung ab, die unabhängig von der Datenlage sondern aufgrund des Verletzungspotenzials von rauen Steinen gegeben wurde. Bei der Zwischenvorstellung wurde jeweils nur ein Aspekt (Farbe oder Oberfläche der Steine) des Datensatzes betrachtet. Auf dem Niveau der wissenschaftlichen Vorstellung wurde erkannt, dass aufgrund des kovariierenden Datenmusters (Farbe und Oberfläche) keine eindeutige Schlussfolgerung getroffen werden konnte.

In der Aufgabe *a76_Diagramm* wurde das Verständnis der Evaluation von Daten, dargestellt in Form eines Diagramms, untersucht. In dieser Aufgabe wurde ein linearer Zusammenhang („Je weniger Zeit ich mit Fernsehen verbringe, desto weniger Filmtitel kann ich mir merken“) formuliert und anhand eines Graphen dargestellt. Die Kinder sollten in einem ja/nein-Item beurteilen, ob der Graph den beschriebenen Zusammenhang zeigte. In diesem Fall handelte es sich bei dem Graphen um eine fehlerhafte, naive Darstellungsform,

da statt einer Zusammenhangsgeraden nur ein Punkt in das Koordinatensystem eingezeichnet war.

Die Aufgabe *a82_Tabelle* untersuchte das Verständnis von Tabellen. In der Aufgabe wurde eine Null-Korrelation („Die Löcher in den Zähnen haben nichts damit zu tun, ob man Gummibärchen isst oder nicht.“) beschrieben. Anhand zweier Tabellen wurden unterschiedliche Datenlagen präsentiert. Es sollte beurteilt werden, welche Tabelle die beschriebene Null-Korrelation repräsentiert. Tabelle 1 entsprach einer naiven Vorstellung, da hier ein positiver Zusammenhang zwischen der Anzahl der Löcher und der Anzahl der Gummibärchen entsprechend dem Vorwissen der Kinder dargestellt wurde. In Tabelle 2 hingegen wurde die beschriebene Null-Korrelation zwischen dem Vorhandensein von Löchern und dem Essen von Gummibärchen abgebildet und stellte daher das Niveau der wissenschaftlichen Vorstellung dar.

3.2.1.3 Durchführung

Alle Aufgaben des Testheftes zum wissenschaftlichen Denken wurden vom Versuchsleiter aus den Testanweisungen wörtlich vorgelesen und mit einer Powerpoint-Präsentation visualisiert. Die Kinder wurden gebeten, die Antwortalternativen im Forced-Choice-, Multiple-Choice- oder ja/nein-Format anzukreuzen bzw. gegebenenfalls bei einem offenen Format eine Kurzantwort zu formulieren. Das Umblättern nach der Bearbeitung einer Aufgabe erfolgte auf Anweisung der Versuchsleiterin im Klassenverband. Ein bis zwei zusätzliche Schreibassistentinnen achteten darauf, dass jedes Kind eigenständig jeweils die entsprechende Testaufgabe beantwortete. Sie standen bei Verständnisfragen und Fragen zur Verschriftlichung zur Verfügung.

3.2.1.4 Kodierung, fehlende Werte und Skalierung

Kodierung

Da die Grundschul Kinder das Niveau der Zwischenvorstellung selten wählten, wurde für die nachfolgenden Analysen jedes Item separat dichotom (0 = falsch, 1 = richtig) bepunktet: Im Falle der Forced-Choice-Aufgaben wurde die Wahl der Antwortalternative bepunktet, die dem höheren Kompetenzniveau entsprach (dichotome Kodierung, 0-1). Bei den

Multiple-Choice-Aufgaben wurde die Wahl der Antwortalternative, die die wissenschaftlichen Vorstellung darstellte, für die Bepunktung herangezogen (dichotome Kodierung: naive Vorstellung, Zwischenvorstellung = 0, wissenschaftliche Vorstellung = 1). Bei ja/nein-Aufgaben wurde die Ablehnung der naiven Vorstellung mit einem Punkt und die Annahme der naiven Vorstellung mit 0 Punkten bewertet. Die offenen Antworten wurden nach einem Kodiermanual (vgl. Anhang A.2) als richtig (= 1 Punkt) oder falsch (= 0 Punkte) kodiert.

Umgang mit fehlenden Werten

Fehlende Werte kommen dadurch zustande, dass Testitems nicht beantwortet werden. Nach Lord (1980) können folgende fehlende Werte klassifiziert werden: omitted = Items werden versehentlich oder absichtlich ausgelassen, not reached = am Ende stehende Testitems werden aus zeitlichen Gründen und/oder aufgrund einer Schwierigkeitszunahme nicht beantwortet, not administered (missings by design) = der Testperson wurden beispielsweise aufgrund eines Multimatrix-Testdesigns nicht alle Aufgaben vorgelegt. Der Test zum wissenschaftlichen Denken wurde als Gruppentest administriert. Den Versuchspersonen wurde je Aufgabe genügend Zeit zur Bearbeitung gegeben. Der Testleiter ging erst dann zur nächsten Aufgabe über, wenn alle Kinder die entsprechende Aufgabe bearbeitet hatten. Dennoch konnte es zu fehlenden Werten kommen. Pro Item wurde nur ein Anteil fehlender Werte unter dem 5%-Kriterium akzeptiert (Lüdtke, Robitzsch, Trautwein & Köller, 2007). Da für alle Items der Anteil fehlender Werte unter diesem 5%-Kriterium blieb (vgl. Tabelle 4.5, S. 89), musste kein Item ausgeschlossen werden. Für die Rasch-Analysen wurden diese Werte als fehlend aufgenommen und in der Berechnung der Personenparameter berücksichtigt. Bei der Bildung von Summenwerten wurden fehlende Werte als falsche Antwort mit 0 Punkten bewertet.

Skalierung

Zur psychometrischen Leistungsmessung der Variable wissenschaftliches Denken wurde das dichotome Rasch-Modell nach Georg Rasch (1960) verwendet. Eine genauere Beschreibung des psychometrischen Modells findet sich in Abschnitt 3.4.1 (S. 71).

3.2.2 Fragebögen zur Erfassung ausgewählter kognitiver Fähigkeiten

Zur Erfassung der kognitiven Variablen kamen standardisierte, psychologische Tests zum Einsatz. Des Weiteren wurden Tests verwendet, die aus bereits bestehenden Testverfahren adaptiert und in Vorstudien vorgetestet worden waren, um eine gruppentestfähige Testbatterie im Multiple-Choice-Format zu Erfassung kognitiver Variablen für das Grundschulalter zu erhalten (siehe Anhang A).

3.2.2.1 Leseverständnis

Das Leseverständnis wurde mit dem Untertest Textverständnis des gruppentestfähigen Lesetest ELFE 1-6 Ein Leseverständnistest für Erst- bis Sechstklässler (Lenhard & Schneider, 2006) erhoben. Dieser Untertest erfasste das Leseverständnis auf Textniveau. Bei jeder der 20 Aufgaben wurden ein Text sowie eine dazugehörige Frage dargeboten. Aus den vier möglichen Antworten im Multiple-Choice-Format sollte jeweils die richtige Alternative ausgewählt werden. Die Bearbeitungszeit war mit einem Abbruchkriterium von sieben Minuten bei den Klassen 1 bis 4 begrenzt. Jede richtige Antwort erhielt einen Punkt. Fehlende Werte aufgrund absichtlichen oder versehentlichen Auslassens sowie Nicht-Erreichens aus zeitlichen Gründen und/oder aufgrund einer Schwierigkeitszunahme wurden mit 0 Punkten bewertet. Die erzielten Punkte wurden zu einem Rohwert (Cronbachs $\alpha = 0.90$) summiert bzw. raschskaliert (EAP/PV-Reliabilität = 0.90).

3.2.2.2 Intelligenz

Die allgemeine Intelligenz wurde anhand zweier Untertests des CFT 20-R (Weiß, 2006) im Gruppentest erfasst. Die sprachfreien Aufgaben bestanden aus figuralen Darstellungen. Die Antworten wurden im Multiple-Choice-Format als fünf Alternativen präsentiert. Die Probanden trugen die Antworten in einem Antwortbogen ein. Der Subtest 1 Reihenfortsetzen bestand aus 15 Items zum Bereich Erkennung von Gesetzmäßigkeiten und Regelmäßigkeiten. In einer Bearbeitungszeit von fünf Minuten sollte jeweils aus fünf vorgegebenen Figu-

ren die Alternative erkannt werden, die eine Reihe von Mustern korrekt ergänzte. Der Subtest 4 topologische Schlussfolgerungen behandelte schlussfolgerndes Denken in elf Aufgaben. Aus jeweils fünf Auswahlfiguren sollte bei einer Bearbeitungsdauer von vier Minuten die Alternative ausgewählt werden, bei der ein Punkt genauso liegen konnte wie beim präsentierten Bild. Jede richtige Antwort erhielt einen Punkt. Fehlende Werte aufgrund absichtlichen oder versehentlichen Auslassens sowie Nicht-Erreichens aus zeitlichen Gründen und/oder aufgrund einer Schwierigkeitszunahme wurden mit 0 Punkten bewertet. Die in beiden Untertests erzielten Punkte wurden zu einem Rohwert summiert (Cronbachs $\alpha = 0.61$) bzw. in einem Rasch-Modell skaliert (EAP/PV-Reliabilität = 0.62).

3.2.2.3 Problemlösen

Um strategisch-planerisches Denken zu erheben, wurde eine gruppentestfähige Papierversion aus dem Test TL-D Turm von London – Deutsche Version (Tucha & Lange, 2004) abgeleitet. Die Originalversion des Tests Turm von London zur Erhebung konvergenten, problemlösenden Denkens besteht aus Transformationsaufgaben, die „komplexe Planungsprozesse, bei denen eine Vielzahl möglicher Handlungsoptionen erkannt und in der Vorstellung auf ihre Brauchbarkeit hinsichtlich des erwünschten Zielzustandes geprüft werden muss“ (Tucha & Lange, 2004, S. 16), erfassen. In Form eines Holzmodells wird eine Konstellation von drei verschiedenfarbigen Kugeln (rot, gelb, blau) auf drei vertikalen, nebeneinander angeordneten Stäben unterschiedlicher Länge vorgegeben. Auf den Stäben haben entweder eine, zwei oder drei Kugeln Platz. Die Aufgabe ist es, die Kugeln von dem Ausgangszustand in einen vorgegeben Zielzustand zu bringen. Die Anzahl der erforderlichen Züge ist vorgegeben und variiert je nach Schwierigkeitsgrad der Aufgabe von drei bis sechs Züge. Je Zug darf nur eine Kugel bewegt werden. In der adaptierten Version wurde in acht Aufgaben jeweils in einem Startbild der Ausgangszustand der Kugeln und in einem Zielbild die Endposition der Kugeln dargeboten. Es wurde die Anzahl der Züge, je zwei Aufgaben zu drei, vier, fünf und sechs Züge, angegeben und eine Reihe von Bildern mit den einzelnen Schritten sowie eines Distraktors dargestellt. Die Aufgabe war es jeweils, vom Startbild bis zum Zielbild die Bilder in die richtige Reihenfolge zu bringen und die Bilder in der entsprechenden Reihenfolge zu nummerieren. Nach der gemeinsamen Lösung einer Beispielaufgabe wurde zur selbständigen Bearbeitung der Aufgaben eine Dauer von 20 Minuten festgesetzt. Jede richtige Antwort erhielt einen Punkt. Fehlende Werte auf-

grund absichtlichen oder versehentlichen Auslassens sowie Nicht-Erreichens aus zeitlichen Gründen und/oder aufgrund einer Schwierigkeitszunahme wurden mit 0 Punkten bewertet. Die erzielten Punkte wurden zu einem Rohwert summiert (Cronbachs $\alpha = 0.67$) bzw. raschskaliert (EAP/PV-Reliabilität = 0.79).

3.2.2.4 Inhibition

Zur Erfassung der Inhibitionsleistung wurde der FWIT Farbe-Wort-Interferenztest nach J. R. Stroop (Bäumler, 1985) zu einer gruppentestfähigen Paper-Pencil-Version modifiziert. Der Originaltest FWIT erfasst im mündlichen Einzelverfahren mittels des Farbe-Wort-Inkongruenzprinzips nach J. R. Stroop die selektive Aufmerksamkeit- und Konzentrationsleistung, dadurch dass „konzentrativer Widerstand gegenüber dominierenden Reaktions-tendenzen“ (Bäumler, 1985, S. 7) geleistet werden muss. Der FWIT besteht aus drei Subtests: Lesen von Farbwörtern, Benennen der Farben von Farbstrichen und Benennen von Farben bei Farbe-Wort-Inkongruenz (sogenannter Interferenzversuch). In der adaptierten Testversion mussten ebenfalls drei Arbeitsbögen bearbeitet werden. Dazu wurde jeweils eine Bearbeitungszeit von einer Minute gegeben. Statt das Farbwort vorzulesen musste die Farbe des in schwarz gedruckten Farbwortes in dem entsprechenden Farbkästchen daneben angestrichen werden. Statt die Farbe eines Farbstrichs zu benennen sollte die Druckfarbe des Wortes in dem entsprechenden Farbkästchen angestrichen werden. In der Interferenzbedingung war es die Aufgabe, die Farbe des Farbwortes, das in einer inkongruenten und dadurch interferierenden Farbe gedruckt war, in dem entsprechenden Farbkästchen anzustreichen. Zur Messung der Inhibitionsleistung wurden die Antworten auf den dritten Arbeitsbogen Interferenzbedingung herangezogen. Aus Varianzgründen wurden die ersten 40 Items des Antwortbogens bewertet. Für jede richtige Antwort wurde ein Punkt vergeben. Fehlende Werte aufgrund absichtlichen oder versehentlichen Auslassens sowie Nicht-Erreichens aus zeitlichen Gründen wurden mit 0 Punkten bewertet. Die Berechnung der Reliabilitäten aus der Raschskalierung sowie der internen Konsistenz ergab sehr gute Werte (EAP/PV-Reliabilität ~ 1 ; Cronbachs $\alpha = 0.95$).

3.2.2.5 Räumliches Denken

Um räumliche Fähigkeiten zu erheben, wurden in der Forschung etablierte Aufgaben zur mentalen Rotation (Kerns & Berenbaum, 1991; Vandenberg & Kuse, 1978; Shepard & Metzler, 1971) adaptiert und zu einem Fragebogen mit elf Aufgaben im Forced-Choice- bzw. Multiple-Choice-Format zusammengestellt. Die Probanden sollten jeweils zwei zeichnerische bzw. fotografische Abbildungen von dreidimensionalen Bauklotz-Konstruktionen, die sich in ihrer Raumorientierung unterschieden, vergleichen und entscheiden, ob die Figuren „gleich“ oder „nicht gleich“ waren, sowie von vier Antwortalternativen die zwei Figuren auswählen, die der präsentierten Figur entsprachen. Für jede richtige Antwort wurde ein Punkt erzielt. Fehlende Werte aufgrund absichtlichen oder versehentlichen Auslassens wurden mit 0 Punkten bewertet. Die erreichten Punkte wurden zu einem Rohwert summiert (Cronbachs $\alpha = 0.67$) bzw. raschskaliert (EAP/PV-Reliabilität = 0.64).

3.2.2.6 Formal-operatorische Fähigkeiten

Zur Erfassung formal-operatorischer Fähigkeiten wurde der Arlin Test of Formal Reasoning (ATFR; Arlin, 1984) aus dem Englischen ins Deutsche übersetzt und jeweils eine Aufgabe zu den Unterbereichen multiplikative Kompensation im Bereich Volumen, Korrelationen, Wahrscheinlichkeit, kombinatorisches Denken, proportionales Denken sowie Koordination von Referenzrahmen eingesetzt. Der Arlin-Test erhebt „schlussfolgerndes Denken beim Übergang von der Stufe der konkreten zu den formalen Operationen sensu Piaget“ (Schneider & Stefanek, 2004, S. 150). In dem gruppentestfähigen Paper-Pencil-Fragebogen wurden die Aufgaben als Text mit Bild dargeboten. Je Aufgabe wurden zwei bis vier Fragen gestellt. Aus den vier Antwortalternativen im Multiple-Choice-Format sollte jeweils die richtige Antwort ausgewählt werden. Jede richtige Antwort erhielt einen Punkt. Fehlende Werte aufgrund von absichtlichen oder versehentlichen Auslassens sowie Nicht-Erreichens aus zeitlichen Gründen und/oder aufgrund einer Schwierigkeitszunahme wurden mit 0 Punkten bewertet. Die erzielten Punkte wurden zu einem Rohwert summiert (Cronbachs $\alpha = 0.25$) bzw. anhand des Rasch-Modells skaliert (EAP/PV-Reliabilität = 0.22). Des Weiteren wurde jedem Kind entsprechend des Summenscores nach den Krite-

rien nach Arlin (1984) ein operatorisches Niveau (konkret, hoch konkret, Übergang, gering formal, formal) zugewiesen. Im Originaltest berichtete Arlin (1984) Punktkriterien für die Klassifikation. Da in der verwendeten Testversion nur ein Teil der Aufgaben ausgewählt wurde, wurden die Punktkriterien in Prozentwerte umgewandelt und diese wurden als Klassifikationskriterien verwendet.

3.2.3 Fragebögen zur Erfassung ausgewählter motivationaler Orientierungen

3.2.3.1 Naturwissenschaftsbezogene Einstellungen und Fähigkeitsselbstkonzept

Zur Erfassung der Einstellungen zu Naturwissenschaften und des naturwissenschaftsbezogenen Fähigkeitsselbstkonzepts wurden die im Rahmen von TIMSS 2007 (vgl. Bos et al., 2008) eingesetzten Items verwendet. Auf einer vierstufigen Antwortskala sollte für jedes der sieben Items (drei Items zu Einstellungen, vier Items zu Selbstkonzept) das Ausmaß der Zustimmung zu den Aussagen angekreuzt werden („stimme völlig zu“, „stimme eher zu“, „stimme eher nicht zu“ oder „stimme überhaupt nicht zu“). Da IRT-Modelle bei der Berechnung der Personenparameter fehlende Werte berücksichtigen, wurden diese Werte als fehlend aufgenommen. Die Itemantworten auf die Skala naturwissenschaftliche Einstellungen und die Skala naturwissenschaftliches Selbstkonzept wurden mit dem Partial-Credit-Modell (Masters, 1982) skaliert (EAP/PV-Reliabilität = 0.73 bzw. 0.66). Bei der Berechnung von Summenwerten wurden nur vollständig bearbeitete Fragebögen berücksichtigt (Cronbachs α = 0.82 bzw. 0.74).

3.2.3.2 Interesse an forschenden Aktivitäten

Das Interesse an forschenden Aktivitäten wurde mittels der Subskala Investigative des Fragebogens ICA-R Inventory of Childrens' Activities – Revised (Tracey & Ward, 1998) erhoben, der die RIASEC (Realistic, Investigative, Artistic, Social, Enterprising und Conventional) Interessentypen nach J. L. Holland (1973, 1985) erfasst. Der Fragebogen wurde aus dem Englischen ins Deutsche übersetzt. Die verwendete Subskala bestand aus fünf Items, die jeweils eine Aktivität abbildeten. Die Kinder sollten für jede Aktivität auf einer adaptierten vierstufigen Likert-Skala („mag ich gar nicht“, „mag ich nicht“, „mag ich“,

„mag ich sehr“) das Ausmaß ihres Interesses ankreuzen. Da IRT-Modelle bei der Berechnung der Personenparameter fehlende Werte berücksichtigen, wurden diese Werte als fehlend aufgenommen. Die Itemantworten wurden mit dem Partial-Credit-Modell skaliert (EAP/PV-Reliabilität = 0.60) Bei der Berechnung von Summenwerten wurden nur vollständig bearbeitete Fragebögen berücksichtigt (Cronbachs α = 0.65).

3.3 Ablauf der Studie

Vor der Datenerhebung wurden an die Schulklassen Elternbriefe verteilt. Die von den Eltern unterschriebenen Erklärungen zur Teilnahme bzw. Nicht-Teilnahme wurden von den Lehrkräften eingesammelt und an die Testleitung übergeben. Die Erhebungen fanden in Gruppentestungen an drei Testterminen à ca. 90 Minuten in den Klassenräumen der Schulen statt. Die Testtage wurden so gestaltet, dass neben der Bearbeitung der Testaufgaben auch Zeit für Erholungspausen blieb (vgl. Tabellen 3.9 bis 3.11). Um die Vergleichbarkeit der Durchführungsbedingungen sicherzustellen, folgten die geschulten Versuchsleiterinnen dem in einem Testmanual festgelegten Testablauf. Zu Beginn der Testungen gaben die Versuchsleiterinnen jeweils eine kurze Einführung und Schilderung der Testabläufe und erklärten die Bearbeitung der Fragebögen mithilfe von Beispielaufgaben. In jeder Klasse achteten ein bis zwei zusätzliche Schreibassistentinnen darauf, dass jedes Kind eigenständig jeweils die entsprechende Testaufgabe beantwortete. Sie standen bei Verständnisfragen und Fragen zur Verschriftlichung zur Verfügung. Um Anonymität der Kinder zu gewährleisten, wurde jedem Kind eine Identifikationsnummer (ID) zugeordnet und für die Datenerhebung und -verarbeitung verwendet.

Am ersten Testtermin wurde der Fragebogen zum wissenschaftlichen Denken durchgeführt. Am zweiten und dritten Testtag bearbeiteten die Schüler die Fragebögen zu den kognitiven und motivationalen Referenzvariablen. Am zweiten Testtag wurden der Test zum Testverständnis des ELFE Leseverständnistests, die Skalen zu naturwissenschaftsbezogenen Einstellungen und Fähigkeitsselbstkonzept, die beiden Subtests des Intelligenztests CFT 20-R, sowie die gruppentestfähige Version des Stroop-Tests durchgeführt. Am dritten Termin bearbeiteten die Kinder die Aufgaben zum räumlichen Denken, die Gruppentestversion des Turms von London zur Erfassung der Problemlösekompetenz sowie die Skala zum Interesse an forschenden Aktivitäten aus dem Interesse-Fragebogen ICA-R. Zusätzlich bearbeiteten die Viertklässler den Arlin-Test zur Erhebung der formal-

operatorischen Fähigkeiten. Am Ende des dritten Testtages wurde an die Klassenleitung ein Geldbetrag für die Klassenkasse überreicht. Außerdem erhielt jedes Kind eine kleine Belohnung in Form einer Süßigkeit.

Tabelle 3.9

Ablauf der Untersuchung und eingesetzte Instrumente am ersten Testtag

Minuten	1. Testtag
40	Testheft – wissenschaftliches Denken (Teil 1)
10	Pause
40	Testheft – wissenschaftliches Denken (Teil 2)

Tabelle 3.10

Ablauf der Untersuchung und eingesetzte Instrumente am zweiten Testtag

Minuten	2. Testtag
15	Leseverständnis (ELFE 1-6)
10	Naturwissenschaftsbezogenen Einstellungen und Fähigkeitsselbstkonzept (TIMSS-Skalen)
10	Pause
20	Intelligenz (CFT 20-R)
10	Inhibition (Stroop-Test)

Tabelle 3.11

Ablauf der Untersuchung und eingesetzte Instrumente am dritten Testtag

Minuten	3. Testtag
20	Räumliches Denken
25	Problemlösen (Turm von London)
10	Pause
25	Formal-operatorische Fähigkeiten (Arlin-Test)*
10	Interesse an forschenden Aktivitäten (ICA-R)

Anmerkung: *Der Arlin-Test wurde nur in Klassenstufe 4 durchgeführt.

3.4 Statistische Verfahren

3.4.1 Item-Response-Theorie

Rasch- Modell und Partial-Credit-Modell

Zur psychometrischen Leistungsmessung der kognitiven Variablen (wissenschaftliches Denken, Leseverständnis, Intelligenz, Problemlösen, Inhibition, räumliches Denken, formal-operatorische Fähigkeiten) wurde das dichotome Rasch-Modell nach Georg Rasch (1960) verwendet, das der probabilistischen Testtheorie bzw. Item-Response-Theorie (IRT) zuzuordnen ist (Moosbrugger, 2008). Die zentrale Annahme besteht darin, dass die beobachteten Antworten auf Testitems Indikatoren für ein dahinterliegendes latentes Konstrukt darstellen (Bühner, 2004). Die latente Variable wird anhand von manifesten, dichotomen Variablen in Form von gelösten bzw. nicht gelösten Items erfasst. Die Lösungswahrscheinlichkeit eines bestimmten Items hängt von der Fähigkeit einer Person und von der Schwierigkeit des Items ab. Die Wahrscheinlichkeit, mit der eine Person v mit einer bestimmten Fähigkeit θ_v ein Item i mit einer bestimmten Schwierigkeit σ_i löst, wird in Form einer logistischen Funktion der Differenz zwischen Personenparameter θ_v und Itemparameter σ_i beschrieben (vgl. Gleichung 3.1).

$$p(X_{vi} = 1) = \frac{\exp(\theta_v - \sigma_i)}{1 + \exp(\theta_v - \sigma_i)} \quad (3.1)$$

Das Rasch-Modell sagt somit in dieser monoton steigenden itemcharakteristischen Funktion (ICF) oder itemcharacteristic curve (ICC) die Lösungswahrscheinlichkeit eines Items durch die beiden Modellparameter, Personenfähigkeit und Itemschwierigkeit, vorher (vgl. Abbildung 3.1). Personenfähigkeiten und Itemschwierigkeiten werden auf einer gemeinsamen Skala (Logit-Skala) verortet (Moosbrugger, 2008). Die Itemschwierigkeit ist dabei als die Stelle auf der Skala definiert, an der die Lösungswahrscheinlichkeit 50 % beträgt. An diesem Punkt sind Personenfähigkeit und Itemschwierigkeit gleich. Ist die Personenfähigkeit größer als die Itemschwierigkeit, ist die Wahrscheinlichkeit für das Lösen des

Items höher als für das Nicht-Lösen. Bei gegebener Itemschwierigkeit steigt die Wahrscheinlichkeit, dass das Item gelöst wird, mit zunehmender Personenfähigkeit an.

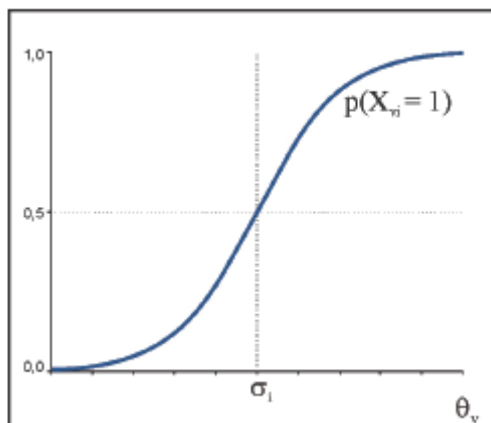


Abbildung 3.1. *Itemcharakteristische Funktion (ICF) eines Items*

Das Rasch-Modell wird auch Einparameter-Logistisches Modell oder 1PL-Modell genannt, da es nur einen Itemparameter, die Itemschwierigkeit, enthält. Bezüglich des Trennschärfeparameters wird angenommen, dass dieser für alle Items gleich ist. Daher sollen die IC-Funktionen parallel verlaufen und nur aufgrund des Schwierigkeitsparameters auf der x-Achse verschoben sein.

Im Rasch-Modell werden die Ausprägungen auf der latenten Variable bzw. die Personenfähigkeiten sowie die Itemschwierigkeiten geschätzt (Embretson & Reise, 2000). Die Parameter werden mit Hilfe der Maximum-Likelihood-Methode so geschätzt, dass unter Geltung der Modellannahmen für die empirisch beobachtete Datenmatrix die beste Passung erzielt wird. Die Zeilen- und Spaltensummenscores dienen dabei als suffiziente Statistiken (Moosbrugger, 2008). Die Schwierigkeit eines Items wird nur dadurch ermittelt, wie viele Personen das Item gelöst haben und nicht welche. Dementsprechend wird die Fähigkeit einer Person dadurch bestimmt, wie viele Items die Person gelöst hat und nicht welche. Bei der Berechnung der Personen- und Itemparameter werden fehlende Werte berücksichtigt. Auf Grundlage der gegebenen Antwortmuster auf die Testitems werden die Personenfähigkeiten in der latenten Variable geschätzt. Die Itemschwierigkeiten werden basierend auf den jeweiligen Lösungshäufigkeiten geschätzt. Bei Modellkonformität sind die Parameter intervallskaliert (Moosbrugger, 2008).

Zur Modellierung der Daten wird das Statistikprogramm ConQuest (Wu, Adams & Wilson, 2005) verwendet. ConQuest berechnet je Item eine Fit-Statistik (Infit, weighted mean square (MNSQ)), die die beobachteten Lösungshäufigkeiten der Items mit den auf Basis des Modells berechneten Lösungshäufigkeiten vergleicht (Wu et al., 2005). Es wird

geprüft, ob die aus dem Modell vorhergesagten Lösungshäufigkeiten mit den beobachteten Häufigkeiten übereinstimmen. Item-Fit-Werte beschreiben demnach, wie gut ein Item zu dem Modell passt (Rost, 2004). Fit-Werte von 1 bezeichnen einen exakten Fit. Das bedeutet, dass die beobachtete IC-Funktion weitgehend der erwarteten Lösungshäufigkeit entspricht. Bei Fit-Werten kleiner als 1 ist die beobachtete Varianz der Lösungshäufigkeiten kleiner als die theoretisch erwartete Verteilung. Somit ist die IC-Funktion des Items steiler als erwartet. Fit-Werte größer als 1 beschreiben einen flacheren Verlauf der beobachteten IC-Funktion als erwartet, da die beobachtete Varianz der Lösungshäufigkeiten größer ist als erwartet. Den Richtlinien von PISA 2000 entsprechend kann von einem akzeptablen Item-Fit gesprochen werden, wenn die Infit-Werte zwischen 0.80 und 1.20 liegen (Adams, 2002). Zusätzlich wird je Item ein T-Wert berechnet, der als Prüfgröße für das Infit-Maß dient. Werte größer als 2.0 und kleiner als -2.0 werden als signifikante Abweichung bewertet. Außerdem berechnet ConQuest pro Item eine punktbiseriale Korrelation, die die Korrelation des dichotomen Items mit dem Gesamtscore der Skala beschreibt und der klassischen Trennschärfe des Items entspricht.

Die Modellparameter werden auf Grundlage der Antwortmuster nach dem Maximum-Likelihood-Prinzip geschätzt. ConQuest bietet dazu die Bock-Aitkin Quadrature Schätzmethode als Standard an. Bei mehrdimensionalen Modellen wird bei drei oder mehr Dimensionen die Monte Carlo Schätzmethode vorgeschlagen (Wu et al., 2005). Den besten Punktschätzer für die Personenfähigkeit liefert dabei der Weighted-Likelihood-Estimate (WLE) nach Warm (1989; Rost, 2004). ConQuest berechnet zusätzlich eine Varianz der latenten Variablen. Die mittels ConQuest ermittelte EAP/PV-Reliabilität ist mit Cronbachs Alpha vergleichbar (Rost, 2004). Die EAP/PV-Reliabilität ist ein Schätzer für die Testreliabilität und wird dadurch berechnet, dass die Varianz der individuellen erwarteten posteriori Fähigkeitsschätzer durch die geschätzte Gesamtvarianz der latenten Fähigkeit geteilt wird.

Im Rahmen der Skalierung der Daten mithilfe von ConQuest werden in einer sogenannten Wright map (vgl. Abbildung 4.1, S. 88) die Itemschwierigkeiten und Personenfähigkeiten auf einer gemeinsamen Skala abgetragen. Die Personenparameter stehen auf der linken Seite der Skala und Itemparameter auf der rechten Seite der Skala. Die Metrik dieser Logit-Skala, die in der Regel von -3 bis +3 Logits reicht, wird dadurch festgelegt, dass die mittlere Itemschwierigkeit auf Null restringiert wird (Moosbrugger, 2008). Negative Werte charakterisieren leichte Items bzw. geringe Personenfähigkeiten. Positive Werte bezeichnen schwere Items bzw. hohe Merkmalsausprägungen. Personen, deren Fähigkeit größer ist als die Schwierigkeit des Items, lösen das Item mit einer Wahrscheinlichkeit von

mehr als 50 %. Dagegen lösen Personen, deren Fähigkeit geringer ist als die Itemschwierigkeit, das Item mit einer geringeren Wahrscheinlichkeit als 50 %.

Zur Skalierung der motivationalen Variablen (naturwissenschaftbezogene Einstellungen, Fähigkeitsselbstkonzept, Interesse an forschenden Aktivitäten) wurde das Partial-Credit-Modell nach Masters (1982) verwendet. Das Partial-Credit-Modell ist eine Erweiterung des Rasch-Modells und ermöglicht die Analyse von Kompetenz- und Einstellungsitems mit mehr als zwei Antwortkategorien.

Neuere Entwicklungen in der IRT ermöglichen weitere Aufschlüsse in der Kompetenzmodellierung (vgl. Koeppen et al., 2008): Beispielsweise können durch mehrdimensionale IRT-Modelle die latente Struktur bzw. die Dimensionalität einer Kompetenz überprüft sowie Korrelationen der latenten Dimensionen berechnet werden. Des Weiteren können Modelle über latente Regressionen analysiert werden, in denen beobachtete Personenmerkmale als Prädiktoren für latente Fähigkeiten genutzt werden.

Ein- und mehrdimensionale IRT-Modelle

Das mehrdimensionale Rasch-Modell kam zum Einsatz, da bei der Entwicklung des Tests zum wissenschaftlichen Denken zwei Kompetenzkomponenten bzw. vier Teilkompetenzen des wissenschaftlichen Denkens operationalisiert wurden und die Dimensionalität des Konstrukts wissenschaftliches Denken in der vorliegenden Studie überprüft werden sollte. Die Dimensionalität eines Konstrukts kann anhand einer globalen Modellprüfung getestet werden. Dabei wird geprüft, ob die theoretisch angenommene Struktur der Kompetenz sich anhand der Daten bestätigen lässt.

Daher wurde ein Modellvergleich durchgeführt: Das eindimensionale Modell, das das Konstrukt wissenschaftliches Denken als eine latente Dimension betrachtet, wurde mit einem zweidimensionalen Modell mit den Dimensionen Wissenschaftsverständnis und Methodenkompetenz bzw. mit einem vierdimensionalen Modell, das die Teilkompetenzen Verständnis der Ziele von Wissenschaften, Theorieverständnis, Experimentierstrategien und Dateninterpretation als einzelne Dimensionen annahm, verglichen. Da die Komponenten bzw. Teilkompetenzen jeweils eine separate latente Dimension erfassten, wurden eine Einfachstruktur und eine between-item-Mehrdimensionalität angenommen (Hartig & Höhler, 2008; Adams, Wilson & Wang, 1997), und derartige Modelle gerechnet.

Für den Vergleich der Passung der verschiedenen Modelle wurden der BIC (Bayesian Information Criterion) und der CAIC (Consistent Akaike's Information Criterion) als Modellselektionsmaße verwendet. Der BIC und CAIC sind Informationskriterien, die

sowohl die Likelihood L , definiert als die Wahrscheinlichkeit der Daten unter der Annahme, dass das Modell gilt, als auch die Anzahl der zu schätzenden Parameter k und dadurch die Sparsamkeit der Modellierung sowie die Größe der Stichprobe N bei der Beurteilung der Passung des Modells zu den Daten berücksichtigen (vgl. Gleichungen 3.2 und 3.3). Der Vorteil dieser Kriterien liegt darin, dass die Modellkomplexität und der Fit der Daten auf das Modell in die Entscheidung eingehen.

$$\text{BIC} = -2 \cdot \log L + (\ln N) \cdot k \quad (3.2)$$

$$\text{CAIC} = -2 \cdot \log L + (\ln N) \cdot k + k \quad (3.3)$$

Es wird dabei ermittelt, welches Modell unter konkurrierenden und unterschiedlich komplexen Modellen als das Angemessenste gilt. Daher sind nur relative Vergleiche zwischen den Modellen möglich. Das Modell mit dem geringeren BIC- und CAIC-Wert beschreibt die Daten besser. Bei der Prüfung der Modelle muss man sich damit zufrieden geben, das relativ beste Modell zu finden (Rost, 2004).

Modellvergleiche der mittels des Programms ConQuest geschätzten ein- und mehrdimensionalen Modelle wurden zum einen zur Überprüfung der latenten Struktur innerhalb der Kompetenz wissenschaftliches Denken herangezogen. Zum anderen wurden ein- und zweidimensionale Modelle unter Einschluss der kognitiven Variablen berechnet, um die Kompetenz wissenschaftliches Denken zu den anderen kognitiven Fähigkeiten abzugrenzen.

Latente Korrelationen

Neben dem Vergleich der Modellgütekriterien der ein- und mehrdimensionalen IRT-Modelle ist die Höhe der latenten Korrelationen zwischen den Dimensionen ein weiteres Kriterium dafür, ob eine Dimension oder mehrere Dimensionen differenzierbar sind. In mehrdimensionalen Modellen werden die Korrelationen der latenten Dimensionen direkt als Modellparameter geschätzt. Im Gegensatz zu dem gestuften Prozess der Berechnung der manifesten Produkt-Moment-Korrelationen aus den in eindimensionalen Modellen geschätzten Personenfähigkeiten sind die latenten Korrelationen durch die direkte Schätzung nicht durch Messfehler vermindert (Adams et al., 1997). Die direkte Schätzung der messfehlerbereinigten latenten Korrelationen ist besonders bei kurzen Testinstrumenten effizienter als der Umweg über die gestufte Berechnung durch die Produkt-Moment-

Korrelationen der messfehlerbehafteten Personenfähigkeiten (Wang, 1999). Sehr hohe Schätzungen der latenten Korrelationen sind dabei konsistent mit der Angemessenheit eines eindimensionalen Modells (Adams et al., 1997).

Latente Korrelationen wurden anhand des Item-Response-Modellierungsprogramms ConQuest in mehrdimensionalen Modellen berechnet, um die latente Struktur innerhalb der Kompetenz wissenschaftliches Denken zu betrachten. Außerdem wurden latente Korrelationen aus mehrdimensionalen Modellen geschätzt, mit dem Ziel die empirischen Zusammenhänge zwischen wissenschaftlichem Denken, respektive der Kompetenzkomponenten Wissenschaftsverständnis und Methodenkompetenz, und den kognitiven sowie motivationalen Variablen zu bestimmen.

Um die Interkorrelationen der Kompetenzkomponenten unter Ausschluss der Kontrollvariablen Leseverständnis und Intelligenz zu betrachten, wurden die bedingten latenten Korrelationen zwischen den Dimensionen anhand eines mehrdimensionalen latenten Regressionsmodells berechnet (Wu et al., 2005). Mittels des gleichen Vorgehens wurden unter Kontrolle von Leseverständnis und Intelligenz die bedingten latenten Korrelationen zwischen wissenschaftlichem Denken, respektive der Komponenten Wissenschaftsverständnis und Methodenkompetenz, zu den kognitiven Variablen Problemlösen, Inhibition, räumliches Denken und formal-operatorische Fähigkeiten ermittelt.

Latente Klassenanalysen

Außerdem wurden Latent-Class-Analysen (LCA) mit dem Statistikprogramm R 2.13.0 mithilfe des Pakets depmixS4 (Visser & Speekenbrink, 2010) durchgeführt, um die Anzahl latenter Klassen im Antwortverhalten der Stichprobe der Grundschüler auf die Items des Tests zum wissenschaftlichen Denken zu ermitteln. Damit sollte überprüft werden, ob sich das Antwortverhalten der Grundschüler qualitativ unterscheidet und dadurch unterschiedliche Kompetenzprofile beschrieben werden können.

Die Grundannahme von LCA-Modellen besteht darin, dass jede Person einer von mehreren latenten Klassen in der Stichprobe angehört (Gollwitzer, 2008). Dabei ist unbekannt, wie viele latente Klassen in einer Stichprobe existieren, wie viele Personen zu den einzelnen Subgruppen angehören und welche Personen welcher latenten Klasse zuzuordnen sind. Ausgehend von der Item-Response-Theorie ist die Zugehörigkeit einer Person zu einer latenten Klasse probabilistisch. Daher wird für das Antwortverhalten einer Person die Wahrscheinlichkeit für die Zugehörigkeit zu einer Klasse berechnet und innerhalb eines LCA-Modells werden die relativen Klassengrößen geschätzt.

Anhand der Itemprofile der latenten Klassen ist ebenfalls eine Dimensionalitätsprüfung möglich: Parallele Antwortprofile auf unterschiedlichem Fähigkeitsniveau sind gleichbedeutend mit der Eindimensionalität des Konstrukts (Rost & Georg, 1991).

Bei der Modellprüfung wurden mittels der Modellgütekriterien BIC und CAIC Modelle mit unterschiedlicher Anzahl an latenten Klassen im Hinblick auf ihre Passung auf die Daten miteinander verglichen sowie der Verlauf der Antwortprofile der ermittelten latenten Klassen betrachtet.

Latente Regressionsanalysen

In latenten Regressionen werden Regressionsmodelle direkt basierend auf dem Antwortverhalten auf die Items geschätzt, ohne zuerst die individuellen Personenscores zu berechnen (Wu et al., 2005).

Das Antwortverhalten auf die Items des Tests zum wissenschaftlichen Denken wurde zunächst in einfachen latenten Regressionen jeweils auf die entsprechenden kognitiven bzw. motivationalen Prädiktoren (Summenwerte⁶ aus den entsprechenden Testinstrumenten) zurückgeführt, um zu ermitteln, welche Variablen einen signifikanten Einfluss auf die Kompetenz im wissenschaftlichen Denken haben. Die aus einfachen latenten Regressionen ermittelten signifikanten Prädiktoren wurden in latente multiple Regressionsanalysen sukzessiv entsprechend der hierarchischen Vorgehensweise eingeführt, um die Kompetenz im wissenschaftlichen Denken durch die kognitiven und motivationalen Prädiktoren vorherzusagen. Die Multikollinearität der Prädiktoren wurde anhand der Produkt-Moment-Interkorrelationsmatrix, des Varianzinflationsfaktors und der Toleranz-Statistik mittels des SPSS-Statistik-Programms PSAW Statistics Version 18 geprüft.

In der latenten Regressionsanalyse kann anhand des Konfidenzintervalls (*KI*) um das Regressionsgewicht beurteilt werden, ob ein Prädiktor einen signifikanten Einfluss auf die abhängige Variable hat. Durch die Multiplikation des Standardfehlers (*SE B*) des Regressionskoeffizienten mit dem Wert 1.96 kann das Konfidenzintervall um den Regressionskoeffizienten *B* gebildet werden (vgl. Gleichung 3.4).

⁶ Da eine Berechnung anhand der WLE-Personenfähigkeiten aufgrund der Dezimalzahlen vom negativen bis positiven Zahlenbereich in den latenten Regressionen mittels ConQuest nicht möglich war, wurden die Summenwerte statt der WLE-Personenfähigkeiten in die latenten Regressionen als Prädiktoren eingeführt. Die Berechnung der WLE-Personenfähigkeiten beruht allerdings auf den Summenwerten als suffiziente Statistiken. Daher sind diese beiden Leistungsmaße mit Werten zwischen $r = 0.92$ und $r = 1.00$ sehr hoch korreliert.

$$KI = B \pm 1.96 * SE B \quad (3.4)$$

Umschließt das Konfidenzintervall nicht den Wert Null, so ist der Einfluss des Prädiktors signifikant von Null verschieden und hat damit einen signifikanten Effekt auf die abhängige Variable. Je enger das Konfidenzintervall und je weiter es von Null entfernt ist, umso größer ist der Effekt des Prädiktors.

Der unstandardisierte Regressionskoeffizient B beschreibt dabei eine Veränderung in der abhängigen Variablen um B -Einheiten (logits) durch eine Veränderung um eine Einheit im Prädiktor.

Da ConQuest bei latenten Regressionen nur unstandardisierte Regressionsgewichte sowie die dazugehörigen Standardfehler berichtet, müssen standardisierte Regressionsgewichte wie folgt berechnet werden: Der standardisierte Regressionskoeffizient $Beta$ kann dadurch ermittelt werden, dass das Verhältnis zwischen der Standardabweichung des Prädiktors (x) und der Standardabweichung der abhängigen Variable (y) mit dem unstandardisierten Regressionsgewicht B des Prädiktors multipliziert wird (vgl. Gleichung 3.5). Die Standardabweichung der abhängigen Variable wird dabei über plausible values ermittelt (Wu et al., 2005).

$$Beta = B * std(x)/std(y) \quad (3.5)$$

Der Wert des standardisierte Regressionsgewicht $Beta$ besagt, das sich die abhängige Variable um $Beta$ Standardabweichungen ändert, wenn sich der Prädiktor um eine Standardabweichung verändert.

Item- und Personenstichprobe

Um konsistente Schätzer für die Personenfähigkeiten in der latenten Variable zu erhalten, müssen die Testverfahren über eine ausreichend große Anzahl an Aufgaben verfügen (Rost, 2004). Für präzise Schätzungen sind 15 bis 20 Items pro Skala erforderlich (Adams et al., 1997; Wu et al., 2005). Die in dem Fragebogen zum wissenschaftlichen Denken enthaltenen 20 Items ermöglichten daher eine konsistente und präzise Schätzung der Personenfähigkeiten für die Variable wissenschaftliches Denken. Aufgrund der geringen Itemzahl für die Komponenten bzw. Teilkompetenzen des wissenschaftlichen Denkens

sind separate Rasch-Modelle für die Komponenten bzw. Teilkompetenzen des wissenschaftlichen Denkens mit Vorsicht zu interpretieren.

Für Rasch-Modelle sind Stichprobengrößen von ca. 200 für präzise Schätzungen notwendig (Wright & Stone, 1979). Da die Stichproben der Klassenstufe 2 ($n = 72$) und der Jahrgangsstufe 3 ($n = 56$) für Raschskalierungen relativ klein waren, sind die Parameterschätzungen für separate Skalierungen stärker von Messfehlern beeinflusst und müssen mit Vorsicht interpretiert werden. Deshalb wurden die Kinder der zweiten und dritten Klassenstufe für Analysen, die auf IRT-Modellen basieren, zu einer Stichprobe des frühen Grundschulalters zusammengefasst. Die Analysen der Gesamtstichprobe ($N = 285$), der vierten Klassenstufe (spätes Grundschulalter: $n = 155$) sowie der Zusammenfassung von zweiter und dritter Jahrgangsstufe (frühes Grundschulalter: $n = 130$) können als präzise bewertet werden.

3.4.2 Prüfung von Gruppenunterschieden

Die Prüfung signifikanter Leistungsunterschiede zwischen den einzelnen Klassenstufen sowie zwischen Mädchen und Jungen erfolgte anhand inferenzstatistischer Verfahren mittels des SPSS-Statistikprogramms PASW Statistics Version 18. Da es sich bei den verwendeten statistischen Testverfahren um gängige Auswertungsmethoden handelt, wird im Folgenden nur die Wahl der eingesetzten Signifikanztests kurz begründet. Eine detaillierte Beschreibung der statistischen Methoden findet sich beispielsweise bei Bühner und Ziegler (2009), Field (2009) und Bortz (1999).

Im Allgemeinen wird der χ^2 -Test als Signifikanztest zur Analyse von Häufigkeitsunterschieden verwendet. Die empirisch ermittelte Prüfgröße χ^2 wird mit dem kritischen χ^2 -Wert verglichen, der sich nach der Anzahl der Freiheitsgrade und dem α -Signifikanzniveau richtet. In der vorliegenden Arbeit kam der χ^2 -Test zum Einsatz, um die Lösungshäufigkeiten der Zweit-, Dritt- und Viertklässler je Aufgabe zu vergleichen. In Voraussetzungsanalysen zur Untersuchung der Beziehungen zwischen Theorieverständnis, Verständnis der experimentellen Methode und der Evaluation von Daten sowie zwischen wissenschaftlichem Denken und formal-operatorischem Niveau wurden die Häufigkeitsunterschiede der jeweils zweifachgestuften Variablen (Verständnis vs. kein Verständnis, bzw. Kompetenz vs. keine Kompetenz und konkret vs. operatorisches Niveau) mittels Kontingenztafeln betrachtet und anhand des χ^2 -Test auf Signifikanz geprüft.

Mithilfe von Varianzanalysen (ANOVA) können Mittelwertsunterschiede intervallskaliertter Variablen in mehr als zwei Gruppen auf Signifikanz getestet werden. Die univariate Varianzanalyse überprüft, in welchem Ausmaß die Gesamtvarianz der abhängigen Variable auf den systematischen und überzufälligen Einfluss der unterschiedlichen Stufen der unabhängigen Variable zurückgeführt und ob ein statistisch signifikanter Varianzanteil durch die Ausprägungen der unabhängigen Variable erklärt werden kann. Zur Signifikanzprüfung, ob die Mittelwertunterschiede systematisch zustande gekommen und damit statistisch bedeutsam sind, wird der empirische F-Wert mit dem kritischen F-Wert, in Abhängigkeit von der Anzahl der Freiheitsgrade und dem α -Irrtumswahrscheinlichkeitsniveau, verglichen. Ein signifikantes Ergebnis besagt, dass mindestens zwei Mittelwerte sich signifikant unterscheiden. Die Kompetenz in der abhängigen Variable wissenschaftliches Denken wurde in Abhängigkeit der beiden unabhängigen Variablen Geschlecht (zweifach gestuft: Mädchen, Jungen) und Jahrgangsstufe (dreifach gestuft: Klasse 2, Klasse 3, Klasse 4) mittels univariater zweifaktorieller Varianzanalysen untersucht. Um im Falle der dreifach gestuften Variable Klassenstufe Aussagen darüber zu treffen, welche Mittelwerte sich signifikant voneinander unterscheiden, wurden a-priori festgelegte Einzelvergleiche berechnet, um zu überprüfen, ob sich Zweit- und Drittklässler sowie Dritt- und Viertklässler bzw. die aus Klassenstufe 2 und 3 zusammengefasste Stichprobe des frühen Grundschulalters von der Klassenstufe 4 als spätes Grundschulalter in ihren Leistungen signifikant voneinander unterscheiden.

Zur Prüfung von Gruppenunterschieden hinsichtlich der Leistungen in den Kompetenzkomponenten Wissenschaftsverständnis und Methodenkompetenz bzw. in den Teilkompetenzen Verständnis der Ziele von Wissenschaften, Theorieverständnis, Experimentierstrategien und Dateninterpretation kamen zunächst multivariate Varianzanalysen (MANOVA) zum Einsatz, um die Beziehungen zwischen den abhängigen Variablen zu berücksichtigen. Anschließend wurden je Komponente bzw. Teilkompetenz univariate Varianzanalysen gerechnet.

Um den Einfluss der Kontrollvariablen Leseverständnis und Intelligenz auf die Gruppenunterschiede im wissenschaftlichen Denken zu kontrollieren, wurden diese Variablen als Kovariaten zusätzlich in die beschriebenen Varianzanalysen eingeführt und Kovarianzanalysen (ANCOVA) mit geplanten Kontrasten berechnet.

Die Leistungen in den abhängigen Variablen Leseverständnis, Intelligenz, Problemlösen, Inhibition, räumliches Denken, naturwissenschaftsbezogene Einstellungen, Fähigkeitsselbstkonzept sowie Interesse an forschenden Aktivitäten wurden ebenfalls mittels

univariater zweifaktorieller Varianzanalysen (Klassenstufe, Geschlecht) auf Mittelwertsunterschiede überprüft.

Da formal-operatorische Fähigkeiten nur in Klassenstufe 4 erhoben wurden, erfolgte die Prüfung hinsichtlich unterschiedlicher Leistungen zwischen Mädchen und Jungen anhand eines t-Tests für unabhängige Stichproben. Ob ein statistisch signifikanter Mittelwertunterschied vorliegt, wird im Rahmen des t-Tests durch den Vergleich des empirisch ermittelten t-Werts mit dem kritischen t-Wert, in Abhängigkeit von der Anzahl der Freiheitsgrade und dem α -Signifikanzniveau, beurteilt.

Die Voraussetzungen für die Durchführung von Varianzanalysen und t-Tests sind Normalverteilung und Varianzhomogenität der Daten. Da sich diese statistischen Verfahren, insbesondere bei großen Stichproben, allerdings relativ robust gegenüber der Verletzung dieser Annahmen verhalten, sind diese parametrischen Verfahren in der Regel den nonparametrischen Tests vorzuziehen (Bühner & Ziegler, 2009). Im Falle inhomogener Varianzen wurde der Welch-Test mit der entsprechenden Korrektur der Freiheitsgrade verwendet.

4. Ergebnisse

4.1 Skalierung der Kompetenz wissenschaftliches Denken

Für die Analysen wurden die 20 Items des Testhefts dichotom kodiert. Die erhobenen Daten der gesamten Stichprobe von 285 Kindern sowie der aus Klassenstufe 2 ($n = 72$) und Klassenstufe 3 ($n = 58$) zusammengefassten Stichprobe des frühen Grundschulalters und der Stichprobe des späten Grundschulalters (Klassenstufe 4: $n = 155$) wurden nach dem Rasch-Modell skaliert. Die Modellparameter wurden mithilfe des Statistikprogramms ConQuest (Wu et al., 2005) nach der Maximum-Likelihood-Methode geschätzt.

4.1.1 Modellprüfung

4.1.1.1 Modellvergleiche

Zur Überprüfung der Dimensionalität des Konstrukts wissenschaftliches Denken war es zunächst notwendig zu bestimmen, welches Modell die Daten am besten beschreibt. Dazu wurden Modellvergleiche anhand der informationstheoretischen Kriterien BIC und CAIC angestellt. Neben dem eindimensionalen Rasch-Modell für das Konstrukt wissenschaftliches Denken wurden ein zweidimensionales Rasch-Modell mit den Dimensionen Wissenschaftsverständnis und Methodenkompetenz sowie ein vierdimensionales Rasch-Modell mit den Dimensionen Verständnis wissenschaftlicher Konzepte und der Ziele von Wissenschaften, Theorieverständnis, Experimentierstrategien und Dateninterpretation berechnet.

Tabelle 4.1 stellt die Modellfit-Indizes der drei Testmodelle je Stichprobe gegenüber. Sowohl der BIC als auch der CAIC fielen jeweils für das eindimensionale Rasch-Modell am kleinsten aus.

Tabelle 4.1

Modellvergleich des eindimensionalen, zweidimensionalen und vierdimensionalen Rasch-Modells der Kompetenz wissenschaftliches Denken

Modell	Stufe	Deviance	#Par	BIC	CAIC
eindimensional	2-3	3159		3261	3282
	4	3443	21	3549	3570
	gesamt	6702		6821	6842
zweidimensional	2-3	3182		3293	3316
	4	3479	23	3595	3618
	gesamt	6749		6879	6902
vierdimensional	2-3	3257		3403	3433
	4	3605	30	3756	3786
	gesamt	6915		7085	7115

Anmerkungen: Deviance = $-2 \cdot \text{LogLikelihood}$, #Par = Anzahl der geschätzten Parameter, BIC = Bayes Information Criterion, CAIC = Consistent Akaike's Information Criterion.

Dem Vergleich der Informationskriterien BIC und CAIC zufolge war ein eindimensionales gegenüber einem zweidimensionalen Modell und einem vierdimensionalen Modell vorzuziehen. Insgesamt hat sich sowohl in der Gesamtstichprobe der Grundkinder als auch in den einzelnen Jahrgangsstufen das eindimensionale Kompetenzmodell auf Grundlage der informationstheoretischen Kriterien im Vergleich zu den Alternativmodellen als das am besten passende Modell herausgestellt.

4.1.1.2 Latente Korrelationen

Die zweidimensionalen Skalierungen ergaben hohe messfehlerbereinigte latente Korrelationen der Komponenten des wissenschaftlichen Denkens, sowohl für die Gesamtstichprobe als auch für das frühe (2. und 3. Jahrgangsstufe) und späte Grundschulalter (4. Klassenstufe). Wie in Tabelle 4.2 dargestellt betrug die latente Korrelation zwischen den beiden Dimensionen Wissenschaftsverständnis und Methodenkompetenz des zweidimensionalen Rasch-Modells $r = 0.85$ in der Gesamtstichprobe, respektive $r = 0.84$ im frühen und $r = 0.71$ im späten Grundschulalter. Die in zweidimensionalen latenten Regressionen ermittelten latenten Korrelationen zwischen den Kompetenzkomponenten des wissenschaftlichen Denkens unter Kontrolle von Leseverständnis und Intelligenz blieben in ihrer Höhe relativ konstant (vgl. Tabelle 4.2). Nur in Klassenstufe 2-3 verminderte sich die latente Korrelation unter Kontrolle von Leseverständnis und Intelligenz auf $r = 0.68$.

Tabelle 4.2

Latente Korrelationen der Komponenten des wissenschaftlichen Denkens

	1.	2.
Gesamtstichprobe (N = 285)		
1. Wissenschaftsverständnis	1	0.88
2. Methodenkompetenz	0.85	1
Jahrgangsstufe 2-3 (n = 130)		
1. Wissenschaftsverständnis	1	0.68
2. Methodenkompetenz	0.84	1
Jahrgangsstufe 4 (n = 155)		
1. Wissenschaftsverständnis	1	0.86
2. Methodenkompetenz	0.71	1

Anmerkung: Latente Korrelationen basierend auf zweidimensionalen Rasch-Skalierungen stehen unterhalb der Diagonale. Bedingte latente Korrelationen unter Kontrolle von Leseverständnis und Intelligenz basierend auf zweidimensionalen latenten Regressionen stehen oberhalb der Diagonale.

In Tabelle 4.3 sind die latenten, messfehlerbereinigten Korrelationen der vierdimensionalen Rasch-Modelle gegenübergestellt. In der Gesamtstichprobe wiesen die Teilkompetenzen des wissenschaftlichen Denkens ebenfalls hohe latente Korrelationen auf und lagen zwischen $r = 0.82$ und $r = 0.90$. Differenziert für das frühe und späte Grundschulalter betrachtet konnten die latenten Korrelationen der Teilkompetenzen des wissenschaftlichen Denkens als moderat bis hoch bewertet werden. In der Stichprobe der Zweit- und Drittklässler schwankten die latenten Korrelationen von $r = 0.68$ bis $r = 0.94$, in Jahrgangsstufe 4 von $r = 0.43$ bis $r = 0.90$. Die Berechnung der bedingten latenten Korrelationen der Teilkompetenzen unter Kontrolle von Leseverständnis und Intelligenz basierend auf vierdimensionalen latenten Regressionen war aufgrund der Datenbasis nicht möglich.

Tabelle 4.3

Latente Korrelationen der Teilkompetenzen des wissenschaftlichen Denkens

	1.	2.	3.	4.
Gesamtstichprobe (N = 285)				
1. Verständnis der Ziele von Wissenschaften	1			
2. Theorieverständnis	0.88	1		
3. Experimentierstrategien	0.82	0.90	1	
4. Dateninterpretation	0.90	0.89	0.90	1
Jahrgangsstufe 2-3 (n = 130)				
1. Verständnis der Ziele von Wissenschaften	1			
2. Theorieverständnis	0.72	1		
3. Experimentierstrategien	0.90	0.82	1	
4. Dateninterpretation	0.92	0.68	0.94	1
Jahrgangsstufe 4 (n = 155)				
1. Verständnis der Ziele von Wissenschaften	1			
2. Theorieverständnis	0.58	1		
3. Experimentierstrategien	0.43	0.77	1	
4. Dateninterpretation	0.48	0.90	0.75	1

Anmerkung: Latente Korrelationen der Teilkompetenzen des wissenschaftlichen Denkens basierend auf der vierdimensionalen Rasch-Skalierungen stehen unterhalb der Diagonale. Die Berechnung der bedingten latenten Korrelationen der Teilkompetenzen unter Kontrolle von Leseverständnis und Intelligenz basierend auf vierdimensionalen latenten Regressionen war aufgrund der Datenbasis nicht möglich.

Die dargestellten Ergebnisse können wie folgt zusammengefasst werden: Die Modellvergleiche stellten heraus, dass das eindimensionale Modell des wissenschaftlichen Denkens dem zwei- und vierdimensionalen Modell sowohl in der Gesamtstichprobe der Grundschul Kinder als auch in den Stichproben der Zweit- und Dritt- sowie Viertklässler überlegen war. Die sowohl in der Gesamtstichprobe als auch auf Klassenebene größtenteils hohen latenten Korrelationen zwischen den Dimensionen der mehrdimensionalen Skalierungen sind mit der Angemessenheit eines eindimensionalen Modells konsistent (Adams et al., 1997). Daher scheint es angemessen, die eng miteinander verknüpften Teilkompetenzen des wissenschaftlichen Denkens zu einem eindimensionalen Konstrukt zusammenzufassen.

4.1.1.3 Reliabilität

Die Reliabilität wurde anhand der von ConQuest berechneten EAP/PV-Reliabilität sowie anhand der internen Konsistenz (Cronbachs Alpha) geprüft. Tabelle 4.4 stellt die Varianzen der latenten Verteilungen und Reliabilitätsmaße der Rasch-Skalierungen sowie die Maße der internen Konsistenz nach Cronbachs Alpha für die Kompetenz wissenschaftliches Denken respektive der Komponenten bzw. Teilkompetenzen für die einzelnen Stichproben gegenüber. Die EAP/PV-Reliabilität des eindimensionalen Konstrukts wissenschaftliches Denken bezogen auf die Gesamtstichprobe erreichte mit 0.75 bei einer Varianz von 0.82 einen zufrieden stellenden Wert. Für die einzelnen Klassenstufen zeigten sich für die Gesamtskala wissenschaftliches Denken ebenfalls zufrieden stellende Varianzen und Reliabilitäten (vgl. Tabelle 4.4).

Bei der differenzierten Erfassung der zwei Komponenten bzw. vier Teilkompetenzen des wissenschaftlichen Denkens als einzelne Dimensionen bestanden zum Teil Reliabilitätsprobleme (vgl. Tabelle 4.4).

Tabelle 4.4

Varianzen und EAP/PV-Reliabilitäten der eindimensionalen Rasch-Modelle sowie Cronbachs Alpha für die Gesamtskala respektive Subskalen der Kompetenz wissenschaftliches Denken

Konstrukt	Varianz			EAP/PV-Reliabilität			Cronbachs Alpha		
	Jahrgangsstufe								
	2-3	4	gesamt	2-3	4	gesamt	2-3	4	gesamt
Wissenschaftliches Denken	0.54	0.60	0.82	0.69	0.68	0.75	0.69	0.68	0.76
Wissenschaftsverständnis	0.24	0.26	0.42	0.27	0.26	0.39	0.25	0.22	0.39
Methodenkompetenz	0.92	1.22	1.35	0.68	0.70	0.74	0.68	0.70	0.74
Verständnis der Ziele von Wissenschaften	0.27	0.43	0.47	0.21	0.24	0.30	0.15	0.23	0.30
Theorieverständnis	0.34	0.09	0.24	0.13	0.05	0.13	0.11	0.00	0.13
Experimentierstrategien	1.20	1.66	1.73	0.64	0.65	0.69	0.63	0.66	0.69
Dateninterpretation	0.43	1.22	1.03	0.26	0.45	0.44	0.26	0.46	0.44

Die latente Variable wissenschaftliches Denken konnte folglich nur hinreichend reliabel in einem eindimensionalen Rasch-Modell und dadurch in einer aussagekräftigen Gesamtskala erfasst werden. Wie auch die Modellvergleiche und die hohen latenten Korrelationen der mehrdimensionalen Skalierungen zeigten, handelte es sich bei den Komponenten bzw.

Teilkompetenzen des wissenschaftlichen Denkens um eng miteinander verknüpfte Konstrukte, die sich am besten in einem eindimensionalen Testmodell abbilden ließen.

Um allerdings ein differenziertes Bild des wissenschaftlichen Denkens zu beschreiben, wird in den folgenden Abschnitten neben der Gesamtskala auch auf die einzelnen Teilbereiche des wissenschaftlichen Denkens eingegangen.

4.1 2 Statistische Kennwerte des eindimensionalen Rasch-Modells

4.1.2.1 Wright map

Item- und Personenparameter des eindimensionalen Rasch-Modells auf Basis der gesamten Stichprobe sind in der Abbildung 4.1 in einer Wright map, die die Modellparameter auf einer gemeinsamen Skala abbildet, dargestellt. Die Personenparameter sind auf der rechten Seite, die Itemparameter auf der linken Seite der Wright map abgetragen.

Die mittlere Itemschwierigkeit wurde auf 0 logits ($SD = 0.92$) restringiert und stellt den Nullpunkt der Skala dar. Die Itemschwierigkeiten reichten von -1.70 bis 1.62 logits. Die Personenfähigkeiten der 285 Schüler der zweiten bis vierten Klassenstufe schwankten von -2.92 bis 4.11 logits. Die mittlere Personfähigkeit lag bei 0.41 logits ($SD = 1.04$). Die Wright map lässt erkennen, dass für den mittleren Fähigkeitsbereich ausreichend Items zur Verfügung standen. Allerdings fehlten Items zur Differenzierung der Personen im unteren und oberen Randbereich der Fähigkeitsskala. Insgesamt kann der Test bezogen auf die gesamte Stichprobe der Zweit-, Dritt- und Viertklässler mit einer mittleren Itemschwierigkeit von 0 logits im Vergleich zu einer durchschnittlichen Personenfähigkeit von 0.41 logits als ein wenig zu leicht, aber angemessen bezeichnet werden.

4.1.2.2 Itemkennwerte

Tabelle 4.5 fasst die statistischen Itemkennwerte der eindimensionalen Rasch-Skalierung der Gesamtskala zur Erfassung des wissenschaftlichen Denkens bezogen auf die Gesamtstichprobe zusammen.

Tabelle 4.5

Statistische Kennwerte der eindimensionalen Rasch-Skalierung in absteigender Schwierigkeitsreihenfolge

Item	Teilkompetenz	N	Lösungshäufigkeit (%)	Schwierigkeit	Trennschärfe	Infit MNSQ	T-Wert
a23_Wichtel	2	284	26.4	1.62	0.22	1.13	1.9
a24_Mittelalter	2	285	30.9	1.36	0.33	1.07	1.2
a27_Geburtstagsparty	2	285	37.2	1.04	0.46	0.97	-0.7
a64_Zähne	4	284	40.9	0.86	0.54	0.93	-1.6
a57_2_Fußpilz	3	285	42.1	0.80	0.57	0.90	-2.1
a35_Flugzeug	3	282	43.6	0.73	0.50	0.96	-0.8
a44_2_Ärzte	3	285	45.6	0.63	0.48	0.97	-0.8
a66_Ravianer	4	282	52.5	0.31	0.39	1.06	1.2
a21_Fehler	1	285	54.0	0.23	0.35	1.09	2.0
a53_Spürnase	3	284	57.0	0.09	0.54	0.92	-1.7
a57_1_Fußpilz	3	284	58.1	0.04	0.57	0.89	-2.2
a76_Diagramm	4	285	66.7	-0.39	0.30	1.11	1.9
a44_1_Ärzte	3	284	68.0	-0.46	0.43	1.00	0.0
a82_Tabelle	4	284	70.4	-0.59	0.55	0.90	-1.7
a16_Wissenschaftler	1	285	70.5	-0.60	0.32	1.09	1.3
a02_Hypothese	1	284	71.8	-0.67	0.26	1.14	2.0
a49_Pizzabacken	3	283	76.3	-0.94	0.45	0.95	-0.6
a20_Pflanzemittel	1	283	78.8	-1.11	0.22	1.11	1.3
a08_Theorie	1	285	80.7	-1.24	0.49	0.90	-1.1
a51_Maus Haus	3	285	86.3	-1.70	0.41	0.94	-0.5
EAP/PV-Reliabilität: 0.75							
Varianz: 0.82							
Trennschärfe: $M = 0.42$, $SD = 0.12$, $MIN = 0.22$, $MAX = 0.57$							
Schwierigkeit: $M = 0$, $SD = 0.92$, $MIN = -1.70$, $MAX = 1.62$							
Infit: $M = 1.00$, $SD = 0.09$, $MIN = 0.89$, $MAX = 1.14$							

Anmerkung: Teilkompetenz: 1 = Verständnis wissenschaftlicher Konzepte und der Ziele von Wissenschaften, 2 = Theorieverständnis, 3 = Experimentierstrategien, 4 = Dateninterpretation.

Die Fit-Statistiken der Itemparameter des eindimensionalen Rasch-Modells waren zufriedenstellend. Es konnten keine signifikanten T-Werte ($-2.0 < T\text{-Wert} < 2.0$) festgestellt werden. Der Infit (Weighted Mean Square, MNSQ) lag für alle Items innerhalb des von den Richtlinien vorgegebenen Bereichs mit einer Untergrenze von 0.80 und einer Obergrenze von 1.20 (Adams, 2002). Die Infit-Werte reichten von 0.89 bis 1.14. Der Mittelwert der MNSQ-Werte aller Items lag bei 1.00 ($SD = 0.09$) und erreichte damit einen optimalen Wert. Die Fit-Statistiken zeigten, dass die Daten gut durch das Rasch-Modell modelliert werden konnten und sich die Items modellkonform verhielten.

Die punktbiserialen Korrelationen bzw. die klassischen Trennschärfen lagen zwischen 0.22 und 0.57. Die mittlere Trennschärfe aller Items betrug 0.42 ($SD = 0.12$) und lag in dem als gut zu bewertenden Bereich zwischen 0.4 und 0.7 (Kelava & Moosbrugger, 2008).

Die Varianz der latenten Variablen lag bei 0.82. Die EAP/PV-Reliabilität, die mit Cronbachs Alpha vergleichbar ist (Rost, 2004), erreichte einen zufriedenstellenden Wert von 0.75 (vgl. Tabelle 4.4, S. 86).

4.1.3 Latente Klassenanalysen

Zusätzlich zu dem eindimensionalen Rasch-Modell wurden Latente Klassenanalysen (LCA-Modelle) mit dem Statistikprogramm R 2.13.0 mithilfe des Pakets depmixS4 (Visser & Speekenbrink, 2010) berechnet. Es wurde damit überprüft, ob in der gesamten Stichprobe qualitative Personenunterschiede hinsichtlich des Profils des Antwortverhaltens auf die Items respektive Teilkompetenzen des wissenschaftlichen Denkens vorliegen.

Da für diese Analysen nur Versuchspersonen mit vollständigen Daten verwendet werden können, mussten elf Kinder mit fehlenden Werten aus den Analysen ausgeschlossen werden, so dass 274 Schüler (70 Zweitklässler, 56 Drittklässler und 148 Viertklässler) in die Berechnung der LCA-Modelle eingingen.

Zum Vergleich der Passung der verschiedenen LCA-Modelle wurden ebenfalls der BIC und CAIC herangezogen. In Tabelle 4.6 sind die Fit-Indizes der LCA-Modelle mit einer bis fünf latenter Klassen berichtet.

Tabelle 4.6

Vergleich der Modellpassung der LCA-Modelle

Modell	LogL	#Par	BIC	CAIC
1 latente Klasse	-3381	20	6875	6895
2 latente Klassen	-3161	41	6551	6592
3 latente Klassen	-3107	62	6561	6623
4 latente Klassen	-3070	83	6606	6689
5 latente Klassen	-3046	104	6676	6780

Anmerkungen: LogL = LogLikelihood, #Par = Anzahl der geschätzten Parameter, BIC = Bayes Information Criterion, CAIC = Consistent Akaike's Information Criterion.

Die informationstheoretischen Kriterien BIC und CAIC sprachen für das 2-Klassen-Modell. Die latente Klasse 2 umfasst mit 61.0 % die Mehrheit der Stichprobe. 39.0 % der Versuchspersonen wurden der latenten Klasse 1 zugeordnet.

Betrachtet man die Profile des Antwortverhaltens der beiden latenten Klassen auf die einzelnen Items in Abbildung 4.2, so können keine Überschneidungen festgestellt werden, sondern lediglich Niveauunterschiede.

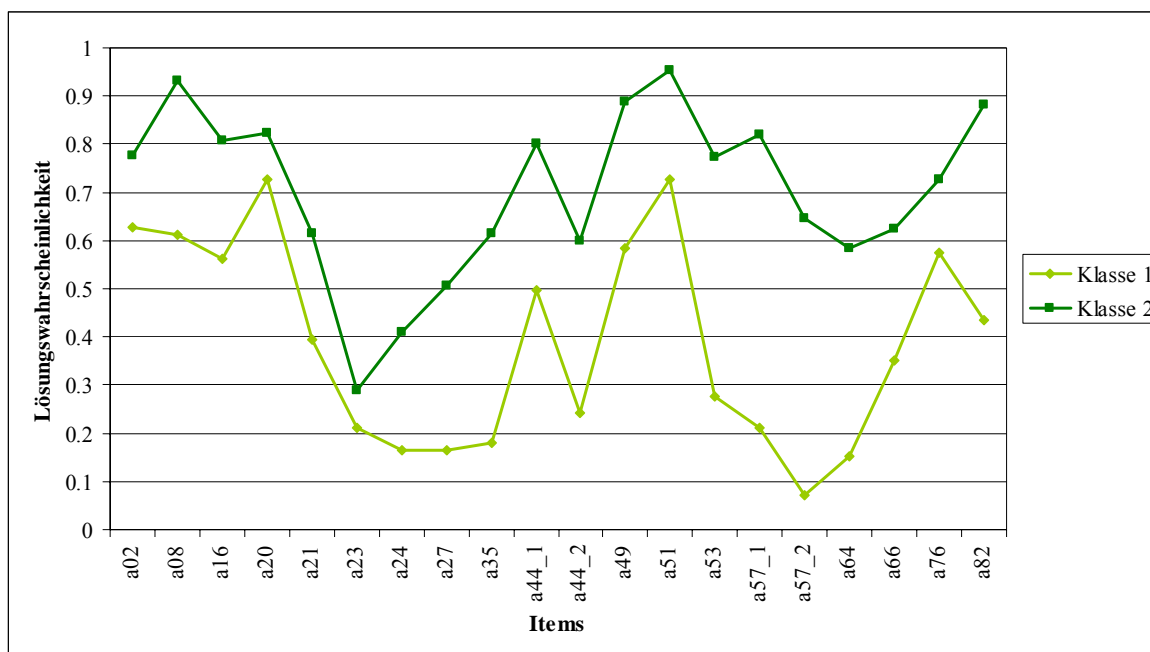


Abbildung 4.2. Itemprofile der latenten Klassen des 2-Klassen-LCA-Modells

Die Mittelwerte der Lösungswahrscheinlichkeiten für die einzelnen Teilkomponenten des wissenschaftlichen Denkens sind getrennt für die beiden latenten Klassen in Abbildung 4.3 dargestellt.

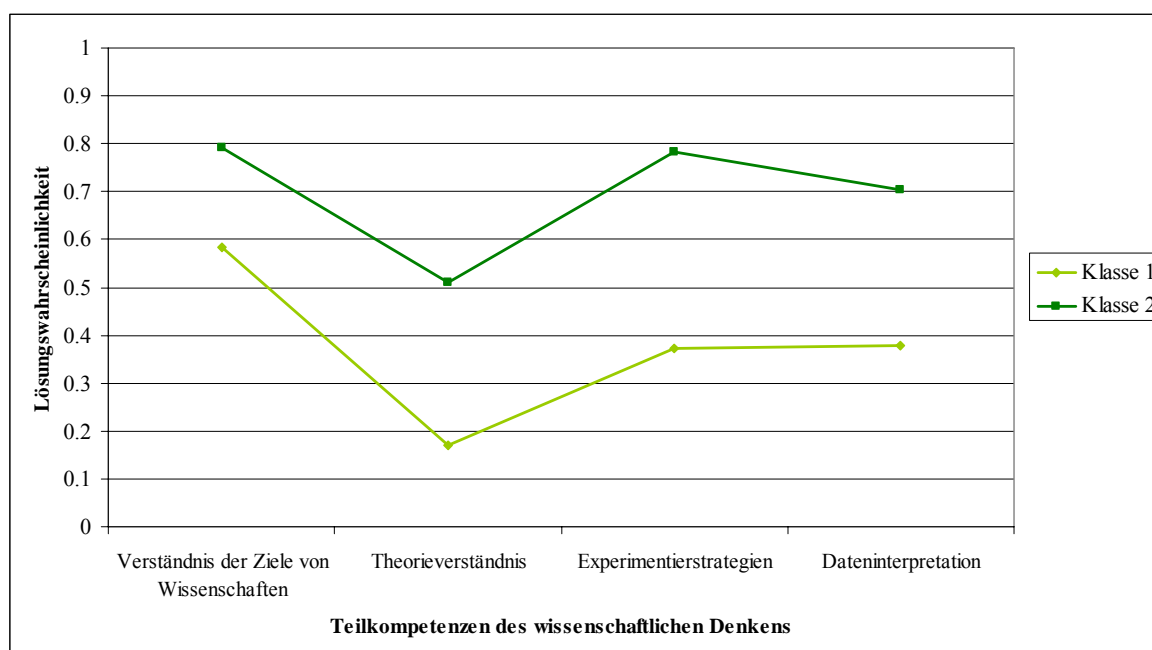


Abbildung 4.3. Kompetenzprofile der latenten Klassen des 2-Klassen-LCA-Modells für die Teilkomponenten des wissenschaftlichen Denkens

Die Kompetenzprofile hinsichtlich der Teilkompetenzen des wissenschaftlichen Denkens, Verständnis der Ziele von Wissenschaften und Theorieverständnis sowie Experimentierstrategien und Dateninterpretation, wiesen ebenfalls keine Überschneidungen auf. Der parallele Verlauf der Kompetenzprofile zeigte, dass keine qualitativ unterschiedlichen Antwortmuster in den latenten Klassen vorlagen. Vielmehr kann die Staffelung der Kompetenzprofile als unterschiedliche Leistungsniveaus angesehen werden.

Das Ergebnis der latenten Klassenanalyse mit zwei latenten Klassen auf unterschiedlichem Fähigkeitsniveau war konsistent mit der Eindimensionalität des latenten Konstrukts wissenschaftliches Denken der Raschanalysen (Rost & Georg, 1991).

Tabelle 4.7 stellt die Klassenzugehörigkeit je Jahrgangsstufe der Grundschule gegenüber. Es wurde deutlich, dass Kinder der Jahrgangsstufe 2 mit 64.3 % noch deutlich häufiger der inkompetenteren latenten Klasse 1 angehörten, während die Schüler der vierten Jahrgangsstufe mit 77.7 % mehrheitlich in der kompetenteren latenten Klasse 2 waren. Die Jahrgangsstufe 3 verteilte sich mit je 50.0 % gleichmäßig auf die beiden latenten Klassen. Die Verteilung der Zweit-, Dritt- und Viertklässler auf die latenten Klassen unterschied sich signifikant ($\chi^2(2) = 39.12, p < .001$).

Tabelle 4.7

Klassenzugehörigkeit nach Jahrgangsstufe

<i>Häufigkeit</i>	Jahrgangsstufe 2		Jahrgangsstufe 3		Jahrgangsstufe 4	
	<i>absolut</i>	<i>relativ</i>	<i>absolut</i>	<i>relativ</i>	<i>absolut</i>	<i>relativ</i>
Latente Klasse 1	45	.64	28	.50	33	.22
Latente Klasse 2	25	.36	28	.50	115	.78

Wird die Variable Jahrgangsstufe als Kovariate für die Übergangswahrscheinlichkeit zwischen den latenten Klassen in das 2-Klassen-Modell aufgenommen, so zeigte sich, dass am Ende der dritten Klassenstufe bzw. mit dem Alter von ca. neun Jahren die Wahrscheinlichkeit für den Wechsel von der inkompetenteren in die kompetentere latente Klasse zunahm (vgl. Abbildung 4.4 und 4.5).

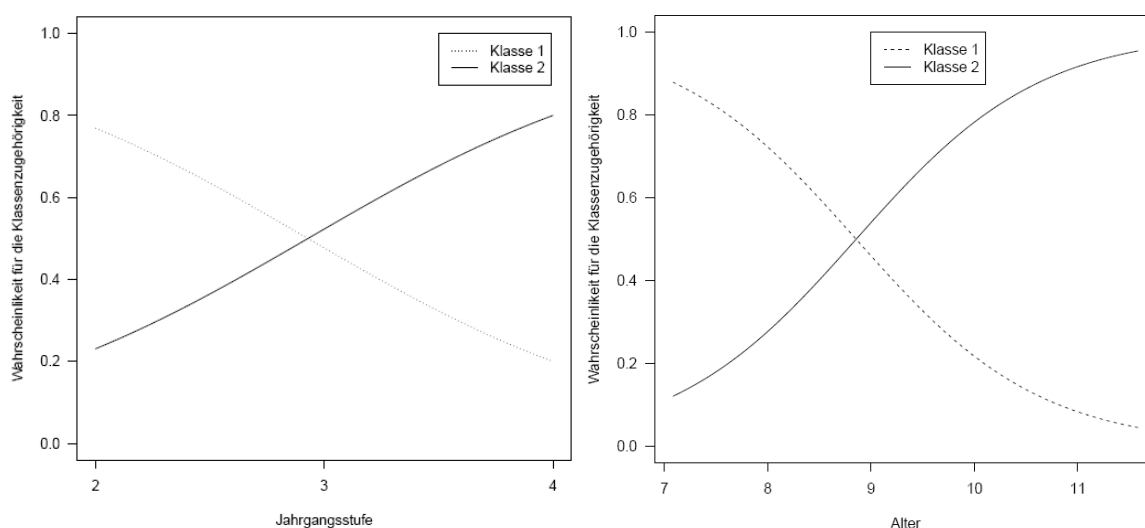


Abbildung 4.4 und 4.5. *Wahrscheinlichkeit für die Klassenzugehörigkeit nach Jahrgangsstufe und Alter*

4.2 Entwicklung der Kompetenz wissenschaftliches Denken im Grundschulalter

4.2.1 Deskriptive Statistiken

Es gab sowohl Evidenzen für Defizite als auch für Kompetenzen im wissenschaftlichen Denken im Grundschulalter, betrachtet an den mittleren relativen Lösungshäufigkeiten für die Gesamtheit der Aufgaben. Wie in Abbildung 4.6 dargestellt, hatte die Mehrheit der Zweitklässler mit 55,5 % noch Fehlvorstellungen, wohingegen die Mehrheit der Viert-

klässler mit 66.6 % schon wissenschaftlich adäquate Vorstellungen besaß. In Klassenstufe 3 verteilten sich Fehlvorstellungen und wissenschaftliche Vorstellung mit 48.7 % bzw. 51.3 % in etwa gleichmäßig.

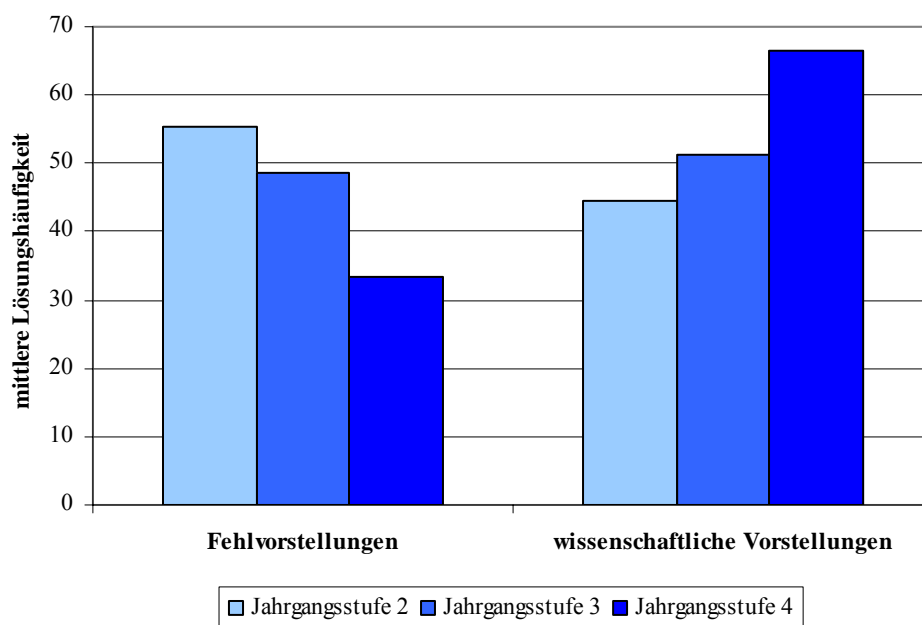


Abbildung 4.6. Mittlere relative Häufigkeiten für Fehlvorstellungen und wissenschaftliche Vorstellungen dargestellt getrennt nach Jahrgangsstufen

In Tabelle 4.8 sind die Aufgaben der Teilkompetenz Verständnis wissenschaftlicher Konzepte und der Ziele von Wissenschaften nach der berechneten Aufgabenschwierigkeit aus der eindimensionalen Raschskalierung der Daten der Gesamtstichprobe in aufsteigender Reihenfolge angeordnet. In dieser Teilkompetenz war für die Grundschul Kinder die Aufgabe zum intuitiven Verständnis des Konzeptes Theorie (Aufgabe a08_Theorie) am einfachsten, gefolgt von den Aufgaben zum intuitiven Verständnis des Konzeptes Experiment (Aufgabe a20_Pflanzenmittel) und des Konzeptes Hypothese (Aufgabe a02_Hypothese). Die Aufgaben über die Ziele von Wissenschaften (Aufgabe a16_Wissenschaftler) und wissenschaftlicher Fehler (Aufgabe a21_Fehler) hatten eine leichte bis mittlere Aufgabenschwierigkeit.

Tabelle 4.8

Lösungshäufigkeiten (in Prozent) der Aufgaben der Teilkompetenz Verständnis der Ziele von Wissenschaften geordnet nach ansteigender Aufgabenschwierigkeit getrennt nach Jahrgangsstufe

Item	Schwierigkeit	Jahrgangsstufe				χ^2 -Test
		2	3	4	gesamt	
a08_Theorie	-1.24	58.3	79.3	91.6	80.7	$\chi^2(2) = 35.05$, $p < .001$
a20_Pflanzenmittel	-1.11	70.8	68.4	86.4	78.8	$\chi^2(2) = 11.68$, $p < .01$
a02_Hypothese	-0.67	61.1	67.2	78.6	71.8	$\chi^2(2) = 8.15$, $p < .05$
a16_Wissenschaftler	-0.60	56.9	72.4	76.1	70.5	$\chi^2(2) = 8.83$, $p < .05$
a21_Fehler	0.23	45.8	41.4	62.6	54.0	$\chi^2(2) = 10.25$, $p < .01$

Tabelle 4.9 fasst die Lösungshäufigkeiten und Aufgabenschwierigkeiten der Aufgaben der Teilkompetenz Theorieverständnis zusammen. Diese drei Aufgaben waren für die Grundschul Kinder insgesamt am schwierigsten.

Tabelle 4.9

Lösungshäufigkeiten (in Prozent) der Aufgaben der Teilkompetenz Theorieverständnis geordnet nach ansteigender Aufgabenschwierigkeit getrennt nach Jahrgangsstufe

Item	Schwierigkeit	Jahrgangsstufe				χ^2 -Test
		2	3	4	gesamt	
a27_Geburtstagsparty	1.04	12.5	27.6	52.3	37.2	$\chi^2(2) = 36.14$, $p < .001$
a24_Mittelalter	1.36	22.2	29.3	35.5	30.9	$\chi^2(2) = 4.14$, $p > .05$
a23_Wichtel	1.62	16.7	22.4	32.5	26.4	$\chi^2(2) = 6.90$, $p < .05$

In Tabelle 4.10 sind die Aufgaben der Teilkompetenz Experimentierstrategien nach zunehmender Aufgabenschwierigkeit angeordnet. Die Beurteilung eines einfachen konklusiven Experiments (Aufgabe a51_Maus-Haus) stellte für die Grundschul Kinder die leichteste Aufgabe dar. Die Kinder schnitten auch gut ab, wenn sie das Verständnis der Variation der fokalen Variablendimension beurteilen sollten (Aufgabe a49_Pizza und die erste Teilauf-

gabe der Aufgabe a44_Ärzte). Von mittlerer Schwierigkeit war die Beurteilung eines Experiments unter Berücksichtigung der Basisrate bei einem Prä-Post-Design (Teilaufgabe 1 der Aufgabe a57_Fußpilz) sowie die Wahl eines konklusiven Tests (Aufgabe a53_Spürnase). Eine mittlere bis hohe Aufgabenschwierigkeit besaßen Aufgaben, die die Produktion einer Begründung sowohl für die Notwendigkeit der Variation der fokalen Variablendimension (Aufgabe a44_2_Ärzte) als auch für die Bedeutung der Berücksichtigung der Baseline (Aufgabe a57_2_Fußpilz) erforderten. Die Wahl der Variablenkontrollstrategie (Aufgabe a35_Flugzeug), um valide Schlussfolgerung aus experimentellen Versuchsanordnungen ziehen zu können, hatte ebenfalls eine relativ hohe Aufgabenschwierigkeit.

Tabelle 4.10

Lösungshäufigkeiten (in Prozent) der Aufgaben der Teilkompetenz Experimentierstrategien geordnet nach ansteigender Aufgabenschwierigkeit getrennt nach Jahrgangsstufe

Item	Schwierigkeit	Jahrgangsstufe				χ^2 -Test
		2	3	4	gesamt	
a51_Maus Haus	-1.70	68.1	86.2	94.8	86.3	$\chi^2(2) = 29.86$, $p < .001$
a49_Pizzabacken	-0.94	62.5	70.2	85.1	76.3	$\chi^2(2) = 15.32$, $p < .001$
a44_1_Ärzte	-0.46	65.3	60.3	72.1	68.0	$\chi^2(2) = 2.98$, $p > .05$
a57_1_Fußpilz	0.04	44.4	41.4	70.8	58.1	$\chi^2(2) = 22.35$, $p < .001$
a53_Spürnase	0.09	41.7	41.4	70.1	57.0	$\chi^2(2) = 23.52$, $p < .001$
a44_2_Ärzte	0.63	36.1	50.0	48.4	45.6	$\chi^2(2) = 3.55$, $p > .05$
a35_Flugzeug	0.73	29.6	27.6	56.2	43.6	$\chi^2(2) = 21.62$, $p < .001$
a57_2_Fußpilz	0.80	25.0	34.5	52.9	42.1	$\chi^2(2) = 17.44$, $p < .001$

Die Schwierigkeiten der Aufgaben zur Dateninterpretation sind in Tabelle 4.11 gegenübergestellt. Die Richtigkeit der Darstellung einer Null-Korrelation zwischen zwei Variablen entgegen den eigenen Überzeugungen in einer Tabelle (Aufgabe a82_Tabelle) sowie der Darstellung eines linearen Zusammenhangs zweier Variablen in der Repräsentationsform

eines Diagramms (Aufgabe a76_Diagramm) konnten die Kinder gut beurteilen. Mittelschwer bis schwer waren dagegen die Aufgaben zur Interpretation kovariierender Daten von Variablen mit je zwei Variablen Dimensionen. Dabei war die Aufgabe zur Interpretation von Kovariationsdaten im artifiziellen Kontext (Aufgabe a66_Ravianer), wodurch keine Interferenz zu den Überzeugungen der Kinder vorlag, leichter als die Beurteilung von Kovariationsdaten, bei denen die Kinder Annahmen über Kausalzusammenhänge hatten (Aufgabe a64_Zähne).

Tabelle 4.11

Lösungshäufigkeiten in Prozent der Aufgaben der Teilkompetenz Dateninterpretation geordnet nach ansteigender Aufgabenschwierigkeit getrennt nach Jahrgangsstufe

Item	Schwierigkeit	Jahrgangsstufe				χ^2 -Test
		2	3	4	gesamt	
a82_Tabelle	-0.59	56.9	63.8	79.2	70.4	$\chi^2(2) = 13.23$, $p < .001$
a76_Diagramm	-0.39	63.9	56.9	71.6	66.7	$\chi^2(2) = 4.45$, $p > .05$
a66_Ravianer	0.31	35.2	51.7	60.8	52.5	$\chi^2(2) = 12.73$, $p < .01$
a64_Zähne	0.86	16.7	34.5	54.5	40.8	$\chi^2(2) = 30.36$, $p < .001$

Werden die Schwierigkeiten der Items je Teilkompetenz des wissenschaftlichen Denkens zu einem Mittelwert zusammengefasst, ergab sich folgende ansteigende Schwierigkeitsreihenfolge: Verständnis der Ziele von Wissenschaften ($M = -0.68$, $SD = 0.58$), Experimentierstrategien ($M = -0.10$, $SD = 0.89$), Dateninterpretation ($M = 0.05$, $SD = 0.67$) und Theorieverständnis ($M = 1.34$, $SD = 0.29$).

Eine detaillierte Betrachtung der Schwierigkeiten der einzelnen Teilkompetenzen des wissenschaftlichen Denkens ist in Abschnitt 4.2.3 (S. 107) beschrieben.

4.2.2 Kompetenzwerte

Die Berechnung der Personenfähigkeiten nach dem Weighted-Likelihood-Schätzer (WLE) erfolgte im Rahmen der eindimensionalen Rasch-Analyse für die erhobenen Daten der gesamten Stichprobe.

Zur Interpretation der Personenfähigkeit kann die in Abbildung 4.1 (S. 88) dargestellte Wright map, die die Personenfähigkeiten und Itemschwierigkeiten auf einer gemeinsamen Logit-Skala verortet, herangezogen werden. Die mittlere Itemschwierigkeit wurde auf Null restringiert und stellte die Metrik der Skala dar. Dadurch konnten die Personenfähigkeiten und Itemschwierigkeiten in Beziehung gesetzt werden. Allgemein beschreiben negative WLE-Werte geringe Personenfähigkeiten und positive Werte hohe Merkmalsausprägungen auf dem Fähigkeitskontinuum.

4.2.2.1 Kompetenz wissenschaftliches Denken – Gesamttest

In Tabelle 4.12 sind die Personenfähigkeiten in der Kompetenz wissenschaftliches Denken für die Gesamtstichprobe und die einzelnen Klassenstufen getrennt nach Geschlecht dargestellt.

Tabelle 4.12

Mittlere Personenfähigkeiten (WLE) im wissenschaftlichen Denken aufgeteilt nach Geschlecht für die Gesamtstichprobe und die einzelnen Klassenstufen

Jahrgangsstufe	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Minimum</i>	<i>Maximum</i>
2	72	-0.29	0.87	-2.92	1.53
Jungen	31	-0.14	0.72	-1.53	1.23
Mädchen	41	-0.40	0.96	-2.92	1.53
3	58	0.06	0.89	-1.88	1.88
Jungen	35	-0.01	0.91	-1.88	1.88
Mädchen	23	0.17	0.88	-1.88	1.23
4	155	0.87	0.93	-1.88	4.11
Jungen	74	0.88	0.94	-1.53	4.11
Mädchen	78	0.86	0.93	-1.88	2.92
Gesamt	285	0.41	1.04	-2.92	4.11
Jungen	140	0.43	1.00	-1.88	4.11
Mädchen	142	0.39	1.08	-2.92	2.92

Anmerkung: *N* = Anzahl der Schüler, *M* = Mittelwert, *SD* = Standardabweichung. Bei drei Viertklässlern fehlte die Angabe des Geschlechts.

Um die Kompetenz wissenschaftliches Denken hinsichtlich Geschlechterdifferenzen und Altersunterschiede zu untersuchen, wurden die Kompetenzwerte der Mädchen und Jungen sowie der einzelnen Klassenstufen miteinander verglichen.

Eine univariate zweifaktorielle Varianzanalyse mit Geschlecht (Jungen, Mädchen) und Klassenstufe (Klasse 2, Klasse 3, Klasse 4) als unabhängige Variablen ergab einen signifikanten Haupteffekt für Klassenstufe ($F(2,276) = 43.17, p < .001$). Der Haupteffekt Geschlecht ($F(1,276) = 0.07, p > .05, ns$) sowie die Interaktion zwischen den Faktoren Geschlecht und Klassenstufe ($F(2,276) = 0.91, p > .05, ns$) waren nicht signifikant.

In der Gesamtstichprobe betrug die mittlere Kompetenz der Grundschüler 0.43 logits ($SD = 1.00$) und der Grundschülerinnen 0.39 logits ($SD = 1.08$). Auch in den einzelnen Jahrgangsstufen erreichten die Mädchen und Jungen im Durchschnitt vergleichbare Kompetenzwerte (vgl. Tabelle 4.12). Es zeigten sich also weder in der gesamten Stichprobe noch in den einzelnen Jahrgangsstufen signifikante Mittelwertsunterschiede zwischen Mädchen und Jungen in der Kompetenz wissenschaftliches Denken.

Betrachtet man die WLE-Personenfähigkeiten im wissenschaftlichen Denken der eindimensionalen Raschskalierung getrennt für die einzelnen Klassenstufen, um die Kompetenzentwicklung von der zweiten bis zur vierten Klassenstufe der Grundschule nachzuzeichnen, zeigte sich, dass die Personenfähigkeiten von der zweiten Klassenstufe mit einem WLE-Mittelwert von -0.29 logits ($SD = 0.87$) zur dritten Klassenstufe mit einer durchschnittlichen Kompetenz von 0.06 logits ($SD = 0.89$) bis zur vierten Jahrgangsstufe, die im Durchschnitt eine Fähigkeit von 0.87 logits ($SD = 0.93$) erreichte, anstieg. Zweitklässler lösten durchschnittlich 44.5% bzw. 8.9 Items. Drittklässler beantworteten im Mittel 51.3% der Items bzw. 10.2 Items richtig. Mit 66.6% bzw. 13.3 gelösten Items schnitt die Gruppe der Viertklässler am besten ab.

Zusätzlich wurden a-priori-Kontraste berechnet, um Gruppenunterschiede zwischen den Klassenstufen, insbesondere zwischen Klassenstufe 2 und 3 sowie zwischen Klassenstufe 3 und 4 zu ermitteln. Der Vergleich der Personenfähigkeiten der zweiten und dritten Klassenstufe zeigte auf einem Signifikanzniveau von 5% einen signifikanten Unterschied, der sich allerdings auf einem strengeren Signifikanzniveau von 1% nicht mehr als signifikant erwies ($t(282) = -2.17, p = .031$). Zwischen Jahrgangsstufe 3 und 4 konnte ein signifikanter Kompetenzunterschied festgestellt werden ($t(282) = -5.84, p < .001$). Werden Klassenstufe 2 und 3 zu der Stichprobe des frühen Grundschulalters zusammengefasst und mit der Klassenstufe 4 als spätes Grundschulalter verglichen, ergab sich ebenfalls ein signifikanter Gruppenunterschied ($t(282) = -9.14, p < .001$).

Werden die Leistungen im Test zum Leseverständnis und zur nonverbalen Intelligenz als Kovariaten in die ANOVA mit dem Faktor Jahrgangsstufe aufgenommen, um den Effekt dieser Korrelate auf die Kompetenz im wissenschaftlichen Denken zu kontrollieren, zeigten sich folgende Ergebnisse: Die Kovariaten Leseverständnis ($F(1,265) = 34.00, p < .001$) und Intelligenz ($F(1,265) = 6.84, p < .01$) hingen signifikant mit der Kompetenz im wissenschaftlichen Denken zusammen. Der Effekt von Klassenstufe auf die Kompetenz im wissenschaftlichen Denken blieb signifikant, nachdem der Effekt des Leseverständnisses und der Intelligenz kontrolliert worden war ($F(2,265) = 15.85, p < .001$). Geplante Kontraste ergaben keinen signifikanten Unterschied in der Leistung im wissenschaftlichen Denken zwischen Klassenstufe 2 und 3 ($t(265) = 0.31, p > .05, ns$), aber einen signifikanten Unterschied zwischen Klassenstufe 3 und 4 ($t(265) = 4.91, p < .001$).

4.2.2.2 Kompetenz wissenschaftliches Denken – Subskalen Komponenten

Die Personenfähigkeiten in den Komponenten Wissenschaftsverständnis und Methodenkompetenz sind für die Gesamtstichprobe und die einzelnen Klassenstufen getrennt nach Geschlecht in Tabelle 4.13 zusammengefasst.

Tabelle 4.13

Mittlere Personenfähigkeiten (WLE) in den Komponenten des wissenschaftlichen Denkens aufgeteilt nach Jahrgangsstufe und Geschlecht

	Jahrgangsstufe							
	2		3		4		gesamt	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
	Wissenschaftsverständnis							
Jungen	-0.23	0.84	0.08	1.02	0.77	0.87	0.37	0.99
Mädchen	-0.45	0.86	0.00	0.88	0.69	1.00	0.25	1.06
gesamt	-0.35	0.85	0.05	0.96	0.73	0.93	0.31	1.03
	Methodenkompetenz							
Jungen	-0.08	0.98	-0.04	1.31	0.98	1.26	0.49	1.32
Mädchen	-0.38	1.19	0.30	1.10	1.02	1.27	0.50	1.36
gesamt	-0.25	1.11	0.09	1.23	1.00	1.26	0.49	1.34

Da für die beiden Komponenten Wissenschaftsverständnis und Methodenkompetenz WLE-Personenparameter vorliegen, wurde zunächst eine multivariate zweifaktorielle Varianzanalyse mit Geschlecht und Klassenstufe als unabhängige Variablen gerechnet. Der Pillai Spur-Test deckte einen signifikanten Effekt von Klassenstufe ($F(4,552) = 20.36, p < .001$) auf. Es zeigte sich weder ein Geschlechtereffekt ($F(2,275) = 0.66, p > .05, ns$) noch eine Interaktion zwischen den Faktoren Geschlecht und Klassenstufe ($F(4,552) = 0.57, p > .05, ns$). Anschließend gerechnete univariate zweifaktorielle Varianzanalysen ergaben je Komponente das gleiche Ergebnismuster (vgl. Tabelle 4.14): Es fanden sich signifikante Effekte für Klassenstufe. Weder Geschlechtereffekte noch Interaktionseffekte traten auf.

Tabelle 4.14

Statistische Kennwerte der univariaten zweifaktoriellen (Geschlecht, Klassenstufe) Varianzanalysen der Komponenten des wissenschaftlichen Denkens

	Teststatistik <i>F</i>	Freiheitsgrade <i>df (276)</i>	Signifikanzniveau <i>p</i>
Wissenschaftsverständnis			
Klassenstufe	35.39	2	< .001
Geschlecht	1.05	1	> .05
Geschlecht x Klassenstufe	0.16	2	> .05
Methodenkompetenz			
Klassenstufe	28.00	2	< .001
Geschlecht	0.03	1	> .05
Geschlecht x Klassenstufe	1.09	2	> .05

Die Betrachtung der Mittelwerte zeigte, dass der signifikante Haupteffekt Klassenstufe für einen deutlichen Anstieg der Leistung in den Komponenten Wissenschaftsverständnis und Methodenkompetenz von der zweiten zur vierten Klassenstufe steht.

Um die Alterseffekte differenziert je Komponente des wissenschaftlichen Denkens zu untersuchen, wurden anschließend univariate Varianzanalysen mit dem Faktor Klassenstufe und geplante Kontraste gerechnet.

Der Vergleich der Leistungen in der Komponente Wissenschaftsverständnis zeigte signifikante Unterschiede zwischen den drei Jahrgangsstufen ($F(2,282) = 37.53, p < .001$). Die Drittklässler waren den Zweitklässlern signifikant überlegen ($t(282) = -2.47, p < .05$). Auch die Viert- und Drittklässler unterschieden sich signifikant in ihrer Leistung im Wissenschaftsverständnis ($t(282) = -4.85, p < .001$). Der Vergleich des frühen versus späten Grundschulalters ergab einen signifikanten Wert ($t(282) = 8.08, p < .001$).

Im Bereich der Methodenkompetenz bestanden signifikante Leistungsunterschiede zwischen den drei Klassenstufen ($F(2,282) = 30.02, p < .001$). Es konnte kein signifikanter Unterschied in den Leistungswerten der Komponente Methodenkompetenz zwischen den Schülern der zweiten und dritten Jahrgangsstufe festgestellt werden ($t(282) = -1.60, p > .05, ns$). Im Gegensatz dazu zeigte sich ein signifikanter Vorteil der Viertklässler gegenüber den Drittklässlern ($t(282) = -4.84, p < .001$) sowie gegenüber der aus Klassenstufe 2 und 3 zusammengefassten Stichprobe ($t(282) = 7.43, p < .001$).

Wenn die Leistungen im Test zum Leseverständnis und zur nonverbalen Intelligenz als Kovariaten in die ANOVA mit dem Faktor Jahrgangsstufe aufgenommen werden, um den Effekt dieser Korrelate auf die Kompetenz im wissenschaftlichen Denken zu kontrollieren, waren folgende Ergebnisse festzustellen: Die Kovariate Leseverständnis ($F(1,265) = 9.55, p < .01$) hing signifikant mit der Kompetenz im Wissenschaftsverständnis zusammen. Für die Kovariate Intelligenz zeigte sich kein signifikanter Effekt ($F(1,265) = 1.89, p > .05$). Der Effekt des Faktors Klassenstufe auf die Kompetenz im wissenschaftlichen Denken blieb nach Kontrolle der Effekte von Leseverständnis und Intelligenz bestehen ($F(2,265) = 16.00, p < .001$). Geplante Kontraste zeigten keinen signifikanten Unterschied in der Leistung im wissenschaftlichen Denken zwischen Klassenstufe 2 und 3 ($t(265) = -1.12, p > .05, ns$), aber einen signifikanten Unterschied zwischen Klassenstufe 3 und 4 ($t(265) = 4.08, p < .001$).

Die Methodenkompetenz hing sowohl mit der Kovariate Leseverständnis ($F(1,265) = 30.26, p < .001$) als auch mit der Kovariate Intelligenz ($F(1,265) = 6.83, p < .01$) signifikant zusammen. Nach Kontrolle der Effekte von Leseverständnis und Intelligenz blieb der Effekt des Faktors Klassenstufe auf die Methodenkompetenz signifikant ($F(2,265) = 8.65, p < .001$). Geplante Kontraste zeigten keinen signifikanten Unterschied in der Leistung im wissenschaftlichen Denken zwischen Jahrgangsstufe 2 und 3 ($t(265) = 0.87, p > .05, ns$), aber einen signifikanten Unterschied zwischen Klassenstufe 3 und 4 ($t(265) = 3.90, p < .001$).

4.2.2.3 Kompetenz wissenschaftliches Denken – Subskalen Teilkompetenzen

Im Folgenden werden die einzelnen Teilkompetenzen des wissenschaftlichen Denkens im Hinblick auf Geschlechter- und Altersunterschiede betrachtet. Je Klassenstufe sind die Mittelwerte der Mädchen und Jungen für die vier Teilkompetenzen Verständnis der Ziele

von Wissenschaften und Theorieverständnis sowie Experimentierstrategien und Dateninterpretation in Tabelle 4.15 dargestellt. Die Überlegenheit der Viertklässler spiegelte sich in allen Teilbereichen des wissenschaftlichen Denkens wieder.

Tabelle 4.15

Mittlere Personenfähigkeiten (WLE) in den Teilkompetenzen des wissenschaftlichen Denkens aufgeteilt nach Jahrgangsstufe und Geschlecht

	Jahrgangsstufe							
	2		3		4		gesamt	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
	Verständnis der Ziele von Wissenschaften							
Jungen	0.54	0.91	0.71	1.03	1.37	0.93	1.02	1.01
Mädchen	0.27	1.09	0.61	0.85	1.28	1.08	0.88	1.14
gesamt	0.39	1.02	0.67	0.95	1.32	1.00	0.95	1.08
	Theorieverständnis							
Jungen	-1.33	0.94	-0.88	1.04	-0.37	1.11	-0.71	1.12
Mädchen	-1.22	0.86	-0.89	1.05	-0.35	1.02	-0.69	1.05
gesamt	-1.27	0.89	-0.88	1.04	-0.36	1.06	-0.70	1.08
	Experimentierstrategien							
Jungen	0.02	1.24	-0.03	1.45	0.99	1.40	0.52	1.46
Mädchen	-0.36	1.45	0.29	1.17	1.13	1.32	0.56	1.48
gesamt	-0.19	1.37	0.10	1.34	1.06	1.36	0.54	1.47
	Dateninterpretation							
Jungen	-0.33	1.02	-0.07	1.25	0.84	1.29	0.35	1.33
Mädchen	-0.28	1.07	0.33	1.38	0.65	1.35	0.33	1.33
gesamt	-0.30	1.04	0.09	1.31	0.75	1.32	0.34	1.33

Anmerkung: *M* = Mittelwert, *SD* = Standardabweichung.

Da für jeden der vier Teilbereiche des wissenschaftlichen Denkens WLE-Personenparameter vorlagen, wurde zunächst eine multivariate zweifaktorielle Varianzanalyse mit Geschlecht und Klassenstufe als unabhängige Variablen gerechnet. Die Verwendung des Pillai Spur-Tests zeigte einen signifikanten Effekt von Klassenstufe ($F(8,548) = 10.90, p < .001$). Der Geschlechtereffekt ($F(4,273) = 0.51, p > .05, ns$) sowie die Interaktion zwischen den Faktoren Geschlecht und Klassenstufe ($F(8,548) = 0.78, p > .05, ns$) waren nicht signifikant. Anschließend gerechnete univariate zweifaktorielle Varianzanalysen stellten je Teilkomponente das gleiche Ergebnismuster heraus (vgl. Tabelle 4.16): Es

fanden sich signifikante Effekte für Klassenstufe. Weder Geschlechtereffekte noch Interaktionseffekte traten auf.

Tabelle 4.16

Statistische Kennwerte der univariaten zweifaktoriellen (Geschlecht, Klassenstufe) Varianzanalysen der Teilkompetenzen des wissenschaftlichen Denkens

	Teststatistik <i>F</i>	Freiheitsgrade <i>df</i> (276)	Signifikanzniveau <i>p</i>
Verständnis der Ziele von Wissenschaften			
Klassenstufe	23.39	2	< .001
Geschlecht	1.39	1	> .05
Geschlecht x Klassenstufe	0.21	2	> .05
Theorieverständnis			
Klassenstufe	20.44	2	< .001
Geschlecht	0.08	1	> .05
Geschlecht x Stufe	0.05	2	> .05
Experimentierstrategien			
Klassenstufe	23.23	2	< .001
Geschlecht	0.02	1	> .05
Geschlecht x Stufe	1.26	2	> .05
Dateninterpretation			
Klassenstufe	18.08	2	< .001
Geschlecht	0.27	1	> .05
Geschlecht x Stufe	1.14	2	> .05

Betrachtet man die Mittelwerte, dann verbirgt sich hinter dem Haupteffekt Klassenstufe ein deutlicher Anstieg der Leistung in allen Teilkompetenzen von der zweiten zur vierten Klassenstufe.

Die Ergebnisse der anschließend gerechneten univariaten Varianzanalysen je Teilkompetenz mit dem Faktor Klassenstufe und den a-priori-Kontrasten zwischen Jahrgangsstufe 2 und 3 sowie zwischen Klassenstufe 3 und 4, um die Alterseffekte differenziert je Teilbereich zu betrachten, sind im Folgenden dargestellt:

Für die Teilkompetenz Verständnis der Ziele von Wissenschaften ergab die univariate Varianzanalyse mit dem dreifachgestuften Faktor Klassenstufe (Klasse 2, Klasse 3, Klasse 4) einen signifikanten Haupteffekt für Klassenstufe ($F(2,282) = 25.03, p < .001$). Während die a-priori-Kontraste keinen signifikanten Unterschied zwischen Zweit- und Drittklässlern in ihrer Leistung im Verständnis der Ziele von Wissenschaften zeigten ($t(282) = -1.58, p >$

.05, ns), erzielten die Viertklässler signifikant höhere Werte als die Drittklässler ($t(282) = -4.32, p < .001$). Der Vergleich der Gruppen frühes Grundschul- (Klassenstufe 2 und 3) und spätes Grundschulalter (Klasse 4) wurde ebenfalls signifikant ($t(282) = 6.75, p < .001$).

In der Teildimension Theorieverständnis konnten signifikante Unterschiede zwischen den drei Klassenstufen festgestellt werden ($F(2,282) = 21.16, p < .001$). Die Drittklässler erreichten signifikant bessere Werte im Theorieverständnis als die Zweitklässler ($t(282) = -2.15, p < .05$). Auch die Viertklässler unterschieden sich signifikant von den Drittklässlern ($t(282) = -3.38, p < .01$). Die aus Klassenstufe 2 und 3 zusammengefasste Stichprobe erbrachte eine signifikant schlechtere Leistung als Klassenstufe 4 ($t(282) = 5.95, p < .001$).

Der Vergleich der Leistungen in der Teilkomponente Experimentierstrategien zeigte signifikante Unterschiede zwischen den drei Jahrgangsstufen ($F(2,282) = 24.88, p < .001$). Die Viertklässler waren den Drittklässlern signifikant überlegen ($t(282) = -4.60, p < .001$), während sich die Zweit- und Drittklässler nicht signifikant in ihrer Leistung im Experimentieren unterschieden ($t(282) = -1.21, p > .05, ns$). Auch der Vergleich des frühen versus späten Grundschulalters ergab einen signifikanten Wert ($t(282) = 6.83, p < .001$).

Im Bereich der Dateninterpretation ergaben sich signifikante Leistungsunterschiede zwischen den drei Klassenstufen ($F(2,136) = 21.63, p < .001$). Es konnte ein tendenzieller Unterschied in den Leistungswerten bei der Interpretation von Daten zwischen den Schülern der zweiten und dritten Jahrgangsstufe festgestellt werden ($t(108) = -1.85, p = .067, ns$). Im Gegensatz dazu zeigte sich ein signifikanter Vorteil der Viertklässler gegenüber den Drittklässlern ($t(103) = -3.29, p < .001$) sowie gegenüber der aus Klassenstufe 2 und 3 zusammengefassten Stichprobe ($t(254) = 5.75, p < .001$).

Um den Effekt der Korrelate Leseverständnis und Intelligenz auf die Leistungen in den einzelnen Teilkompetenzen des wissenschaftlichen Denkens zu berücksichtigen, wurden die Leistungen in den Tests zum Leseverständnis und zur nonverbalen Intelligenz als Kovariaten in die ANOVA mit dem Faktor Jahrgangsstufe aufgenommen. Nachfolgend werden die Ergebnisse je Teilkompetenz dargestellt:

Die Kovariate Leseverständnis hing nur tendenziell mit der Kompetenz im Verständnis der Ziele von Wissenschaften zusammen ($F(1,265) = 3.26, p = .072, ns$). Für die Kovariate Intelligenz ($F(1,265) = 1.04, p > .05, ns$) zeigte sich kein signifikanter Effekt auf die Leistung im Verständnis der Ziele von Wissenschaften. Der Effekt des Faktors Klassenstufe auf die Kompetenz im Verständnis der Ziele von Wissenschaften blieb nach der Kontrolle von Leseverständnis und Intelligenz bestehen ($F(2,265) = 12.08, p < .001$). Geplante Kontraste stellten keinen signifikanten Unterschied in der Leistung im Verständnis der Ziele

von Wissenschaften zwischen Klassenstufe 2 und 3 fest ($t(265) = -0.76, p > .05, ns$), aber einen signifikanten Unterschied zwischen Klassenstufe 3 und 4 ($t(265) = 3.69, p < .001$).

Für die Leistung im Theorieverständnis trat ein signifikanter Zusammenhang zur Lesekompetenz auf ($F(1,265) = 10.89, p < .001$). Die Kovariate Intelligenz ($F(1,265) = 0.31, p > .05, ns$) hing nicht signifikant mit der Kompetenz im Theorieverständnis zusammen. Nach Kontrolle der Effekte von Leseverständnis und Intelligenz blieb der Effekt des Faktors Klassenstufe auf das Theorieverständnis bestehen ($F(2,265) = 6.71, p < .001$). Geplante Kontraste zeigten keinen signifikanten Unterschied in der Leistung im Theorieverständnis zwischen Jahrgangsstufe 2 und 3 ($t(265) = -0.76, p > .05, ns$). Der Unterschied zwischen Klassenstufe 3 und 4 ($t(265) = 2.62, p < .01$) wurde signifikant.

Die Kovariaten Leseverständnis ($F(1,265) = 30.63, p < .001$) und Intelligenz ($F(1,265) = 7.61, p < .01$) hatten signifikante Effekte auf das Verständnis der experimentellen Methodik. Der Effekt des Faktors Klassenstufe auf die Leistung in der Teilkompetenz Experimentierstrategien wurde auch nach Kontrolle von Leseverständnis und Intelligenz signifikant ($F(2,265) = 7.86, p < .001$). Geplante Kontraste stellten keinen signifikanten Unterschied im Verständnis der experimentellen Arbeitsweise zwischen Zweit- und Drittklässlern fest ($t(265) = 1.51, p > .05, ns$), aber einen signifikanten Unterschied zwischen Klassenstufe 3 und 4 ($t(265) = 3.90, p < .001$).

Bei der Leistung im Bereich Dateninterpretation zeigte sich ein signifikanter Zusammenhang zur Lesekompetenz ($F(1,265) = 14.19, p < .001$), aber nicht zur nonverbalen Intelligenz ($F(1,265) = 1.70, p > .05, ns$). Der Effekt des Faktors Jahrgangsstufe auf die Dateninterpretationskompetenz blieb nach Kontrolle der Effekte von Leseverständnis und Intelligenz signifikant ($F(2,265) = 4.47, p < .05$). Geplante Kontraste ergaben keinen signifikanten Unterschied in der Interpretation von Evidenz zwischen Jahrgangsstufe 2 und 3 ($t(265) = -0.43, p > .05, ns$). Der Unterschied zwischen Klassenstufe 3 und 4 ($t(265) = 2.27, p < .05$) wurde signifikant.

Die zusammenfassende Betrachtung des Altersvergleichs im Hinblick auf die Leistung der einzelnen Jahrgangsstufen in den Teilkompetenzen machte deutlich, dass die Schüler der vierten Klassenstufe konsistent in allen Teilbereichen den Zweit- und Drittklässlern überlegen waren. Die Schüler der zweiten und dritten Jahrgangsstufe unterschieden sich nur signifikant in der Leistung im Theorieverständnis, wobei dieser Unterschied auf einem strengeren Signifikanzniveau von 1 % bzw. unter Berücksichtigung der Effekte von Leseverständnis und Intelligenz nicht signifikant blieb. Die durchschnittlichen Leistungen

(WLE-Personenparameter) der einzelnen Klassenstufen in den Teilkompetenzen des wissenschaftlichen Denkens sind in Abbildung 4.7 zusammengefasst.

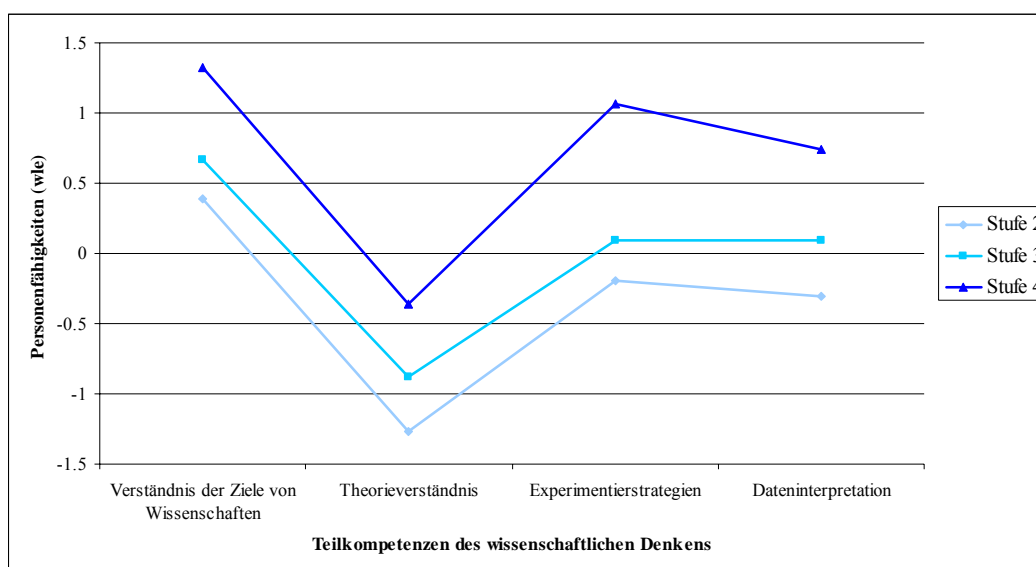


Abbildung 4.7. Durchschnittliche Leistung der einzelnen Klassenstufen je Teilkompetenz des wissenschaftlichen Denkens

Insgesamt konnte anhand der ähnlichen Kompetenzprofile auf unterschiedlichem Leistungsniveau festgestellt werden, dass den Kindern der drei Jahrgangsstufen die Aufgaben im Teilbereich Verständnis der Ziele von Wissenschaften am leichtesten fielen, gefolgt von Aufgaben zu Experimentierstrategien und Dateninterpretation. Allerdings zeigten sowohl Zweit-, Dritt- als auch Viertklässler schwächere Leistungen in der Teilkompetenz Theorieverständnis. Der Entwicklungsverlauf in den einzelnen Bereichen wird im folgenden Abschnitt genauer betrachtet.

4.2.3 Sequenzieller Entwicklungsverlauf

Um Entwicklungssequenzen im wissenschaftlichen Denken zu untersuchen, wurden die Antwortmuster auf die nach Teilkompetenzen zusammengefassten Aufgabensets bewertet (Wu & Adams, 2007). Entsprechend der Operationalisierung der Aufgabenentwicklung konnten vier Aufgabensets zu den Teilkompetenzen Verständnis der Ziele von Wissenschaften und Theorieverständnis sowie Experimentierstrategien und Dateninterpretation zusammengefasst werden (vgl. Tabelle 3.2, S. 51).

Bei jedem Aufgabenset mussten mindestens 65 % der Aufgaben korrekt beantwortet werden, damit das Aufgabenset der Teilkompetenz als richtig gewertet wurde. Für jedes richtig gelöste Aufgabenset wurde ein Punkt vergeben, so dass jedes Kind im Hinblick auf die Teilkompetenzen zwischen 0 und 4 Punkte erreichen konnte. Die vier Teilkompetenz-Aufgabensets wurden in einem Rasch-Modell skaliert.

Tabelle 4.17 zeigt die dichotom kodierten Aufgabensets der Teilkompetenzen des wissenschaftlichen Denkens angeordnet nach ansteigender Schwierigkeit aus der Rasch-Analyse. Ein höherer Wert beschreibt eine größere Schwierigkeit.

Tabelle 4.17

Statistische Kennwerte des eindimensionalen Rasch-Modells der Aufgabensets der Teilkompetenzen des wissenschaftlichen Denkens

Teilkompetenz-Aufgabenset	N	Lösungshäufigkeit (%)	Schwierigkeit	Trennschärfe	Infit MNSQ	T-Wert
Theorieverständnis	285	23.16	1.75	0.57	1.08	0.9
Dateninterpretation	285	41.75	0.57	0.73	0.96	-0.6
Experimentierstrategien	285	57.19	-0.31	0.75	0.96	-0.6
Verständnis der Ziele von Wissenschaften	285	82.46	-2.02	0.54	1.01	0.2
Skala	EAP/PV-Reliabilität: 0.56 Varianz: 2.18 Trennschärfe: $M = 0.65$, $SD = 0.11$, $MIN = 0.54$, $MAX = 0.75$ Schwierigkeit: $M = 0$, $SD = 1.59$, $MIN = -2.02$, $MAX = 1.75$ Infit: $M = 1.00$, $SD = 0.06$, $MIN = 0.96$, $MAX = 1.08$					

Der Nullpunkt der Skala wurde durch die Restriktion der mittleren Schwierigkeit auf 0 logits ($SD = 1.59$) festgelegt. Die Schwierigkeiten reichten von -2.02 bis 1.75 logits. Die Fit-Statistiken für die Aufgabenparameter waren zufrieden stellend. Es konnten keine signifikanten T-Werte ($-2.0 < T\text{-Wert} < 2.0$) festgestellt werden. Der Infit (Weighted Mean Square, MNSQ) lag für alle Aufgabensets innerhalb des von den Richtlinien vorgegebenen Bereichs mit einer Untergrenze von 0.96 und einer Obergrenze von 1.08 (Adams, 2002). Der Mittelwert der MNSQ-Werte aller Items lag bei 1.00 ($SD = 0.06$). Die Fit-Statistiken zeigten, dass die Daten gut durch das Rasch-Modell modelliert werden konnten und sich die Aufgabensets modellkonform verhielten. Die punktbiserialen Korrelationen bzw. klassischen Trennschärfen lagen zwischen 0.54 und 0.75. Die mittlere Trennschärfe der Aufgabensets betrug 0.65 ($SD = 0.11$). Die Varianz der latenten Variablen lag bei 2.18. Die EAP/PV-Reliabilität erreichte einen Wert von 0.56.

Die Itemcharakteristischen Funktionen der Aufgabensets der einzelnen Teilkompetenzen sind in der nachfolgenden Abbildung 4.8 dargestellt. Ein hoher Wert auf der x-Achse steht für eine hohe Aufgabenschwierigkeit. Es bestätigte sich die in Abschnitt 4.2.1 (S. 97) aufgrund der gemittelten Aufgabenschwierigkeiten beschriebene Reihenfolge der Schwierigkeiten der Teilkompetenzen: Die Aufgaben zum Verständnis der Ziele von Wissenschaften waren für die Grundschul Kinder am leichtesten, gefolgt vom Aufgabenset Experimentierstrategien. Die Aufgaben zur Interpretation von Daten hatten die nächst höhere Schwierigkeit. Am schwersten waren die Aufgaben zum Theorieverständnis.

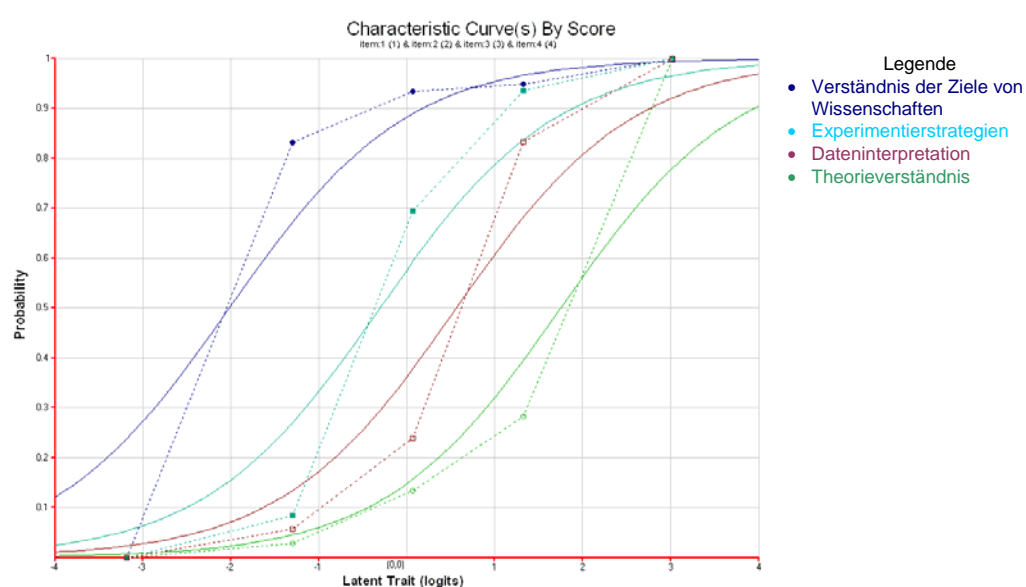


Abbildung 4.8. *Itemcharakteristische Funktionen der Aufgabensets der Teilkompetenzen des wissenschaftlichen Denkens*

Tabelle 4.18 zeigt, dass das Lösen der Teilkompetenz-Aufgabensets mit dem Alter von der zweiten zur vierten Klasse signifikant zunahm.

Tabelle 4.18

Anteil der Kinder in Prozent, die die Teilkompetenz-Aufgabensets richtig beantworteten

Aufgabenset	Schwierigkeit	Jahrgangsstufe				gesamt	χ^2 -Test
		2	3	4			
Verständnis der Ziele von Wissenschaften	-2.02	61.1	81.0	92.9	82.5	$\chi^2(2) = 34.45$, $p < .001$	
Experimentierstrategien	-0.31	38.9	44.8	70.3	57.2	$\chi^2(2) = 24.39$, $p < .001$	
Dateninterpretation	0.57	18.1	34.5	55.5	41.8	$\chi^2(2) = 29.90$, $p < .001$	
Theorieverständnis	1.75	9.7	15.5	32.3	23.2	$\chi^2(2) = 16.42$, $p < .001$	

Zusätzlich wurde eine Skalogramm-Analyse nach Guttman durchgeführt. Dieses deterministische Testmodell stellt striktere Anforderungen an die Daten. Es wird angenommen, dass eine Person, die eine bestimmte Aufgabe löst, auch alle leichteren Aufgaben löst. Daher dürfen nur bestimmte Antwortmuster, sogenannte Guttman-Pattern, auftreten. Andere Antwortmuster sind nicht zulässig. Für die Skala der vier Teilkompetenz-Aufgabensets ergaben sich fünf zulässige Antwortmuster. Um das Modell beim Auftreten eines unzulässigen Antwortmusters nicht verwerfen zu müssen, kann das Reproduzierbarkeitsmaß nach Green (1956) berechnet werden. Dieser Index beschreibt basierend auf der Anzahl unzulässiger Itemantworten den relativen Anteil modellkonformer Itemantworten in der Datenmatrix. Das Reproduzierbarkeitsmaß zeigt daher das Ausmaß an, inwieweit die Daten Modellpassung zeigen (Rost, 2004). Die Aufgaben können als skalierbar bewertet werden, wenn dieser Index einen Wert größer als .90 erreicht. Der relative Anteil der Klasse der „Skalierbaren“ ist ebenfalls ein Indikator für Modellpassung.

Für die Teilkompetenz-Aufgabensets zeigten 78.2% der Kinder die deterministische Guttman-Pattern (vgl. Tabelle 4.19). Das Reproduzierbarkeitsmaß mit $Rep = .94$ bestätigte, dass das Modell der Guttman-Skala auf die Aufgabensets passt. Es zeigte sich ein signifikanter Zusammenhang der sukzessiv gelösten Aufgabenmuster nach Guttman mit steigender Klassenstufe ($r = .48$, $p < .001$; $\chi^2(10) = 64.87$, $p < .001$).

Tabelle 4.19

Anteil der Kinder in Prozent, die die möglichen Antwortmuster auf die 4 Aufgabensets aufwiesen

Jahrgangsstufe	Keine Passung	0	1	2	3	4
2 (n = 72)	27.8	26.4	20.8	19.4	5.6	0
3 (n = 58)	13.8	13.8	31.0	13.8	20.7	6.9
4 (n = 155)	21.9	1.3	16.8	16.1	25.8	18.1
Gesamt (N=285)	21.8	10.2	20.7	16.5	19.6	11.2

Anmerkungen: Keine Passung: Lösungsmuster, die keinem der Muster 0 bis 4 zuzuordnen sind (z. B. Lösung des Aufgabensets Experimentierstrategien aber nicht des Aufgabensets Verständnis der Ziele von Wissenschaften); 0: kein Aufgabenset gelöst; 1: nur Aufgabenset Verständnis der Ziele von Wissenschaften gelöst; 2: Aufgabenset Verständnis der Ziele von Wissenschaften und Experimentierstrategien gelöst; 3: Aufgabenset Verständnis der Ziele von Wissenschaften, Experimentierstrategien und Dateninterpretation gelöst; 4: alle Aufgabensets gelöst.

4.2.4 Wechselseitige Beziehungen zwischen den Teilkompetenzen des wissenschaftlichen Denkens

Um die Rolle des metakonzeptuellen Theorieverständnisses auf die Leistung in den Aufgaben zu Experimentierstrategien und Dateninterpretation zu untersuchen, wurden mittels Kontingenztafeln Voraussetzungsanalysen durchgeführt. Die Tabelle 4.20 beschreibt den Anteil an Kindern, die das Experimentierstrategien- bzw. Dateninterpretations-Aufgabenset richtig gelöst haben (mindestens 65 % der Aufgaben korrekt beantwortet), als eine Funktion des metakonzeptuellen Theorieverständnisses.

Tabelle 4.20

Prozentsatz an Kindern, die das Experimentierstrategien-Aufgabenset bzw. das Dateninterpretation-Aufgabenset richtig gelöst haben, als eine Funktion des metakonzeptuellen Theorieverständnisses

	Klassifikation Theorieverständnis		χ^2 -Test
	Fehlvorstellung (n = 219)	Verständnis (n = 66)	
Verständnis von Experimentierstrategien	51.1	77.3	$\chi^2(1) = 14.15, p < .001$
Verständnis von Dateninterpretation	35.6	62.1	$\chi^2(1) = 14.65, p < .001$

Es wurde deutlich, dass Kinder, die ein höher ausgeprägtes metakonzeptuelles Theorieverständnis besaßen, signifikant bessere Leistungen in Aufgaben zu Experimentierstrategien und Dateninterpretation erzielten als Kinder, die kein Theorieverständnis zeigten.

Zur Überprüfung der Bedeutung des Verständnisses der experimentellen Methodik auf die Leistung in den Aufgaben zum Theorieverständnis und zur Dateninterpretation wurden Voraussetzungsanalysen mittels Kreuztabellen betrachtet. Der Anteil an Kindern, die das Aufgabenset zum Theorieverständnis bzw. das Dateninterpretations-Aufgabenset richtig gelöst haben (mindestens 65 % der Aufgaben korrekt beantwortet), als eine Funktion des Verständnisses von Experimentierstrategien ist in nachfolgender Tabelle 4.21 dargestellt.

Tabelle 4.21

Prozentsatz an Kindern, die das Theorieverständnis-Aufgabenset bzw. das Dateninterpretation-Aufgabenset richtig gelöst haben, als eine Funktion des Verständnisses der experimentellen Methodik

	Klassifikation Verständnis von Experimentierstrategien		χ^2 -Test
	Fehlvorstellung (n = 122)	Verständnis (n = 163)	
Theorieverständnis	12.3	31.3	$\chi^2(1) = 14.15, p < .001$
Verständnis von Dateninterpretation	19.7	58.3	$\chi^2(1) = 42.77, p < .001$

Kinder mit einem höher ausgeprägten Verständnis der experimentellen Methodik erreichten signifikant bessere Leistungen in Aufgaben zum Theorieverständnis respektive Dateninterpretation als Kinder, die kein ausgeprägtes Verständnis von wissenschaftlichen Arbeitsweisen besaßen.

Des Weiteren wurden Voraussetzungsanalysen mittels Kontingenztabelle durchgeführt, um die Rolle von Evidenzevaluationsfähigkeiten auf die Leistung in den Aufgaben zum Theorieverständnis und Experimentierstrategien zu betrachten. Tabelle 4.22 zeigt den prozentualen Anteil an Kindern, die das Aufgabenset zum Theorieverständnis bzw. das Experimentierstrategien-Aufgabenset richtig gelöst haben (mindestens 65 % der Aufgaben korrekt beantwortet), als eine Funktion des Verständnisses der Evidenzevaluation.

Tabelle 4.22

Prozentsatz an Kindern, die das Theorieverständnis-Aufgabenset bzw. das Experimentierstrategien-Aufgabenset richtig gelöst haben, als eine Funktion der Evidenzevaluationsfähigkeiten

	Klassifikation Verständnis von Dateninterpretation		χ^2 -Test
	Fehlvorstellung (n = 166)	Verständnis (n = 119)	
Theorieverständnis	15.1	34.5	$\chi^2(1) = 14.65, p < .001$
Verständnis von Experimentierstrategien	41.0	79.8	$\chi^2(1) = 42.77, p < .001$

Die Ergebnisse zeigten, dass Kinder mit einem höher ausgeprägten Verständnis der Evidenzevaluation signifikant besser in Aufgaben zum Theorieverständnis respektive Experimentierstrategien abschnitten als Kinder mit einem geringen Verständnis der Interpretation von Daten.

4.3 Abgrenzung und Zusammenhang des wissenschaftlichen Denkens zu ausgewählten kognitiven Fähigkeiten

4.3.1 Leseverständnis

Rasch-Skalierung

Insgesamt hatten 271 Kinder den Fragebogen zum Leseverständnis bearbeitet (Klasse 2: $n = 70$; Klasse 3: $n = 56$; Klasse 4: $n = 145$). Für die Analysen wurden die 20 Items des Tests zum Textverständnis des ELFE-Lesetests dichotom kodiert und in einem eindimensionalen Rasch-Modell mithilfe des Computerprogramms ConQuest skaliert.

Da es sich bei diesem Fragebogen um einen standardisierten Test handelte und für die Gesamtskala zufrieden stellende statistische Kennwerte erzielt werden konnten, wurden trotz zum Teil unzureichender statistischer Itemkennwerte keine Items aus den Analysen ausgeschlossen. Eine detaillierte Zusammenfassung der statistischen Kennwerte des eindimensionalen Rasch-Modells des ELFE-Textverständnistests findet sich in Tabelle B.1 im Anhang B.

Die Reliabilitätsmaße erreichten zufrieden stellende Werte: EAP/PV-Reliabilität = 0.90, Cronbachs Alpha = 0.90. Die latente Verteilung verfügte mit einem Wert von 4.41 über eine hohe Varianz. Die mittlere Itemschwierigkeit wurde auf 0 logits ($SD = 1.85$) restringiert und stellte den Nullpunkt der Skala dar. Die Itemschwierigkeiten reichten von -3.68 bis 3.28 logits. Die Personenfähigkeiten der 271 Grundschüler schwankten von -5.22 bis 5.07 logits. Die mittlere Personenfähigkeit lag bei 1.55 logits ($SD = 2.13$). Insgesamt konnte die mittlere Lesekompetenz der Stichprobe der Grundschulkindern im Vergleich zur durchschnittlichen Itemschwierigkeit als relativ hoch eingestuft werden.

Lesekompetenzwerte

Die Berechnung der Personenfähigkeiten in der Lesekompetenz nach dem Weighted-Likelihood-Schätzer (WLE) erfolgte im Rahmen der eindimensionalen Rasch-Analyse für die erhobenen Daten der gesamten Stichprobe. Um das Leseverständnis der Grundschulkindern hinsichtlich Geschlechterdifferenzen und Altersunterschiede zu untersuchen, wurden die Kompetenzwerte der Mädchen und Jungen sowie der Klassenstufen verglichen. Die Personenparameter sind in Tabelle 4.23 zusammengefasst.

Tabelle 4.23

Personenfähigkeiten (WLE) im Leseverständnis aufgeteilt nach Geschlecht für die Gesamtstichprobe und die einzelnen Klassenstufen

	Jahrgangsstufe							
	2		3		4		gesamt	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Jungen	-0.67	1.52	1.51	1.40	2.15	1.99	1.36	2.08
Mädchen	0.04	1.67	1.85	2.18	2.62	1.88	1.75	2.17
gesamt	-0.26	1.64	1.65	1.76	2.39	1.94	1.55	2.13

Anmerkung: M = Mittelwert, SD = Standardabweichung.

Eine univariate zweifaktorielle Varianzanalyse mit Geschlecht (Jungen, Mädchen) und Klassenstufe (Klasse 2, Klasse 3, Klasse 4) als unabhängige Variablen ergab einen signifikanten Haupteffekt Klassenstufe ($F(2,265) = 51.28, p < .001$) und einen signifikanten Haupteffekt Geschlecht ($F(1,265) = 4.37, p < .05$) auf die Leistung im Test zum Leseverständnis. Die Interaktion zwischen den Faktoren Geschlecht und Klassenstufe ($F(2,265) = 0.17, p > .05, ns$) wurde nicht signifikant.

In der Gesamtstichprobe betrug die mittlere Lesekompetenz der Grundschüler 1.36 logits ($SD = 2.08$) und der Grundschülerinnen 1.75 logits ($SD = 2.17$). Auch in den einzelnen Jahrgangsstufen erzielten die Mädchen durchschnittlich bessere Leistungen als die Jungen (vgl. Tabelle 4.23).

Die mittleren WLE-Personenfähigkeiten im Leseverständnis stiegen von der zweiten Klassenstufe mit -0.26 logits ($SD = 1.64$) zur dritten Klassenstufe mit einer durchschnittlichen Kompetenz von 1.65 logits ($SD = 1.76$) bis zur vierten Jahrgangsstufe, die im Durchschnitt eine Lesefähigkeit von 2.39 logits ($SD = 1.94$) erreichte, signifikant an. Zusätzlich wurden a-priori-Kontraste berechnet, um Gruppenunterschiede zwischen die Klassenstufen, insbesondere zwischen Klassenstufe 2 und 3 sowie zwischen Klassenstufe 3 und 4 zu ermitteln. Der Vergleich der Personenfähigkeiten der zweiten und dritten Klassenstufe zeigte einen signifikanten Unterschied ($t(114) = -6.24, p < .001$). Auch zwischen Jahrgangsstufe 3 und 4 konnte ein signifikanter Kompetenzunterschied festgestellt werden ($t(110) = -2.63, p < .01$). Der Vergleich der Leseleistungen des frühen (Klassenstufe 2 und 3) und späten Grundschulalters (Klassenstufe 4) ergab ebenfalls einen signifikanten Unterschied ($t(257) = 7.66, p < .001$).

Beziehung zwischen Leseverständnis und wissenschaftlichem Denken

Um die Kompetenz wissenschaftliches Denken von der Lesekompetenz der Grundschüler abzugrenzen, wurden das eindimensionale Modell und das zweidimensionale Modell mit den Dimensionen wissenschaftliches Denken und Leseverständnis miteinander verglichen. Die Modellselektionsmaße BIC und CAIC der zweidimensionalen Modelle der Gesamtstichprobe und der Klassenstufen fielen jeweils geringer aus (vgl. Tabelle 4.24). Daher konnte das zweidimensionale Modell dem eindimensionalen Modell als überlegen bezeichnet werden.

Tabelle 4.24

*Modellvergleich des eindimensionalen und zweidimensionalen Rasch-Modells
(Dimension 1 = Leseverständnis, Dimension 2 = wissenschaftliches Denken)*

Modell	Stufe	Deviance	#Par	BIC	CAIC
eindimensional	2/3	5477		5677	5718
	4	5672	41	5879	5920
	gesamt	11361		11592	11633
zweidimensional	2/3	5240		5449	5492
	4	5490	43	5707	5750
	gesamt	10951		11194	11237

Anmerkungen: Deviance = $-2 \cdot \text{LogLikelihood}$, #Par = Anzahl der geschätzten Parameter, BIC = Bayes Information Criterion, CAIC = Consistent Akaike's Information Criterion.

Wie die Modellprüfung zeigte, stellten Leseverständnis und wissenschaftliches Denken zwei separate Konstrukte dar. Wissenschaftliches Denken und Leseverständnis waren dabei moderat bis stark korreliert, sowohl in der Gesamtstichprobe als auch in den einzelnen Klassenstufen (vgl. Tabelle 4.25).

Tabelle 4.25

Latente Korrelationen zwischen wissenschaftlichem Denken und Leseverständnis

	1.	2.
Gesamtstichprobe (N = 271)		
1. Leseverständnis	1	
2. wissenschaftliches Denken	0.66	1
Jahrgangsstufe 2-3 (n = 126)		
1. Leseverständnis	1	
2. wissenschaftliches Denken	0.52	1
Jahrgangsstufe 4 (n = 145)		
1. Leseverständnis	1	
2. wissenschaftliches Denken	0.57	1

Anmerkung: Latente Korrelationen basierend auf den zweidimensionalen Rasch-Skalierungen stehen unterhalb der Diagonale.

Wird wissenschaftliches Denken differenziert in Wissenschaftsverständnis und Methodenkompetenz, zeigten sich hohe latente Interkorrelationen der Komponenten des wissen-

schaftlichen Denkens und jeweils geringere latente Korrelationen von moderater bis großer Stärke zum Leseverständnis (vgl. Tabelle 4.26).

Tabelle 4.26

Latente Korrelationen zwischen Wissenschaftsverständnis bzw. Methodenkompetenz und Leseverständnis

	1.	2.	3.
Gesamtstichprobe (N = 271)			
1. Leseverständnis	1		
2. Wissenschaftsverständnis	0.67	1	
3. Methodenkompetenz	0.62	0.86	1
Jahrgangsstufe 2-3 (n = 126)			
1. Leseverständnis	1		
2. Wissenschaftsverständnis	0.54	1	
3. Methodenkompetenz	0.49	0.96	1
Jahrgangsstufe 4 (n = 145)			
1. Leseverständnis	1		
2. Wissenschaftsverständnis	0.54	1	
3. Methodenkompetenz	0.54	0.74	1

Anmerkung: Latente Korrelationen der Komponenten des wissenschaftlichen Denkens und des Leseverständnisses basierend auf der dreidimensionalen Rasch-Skalierungen stehen unterhalb der Diagonale.

4.3.2 Intelligenz

Rasch-Skalierung

Von 272 Kindern lagen bearbeitete Intelligenztestbögen vor (Klasse 2: $n = 71$; Klasse 3: $n = 55$; Klasse 4: $n = 146$). Für die Analysen wurden die 26 Items der beiden Subtests des CFT-20R dichotom kodiert und in einem eindimensionalen Rasch-Modell mithilfe des Computerprogramms ConQuest skaliert.

Da die verwendeten Subtests Teile einer standardisierten Testbatterie waren und für die Gesamtskala aus beiden Subtests zufrieden stellende statistische Kennwerte erzielt werden konnten, wurden trotz zum Teil unzureichender statistischer Itemkennwerte einzelner Items keine Aufgaben aus den Analysen ausgeschlossen. Eine detaillierte Beschreibung der statistischen Kennwerte des eindimensionalen Rasch-Modells der ausgewählten Skala des CFT-20R-Intelligenztest steht in Tabelle B.2 im Anhang B.

Die Reliabilitätsmaße erreichten befriedigende Werte: EAP/PV-Reliabilität = 0.62, Cronbachs Alpha = 0.61. Die latente Verteilung verfügte mit einem Wert von 0.42 über eine relativ geringe Varianz. Die mittlere Itemschwierigkeit wurde auf 0 logits ($SD = 1.64$) restringiert und stellte den Nullpunkt der Skala dar. Die Itemschwierigkeiten reichten von -3.09 bis 4.24 logits. Die Personenfähigkeiten der 272 Grundschüler schwankten von -3.89 bis 1.93 logits. Die mittlere Intelligenz der Stichprobe lag bei 0.28 logits ($SD = 0.83$), so dass der Test für die Stichprobe der Grundschüler als angemessen bewertet werden kann.

Intelligenzwerte

Die WLE-Personenfähigkeiten nach dem Weighted-Likelihood-Schätzer im Intelligenztest wurden im Rahmen der eindimensionalen Rasch-Analyse für die erhobenen Intelligenzdaten der gesamten Stichprobe ermittelt. Um die Intelligenz der Grundschulkinder hinsichtlich Geschlechterdifferenzen und Altersunterschiede zu untersuchen, wurden die Intelligenzwerte der Mädchen und Jungen sowie der Klassenstufen verglichen. Die Personenfähigkeiten sind in Tabelle 4.27 zusammengefasst.

Tabelle 4.27

Personenfähigkeiten (WLE) im Intelligenztest aufgeteilt nach Geschlecht für die Gesamtstichprobe und die einzelnen Klassenstufen

	Jahrgangsstufe							
	2		3		4		gesamt	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Jungen	-0.09	0.76	0.32	0.54	0.47	0.82	0.31	0.78
Mädchen	-0.30	0.95	0.37	0.69	0.52	0.75	0.25	0.88
gesamt	-0.21	0.88	0.34	0.60	0.50	0.78	0.28	0.83

Anmerkung: *M* = Mittelwert, *SD* = Standardabweichung.

Es wurde eine univariate zweifaktorielle Varianzanalyse mit Geschlechtszugehörigkeit (Jungen, Mädchen) und Jahrgangsstufe (Klasse 2, Klasse 3, Klasse 4) als unabhängige Variablen gerechnet. Die Ergebnisse zeigten einen signifikanten Haupteffekt für Klassenstufe ($F(2,266) = 18.76, p < .001$). Weder der Haupteffekt Geschlecht ($F(1,266) = 0.14, p > .05, ns$) noch die Interaktion zwischen den Faktoren Geschlecht und Klassenstufe ($F(2,266) = 0.65, p > .05, ns$) wurden signifikant.

Betrachtet man die Mittelwerte erzielten Mädchen und Jungen sowohl in der Gesamtstichprobe als auch in den einzelnen Klassenstufen vergleichbare Werte (vgl. Tabelle

4.27). Hinter dem Haupteffekt Klassenstufe verbirgt sich ein Anstieg der Intelligenzwerte über die Klassenstufen hinweg: Die Intelligenzwerte stiegen von der zweiten Klassenstufe mit -0.21 logits ($SD = 0.88$), zur dritten Klasse mit einer durchschnittlichen Intelligenz von 0.34 logits ($SD = 0.60$) bis zur vierten Jahrgangsstufe an, die im Durchschnitt Intelligenzwerte von 0.50 logits ($SD = 0.78$) erzielte. Geplante Kontraste ergaben einen signifikanten Unterschied zwischen den Zweit- und Drittklässlern ($t(269) = -3.97, p < .001$), aber keinen signifikanten Effekt zwischen Klassenstufe 3 und 4 ($t(269) = -1.26, p > .05, ns$). Werden die Intelligenzleistungen des frühen (Klassenstufe 2 und 3) und späten Grundschulalters (Klassenstufe 4) verglichen, zeigte sich ein signifikanter Unterschied ($t(269) = 4.56, p < .001$).

Beziehung zwischen Intelligenz und wissenschaftlichem Denken

Zur Differenzierung zwischen der Kompetenz wissenschaftliches Denken von dem Intelligenzkonstrukt wurden das eindimensionale Modell und das zweidimensionale Modell mit den latenten Dimensionen wissenschaftliches Denken und Intelligenz gegenüber gestellt (vgl. Tabelle 4.28). Entsprechend der Informationskriterien BIC und CAIC beschrieb sowohl in der Gesamtstichprobe als auch in der Stichprobe der Zweit- und Drittklässler das zweidimensionale Modell die Daten besser. In der Jahrgangsstufe 4 konnte das eindimensionale dem zweidimensionalen Modell als überlegen angesehen werden. Die Modellentscheidungen fielen sehr eng aus.

Tabelle 4.28

Modellvergleich des eindimensionalen und zweidimensionalen Rasch-Modells (Dimension 1 = Intelligenz, Dimension 2 = wissenschaftliches Denken)

Modell	Stufe	Deviance	#Par	BIC	CAIC
eindimensional	2/3	6540		6549	6816
	4	7363	47	7600	7647
	gesamt	14096		14362	14409
zweidimensional	2/3	6524		6532	6811
	4	7382	49	7629	7678
	gesamt	14081		14358	14407

Anmerkungen: Deviance = $-2 \cdot \text{LogLikelihood}$, #Par = Anzahl der geschätzten Parameter, BIC = Bayes Information Criterion, CAIC = Consistent Akaike's Information Criterion.

Allerdings war aufgrund der moderaten bis hohen latenten Korrelationen zwischen wissenschaftlichem Denken und Intelligenz jeweils das zweidimensionale Modell, in dem wis-

senschaftliches Denken und Intelligenz als separate Konstrukte beschrieben wurden, als angemessener zu bezeichnen (vgl. Tabelle 4.29).

Tabelle 4.29

Latente Korrelation zwischen wissenschaftlichem Denken und Intelligenz

	1.	2.
Gesamtstichprobe (N = 272)		
1. Intelligenz	1	
2. wissenschaftliches Denken	0.59	1
Jahrgangsstufe 2-3 (n = 126)		
1. Intelligenz	1	
2. wissenschaftliches Denken	0.33	1
Jahrgangsstufe 4 (n = 146)		
1. Intelligenz	1	
2. wissenschaftliches Denken	0.62	1

Anmerkung: Latente Korrelationen basierend auf den zweidimensionalen Rasch-Skalierungen stehen unterhalb der Diagonale.

Die differenzierte Betrachtung der Zusammenhänge zwischen Intelligenz und den Komponenten des wissenschaftlichen Denkens ließ moderate bis hohe latente Korrelationen sowohl zum Wissenschaftsverständnis als auch zur Methodenkompetenz erkennen (vgl. Tabelle 4.30). Die latenten Interkorrelationen der Komponenten erreichten ein höheres Ausmaß als die Korrelationen zwischen Intelligenz und Wissenschaftsverständnis respektive Methodenkompetenz.

Tabelle 4.30

Latente Korrelationen zwischen Wissenschaftsverständnis bzw. Methodenkompetenz und Intelligenz

	1.	2.	3.
Gesamtstichprobe (N = 272)			
1. Intelligenz	1		
2. Wissenschaftsverständnis	0.65	1	
3. Methodenkompetenz	0.57	0.90	1
Jahrgangsstufe 2-3 (n = 126)			
1. Intelligenz	1		
2. Wissenschaftsverständnis	0.50	1	
3. Methodenkompetenz	0.27	0.89	1
Jahrgangsstufe 4 (n = 146)			
1. Intelligenz	1		
2. Wissenschaftsverständnis	0.67	1	
3. Methodenkompetenz	0.64	0.78	1

Anmerkung: Latente Korrelationen der Komponenten des wissenschaftlichen Denkens und der Intelligenz basierend auf der dreidimensionalen Rasch-Skalierungen stehen unterhalb der Diagonale.

4.3.3 Problemlösen

Rasch-Skalierung

In die Auswertungen gingen die Daten von 270 Kindern ein (Klasse 2: $n = 71$; Klasse 3: $n = 54$; Klasse 4: $n = 145$). Für die Analysen wurden die acht Items der Papierversion des Turm von Londons dichotom kodiert und in einem eindimensionalen Rasch-Modell mithilfe des Computerprogramms ConQuest skaliert. Eine detaillierte Beschreibung der statistischen Kennwerte des eindimensionalen Rasch-Modells des Tests zu Erfassung der Problemlösekompetenz findet sich in Tabelle B.3 im Anhang B.

Die Varianz der latenten Variable lag bei 4.60. Die EAP/PV-Reliabilität erreichte einen zufrieden stellenden Wert von 0.79. Für das Maß der internen Konsistenz wurde mittels Cronbachs Alpha der Wert 0.80 bestimmt. Die Fit-Statistiken der Itemparameter waren zufrieden stellend. Es konnten keine signifikanten T-Werte ($-2.0 < T\text{-Wert} < 2.0$) festgestellt werden. Der Infit (Weighted Mean Square, MNSQ) lag für alle Items mit Werten zwischen 0.85 und 1.15 innerhalb des von den Richtlinien vorgegebenen Bereichs (Adams, 2002). Der Mittelwert der MNSQ-Werte aller Items lag bei 0.96 ($SD = 0.09$). Die Fit-

Statistiken zeigten, dass die Daten gut durch das Rasch-Modell modelliert werden konnten und sich die Items modellkonform verhielten. Die punktbiserialen Korrelationen bzw. die klassischen Trennschärfen lagen zwischen 0.49 und 0.74, mit einem Mittelwert von 0.65 ($SD = 0.09$). Die Personenfähigkeiten der 270 Grundschul Kinder schwankten von -4.15 bis 4.49 logits. Die mittlere Problemlösekompetenz der Stichprobe lag bei -1.51 logits ($SD = 2.14$), so dass der Test für die Stichprobe der Grundschüler als relativ schwer beurteilt werden muss.

Problemlösekompetenzwerte

Die Problemlösefähigkeiten (WLE, Weighted-Likelihood-Schätzer) der Kinder wurden mittels der eindimensionalen Rasch-Analyse für die erhobenen Daten der Gesamtstichprobe berechnet und hinsichtlich Geschlechterdifferenzen und Altersunterschiede betrachtet. Die Personenparameter im Problemlösetest sind für Mädchen und Jungen sowie für die einzelnen Klassenstufen und die Gesamtstichprobe in Tabelle 4.31 zusammengefasst.

Tabelle 4.31

Personenfähigkeiten (WLE) im Problemlösen aufgeteilt nach Geschlecht für die Gesamtstichprobe und die einzelnen Klassenstufen

	Jahrgangsstufe							
	2		3		4		gesamt	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Jungen	-2.47	1.32	-1.46	2.09	-0.96	2.42	-1.43	2.21
Mädchen	-2.80	1.32	-1.60	1.67	-0.93	2.26	-1.59	2.09
gesamt	-2.66	1.32	-1.52	1.92	-0.95	2.33	-1.51	2.14

Anmerkung: M = Mittelwert, SD = Standardabweichung.

Die Auswertung erfolgte auf der Grundlage einer univariaten zweifaktoriellen Varianzanalyse mit Geschlecht (Jungen, Mädchen) und Klassenzugehörigkeit (Klasse 2, Klasse 3, Klasse 4) als unabhängige Variablen. Wie die Ergebnisse zeigten, fand sich bei dem Faktor Klassenzugehörigkeit ein signifikanter Haupteffekt ($F(2,264) = 16.05, p < .001$). Weder der Haupteffekt Geschlecht ($F(1,264) = 0.28, p > .05, ns$) noch die Interaktion zwischen den Faktoren Geschlecht und Klassenstufe ($F(2,264) = 0.19, p > .05, ns$) wurden signifikant.

Mädchen und Jungen schnitten im Problemlösetest sowohl in der Gesamtstichprobe als auch in den einzelnen Klassenstufen gleich gut ab (vgl. Tabelle 4.31). Mit steigender Klas-

senstufe nahm die Problemlösefähigkeit von der zweiten Klassenstufe mit einer mittleren Problemlösekompetenz von -2.66 logits ($SD = 1.32$), zur dritten Klasse mit einer durchschnittlichen Fähigkeit von -1.52 logits ($SD = 1.92$) bis zur vierten Jahrgangsstufe, die im Durchschnitt Kompetenzwerte von -0.95 logits ($SD = 2.33$) erreichte, zu. A-priori Kontraste zeigten einen signifikanten Unterschied zwischen den Klassenstufen 2 und 3 ($t(89) = -3.76, p < .001$), aber nur einen tendenziellen Unterschied zwischen Klasse 3 und 4 ($t(115) = -1.75, p = .083$). Werden Klassenstufe 2 und 3 als frühes Grundschulalter zusammengefasst und deren Leistungen mit denen der Kinder aus Jahrgangsstufe 4 verglichen, konnte ein signifikanter Unterschied ($t(233) = 4.63, p < .001$) festgestellt werden.

Beziehung zwischen Problemlösen und wissenschaftlichem Denken

Um die Beziehung zwischen wissenschaftlichem Denken und Problemlösen voneinander abgrenzend zu betrachten, wurden das eindimensionale Modell und das zweidimensionale Modell mit den Dimensionen wissenschaftliches Denken und Problemlösen miteinander verglichen. Die Modellselektionsmaße BIC und CAIC der zweidimensionalen Modelle fielen sowohl in der Gesamtstichprobe als auch in den Klassenstufen geringer aus (vgl. Tabelle 4.32), daher konnte das zweidimensionale Modell die Daten besser beschreiben als das eindimensionale Modell.

Tabelle 4.32

Modellvergleich des eindimensionalen und zweidimensionalen Rasch-Modells (Dimension 1 = Problemlösen, Dimension 2 = wissenschaftliches Denken)

Modell	Stufe	Deviance	#Par	BIC	CAIC
eindimensional	2/3	3929		4070	4099
	4	4539	29	4685	4714
	gesamt	8593		8757	8786
zweidimensional	2/3	3915		4066	4097
	4	4440	31	4596	4627
	gesamt	8485		8660	8691

Anmerkungen: Deviance = $-2 \cdot \text{LogLikelihood}$, #Par = Anzahl der geschätzten Parameter, BIC = Bayes Information Criterion, CAIC = Consistent Akaike's Information Criterion.

Die latenten Korrelationen zwischen der Kompetenz wissenschaftliches Denken und der Problemlösekompetenz fielen moderat bis hoch aus, auch unter Kontrolle des Einflusses

von Leseverständnis und Intelligenz. Die Höhe der latenten Korrelationen deutete auf zwar korrelierte, aber getrennte Konstrukte hin (vgl. Tabelle 4.33).

Tabelle 4.33

Latente Korrelationen zwischen wissenschaftlichem Denken und Problemlösen

	1.	2.
Gesamtstichprobe (N = 270)		
1. Problemlösen	1	0.58
2. wissenschaftliches Denken	0.66	1
Jahrgangsstufe 2-3 (n = 125)		
1. Problemlösen	1	0.68
2. wissenschaftliches Denken	0.59	1
Jahrgangsstufe 4 (n = 145)		
1. Problemlösen	1	0.48
2. wissenschaftliches Denken	0.62	1

Anmerkung: Latente Korrelationen basierend auf zweidimensionalen Rasch-Skalierungen stehen unterhalb der Diagonale. Bedingte latente Korrelationen unter Kontrolle von Leseverständnis und Intelligenz basierend auf zweidimensionalen latenten Regressionen stehen oberhalb der Diagonale.

Das Niveau der latenten Interkorrelationen der Komponenten des wissenschaftlichen Denkens war relativ gesehen höher als die Korrelationen von Wissenschaftsverständnis respektive Methodenkompetenz zur Problemlösefähigkeit, auch unter Kontrolle von Leseverständnis und Intelligenz (vgl. Tabelle 4.34). Nur im frühen Grundschulalter (Klassenstufe 2-3) zeigte sich unter Kontrolle von Leseverständnis eine hohe bedingte latente Korrelation von $r = 0.84$ zwischen Methodenkompetenzen und planerisch-problemlösendem Denken.

Tabelle 4.34

Latente Korrelationen zwischen Wissenschaftsverständnis bzw. Methodenkompetenz und Problemlösen

	1.	2.	3.
Gesamtstichprobe (N = 270)			
1. Problemlösen	1	0.61	0.66
2. Wissenschaftsverständnis	0.67	1	0.88
3. Methodenkompetenz	0.61	0.88	1
Jahrgangsstufe 2-3 (n = 125)			
1. Problemlösen	1	0.43	0.84
2. Wissenschaftsverständnis	0.47	1	0.66
3. Methodenkompetenz	0.58	0.86	1
Jahrgangsstufe 4 (n = 145)			
1. Problemlösen	1	0.57	0.43
2. Wissenschaftsverständnis	0.80	1	0.91
3. Methodenkompetenz	0.53	0.80	1

Anmerkung: Latente Korrelationen basierend auf dreidimensionalen Rasch-Skalierungen stehen unterhalb der Diagonale. Bedingte latente Korrelationen der Komponenten des wissenschaftlichen Denkens unter Kontrolle von Leseverständnis und Intelligenz basierend auf dreidimensionalen latenten Regressionen stehen oberhalb der Diagonale.

4.3.4 Inhibition

Rasch-Skalierung

Insgesamt bearbeiteten 276 Kinder die Papierversion des Stroop-Tests zur Erfassung der Inhibitionsleistung (Klasse 2: $n = 71$; Klasse 3: $n = 53$; Klasse 4: $n = 152$). Für die Analysen wurden die ersten 40 Items des 72 Items umfassenden Aufgabenbogens zum Interferenzversuch ausgewählt, da diese Aufgaben über genügend Varianz verfügten. Die dichotom kodierten Items wurden in einem eindimensionalen Rasch-Modell mithilfe des Computerprogramms ConQuest skaliert.

Aufgrund zufrieden stellender statistischer Kennwerte der Skala der 40 Inhibitionitems des Speedtests wurden trotz zum Teil abweichender statistischer Itemkennwerte keine Items aus den Analysen ausgeschlossen. Eine detaillierte Zusammenfassung der statistischen Kennwerte des eindimensionalen Rasch-Modells des Stroop-Inhibitionstests findet sich in Tabelle B.4 im Anhang B.

Die Skalierung ergab sehr hohe Reliabilitätswerte: EAP/PV-Reliabilität $> 1^7$, Cronbachs Alpha = 0.95. Die latente Verteilung verfügte über eine hohe Varianz von 6.21. Die Fit-Statistiken erreichten einen Mittelwert von 0.79 ($SD = 0.23$) und zeigten, dass die Daten durch das Rasch-Modell modelliert werden konnten. Die punktbiserialen Korrelationen bzw. die klassischen Trennschärfen lagen zwischen 0.24 und 0.77, mit einem Mittelwert von 0.60 ($SD = 0.14$). Die mittlere Itemschwierigkeit wurde auf 0 logits ($SD = 2.81$) restringiert und stellte den Nullpunkt der Skala dar. Die Itemschwierigkeiten reichten von -5.09 bis 5.92 logits. Die Personenfähigkeiten der 282 Grundschüler schwankten von -5.24 bis 7.84 logits. Die mittlere Personfähigkeit lag bei 2.66 logits ($SD = 2.86$). Insgesamt konnte die mittlere Inhibitionsleistung der Stichprobe im Vergleich zur durchschnittlichen Itemschwierigkeit als relativ hoch bewertet werden.

Inhibitionswerte

Die Inhibitionsfähigkeiten (WLE, Weighted-Likelihood-Schätzer) der Grundschul Kinder wurden durch die eindimensionale Rasch-Analyse der erhobenen Daten der Gesamtstichprobe berechnet. Im Folgenden werden die Leistungen im Inhibitionstest im Hinblick auf Geschlechter- und Altersunterschiede betrachtet. Die Personenfähigkeiten in der Inhibition aufgabenirrelevanter Reize sind in Tabelle 4.35 getrennt nach Geschlecht und Klassenstufe dargestellt.

Tabelle 4.35

Personenfähigkeiten (WLE) im Inhibitionstest aufgeteilt nach Geschlecht für die Gesamtstichprobe und die einzelnen Klassenstufen

	Jahrgangsstufe							
	2		3		4		gesamt	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Jungen	0.40	2.13	1.37	2.29	3.25	2.83	2.17	2.83
Mädchen	0.89	2.41	2.82	1.80	4.42	2.48	3.14	2.82
gesamt	0.68	2.29	1.97	2.21	3.85	2.70	2.66	2.86

Anmerkung: *M* = Mittelwert, *SD* = Standardabweichung.

Eine univariate zweifaktorielle Varianzanalyse mit Geschlecht (Jungen, Mädchen) und Klassenstufe (Klasse 2, Klasse 3, Klasse 4) als unabhängige Variablen ergab signifikante Effekte des Faktors Klassenstufe ($F(2,267) = 41.39, p < .001$) und des Faktors Geschlecht

⁷ Aufgrund der Berechnungsweise der EAP/PV-Reliabilität (vgl. Abschnitt 3.4.1, S. 75) sind Werte über 1 möglich.

($F(1,267) = 9.80, p < .01$) auf die Leistung im Test zu inhibitorischen Fähigkeiten. Die Interaktion zwischen den Faktoren Geschlecht und Klassenstufe ($F(2,267) = 0.65, p > .05, ns$) wurde nicht signifikant.

In der Gesamtstichprobe betrug die mittlere Inhibitionsleistung der Grundschüler 2.17 logits ($SD = 2.83$) und der Grundschülerinnen 3.14 logits ($SD = 2.82$). Auch in den einzelnen Klassenstufen erzielten die Mädchen durchschnittlich bessere Leistungen als die Jungen (vgl. Tabelle 4.35). Die mittleren WLE-Personenfähigkeiten im Inhibitionstest stiegen von der zweiten Klassenstufe mit 0.68 logits ($SD = 2.29$) zur dritten Klassenstufe mit einer durchschnittlichen Kompetenz von 1.97 logits ($SD = 2.21$) bis zur vierten Jahrgangsstufe, die im Mittel eine Inhibitionsfähigkeit von 3.85 logits ($SD = 2.70$) erreichte, signifikant an. Zusätzlich gerechnete a-priori-Kontraste zeigten, dass sich Zweit- und Drittklässler signifikant in ihrer Inhibitionsleistung unterschieden ($t(114) = -3.17, p < .01$). Der Vergleich der Jahrgangsstufe 3 und 4 ergab ebenfalls einen signifikanten Kompetenzunterschied ($t(110) = -5.08, p < .001$). Werden die Inhibitionsleistungen des frühen (Klassenstufe 2 und 3) und späten Grundschulalters (Klassenstufe 4) gegenübergestellt, konnte ebenfalls ein signifikanter Unterschied ($t(264) = 8.50, p < .001$) festgestellt werden.

Beziehung zwischen Inhibitionsleistung und wissenschaftlichem Denken

Um die Relation zwischen wissenschaftlichem Denken und der Fähigkeit zur Inhibition von aufgabenirrelevanten Informationen zu untersuchen, wurden das eindimensionale Modell und das zweidimensionale Modell mit den Dimensionen wissenschaftliches Denken und Inhibition miteinander verglichen. Die Modellselektionsmaße BIC und CAIC der zweidimensionalen Modelle der Gesamtstichprobe sowie der Klassenstufen fielen jeweils geringer aus (vgl. Tabelle 4.36), daher war das zweidimensionale Modell dem eindimensionalen Modell als überlegen anzusehen.

Tabelle 4.36

*Modellvergleich des eindimensionalen und zweidimensionalen Rasch-Modells
(Dimension 1 = Inhibition, Dimension 2 = wissenschaftliches Denken)*

Modell	Stufe	Deviance	#Par	BIC	CAIC
eindimensional	2/3	6339		6636	6697
	4	6878	61	7186	7247
	gesamt	13582		13927	13988
zweidimensional	2/3	5502		5808	5871
	4	6096	63	6414	6477
	gesamt	12030		12386	12449

Anmerkungen: Deviance = $-2 \cdot \text{LogLikelihood}$, #Par = Anzahl der geschätzten Parameter, BIC = Bayes Information Criterion, CAIC = Consistent Akaike's Information Criterion.

Es konnte kein substantieller Zusammenhang zwischen der Inhibitionsleistung und der Kompetenz im wissenschaftlichen Denken festgestellt werden. Es fanden sich keine bis kleine Effekte, insbesondere unter Kontrolle von Leseverständnis und Intelligenz (vgl. Tabelle 4.37).

Tabelle 4.37

Latente Korrelationen zwischen wissenschaftlichem Denken und Inhibition

	1.	2.
Gesamtstichprobe (N = 276)		
1. Inhibition	1	0.14
2. wissenschaftliches Denken	0.35	1
Jahrgangsstufe 2-3 (n = 124)		
1. Inhibition	1	0.09
2. wissenschaftliches Denken	0.18	1
Jahrgangsstufe 4 (n = 152)		
1. Inhibition	1	-0.07
2. wissenschaftliches Denken	0.05	1

Anmerkung: Latente Korrelationen basierend auf zweidimensionalen Rasch-Skalierungen stehen unterhalb der Diagonale. Bedingte latente Korrelationen unter Kontrolle von Leseverständnis und Intelligenz basierend auf zweidimensionalen latenten Regressionen stehen oberhalb der Diagonale.

Auch die Betrachtung der Zusammenhänge der Inhibitionsfähigkeit im Hinblick auf die Komponenten Wissenschaftsverständnis und Methodenkompetenz zeigte keine bis geringe Effekte (vgl. Tabelle 4.38).

Tabelle 4.38

Latente Korrelationen zwischen Wissenschaftsverständnis bzw. Methodenkompetenz und Inhibition

	1.	2.	3.
Gesamtstichprobe (N = 276)			
1. Inhibition	1	0.04	0.23
2. Wissenschaftsverständnis	0.30	1	0.94
3. Methodenkompetenz	0.35	0.87	1
Jahrgangsstufe 2-3 (n = 124)			
1. Inhibition	1	-0.12	0.20
2. Wissenschaftsverständnis	-0.09	1	0.78
3. Methodenkompetenz	0.23	0.93	1
Jahrgangsstufe 4 (n = 152)			
1. Inhibition	1	-0.16	0.02
2. Wissenschaftsverständnis	0.06	1	0.98
3. Methodenkompetenz	0.07	0.92	1

Anmerkung: Latente Korrelationen basierend auf dreidimensionalen Rasch-Skalierungen stehen unterhalb der Diagonale. Bedingte latente Korrelationen unter Kontrolle von Leseverständnis und Intelligenz basierend auf dreidimensionalen latenten Regressionen stehen oberhalb der Diagonale.

4.3.5 Räumliches Denken

Rasch-Skalierung

Die Daten von 271 Kindern stellten die Basis der Skalierung des Tests zur Erfassung des räumlichen Denkens dar (Klasse 2: $n = 64$; Klasse 3: $n = 58$; Klasse 4: $n = 149$). Item 12 musste aufgrund schlechter statistischer Kennwerte aus der Skalierung ausgeschlossen werden, so dass die elf verbleibenden Items in dichotomer Kodierung in die Berechnung des eindimensionalen Rasch-Modells mithilfe des Computerprogramms ConQuest eingingen. Eine detaillierte Beschreibung der statistischen Kennwerte der eindimensionalen Skala zum räumlichen Denken steht in Tabelle B.5 im Anhang B.

Es ergab sich eine Varianz der latenten Variable von 1.39. Die Reliabilitäten (EAP/PV-Reliabilität = 0.64, Cronbachs Alpha = 0.67) erreichten befriedigende Werte. Die Fit-Statistiken für die Itemparameter waren zufrieden stellend. Es konnten keine signifikanten T-Werte ($-2.0 < \text{T-Wert} < 2.0$) festgestellt werden. Mit Werten zwischen 0.89 und 1.09 lag der Infit (Weighted Mean Square, MNSQ) für alle Items innerhalb des von den Richtlinien

vorgegebenen Bereichs (Adams, 2002). Der mittlere Infit mit einem Wert von 1.00 ($SD = 0.07$) erreichte einen optimalen Wert. Die Ergebnisse zeigten, dass sich die Items modellkonform verhielten und die Daten gut durch das Rasch-Modell modelliert werden konnten. Die punktbiserialen Korrelationen bzw. die klassischen Trennschärfen reichten von 0.11 bis 0.67, mit einem Mittelwert von 0.46 ($SD = 0.18$). Die Personenfähigkeiten der 271 Grundschul Kinder schwankten von -2.03 bis 4.12 logits, mit einem Mittelwert von -0.88 logits ($SD = 1.42$), so dass der Test für die Stichprobe der Grundschüler als angemessen eingestuft werden konnte.

Kompetenzwerte im räumlichen Denken

Die WLE-Personenfähigkeiten (Weighted-Likelihood-Schätzer) im Test zum räumlichen Denken wurden im Rahmen der eindimensionalen Rasch-Analyse der erhobenen Daten der Gesamtstichprobe berechnet und im Hinblick auf Geschlechterunterschiede und Alterseffekte untersucht. In Tabelle 4.39 sind die Kompetenzwerte im räumlichen Denken für die einzelnen Klassenstufen und die Gesamtstichprobe nach Mädchen und Jungen getrennt dargestellt.

Tabelle 4.39

Personenfähigkeiten (WLE) im räumlichen Denken aufgeteilt nach Geschlecht für die Gesamtstichprobe und die einzelnen Klassenstufen

	Jahrgangsstufe							
	2		3		4		gesamt	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Jungen	0.29	1.17	1.20	1.59	1.53	1.45	1.19	1.51
Mädchen	0.15	0.98	0.28	0.93	0.89	1.40	0.59	1.27
gesamt	0.21	1.06	0.83	1.43	1.19	1.46	0.88	1.42

Anmerkung: M = Mittelwert, SD = Standardabweichung.

Zur Ermittlung der Gruppenunterschiede wurde eine univariate zweifaktorielle Varianzanalyse mit Geschlechtszugehörigkeit (Jungen, Mädchen) und Jahrgangsstufe (Klasse 2, Klasse 3, Klasse 4) als unabhängige Variablen gerechnet. Wie die Ergebnisse zeigten, fanden sich bei der Klassenstufe ($F(2,263) = 12.45, p < .001$) und dem Geschlecht ($F(1,263) = 9.86, p < .01$) signifikante Effekte. Ein Interaktionseffekt trat nicht auf ($F(2,263) = 1.34, p > .05, ns$).

Betrachtet man die Mittelwerte, zeigte sich ein deutlicher Vorsprung der Jungen gegenüber den Mädchen im räumlichen Denken. Dieser Unterschied fand sich sowohl in der Gesamtstichprobe als auch in den einzelnen Jahrgangsstufen (vgl. Tabelle 4.39). Im Hinblick auf die Klassenstufen verbarg sich hinter dem Haupteffekt ein deutlicher Anstieg der Fähigkeiten über die Klassenstufen hinweg: Mit steigender Klassenstufe nahm die Kompetenz von Klasse 2 mit 0.21 logits ($SD = 1.06$), zur dritten Klasse mit einer mittleren Leistung von 0.83 logits ($SD = 1.43$) bis zur vierten Jahrgangsstufe, die durchschnittlich 1.19 logits ($SD = 1.46$) erzielte, zu. Geplante Kontraste ergaben einen signifikanten Unterschied zwischen den Zweit- und Drittklässlern ($t(105) = -2.71, p < .01$), aber keinen signifikanten Effekt zwischen Klassenstufe 3 und 4 ($t(105) = -1.61, p > .05, ns$). Werden die räumlichen Fähigkeiten des frühen (Klassenstufe 2 und 3) und späten Grundschulalters (Klassenstufe 4) gegenüber gestellt, konnte ein signifikanter Unterschied ($t(248) = 4.04, p < .001$) festgestellt werden.

Beziehung zwischen räumlichem Denken und wissenschaftlichem Denken

Um die Beziehung zwischen wissenschaftlichem Denken und räumlichem Denken zu betrachten, wurde ein Modellvergleich zwischen dem eindimensionalen Modell und dem zweidimensionalen Modell mit den Dimensionen wissenschaftliches Denken und räumliche Fähigkeiten angestellt. Die Informationskriterien BIC und CAIC der eindimensionalen Modelle fielen jeweils geringer aus, allerdings nur minimal (vgl. Tabelle 4.40).

Tabelle 4.40

Modellvergleich des eindimensionalen und zweidimensionalen Rasch-Modells (Dimension 1 = räumliches Denken, Dimension 2 = wissenschaftliches Denken)

Modell	Stufe	Deviance	#Par	BIC	CAIC
eindimensional	2/3	4568		4724	4756
	4	4998	32	5160	5192
	gesamt	9720		9901	9933
zweidimensional	2/3	4567		4732	4766
	4	4994	34	5166	5200
	gesamt	9712		9905	9939

Anmerkungen: Deviance = $-2 \cdot \text{LogLikelihood}$, #Par = Anzahl der geschätzten Parameter, BIC = Bayes Information Criterion, CAIC = Consistent Akaike's Information Criterion.

Anhand der Modellprüfung war eine knappe Überlegenheit des eindimensionalen Modells deutlich geworden. Diese Überlegenheit konnte allerdings aufgrund der nur moderaten

Korrelationen der latenten Dimensionen wissenschaftliches Denken und räumliches Denken nicht aufrechterhalten werden (vgl. Tabelle 4.41).

Tabelle 4.41

Latente Korrelationen zwischen wissenschaftlichem Denken und räumlichem Denken

	1.	2.
Gesamtstichprobe (N = 271)		
1. Räumliches Denken	1	0.46
2. wissenschaftliches Denken	0.55	1
Jahrgangsstufe 2-3 (n = 122)		
1. Räumliches Denken	1	0.49
2. wissenschaftliches Denken	0.41	1
Jahrgangsstufe 4 (n = 149)		
1. Räumliches Denken	1	0.35
2. wissenschaftliches Denken	0.53	1

Anmerkung: Latente Korrelationen basierend auf zweidimensionalen Rasch-Skalierungen stehen unterhalb der Diagonale. Bedingte latente Korrelationen unter Kontrolle von Leseverständnis und Intelligenz basierend auf zweidimensionalen latenten Regressionen stehen oberhalb der Diagonale.

Auch die im Verhältnis höheren Interkorrelationen zwischen Wissenschaftsverständnis und Methodenkompetenz im Vergleich zur Höhe der Korrelationen zwischen den Komponenten des wissenschaftlichen Denkens und den räumlichen Fähigkeiten ließen das zweidimensionale Modell und damit zwei separate Konstrukte als angemessen erscheinen (vgl. Tabelle 4.42).

Tabelle 4.42

Latente Korrelationen zwischen Wissenschaftsverständnis bzw. Methodenkompetenz und räumlichem Denken

	1.	2.	3.
Gesamtstichprobe (N = 271)			
1. Räumliches Denken	1	0.57	0.48
2. Wissenschaftsverständnis	0.54	1	0.87
3. Methodenkompetenz	0.54	0.91	1
Jahrgangsstufe 2-3 (n = 122)			
1. Räumliches Denken	1	0.35	0.55
2. Wissenschaftsverständnis	0.20	1	0.67
3. Methodenkompetenz	0.48	0.94	1
Jahrgangsstufe 4 (n = 149)			
1. Räumliches Denken	1	0.64	0.33
2. Wissenschaftsverständnis	0.68	1	0.86
3. Methodenkompetenz	0.46	0.79	1

Anmerkung: Latente Korrelationen basierend auf dreidimensionalen Rasch-Skalierungen stehen unterhalb der Diagonale. Bedingte latente Korrelationen unter Kontrolle von Leseverständnis und Intelligenz basierend auf dreidimensionalen latenten Regressionen stehen oberhalb der Diagonale.

4.3.6 Formal-operatorische Fähigkeiten

Skalierung

Die Fähigkeiten im formal-operatorischen Denken wurden nur in Klassenstufe 4 erhoben. Insgesamt hatten 141 Viertklässler den Arlin-Test zur Erfassung der formal-operatorischen Fähigkeiten bearbeitet. Für die Analysen wurden die 16 Items dichotom kodiert und in einem eindimensionalen Rasch-Modell mithilfe des Computerprogramms ConQuest skaliert. Eine detaillierte Zusammenfassung der statistischen Kennwerte des eindimensionalen Rasch-Modells des Arlin-Tests findet sich in Tabelle B.6 im Anhang B.

Die latente Verteilung verfügte über eine sehr eingeschränkte Varianz von 0.10. Die Reliabilitätsmaße erreichten ebenfalls unzureichende Werte: EAP/PV-Reliabilität = 0.22, Cronbachs Alpha = 0.29. Die Trennschärfen der Items erreichten durchschnittlich nur einen Wert von 0.29 (Minimum = 0.10, Maximum = 0.39) und waren daher als ungenügend einzustufen. Hinsichtlich der Fit-Statistiken konnten keine signifikanten T-Werte ($-2.0 < T\text{-Wert} < 2.0$) festgestellt werden. Mit Werten zwischen 0.97 und 1.04 lag der Infit (Weighted Mean Square, MNSQ) ($M = 1.00$, $SD = 0.02$) für alle Items innerhalb des von

den Richtlinien vorgegebenen Bereichs (Adams, 2002). Die mittlere Itemschwierigkeit wurde auf 0 logits ($SD = 0.56$) restringiert und stellte den Nullpunkt der Skala dar. Die Itemschwierigkeiten reichten von -0.95 bis 1.11 logits. Die Personenfähigkeiten der 141 Viertklässler schwankten von -2.47 bis 1.44 logits. Die mittlere Personfähigkeit lag bei -0.60 logits ($SD = 0.61$). Aufgrund der durchschnittlich geringen Anzahl an erreichten Punkten und damit vorwiegenden Einstufung in das konkret-operatorische Niveau sowie der dadurch bedingten sehr eingeschränkten Varianz war der Test zur Erfassung des formal-operatorischen Denkens für die Stichprobe der Viertklässler als zu schwer zu beurteilen. Das formal-operatorische Denken konnte in Klassenstufe 4 nicht reliabel erfasst werden.

Kompetenzwerte im formal-operatorischen Denken

Die WLE-Personenfähigkeiten (Weighted-Likelihood-Schätzer) im Test zur Erfassung der formal-operatorischen Fähigkeiten wurden im Rahmen der eindimensionalen Rasch-Analyse der Daten der Viertklässler berechnet und im Hinblick auf Geschlechterunterschiede untersucht. In Tabelle 4.43 sind die Kompetenzwerte des formal-operatorischen Denkens für die Mädchen und Jungen der vierten Klassenstufe dargestellt.

Tabelle 4.43

Personenfähigkeiten (WLE) der Viertklässler im formal-operatorischen Denken

	Jahrgangsstufe 4	
	<i>M</i>	<i>SD</i>
Jungen (n = 67)	-0.62	0.56
Mädchen (n = 73)	-0.58	0.66
Gesamt (n = 140)	-0.60	0.61

Anmerkung: *M* = Mittelwert, *SD* = Standardabweichung.

Es wurde ein Vergleich zwischen den Mädchen und Jungen der vierten Jahrgangsstufe durchgeführt, indem die Mittelwerte der Kompetenzwerte im formal-operatorischen Test beider Geschlechter anhand des t-Tests für unabhängige Stichproben miteinander verglichen wurden. Das Ergebnis zeigte, dass sich Mädchen mit einer mittleren Fähigkeit von -0.58 logits ($SD = 0.66$) und Jungen mit einer durchschnittlichen Leistung von -0.62 logits ($SD = 0.56$) nicht signifikant in ihren formal-operatorischen Fähigkeiten unterschieden ($t(138) = -0.46, p > .05$).

Beziehung zwischen formal-operatorischem Denken und wissenschaftlichem Denken

Mittels des Vergleichs des eindimensionalen und des zweidimensionalen Modells mit den Dimensionen wissenschaftliches Denken und formal-operatorische Fähigkeiten wurde das eindimensionale Modell aufgrund des geringeren BIC und CAIC als das angemessenere Modell identifiziert (vgl. Tabelle 4.44).

Tabelle 4.44

Modellvergleich des eindimensionalen und zweidimensionalen Rasch-Modells (Dimension 1 = formal-operatorische Fähigkeiten, Dimension 2 = wissenschaftliches Denken) bezogen auf die Jahrgangsstufe 4

Modell	Deviance	#Par	BIC	CAIC
eindimensional	6498	37	6684	6721
zweidimensional	6725	39	6922	6961

Anmerkungen: Deviance = $-2 \cdot \text{LogLikelihood}$, #Par = Anzahl der geschätzten Parameter, BIC = Bayes Information Criterion, CAIC = Consistent Akaike's Information Criterion.

Allerdings sprach die moderate bis hohe Korrelation zwischen wissenschaftlichem Denken und formal-operatorischem Denken, die unter Kontrolle von Leseverständnis und Intelligenz verschwand, dafür, von zwei separaten Konstrukten auszugehen (vgl. Tabelle 4.45).

Tabelle 4.45

Latente Korrelation zwischen wissenschaftlichem Denken und formal operatorische Fähigkeiten

	1	2
Jahrgangsstufe 4 (n = 141)		
1. formal operatorische Fähigkeiten	1	-0.05
2. wissenschaftliches Denken	0.62	1

Anmerkung: Latente Korrelationen basierend auf zweidimensionalen Rasch-Skalierungen stehen unterhalb der Diagonale. Bedingte latente Korrelationen unter Kontrolle von Leseverständnis und Intelligenz basierend auf zweidimensionalen latenten Regressionen stehen oberhalb der Diagonale.

Die moderaten bis hohen latenten Korrelationen der Komponenten Wissenschaftsverständnis bzw. Methodenkompetenz zu formal-operatorischen Fähigkeiten wurden unter Kontrolle von Leseverständnis und Intelligenz auf das Ausmaß von kleinen Korrelationen vermindert (vgl. Tabelle 4.46). Im Gegensatz dazu war das Niveau der Interkorrelation der Kom-

ponenten des wissenschaftlichen Denkens höher als die Korrelationen der Komponenten zu den formal-operatorischen Fähigkeiten.

Tabelle 4.46

Latente Korrelationen zwischen Wissenschaftsverständnis bzw. Methodenkompetenz und formal-operatorischen Fähigkeiten

	1.	2.	3.
Jahrgangsstufe 4 (n = 141)			
1. formal operatorische Fähigkeiten	1	0.14	0.25
2. Wissenschaftsverständnis	0.55	1	0.90
3. Methodenkompetenz	0.61	0.79	1

Anmerkung: Latente Korrelationen basierend auf dreidimensionalen Rasch-Skalierungen stehen unterhalb der Diagonale. Bedingte latente Korrelationen unter Kontrolle von Leseverständnis und Intelligenz basierend auf dreidimensionalen latenten Regressionen stehen oberhalb der Diagonale.

Um den Zusammenhang zwischen wissenschaftlichem Denken und operatorischem Niveau zu betrachten, wurden Voraussetzungsanalysen mittels Kontingenztafeln durchgeführt. Zunächst wurde den Kindern nach den angepassten Punktkriterien (vgl. Abschnitt 3.2.2.6, S. 68) ihr operatorisches Level zugeordnet. Dabei wurden die Kategorien „konkret“ und „hoch konkret“ zur Stufe der konkreten Operationen und die Kategorien „Übergang“, „gering formal“ und „formal“ zur Stufe der formalen Operationen zusammengefasst. Mit 79.4 % wurde die Mehrheit der Kinder in die Stufe der konkreten Operationen klassifiziert. Nur 20.6 % der Kinder erreichten das formal-operatorische Niveau. Tabelle 4.47 zeigt die Klassifikation der Viertklässler in die operatorischen Niveaus sowie den Prozentsatz der Kinder mit Kompetenz im wissenschaftlichen Denken in Abhängigkeit von ihrem operatorischem Niveau. In diesem Fall wurde dann Kompetenz im wissenschaftlichen Denken angenommen, wenn die Kinder mindestens 65 % der 20 Aufgaben richtig beantwortet hatten. Wie in Tabelle 4.47 zu sehen ist, erzielten die Kinder, die als formal-operatorisch eingestuft wurden, bessere Leistungen im wissenschaftlichen Denken, allerdings wurde der Unterschied nicht signifikant ($\chi^2(1) = 0.99, p > .05, ns$).

Tabelle 4.47

Prozentsatz der Viertklässler (N = 141), die mindestens 65 % der Aufgaben zum wissenschaftlichen Denken richtig gelöst haben, als einen Funktion des operatorischen Niveaus

Kompetenz	Klassifikation operatorisches Niveau		χ^2 -Test
	konkret (n = 112)	formal (n = 29)	
Wissenschaftliches Denken	62.5	72.4	$\chi^2(1) = 0.99, p > .05$

Eine differenzierte Betrachtung des wissenschaftlichen Denkens untergliedert in die Komponenten Wissenschaftsverständnis und Methodenkompetenz in Abhängigkeit von der Zuordnung des operatorischen Niveaus findet sich in Tabelle 4.48. Kinder, die als formal-operatorisch klassifiziert wurden, erzielten nicht signifikant bessere Leistungen in den Aufgaben zum Wissenschaftsverständnis (mindestens 65 % der Aufgaben korrekt beantwortet; $\chi^2(1) = 0.09, p > .05, ns$). In der Komponente Methodenkompetenz zeigte sich ein signifikanter Unterschied: Kinder auf dem formal-operatorischem Niveau scorten signifikant höher in Aufgaben zur Methodenkompetenz (mindestens 65 % der Aufgaben richtig gelöst) als Kinder auf dem konkret-operatorischem Niveau ($\chi^2(1) = 3.98, p < .05$).

Tabelle 4.48

Prozentsatz der Viertklässler (N = 141), die Komponenten-Aufgabensets richtig (mindestens 65 %) gelöst haben, als einen Funktion des operatorischen Niveaus

Kompetenz	Klassifikation operatorisches Niveau		χ^2 -Test
	konkret (n = 112)	formal (n = 29)	
Wissenschaftsverständnis	69.6	72.4	$\chi^2(1) = 0.09, p > .05$
Methodenkompetenz	48.2	69.0	$\chi^2(1) = 3.98, p < .05$

4.4 Zusammenhang zwischen wissenschaftlichem Denken und ausgewählten motivationalen Orientierungen

4.4.1 Naturwissenschaftsbezogene Einstellungen

Skalierung mit dem Partial-Credit-Modell

Die Daten von 272 Kindern stellten die Basis der Skalierung des Fragebogens zur Erfassung der naturwissenschaftsbezogenen Einstellungen dar (Klasse 2: $n = 71$; Klasse 3: $n = 55$; Klasse 4: $n = 146$). Die Antworten auf die drei Items wurden entsprechend der vierstufigen Likert-Skala kodiert und mithilfe des Computerprogramms ConQuest anhand des Partial-Credit-Modells skaliert. Eine detaillierte Beschreibung der statistischen Kennwerte der eindimensionalen Skala zu den naturwissenschaftsbezogenen Einstellungen findet sich in Tabelle B.7 im Anhang B.

Die Varianz der latenten Variable war mit 3.28 als hoch zu bewerten. Die Reliabilitäten (EAP/PV-Reliabilität = 0.73; Cronbachs Alpha = 0.82) erreichten zufrieden stellende Werte. Die Fit-Statistiken für die Itemparameter waren ebenfalls zufrieden stellend. Es konnten keine signifikanten T-Werte ($-2.0 < T\text{-Wert} < 2.0$) festgestellt werden. Mit Werten zwischen 0.82 und 1.08 lag der Infit (Weighted Mean Square, MNSQ) für alle Items innerhalb des von den Richtlinien vorgegebenen Bereichs (Adams, 2002). Der mittlere Infit betrug 0.97 ($SD = 0.13$). Die Ergebnisse zeigten, dass sich die Items modellkonform verhielten und die Daten gut anhand des Partial-Credit-Modells modelliert werden konnten. Die punktbiserialen Korrelationen bzw. die klassischen Trennschärfen reichten von 0.83 bis 0.89, mit einem Mittelwert von 0.86 ($SD = 0.03$). Die Personenparameter der 272 Grundschul Kinder schwankten von -3.35 bis 3.64 logits, mit einem Mittelwert von 1.57 logits ($SD = 1.72$), so dass die Stichprobe der Grundschüler im Verhältnis zur mittleren Zustimmung auf die Items eine relativ hohe positive Einstellung zu Naturwissenschaften hatte.

Einstellungswerte

Die WLE-Personenparameter (Weighted-Likelihood-Schätzer) im Test zu naturwissenschaftsbezogenen Einstellungen wurden anhand des eindimensionalen Partial-Credit-Modells bezogen auf die erhobenen Daten der Gesamtstichprobe ermittelt und im Hinblick

auf Geschlechterunterschiede und Alterseffekte untersucht. In Tabelle 4.49 sind die Einstellungswerte für die einzelnen Klassenstufen und die Gesamtstichprobe nach Mädchen und Jungen getrennt dargestellt.

Tabelle 4.49

Personenparameter (WLE) in naturwissenschaftsbezogenen Einstellungen aufgeteilt nach Geschlecht für die Gesamtstichprobe und die einzelnen Klassenstufen

	Jahrgangsstufe							
	2		3		4		gesamt	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Jungen	1.34	1.98	1.32	2.17	1.45	1.59	1.39	1.82
Mädchen	1.85	1.59	1.82	1.75	1.64	1.59	1.73	1.61
gesamt	1.63	1.77	1.53	2.00	1.55	1.59	1.57	1.72

Anmerkung: *M* = Mittelwert, *SD* = Standardabweichung.

Zur Berechnung der Gruppenunterschiede wurde eine univariate zweifaktorielle Varianzanalyse mit Geschlecht (Jungen, Mädchen) und Klassenzugehörigkeit (Klasse 2, Klasse 3, Klasse 4) als unabhängige Variablen durchgeführt. Dabei fand sich ein tendenzieller Geschlechtereffekt ($F(1,266) = 3.03, p = .083, ns$). Es trat kein Effekt der Klassenstufe ($F(2,266) = 0.02, p > .05, ns$) und auch kein Interaktionseffekt auf ($F(2,266) = 0.27, p > .05, ns$). Die Betrachtung der Mittelwerte zeigte, dass die Mädchen über eine positivere Einstellungen zu Naturwissenschaften bzw. zum Sachunterricht verfügten als die Jungen. Dieser Unterschied war sowohl in der Gesamtstichprobe als auch in den einzelnen Jahrgangsstufen feststellbar (vgl. Tabelle 4.49).

Beziehung zwischen naturwissenschaftsbezogenen Einstellungen und wissenschaftlichem Denken

Es konnte nur für das frühe Grundschulalter (Klassenstufe 2-3) ein geringer bis mittlerer Zusammenhang zwischen positiven Einstellungen zum Sachunterricht und wissenschaftlichem Denken gefunden werden (vgl. Tabelle 4.50). Weder in der Gesamtstichprobe noch in Klassenstufe 4 trat ein Zusammenhang auf.

Tabelle 4.50

Latente Korrelationen zwischen wissenschaftlichem Denken und naturwissenschaftsbezogenen Einstellungen

	1.	2.
Gesamtstichprobe (N = 272)		
1. Einstellungen	1	
2. wissenschaftliches Denken	0.05	1
Jahrgangsstufe 2-3 (n = 126)		
1. Einstellungen	1	
2. wissenschaftliches Denken	0.30	1
Jahrgangsstufe 4 (n = 146)		
1. Einstellungen	1	
2. wissenschaftliches Denken	-0.17	1

Anmerkung: Latente Korrelationen basierend auf zweidimensionalen Partial-Credit-Modellen stehen unterhalb der Diagonale.

Eine differenzierte Betrachtung der Zusammenhänge zwischen naturwissenschaftsbezogenen Einstellungen und den Komponenten des wissenschaftlichen Denkens ergab ebenfalls nur in Klassenstufe 2-3 einen kleinen bis mittleren positiven Effekt zur Methodenkompetenz und einen moderaten Effekt zum Wissenschaftsverständnis (vgl. Tabelle 4.51). In Klassenstufe 4 zeigte sich dagegen ein mittlerer negativer Zusammenhang zum Verständnis von Wissenschaften.

Tabelle 4.51

Latente Korrelationen zwischen Wissenschaftsverständnis bzw. Methodenkompetenz und naturwissenschaftsbezogenen Einstellungen

	1.	2.	3.
Gesamtstichprobe (N = 272)			
1. Einstellungen	1		
2. Wissenschaftsverständnis	0.03	1	
3. Methodenkompetenz	0.05	0.88	1
Jahrgangsstufe 2-3 (n = 126)			
1. Einstellungen	1		
2. Wissenschaftsverständnis	0.46	1	
3. Methodenkompetenz	0.24	0.95	1
Jahrgangsstufe 4 (n = 146)			
1. Einstellungen	1		
2. Wissenschaftsverständnis	-0.35	1	
3. Methodenkompetenz	-0.11	0.79	1

Anmerkung: Latente Korrelationen basierend auf dreidimensionalen Partial-Credit-Modellen stehen unterhalb der Diagonale.

4.4.2 Naturwissenschaftsbezogenes Fähigkeitsselbstkonzept

Skalierung mit dem Partial-Credit-Modell

Die Daten von 272 Kindern bildeten die Grundlage zur Skalierung des Fragebogens zur Erfassung des naturwissenschaftsbezogenen Fähigkeitsselbstkonzepts (Klasse 2: $n = 71$; Klasse 3: $n = 55$; Klasse 4: $n = 146$). Die Antworten auf die vier Items wurden entsprechend der vierstufigen Likert-Skala kodiert und anhand des Partial-Credit-Modells mithilfe des Computerprogramms ConQuest analysiert. Im Anhang B in Tabelle B.8 steht eine detaillierte Beschreibung der statistischen Kennwerte der eindimensionalen Skala zur Erfassung des naturwissenschaftsbezogenen Fähigkeitsselbstkonzepts.

Die Varianz der latenten Variable war mit 1.66 ausreichend hoch. Die Reliabilitäten (EAP/PV-Reliabilität = 0.66; Cronbachs Alpha = 0.74) erreichten befriedigende Werte. Die Fit-Statistiken für die Itemparameter waren zufrieden stellend. Die T-Werte ($-2.0 < T\text{-Wert} < 2.0$) wurden nicht signifikant. Mit Werten zwischen 0.88 und 1.08 lag der Infit (Weighted Mean Square, MNSQ) für alle Items innerhalb des von den Richtlinien vorgegebenen Bereichs (Adams, 2002). Der mittlere Infit betrug 0.99 ($SD = 0.09$). Somit ver-

hielten sich die Items modellkonform und die Daten waren gut anhand des Partial-Credit-Modells modellierbar. Die punktbiserialen Korrelationen bzw. die klassischen Trennschärfen schwankten von 0.72 bis 0.77, mit einem Mittelwert von 0.75 ($SD = 0.02$). Die Personenparameter der 272 Grundschul Kinder reichten von -2.92 bis 3.61 logits, mit einem Mittelwert von 1.40 logits ($SD = 1.39$), so dass die Stichprobe der Grundschüler im Verhältnis zur mittleren Zustimmung auf die Items eine relativ hohes naturwissenschaftsbezogenes Fähigkeitsselbstkonzept besaß.

Fähigkeitsselbstkonzeptwerte

Anhand des eindimensionalen Partial-Credit-Modells wurden die WLE-Personenparameter (Weighted-Likelihood-Schätzer) bezogen auf die Daten der gesamten Stichprobe im Fragebogen zum naturwissenschaftsbezogenen Fähigkeitsselbstkonzept berechnet und im Hinblick auf Geschlechterunterschiede und Alterseffekte überprüft. In Tabelle 4.52 sind die Fähigkeitsselbstkonzeptwerte für die einzelnen Klassenstufen und die Gesamtstichprobe nach Mädchen und Jungen getrennt dargestellt.

Tabelle 4.52

Personenparameter (WLE) im naturwissenschaftsbezogenem Fähigkeitsselbstkonzept aufgeteilt nach Geschlecht für die Gesamtstichprobe und die einzelnen Klassenstufen

	Jahrgangsstufe							
	2		3		4		gesamt	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Jungen	1.21	1.54	1.07	1.23	1.28	1.21	1.21	1.29
Mädchen	1.42	1.46	1.85	1.42	1.58	1.47	1.58	1.46
gesamt	1.33	1.48	1.40	1.36	1.43	1.36	1.40	1.39

Anmerkung: M = Mittelwert, SD = Standardabweichung.

Zur Überprüfung auf Gruppenunterschiede wurde eine univariate zweifaktorielle Varianzanalyse mit Geschlecht (Jungen, Mädchen) und Klassenstufe (Klasse 2, Klasse 3, Klasse 4) als unabhängige Variablen berechnet. Die Ergebnisse zeigten einen signifikanten Geschlechtereffekt ($F(1,266) = 5.43, p < .05$). Es fand sich kein Effekt der Klassenstufe ($F(2,266) = 0.22, p > .05, ns$) und auch kein Interaktionseffekt ($F(2,266) = 0.73, p > .05, ns$). Die Betrachtung der Mittelwerte zeigte, dass die Mädchen ein höheres naturwissenschaftsbezogenes Fähigkeitsselbstkonzept besaßen als die Jungen. Dieser Unterschied fand

sich sowohl in der Gesamtstichprobe als auch in den einzelnen Jahrgangsstufen (vgl. Tabelle 4.52).

Beziehung zwischen naturwissenschaftsbezogenem Fähigkeitsselbstkonzept und wissenschaftlichem Denken

Sowohl in der Gesamtstichprobe als auch getrennt betrachtet für das frühe und späte Grundschulalter konnte ein mittlerer positiver Zusammenhang zwischen dem naturwissenschaftsbezogenem Fähigkeitsselbstkonzept und der Kompetenz der Kinder im wissenschaftlichen Denken festgestellt werden (vgl. 4.53).

Tabelle 4.53

Latente Korrelation zwischen wissenschaftlichem Denken und naturwissenschaftsbezogenem Fähigkeitsselbstkonzept

	1.	2.
Gesamtstichprobe (N = 272)		
1. Fähigkeitsselbstkonzept	1	
2. wissenschaftliches Denken	0.33	1
Jahrgangsstufe 2-3 (n = 126)		
1. Fähigkeitsselbstkonzept	1	
2. wissenschaftliches Denken	0.41	1
Jahrgangsstufe 4 (n = 146)		
1. Fähigkeitsselbstkonzept	1	
2. wissenschaftliches Denken	0.35	1

Anmerkung: Latente Korrelationen basierend auf zweidimensionalen Partial-Credit-Modellen stehen unterhalb der Diagonale.

Wird die Kompetenz im wissenschaftlichen Denken differenziert in Wissenschaftsverständnis und Methodenkompetenz, traten ebenfalls mittlere Zusammenhänge auf (vgl. Tabelle 4.54). Während im frühen Grundschulalter der Zusammenhang des Fähigkeitsselbstkonzepts mit dem Wissenschaftsverständnis der Kinder recht hoch und verhältnismäßig höher als zur Methodenkompetenz ausfiel, drehte sich das Bild in Klassenstufe 4 um und es zeigte sich ein mittlerer Zusammenhang zur Methodenkompetenz, allerdings nur eine Korrelation mit einem kleinen Effekt zum Wissenschaftsverständnis.

Tabelle 4.54

Latente Korrelationen zwischen Wissenschaftsverständnis bzw. Methodenkompetenz und naturwissenschaftsbezogenem Fähigkeitsselbstkonzept

	1.	2.	3.
Gesamtstichprobe (N = 272)			
1. Fähigkeitsselbstkonzept	1		
2. Wissenschaftsverständnis	0.28	1	
3. Methodenkompetenz	0.32	0.90	
Jahrgangsstufe 2-3 (n = 126)			
1. Fähigkeitsselbstkonzept	1		
2. Wissenschaftsverständnis	0.58	1	
3. Methodenkompetenz	0.33	0.90	1
Jahrgangsstufe 4 (n = 146)			
1. Fähigkeitsselbstkonzept	1		
2. Wissenschaftsverständnis	0.13	1	
3. Methodenkompetenz	0.38	0.80	1

Anmerkung: Latente Korrelationen basierend auf dreidimensionalen Partial-Credit-Modellen stehen unterhalb der Diagonale.

4.4.3 Interesse an forschenden Aktivitäten

Skalierung mit dem Partial-Credit-Modell

Den Fragebogen zum Interesse an forschenden Aktivitäten hatten 269 Kinder (Klasse 2: $n = 68$; Klasse 3: $n = 54$; Klasse 4: $n = 147$) ausgefüllt. Die Antworten der Gesamtstichprobe auf die fünf Items wurden entsprechend der vierstufigen Likert-Skala kodiert und anhand des Partial-Credit-Modells mithilfe des Computerprogramms ConQuest skaliert. Eine detaillierte Beschreibung der statistischen Kennwerte der eindimensionalen Skala zum Interesse an forschenden Aktivitäten ist in Tabelle B.9 im Anhang B zusammengefasst.

Die latente Verteilung wies mit einem Wert von 0.69 eine ausreichend hohe Varianz auf. Die Reliabilitäten (EAP/PV-Reliabilität = 0.60, Cronbachs Alpha = 0.65) erreichten befriedigende Werte. Die Itemparameter zeigten zufrieden stellende Fit-Statistiken. Es konnten keine signifikanten T-Werte ($-2.0 < \text{T-Wert} < 2.0$) festgestellt werden. Der Infit ($M = 1.01$, $SD = 0.07$) lag mit Werten zwischen 0.91 und 1.07 für alle Items innerhalb des als gut zu bewertenden Bereichs (Adams, 2002). Somit verhielten sich die Items modellkonform und die Daten waren gut anhand des Partial-Credit-Modells modellierbar. Die

punktbiserialen Korrelationen bzw. die klassischen Trennschärfen schwankten von 0.55 bis 0.71, mit einem Mittelwert von 0.64 ($SD = 0.06$). Die Personenparameter der 269 Grundschul Kinder lagen durchschnittlich bei 0.80 logits ($SD = 1.09$) und reichten von -2.21 bis 3.67 logits, so dass die Stichprobe der Grund Kinder im Verhältnis zur mittleren Zustimmung auf die Items eine durchschnittliches Interesse an forschenden Aktivitäten aufwies.

Interessewerte

Das Interesse an forschenden Aktivitäten (WLE, Weighted-Likelihood-Schätzer) der Grundschul Kinder wurde durch das eindimensionale Partial-Credit-Modell der erhobenen Daten der Gesamtstichprobe berechnet. Im Folgenden werden die Interessewerte im Hinblick auf Geschlechter- und Altersunterschiede betrachtet. Die Personenparameter im Fragebogen zum Interesse an forschenden Aktivitäten sind in Tabelle 4.55 getrennt nach Geschlecht für die einzelnen Klassenstufen sowie für die Gesamtstichprobe gegenübergestellt.

Tabelle 4.55

Personenparameter (WLE) im Interesse an forschenden Aktivitäten aufgeteilt nach Geschlecht für die Gesamtstichprobe und die einzelnen Klassenstufen

	Jahrgangsstufe							
	2		3		4		gesamt	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Jungen	1.16	1.16	1.33	0.99	1.02	1.20	1.13	1.14
Mädchen	0.38	1.10	0.68	0.87	0.48	0.88	0.48	0.95
gesamt	0.70	1.18	1.07	0.99	0.74	1.08	0.80	1.09

Anmerkung: *M* = Mittelwert, *SD* = Standardabweichung.

Zur Überprüfung auf Geschlechter- und Altersunterschiede wurde eine univariate zweifaktorielle Varianzanalyse mit Geschlecht (Jungen, Mädchen) und Klassenstufe (Klasse 2, Klasse 3, Klasse 4) als unabhängige Variablen berechnet. Die Ergebnisse stellten einen signifikanten Geschlechtereffekt heraus ($F(1,261) = 21.27, p < .001$). Der Haupteffekt des Faktors Klassenstufe ($F(2,261) = 1.21, p > .05, ns$) wurde nicht signifikant. Es trat auch keine Interaktion zwischen den Faktoren Geschlecht und Klassenstufe auf ($F(2,261) = 0.30, p > .05, ns$). Der Vergleich der Mittelwerte ergab, dass die Jungen ein höheres Interesse an forschenden Aktivitäten hatten als die Mädchen. Dieser Unterschied fand sich

sowohl in der Gesamtstichprobe als auch in den einzelnen Jahrgangsstufen (vgl. Tabelle 4.55).

Beziehung zwischen Interesse an forschenden Aktivitäten und wissenschaftlichem Denken

Die Betrachtung der Relation zwischen der Kompetenz im wissenschaftlichen Denken und dem Interesse an forschenden Aktivitäten zeigte kleine bis mittlere positive Zusammenhänge in der Gesamtstichprobe und in Klassenstufe 4 (vgl. Tabelle 4.56). Im frühen Grundschulalter fiel die Korrelation etwas höher aus.

Tabelle 4.56

Latente Korrelationen zwischen wissenschaftlichem Denken und Interesse an forschenden Aktivitäten

	1.	2.
Gesamtstichprobe (N = 269)		
1. Interesse an forschenden Aktivitäten	1	
2. wissenschaftliches Denken	0.20	1
Jahrgangsstufe 2-3 (n = 122)		
1. Interesse an forschenden Aktivitäten	1	
2. wissenschaftliches Denken	0.39	1
Jahrgangsstufe 4 (n = 147)		
1. Interesse an forschenden Aktivitäten	1	
2. wissenschaftliches Denken	0.22	1

Anmerkung: Latente Korrelationen basierend auf zweidimensionalen Partial-Credit-Modellen stehen unterhalb der Diagonale.

Werden die Zusammenhänge des Interesses an forschenden Aktivitäten zu beiden Komponenten des wissenschaftlichen Denkens untersucht, war in der Gesamtstichprobe und der Klassenstufe 4 nur ein kleiner bis moderater Effekt zur Methodenkompetenz festzustellen (vgl. Tabelle 4.57). Im frühen Grundschulalter ergaben sich dagegen mittlere Zusammenhänge zu beiden Kompetenzkomponenten, wobei der Effekt für das Wissenschaftsverständnis etwas höher ausfiel.

Tabelle 4.57

Latente Korrelationen zwischen Wissenschaftsverständnis bzw. Methodenkompetenz und Interesse an forschenden Aktivitäten

	1.	2.	3.
Gesamtstichprobe (N = 269)			
1. Interesse an forschenden Aktivitäten	1		
2. Wissenschaftsverständnis	0.10	1	
3. Methodenkompetenz	0.23	0.86	
Jahrgangsstufe 2-3 (n = 122)			
1. Interesse an forschenden Aktivitäten	1		
2. Wissenschaftsverständnis	0.48	1	
3. Methodenkompetenz	0.35	0.95	1
Jahrgangsstufe 4 (n = 147)			
1. Interesse an forschenden Aktivitäten	1		
2. Wissenschaftsverständnis	0.02	1	
3. Methodenkompetenz	0.30	0.72	1

Anmerkung: Latente Korrelationen basierend auf dreidimensionalen Partial-Credit-Modellen stehen unterhalb der Diagonale.

4.5 Latente hierarchische Regressionsanalysen zur Vorhersage der Kompetenz im wissenschaftlichen Denken

Die Prognostizierbarkeit der Kompetenz im wissenschaftlichen Denken durch die untersuchten kognitiven und motivationalen Korrelate wurde mit latenten hierarchischen Regressionsanalysen untersucht. Die latenten Regressionen wurden sowohl für die Gesamtstichprobe als auch für die einzelnen Klassenstufen berechnet. In die latenten multiplen Regressionsanalysen wurden die aus einfachen latenten Regressionen ermittelten signifikanten Prädiktoren⁸ (vgl. Tabellen B.11-B.13 im Anhang B) sukzessiv eingeführt.

⁸ Da eine Berechnung mittels der WLE-Personenfähigkeiten aufgrund der Dezimalzahlen vom negativen bis positiven Zahlenbereich in den latenten Regressionen mittels ConQuest nicht möglich war, wurden die Summenwerte statt der WLE-Personenfähigkeiten in die latenten Regressionen als Prädiktoren eingeführt. Die Berechnung der WLE-Personenfähigkeiten beruht allerdings auf den Summenwerten als suffiziente Statistiken. Daher sind diese beiden Leistungsmaße mit Werten zwischen $r = 0.95$ und $r = 1.00$ sehr hoch korreliert (vgl. Tabelle B.10 im Anhang B).

4.5.1. Gesamtstichprobe

Die Überprüfung auf Multikollinearität der Prädiktoren anhand der Interkorrelationen (vgl. Tabelle 4.58), des Varianzinflationsfaktors und der Toleranz-Statistiken ergab keine problematischen Werte, so dass das Auftreten von Suppressionseffekten nicht anzunehmen war.

Tabelle 4.58

Pearson-Interkorrelationen der Prädiktoren in der Gesamtstichprobe

	1.	2.	3.	4.	5.	6.	7.	8.	9.
1. Klassenstufe	1								
2. Leseverständnis	0.52**	1							
3. Intelligenz	0.36**	0.49**	1						
4. Problemlösen	0.34**	0.45**	0.48**	1					
5. Inhibition	0.46**	0.34**	0.37**	0.27**	1				
6. räumliches Denken	0.28**	0.24**	0.39**	0.43**	0.20**	1			
7. Einstellungen	0.01	-0.04	-0.04	-0.07	0.13*	-0.01	1		
8. Selbstkonzept	0.04	0.23**	0.16**	0.15*	0.21**	0.06	0.50**	1	
9. Interesse	-0.03	-0.03	0.12	0.14*	-0.01	0.03	0.21**	0.17**	1

*Anmerkung:*** Die Korrelation ist auf einem Niveau von $p = .01$ signifikant. * Die Korrelation ist auf einem Niveau von $p = .05$ signifikant.

In Tabelle 4.59 sind die Ergebnisse der einzelnen Modelle der latenten hierarchischen Regressionen zur Vorhersage der Kompetenz im wissenschaftlichen Denken im Grundschulalter zusammengefasst.

Entsprechend der hierarchischen Vorgehensweise wurde im Modell 1 zunächst nur die Klassenstufe als Prädiktor berücksichtigt. Die Variable Klassenstufe hatte ein signifikantes Regressionsgewicht und es wurden 33 Prozent der Varianz der latenten Variable wissenschaftliches Denken erklärt. Der Regressionskoeffizient B bedeutet, dass bei einem Anstieg der Klassenstufe um eine Einheit (z. B. von Klassenstufe 2 auf 3) die Kompetenz im wissenschaftlichen Denken um 0.60 logits zunahm.

In Modell 2 wurden zusätzlich das Leseverständnis und die Intelligenz als Prädiktoren eingeführt. Alle drei Variablen hatten signifikante Regressionsgewichte und zusätzliche 18 Prozent der Varianz konnten erklärt werden. Das Leseverständnis hatte dabei einen großen Effekt auf die Kompetenz im wissenschaftlichen Denken. Das standardisierte Regressionsgewicht der Prädiktoren Leseverständnis und Intelligenz gibt an, dass bei einer Zunahme

des Leseverständnisses bzw. der Intelligenz um eine Standardabweichung sich die Kompetenz im wissenschaftlichen Denken um 0.33 bzw. 0.16 Standardabweichungen änderte.

In Modell 3 wurden als weitere Prädiktoren die kognitiven Variablen Problemlösen, Inhibition und räumliches Denken aufgenommen. Der Anteil der erklärten Varianz stieg erneut an. Es konnten insgesamt 57 Prozent der Varianz der latenten Verteilung erklärt werden. Die Regressionsgewichte von Klassenstufe und Leseverständnis blieben signifikant, nahmen aber ab. Das Gewicht von Intelligenz wurde nicht mehr signifikant. Auffallend war der Effekt des planerisch-problemlösenden Denkens auf die Kompetenz im wissenschaftlichen Denken. Die Inhibitionsleistung lieferte unter Berücksichtigung des Alters und der anderen kognitiven Variablen keinen signifikanten Einfluss. Die Variable räumliches Denken hatte ebenfalls ein signifikantes Regressionsgewicht. Unter Kontrolle der kognitiven Variablen lieferte der Prädiktor Intelligenz keinen weiteren eigenen Anteil zur Varianzerklärung. Problemlösefähigkeiten und Intelligenz trugen vermutlich zu einem ähnlichen Varianzanteil im wissenschaftlichen Denken bei.

Die Variablen naturwissenschaftsbezogenes Fähigkeitsselbstkonzept und Interesse an forschenden Aktivitäten wurden in Modell 4 als zusätzliche Prädiktoren eingeführt. Das Fähigkeitsselbstkonzept lieferte einen signifikanten Beitrag zur weiteren Varianzerklärung von allerdings nur zwei Prozent. Das Interesse an forschenden Aktivitäten hatte keinen signifikanten Effekt. Die anderen Regressionsgewichte sanken etwas ab.

Werden die Analysen zusammengefasst, so ist festzustellen, dass insgesamt 59 Prozent der Varianz der latenten Verteilung durch die Prädiktoren Klassenstufe, Leseverständnis, Intelligenz, Problemlösen, räumliches Denken und Fähigkeitsselbstkonzept erklärt wurden, wobei Intelligenz nach Aufnahme der Prädiktoren Problemlösefähigkeiten, räumliche Fähigkeiten und Inhibitionsfähigkeiten, keinen signifikanten Beitrag zur Varianzaufklärung mehr leistete.

Tabelle 4.59

Ergebnisse latenter hierarchischer Regressionen zur Vorhersage der Kompetenz im wissenschaftlichen Denken bezogen auf die Gesamtstichprobe

Modell	M1			M2			M3			M4		
Prädiktor	<i>B</i>	<i>SE B</i>	<i>Beta</i>	<i>B</i>	<i>SE B</i>	<i>Beta</i>	<i>B</i>	<i>SE B</i>	<i>Beta</i>	<i>B</i>	<i>SE B</i>	<i>Beta</i>
Klassenstufe	0.60	0.05	0.56*	0.34	0.05	0.29*	0.26	0.06	0.21*	0.28	0.06	0.16*
Leseverständnis				0.06	0.01	0.33*	0.05	0.01	0.24*	0.06	0.01	0.19*
Intelligenz				0.04	0.01	0.16*	0.01	0.01	0.02	0.01	0.02	0.02
Problemlösen							0.12	0.02	0.24*	0.11	0.02	0.16*
Inhibition							0.00	0.01	0.02	-0.00	0.01	-0.02
Räumliches Denken							0.06	0.02	0.14*	0.07	0.02	0.11*
Fähigkeitsselbstkonzept										0.06	0.02	0.11*
Interesse an forschenden Aktivitäten										0.01	0.01	0.03

Anmerkungen: $R^2 = 0.33$ für M1, $\Delta R^2 = 0.18$ für M2, $\Delta R^2 = 0.06$ für M3, $\Delta R^2 = 0.02$ für M4; * = signifikant.

4.5.2 Frühes Grundschulalter

Nach Prüfung der Multikollinearität der Prädiktoren anhand der Interkorrelationen (vgl. Tabelle 4.60), des Varianzinflationsfaktors und der Toleranz-Satistiken waren Suppressions-effekte auszuschließen.

Tabelle 4.60

Pearson-Interkorrelationen der Prädiktoren im frühen Grundschulalter

	1.	2.	3.	4.	5.	6.	7.	8.
1. Leseverständnis	1							
2. Intelligenz	0.45**	1						
3. Problemlösen	0.42**	0.50**	1					
4. Inhibition	0.26**	0.35**	0.23**	1				
5. räumliches Denken	0.11	0.25**	0.48**	0.14	1			
6. Einstellungen	0.04	-0.06	0.03	0.15	-0.07	1		
7. Selbstkonzept	0.23*	0.08	0.08	0.29**	0.00	0.55**	1	
8. Interesse	0.09	0.28**	0.25**	0.08	0.15	0.10	0.04	1

*Anmerkung:*** Die Korrelation ist auf einem Niveau von $p = .01$ signifikant. * Die Korrelation ist auf einem Niveau von $p = .05$ signifikant.

Nach der analogen hierarchischen Vorgehensweise der latenten Regressionsanalyse wurden in Modell 1 zunächst die Prädiktoren Leseverständnis und Intelligenz eingeführt (vgl. Tabelle 4.61). Nur die Variable Leseverständnis hatte ein signifikantes Regressionsgewicht mit einem starken Effekt auf die Kompetenz im schriftlichen Test zur Erfassung des wissenschaftlichen Denkens. Das standardisierte Regressionsgewicht des Prädiktors Leseverständnis besagt, dass die Kompetenz im wissenschaftlichen Denken um 0.48 Standardabweichungen anstieg, wenn sich das Leseverständnis um eine Standardabweichung erhöhte. Intelligenz trug unter Kontrolle von Leseverständnis im frühen Grundschulalter nicht zur Varianzerklärung bei. Insgesamt konnten durch das Modell 1 33 Prozent der Varianz der latenten Variable erklärt werden.

Im zweiten Modell wurden zusätzlich die weiteren kognitiven Prädiktoren in die multiple Regressionsanalyse aufgenommen. Nur die Variablen Leseverständnis und Problemlösen hatten signifikante, starke Effekte. Das Regressionsgewicht von Leseverständnis sank unter Kontrolle der weiteren kognitiven Variablen allerdings etwas ab. Die Intelligenz, die Inhibitionsleistung und das räumliche Denken lieferten unter Berücksichtigung

aller erhobenen kognitiven Variablen keinen signifikanten Beitrag zur Varianzaufklärung. Durch Modell 2 konnten weitere fünf Prozent der Varianz der latenten Verteilung erklärt werden.

Modell 3 berücksichtigte darüber hinaus noch die motivationalen Variablen als Prädiktoren für die Leistung im wissenschaftlichen Denken. Erneut verbesserte sich der Anteil der erklärten Varianz um 17 Prozent. Das Regressionsgewicht der Variable Leseverständnis sank wieder etwas ab, ebenso der Regressionskoeffizient von Problemlösen. Zusätzlich hatten die Einstellungen zu Naturwissenschaften im frühen Grundschulalter einen signifikanten Einfluss auf die Kompetenz im wissenschaftlichen Denken. Die Regressionsgewichte der weiteren kognitiven Variablen waren nicht signifikant. Das naturwissenschaftsbezogene Fähigkeitsselbstkonzept und das Interesses an forschenden Aktivitäten hatten nur tendenziell signifikante Effekte.

Eine zusammenfassende Betrachtung der Ergebnisse zeigte, dass im frühen Grundschulalter insgesamt 55 Prozent der Varianz der latenten Verteilung durch die Prädiktoren Leseverständnis, Problemlösen und Einstellungen zu Naturwissenschaften erklärt werden konnten.

Tabelle 4.61

Ergebnisse latenter hierarchischer Regressionen zur Vorhersage der Kompetenz im wissenschaftlichen Denken bezogen auf das frühe Grundschulalter

Modell Prädiktor	M1			M2			M3		
	<i>B</i>	<i>SE B</i>	<i>Beta</i>	<i>B</i>	<i>SE B</i>	<i>Beta</i>	<i>B</i>	<i>SE B</i>	<i>Beta</i>
Leseverständnis	0.07	0.01	0.48*	0.06	0.01	0.37*	0.05	0.01	0.26*
Intelligenz	0.01	0.02	0.03	-0.03	0.02	-0.15	-0.03	0.02	-0.10
Problemlösen				0.13	0.04	0.29*	0.12	0.04	0.21*
Inhibition				0.00	0.01	0.04	-0.01	0.01	-0.06
Räumliches Denken				0.04	0.03	0.10	0.04	0.03	0.08
Einstellungen							0.05	0.02	0.13*
Fähigkeitsselbstkonzept							0.05	0.03	0.12
Interesse an forschenden Aktivitäten							0.03	0.02	0.09

*Anmerkungen: $R^2 = 0.33$ für M1, $\Delta R^2 = 0.05$ für M2, $\Delta R^2 = 0.17$ für M3; * = signifikant.*

4.5.3 Spätes Grundschulalter

Nach Betrachtung der Interkorrelationsmatrix der Prädiktoren (vgl. Tabelle 4.62) sowie der statistischen Kennwerte Varianzinflationsfaktor und Toleranz-Statistiken bestand keine Multikollinearitätsproblematik der Prädiktoren im späten Grundschulalter.

Tabelle 4.62

Pearson-Interkorrelationen der Prädiktoren im späten Grundschulalter

	1.	2.	3.	4.	5.	6.	7.	8.	9.
1. Leseverständnis	1								
2. Intelligenz	0.40**	1							
3. Problemlösen	0.36**	0.39**	1						
4. Inhibition	0.10	0.21**	0.13	1					
5. räumliches Denken	0.22**	0.42**	0.35**	0.10	1				
6. formal-operator. Fähigkeiten	0.31**	0.27**	0.37**	0.12	0.25**	1			
7. Einstellungen	-0.16	-0.05	-0.17*	0.09	0.03	0.09	1		
8. Selbstkonzept	0.25**	0.24**	0.20*	0.13	0.10	0.29**	0.44**	1	
9. Interesse	-0.06	0.02	0.13	-0.03	-0.03	0.13	0.33**	0.30**	1

*Anmerkung:*** Die Korrelation ist auf einem Niveau von $p = .01$ signifikant. * Die Korrelation ist auf einem Niveau von $p = .05$ signifikant.

In Tabelle 4.63 sind die Ergebnisse der latenten hierarchischen Regressionsanalysen zur Vorhersage der Kompetenz im wissenschaftlichen Denken für das späte Grundschulalter dargestellt.

Das erste Modell berücksichtigte zunächst nur die Variablen Leseverständnis und Intelligenz als Prädiktoren. Beide Variablen hatten signifikante Regressionskoeffizienten. Die Standardisierung zeigte, dass Intelligenz und Leseverständnis einen Einfluss von etwa gleicher Höhe auf die Leistung im wissenschaftlichen Denken hatten. Bei einer Änderung um eine Standardabweichung in den Prädiktoren Leseverständnis bzw. Intelligenz stieg die Kompetenz im wissenschaftlichen Denken um 0.29 bzw. 0.28 Standardabweichungen an. Die beiden Prädiktoren erklärten 33 Prozent der Varianz der latenten Variable wissenschaftliches Denken.

In Modell 2 wurden zusätzlich problemlösende, räumliche und formal-operatorische Fähigkeiten als weitere Prädiktoren eingeführt. Die Inhibitionsleistung wurde nicht in die multiple Regression aufgenommen, da sich für Klassenstufe 4 in der einfachen latenten Regressionsanalyse kein signifikanter Effekt auf die Leistung im wissenschaftlichen Denken gezeigt hatte (vgl. Tabelle B.13 im Anhang B). Die Regressionsgewichte von Leseverständnis und Intelligenz blieben signifikant, nahmen allerdings etwas ab. Darüber hinaus hatte die Problemlösefähigkeit den relativ größten Einfluss. Auch der Effekt der Variable räumliches Denken wurde signifikant. Formal-operatorische Fähigkeiten lieferten unter Kontrolle der anderen kognitiven Variablen keinen signifikanten Beitrag zur Varianzerklärung. Durch die Aufnahme von Problemlösen und räumlichem Denken in das Regressionsmodell 2 konnten weitere acht Prozent der latenten Varianz im wissenschaftlichen Denken erklärt werden.

Die Variablen Einstellungen zu Naturwissenschaften und naturwissenschaftsbezogenes Fähigkeitsselbstkonzept wurden in Modell 4 als zusätzliche Prädiktoren eingeführt. Keine motivationale Variable hatte im späten Grundschulalter einen signifikanten Effekt auf die Leistung im wissenschaftlichen Denken. Die Effekte der in Modell 2 ermittelten signifikanten kognitiven Prädiktoren blieben unverändert. Daher konnte nur ein weiteres Prozent der Varianz der latenten Verteilung erklärt werden.

Fasst man die Ergebnisse zusammen, so ist festzustellen, dass in Modell 3 insgesamt 42 Prozent der Varianz der latenten Verteilung durch die Prädiktoren Leseverständnis, Intelligenz, Problemlösen und räumliches Denken erklärt werden konnten.

Tabelle 4.63

Ergebnisse latenter hierarchischer Regressionen zur Vorhersage der Kompetenz im wissenschaftlichen Denken bezogen auf das späte Grundschulalter

Modell Prädiktor	M1			M2			M3		
	<i>B</i>	<i>SE B</i>	<i>Beta</i>	<i>B</i>	<i>SE B</i>	<i>Beta</i>	<i>B</i>	<i>SE B</i>	<i>Beta</i>
Leseverständnis	0.07	0.01	0.29*	0.05	0.01	0.18*	0.05	0.01	0.16*
Intelligenz	0.09	0.02	0.28*	0.05	0.02	0.14*	0.05	0.02	0.13*
Problemlösen				0.12	0.03	0.23*	0.12	0.03	0.21*
Räumliches Denken				0.07	0.03	0.13*	0.07	0.03	0.12*
Formal-operatorische Fähigkeiten				0.00	0.03	0.01	0.00	0.03	0.00
Fähigkeitsselbstkonzept							0.03	0.02	0.06

Anmerkungen: $R^2 = 0.33$ für M1, $\Delta R^2 = 0.08$ für M2, $\Delta R^2 = 0.01$ für M3; * = signifikant.

5 Diskussion

Im Rahmen der Diskussion werden die Ergebnisse der vorliegenden Studie zusammengefasst und im Hinblick auf die aufgestellten Hypothesen interpretiert und diskutiert. Die Diskussion gliedert sich daher in die Interpretation, kritische Bewertung und Einbettung der Ergebnisse in den Gesamtzusammenhang der Forschungsliteratur zum wissenschaftlichen Denken. Abschließend werden die Grenzen der Studie diskutiert sowie Implikationen für die weitere Forschung und die pädagogische Praxis beschrieben.

Reliable Erfassung des wissenschaftlichen Denkens mittels eines schriftlichen Gruppentests im Grundschulalter

Die Ergebnisse der Rasch-Analysen zeigten, dass bereits im Grundschulalter Kompetenzen im wissenschaftlichen Denken anhand schriftlicher Aufgaben im geschlossenen Antwortformat reliabel erfasst werden konnten. Es ergaben sich zufrieden stellende Reliabilitäten der Gesamtskala des wissenschaftlichen Denkens sowohl für das gesamte Grundschulalter von Klassenstufe 2 bis 4 als auch für die differenzierte Betrachtung des frühen und späten Grundschulalters (Klassenstufe 2-3 und Klassenstufe 4). Anhand der Itemstatistiken wurde die Modellkonformität der Items geprüft und es konnte gezeigt werden, dass alle 20 Items des Testheftes sich modellkonform verhielten und die Daten gut durch das Rasch-Modell modelliert werden konnten. Insgesamt konnte der Test für die vorliegende Stichprobe der Zweit-, Dritt- und Viertklässler in der Schwierigkeit als ein wenig zu leicht, aber angemessen beurteilt werden. Die in Abbildung 4.1 (S. 88) dargestellte Wright map, in der Personenfähigkeiten und Itemschwierigkeiten gegenübergestellt sind, ließ erkennen, dass genügend Aufgaben zur Differenzierung der Personenfähigkeiten im mittleren Fähigkeitsbereich vorhanden waren. Es fehlten jedoch Items, um Personen im unteren und insbesondere im oberen Randbereich des Fähigkeitskontinuums differenziert genug erfassen zu können. Daher konnte die Kompetenz der Grundschulkinder im mittleren Leistungsbereich sehr gut beurteilt werden, während bei der Beurteilung von Personen mit extremen Fähigkeitsausprägungen Einschränkungen gegeben waren.

Bei der differenzierten Erfassung der Kompetenzkomponente Wissenschaftsverständnis, respektive der Teilkompetenzen Verständnis der Ziele von Wissenschaften und Theorieverständnis, ergaben sich allerdings Reliabilitätsprobleme. Auch der Teilbereich Dateninterpretation konnte nicht hinreichend reliabel erfasst werden. Daher müssen diese Einschränkungen bei der differenzierten Beurteilung dieser Kompetenzbereiche des wissenschaftlichen Denkens beachtet werden. Sehr gute statistische Kennwerte konnten dagegen für die Kompetenzkomponente Methodenkompetenz sowie deren Teilkompetenz Experimentierstrategien erreicht werden.

Es ist allerdings zu berücksichtigen, dass es aufgrund der geringen Itemzahlen erschwert war, die Rasch-Modelle der Kompetenzkomponenten und Teilkompetenzen des wissenschaftlichen Denkens zu berechnen. Denn nach Adams und Kollegen (Adams et al., 1997; Wu et al., 2005) sind mindestens 15 bis 20 Items pro Skala notwendig, um präzise Schätzungen und hohe Reliabilitäten zur Erfassung komplexer, kognitiver Fähigkeiten zu erhalten.

Latente Struktur der Kompetenz wissenschaftliches Denken

Ein Ziel der vorliegenden Arbeit war es, die latente Struktur der Kompetenz wissenschaftliches Denken im Grundschulalter zu überprüfen. Ausgehend von der Forschungsliteratur zum wissenschaftlichen Denken (z. B. Bullock et al., 2009) und dem dargestellten Kompetenzstrukturmodell (Hardy et al., 2010) wurde angenommen, dass sich die Kompetenz im wissenschaftlichen Denken besser als mehrdimensionales Konstrukt von miteinander verknüpften Teilfähigkeiten statt als ein eindimensionales, einheitliches Konstrukt abbilden lässt.

Zur Überprüfung der Dimensionalität des Konstrukts wurden ein- und mehrdimensionale Rasch-Modelle für die Gesamtstichprobe der Grundschulkinder sowie für die Teilstichproben frühes und spätes Grundschulalter gegenübergestellt. Der Vergleich des eindimensionalen Rasch-Modells, des zweidimensionalen Modells mit den Dimensionen Wissenschaftsverständnis und Methodenkompetenz sowie des vierdimensionalen Modells, das die vier Teilkompetenzen Verständnis der Ziele von Wissenschaften, Theorieverständnis, Experimentierstrategien und Dateninterpretation als separate Dimensionen annahm, stellte entgegen der Erwartung jeweils das eindimensionale Modell als das am besten geeignete Modell zur Erklärung der Daten heraus.

Ferner wurde die Dimensionalität der Kompetenz wissenschaftliches Denken durch die Betrachtung der Höhe der latenten, messfehlerbereinigten Korrelationen der Dimensionen geprüft. Entsprechend der Hypothese fanden sich substantielle Zusammenhänge zwischen den Komponenten respektive Teilkompetenzen des wissenschaftlichen Denkens. Die latenten Korrelationen der Kompetenzkomponenten sowie der Teilkompetenzen fielen insgesamt sehr hoch aus, auch wenn die Effekte der Kontrollvariablen Leseverständnis und Intelligenz berücksichtigt wurden. Dieses Ergebnis der hohen latenten Korrelationen war konsistent mit der Angemessenheit der eindimensionalen latenten Struktur der Modellvergleiche (Adams et al., 1997).

Die latenten Klassenanalysen ergaben zwei latente Klassen auf unterschiedlichem Fähigkeitsniveau, so dass eine latente Klasse von inkompetenteren Kindern von einer Klasse von kompetenteren Kindern unterschieden werden konnte. Die Antwortprofile der latenten Klassen zeigten einen parallelen Verlauf ohne Überschneidungen. Daher waren entgegen der Erwartung keine qualitativ unterschiedlichen Kompetenzprofile festzustellen und keine separaten Subskalen zur differenzierten Erfassung des wissenschaftlichen Denkens notwendig. Das Resultat der parallelen, gestaffelten Kompetenzprofile auf unterschiedlichem Leistungsniveau entsprach dem Ergebnis der Angemessenheit eines eindimensionalen Rasch-Modells (Rost & Georg, 1991).

Zusammenfassend ist festzustellen, dass die Vergleiche der ein- und mehrdimensionalen IRT-Modelle, die hohen latenten Korrelationen der Komponenten bzw. Teilkompetenzen des wissenschaftlichen Denkens sowie die parallelen Verläufe der Kompetenzprofile der latenten Klassenanalysen entgegen der Erwartung auf eine Eindimensionalität der Kompetenz wissenschaftliches Denken deuteten. Es scheint angemessen, die eng miteinander verknüpften Teilkompetenzen des wissenschaftlichen Denkens zu einem einheitlichen, eindimensionalen Konstrukt und einer aussagekräftigen Gesamtskala zusammenzufassen. Daher konnte die aufgestellte Hypothese über eine mehrdimensionale Struktur der Kompetenz wissenschaftliches Denken nicht bestätigt werden.

Eine mögliche Ursache für die entgegen der aufgestellten Hypothese gefundene Eindimensionalität der Kompetenz wissenschaftliches Denken könnte in den geringen Reliabilitäten der Subskalen der einzelnen Komponenten bzw. Teilkompetenzen liegen. Aufgrund des niedrigen Reliabilitätswertes für die Komponente Wissenschaftsverständnis könnte es daher sein, dass es nur möglich war, die Dimension Methodenkompetenz zu erfassen, während die andere Dimension Wissenschaftsverständnis nicht gemessen werden konnte. Im Hinblick auf die einzelnen Teilkompetenzen konnte nur für die latente Dimension

Experimentierstrategien eine zufrieden stellende Reliabilität erzielt werden. Daher könnte es sein, dass nur ein eindimensionales Modell die Daten am besten erklären konnte, da die Teilkompetenzen Verständnis der Ziele von Wissenschaften und Theorieverständnis sowie Dateninterpretation aufgrund der niedrigen Reliabilitäten nicht als separate Dimensionen erfasst werden konnten. Die Konsequenz war ein eindimensionales Konstrukt wissenschaftliches Denken.

Es kann argumentiert werden, dass es nicht möglich war, die verschiedenen Dimensionen des wissenschaftlichen Denkens reliabel zu messen, weil nicht ausreichend Items für eine reliable Erfassung der einzelnen Teilkompetenzen vorhanden waren. Für präzise Parameterschätzungen und hohe Reliabilitäten sind mindestens 15 bis 20 Items pro Skala notwendig (Adams et al., 1997; Wu et al., 2005). Daher wäre es eventuell möglich, mit genügend langen Subskalen die angenommene Mehrdimensionalität des Konstrukts wissenschaftliches Denken abzubilden.

Allerdings sind sowohl in den Daten der vorliegenden Studie, in dem Gesamtdatensatz des Projektes Science-P als auch in anderen Studien Begründungen und Evidenz für eine mögliche Eindimensionalität der Kompetenz wissenschaftliches Denken zu finden: In der Querschnittstudie des Projekts Science-P wurden in einem rotierten Testheftdesign 92 Items in 66 Aufgaben von insgesamt 1916 Grundschulern der zweiten, dritten und vierten Klassenstufe bearbeitet (Möller et al., 2011). Modellvergleiche von ein- und mehrdimensionalen Rasch-Modellen ergaben ebenfalls eine eindimensionale Struktur der Kompetenz wissenschaftliches Denken. Zwischen den Komponenten und Teilkompetenzen des wissenschaftlichen Denkens bestanden hohe latente Interkorrelationen.

In der LOGIC-Studie (Bullock et al., 2009), auf deren Ergebnisse zum wissenschaftlichen Denken ein Großteil der in der vorliegenden Arbeit untersuchten Hypothesen beruhte, wurden nur einzelne Aufgaben und Interviews zu verschiedenen Komponenten des wissenschaftlichen Denkens durchgeführt. Daher waren keine Faktorenanalysen oder Raschskalierungen möglich, um die Dimensionalität des Konstrukts zu untersuchen. Die Schlussfolgerung, dass das Konstrukt wissenschaftliches Denken als ein Bündel an mehr oder weniger verknüpften Teilfähigkeiten beschrieben werden kann, wurde auf den moderaten Korrelationen zwischen den Teilkompetenzen, gemessen durch Einzelaufgaben und Interviews, gezogen. Daher könnte es sein, dass in der vorliegenden Studie die breitere Erfassung des Konstrukts wissenschaftliches Denken anhand einer Skala eine bessere Datenbasis zur Beurteilung der Dimensionalität und der Kompetenzstruktur darstellte.

Außerdem beschrieben auch Studien zur Entwicklung von Messinstrumenten zur Erfassung der Kompetenz wissenschaftliches Denken im Sekundarschulalter Evidenzen für eine eindimensionale Kompetenzstruktur. Hammann und Kollegen (Hammann, Phan & Bayrhuber, 2007) entwickelten einen Multiple-Choice-Test auf Grundlage des Scientific Discovery as Dual Search (SDDS)-Modells mit den drei Subskalen Suche im Hypothesenraum, Testen von Hypothesen und Analyse von Evidenzen für die Klassenstufen 5-6. Die ermittelten hohen, manifesten Interkorrelationen zwischen den drei Dimensionen ($r = 0.64 - 0.78$) sowie die gestaffelten, überschneidungsfreien Kompetenzprofile der latenten Klassenanalyse wiesen auf eine Eindimensionalität des Konstrukts wissenschaftliches Denken im Sekundarschulalter hin.

Glug (2009) beschrieb in ihrer Dissertation die Entwicklung eines Tests zur Erfassung der als prozessbezogenen bezeichneten naturwissenschaftlichen Grundbildung in der neunten Klassenstufe mit den drei Subskalen Identifizieren wissenschaftlicher Hypothesen, Planen einer wissenschaftlichen Untersuchung und Nutzen wissenschaftlicher Ergebnisse. Die Vergleiche von ein- und mehrdimensionalen Rasch-Modellen sowie die hohen latenten Korrelationen der Teildimensionen ergaben eine Eindimensionalität des Konstrukts.

Darüber hinaus könnte man die Ergebnisse von Trainingsstudien als Hinweis für ein einheitliches, kohärentes Konstrukt wissenschaftliches Denken deuten: Beispielsweise konnte eine Intervention mit einem wissenschaftstheoretischem Curriculum bei Viertklässlern sowohl Effekte auf das explizite, verbalisierte Wissen über Wissenschaft als auch auf die Fähigkeit zum Experimentieren erzielen, obwohl es kein explizites Strategietraining gab (Sodian et al., 2006). Daher sind enge Zusammenhänge zwischen Wissenschaftsverständnis und Methodenkompetenz anzunehmen, eventuell sogar im Sinne eines eindimensionalen Konstrukts.

Kompetenzentwicklung im wissenschaftlichen Denken im Grundschulalter

Ein weiteres Ziel der vorliegenden Arbeit war es, individuelle Unterschiede im wissenschaftlichen Denken für das Alter von acht bis zehn Jahren sowie frühe Entwicklungsveränderungen zu beschreiben.

Insgesamt ist festzustellen, dass in der Studie bereits im Grundschulalter Kompetenzen im wissenschaftlichen Denken gefunden werden konnten. Es zeigte sich eine Abnahme der Fehlvorstellungen und Zunahme der wissenschaftlich adäquaten Vorstellungen von der zweiten zur vierten Klassenstufe. Ein substantieller Anteil der Kinder, hauptsächlich der

zweiten und dritten Klassenstufe, wählte diese von der wissenschaftlichen Vorstellung qualitativ unterschiedlichen naiven Vorstellungen und Zwischenvorstellungen. Konform mit den Befunden von Koerber und Kollegen (2011) wurden erst in Klassenstufe 4 mehrheitlich Antworten auf dem Niveau der wissenschaftlichen Vorstellung gewählt.

Um die Ergebnisse der latenten Klassenanalysen nochmals aufzugreifen, in denen zwei latente Klassen auf unterschiedlichem Fähigkeitsniveau gefunden wurden, ist festzuhalten, dass die Zweitklässler noch mehrheitlich der inkompetenteren und die Viertklässler vorwiegend der kompetenteren latenten Klasse angehörten. Die Drittklässler verteilten sich in etwa gleichmäßig auf die beiden latenten Klassen. Mit dem Ende der dritten Jahrgangsstufe bzw. dem Alter von ca. neun Jahren nahm die Wahrscheinlichkeit für den Wechsel von der inkompetenten in die kompetente latente Klasse zu. Damit scheinen, markante Veränderungen im wissenschaftlichen Denken in der dritten Jahrgangsstufe stattzufinden und das Alter von ca. neun Jahren einen entscheidenden Zeitpunkt in der Entwicklung des wissenschaftlichen Denkens im Grundschulalter darzustellen. In einer Studie von Sodian und Koerber (2011) wurden ebenfalls in der dritten Klassenstufe markante Entwicklungsschritte in der Hypothesenprüfung und Evidenzevaluation beobachtet.

Insgesamt stieg entsprechend der Erwartung die Kompetenz im wissenschaftlichen Denken signifikant von der zweiten und dritten zur vierten Klassenstufe an, auch unter Kontrolle von Leseverständnis und Intelligenz. Diese Altersunterschiede fanden sich sowohl in den beiden Kompetenzkomponenten Wissenschaftsverständnis und Methodenkompetenz als auch in den vier Teilkompetenzen Verständnis der Ziele von Wissenschaften, Theorieverständnis, Experimentierstrategien und Dateninterpretation. Dabei wurde deutlich, dass die Schüler der vierten Klassenstufe konsistent in allen Teilbereichen den Zweit- und Drittklässlern überlegen waren. Mit den gefunden Altersunterschieden wurden Ergebnisse aus früheren Studien repliziert (z. B. Koerber et al., 2011; Chen & Klahr, 1999). Im Einklang mit früheren Studien zum wissenschaftlichen Denken traten keine signifikanten Geschlechterdifferenzen im wissenschaftlichen Denken auf (z. B. Bullock et al., 2009; Koerber et al., 2011).

Insgesamt konnte anhand der ähnlichen Kompetenzprofile auf unterschiedlichem Leistungsniveau festgestellt werden, dass den Kindern der drei Jahrgangsstufen die Aufgaben im Teilbereich Verständnis der Ziele von Wissenschaften am leichtesten fielen, gefolgt von Aufgaben zu Experimentierstrategien und Dateninterpretation. Über alle Klassenstufen hinweg zeigten die Grundschüler schwächere Leistungen in der Teilkompetenz Theorieverständnis. Dieses Ergebnis passt zu den längsschnittlichen Befunden der LOGIC-Studie

(Bullock et al., 2009), in der sich das metakonzeptuelle Theorieverständnis ebenfalls am spätesten entwickelte. Dieser deskriptive Entwicklungsverlauf, ermittelt anhand der mittleren Itemschwierigkeiten bzw. Lösungshäufigkeiten der einzelnen Teilkompetenzen, bestätigte sich für die aus den Aufgaben der Teilkompetenzen gebildeten Itemsets mittels des Rasch-Modells sowie der strengeren Guttman-Skala.

Entgegen der Erwartung ergab sich allerdings eine andere Reihenfolge im Entwicklungsverlauf der vier Teilkompetenzen des wissenschaftlichen Denkens, so dass die aufgestellte Hypothese nur teilweise bestätigt werden konnte. Die Analysen zu den Schwierigkeiten bzw. zur Sequenz der einzelnen Teilkompetenzen ergaben das überraschende Ergebnis, dass die Aufgaben zum Verständnis der Ziele von Wissenschaften den Kindern verhältnismäßig leicht fielen. Insgesamt waren diese Aufgaben im Vergleich zu den Aufgaben zum Verständnis der experimentellen Methode, zum Verständnis der Interpretation von Daten und zum Theorieverständnis sogar am leichtesten.

Eine mögliche Erklärung ist, dass die Multiple-Choice-Aufgaben zum Verständnis wissenschaftlicher Konzepte und der Ziele von Wissenschaften nur ein Erkennen, das bedeutet implizites Verständnis der Konzepte Theorie, Hypothese und Experiment sowie der Ziele von Wissenschaften erfassten, dadurch dass aus vorgegebenen Antwortalternativen gewählt werden musste. Bisher wurden zum deklarativen Wissenschaftsverständnis im Grundschulalter nur Interviews durchgeführt, die eine aktive Produktion der Konzepte und damit ein explizites Verständnis erforderten. In derartigen Interviews konnten im Grundschulalter nur Bodeneffekte berichtet werden (Sodian et al., 2006; Grygier, 2008). Die Validierungsstudie des Projektes Science-P zeigte, dass die Kinder im schriftlichen Test mit geschlossenem Antwortformat bessere Leistungen erzielten als in den entsprechenden Interviews (Kropf, 2010). Dieser Befund der besseren Leistungen in Aufgaben mit geschlossenem Antwortformaten ist vergleichbar mit der von Bullock und Ziegler (1999) gefundenen Leistungsdifferenz in Produktion- und Wahlaufgaben im Bereich der Variablenkontrollstrategie. Die Ergebnisse können dahingehend interpretiert werden, dass Grundschulkinder scheinbar schon früher als bisher aufgrund von Interviewstudien angenommen in schriftlichen, kontextualisierten Aufgaben, in denen eine Antwort aus vorgegebenen Alternativen ausgewählt werden muss, ein beginnendes, implizites Verständnis der Konzepte Theorie, Hypothese, Experiment und der Ziele von Wissenschaften zeigen, auch wenn sie noch kein explizites, epistemologisches Verständnis der Konzepte in Produktionsaufgaben besitzen. In den Interviewstudien wurde die Kompetenz der Kinder eventuell

unterschätzt, während die Multiple-Choice-Aufgaben das Wissenschaftsverständnis der Kinder möglicherweise überschätzten.

Im Gegensatz zu den Aufgaben zum Verständnis der Ziele von Wissenschaften erfassen die Aufgaben zum Theorieverständnis ein tieferes, metakonzeptuelles Wissenschaftsverständnis, nämlich das Verständnis von kohärenten, alternativen Interpretationsrahmen und die Theorieabhängigkeit der Interpretation von Phänomenen und Evidenzen. Entsprechend der Erwartung basierend auf Befunden der LOGIC-Studie (Bullock et al., 2009) waren diese Aufgaben zum Verständnis von Theoriegebäuden für die Kinder am schwierigsten und Kompetenz darin entwickelte sich am spätesten.

Die Kinder lösten die Aufgaben der Teilkompetenz Experimentierstrategien mit einer höheren Wahrscheinlichkeit als die Aufgaben zur Dateninterpretation. Dies wurde ebenfalls durch das Guttman-Antwortmuster bestätigt. Daher konnte gefolgert werden, dass sich das Verständnis der experimentellen Methodik früher entwickelte als die Fähigkeit zur evidenzbasierten Interpretation von Daten. Bei einer genaueren Betrachtung der Inhalte der Aufgaben der beiden Teilbereiche ist diese Schlussfolgerung allerdings nur bedingt haltbar: Beispielsweise drehte es sich bei den Aufgaben a51_Maus-Haus und a53_Spürnase zum Verständnis der experimentellen Methode nicht um Kausalhypothesen, sondern um Hypothesen über singuläre Sachverhalte („die Maus ist groß oder klein“, „das Haustier kann gut oder schlecht riechen“). Bei der Evidenz handelte es sich um eindeutige bzw. nicht eindeutige Ergebnisse von Handlungen und nicht um Kovariationsdaten. Die Beurteilung eines inkonklusiven Tests konnte auch schon ein Großteil der Zweitklässler richtig lösen. Dieses Ergebnis replizierte die Interview-Befunde von Sodian et al. (1991). Wenn nur eine Variablendimension in einem Multiple-Choice-Item über eine Kausalhypothese mit Evidenz in Form von Kovariationsdaten berücksichtigt werden musste, lösten die Kinder sowohl Aufgaben zu Experimentierstrategien (drei Items: a44_1_Ärzte, a57_1_Fußpilz, a49_Pizzabacken) als auch Aufgaben zur Dateninterpretation (zwei Aufgaben: a76_Diagramm, a82_Tabelle) häufiger, als wenn mehrere Variablendimensionen bei der Lösung der Aufgabe einbezogen werden mussten. Die Interpretation eines Zusammenhangs zweier Variablen aus Diagrammen konnten die Kinder schon gut verstehen. Dies passt zu den Ergebnissen früherer Studien (Koerber, 2006; Koerber & Sodian, 2008, 2009; Koerber, 2011), die bereits im Vorschulalter ein Grundverständnis von graphischen Repräsentationsformen und der Interpretation von Kovariationsdaten aus Graphen fanden. Die Berücksichtigung mehrerer Variablendimensionen war im Bereich des Verständnisses der experimentellen Methode bei der Aufgabe a35_Flugzeug eine Anforderung, da die

fokale Variablendimension (Höhenruder) manipuliert werden musste, während die beiden anderen Variablendimensionen (Flugzeugnase und -flügel) konstant zu halten waren. In der Teilkompetenz Dateninterpretation mussten die Kinder in den Aufgaben a64_Zähne und a66_Ravianer bei der Ableitung von Inferenzen aus den präsentierten Daten ebenfalls jeweils eine zweite Variablendimension mit einbeziehen, um die richtige Lösung zu erreichen. Aufgaben mit mehreren Variablendimensionen waren sowohl in der Teilkompetenz Experimentierstrategien als auch Dateninterpretation für die Kinder vergleichbar schwer. Außerdem bestätigten sich im Bereich der Dateninterpretation frühere Befunde der Kontextabhängigkeit von Aufgaben (z. B. Koerber et al., 2005): Kinder lösten häufiger Aufgaben in einem artifiziellen Kontext (Aufgabe a66_Ravianer), in dem keine Voreingenommenheiten aufgrund von Vorwissen oder intuitiven Theorien vorhanden waren, als in einem Kontext, in dem sie Annahmen über Kausalzusammenhänge hatten (Aufgabe a64_Zähne).

Daher sollte der sequenzielle Entwicklungsverlauf der beiden Teilkompetenzen Experimentierstrategien und Dateninterpretation nochmals mit einer größeren Anzahl an Aufgaben, in denen je Teilkompetenz sowohl nur eine als auch mehrere Variablendimensionen in Betracht gezogen werden müssen, überprüft werden, um einen Artefakt der Aufgaben auszuschließen.

Ausgehend von den dargestellten Voraussetzungsanalysen innerhalb der Kompetenz wissenschaftliches Denken scheinen bidirektionale Beziehungen zwischen den Teilkompetenzen Theorieverständnis, Verständnis von Experimentierstrategien und Verständnis der Evidenzevaluation zu bestehen. Eine hohe Fähigkeitsausprägung in einer der Teilkompetenzen bedingte scheinbar jeweils ein ausgeprägtes Verständnis in den anderen Teilkompetenzen. Diese Ergebnisse sind konsistent mit den in der LOGIC-Studie (Bullock et al., 2009) gefundenen wechselseitigen Beziehungen zwischen dem Verständnis von Experimenten und Theorien und der Strategieproduktion. Die längsschnittlichen Befunde wurden dahingehend diskutiert, dass das metakonzeptuelle Verständnis der Hypothesenprüfung die Verwendung von intentionalen Strategien zum Erreichen des Erkenntnisgewinns beeinflusst und die erfolgreiche Verwendung von Experimentierstrategien das metakonzeptuelle Verständnis der Hypothesentestung verbessert. Ein breites Verständnis der Theorie-Evidenz-Beziehung scheint eine bedeutsame Rolle bei der Entwicklung von Experimentierstrategien zu spielen. Auch in mikrogenetischen Studien zur selbstgesteuerten Exploration wurden bidirektionale Beziehungen gefunden (z. B. Kuhn et al., 1995; Kuhn & Pearl, 1998; Kuhn, 2002; Kuhn & Franklin, 2006; Schauble, Klopfer & Raghavan, 1991).

Daher werden metakognitive Reflexionsfähigkeiten, im Sinne eines Wissens über Ziele und Vorgehensweisen von Wissenschaften, der Unterscheidung von Theorie und Evidenz sowie der Inhibition eigener Vorstellungen und deren Einfluss auf die Interpretation von neuen Daten, als Erfolgsfaktoren für Experimentieren und Interpretieren von Daten erachtet.

Reliable Erfassung der kognitiven Fähigkeiten und motivationalen Orientierungen mittels schriftlicher Gruppentests im Grundschulalter

Für den Subtest Textverständnis des standardisierten Multiple-Choice-Tests ELFE 1-6 Ein Leseverständnistest für Erst- bis Sechstklässler (Lenhard & Schneider, 2006) konnte gezeigt werden, dass die Items sich raschkonform verhielten und der Test durch das Rasch-Modell skalierbar war.

Für die beiden Untertests topologisches Schlussfolgern und Fortsetzen von Reihenfolgen der Intelligenzbatterie CFT 20-R (Weiß, 2006) galt ebenfalls das Rasch-Modell.

Um die kognitiven Fähigkeiten Problemlösen, Inhibition und räumliches Denken im gleichen Bearbeitungsmodus zu erheben, waren bereits bestehende Testverfahren modifiziert und an das geschlossene Antwortformat angepasst worden: Die Items der adaptierten Papierversion des Tests TL-D Turm von London (Tucha & Lange, 2004) ließen sich gut anhand des Rasch-Modells skalieren. Das galt auch für die schriftliche Form des FWIT Farbe-Wort-Interferenztest nach J. R. Stroop (Bäumler, 1985) zur Erfassung der Inhibitionsleistung. Für die an das schriftliche Format und das Grundschulalter adaptierten Aufgaben zur mentalen Rotation (Kerns & Berenbaum, 1991; Vandenberg & Kuse, 1978; Shepard & Metzler, 1971) konnte nachgewiesen werden, dass sie nach dem Rasch-Modell skaliert werden konnten. Lediglich der ins Deutsche übersetzte Arlin-Test (ATFR; Arlin, 1984) zur Messung formal-operatorischer Fähigkeiten zeigte weder nach klassischer noch nach probabilistischer Testtheorie befriedigende statistische Kennwerte.

Die Skalen zur Erfassung naturwissenschaftsbezogener Einstellungen und des Fähigkeitsselbstkonzepts (TIMSS 2007 / Bos et al., 2008) sowie des Interesses an forschenden Aktivitäten (Subskala des ICA-R Inventory of Childrens' Activities – Revised / Tracey & Ward, 1998) konnten gut anhand des Partial-Credit-Modells gemessen werden.

Zusammenfassend kann festgestellt werden, dass für bereits etablierte Instrumente die Gültigkeit des Rasch-Modells nachgewiesen wurde und reliable Instrumente im geschlossenen Antwortformat zur Erfassung von problemlösenden, inhibitorischen und räumlichen

Fähigkeiten im Grundschulalter entwickelt wurden. Dadurch konnten alle Variablen im gleichen Bearbeitungsmodus erfasst werden, wodurch Konfundierungen durch unterschiedliche Antwortformate auszuschließen waren.

Abgrenzung und Zusammenhang des wissenschaftlichen Denkens zu ausgewählten kognitiven Fähigkeiten

Ein weiteres Ziel der vorliegenden Arbeit bestand darin, den Zusammenhang zwischen wissenschaftlichem Denken und möglichen zugrunde liegenden kognitiven Fähigkeiten zu beleuchten. Neben den Referenzvariablen Leseverständnis und Intelligenz wurden die beiden exekutiven Funktionen Problemlösen und Inhibition sowie räumliches Denken und formal-operatorische Fähigkeiten betrachtet. Im Folgenden werden die Beziehungen zu den erfassten kognitiven Fähigkeiten zunächst einzeln diskutiert, bevor in den Ergebnissen der latenten hierarchischen Regressionsanalysen der Einfluss aller Prädiktoren zur Vorhersage der Kompetenz im wissenschaftlichen Denken im Grundschulalter gemeinsam betrachtet wird.

Einfluss von Leseverständnis

Die noch nicht ausgereifte Lesekompetenz der Grundschul Kinder sowie die Verwendung von Multiple-Choice-Tests sind Kritikpunkte der schriftlichen Kompetenzerfassung im Grundschulalter (z. B. Paris et al., 2006; Duit et al., 2001). Daher könnte bei einem schriftlichen Test zur Erfassung des wissenschaftlichen Denkens im Grundschulalter das Problem bestehen, dass weniger die Kompetenz wissenschaftliches Denken als eigenständiges Konstrukt gemessen wird, sondern vielmehr die Lesekompetenz der Kinder die Leistung beeinflusst.

Im Hinblick auf die Lesekompetenzen der Grundschüler ergaben sich mit Studien wie IGLU (Internationale Grundschul-Lese-Untersuchung / Bos et al., 2003) vergleichbare Ergebnisse: Es fand sich ein Leistungsvorteil der Mädchen in allen Klassenstufen sowie über alle Klassenstufen hinweg. Das Leseverständnis nahm signifikant von der zweiten zur vierten Klassenstufe zu.

Diesem Aspekt des Einflusses der Lesekompetenz auf die Bearbeitung der Aufgaben zum wissenschaftlichen Denken wurde nachgegangen: Die Variablen Leseverständnis und wissenschaftliches Denken wurden in einem eindimensionalen und zweidimensionalen Rasch-Modell skaliert, um die Kompetenz im wissenschaftlichen Denken von der Lese-

kompetenz abzugrenzen. Es ergab sich eine zweidimensionale Struktur mit moderaten bis hohen latenten Korrelationen zwischen den Dimensionen wissenschaftliches Denken und Leseverständnis, sowohl in der Gesamtstichprobe als auch im frühen und späten Grundschulalter. Darüber hinaus lagen die Interkorrelationen zwischen Wissenschaftsverständnis und Methodenkompetenz jeweils auf einem höheren Niveau als die Korrelationen der beiden Kompetenzkomponenten zu der Kontrollvariable Leseverständnis. Die Befunde deuteten auf separate Konstrukte und diskriminante Validität hin. Daraus kann geschlossen werden, dass der schriftliche Test wissenschaftliches Denken erfasste und nicht nur Leseverständnis. Lesekompetenz war allerdings durchaus ein wichtiger prädiktiver Faktor, der zur erfolgreichen Bearbeitung der schriftlichen Aufgaben zum wissenschaftlichen Denken beitrug.

Einfluss von Intelligenz

Generell spielen Intelligenzaspekte bei der Bearbeitung von Instrumenten zur Erfassung von Kompetenzen eine wichtige Rolle (Hartig & Klieme, 2006). Daher wurden die Begabungsaspekte topologisches Schlussfolgern und Fortsetzen von Reihenfolgen als Intelligenzmaß im Sinne einer Kontrollvariable erfasst, um die Kompetenz im wissenschaftlichen Denken von allgemeiner Intelligenz zu unterscheiden. Es wurde damit überprüft, inwiefern der schriftliche Test zur Erfassung des wissenschaftlichen Denkens im Grundschulalter die Kompetenz wissenschaftliches Denken als eigenständiges Konstrukt misst oder ob vielmehr allgemeine intellektuelle Fähigkeiten die Testleistung bestimmen.

Die Leistung im Intelligenztest stieg vom frühen zum späten Grundschulalter signifikant an, wobei sich Mädchen und Jungen nicht signifikant unterschieden.

Die Variablen Intelligenz und wissenschaftliches Denken wurden als eindimensionales und zweidimensionales Rasch-Modell skaliert, um die Kompetenz im wissenschaftlichen Denken von allgemeiner Intelligenz zu differenzieren. In den einzelnen Stichproben zeigte sich eine zweidimensionale Struktur mit moderaten bis hohen latenten Korrelationen zwischen den Dimensionen wissenschaftliches Denken und Intelligenz. Darüber hinaus waren die ermittelten Interkorrelationen zwischen Wissenschaftsverständnis und Methodenkompetenz jeweils höher als die Korrelationen der beiden Kompetenzkomponenten zu der Referenzvariable Intelligenz. Die Ergebnisse sprechen für zwei separate Konstrukte und können dahingehend interpretiert werden, dass der schriftliche Kompetenztest nicht nur allgemeine Intelligenzaspekte sondern wissenschaftliches Denken erfasste. Intelligenz war allerdings durchaus von prädiktiver Bedeutung und spielte bei der erfolgreichen Bearbei-

tung der schriftlichen Aufgaben zum wissenschaftlichen Denken wohl eine bedeutsame Rolle.

Zusammenhang zu Problemlösen

Das Scientific Discovery als Dual Search (SDDS)-Modell von Klahr (2000) beschreibt den wissenschaftlichen Erkenntnisprozess als Problemlösen. Darüber hinaus stellten Dunbar und Klahr (1989) die These auf, dass die Repräsentation des wissenschaftlichen Problemraums wichtig für wissenschaftliches Denken ist und die Schwierigkeiten von Kindern im wissenschaftlichen Denken daher herrühren, dass sie die Problemlösesituation im wissenschaftlichen Erkenntnisprozess nicht vollständig repräsentieren können. Deshalb wurde in der vorliegenden Studie der Zusammenhang zwischen wissenschaftlichem Denken und Problemlösekompetenz untersucht.

Hinsichtlich der Problemlösekompetenzen der Grundschul Kinder konnte generell festgestellt werden, dass die Kompetenz von der zweiten zur vierten Grundschulklasse signifikant anstieg und keine Geschlechterdifferenzen bestanden. Die Ergebnisse entsprechen den Befunden zu alterskorrelierten Kompetenzzuwächsen im Problemlösen (Müller & Zelazo, 2002).

Die Modellprüfung stellte heraus, dass das zweidimensionale Rasch-Modell, welches die Kompetenz wissenschaftliches Denken und Problemlösekompetenz als zwei separate Dimensionen abbildete, die Daten am besten erklärte. Erwartungsgemäß deckte die geschätzte latente Korrelation einen hohen Zusammenhang von wissenschaftlichem Denken zu planerisch-problemlösendem Denken auf. Dieser Befund ist konsistent mit den von Kwon und Lawson (2000) im Sekundarschulalter gefundenen Zusammenhängen zwischen Problemlösefähigkeiten und wissenschaftlichem Denken. Die ermittelten Interkorrelationen der Dimensionen des wissenschaftlichen Denkens lagen in der Gesamtstichprobe und in der Klassenstufe 4 jeweils auf einem höheren Niveau als die Korrelationen der Kompetenzkomponenten zur Problemlösekompetenz. Lediglich in Klassenstufe 2-3 fiel unter Kontrolle der Variablen Leseverständnis und Intelligenz die Korrelation zwischen Methodenkompetenz und Problemlösekompetenz höher aus. Dies lässt eine große Bedeutung des problemlösenden Denkens bei der erfolgreichen Bearbeitung von Aufgaben zur Methodenkompetenz vor allem im frühen Grundschulalter erkennen. Somit stellten problemlösendes Denken und wissenschaftliches Denken korrelierte, aber separate und eigenständige Konstrukte dar.

Zusammenhang zu Inhibition

Befunde aus Studien zur selbstgesteuerten Exploration komplexer Mikrowelten deckten die Fähigkeit zur Inhibition des Einflusses von eigenen Vorstellungen über Ursachefaktoren auf die Generierung von Hypothesen, Planung von Experimenten und auf die Interpretation von Daten als Erfolgsfaktor für wissenschaftliches Denken auf (Kuhn & Franklin, 2006). Auch die Studie von Kwon und Lawson (2000) fand einen Einfluss der Fähigkeit, aufgabenirrelevante Informationen zu inhibieren, auf die Leistung von Jugendlichen im Test zum wissenschaftlichen Denken (Kwon & Lawson, 2000). Daher wurde in der vorliegenden Studie der Frage nachgegangen, ob derartige Beziehungen zwischen Inhibitionsleistung und wissenschaftlichem Denken auch im Grundschulalter gefunden werden können.

Konsistent mit den Befunden über alterskorrelierte Leistungszuwächse (Zelazo & Müller, 2002) nahm die Inhibitionsfähigkeit von der zweiten zur vierten Grundschulklasse signifikant zu. Dabei erzielten die Mädchen bessere Leistungen als die Jungen.

Der Vergleich von ein- und zweidimensionalen Rasch-Modellen stellte die Konstrukte wissenschaftliches Denken und Inhibition als zwei separate Dimensionen heraus. Entgegen der Erwartung bestanden nur geringe bis keine Zusammenhänge zum wissenschaftlichen Denken, insbesondere nach Berücksichtigung des Leseverständnisses und der Intelligenz. Daneben zeigten sich deutlich höhere Interkorrelationen der Komponenten Wissenschaftsverständnis und Methodenkompetenz im Vergleich zum Ausmaß der Korrelation der Komponenten zur Inhibitionsleistung. Entgegen der Befunde im Sekundarschulalter (Kwon & Lawson, 2000) hatten in der vorliegenden Studie Inhibitionsfähigkeiten im Grundschulalter geringe bis keine Effekte auf die Leistung im wissenschaftlichen Denken.

Zusammenhang zum räumlichem Denken

Die Beziehung zwischen räumlichem und wissenschaftlichem Denken wurde untersucht, da räumliches Denken für wissenschaftliches Denken, insbesondere im Hinblick auf die Fähigkeit zur Interpretation von Graphen, als notwendig erachtet wird (Koerber & Sodian, 2009; Koerber & Sodian, 2008; Mathewson, 1999). Denn in diesen Repräsentationsformen müssen Relationen zwischen Variablen auf räumliche Relationen abgebildet werden.

Gemäß den berichteten alterskorrelierten Unterschieden (Liben, 2002) konnte hinsichtlich der räumlichen Fähigkeiten ein signifikanter Unterschied zwischen frühem und spätem Grundschulalter festgestellt werden. Der gefundene Geschlechterunterschied zugunsten der

Jungen ist konsistent mit den Ergebnissen aus Studien zu Geschlechterdifferenzen im räumlichen Denken (Kerns & Berenbaum, 1991; Lohaus et al., 1999).

Aufgrund der moderaten latenten Zusammenhänge zwischen wissenschaftlichem Denken und räumlichen Fähigkeiten wurde das zweidimensionale Rasch-Modell als geeigneter zur Beschreibung der Daten erachtet. Auch die latenten Korrelationen des räumlichen Denkens zu den beiden Kompetenzkomponenten des wissenschaftlichen Denkens fielen geringer aus als die Interkorrelation zwischen Wissenschaftsverständnis und Methodenkompetenz. Folglich stellten wissenschaftliches Denken und räumliche Fähigkeiten separate Konstrukte dar.

Die Befunde bestätigen die Annahme, dass räumliche Fähigkeiten bedeutsam sind für Leistungen im wissenschaftlichen und technischen Bereich sowie im wissenschaftlichen Denken (Koerber & Sodian, 2009; Mathewson, 1999; Shea et al., 2001).

Zusammenhang zu formal-operatorischen Fähigkeiten

Formal-operatorische Fähigkeiten wurden erhoben, um die These von Inhelder und Piaget (1958) zu überprüfen, dass die Fähigkeit zu formalen Operationen eine Voraussetzung für das Verständnis der experimentellen Methodik sei.

Die Ergebnisse zu formal-operatorischen Fähigkeiten müssen mit Vorsicht interpretiert werden, da bei der Erfassung des formal-operatorischen Denkens mithilfe des Arlin-Tests (ATFR, Arlin, 1984) Varianz- und Reliabilitätsprobleme bestanden. Daher sollten diese Einschränkungen bei der Beurteilung der Befunde berücksichtigt werden. Der Test war für die Viertklässler zu schwer. Die Mehrheit der Viertklässler wurde der Stufe der konkreten Operationen zugeordnet. Entgegen Befunden im Sekundarschulalter (Bullock et al., 2009) unterschieden sich Mädchen und Jungen nicht signifikant in ihrer Leistung.

Aufgrund der moderaten bis hohen latenten Korrelation zwischen wissenschaftlichem Denken und formal-operatorischen Fähigkeiten, die unter Kontrolle von Leseverständnis und Intelligenz verschwand, war von zwei separaten Dimensionen auszugehen. Dies wurde durch die niedrigeren latenten Korrelationen der formal-operatorischen Fähigkeiten zu Wissenschaftsverständnis und Methodenkompetenz, insbesondere unter Berücksichtigung der Kontrollvariablen, im Vergleich zur Höhe der latenten Interkorrelationen der Kompetenzkomponenten des wissenschaftlichen Denkens untereinander bestätigt. Die Hypothese über einen positiven Zusammenhang zwischen wissenschaftlichem Denken und formal-operatorischen Fähigkeiten konnte nur teilweise belegt werden, da entgegen den Befunden

im Jugendalter (Bullock et al., 2009; Bullock & Ziegler, 1994) unter Kontrolle von Intelligenz und Leseverständnis kein substantieller Zusammenhang mehr bestand.

Voraussetzungsanalysen ergaben, dass Kinder, die als formal-operatorisch klassifiziert wurden, bessere Leistungen im wissenschaftlichen Denken erzielten. Allerdings zeigte sich nur für die Komponente Methodenkompetenz ein signifikanter Unterschied zwischen Kindern auf dem konkreten und dem formalem Niveau.

Im Hinblick auf formal-operatorische Fähigkeiten ist zusammenfassend zu sagen, dass in der vorliegenden Arbeit gezeigt werden konnte, dass entgegen der Annahme von Inhelder und Piaget (1958) wissenschaftliches Denken nicht erst im Jugendalter mit dem formal-operatorischen Stadium auftrat, sondern bereits Grundschul Kinder wissenschaftlich denken konnten. Kinder auf dem formal-operatorischen Niveau erzielten zwar durchaus bessere Leistungen in den Aufgaben zum wissenschaftlichen Denken, jedoch zeigten sie auch schon Kompetenzen im wissenschaftlichen Denken, bevor sie das Stadium der formalen Operationen erreichten. Die Fähigkeit zum wissenschaftlichen Denken scheint somit nicht notwendigerweise an die Entwicklung formal-operatorischer Fähigkeiten sensu Piaget gebunden zu sein. Ein Ansatz nach Piaget mit strukturellen Einschränkungen ist scheinbar zur Erklärung der Entwicklung des wissenschaftlichen Denkens weniger geeignet.

Zusammenhang der Kompetenz im wissenschaftlichen Denken zu ausgewählten motivationalen Orientierungen

Im Allgemeinen wird von wechselseitigen Beziehungen zwischen Kompetenzentwicklung und kompetenzbezogenen Einstellungen, Fähigkeitsselbstkonzept und Interesse ausgegangen (Weinert, 1999; Krapp et al., 1992).

Ein weiteres Ziel der vorliegenden Studie bestand darin, die Bedeutung motivationaler Variablen, wie Einstellungen zu Naturwissenschaften, naturwissenschaftsbezogenes Fähigkeitsselbstkonzept und Interesse an forschenden Aktivitäten, für die Kompetenz der Grundschul Kinder im wissenschaftlichen Denken zu untersuchen.

Zusammenhang zu naturwissenschaftsbezogenen Einstellungen

Konsistent mit den Ergebnissen von TIMSS 2007 (Wittwer et al., 2008) besaßen die Grundschul Kinder der vorliegenden Stichprobe hohe, positive Einstellungen zum Sachunterricht und zu Naturwissenschaften. Die Zweit-, Dritt- und Viertklässler unterschieden

sich nicht signifikant. Es trat ein tendenzieller Geschlechterunterschied dahingehend auf, dass Mädchen positivere Einstellungen aufwiesen als Jungen.

Bezogen auf die Kompetenz wissenschaftliches Denken konnte lediglich im frühen Grundschulalter ein substantieller positiver latenter Zusammenhang zu naturwissenschaftsbezogenen Einstellungen festgestellt werden. Die differenzierte Betrachtung zeigte, dass der Zusammenhang zum Wissenschaftsverständnis etwas höher ausfiel als zur Methodenkompetenz. Im späten Grundschulalter sowie in der Gesamtstichprobe der Grundschulkinder hatten die Einstellungen zu Naturwissenschaften keinen signifikanten Einfluss auf die Kompetenz im wissenschaftlichen Denken. Die in TIMSS 2007 gefundenen Zusammenhänge zwischen positiven Einstellungen zu Naturwissenschaften und naturwissenschaftlichen Kompetenz bei Grundschulkindern (Wittwer et al., 2008) konnten daher nur teilweise bestätigt werden.

Zusammenhang zum naturwissenschaftsbezogenem Fähigkeitsselbstkonzept

Die an dieser Studie teilgenommenen Grundschulkinder verfügten über ein hohes, positives naturwissenschaftsspezifisches Begabungsselbstkonzept. Dieses Ergebnis passt zu den Befunden von TIMSS 2007 (Wittwer et al., 2008). Während sich keine Altersunterschiede zeigten, trat ein signifikanter Geschlechtereffekt zugunsten der Mädchen auf.

Im Hinblick auf die Kompetenz im wissenschaftlichen Denken ergab sich sowohl in der Gesamtstichprobe als auch aufgegliedert in frühes und spätes Grundschulalter ein latenter Zusammenhang mittlerer Größe. Dieser Befund ist konsistent mit dem Ergebnis der TIMSS-Grundschulstudie, die für deutsche Grundschulkinder ebenfalls positive Zusammenhänge zwischen naturwissenschaftlicher Kompetenz und dem naturwissenschaftsbezogenem Begabungsselbstkonzept berichtete (Wittwer et al., 2008). Auch für die beiden Kompetenzkomponenten zeigten sich im Grundschulalter mittlere Zusammenhänge zum Fähigkeitsselbstkonzept, wobei der Effekt im frühen Grundschulalter größer hinsichtlich des Wissenschaftsverständnisses ausfiel, im späten Grundschulalter dagegen im Hinblick auf die Methodenkompetenzen.

Zusammenhang zu Interesse an forschenden Aktivitäten

Die Kinder der vorliegenden Stichprobe zeigten ein durchschnittliches Interesse an forschenden Aktivitäten. Während keine Altersunterschiede auftraten, unterschieden sich Mädchen und Jungen signifikant in ihrem Interesse an wissenschaftlichen Aktivitäten.

Konsistent mit den Befunden von Tracey (2002) bei Fünft- bis Achtklässlern zeigten die Jungen ein größeres Interesse an forschenden Aktivitäten.

Es bestand ein kleiner bis mittlerer latenter Zusammenhang zwischen wissenschaftlichem Denken und dem Interesse an forschenden Aktivitäten. Dieser Befund ist konsistent mit der im Sekundarschulalter gefundenen positiven Beziehung zwischen wissenschaftlichem Denken und naturwissenschaftlichem Fachinteresse sowie Interesse an naturwissenschaftsbezogenen Aktivitäten und Tätigkeiten (Glug, 2009). Während sich im frühen Grundschulalter für beide Kompetenzkomponenten mittlere Effekte zeigten, fanden sich im späten Grundschulalter und in der Stichprobe der gesamten Grundschul Kinder nur Zusammenhänge zur Methodenkompetenz.

Prognostizierbarkeit der Kompetenz im wissenschaftlichen Denken durch die erhobenen kognitiven und motivationalen Korrelate

In der vorliegenden Arbeit konnten im Grundschulalter (Klassenstufe 2 bis 4) anhand von latenten hierarchischen Regressionsanalysen das Alter sowie das Leseverständnis und die Intelligenz der Kinder als wichtige Determinanten zur erfolgreichen Bearbeitung der schriftlichen Aufgaben zum wissenschaftlichen Denken herausgestellt werden. Unter zusätzlicher Berücksichtigung der weiteren kognitiven und motivationalen Variablen erwiesen sich weiterhin das Alter und die Lesekompetenz der Kinder als bedeutsame Prädiktoren, sowie Problemlöse- und räumlichen Fähigkeiten als auch das naturwissenschaftsbezogene Fähigkeitsselbstkonzept.

Dass die Effekte von Intelligenz und Inhibition nach Berücksichtigung der weiteren kognitiven Variablen in der latenten multiplen Regressionsanalyse nicht mehr signifikant wurden, kann damit erklärt werden, dass der Multiple-Choice-Test nach dem Turm von London zur Erfassung der Problemlösekompetenz scheinbar ein globales Intelligenzmaß darstellte. Neben planerisch-problemlösenden Fähigkeiten erfasste er wohl auch allgemeine Intelligenzaspekte wie schlussfolgerndes Denken, das durch die beiden Untertests des CFT 20-R erhoben wurde, und inhibitorische Kontrolle. Dies ist in Einklang mit der Auffassung, dass die Leistung im Turm von London nicht nur abhängig ist von Planungsfähigkeiten, sondern auch von den Einflussgrößen Arbeitsgedächtnis, Inhibition und räumlichen Fähigkeiten (z. B. Renner, 2005). Auch die schriftlichen Problemlösetests der PISA-Studie erfassten zu einem beträchtlichen Teil schlussfolgerndes Denken (Klieme et al., 2001).

Die große prädiktive Bedeutung des Leseverständnisses passt zu dem Kritikpunkt, dass unausgereifte Lesekompetenzen die Bearbeitung von Multiple-Choice-Tests im Grundschulalter erschweren und Kinder mit einem ausgereifteren Sprach- und Leseverständnis bessere Leistungen erzielen. Allerdings konnte auch für Interviews zum wissenschaftlichen Denken ein Zusammenhang zum Leseverständnis gezeigt werden (Kropf, 2010). Daher scheinen zur Beantwortung von Aufgaben zum wissenschaftlichen Denken generell ein elaboriertes Textverständnis notwendig zu sein, um die Anforderungen der Aufgaben beider Bearbeitungsmodi begreifen zu können. Dies ist konsistent mit der allgemeinen Annahme, dass das Sprach- und Leseniveau eine entscheidende Rolle in der Leistungs- und Kompetenzentwicklung spielen (Schrader et al., 2008).

Eine herausragende prädiktive Rolle hatte ebenfalls die Problemlösekompetenz der Kinder. Dieser Befund bestätigt die Auffassung des Scientific Discovery as Dual Search (SDDS)-Modells (Klahr, 2000), das wissenschaftliches Denken als Problemlöseprozess beschreibt. Um die Aufgaben zum wissenschaftlichen Denken erfolgreich lösen zu können, scheint es notwendig, den Problemraum des wissenschaftlichen Erkenntnisproblems vollständig repräsentieren zu können (Dunbar & Klahr, 1989). Daher war es zu erwarten und konnte anhand der vorliegenden Arbeit für das Grundschulalter nachgewiesen werden, dass allgemeine Problemlösekompetenzen die Beantwortung der Aufgaben zum wissenschaftlichen Denken beeinflussten (vgl. Kwon & Lawson, 2000, für das Sekundarschulalter).

Fasst man die Ergebnisse zusammen, so können im Grundschulalter die kognitiven Fähigkeiten Leseverständnis, räumliches Denken und insbesondere Problemlösefähigkeiten, welche Intelligenzaspekte und inhibitorische Kontrolle beinhalten, wohl als Voraussetzungen für die erfolgreiche Bearbeitung der Testaufgaben bzw. für die Kompetenz im wissenschaftlichen Denken gelten. Auf motivationaler Ebene hatte das naturwissenschaftsbezogene Fähigkeitsselbstkonzept einen substantiellen Einfluss.

Die Ergebnisse der altersspezifisch für das frühe und späte Grundschulalter durchgeführten latenten hierarchischen Regressionsanalysen stimmten mit den allgemein für das Grundschulalter gefundenen Prädiktoren weitgehend überein. Sowohl im frühen als auch im späten Grundschulalter hatten das Leseverständnis und die Problemlösekompetenz eine starke prädiktive Bedeutung. Während im späten Grundschulalter die motivationalen Variablen keine wichtige Rolle bei der Kompetenzentwicklung zu spielen schienen, hatten die positiven Einstellungen zu Naturwissenschaften der Zweit- und Drittklässler einen substantiellen Einfluss auf die Kompetenz im wissenschaftlichen Denken. Die Effekte des natur-

wissenschaftsbezogenen Fähigkeitsselbstkonzepts und des Interesses an forschenden Aktivitäten waren im frühen Grundschulalter stärker ausgeprägt als am Ende der Grundschulzeit, verfehlten allerdings knapp die Signifikanzgrenze. Demnach tragen motivationale Aspekte im frühen Grundschulalter wohl zu einer häufigeren und intensiveren Auseinandersetzung mit naturwissenschaftlichen Inhalten und wissenschaftlichen Aktivitäten bei und beeinflussen scheinbar dadurch positiv die Kompetenzentwicklung im wissenschaftlichen Denken. Dagegen traten motivationale Variablen in der vierten Klassenstufe in den Hintergrund und lediglich kognitive Fähigkeiten hatten einen Einfluss auf die Kompetenz im wissenschaftlichen Denken.

Insgesamt bestätigte sich auch für die Kompetenz im wissenschaftlichen Denken der dominierende Einfluss kognitiver Fähigkeiten gegenüber motivationalen Bedingungsfaktoren (vgl. Weinert & Helmke, 1997), so dass kognitive Fähigkeiten zu einem größeren Anteil als motivationale Orientierungen zur Erklärung individueller Kompetenzunterschiede im wissenschaftlichen Denken beitrugen.

Grenzen der vorliegenden Studie

Aufgrund von geringen Itemzahlen je Komponente bzw. Teilkompetenz des wissenschaftlichen Denkens ($n < 15-20$) war eine reliable Erfassung der einzelnen Teilkompetenzen des wissenschaftlichen Denkens schwierig. Denn es sind mindestens 15 bis 20 Items je Skala notwendig, um das interessierende Konstrukt reliabel erfassen zu können (Adams et al., 1997; Wu et al., 2005). Das könnte dazu geführt haben, dass die Kompetenz wissenschaftliches Denken nur in einer Gesamtskala reliabel und als eindimensionales Konstrukt gemessen werden konnte.

Eine separate Betrachtung der zweiten und dritten Jahrgangsstufe war, sofern nicht auf Skalierungen der Gesamtstichprobe zurückgegriffen werden konnte, nicht möglich, da die Stichproben zu klein (Klassestufe 2: $n = 58$, Klassestufe 3: $n = 72$) und die Parameterschätzungen dadurch stark von Messfehlern beeinflusst waren. Für Rasch-Modelle sind Stichprobengrößen von ca. $n = 200$ für präzise Parameterschätzungen notwendig (Wright & Stone, 1979). Daher wurden in der vorliegenden Studie die Klassenstufe 2 und 3 zu einer Stichprobe des frühen Grundschulalters ($n = 130$) zusammengefasst. Die IRT-Analysen der Gesamtstichprobe der Grundschulkinder ($n = 285$), des frühen Grundschulalters und der Klassenstufe 4 als spätes Grundschulalters ($n = 155$) können allerdings als präzise bewertet werden.

Da es sich um eine angefallene Stichprobe von Kindern aus Grundschulen im Raum München handelte, sind Generalisierung der Ergebnisse auf die Gesamtpopulation der Grundschulkinder nur mit Einschränkungen möglich und müssen kritisch betrachtet werden. Jedoch gibt es aus der Querschnittstudie des Gesamtprojektes Science-P mit 1916 Schülern Hinweise auf vergleichbare Ergebnisse hinsichtlich der Kompetenz wissenschaftliches Denken für Schüler aus Bayern und Nordrhein-Westfalen, so dass grundsätzlich von generalisierbaren Befunden auszugehen ist (Möller et al., 2011).

Des Weiteren untersuchte die vorliegende Studie die Entwicklung des wissenschaftlichen Denkens lediglich in einem querschnittlichen Design. Zur besseren Beschreibung der Entwicklungsveränderungen im wissenschaftlichen Denken und Entwicklungszusammenhänge mit kognitiven und motivationalen Variablen sollte zusätzlich eine Längsschnittstudie mit größeren Zahlen an Kindern und Items durchgeführt werden.

Außerdem fanden Variablen auf Seite der Eltern in der vorliegenden Studie keine Beachtung, da der Fokus auf der Untersuchung der individuellen Bedingungsfaktoren, die zur Kompetenzentwicklung im wissenschaftlichen Denken beitragen, lag. In der Regel stellten Schulleistungsstudien einen engen Zusammenhang zwischen dem sozioökonomischen Status der Eltern und dem Kompetenzniveau der Grundschulkinder fest (z. B. Schwippert, Bos & Lankes, 2004). Auch in der Validierungsstudie und der Querschnittstudie des Projektes Science-P zeigten sich signifikante Effekte des sozioökonomischen Status, gemessen über den höchsten Bildungsabschluss und den Beruf bzw. die Arbeit der Eltern, auf die Kompetenz im wissenschaftlichen Denken der Kinder (Kropf, 2010).

Für die Kompetenz wissenschaftliches Denken gibt es bisher keine etablierten Curricula in der Grundschule. Daher wird wissenschaftliches Denken nicht explizit im Unterricht gefördert. Dennoch ist anzunehmen, dass im Rahmen des naturwissenschaftlichen Unterrichts das Wissenschaftsverständnis und die Methodenkompetenzen der Lehrkräfte die Kompetenzentwicklung der Kinder beeinflussen (vgl. Lederman, 1992). Instrumente zur Erfassung des Wissens über Naturwissenschaften von Lehrkräften sind im Rahmen des Projektes Science-P in Entwicklung, um diesem Aspekt nachzugehen. Erste Ergebnisse deuten auf einen förderlichen Einfluss eines reflektierten Wissens über Naturwissenschaften der Lehrkräfte auf die Kompetenzentwicklung der Kinder im wissenschaftlichen Denken hin (Osterhaus, Koerber, Mayer & Sodian, 2011).

Ein weiterer möglicher Einflussfaktor auf die Entwicklung wissenschaftlichen Denkens, der in der vorliegenden Studie keine Berücksichtigung fand, sind außerschulische Angebote. In naturwissenschaftlichen Programmen werden Methodenkompetenzen und

das Verständnis von Wissenschaften zwar meist nicht explizit unterrichtet, allerdings könnten die Programme durch die Auseinandersetzung mit naturwissenschaftlichen Phänomenen und Experimenten einen Effekt auf die Kompetenzentwicklung der Kinder im wissenschaftlichen Denken haben.

Fazit

In der vorliegenden Arbeit konnte gezeigt werden, dass es dem Forschungsprojekt Science-P erfolgreich gelungen ist, ein reliables gruppentestfähiges Messinstrument zur Erfassung der Kompetenz wissenschaftliches Denken für das Grundschulalter zu entwickeln. Dabei bestanden hohe Interkorrelationen zwischen den Komponenten bzw. Teilkompetenzen des wissenschaftlichen Denkens. Die Kompetenz wissenschaftliches Denken ließ sich am besten als eindimensionales Konstrukt abbilden und scheint daher ein kohärentes, einheitliches Konstrukt darzustellen.

Es konnten bereits im Grundschulalter Kompetenzen im wissenschaftlichen Denken nachgewiesen werden. Daher ist die Fähigkeit zum wissenschaftlichen Denken wohl nicht notwendigerweise an die Entwicklung formal-operatorischer Fähigkeiten sensu Piaget gebunden und wird nicht erst im Jugendalter erworben.

Zwischen den Grundschulkindern traten signifikante Leistungsunterschiede dahingehend auf, dass in der zweiten Grundschulklasse noch mehrheitlich Fehlvorstellungen bestanden, während im späten Grundschulalter schon wissenschaftlich adäquate Vorstellungen vorherrschten. Die Kompetenz im wissenschaftlichen Denken nahm signifikant von der zweiten zur vierten Klassenstufe zu. In der dritten Jahrgangsstufe traten scheinbar markante Veränderungen in der Entwicklung des wissenschaftlichen Denkens auf, so dass die Wahrscheinlichkeit für den Wechsel von der inkompetenteren in die kompetentere latente Klasse anstieg.

Darüber hinaus wurde ein sequenzieller Entwicklungsverlauf innerhalb der Teilkompetenzen des wissenschaftlichen Denkens festgestellt. Das Verständnis der Konzepte Hypothese, Theorie, Experiment und der Ziele von Wissenschaften schien sich als erstes zu entwickeln, gefolgt von dem Verständnis der experimentellen Methode sowie der Interpretation von Daten. Ein Theorieverständnis, im Sinne eines metakonzeptuellen Verständnisses alternativer Interpretationsrahmen, entwickelte sich zuletzt.

Zwischen den Teilkompetenzen metakonzeptuelles Theorieverständnis, Experimentierstrategien und Dateninterpretation bestanden bidirektionale Beziehungen.

Des Weiteren ist die diskriminante Validität hervorzuheben, so dass wissenschaftliches Denken als separates, eigenständiges Konstrukt gemessen werden konnte. Leseverständnis und Intelligenz hatten zwar einen Einfluss auf die erfolgreiche Bearbeitung der Aufgaben, allerdings konnte gezeigt werden, dass die im schriftlichen Testverfahren gemessene Kompetenz im wissenschaftlichen Denken durchaus eine separate Dimension darstellte.

Außerdem wurde ein substantieller Einfluss der angenommenen zugrundeliegenden kognitiven Fähigkeiten Problemlösen und räumliches Denken, auch unter Kontrolle von Leseverständnis und Intelligenz, auf die Kompetenz im wissenschaftlichen Denken festgestellt. Unter Berücksichtigung der Kontrollvariablen zeigten sich nur geringe Zusammenhänge zu formal-operatorischen und inhibitorischen Fähigkeiten.

Besonders im frühen Grundschulalter bestanden positive Zusammenhänge zwischen der Kompetenz im wissenschaftlichen Denken und den motivationalen Variablen naturwissenschaftsbezogene Einstellungen, Fähigkeitsselbstkonzept und Interesse an forschenden Aktivitäten.

6 Ausblick

Die neuere entwicklungspsychologische Forschung zum wissenschaftlichen Denken (siehe Zimmerman, 2007, für einen Überblick) sowie die vorliegende Arbeit konnten bereits im Grundschulalter ein beginnendes Verständnis von Wissenschaften und Theorien, der experimentellen Methodik sowie der Evaluation von Daten zeigen.

Es sei hier aber erwähnt, dass Basisfähigkeiten im formal-wissenschaftlichen Denken zwar schon im Grundschulalter vorhanden sind, jedoch markante Entwicklungen zwischen dem Grundschul- und Jugendalter und noch bis ins Erwachsenenalter stattfinden: Entwicklungsfortschritte treten beispielsweise im Hinblick auf die spontane Verwendung der Variablenkontrollstrategie zur Produktion kontrollierter Experimente auf (Bullock et al., 2009). Des Weiteren zeigen sich Veränderungen in der Interpretation von komplexen Kovariationsmustern in multivariaten Aufgaben (Kuhn et al., 1988) sowie in der Berücksichtigung und Koordination der Effekte mehrerer Kausalfaktoren (Kuhn & Dean, 2004; Kuhn, Iordanou, Pease & Wirkala, 2008; Kuhn, Pease & Wirkala, 2009). Auch hinsichtlich des expliziten, deklarativen Wissenschaftsverständnisses und des metakonzeptuellen Theorieverständnisses kommt es zu Verbesserungen (Thoermer & Sodian, 2002; Bullock et al., 2009).

Interventionsstudien belegten, dass schon bei Grundschulkindern durch einfache Trainings der Variablenkontrollstrategie die Fähigkeit, unkonfundierte Experimente zu produzieren und valide Schlussfolgerungen zu ziehen, verbessert werden konnte (Bullock & Ziegler, 1999; Chen & Klahr, 1999; Klahr & Nigam, 2004). Ein explizit konstruktivistisches und wissenschaftstheoretisches Unterrichtscurriculum in der vierten Klassenstufe der Grundschule hatte positive Effekte sowohl auf das Wissenschaftsverständnis, im Sinne einer konstruktivistischen Vorstellung wissenschaftlicher Erkenntnis, als auch auf die Produktion der Variablenkontrollstrategie, obwohl das Curriculum kein Strategietraining beinhaltete (Sodian et al., 2006). Dabei traten sowohl kurzfristige Effekte als auch Langzeiteffekte in follow-up-Messungen nach einem Jahr auf. Smith und Kollegen (2000) berichteten ebenfalls positive Auswirkungen eines langfristigen, wissenschaftstheoretisch reflektierten Curriculums auf das Wissenschaftsverständnis der Schüler der ersten bis zur sechsten Jahrgangsstufe.

Des Weiteren werden positive Zusammenhänge zwischen wissenschaftlichem Denken und inhaltsbezogenem naturwissenschaftlichem Wissen angenommen (Zimmermann, 2007). Grygier (2008) fand für das Grundschulalter positive Effekte eines reflektierten Wissenschaftsverständnisses auf den Erwerb fachinhaltlichen, naturwissenschaftlichen Wissens. Auch wirkte sich ein gezieltes Training der bereichsübergreifenden Variablenkontrollstrategie bei amerikanischen Grundschulkindern positiv auf den Erwerb domänen-spezifischen naturwissenschaftlichen Inhaltswissens aus (Chen & Klahr, 1999).

Die dargestellten Studien zeigten auf, dass Kompetenzen im wissenschaftlichen Denken nicht erst, wie bisher häufig angenommen, im Sekundarschulalter bestehen und vermittelt werden können, sondern bereits im Grundschulalter. Daher besteht die Notwendigkeit formal-wissenschaftliches Denken als domänenübergreifende Kompetenz bereits in der Grundschule in die Lehrpläne des Sachunterrichts zu integrieren. Eine erfolgreiche Implementierung von Curricula zum wissenschaftlichen Denken setzt eine gezielte Ausbildung der Grundschullehrkräfte im Hinblick auf ein elaboriertes und reflektiertes Wissenschaftsverständnis sowie ausgeprägte Methodenkompetenzen voraus. Denn Interviewstudien stellten fest, dass Lehrkräfte in der Primarstufe noch nicht im ausreichenden Maße ein Verständnis der Konstruktion wissenschaftlicher Erkenntnis und wissenschaftlichen Wissens besitzen (Günther, Grygier, Kircher, Sodian & Thoermer, 2004). Generell wird davon ausgegangen, dass Lehrermerkmale das Wissen und den Lernerfolg von Schülern beeinflussen (Helmke, 2003). Ein adäquates Verständnis von Wissenschaft und wissenschaftlicher Denk- und Arbeitsweisen der Lehrkräfte wird als Bedingung für eine erfolgreiche Vermittlung einer konstruktivistischen Vorstellung von Wissenschaften und naturwissenschaftlichen Wissens gesehen (Grygier, 2008). Im Bereich des Wissens über Naturwissenschaften gibt es dazu derzeit nur wenige Studien. Dogan und Abd-El-Khalick (2008) fanden in der Sekundarstufe einen Zusammenhang zwischen dem Wissenschaftsverständnis von Lehrkräften und den wissenschaftstheoretischen Überzeugungen ihrer Schüler. Auch im Primarbereich gibt es Hinweise auf positive Effekte des Wissens über Naturwissenschaften der Lehrkräfte auf die Kompetenz im wissenschaftlichen Denken ihrer Schüler (Osterhaus et al., 2011). Daher ist eine Förderung sowohl des Wissenschaftsverständnisses, der Methodenkompetenzen und des fachspezifisch-pädagogischen Wissens der Lehrkräfte notwendig, damit wissenschaftliches Denken als wichtiger Aspekt der scientific literacy gezielt und effektiv in der Grundschule vermittelt werden kann.

7 Literaturverzeichnis

- Adams, R. (2002). Scaling PISA cognitive data. In R. Adams & M. Wu (Eds.), *PISA 2000 technical report* (pp. 99-108). Paris: OECD.
- Adams, R. J., Wilson, M. R. & Wang, W.-C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, 21, 1-23.
- American Association for the Advancement of Science (Eds.) (1993). *Benchmarks for science literacy. Project 2061*. New York: Oxford University Press.
- Amsel, E. & Brock, S. (1996). The development of evidence evaluation skills. *Cognitive Development*, 11, 523–550.
- Arlin, P. K. (1984). *Arlin Test of Formal Reasoning (AFTR). Test manual for middle school, high school, and adult levels*. New York: Slossen Publications.
- Baumert, J., Artelt, C., Klieme, E. & Stanat, P. (2001). PISA – Programme for International Student Assessment. Zielsetzung, theoretische Konzeption und Entwicklung von Messverfahren. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 285-310). Weinheim: Beltz.
- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J. & Weiß, M. (Hrsg.) (2001). *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.
- Baumert, J., Lehmann, R. H., Lehrke, M., Schmitz, B., Clausen, M., Hosenfeld, I., Köller, O. & Neubrand, J. (1997). *TIMSS - Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich: Deskriptive Befunde*. Opladen: Leske + Budrich.
- Bäumler, G. (1985). *Farbe-Wort-Interferenztest nach J. R. Stroop (FWIT)*. Göttingen: Hogrefe.
- Bortz, J. (1999). *Statistik für Sozialwissenschaftler* (5. Aufl.). Heidelberg: Springer.

- Bos, W., Bensen, M., Baumert, J., Prenzel, M., Selter, C. & Walther, G. (Hrsg.) (2008). *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Walther, G. & Valtin, R. (Hrsg.) (2003). *Erste Ergebnisse aus IGLU – Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*. Münster: Waxmann.
- Bühner, M. (2004). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson.
- Bühner, M. & Ziegler, M. (2009). *Statistik für Psychologen und Sozialwissenschaftler*. München: Pearson.
- Bullock, M., Sodian, B. & Koerber, S. (2009). Doing experiments and understanding science: Development of scientific reasoning from childhood to adulthood. In W. Schneider & M. Bullock (Eds.), *Human development from early childhood to early adulthood. Findings from the Munich Longitudinal Study* (pp. 173-197). Mahwah, NJ: Erlbaum.
- Bullock, M. & Ziegler, A. (1994). Scientific thinking. In F. E. Weinert & W. Schneider (Eds.), *The Munich Longitudinal Study on the Genesis of Individual Competencies (LOGIC): Report No. 11* (pp. 56-76). Munich: Max Planck Institute for Psychological Research.
- Bullock, M. & Ziegler, A. (1999). Scientific reasoning: Developmental and individual differences. In F. E. Weinert & W. Schneider (Eds.), *Individual development from 3 to 12. Findings from the Munich Longitudinal Study* (pp. 38-60). Cambridge: Cambridge University Press.
- Bullock, M., Ziegler, A. & Martin, S. (1993). Scientific thinking. In F. E. Weinert & W. Schneider (Eds.), *The Munich Longitudinal Study on the Genesis of Individual Competencies (LOGIC): Report No. 9* (pp. 66-110). Munich: Max Planck Institute for Psychological Research.
- Bybee, R. W. (1997). Toward an understanding of scientific literacy. In W. Gräber & C. Bolte (Eds.), *Scientific Literacy* (pp. 37-68). Kiel: IPN.
- Bybee, R. W. (2002). Scientific Literacy – Mythos oder Realität? In W. Gräber, P. Nentwig, T. Koballa & R. Evans (Hrsg.), *Scientific Literacy. Der Beitrag der Naturwissenschaften zur Allgemeinen Bildung*. (S. 21-43). Opladen: Leske + Budrich.

- Carey, S., Evans, R., Honda, M., Jay, E. & Unger, C. (1989). An experiment is when you try it and see if it works. A study of junior high school students' understanding of the construction of scientific knowledge. *International Journal of Science Education*, 11, 514-529.
- Carey, S. & Smith, C. (1993). On understanding the nature of scientific knowledge. *Educational Psychologist*, 28 (3), 235-251.
- Chen, Z. & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development*, 70 (5), 1098-1120.
- diSessa, A. (2006). A history of conceptual change research. In K. Sawyer (Eds.), *The Cambridge Handbook of the Learning Sciences* (pp. 265-281). Cambridge: Cambridge University Press.
- Dogan, N. & Abd-el-Khalick, F. (2008). Turkish grade 10 students' and science teachers' conceptions of nature of science: A national study. *Journal of Research in Science Teaching*, 45, 1083-1112.
- Driver, R., Leach, J., Millar, R. & Scott, P. (1996). Young people's images of science. Bristol: Open University Press.
- Duit, R., Häußler, P. & Prenzel, M. (2001). Schulleistungen im Bereich der naturwissenschaftlichen Bildung. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 169-185). Weinheim: Beltz.
- Dunbar, K. & Klahr, D. (1989). Developmental differences in scientific discovery processes. In D. Klahr & K. Kotovsky (eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 109-143). Hillsdale, NJ: Erlbaum.
- Embretson, S. E. & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.
- Faulstich-Wieland, H. (2004). *Mädchen und Naturwissenschaften in der Schule. Expertise für das Landesinstitut für Lehrerbildung und Schulentwicklung*. Hamburg.
- Field, A. P. (2009). *Discovering statistics using SPSS (and sex and drugs and rock'n'roll)* (3. Aufl.). Los Angeles: Sage.
- Fraser, B. (1980). Development and validation of a test of enquiry skills. *Journal of Research in Science Teaching*, 17 (1), 7-16.

- Gattis, M. (2002). Structure mapping in spatial reasoning. *Cognitive Development*, 17, 1157–1183.
- Gesellschaft für Didaktik des Sachunterrichts (2002). *Perspektivenrahmen Sachunterricht*. Bad Heilbrunn: Klinkhardt.
- Glug, I. (2009). *Entwicklung und Validierung eines Multiple-Choice-Tests zur Erfassung prozessbezogener naturwissenschaftlicher Grundbildung*. Dissertation: IPN Kiel.
- Gollwitzer, M. (2008). Latent-Class-Analysis. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 279–306). Heidelberg: Springer.
- Green, B. F. (1956). A method of scalogram analysis using summary statistics. *Psychometrika*, 62, 98-110.
- Grygier, P. (2008). *Wissenschaftsverständnis von Grundschulern im Sachunterricht*. Bad Heilbrunn: Klinkhardt.
- Günther, J., Grygier, P., Kircher, E., Sodian, B. & Thoermer, C. (2004). Studien zum Wissenschaftsverständnis von Grundschullehrkräften. In J. Doll & M. Prenzel (Hrsg.), *Bildungsqualität von Schule: Lehrerprofessionalisierung, Unterrichtsentwicklung und Schülerförderung als Strategien der Qualitätsverbesserung* (S. 93-113). Münster: Waxmann.
- Hammann, M. (2004). Kompetenzentwicklungsmodelle. Merkmale und ihre Bedeutung – dargestellt anhand von Kompetenzen beim Experimentieren. *MNU*, 57 (4), 196-203.
- Hammann, M., Phan, T. T. H. & Bayrhuber, H. (2007). Experimentieren als Problemlösen: Lässt sich das SDDS-Modell nutzen, um unterschiedliche Dimensionen beim Experimentieren zu messen? *Zeitschrift für Erziehungswissenschaft, Sonderheft 8*, 33-49.
- Häcker, O. & Stapf, K.-H. (2004) (Hrsg.). *Dorsch Psychologisches Wörterbuch*. Bern: Huber.
- Hardy, I., Kleickmann, T., Koerber, S., Mayer, D., Möller, K., Pollmeier, J., Schwippert, K. & Sodian, B. (2010). Die Modellierung naturwissenschaftlicher Kompetenz im Grundschulalter. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes. 56. Beiheft der Zeitschrift für Pädagogik* (S. 115–125). Weinheim: Beltz.

- Hartig, J. & Höhler, J. (2008). Representation of competencies in multidimensional IRT models with within- and between-item multidimensionality. *Zeitschrift für Psychologie*, 216, 88–100.
- Hartig, J. & Klieme, E. (2006). Kompetenz und Kompetenzdiagnostik. In K. Schweizer (Hrsg.), *Leistung und Leistungsdiagnostik* (S. 127–143). Heidelberg: Springer.
- Hartinger, A. & Fölling-Albers, M. (2002). *Schüler motivieren und interessieren. Ergebnisse aus der Forschung – Anregungen für die Praxis*. Bad Heilbrunn: Klinkhardt.
- Helmke, A. (2003). *Unterrichtsqualität - erfassen, bewerten, verbessern*. Seelze: Kallmeyersche Verlagsbuchhandlung.
- Holland, J. L. (1973). *Making vocational choices: A theory of careers*. Englewood Cliffs, NJ: Prentice Hall.
- Holland, J. L. (1985). *Making vocational choices: A theory of vocational personalities and work environments* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Inhelder, B. & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.
- Kelava, A. & Moosbrugger, H. (2008). Deskriptiv statistische Evaluation von Items (Itemanalyse) und Testwertverteilungen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 73–98). Heidelberg: Springer.
- Kerns, K. A. & Berenbaum, S. A. (1991). Differences in spatial ability in children. *Behavior Genetics*, 21 (4), 383-396.
- Klahr, D. (2000). *Exploring science*. Cambridge, MA: MIT Press.
- Klahr, D. & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, 12 (1), 1-48.
- Klahr, D., Fay, A. L. & Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology*, 25 (1), 111-146.
- Klahr, D. & Nigam, M. (2004). The equivalence of learning paths in early science instruction. *Psychological Science*, 15 (10), 661-667.
- 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, 179-200.

- Klopfer, L. & Cooley, W. (1961). *Test on understanding science, Form W*. Princeton, NJ: educational Testing Service.
- Koepfen, K., Hartig, J., Klieme, E. & Leutner, D. (2008). Current issues in competence modeling and assessment. *Zeitschrift für Psychologie - Journal of Psychology*, 216 (2), 61-73.
- Koerber, S. (2006). Entwicklung des wissenschaftlichen Denkens bei Vier- bis Achtjährigen. *Beiträge zur Lehrerbildung*, 24 (2), 193-201.
- Koerber, S. (2011). Der Umgang mit visuell-grafischen Repräsentationen im Grundschulalter. *Unterrichtswissenschaft*, 39 (1), 49-62.
- Koerber, S., Kropf, N., Mayer, D., Sodian, B. & Schwippert, K. (2009). Entwicklung naturwissenschaftlicher Kompetenz in der Grundschule (Science-P): Wissen über Naturwissenschaften. In C. Röhner, M. Hopf & C. Henrichwark (Hrsg.), *Europäisierung der Bildung. Konsequenzen und Herausforderungen für die Grundschulpädagogik, Jahrbuch Grundschulforschung Band 13* (S. 194-198). Wiesbaden: Verlag für Sozialwissenschaften.
- Koerber, S., Sodian, B., Kropf, N., Mayer, D. & Schwippert, K. (2011). Die Entwicklung des wissenschaftlichen Denkens im Grundschulalter. Theorieverständnis, Experimentierstrategien, Dateninterpretation. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 43 (1), 16-21.
- Koerber, S. & Sodian, B. (2008). Preschool children's ability to visually represent relations. *Developmental Science*, 11 (3), 390-395.
- Koerber, S. & Sodian, B. (2009). Reasoning from graphs in young children: Preschoolers' ability to interpret covariation data from graphs. *Journal of Psychology of Science & Technology*, 2 (2), 73-86.
- Koerber, S., Sodian, B., Thoermer, C. & Nett, U. (2005). Scientific reasoning in young children: Preschoolers' ability to evaluate covariation evidence. *Swiss Journal of Psychology*, 64 (3), 141-152.
- Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning*. Cambridge, MA: MIT Press.
- Krapp, A. (1998). Entwicklung und Förderung von Interessen im Unterricht. *Psychologie in Erziehung und Unterricht*, 45, 186-203.

- Krapp, A., Hidi, S. & Renninger, K. A. (1992). Interest, learning, and development. In K. A. Renninger, S. Hidi & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 3–25). Hillsdale, NJ: Erlbaum.
- Krikorian, R., Bartok, J. & Gay, N. (1994). Tower of london procedure: A standard method and developmental data. *Journal of Clinical and Experimental Neuropsychology*, 16 (6), 840-850.
- Kropf, N. (2010). *Entwicklung und Analyse von Messinstrumenten zur Erfassung des wissenschaftlichen Denkens im Grundschulalter*. Dissertation: Ludwig-Maximilians-Universität München.
- Kuhn, D. (1995). Microgenetic study of change: What Has It Told Us? *Psychological Science*, 6 (3), 133–139.
- Kuhn, D. (2001). How Do People Know? *Psychological Science*, 12 (1), 1-8.
- Kuhn, D. (2002). What is scientific thinking and how does it develop? In U. Goswami (Eds.), *Handbook of childhood cognitive development* (pp. 371-393). Oxford: Blackwell.
- Kuhn, D., Amsel, E. & O'Loughlin, M. (1988). *The development of scientific thinking skills*. Orlando, FL: Academic Press.
- Kuhn, D. & Dean, D. (2004). Connecting scientific reasoning and causal inference. *Journal of Cognition and Development*, 5 (2), 261–288.
- Kuhn, D. & Franklin, S. (2006). The second decade. What develops (and how). In D. Kuhn & R. S. Siegler (Vol. Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception and language* (pp. 953-993). Hoboken, NJ: Wiley.
- Kuhn, D., Garcia-Mila, M., Zohar, A. & Andersen, C. (1995). Strategies of knowledge acquisition. *Monographs of the society for research in child development*, 60, 1-128.
- Kuhn, D., Iordanou, K., Pease, M. & Wirkala, C. (2008). Beyond control of variables: What needs to develop to achieve skilled scientific thinking? *Cognitive Development*, 23 (4), 435–451.
- Kuhn, D. & Pearsall, S. (1998). Relations between metastrategic knowledge and strategic performance. *Cognitive Development*, 13 (2), 227–247.
- Kuhn, D. & Pearsall, S. (2000). Developmental origins of scientific thinking. *Journal of Cognition and Development*, 1, 113–129.

- Kuhn, D., Pease, M. & Wirkala, C. (2009). Coordinating the effects of multiple variables: A skill fundamental to scientific thinking. *Journal of Experimental Child Psychology*, 103 (3), 268–284.
- Kuhn, D., Schauble, L. & Garcia-Mila, M. (1992). Cross-domain development of scientific reasoning. *Cognition and Instruction*, 9, 285-327.
- Kwon, Y.-J. & Lawson, A. E. (2000). Linking brain growth with the development of scientific reasoning ability and conceptual change during adolescence. *Journal of Research in Science Teaching*, 37 (1), 44-62.
- Lederman, N. G. (1992). Students' and teachers' conceptions of the nature of science: A review of the research. *Journal of Research in Science Teaching*, 29 (4), 331-359.
- Lenhard, W. & Schneider, W. (2006). *ELFE 1-6. Ein Leseverständnistest für Erst- bis Sechstklässler*. Göttingen: Hogrefe.
- Liben, L. S. (2002). Spatial development in childhood: Where are we now? In U. Goswami (Eds.), *Handbook of childhood cognitive development* (pp. 326-348). Oxford: Blackwell.
- Linn, M. C. & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability. A meta-analysis. *Child Development*, 56, 1479-1498.
- Lohaus, A., Schumann-Hengsteler, R. & Kessler, T. (1999). *Räumliches Denken im Kindesalter*. Göttingen: Hogrefe.
- Lohaus, A., Vierhaus, M. & Maass, A. (2010). *Entwicklungspsychologie des Kindes- und Jugendalters*. Berlin: Springer.
- Lord, F. M. (1980). *Applications of item response theory to practical test problems*. Hillsdale, NJ: Erlbaum.
- Lüdtke, O., Robitzsch, A., Trautwein, U. & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung. *Psychologische Rundschau*, 58 (2), 103-117.
- Mathewson, J. H. (1999). Visual-spatial thinking: An aspect of science overlooked by educators. *Science Education*, 83, 33-54.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.

- Montada, L. (1998). Die geistige Entwicklung aus der Sicht Jean Piagets. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie* (S. 517-560). Weinheim: Beltz.
- Möller, K., Hardy, I., Jonen, A., Kleickmann, T. & Blumberg, E. (2006). Naturwissenschaften in der Primarstufe. Zur Förderung konzeptuellen Wissens durch Unterricht und zur Wirksamkeit von Lehrerfortbildungen. In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms* (S. 161-193). Münster: Waxmann.
- Möller, K., Sodian, B., Koerber, S., Hardy, I., Schwippert, K., Kleickmann, T., Mayer, D. & Pollmeier, J. (2011). *Arbeitsbericht zum Projekt Entwicklung naturwissenschaftlicher Kompetenz in der Grundschule – Science-P*. Unveröffentlichter Bericht an die Deutsche Forschungsgemeinschaft.
- Moosbrugger, H. (2008). Item-Response-Theorie (IRT). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 215-259). Heidelberg: Springer.
- Osterhaus, C., Koerber, S., Mayer, D. & Sodian, B. (2011). Wissen über Naturwissenschaften im Grundschulalter. Die Bedeutung des Wissenschaftsverständnisses der Lehrkraft. *Vortrag auf der 76. Tagung der Arbeitsgruppe für Empirische Pädagogische Forschung*, Klagenfurt, September.
- Parchmann, I. (2010). Kompetenzmodellierung in den Naturwissenschaften – Vielfalt ist wertvoll, aber nicht ohne ein gemeinsames Fundament. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes. 56. Beiheft der Zeitschrift für Pädagogik* (S. 135–142). Weinheim: Beltz.
- Paris, S. G., Morrison, F. J. & Miller, K. F. (2006). Academic pathways from preschool through elementary school. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (pp. 61-85). Mahwah, N.J.: Erlbaum.
- Pollmeier, J., Kleickmann, T., Hardy, I., Tröbst, S., Möller, K. & Schwippert, K. (2009). Entwicklung naturwissenschaftlicher Kompetenz in der Grundschule (Science-P). Naturwissenschaftliches Wissen. In C. Röhner, M. Hopf & C. Henrichwark (Hrsg.), *Europäisierung der Bildung. Konsequenzen und Herausforderungen für die Grundschulpädagogik, Jahrbuch Grundschulforschung Band 13* (S. 199-203). Wiesbaden: Verlag für Sozialwissenschaften.

- Prenzel, M. (1994). Mit Interesse in das dritte Jahrtausend! Pädagogische Überlegungen. In N. Seibert & H. J. Serve (Hrsg.), *Erziehung und Bildung an der Schwelle zum dritten Jahrtausend. Multidisziplinäre Aspekte, Analysen, Positionen, Perspektiven* (S. 32-44). Bern: Huber.
- Prenzel, M., Rost, J., Senkbeil, M., Häußler, P. & Klopp, A. (2001). Naturwissenschaftliche Grundbildung: Testkonzeption und Ergebnisse. In J. Baumert et al. (Hrsg.), *PISA 2000 – Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 191-248). Opladen: Leske + Budrich.
- Quaiser-Pohl, C. (2001). Raumvorstellungsfähigkeit. In G. Wenninger (Hrsg.), *Lexikon der Psychologie. M-Ref* (S. 422-423). Heidelberg: Spektrum.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Institute of Educational Research.
- Renner, G. (2005). Testbesprechung. TL. Turm von London – Deutsche Version. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 37 (4), 223-225.
- Rost, J. (2004). Lehrbuch Testtheorie – Testkonstruktion. Bern: Huber.
- Rost, J. & Georg, W. (1991). Alternative Skalierungsmöglichkeiten zur klassischen Testtheorie am Beispiel der Skala "Jugendzentrismus". *ZA-Information*, 28, 52-74.
- Ruffman, T., Perner, J., Olson, D. R. & Doherty, M. (1993). Reflecting on scientific thinking: Children's understanding of the hypothesis-evidence relation. *Child Development*, 64, 1617–1636.
- Schauble, L. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology*, 49 (1), 31–57.
- Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology*, 32 (1), 102–119.
- Schauble, L., Klopfer, L. E. & Raghavan, K. (1991). Students' transition from an engineering model to a science model of experimentation. *Journal of Research in Science Teaching*, 28 (9), 859-882.
- Schecker, H. & Parchmann, I. (2006). Modellierung naturwissenschaftlicher Kompetenz. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 45-66.
- Siegler, R., DeLoache, J. & Eisenberg, N. (2008). *Entwicklungspsychologie im Kindes- und Jugendalter*. Heidelberg: Spektrum.

- Schneider, W. & Stefanek, J. (2004). Entwicklungsveränderungen allgemeiner kognitiver Fähigkeiten und schulbezogener Fertigkeiten im Kindes- und Jugendalter. Evidenz für einen Schereneffekt? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 36 (3), 147-159.
- Schrader, F.-W., Helmke, A. & Hosenfeld, I. (2008). Stichwort: Kompetenzentwicklung im Grundschulalter. *Zeitschrift für Erziehungswissenschaft*, 11 (1), 7-29.
- Schwippert, K., Bos, W. & Lankes, E. M. (2004). Heterogenität und Chancengleichheit am Ende der vierten Jahrgangsstufe in den Ländern der Bundesrepublik Deutschland und im internationalen Vergleich. In W. Bos, E. M. Lankes, M. Prenzel, K. Schwippert, R. Valtin & G. Walther (Hrsg.), IGLU. Einige Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich (S. 165-190). Münster: Waxmann.
- Seiferth, N. Y., Thienel, R. & Kircher, T. (2007). Exekutive Funktionen. In F. Schneider & G. R. Fink (Hrsg.), *Funktionelle MRT in Psychiatrie und Neurologie* (S. 265-277). Berlin: Springer.
- Shea, D. L., Lubinski, D. & Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology*, 93 (3), 604-614.
- Shepard, R. N. & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171, 701-703.
- Six, B. (2004). Einstellung. In H. O. Häcker & K.-H. Stapf (Hrsg.), *Dorsch Psychologisches Wörterbuch* (S. 234-235). Bern: Huber.
- Smith, C. L., Maclin, D., Houghton, C. & Hennessey, M. G. (2000). Sixth-grade students' epistemologies of science: The impact of school science experiences on epistemological development. *Cognition & Instruction*, 18 (3), 349-422.
- Sodian, B. (1998). Entwicklung bereichsspezifischen Wissens. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie* (S. 622-653). Weinheim: Beltz.
- Sodian, B. (2005). Entwicklung des Denkens im Alter von vier bis acht Jahren – was entwickelt sich? In T. Guldemann & B. Hauser (Hrsg.), *Bildung 4- bis 8-jähriger Kinder* (S. 9-28). Münster: Waxmann.
- Sodian, B. & Bullock, M. (2008). Scientific reasoning – Where are we now? *Cognitive Development*, 23, 431-434.

- Sodian, B., Bullock, M. & Koerber, S. (2008). Wissenschaftliches Denken und Argumentieren: Was muss Hänschen lernen, damit aus Hans etwas wird? In W. Schneider (Hrsg.), *Entwicklung von der Kindheit bis zum Erwachsenenalter. Befunde der Münchner Längsschnittstudie LOGIK* (S. 67–84). Weinheim: Beltz.
- Sodian, B., Carey, S., Grosslight, L. & Smith, C. (1992). Junior high school students' understanding of the nature of scientific knowledge. The notion of interpretative frameworks. *Paper presented at the Annual Meeting of the American Research Association*, San Francisco, April.
- Sodian, B., Jonen, A., Thoermer, C. & Kircher, E. (2006). Die Natur der Naturwissenschaften verstehen: Implementierung wissenschaftstheoretischen Unterrichts in der Grundschule. In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms* (S. 147-160). Münster: Waxmann.
- Sodian, B. & Koerber, S. (2011). Hypothesenprüfung und Evidenzevaluation im Grundschulalter. *Unterrichtswissenschaft*, 39 (1), 21-34.
- Sodian, B., Thoermer, C., Kircher, E., Grygier, P. & Günther, J. (2002). Vermittlung von Wissenschaftsverständnis in der Grundschule. *Zeitschrift für Pädagogik*, 45. Beiheft, 192-206.
- Sodian, B., Zaitchik, D. & Carey, S. (1991). Young children's differentiation of hypothetical beliefs from evidence. *Child Development*, 6, 753-766.
- Songer, N. B. & Linn, M. C. (1991). How do students' views of science influence knowledge integration? *Journal of Research in Science Teaching*, 28 (9), 761-784.
- St. Clair-Thompson, H. L. & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *Quarterly Journal of Experimental Psychology*, 59 (4), 745-59.
- Tannenbaum, R. S. (1971). The development of the test of science processes. *Journal of Research in Science Teaching*, 8 (2), 123-136.
- Thoermer, C. & Sodian, B. (2002). Science undergraduates' and graduates' epistemologies of science: The notion of interpretive frameworks. *New Ideas in Psychology*, 26, 263-283.

- Tracey, T. J. G. (2002). Development of interests and competency beliefs. A 1-year longitudinal study of fifth- to eighth-grade students using the ICA-R and structural equation modeling. *Journal of Counseling Psychology, 49*, 148-163.
- Tracey, T. J. G. & Ward, C. C. (1998). The structure of children's interests and competence perceptions. *Journal of Counseling Psychology, 45*, 290-303.
- Tucha, O. & Lange, K.W. (2004). *Turm von London – Deutsche Version (TL-D)*. Göttingen: Hogrefe.
- Vandenberg, S. G. & Kuse, A. R. (1978). Mental rotations: A group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47*, 599-604.
- Visser, I. & Speekenbrink, M. (2010). depmixS4: an R package for hidden markov models. *Journal of Statistical Software, 36* (7), 1–21.
- Vosniadou, S., Ioannides, C., Dimitrakopoulou, A. & Papademetriou, E. (2001). Designing learning environments to promote conceptual change in science. *Learning and Instruction, 15*, 317-419.
- Wallace, C. S., Tsoi, M. Y., Calkin, J. & Darley, M. (2003). Learning from inquiry-based laboratories in nonmajor biology: An interpretive study of the relationships among inquiry experience, epistemologies, and conceptual growth. *Journal of Research in Science Teaching, 40* (10), 986–1024.
- Wang, W.-C. (1999). Direct estimation of correlations among latent traits within IRT framework. *Methods of Psychological Research Online, 4*, 63–82.
- Warm, T. A. (1989). Weighted Likelihood estimation of ability in item response models. *Psychometrika, 54*, 427-450.
- Weinert, F. E. (1999). *Concepts of Competence*. München: Max Planck Institut für Psychologische Forschung.
- Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45-65). Seattle: Hogrefe & Huber.
- Weinert, F. E. & Helmke, A. (Hrsg.) (1997). *Entwicklung im Grundschulalter*. Weinheim: Beltz.
- Weinert, F. E. & Schneider, W. (Eds.) (1999). *Individual development from 3 to 12. Findings from the Munich Longitudinal Study*. Cambridge: Cambridge University Press.

- Weiß, R. H. (2006). *Grundintelligenztest Skala 2 - Revision - (CFT 20-R)*. Göttingen: Hogrefe.
- Wilkening, F. & Sodian, B. (Eds.) (2005). Special Issue on "Scientific reasoning in young children". *Swiss Journal of Psychology*, 64 (3), 137-217.
- Wilson, M. (2005). *Constructing Measures. An item response modeling approach*. Mahwah: Erlbaum.
- Wittwer, J., Saß, S. & Prenzel, M. (2008). Naturwissenschaftliche Kompetenz im internationalen Vergleich: Testkonzeption und Ergebnisse. In W. Bos, M. Bensen, J. Baumert, M. Prenzel, C. Seiter & G. Walther (Hrsg.), *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 87-124). Münster: Waxmann.
- Wright, B. D. & Stone, M. H. (1979). *Best test design*. Chicago: Measurement, Evaluation, Statistics, and Assessment Press.
- Wu, M. & Adams, R. (2007). *Applying the Rasch model to psycho-social measurement: A practical approach*. Melbourne: Educational Measurement Solutions.
- Wu, M., Adams, R. & Wilson, M. (2005). *ACER ConQuest*. Australian Council for Educational Research.
- Zelazo, P. D., Carter, A., Reznick, J. S. & Fryne, D. (1997). Early development of executive function: A problem-solving framework. *Review of General Psychology*, 1 (2), 198-226.
- Zelazo, P. & Müller, U. (2002). Executive function in typical and atypical development. In U. Goswami (Eds.), *Handbook of childhood cognitive development* (pp. 445-469). Oxford, England: Blackwell Publishers.
- Zimmerman, C. (2000). The development of scientific reasoning skills. *Developmental Review*, 20 (1), 99-149.
- Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, 27 (2), 172-223.

8 Tabellen- und Abbildungsverzeichnis

Tabellenverzeichnis

- 3.1 Beschreibung der Stichprobe
- 3.2 Übersicht über die Zuordnung der Aufgaben zu Komponenten bzw. Teilkompetenzen des wissenschaftlichen Denkens und über die Grundlagen der Aufgabenentwicklung
- 3.3 Kompetenzniveaus des wissenschaftlichen Denkens
- 3.4 Construct Map der Teilkompetenz Verständnis der Ziele von Wissenschaften
- 3.5 Construct Map der Teilkompetenz Theorieverständnis
- 3.6 Construct Map der Teilkompetenz Experimentierstrategien
- 3.7 Construct Map der Teilkompetenz Dateninterpretation
- 3.8 Anzahl und Verteilung der Items auf Teilkompetenzen und Aufgabenformate
- 3.9 Ablauf der Untersuchung und eingesetzte Instrumente am ersten Testtag
- 3.10 Ablauf der Untersuchung und eingesetzte Instrumente am zweiten Testtag
- 3.11 Ablauf der Untersuchung und eingesetzte Instrumente am dritten Testtag
- 4.1 Modellvergleich des eindimensionalen, zweidimensionalen und vierdimensionalen Rasch-Modells der Kompetenz wissenschaftliches Denken
- 4.2 Latente Korrelationen der Komponenten des wissenschaftlichen Denkens
- 4.3 Latente Korrelationen der Teilkompetenzen des wissenschaftlichen Denkens
- 4.4 Varianzen und EAP/PV-Reliabilitäten der eindimensionalen Rasch-Modelle sowie Cronbachs Alpha für die Gesamtskala respektive Subskalen der Kompetenz wissenschaftliches Denken
- 4.5 Statistische Kennwerte der eindimensionalen Rasch-Skalierung in absteigender Schwierigkeitsreihenfolge
- 4.6 Vergleich der Modellpassung der LCA-Modelle
- 4.7 Klassenzugehörigkeit nach Jahrgangsstufe
- 4.8 Lösungshäufigkeiten (in Prozent) der Aufgaben der Teilkompetenz Verständnis der Ziele von Wissenschaften geordnet nach ansteigender Aufgabenschwierigkeit getrennt nach Jahrgangsstufe
- 4.9 Lösungshäufigkeiten (in Prozent) der Aufgaben der Teilkompetenz Theorieverständnis geordnet nach ansteigender Aufgabenschwierigkeit getrennt nach Jahrgangsstufe

- 4.10 Lösungshäufigkeiten (in Prozent) der Aufgaben der Teilkompetenz Experimentierstrategien geordnet nach ansteigender Aufgabenschwierigkeit getrennt nach Jahrgangsstufe
- 4.11 Lösungshäufigkeiten in Prozent der Aufgaben der Teilkompetenz Dateninterpretation geordnet nach ansteigender Aufgabenschwierigkeit getrennt nach Jahrgangsstufe
- 4.12 Mittlere Personenfähigkeiten (WLE) im wissenschaftlichen Denken aufgeteilt nach Geschlecht für die Gesamtstichprobe und die einzelnen Klassenstufen
- 4.13 Mittlere Personenfähigkeiten (WLE) in den Komponenten des wissenschaftlichen Denkens aufgeteilt nach Jahrgangsstufe und Geschlecht
- 4.14 Statistische Kennwerte der univariaten zweifaktoriellen (Geschlecht, Klassenstufe) Varianzanalysen der Komponenten des wissenschaftlichen Denkens
- 4.15 Mittlere Personenfähigkeiten (WLE) in den Teilkompetenzen des wissenschaftlichen Denkens aufgeteilt nach Jahrgangsstufe und Geschlecht
- 4.16 Statistische Kennwerte der univariaten zweifaktoriellen (Geschlecht, Klassenstufe) Varianzanalysen der Teilkompetenzen des wissenschaftlichen Denkens
- 4.17 Statistische Kennwerte des eindimensionalen Rasch-Modells der Aufgabensets der Teilkompetenzen des wissenschaftlichen Denkens
- 4.18 Anteil der Kinder in Prozent, die die Teilkompetenz-Aufgabensets richtig beantworteten
- 4.19 Anteil der Kinder in Prozent, die die möglichen Antwortmuster auf die 4 Aufgabensets aufwiesen
- 4.20 Prozentsatz an Kindern, die das Experimentierstrategien-Aufgabenset bzw. das Dateninterpretation-Aufgabenset richtig gelöst haben als eine Funktion des meta-konzeptuellen Theorieverständnisses
- 4.21 Prozentsatz an Kindern, die das Theorieverständnis-Aufgabenset bzw. das Dateninterpretation-Aufgabenset richtig gelöst haben als eine Funktion des Verständnisses der experimentellen Methodik
- 4.22 Prozentsatz an Kindern, die das Theorieverständnis-Aufgabenset bzw. das Experimentierstrategien-Aufgabenset richtig gelöst haben als eine Funktion der Evidenzevaluationsfähigkeiten
- 4.23 Personenfähigkeiten (WLE) im Leseverständnis aufgeteilt nach Geschlecht für die Gesamtstichprobe und die einzelnen Klassenstufen
- 4.24 Modellvergleich des eindimensionalen und zweidimensionalen Rasch-Modells (Dimension 1 = Leseverständnis, Dimension 2 = wissenschaftliches Denken)
- 4.25 Latente Korrelationen zwischen wissenschaftlichem Denken und Leseverständnis
- 4.26 Latente Korrelationen zwischen Wissenschaftsverständnis bzw. Methodenkompetenz und Leseverständnis
- 4.27 Personenfähigkeiten (WLE) im Intelligenztest aufgeteilt nach Geschlecht für die Gesamtstichprobe und die einzelnen Klassenstufen

- 4.28 Modellvergleich des eindimensionalen und zweidimensionalen Rasch-Modells (Dimension 1 = Intelligenz, Dimension 2 = wissenschaftliches Denken)
- 4.29 Latente Korrelation zwischen wissenschaftlichem Denken und Intelligenz
- 4.30 Latente Korrelationen zwischen Wissenschaftsverständnis bzw. Methodenkompetenz und Intelligenz
- 4.31 Personenfähigkeiten (WLE) im Problemlösen aufgeteilt nach Geschlecht für die Gesamtstichprobe und die einzelnen Klassenstufen
- 4.32 Modellvergleich des eindimensionalen und zweidimensionalen Rasch-Modells (Dimension 1 = Problemlösen, Dimension 2 = wissenschaftliches Denken)
- 4.33 Latente Korrelationen zwischen wissenschaftlichem Denken und Problemlösen
- 4.34 Latente Korrelationen zwischen Wissenschaftsverständnis bzw. Methodenkompetenz und Problemlösen
- 4.35 Personenfähigkeiten (WLE) im Inhibitionstest aufgeteilt nach Geschlecht für die Gesamtstichprobe und die einzelnen Klassenstufen
- 4.36 Modellvergleich des eindimensionalen und zweidimensionalen Rasch-Modells (Dimension 1 = Inhibition, Dimension 2 = wissenschaftliches Denken)
- 4.37 Latente Korrelationen zwischen wissenschaftlichem Denken und Inhibition
- 4.38 Latente Korrelationen zwischen Wissenschaftsverständnis bzw. Methodenkompetenz und Inhibition
- 4.39 Personenfähigkeiten (WLE) im räumlichen Denken aufgeteilt nach Geschlecht für die Gesamtstichprobe und die einzelnen Klassenstufen
- 4.40 Modellvergleich des eindimensionalen und zweidimensionalen Rasch-Modells (Dimension 1 = räumliches Denken, Dimension 2 = wissenschaftliches Denken)
- 4.41 Latente Korrelationen zwischen wissenschaftlichem Denken und räumlichem Denken
- 4.42 Latente Korrelationen zwischen Wissenschaftsverständnis bzw. Methodenkompetenz und räumlichem Denken
- 4.43 Personenfähigkeiten (WLE) der Viertklässler im formal-operatorischen Denken
- 4.44 Modellvergleich des eindimensionalen und zweidimensionalen Rasch-Modells (Dimension 1 = formal-operatorische Fähigkeiten, Dimension 2 = wissenschaftliches Denken) bezogen auf die Jahrgangsstufe 4
- 4.45 Latente Korrelation zwischen wissenschaftlichem Denken und formal-operatorischen Fähigkeiten
- 4.46 Latente Korrelationen zwischen Wissenschaftsverständnis bzw. Methodenkompetenz und formal operatorische Fähigkeiten
- 4.47 Prozentsatz der Viertklässler (N = 141), die mindestens 65 % der Aufgaben zum wissenschaftlichen Denken richtig gelöst haben, als einen Funktion des operatorischen Niveaus
- 4.48 Prozentsatz der Viertklässler (N = 141), die Komponenten-Aufgabensets richtig (mindestens 65 %) gelöst haben, als einen Funktion des operatorischen Niveaus

- 4.49 Personenparameter (WLE) in naturwissenschaftsbezogenen Einstellungen aufgeteilt nach Geschlecht für die Gesamtstichprobe und die einzelnen Klassenstufen
- 4.50 Latente Korrelationen zwischen wissenschaftlichem Denken und naturwissenschaftsbezogenen Einstellungen
- 4.51 Latente Korrelationen zwischen Wissenschaftsverständnis bzw. Methodenkompetenz und naturwissenschaftsbezogenen Einstellungen
- 4.52 Personenparameter (WLE) im naturwissenschaftsbezogenem Fähigkeitsselbstkonzept aufgeteilt nach Geschlecht für die Gesamtstichprobe und die einzelnen Klassenstufen
- 4.53 Latente Korrelation zwischen wissenschaftlichem Denken und naturwissenschaftsbezogenem Fähigkeitsselbstkonzept
- 4.54 Latente Korrelationen zwischen Wissenschaftsverständnis bzw. Methodenkompetenz und naturwissenschaftsbezogenem Fähigkeitsselbstkonzept
- 4.55 Personenparameter (WLE) im Interesse an forschenden Aktivitäten aufgeteilt nach Geschlecht für die Gesamtstichprobe und die einzelnen Klassenstufen
- 4.56 Latente Korrelationen zwischen wissenschaftlichem Denken und Interesse an forschenden Aktivitäten
- 4.57 Latente Korrelationen zwischen Wissenschaftsverständnis bzw. Methodenkompetenz und Interesse an forschenden Aktivitäten
- 4.58 Pearson-Interkorrelationen der Prädiktoren in der Gesamtstichprobe
- 4.59 Ergebnisse latenter hierarchischer Regressionen zur Vorhersage der Kompetenz im wissenschaftlichen Denken bezogen auf die Gesamtstichprobe
- 4.60 Pearson-Interkorrelationen der Prädiktoren im frühen Grundschulalter
- 4.61 Ergebnisse latenter hierarchischer Regressionen zur Vorhersage der Kompetenz im wissenschaftlichen Denken bezogen auf das frühe Grundschulalter
- 4.62 Pearson-Interkorrelationen der Prädiktoren im späten Grundschulalter
- 4.63 Ergebnisse latenter hierarchischer Regressionen zur Vorhersage der Kompetenz im wissenschaftlichen Denken bezogen auf das späte Grundschulalter
- B.1 ELFE 1-6 – Subtest Textverständnis: Statistische Kennwerte der eindimensionalen Rasch-Skalierung
- B.2 CFT 20-R – Subtest 1 und 4: Statistische Kennwerte der eindimensionalen Rasch-Skalierung
- B.3 Turm von London: Statistische Kennwerte der eindimensionalen Rasch-Skalierung
- B.4 Stroop-Test: Statistische Kennwerte der eindimensionalen Rasch-Skalierung
- B.5 Räumlicher Test: Statistische Kennwerte der eindimensionalen Rasch-Skalierung
- B.6 Arlin-Test: Statistische Kennwerte der eindimensionalen Rasch-Skalierung
- B.7 Skala naturwissenschaftsbezogene Einstellungen (TIMSS): Statistische Kennwerte des eindimensionalen Partial-Credit-Modells

- B.8 Skala naturwissenschaftsbezogenes Fähigkeitsselbstkonzept (TIMSS): Statistische Kennwerte des eindimensionalen Partial-Credit-Modells
- B.9 Skala Interesse an forschenden Aktivitäten (ICA-R): Statistische Kennwerte des eindimensionalen Partial-Credit-Modells
- B.10 Pearson-Korrelationen der WLE-Personenparameter und Summenscores der Prädiktoren
- B.11 Ergebnisse latenter einfacher Regressionen zur Vorhersage der Kompetenz im wissenschaftlichen Denken bezogen auf die Gesamtstichprobe
- B.11 Ergebnisse latenter einfacher Regressionen zur Vorhersage der Kompetenz im wissenschaftlichen Denken bezogen auf das frühe Grundschulalter
- B.12 Ergebnisse latenter einfacher Regressionen zur Vorhersage der Kompetenz im wissenschaftlichen Denken bezogen auf das späte Grundschulalter

Abbildungsverzeichnis

- 1.1 Zyklus des wissenschaftlichen Erkenntnisprozesses
- 1.2 Schematische Darstellung des Strukturmodells der Kompetenz wissenschaftliches Denken
- 1.3 Rahmenmodell der Kompetenz wissenschaftliches Denken
- 3.1 Itemcharakteristische Funktion (ICF) eines Items
- 4.1 Wright map des eindimensionalen Rasch-Modells: Abbildung der Personenfähigkeiten und Itemschwierigkeiten auf der gemeinsamen Skala (in logit-Einheiten)
- 4.2 Itemprofile der latenten Klassen des 2-Klassen-LCA-Modells
- 4.3 Kompetenzprofile der latenten Klassen des 2-Klassen-LCA-Modells für die Teilkomponenten des wissenschaftlichen Denkens
- 4.4-4.5 Wahrscheinlichkeit für die Klassenzugehörigkeit nach Jahrgangsstufe und Alter
- 4.6 Mittlere relative Häufigkeiten für Fehlvorstellungen und wissenschaftliche Vorstellungen dargestellt getrennt nach Jahrgangsstufen
- 4.7 Durchschnittliche Leistung der einzelnen Klassenstufen je Teilkompetenz des wissenschaftlichen Denkens
- 4.8 Itemcharakteristische Funktionen der Aufgabensets der Teilkompetenzen des wissenschaftlichen Denkens

9 Anhang


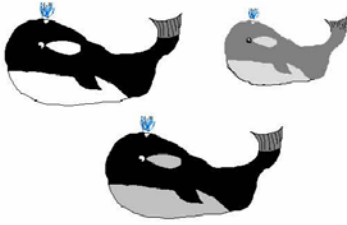

Anhang A

Testmaterial

- A.1 Aufgaben zur Erfassung des wissenschaftlichen Denkens (in der Reihenfolge des präsentierten Testhefts)
- A.2 Manuale zur Kodierung der offenen Antworten
- A.3 Papierversion des Tests Turm von London zur Erfassung von Problemlösefähigkeiten
- A.4 Papierversion des Stroop-Tests zur Erfassung der Inhibition
- A.5 Aufgaben zur Erfassung des räumlichen Denkens
- A.6 Deutsche Version des Arlin-Tests zur Erfassung formal-operatorischer Fähigkeiten
- A.7 Deutsche Version des ICA-R Inventory of Childrens' Activities – Revised zur Erfassung des Interesses an forschenden Aktivitäten

A.1 Aufgaben zur Erfassung des wissenschaftlichen Denkens (in der Reihenfolge des präsentierten Testhefts)

Aufgabe a02_Hypothese

<p>Eine Hypothese ist eine wissenschaftliche Vermutung, die man überprüfen kann.</p>		
<p>Michael überlegt, was das bessere Beispiel für eine Hypothese ist.</p>		
<p>Was ist das <u>bessere</u> Beispiel für eine Hypothese? Kreuze an!</p>		
	<p>Es gibt viele Arten von Walen.</p>	<input type="checkbox"/>
	<p>Sonnenblumen, die gedüngt werden, werden größer.</p>	<input type="checkbox"/>


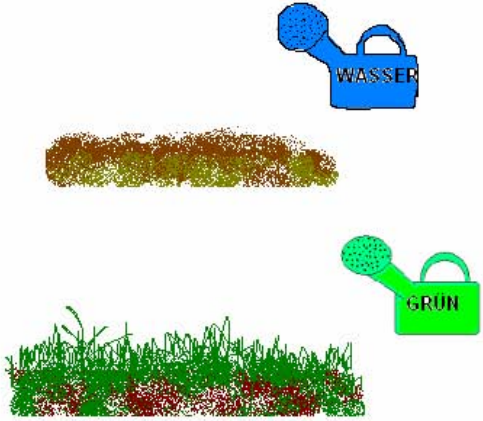
Aufgabe a08_Theorie

<p>Markus, Simon und Thomas wollen schwimmen gehen. Sie glauben, der Badesee ist schön warm.</p> <p>Sie haben unterschiedliche Ideen, warum der See warm ist.</p>	
	<p>Markus sagt:</p> <p>„Der Badesee ist warm, weil es die letzten drei Wochen nicht geregnet hat.“</p>
	<p>Simon sagt:</p> <p>„Der Badesee ist warm, weil die Sonne geschienen hat und der See die Wärme speichert.“</p>
	<p>Thomas sagt:</p> <p>„Der Badesee ist warm, weil es Juli und Hochsommer ist.“</p>
<p>Wer hat die wissenschaftlichste Idee?</p> <p>Kreuze die <u>beste</u> Antwort an!</p>	

Aufgabe a16_Wissenschaftler

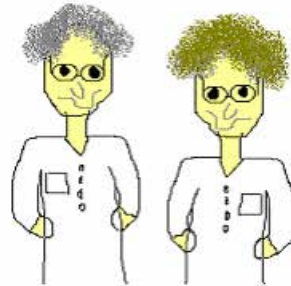
<p>Andrea, Paula und Jana überlegen, was ein Wissenschaftler ist.</p>	
	<p>Jana sagt: „Mein Papa testet verschiedene Gummiarten für Winterreifen, um herauszufinden, mit welchen man am wenigsten rutscht.“</p>
	<p>Paula sagt: „Mein Papa testet mit einer Stoppuhr, wie lange die Autos mit Winterreifen brauchen, um von 0 auf 100 zu kommen.“</p>
	<p>Andrea sagt: „Mein Papa arbeitet in einer großen Firma und testet dort jeden Winterreifen, damit keiner ein Loch hat.“</p>
<p>Welcher Papa ist ein Wissenschaftler? Kreuze die <u>beste</u> Antwort an!</p>	

Aufgabe a20_Pflanzenmittel

<p>Ein Wissenschaftler hat eine Idee: „Felder werden grüner, wenn sie mit dem Pflanzenschutzmittel „GRÜN“ gegossen werden. Sie werden nicht so grün, wenn sie nur Wasser bekommen.“</p>	
<p>Er macht ein Experiment und findet heraus: Die Felder, die mit dem Pflanzenschutzmittel „GRÜN“ gegossen wurden, waren nach einem Jahr grüner als die Felder, die nur Wasser bekommen haben.</p>	
<p>Was wird der Wissenschaftler als nächstes tun? Kreuze die <u>bessere</u> Antwort an!</p>	
<p><input type="radio"/> Er wird das gleiche Experiment noch mal machen, um sicher zu sein, dass er keinen Fehler gemacht hat.</p> <p><input type="radio"/> Sein Experiment war erfolgreich. Er braucht nichts mehr machen.</p>	

Aufgabe a21_Fehler

Wissenschaftler erforschen die Natur und machen viele neue Entdeckungen.







Können Wissenschaftler auch Fehler machen?

Kreuze die beste Antwort an!

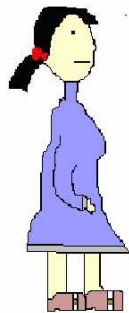
- Ja, ihre grundsätzlichen Annahmen können falsch sein.
- Nein, Wissenschaftler wissen sehr viel, deshalb machen sie normalerweise keine Fehler.
- Ja, sie arbeiten manchmal mit Geräten, die nicht funktionieren. Dann kann es zu Fehlern kommen.

Aufgabe a24_Mittelalter

<p>Vor langer Zeit, im Mittelalter, glaubten die Leute, dass es Hexen gibt, die Leute krank machen können.</p>	
<p>Ein Wissenschaftler aus unserer Zeit flog mit einer Zeitmaschine ins Mittelalter.</p>	
<p>Der Wissenschaftler von damals glaubte, dass Hexen die Menschen krank machen können. Der Wissenschaftler aus unserer Zeit glaubte, dass Bakterien die Menschen krank machen können.</p>	
<p>Der Wissenschaftler aus unserer Zeit zeigte dem Wissenschaftler von damals diese Bakterien unter einem Mikroskop und behauptete: „Diese Bakterien sind der Grund, warum Menschen krank werden!“</p>	
<p>Was würde der <u>Wissenschaftler von damals</u> jetzt sagen? Kreuze die <u>beste</u> Antwort an!</p>	
<p>„Also gut, Bakterien machen die Leute krank, nicht Hexen.“</p>	<input type="checkbox"/>
<p>„Bakterien könnten ja kleine Helfer der Hexen sein.“</p>	<input type="checkbox"/>
<p>„Es kann schon sein, dass es Bakterien gibt, aber die Hexe macht die Menschen krank.“</p>	<input type="checkbox"/>

Aufgabe a27_Geburtstagsparty

Die Schwestern Eva und Susi kommen zerstritten von einer Geburtstagsparty nach Hause. Beide haben Ketchupflecken auf den Kleidern. Sie erzählen ihrer Mutter, warum sie sich gestritten haben.

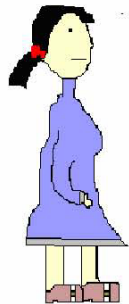


Mama



Eva

Susi hat mich geschubst. Dabei ist mir Ketchup aufs Kleid gekommen.



Mama



Susi


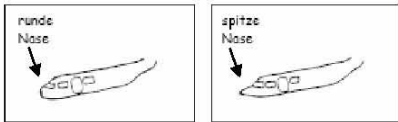
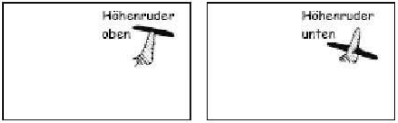
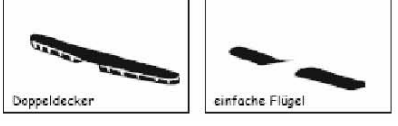

Eva hat Schuld. Ich musste drängeln, weil sie mich nicht vorbei lassen wollte. Dabei ist mir Ketchup aufs Kleid gekommen.

Warum haben Eva und Susi unterschiedliche Meinungen über den Streit?

Kreuze die beste Antwort an!

- Eine von beiden lügt.
- Sie haben das, was passiert ist, verschieden erlebt.
- Eine hat nicht alles gesehen.


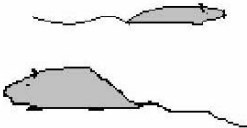
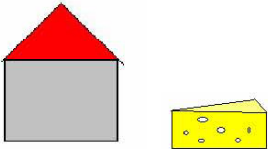
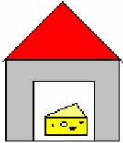
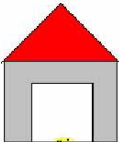
Aufgabe a35_Flugzeug

<p>Herr Müller baut Flugzeuge und möchte, dass sie möglichst wenig Treibstoff verbrauchen. Jetzt hat er verschiedene Ideen, wovon der Treibstoffverbrauch abhängen könnte:</p>	
<p>Er denkt: Ein Flugzeug kann eine runde Nase oder eine spitze Nase haben.</p>	
<p>Er denkt: Die Höhenruder können oben oder unten angebracht werden.</p>	
<p>Er denkt: Ein Flugzeug kann doppelte Flügel oder einfache Flügel haben.</p>	
<p>Herr Müller überlegt: Es könnte daran liegen, ob die Höhenruder oben oder unten angebracht werden.</p>	
<p>Was soll Herr Müller tun, um herauszufinden, ob die Stellung der Höhenruder wichtig oder egal für den Treibstoffverbrauch ist? Kreuze die beste Antwort an!</p>	
<p>1. Herr Müller muss ein paar Flugzeuge bauen und schauen, ob sie wenig Treibstoff verbrauchen.</p>	<input type="checkbox"/>
<p>2. Herr Müller muss zwei Flugzeuge bauen, eines mit dem Höhenruder oben und eines mit dem Höhenruder unten. Sie müssen aber sonst gleich sein.</p>	<input type="checkbox"/>
<p>3. Herr Müller muss zwei ganz unterschiedliche Flugzeuge bauen, bei denen er Nase, Flügel und Höhenruder unterschiedlich macht.</p>	<input type="checkbox"/>

Aufgabe a44_Ärzte

<p>Zwei Ärzte haben ein Mittel gegen Bauchweh entwickelt. Doktor Baum eine grüne Tablette und Doktor Kirsch eine rote Tablette.</p> <p>Sie wollen testen, welche Tablette besser gegen Bauchweh hilft.</p>	
<p>Sie haben 100 Kindern mit Bauchweh die grüne Tablette gegeben.</p> <p>Danach hatte kein Kind mehr Bauchweh.</p>	
<p>Können die Ärzte so herausfinden, welche Tablette besser gegen Bauchweh hilft?</p> <p><u>Kreuze an!</u></p>	
<p><input type="radio"/> ja</p> <p><input type="radio"/> nein</p>	<p>Warum?</p> <p>_____</p> <p>_____</p> <p>_____</p>

Aufgabe a51_Maus-Haus

<p>Tina und Lara haben eine Maus im Keller bemerkt. Sie haben die Maus aber noch nicht gesehen. Denn die Maus kommt immer in der Nacht, wenn alle Menschen schlafen.</p>	
<p>Tina und Lara wollen herausfinden, ob die Maus dick oder dünn ist. Sie wollen sie aber nicht einfangen.</p>	
<p>Sie wollen Käse in ein Häuschen legen und dann sehen, ob die Maus den Käse frisst.</p>	
<p>Sie nehmen ein Häuschen mit einem großen Eingang und legen den Käse hinein.</p>	
<p>Am nächsten Morgen ist der Käse weg.</p>	
<p>War das ein guter Versuch, um herauszufinden, ob die Maus dick oder dünn ist? Kreuze die <u>beste</u> Antwort an!</p>	
<p>1. Ja, weil die Maus etwas zu essen bekommen hat.</p>	<input type="checkbox"/>
<p>2. Ja, weil man etwas herausfinden kann, wenn man ein Haus mit Käse aufstellt.</p>	<input type="checkbox"/>
<p>3. Nein, weil eine dicke und eine dünne Maus durch den großen Eingang im Haus passen.</p>	<input type="checkbox"/>

Aufgabe a64_Zähne

Forscher machen eine Untersuchung in einem Internat.
 Sie wollen herausfinden, ob das, was die Kinder essen, etwas damit zu tun hat, ob die Kinder gute oder schlechte Zähne haben.



Deshalb haben die Forscher die Erzieher gebeten, den Kindern 6 Monate lang immer ganz bestimmtes Essen zu geben:



Maiskekse oder Karotten



Kaugummi oder Schokolade

Die Kinder im Internat wurden auf vier Tische verteilt, so dass an jedem Tisch fünf Kinder saßen. An jedem Tisch gab es ganz bestimmtes Essen.

Die Forscher haben genau beobachtet, was die Kinder gegessen haben und wer schlechte Zähne bekommen hat und wer nicht.

Schau dir die Tische ganz genau an!

<p style="text-align: center;">gute Zähne</p>	<p style="text-align: center;">schlechte Zähne</p>
<p style="text-align: center;">schlechte Zähne</p>	<p style="text-align: center;">gute Zähne</p>

Die Forscher wollen wissen, ob es an der Schokolade liegt, dass die Kinder im Internat schlechte Zähne bekommen.

Sie haben sich die Bilder ganz genau angesehen und sind jetzt der Meinung:

„Wir wissen nicht, ob es an der Schokolade liegt!“

Was glaubst du?

Schau auf die Bilder und kreuze die beste Antwort an!

- | | |
|--|--------------------------|
| 1. Ich weiß es aber schon! Weil Schokolade macht immer schlechte Zähne. | <input type="checkbox"/> |
| 2. Ich bin mir auch nicht sicher, weil die Kinder auch schlechte Zähne haben, wenn sie Maiskekse gegessen haben. | <input type="checkbox"/> |
| 3. Ich glaube, dass die Forscher die Bilder falsch gelesen haben, denn die Maiskekse machen schlechte Zähne. Alle Kinder, die Maiskekse gegessen haben, haben schlechte Zähne. | <input type="checkbox"/> |

Aufgabe a82_Gummibärchen

Der Chef einer Gummibärchenfabrik behauptet, dass die Löcher in den Zähnen **nichts** damit zu tun haben, ob man Gummibärchen isst oder nicht.



Hier siehst du zwei Tabellen, in denen gezeigt wird, wie viele Gummibärchen die Kinder jeden Tag gegessen haben und wie viele Löcher der Zahnarzt beim letzten Mal gefunden hat.

Tabelle 1

	Gummi- bärchen	Löcher
Max	0	0
Nina	1	1
Lisa	2	2

Tabelle 2

	Gummi- bärchen	Löcher
Toni	0	3
Rita	1	3
Andi	2	3

Der Chef der Gummibärchenfabrik behauptet:



Die Löcher in den Zähnen haben **nichts** damit zu tun, ob man Gummibärchen isst oder nicht.

Welche Tabelle zeigt, dass der Chef der Gummibärchenfabrik Recht hat?

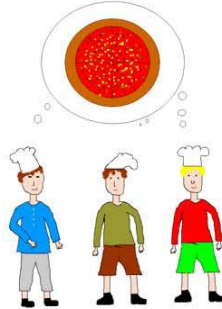
Kreuze an!

Aufgabe a23_Wichtel

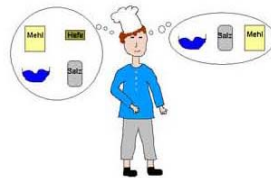
<p>Vor langer Zeit glaubten Menschen an Wichtel, die kranke Menschen gesund machen können.</p>	
<p>Immer wenn Menschen krank waren, haben sie die Wichtel besucht. Nach dem Besuch bei den Wichteln waren die kranken Menschen wieder gesund.</p>	
<p>Auf dem Weg zu den Wichteln haben die kranken Menschen auch immer Kräuter gesammelt und gegessen.</p>	
<p>Was glauben die <u>Menschen von damals</u>, warum sie gesund geworden sind? Kreuze die <u>beste</u> Antwort an!</p>	
<p>Die <u>Menschen von damals</u> glauben:</p> <p><input type="radio"/> „Man braucht beides: Kräuter und Wichtel haben uns geheilt.“</p> <p><input type="radio"/> „Die Wichtel haben uns geheilt.“</p> <p><input type="radio"/> „Die Kräuter, die wir gesammelt haben, haben uns geheilt.“</p>	

Aufgabe a49_Pizzabacken

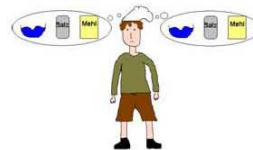
Adrian, Ludwig und Hans backen Pizza. Sie möchten herausfinden, ob man für den Pizzateig Hefe braucht, damit der Pizzateig dick wird.



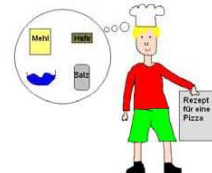
Adrian mischt für eine Pizza Mehl, Salz, Wasser und Hefe. Für eine andere Pizza mischt er nur Mehl, Salz und Wasser.



Ludwig mischt für eine Pizza Mehl, Salz und Wasser. Er probiert es noch bei einer zweiten Pizza aus und mischt noch einmal Mehl, Salz und Wasser.







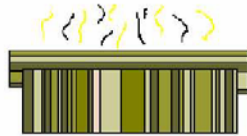
Hans hat ein Pizzarezept und backt eine Pizza ganz genau nach dem Rezept. Er mischt Mehl, Salz, Wasser und Hefe.






Wer arbeitet mehr wie ein Wissenschaftler?

Kreuze die beste Antwort an!


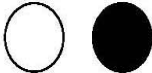



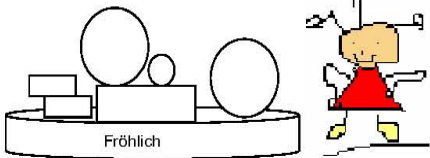
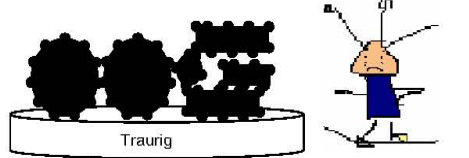
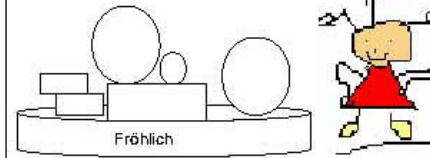
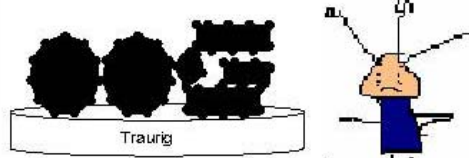
Aufgabe a53_Spürnase

<p>Paula und Susi haben ein neues Haustier. Es ist ein sehr seltenes und unbekanntes Tier, ein Erdferkel.</p> <p>Sie haben unterschiedliche Meinungen, ob das neue Tier eine feine Spürnase hat oder nicht.</p>		
<p>Sie nehmen eine Kiste, stark riechenden Käse und schwach riechendes Brot.</p> <p>Jetzt überlegen sie, wo sie den Käse oder das Brot hinlegen sollen, um herauszufinden, ob das Haustier eine feine Spürnase hat oder nicht.</p>		
<p>Wie können Paula und Susi herausfinden, ob das Haustier eine feine Spürnase hat oder nicht?</p> <p>Kreuze die <u>beste</u> Möglichkeit an!</p>		
 <p>Brot in der Kiste</p>	 <p>Käse auf der Kiste</p>	 <p>Käse in der Kiste</p>

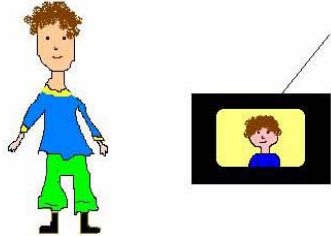
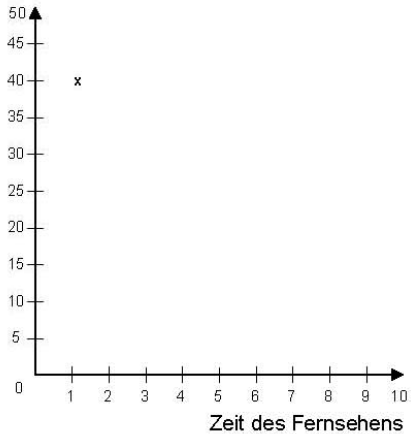
Aufgabe a57_Fußpilz

<p>Herr Sonne ist Apotheker und hat eine Tablette gegen Fußpilz entwickelt.</p>	
<p>Er macht eine Untersuchung, um ihre Wirkung auszuprobieren: Er sucht im Telefonbuch 100 Personen aus: Er ruft sie an und fragt, ob sie an der Untersuchung teilnehmen wollen. Alle 100 Personen nehmen teil. Diese 100 Personen nehmen nun 10 Tage lang jeden Tag eine Tablette.</p>	
<p>Nach den 10 Tagen ruft Herr Sonne diese Menschen wieder an und fragt, ob sie Fußpilz haben. Keiner dieser Menschen hat Fußpilz.</p>	
<p style="text-align: center;">Hat Herr Sonne eine gute Untersuchung gemacht? <u>Kreuze an!</u></p>	
<p><input type="radio"/> ja <input type="radio"/> nein</p>	<p>Warum glaubst du das?</p> <p>-----</p> <p>-----</p> <p>-----</p>

Aufgabe a66_Ravianer

<p>Auf dem Planeten „Ravi“ gibt es Steine, die machen die „Ravianer“ entweder fröhlich oder traurig, wenn sie sie in die Hand nehmen. Forscher wollten herausfinden, woran es liegen kann.</p> <p>Sie überlegen:</p>	
<p>Liegt es an der Farbe?</p>	 <p>hell dunkel</p>
<p>Liegt es an der Größe?</p>	 <p>groß klein</p>
<p>Liegt es an der Form?</p>	 <p>rund eckig</p>
<p>Liegt es an der Oberfläche?</p>	 <p>glatt rau</p>
<p>Die Forscher haben eine Untersuchung gemacht. Sie haben den Ravianern verschiedene Steine in die Hand gegeben. Dann haben sie gesehen, ob sie fröhlich oder traurig geworden sind.</p> <p>Schau dir die zwei Tische auf dem Bild an! Dort siehst du, was die Forscher herausgefunden haben.</p>	
 <p>Fröhlich</p>	 <p>Traurig</p>
<p>Die Forscher haben herausgefunden, dass sie mit den Bildern nicht erklären können, an was es liegt, dass die Ravianer fröhlich oder traurig werden.</p>	
<p style="text-align: center;">Schau dir die Bilder noch einmal an!</p>	
 <p>Fröhlich</p>	 <p>Traurig</p>
<p style="text-align: center;">Wie würden die Forscher ihre Meinung begründen?</p> <p style="text-align: center;">Kreuze die <u>beste</u> Antwort an!</p> <p>Die Forscher sagen:</p>	
<p>1. Man kann es nicht sagen, weil es an der Farbe oder an der Oberfläche liegen könnte.</p>	<input type="checkbox"/>
<p>2. Man kann es nicht sagen, weil helle Steine fröhlich machen und raue Steine traurig.</p>	<input type="checkbox"/>
<p>3. Man kann es nicht sagen, weil raue Steine die Hände verletzen und das traurig macht.</p>	<input type="checkbox"/>

Aufgabe a76_Diagramm

<p>Nick glaubt: „Je weniger Zeit ich mit Fernsehen verbringe, desto weniger Filmtitel kann ich mir merken.“</p>	
<p>Nick möchte das, was er glaubt, in einem Bild zeigen. Hier rechts siehst du sein Bild.</p>	<p>Anzahl der Filmtitel</p>  <p>Zeit des Fernsehens</p>
<p>Zeigt Nicks Bild das, was er glaubt? <u>Kreuze an!</u></p>	
<p><input type="radio"/> ja <input type="radio"/> nein</p>	

A.2 Manuale zur Kodierung der offenen Antworten

Kodiermanual Aufgabe a44_2 Ärzte (offene Antwort)

Der folgende Fehler im experimentellen Design muss erkannt werden: Variation der fokalen Variablendimension wurde nicht berücksichtigt, d. h. man muss die Wirkung beider Tabletten untersuchen, um zu wissen, welche Tablette besser wirkt.

Kodierung	Offene Antwort
------------------	-----------------------

1	Wenn erkannt wurde, dass beide Tabletten getestet werden müssen: z. B. Hat nur 1 Tablette getestet Muss 2 Tabletten ausprobieren
0	Keine Ahnung weiß ich nicht Antworten, die von eigentlicher Fragestellung weggehen

Kodiermanual Aufgabe a57_2 Fußpilz (offene Antwort)

Der folgende Fehler im experimentellen Design muss erkannt werden: Basisrate wurde nicht berücksichtigt, d. h., es ist wichtig zu wissen, wie viele der untersuchten Personen vor der Untersuchung Fußpilz hatten.

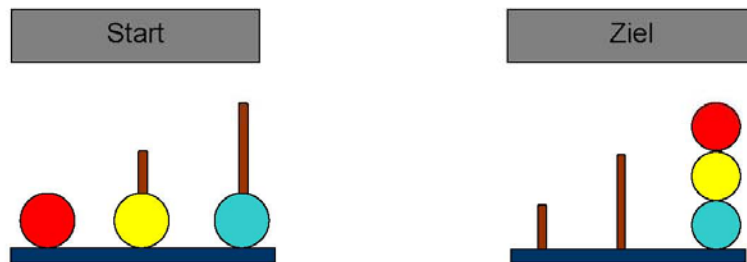
Kodierung	Offene Antwort
------------------	-----------------------

1	Wenn erkannt wurde, dass es wichtig ist zu wissen, wie viele der untersuchten Personen vor der Untersuchung Fußpilz hatten: z. B. Muss wissen, ob die Leute vor der Untersuchung Fußpilz haben Muss wissen, wie viele Leute vor der Untersuchung Fußpilz haben
0	Keine Ahnung weiß ich nicht Antworten, die von eigentlicher Fragestellung weggehen.

A.3 Papierversion des Tests Turm von London zur Erfassung von Problemlösefähigkeiten

Name: _____

Beispielaufgabe

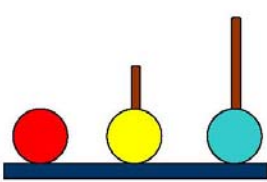
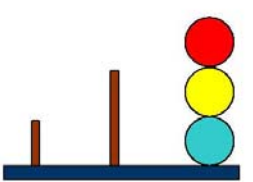
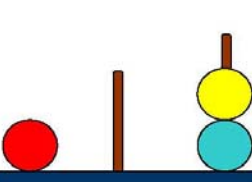
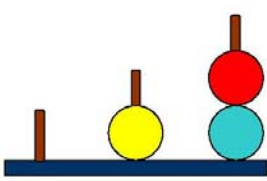


Hier unten siehst du vier Bilder. Diese Bilder zeigen das Startbild und das Zielbild. Außerdem siehst du noch zwei Bilder: Ein Bild, zeigt einen Zug, um vom Start zum Ziel zu kommen. Das andere Bild ist falsch und führt nicht zum Ziel.

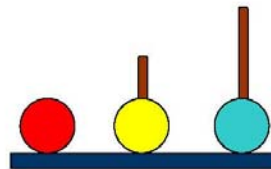
Komme in zwei Zügen vom Start zum Ziel! (Der letzte Zug ist das Zielbild)
Leider sind die Bilder nicht mehr in der richtigen Reihenfolge und ein Bild ist falsch und führt nicht zum Ziel.

Suche die richtigen Bilder und schreibe unter die Bilder die Zahlen von 1 bis 3!
Beginne mit 1 bei dem Startbild.

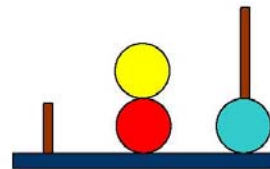
Suche die richtigen Bilder und schreibe unter die Bilder die Zahlen von 1 bis 3!

 <input type="checkbox"/>	 <input type="checkbox"/>	 <input type="checkbox"/>
 <input type="checkbox"/>		

Start



Ziel



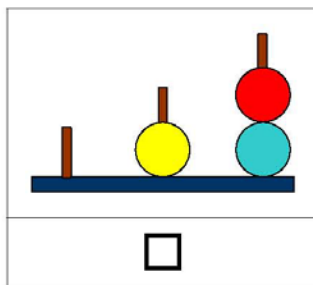
Hier unten siehst du fünf Bilder. Diese Bilder zeigen das Startbild und das Zielbild. Außerdem siehst du noch drei Bilder: Zwei Bilder, zeigen die Züge, um vom Start zum Ziel zu kommen. Ein Bild ist falsch und führt nicht zum Ziel.

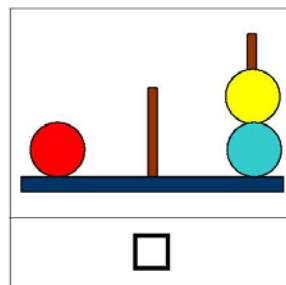
Komme in drei Zügen vom Start zum Ziel! (Der letzte Zug ist das Zielbild)

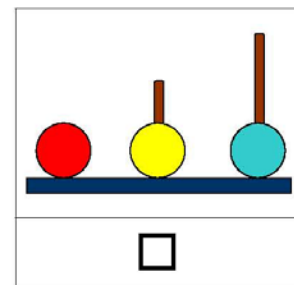
Leider sind die Bilder nicht mehr in der richtigen Reihenfolge und ein Bild ist falsch und führt nicht zum Ziel.

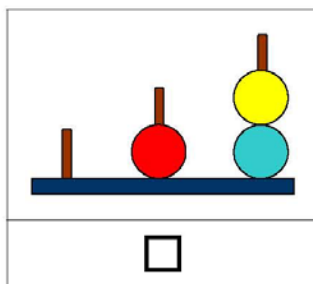
Suche die richtigen Bilder und schreibe unter die Bilder die Zahlen von 1 bis 4!
Beginne mit 1 bei dem Startbild.

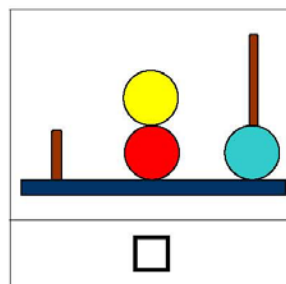
Suche die richtigen Bilder und schreibe unter die Bilder die Zahlen von 1 bis 4!



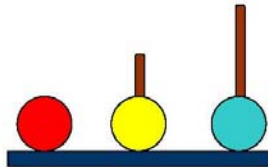




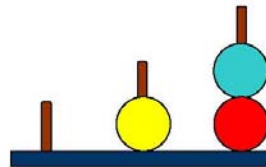




Start



Ziel



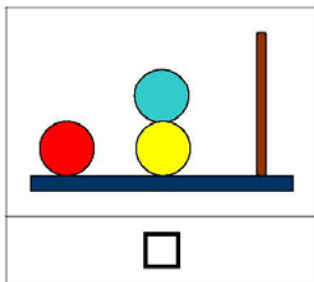
Hier unten siehst du fünf Bilder. Diese Bilder zeigen das Startbild und das Zielbild. Außerdem siehst du noch drei Bilder: Zwei Bilder, zeigen die Züge, um vom Start zum Ziel zu kommen. Ein Bild ist falsch und führt nicht zum Ziel.

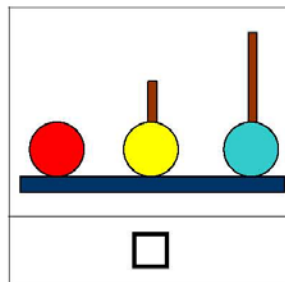
Komme in drei Zügen vom Start zum Ziel! (Der letzte Zug ist das Zielbild)

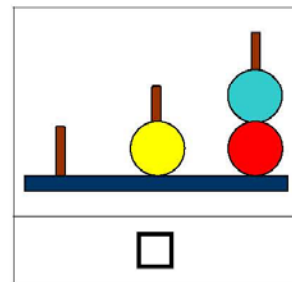
Leider sind die Bilder nicht mehr in der richtigen Reihenfolge und ein Bild ist falsch und führt nicht zum Ziel.

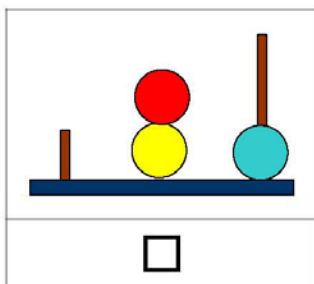
Suche die richtigen Bilder und schreibe unter die Bilder die Zahlen von 1 bis 4! Beginne mit 1 bei dem Startbild.

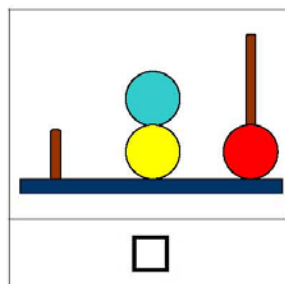
Suche die richtigen Bilder und schreibe unter die Bilder die Zahlen von 1 bis 4!



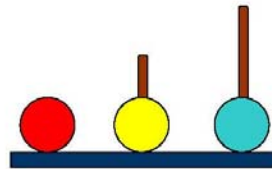




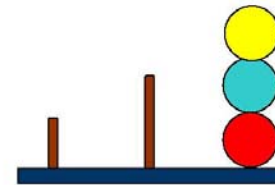




Start



Ziel



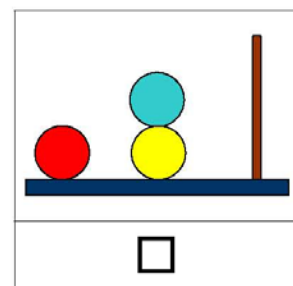
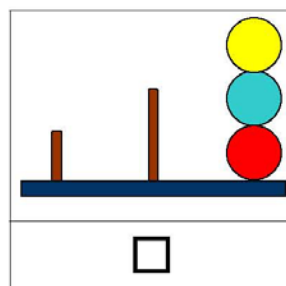
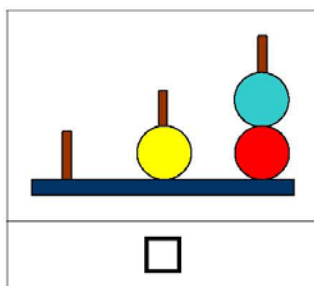
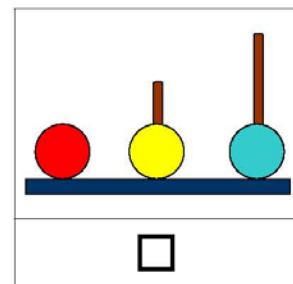
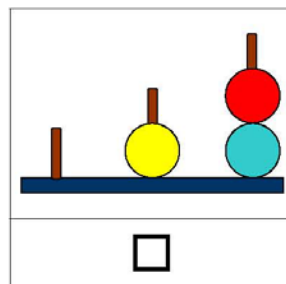
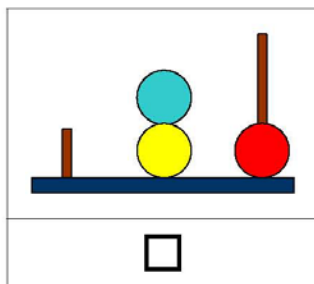
Hier unten siehst du sechs Bilder. Diese Bilder zeigen das Startbild und das Zielbild. Außerdem siehst du noch vier Bilder: Drei Bilder, zeigen die Züge, um vom Start zum Ziel zu kommen. Ein Bild ist falsch und führt nicht zum Ziel.

Komme in vier Zügen vom Start zum Ziel! (Der letzte Zug ist das Zielbild)

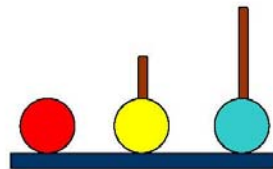
Leider sind die Bilder nicht mehr in der richtigen Reihenfolge und ein Bild ist falsch und führt nicht zum Ziel.

Suche die richtigen Bilder und schreibe unter die Bilder die Zahlen von 1 bis 5!
Beginne mit 1 bei dem Startbild.

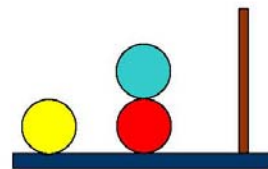
Suche die richtigen Bilder und schreibe unter die Bilder die Zahlen von 1 bis 5!



Start



Ziel



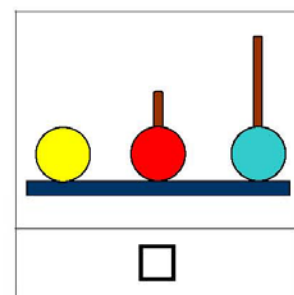
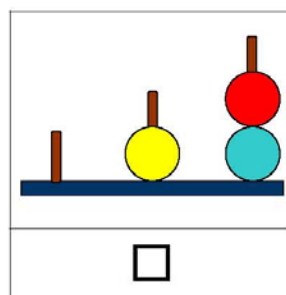
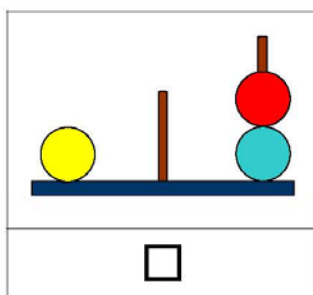
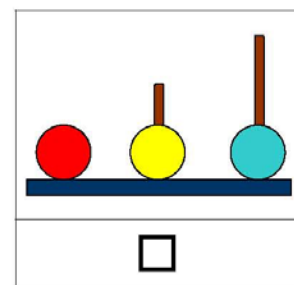
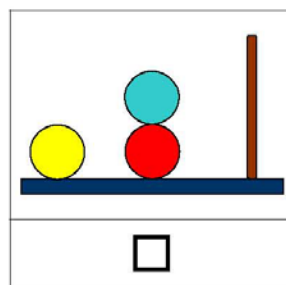
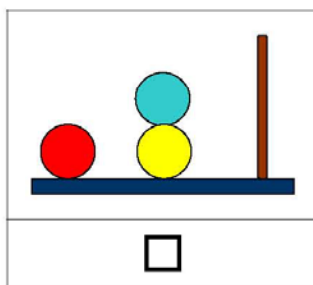
Hier unten siehst du sechs Bilder. Diese Bilder zeigen das Startbild und das Zielbild. Außerdem siehst du noch vier Bilder: Drei Bilder, zeigen die Züge, um vom Start zum Ziel zu kommen. Ein Bild ist falsch und führt nicht zum Ziel.

Komme in vier Zügen vom Start zum Ziel! (Der letzte Zug ist das Zielbild)

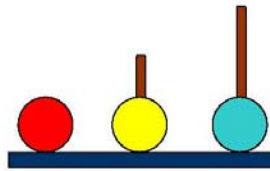
Leider sind die Bilder nicht mehr in der richtigen Reihenfolge und ein Bild ist falsch und führt nicht zum Ziel.

Suche die richtigen Bilder und schreibe unter die Bilder die Zahlen von 1 bis 5! Beginne mit 1 bei dem Startbild.

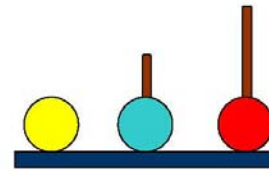
Suche die richtigen Bilder und schreibe unter die Bilder die Zahlen von 1 bis 5!



Start



Ziel



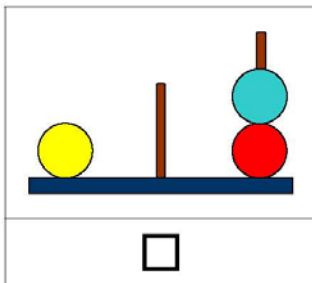
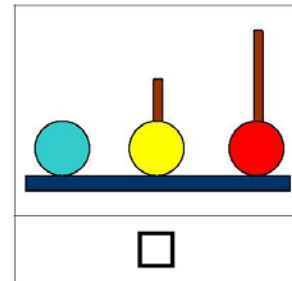
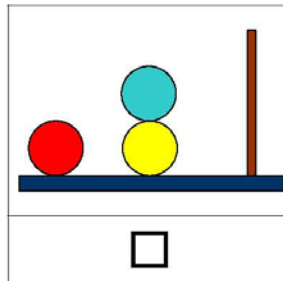
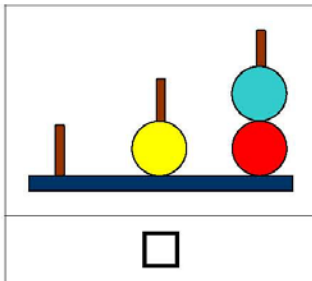
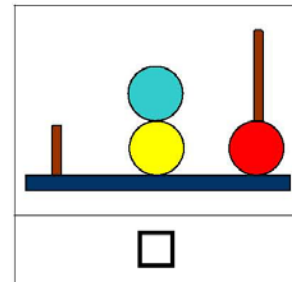
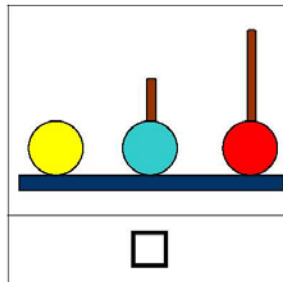
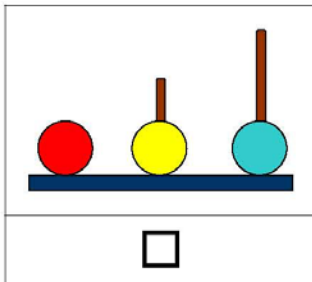
Hier unten siehst du sieben Bilder. Diese Bilder zeigen das Startbild und das Zielbild. Außerdem siehst du noch fünf Bilder: Vier Bilder, zeigen die Züge, um vom Start zum Ziel zu kommen. Ein Bild ist falsch und führt nicht zum Ziel.

Komme in fünf Zügen vom Start zum Ziel! (Der letzte Zug ist das Zielbild)

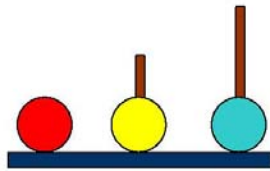
Leider sind die Bilder nicht mehr in der richtigen Reihenfolge und ein Bild ist falsch und führt nicht zum Ziel.

Suche die richtigen Bilder und schreibe unter die Bilder die Zahlen von 1 bis 6!
Beginne mit 1 bei dem Startbild.

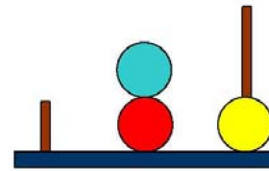
Suche die richtigen Bilder und schreibe unter die Bilder die Zahlen von 1 bis 6!



Start



Ziel



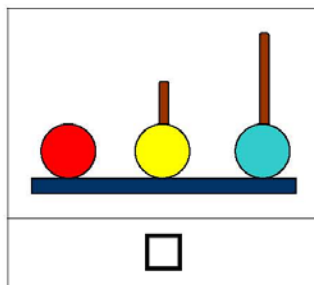
Hier unten siehst du sieben Bilder. Diese Bilder zeigen das Startbild und das Zielbild. Außerdem siehst du noch fünf Bilder: Vier Bilder, zeigen die Züge, um vom Start zum Ziel zu kommen. Ein Bild ist falsch und führt nicht zum Ziel.

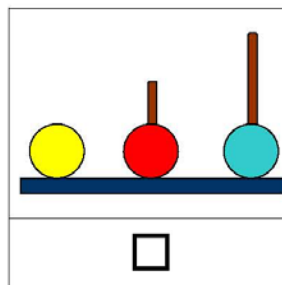
Komme in fünf Zügen vom Start zum Ziel! (Der letzte Zug ist das Zielbild)

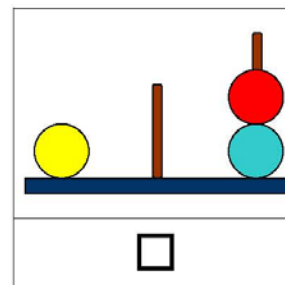
Leider sind die Bilder nicht mehr in der richtigen Reihenfolge und ein Bild ist falsch und führt nicht zum Ziel.

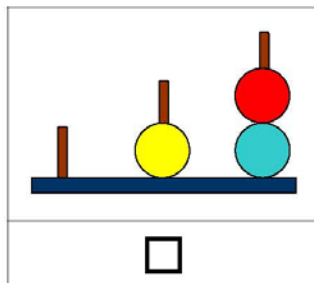
Suche die richtigen Bilder und schreibe unter die Bilder die Zahlen von 1 bis 6! Beginne mit 1 bei dem Startbild.

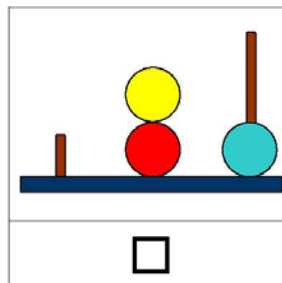
Suche die richtigen Bilder und schreibe unter die Bilder die Zahlen von 1 bis 6!

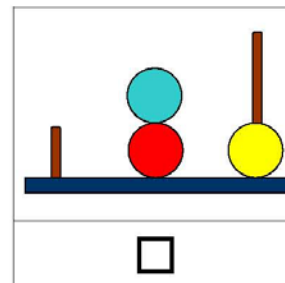


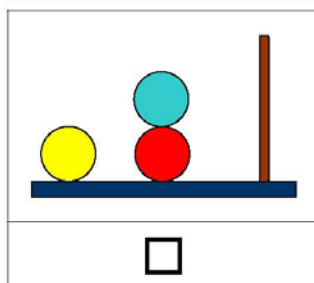




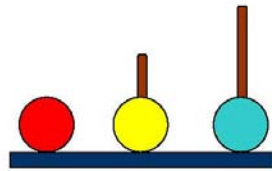




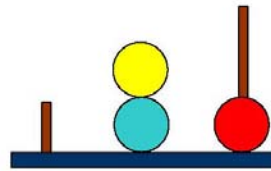




Start



Ziel



Hier unten siehst du acht Bilder. Diese Bilder zeigen das Startbild und das Zielbild. Außerdem siehst du noch sechs Bilder: Fünf Bilder, zeigen die Züge, um vom Start zum Ziel zu kommen. Ein Bild ist falsch und führt nicht zum Ziel.

Komme in sechs Zügen vom Start zum Ziel! (Der letzte Zug ist das Zielbild)

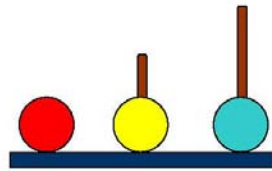
Leider sind die Bilder nicht mehr in der richtigen Reihenfolge und ein Bild ist falsch und führt nicht zum Ziel.

Suche die richtigen Bilder und schreibe unter die Bilder die Zahlen von 1 bis 7!
Beginne mit 1 bei dem Startbild.

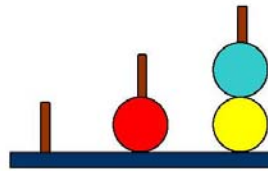
Suche die richtigen Bilder und schreibe unter die Bilder die Zahlen von 1 bis 7!

 <input type="checkbox"/>	 <input type="checkbox"/>	 <input type="checkbox"/>
 <input type="checkbox"/>	 <input type="checkbox"/>	 <input type="checkbox"/>
 <input type="checkbox"/>	 <input type="checkbox"/>	

Start



Ziel



Hier unten siehst du acht Bilder. Diese Bilder zeigen das Startbild und das Zielbild. Außerdem siehst du noch sechs Bilder: Fünf Bilder, zeigen die Züge, um vom Start zum Ziel zu kommen. Ein Bild ist falsch und führt nicht zum Ziel.

Komme in sechs Zügen vom Start zum Ziel! (Der letzte Zug ist das Zielbild)

Leider sind die Bilder nicht mehr in der richtigen Reihenfolge und ein Bild ist falsch und führt nicht zum Ziel.

Suche die richtigen Bilder und schreibe unter die Bilder die Zahlen von 1 bis 7!
Beginne mit 1 bei dem Startbild.

Suche die richtigen Bilder und schreibe unter die Bilder die Zahlen von 1 bis 7!









<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>	

A.4 Papierversion des Stroop-Tests zur Erfassung der Inhibition

Name: _____

Beispiel

Streiche die Farbe des Farbwortes an!

GELB				
BLAU				
GRÜN				
ROT				

Streiche die Farbe des Farbwortes an! Beispiel:

GELB	■ ■ ■ ■
------	---

GELB	ROT	GRÜN
ROT	BLAU	ROT
GRÜN	GRÜN	BLAU
GELB	ROT	GELB
BLAU	GELB	ROT
GRÜN	BLAU	GRÜN
ROT	GRÜN	GELB
BLAU	GELB	BLAU
GRÜN	ROT	GRÜN
GELB	GRÜN	ROT
ROT	GELB	GELB
GRÜN	BLAU	GRÜN
BLAU	ROT	ROT
ROT	GRÜN	BLAU
GELB	BLAU	GELB
GRÜN	GELB	GRÜN
ROT	GRÜN	BLAU
BLAU	BLAU	ROT
GELB	ROT	GRÜN
GRÜN	GELB	GELB
BLAU	GRÜN	BLAU
ROT	ROT	ROT
GELB	BLAU	GELB
BLAU	GELB	BLAU

2

Beispiel

Streiche die Farbe des Wortes an!

GELB	
BLAU	
ROT	
GELB	

Streiche die Farbe des Wortes an! Beispiel: GELB
■ ■ ■ ■

ROT	■ ■ ■ ■	ROT	■ ■ ■ ■	BLAU	■ ■ ■ ■
BLAU	■ ■ ■ ■	GRÜN	■ ■ ■ ■	GRÜN	■ ■ ■ ■
ROT	■ ■ ■ ■	GELB	■ ■ ■ ■	ROT	■ ■ ■ ■
GRÜN	■ ■ ■ ■	BLAU	■ ■ ■ ■	BLAU	■ ■ ■ ■
GELB	■ ■ ■ ■	GRÜN	■ ■ ■ ■	GELB	■ ■ ■ ■
ROT	■ ■ ■ ■	ROT	■ ■ ■ ■	GRÜN	■ ■ ■ ■
BLAU	■ ■ ■ ■	BLAU	■ ■ ■ ■	ROT	■ ■ ■ ■
GRÜN	■ ■ ■ ■	GELB	■ ■ ■ ■	GELB	■ ■ ■ ■
GELB	■ ■ ■ ■	GRÜN	■ ■ ■ ■	GRÜN	■ ■ ■ ■
ROT	■ ■ ■ ■	ROT	■ ■ ■ ■	ROT	■ ■ ■ ■
GELB	■ ■ ■ ■	BLAU	■ ■ ■ ■	GELB	■ ■ ■ ■
BLAU	■ ■ ■ ■	GELB	■ ■ ■ ■	BLAU	■ ■ ■ ■
ROT	■ ■ ■ ■	ROT	■ ■ ■ ■	GRÜN	■ ■ ■ ■
GELB	■ ■ ■ ■	GRÜN	■ ■ ■ ■	GELB	■ ■ ■ ■
GELB	■ ■ ■ ■	GELB	■ ■ ■ ■	GRÜN	■ ■ ■ ■
GRÜN	■ ■ ■ ■	BLAU	■ ■ ■ ■	BLAU	■ ■ ■ ■
BLAU	■ ■ ■ ■	ROT	■ ■ ■ ■	ROT	■ ■ ■ ■
ROT	■ ■ ■ ■	GELB	■ ■ ■ ■	GRÜN	■ ■ ■ ■
GELB	■ ■ ■ ■	BLAU	■ ■ ■ ■	BLAU	■ ■ ■ ■
GRÜN	■ ■ ■ ■	GRÜN	■ ■ ■ ■	ROT	■ ■ ■ ■
ROT	■ ■ ■ ■	ROT	■ ■ ■ ■	GELB	■ ■ ■ ■
GRÜN	■ ■ ■ ■	GELB	■ ■ ■ ■	BLAU	■ ■ ■ ■
ROT	■ ■ ■ ■	ROT	■ ■ ■ ■	GELB	■ ■ ■ ■
GRÜN	■ ■ ■ ■	GRÜN	■ ■ ■ ■	ROT	■ ■ ■ ■

4

Beispiel

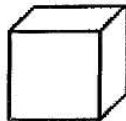
Streiche die Farbe des Farbwortes an!

GELB	   
BLAU	   
ROT	   
GELB	   

Streiche die Farbe des Farbwortes an! Beispiel: GELB 	
ROT	
BLAU	
ROT	
GRÜN	
GELB	
ROT	
BLAU	
GRÜN	
GELB	
ROT	
GELB	
BLAU	
ROT	
GELB	
GELB	
GRÜN	
BLAU	
ROT	
GELB	
GRÜN	
ROT	
BLAU	
ROT	
GRÜN	
ROT	
GRÜN	
ROT	
GRÜN	
BLAU	
GELB	
BLAU	
GRÜN	
ROT	
BLAU	
GELB	
BLAU	
GRÜN	
ROT	
BLAU	
GELB	
BLAU	
GRÜN	
ROT	
BLAU	
GELB	
BLAU	
GRÜN	
ROT	
BLAU	
GELB	
BLAU	
GRÜN	
ROT	
BLAU	
GELB	
BLAU	
GRÜN	
ROT	
BLAU	
GELB	
BLAU	
GRÜN	
ROT	
BLAU	
GELB	

A.5 Aufgaben zur Erfassung des räumlichen Denkens

Hier siehst du einen Würfel, der aus Papier gefaltet und mit Klebeband
zusammengeklebt wurde. Jetzt wird das Klebeband entfernt.
Unten siehst du vier Möglichkeiten, wie das gefaltete Papier ohne Klebeband
aussehen könnte, wenn es flach hingelegt wird.
Nur eine Faltmöglichkeit ist richtig und ergibt gefaltet einen Würfel.

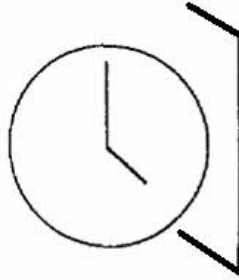


Welche Antwort ist richtig und zeigt ein Papier, das zu einem Würfel
zusammengefalted werden kann?

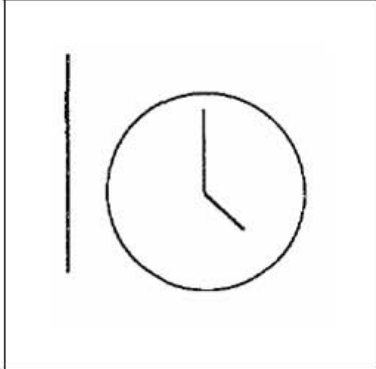
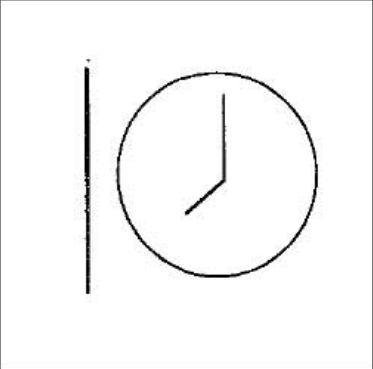
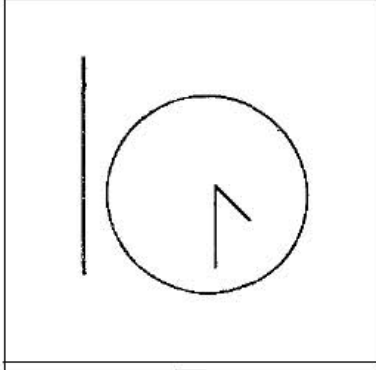
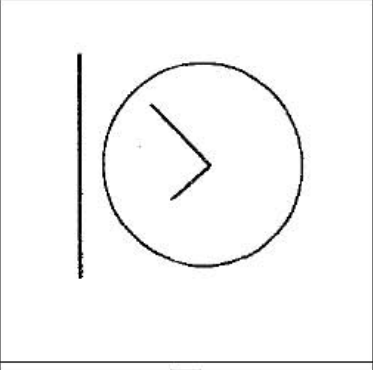
Kreuze an!

<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>

Hier siehst du eine Zeichnung. Neben der Zeichnung steht ein Spiegel. Daher entsteht ein Spiegelbild von der Zeichnung.
 Unten siehst du vier Möglichkeiten, wie das Spiegelbild der Zeichnung aussehen könnte.
 Nur eine Möglichkeit ist richtig und zeigt das Spiegelbild der Zeichnung.



Welche Antwort ist richtig und zeigt das Spiegelbild der Zeichnung?
Kreuze an!

	
<input type="checkbox"/>	<input type="checkbox"/>
	
<input type="checkbox"/>	<input type="checkbox"/>

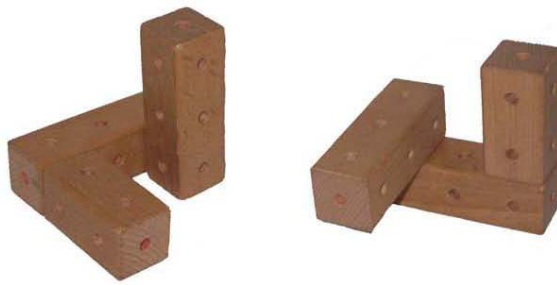
Hier siehst du in jedem Bild zwei Figuren.
Die beiden Figuren können gleich und gedreht sein. Die beiden Figuren können aber auch nicht gleich sein.

Prüfe, ob die beiden Figuren gleich sind.

Deine Aufgabe ist es jetzt, diese Figuren in deinem Kopf zu drehen. Dann sollst du entscheiden, ob die Figuren gleich sind oder nicht.

Sind die beiden Figuren gleich oder nicht gleich?

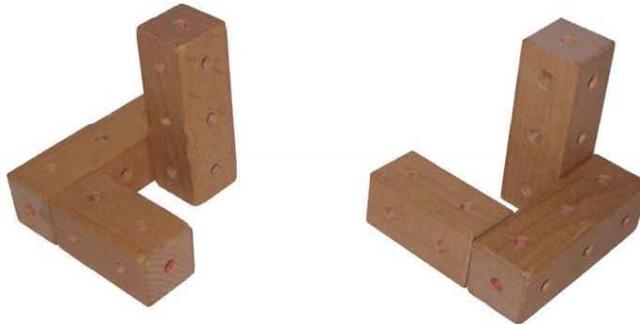
Kreuze an!



- gleich
- nicht gleich

Sind die beiden Figuren gleich oder nicht gleich?

Kreuze an!



- gleich
- nicht gleich

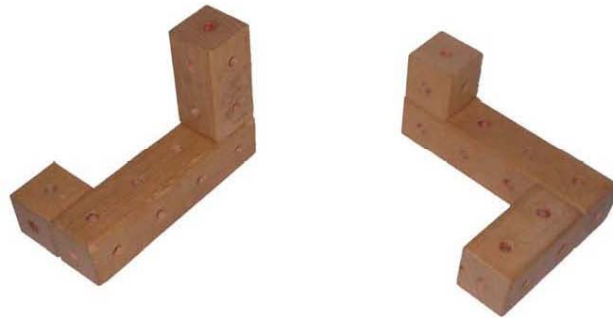
Hier siehst du in jedem Bild zwei Figuren.
Die beiden Figuren können gleich und gedreht sein. Die beiden Figuren können aber auch nicht gleich sein.

Prüfe, ob die beiden Figuren gleich sind.

Deine Aufgabe ist es jetzt, diese Figuren in deinem Kopf zu drehen. Dann sollst du entscheiden, ob die Figuren gleich sind oder nicht.

Sind die beiden Figuren gleich oder nicht gleich?

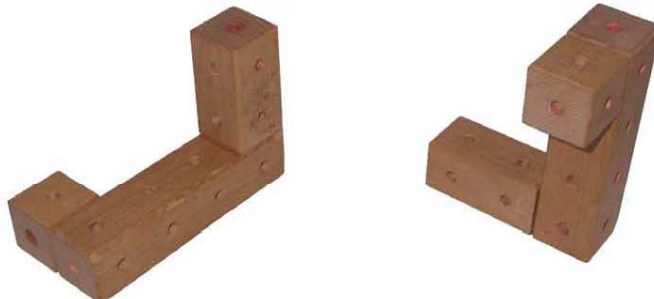
Kreuze an!



- gleich
- nicht gleich

Sind die beiden Figuren gleich oder nicht gleich?

Kreuze an!



- gleich
- nicht gleich

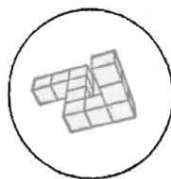
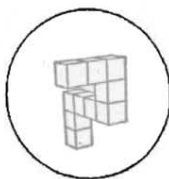
Hier siehst du in jedem Bild zwei Figuren. Die beiden Figuren können gleich und gedreht sein. Die beiden Figuren können aber auch nicht gleich sein.

Prüfe, ob die beiden Figuren gleich sind.

Deine Aufgabe ist es jetzt, diese Figuren in deinem Kopf zu drehen. Dann sollst du entscheiden, ob die Figuren gleich sind oder nicht.

Sind die beiden Figuren gleich oder nicht gleich?

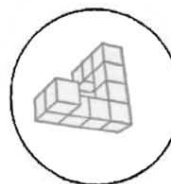
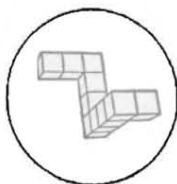
Kreuze an!



- gleich
 nicht gleich

Sind die beiden Figuren gleich oder nicht gleich?

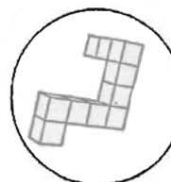
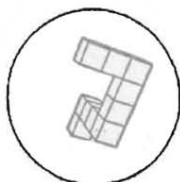
Kreuze an!



- gleich
 nicht gleich

Sind die beiden Figuren gleich oder nicht gleich?

Kreuze an!



- gleich
 nicht gleich

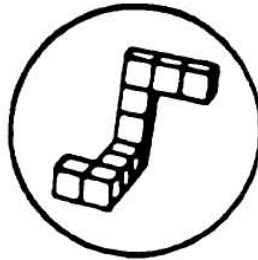
Hier siehst du eine Figur.

Prüfe, ob die Figuren unten gleich sind wie die Figur oben.

Die Figuren unten können gleich und gedreht sein. Die Figuren unten können aber auch ganz andere Figuren sein als die Figur oben.

Deine Aufgabe ist es, diese Figuren in deinem Kopf zu drehen. Dann sollst du entscheiden, ob die Figuren unten gleich sind wie die Figur oben oder nicht.

Bei dieser Aufgabe gibt es zwei richtige Antworten.



Welche zwei Figuren unten sind gleich wie die Figur oben?

Kreuze die zwei Antworten an!

<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>

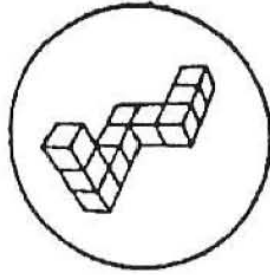
Hier siehst du eine Figur.

Prüfe, ob die Figuren unten gleich sind wie die Figur oben.

Die Figuren unten können gleich und gedreht sein. Die Figuren unten können aber auch ganz andere Figuren sein als die Figur oben.

Deine Aufgabe ist es, diese Figuren in deinem Kopf zu drehen. Dann sollst du entscheiden, ob die Figuren unten gleich sind wie die Figur oben oder nicht.

Bei dieser Aufgabe gibt es zwei richtige Antworten.



Welche zwei Figuren unten sind gleich wie die Figur oben?
Kreuze die zwei Antworten an!

<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>

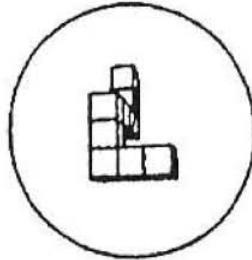
Hier siehst du eine Figur.

Prüfe, ob die Figuren unten gleich sind wie die Figur oben.

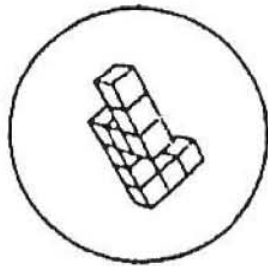
Die Figuren unten können gleich und gedreht sein. Die Figuren unten können aber auch ganz andere Figuren sein als die Figur oben.

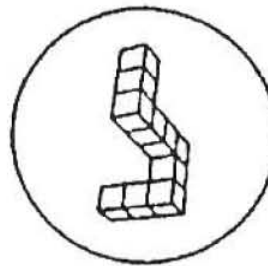
Deine Aufgabe ist es, diese Figuren in deinem Kopf zu drehen. Dann sollst du entscheiden, ob die Figuren unten gleich sind wie die Figur oben oder nicht.

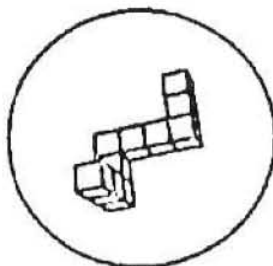
Bei dieser Aufgabe gibt es zwei richtige Antworten.

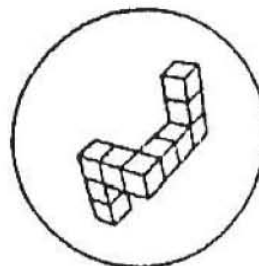


Welche zwei Figuren unten sind gleich wie die Figur oben?
Kreuze die zwei Antworten an!









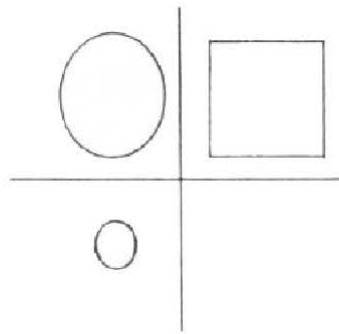
A.6 Deutsche Version des Arlin-Tests zur Erfassung formal-operatorischer Fähigkeiten

Name: _____

Zu jeder Frage in diesem Test gibt es vier verschiedene Antworten. Davon passt immer eine Antwort am besten. Kreuze die beste Antwort an. Wenn du dir bei einer Frage nicht sicher bist, welche die richtige Antwort ist, dann nimm einfach die Antwort, die dir am besten erscheint. Markiere für jede Frage nur eine Antwort. Du kannst eine Antwort auch korrigieren, mache dann in dem entsprechenden Feld ein Kreuz und einen Kreis darum. Arbeite so schnell und so genau du kannst! Für die 16 Fragen in diesem Test hast du 20 Minuten Zeit. Wenn du vor Ablauf der Zeit fertig bist, kannst du deine Antworten nochmals überprüfen.

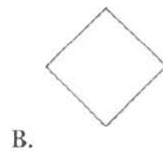
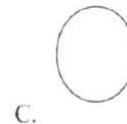
Alle Fragen in diesem Test sind so aufgebaut wie die Beispielaufgabe unten.

Beispielaufgabe

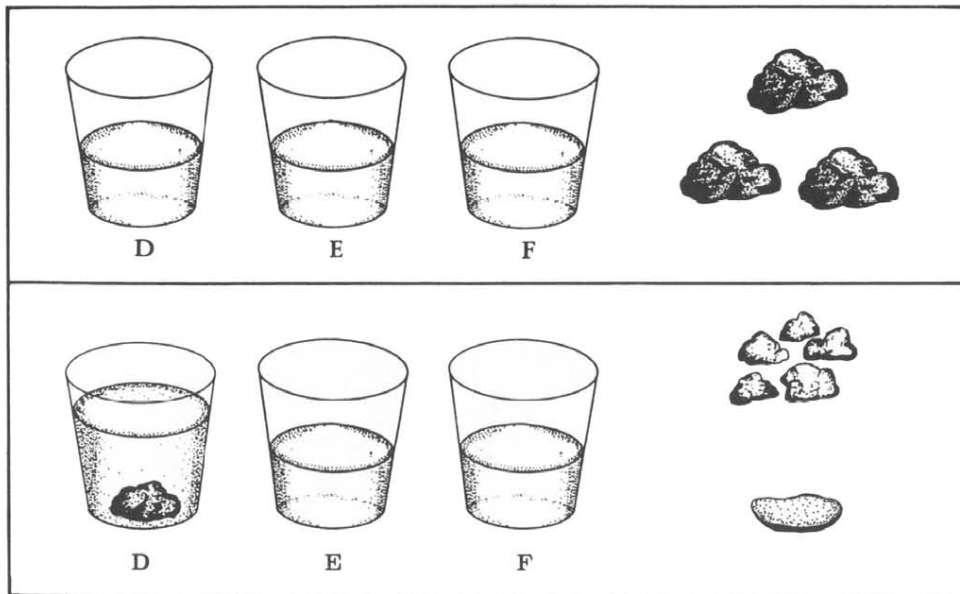


Die drei Formen, die oben abgebildet sind, ergeben ein Muster.

Welche der folgenden Formen passt am besten, um das Muster zu vervollständigen?



Das Muster beginnt mit zwei Kreisen. Ein Kreis ist groß und ein Kreis ist klein. Die zweite Spalte beginnt mit einem großen Viereck. Die richtige Antwort ist also „A“, ein kleines Viereck, weil die Form und die Größe für das Muster entscheidend sind.



Drei Becher (Becher D, Becher E und Becher F) sind zum Teil mit Wasser gefüllt. Neben den drei Bechern sind drei Kugeln Lehm. Diese drei Kugeln sind exakt gleich groß. Die erste Kugel wird, wie oben dargestellt, in den Becher D hineingegeben. Der Wasserstand im Becher D steigt. Bevor nun die zweite Kugel in den Becher E gegeben wird, wird sie zur Form eines Pfannkuchens flachgedrückt (siehe Bild). Die dritte Kugel wird, wie auf dem Bild gezeigt, in 5 Stücke geteilt und dann in den Becher F gegeben.

1. Wie denkst du, wird sich der Wasserstand im Becher E verändern, wenn die Kugel hineingegeben wird, die so flach ist wie ein Pfannkuchen?

- A. Das Wasser steigt höher als das im Becher D.
- B. Das Wasser wird bis zur Hälfte des Wasserstandes im Becher D steigen.
- C. Das Wasser wird auf den gleichen Stand wie im Becher D steigen.
- D. Das Wasser wird auf ein Fünftel des Wasserstandes von Becher D steigen.

2. Wie begründest du deine Antwort aus der ersten Frage?

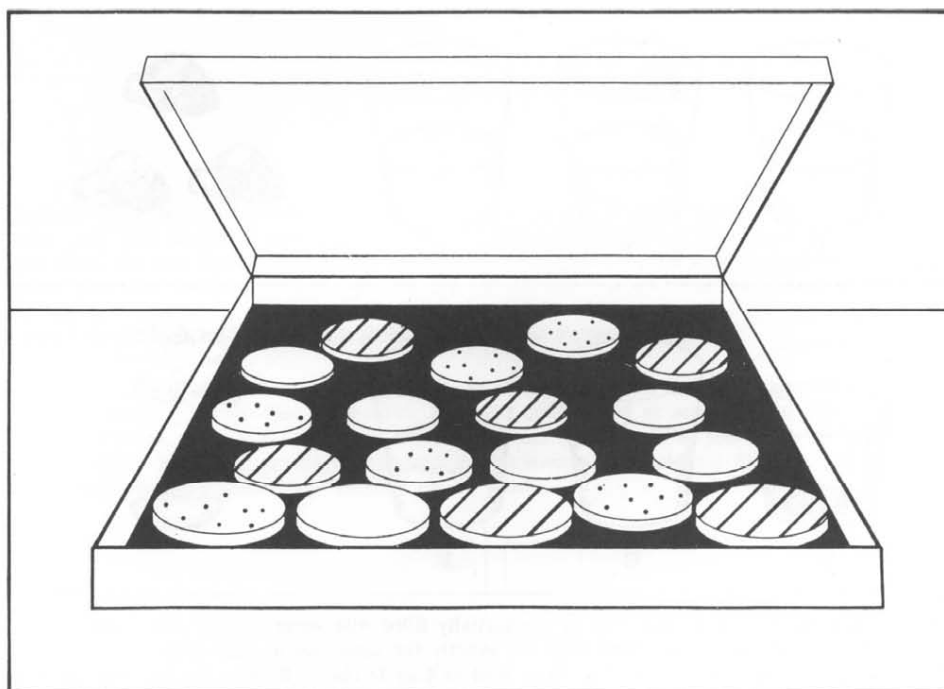
- A. Durch die Form eines Pfannkuchens benötigt die Kugel mehr Platz.
- B. Die Kugeln waren am Anfang gleich groß.
- C. Durch die flache Form eines Pfannkuchens benötigt die Kugel weniger Platz.
- D. Die Kugel und der Pfannkuchen wiegen gleich viel.

3. Wie denkst du, wird sich der Wasserstand im Becher F verändern, wenn die fünf Lehmkugeln in den Becher gegeben werden?

- A. Das Wasser wird auf den gleichen Stand wie im Becher D steigen.
- B. Das Wasser wird NICHT auf den gleichen Stand wie im Becher D steigen.
- C. Das Wasser wird höher steigen, als das Wasser im Becher D.
- D. Das Wasser wird um ein Fünftel des Wasserstandes von Becher D steigen.

4. Wie begründest du deine Antwort zur eben gestellten Frage?

- A. Die fünf Lehmkugeln benötigen mehr Platz.
- B. Die Kugeln waren alle gleich groß, bevor eine Kugel in kleine Stücke geteilt wurde.
- C. Die fünf kleinen Kugeln benötigen weniger Platz.
- D. Die fünf kleinen Kugeln wiegen genauso viel wie eine große Kugel.



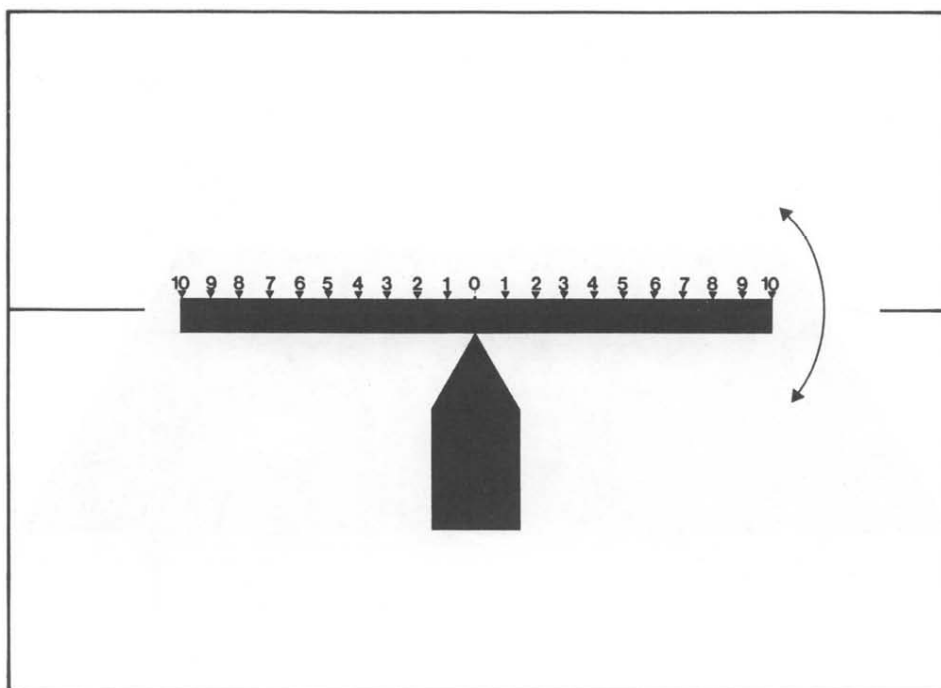
Bei einem Glücksspiel werden sechs einfarbige Spielsteine, sechs gestreifte Spielsteine und sechs gepunktete Spielsteine in eine Kiste gelegt (siehe Bild). Die Kiste wird über deinem Kopf gehalten, so dass du die Spielsteine nicht sehen kannst. Du sollst nun einen der Spielsteine aus der Kiste holen.

5. Wie hoch ist deiner Meinung nach die Chance, dass du bei deinem ersten Versuch einen gestreiften Spielstein ziehst?

- A. Eins zu zwei.
- B. Eins zu achtzehn.
- C. Eins zu zwölf.
- D. Eins zu drei.

6. Wie begründest du deine Antwort zu der oben genannten Frage?

- A. Meine Chancen sind so, als ob ich eine Münze werfen und das Wappen bekommen würde.
- B. Meine Chancen hängen von der Tatsache ab, dass die Anzahl der gestreiften Spielsteine mit der Gesamtzahl der Spielsteine verglichen werden muss.
- C. Meine Chancen, einen gestreiften Stein gleich in den ersten zwei oder drei Versuchen zu ziehen, stehen gut, denn ich bin ein Glückspilz.
- D. Meine Chancen hängen von der Tatsache ab, dass zwölf nicht gestreifte Spielsteine vorhanden sind und ich diese zuerst ausschließen muss.



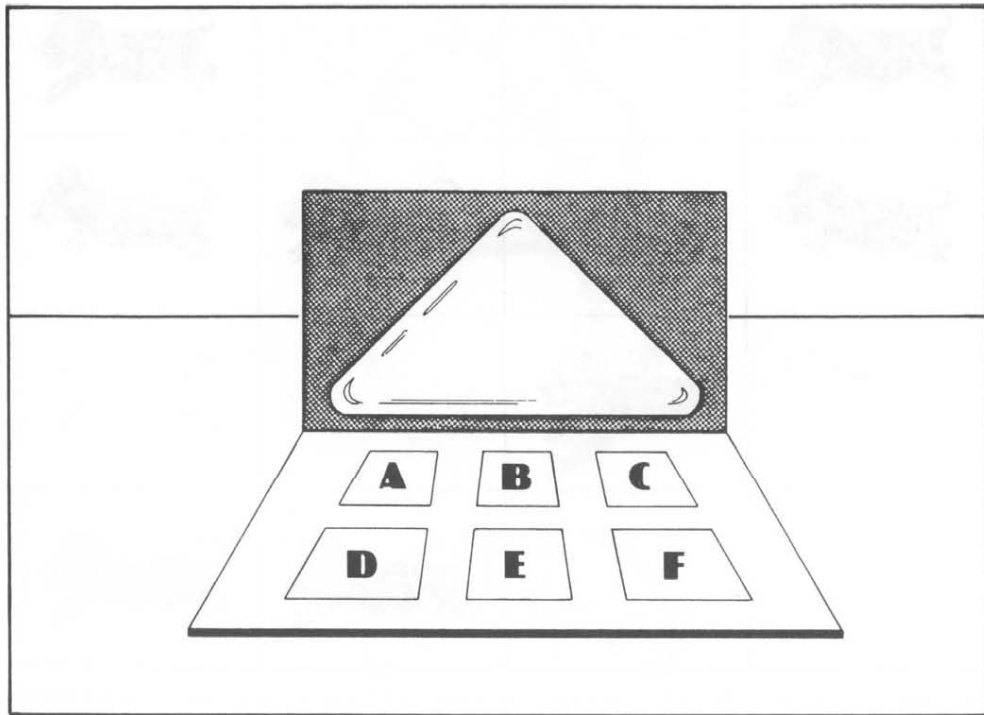
Eine Gruppe von Kindern spielt im Park an einer Wippe. Wenn ZWEI Kinder, die beide in etwa 20 Kilogramm wiegen, links vom Balancepunkt auf dem Platz sitzen, der mit der Markierung 5 gekennzeichnet ist, kippt die Wippe nach links. Um die Wippe wieder ins Gleichgewicht zu bringen, müssen zwei Kinder, die beide in etwa 20 Kilogramm wiegen, rechts vom Balancepunkt auf dem Platz sitzen, der mit der Markierung 5 gekennzeichnet ist. Die nachfolgenden Fragen beziehen sich auf die Wippe und die Markierung, die links und rechts davon angegeben sind.

7. Wenn nun DREI Kinder, die alle ungefähr 20 Kilogramm wiegen, an der Markierung 5 links vom Gleichgewichtspunkt sitzen, wo müssen die ZWEI Kinder auf der rechten Seite sitzen, um die Wippe wieder ins Gleichgewicht zu bringen? (Diese beiden Kinder müssen nicht unbedingt auf der gleichen Markierung sitzen.)

- A. Die zwei Kinder auf der rechten Seite können das Gleichgewicht NICHT wiederherstellen.
- B. Beide Kinder auf der rechten Seite müssen an der Markierung 10 sitzen.
- C. Ein Kind muss an der Markierung 4 sitzen und das andere Kind muss an der Markierung 6 sitzen.
- D. Ein Kind muss an der Markierung 7 sitzen und das andere Kind muss an der Markierung 8 sitzen.

8. Wenn nun VIER Kinder, die alle ungefähr 20 Kilogramm wiegen, an der Markierung 5 links vom Gleichgewichtspunkt sitzen, wo müssen die ZWEI Kinder auf der rechten Seite sitzen, um die Wippe wieder ins Gleichgewicht zu bringen? (Diese beiden Kinder müssen nicht unbedingt auf der gleichen Markierung sitzen.)

- A. Die zwei Kinder auf der rechten Seite müssen an der Markierung 5 sitzen.
- B. Die zwei Kinder auf der rechten Seite müssen an der Markierung 10 sitzen.
- C. Ein Kind muss an der Markierung 6 sitzen und das andere Kind an der Markierung 9.
- D. Ein Kind muss an der Markierung 8 sitzen und ein Kind an der Markierung 10.



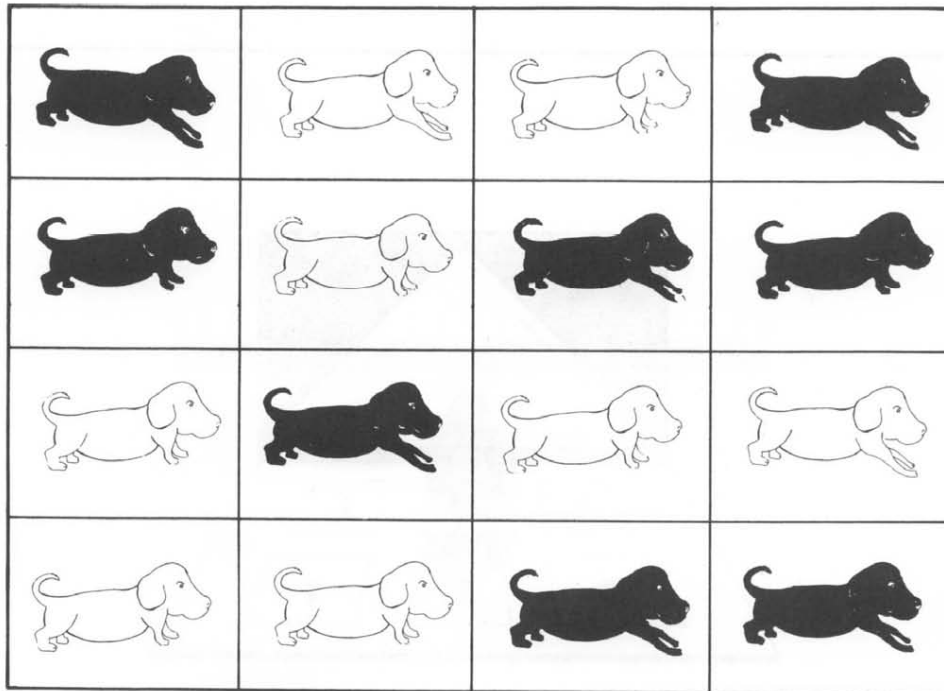
Ein neues Computerspiel ist auf dem Markt. Ziel des Spiels ist es, das Dreieck oberhalb des Spielbretts zum Leuchten zu bringen. Das Licht leuchtet auf, wenn eine oder mehrere der Tasten gleichzeitig gedrückt werden.

9. Wie kannst du herausfinden, welche der Tasten das Licht zum Leuchten bringen, wenn sie gleichzeitig gedrückt werden?

- A. Man kann alle möglichen Tastenpaare drücken, die das Licht zum Leuchten bringen.
- B. Man kann jede Taste einzeln und dann zwei auf einmal drücken, bis das Dreieck leuchtet.
- C. Man kann alle sechs auf einmal und anschließend fünf auf einmal, vier auf einmal und so weiter drücken, bis das Licht angeht.
- D. Man kann alle sechs Tasten auf einmal, fünf auf einmal, vier auf einmal oder drei auf einmal drücken.

10. Wie begründest du deine Antwort?

- A. Um die Aufgabe zu lösen, müssen alle Kombinationen getestet werden, angefangen bei nur einer gedrückten Taste bis hin zu sechs Tasten auf einmal.
- B. Das Wort „Kombination“ setzt ein Paar oder zwei gleichzeitig voraus.
- C. Die Angaben in der Aufgabe erfordern es, dass alle Kombinationen von drei bis sechs gleichzeitig getestet werden.
- D. Mit den vorhandenen Informationen kann die Aufgabe NICHT gelöst werden, selbst wenn man das Spiel vor sich hätte und ausprobieren könnte.



Du bekommst 16 Karten. Auf jeder dieser Karten ist ein Hund abgebildet, entweder ein schwarzer oder ein weißer, mit langen oder kurzen Beinen. Die Karte Nr. 1 zeigt einen schwarzen Hund mit langen Beinen.

Die nachfolgenden Fragen sind auf der Grundlage dieser 16 Karten zu beantworten.

11. Kannst du anhand dieser 16 Karten feststellen, ob zwischen der Fellfarbe und der Länge der Beine dieser Hunde ein Zusammenhang besteht?

- A. Nein, denn es gibt genauso viele schwarze wie weiße Hunde mit kurzen Beinen.
- B. Nein, denn 8 Hunde haben kurze Beine und 8 Hunde haben lange Beine, deshalb besteht kein Zusammenhang.
- C. Ja, denn alle schwarzen Hunde haben kurze Beine.
- D. Ja, denn die meisten schwarzen Hunde haben lange Beine und die meisten weißen Hunde kurze.

12. Wie groß ist anhand der Karten die Wahrscheinlichkeit, dass ein schwarzer Hund lange Beine hat?

- A. Sechs zu acht
- B. Vier zu acht
- C. Eins zu vier
- D. Neun zu sechzehn

13. Wie groß ist anhand der Karten die Wahrscheinlichkeit, dass ein weißer Hund lange Beine hat?

- A. Eins zu sechs
- B. Eins zu acht
- C. Zwei zu acht
- D. Eins zu sechzehn

14. Wie groß ist anhand der Karten die Wahrscheinlichkeit, dass ein schwarzer Hund kurze Beine hat?

- A. Zwei zu acht
- B. Drei zu acht
- C. Drei zu sechzehn
- D. Das ist gar nicht möglich.



Zwei Personen sitzen in diesem Zug, der gerade durch einen Tunnel in einem Berg fährt. Herr Rot sitzt an der Spitze des Zuges und Herr Blau sitzt am Ende des Zuges. Entscheide bei den folgenden Situationen, ob Herr Rot oder Herr Blau über den gleichen Zeitraum hinweg im Tunnel sind.

15. SITUATION 1:

Nachdem der Zug in den Tunnel gefahren ist, verlässt Herr Rot seinen Platz an der Spitze des Zuges und geht zurück, um bei Herrn Blau am Ende des Zuges zu sitzen. Wie viel Zeit verbringt Herr Rot insgesamt im Tunnel?

- A. Weniger Zeit als Herr Blau.
- B. Zweimal so viel Zeit wie Herr Blau.
- C. Genauso viel Zeit wie Herr Blau.
- D. Mehr Zeit als Herr Blau.

16. SITUATION 2:

Nachdem der Zug in den Tunnel gefahren ist, verlässt Herr Blau seinen Platz am Ende des Zuges. Er geht nach vorne, um bei Herrn Rot an der Spitze des Zuges zu sitzen. Auf halber Strecke, beschließt er, zurück zu seinem Platz zu gehen, um seine Zeitung zu holen. Er holt die Zeitung, geht wieder nach vorne und setzt sich zu Herrn Rot, während der Zug immer noch im Tunnel ist. Wie viel Zeit hat Herr Blau im Tunnel verbracht?

- A. Weniger Zeit als Herr Rot.
- B. Mehr Zeit als Herr Rot.
- C. Eineinhalb mal so viel Zeit wie Herr Rot.
- D. Genauso viel Zeit wie Herr Rot.

A.7 Deutsche Version des ICA-R Inventory of Childrens' Activities – Revised⁹ zur Erfassung des Interesses an forschenden Aktivitäten

Lies dir bitte die folgenden Aussagen durch.
 Wenn du einer Aussage zustimmst, dann setze ein Kreuz in den Kreis bei der Antwort, die für dich stimmt:

Bitte beantworte alle Fragen ehrlich. Es gibt keine richtigen oder falschen Antworten, alleine deine Meinung ist wichtig!

Wie gerne magst du ... ?	mag ich gar nicht	mag ich nicht	mag ich	mag ich sehr
1. Etwas bauen	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
2. Verstehen, wie Dinge funktionieren	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
3. Bilder malen	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
4. Mit Freunden reden	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
5. Dinge verkaufen	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
6. Rechnen	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
7. Nägel mit dem Hammer schlagen	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
8. Dinge auseinander bauen	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
9. Musik hören	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
10. Jemandem Lesen beibringen	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
11. Anführer einer Gruppe sein	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
12. Ordnung halten und aufräumen	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>

⁹ Skala Interesse an forschenden Aktivitäten: Item 2, Item 8, Item 14, Item 20, Item 26

Wie gerne magst du ... ?	mag ich gar nicht	mag ich nicht	mag ich	mag ich sehr
13. Zuschauen, wie etwas gebaut wird	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
14. Wissenschaftssendungen anschauen	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
15. Sich eine Geschichte ausdenken	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
16. Anderen helfen	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
17. Anderen sagen, was sie machen sollen	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
18. Dinge zählen und sortieren	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
19. Ein Spielzeug reparieren	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
20. In ein Mikroskop schauen	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
21. Mit Ton und Knetmasse spielen	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
22. Kranken Menschen helfen	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
23. Sich für andere Spiele ausdenken	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
24. Tabellen und Diagramme erstellen	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
25. Zuschauen, wie ein Gerät repariert wird	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>

Wie gerne magst du ... ?	mag ich gar nicht	mag ich nicht	mag ich	mag ich sehr
26. Sachen zusammenmischen, um zu sehen, was passiert	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
27. Zu Musik singen	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
28. Auf kleine Kinder aufpassen	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
29. Deine Eltern zu etwas überreden, was du willst	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
30. Einen Schrank aufräumen	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
31. Schularbeiten	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
32. Sport	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
33. Freundschaften schließen	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
34. Alleine arbeiten	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
35. Mit anderen zusammen arbeiten	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
36. Alleine spielen	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
37. Mit anderen zusammen spielen	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
38. Dinge, die Jungen mögen	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>
39. Dinge, die Mädchen mögen	mag ich gar nicht <input type="radio"/>	mag ich nicht <input type="radio"/>	mag ich <input type="radio"/>	mag ich sehr <input type="radio"/>

Anhang B

Ergebnisse

B.1 Skalendokumentation

B.2 Latente einfache Regressionsanalysen

B.1 Skalendokumentation

Tabelle B.1

ELFE 1-6 – Subtest Textverständnis: Statistische Kennwerte der eindimensionalen Rasch-Skalierung

Item	N	Lösungs- häufigkeit (%)	Schwierig- keit	Trenn- schärfe	Infit MNSQ	T-Wert
1	271	97.42	-3.68	0.28	0.86	-0.4
2	271	86.72	-1.46	0.43	1.23	1.8
3	271	82.29	-0.97	0.42	1.31	2.7
4	271	89.30	-1.82	0.34	1.27	1.9
5	271	92.25	-2.30	0.30	1.21	1.2
6	271	92.25	-2.30	0.47	0.85	-0.9
7	271	86.35	-1.43	0.54	0.93	-0.5
8	271	73.80	-0.20	0.60	1.07	0.8
9	271	76.01	-0.39	0.64	0.91	-1.0
10	271	71.96	-0.05	0.66	0.94	-0.7
11	271	77.49	-0.52	0.61	0.97	-0.3
12	271	52.40	1.40	0.53	1.42	4.5
13	271	64.94	0.50	0.67	0.98	-0.2
14	271	59.78	0.87	0.73	0.88	-1.5
15	271	57.93	1.01	0.75	0.83	-2.1
16	271	50.18	1.56	0.79	0.71	-3.8
17	271	48.34	1.69	0.78	0.71	-3.8
18	271	41.33	2.20	0.70	0.87	-1.5
19	271	35.79	2.62	0.67	0.86	-1.5
20	271	27.68	3.28	0.62	0.86	-1.5
Skala	EAP/PV-Reliabilität: 0.90 Varianz: 4.41 Trennschärfe: $M = 0.58$, $SD = 0.16$, $MIN = 0.28$, $MAX = 0.79$ Schwierigkeit: $M = 0$, $SD = 1.85$, $MIN = -3.68$, $MAX = 3.28$ Infit: $M = 0.98$, $SD = 0.20$, $MIN = 0.71$, $MAX = 1.42$					
Stufe 2-3	EAP/PV-Reliabilität: 0.89 Varianz: 4.14					
Stufe 4	EAP/PV-Reliabilität: 0.85 Varianz: 3.45					

Tabelle B.2

CFT 20-R – Subtest 1 und 4: Statistische Kennwerte der eindimensionalen Rasch-Skalierung

Item	N	Lösungs- häufigkeit (%)	Schwieri- gkeit	Trenn- schärfe	Infit MNSQ	T-Wert
1	272	93.75	-2.61	0.33	0.97	-0.1
2	272	94.85	-2.82	0.20	1.01	0.1
3	272	95.96	-3.09	0.28	0.99	0.0
4	272	75.74	-0.94	0.27	1.03	0.4
5	272	79.04	-1.15	0.39	0.96	-0.5
6	272	81.25	-1.30	0.42	0.95	-0.5
7	272	75.74	-0.94	0.47	0.94	-0.8
8	272	55.51	0.06	0.19	1.11	2.8
9	272	54.04	0.12	0.38	1.00	0.1
10	272	59.56	-0.12	0.32	1.03	0.7
11	272	59.93	-0.14	0.42	0.96	-0.8
12	272	59.93	-0.14	0.39	0.99	-0.1
13	272	26.10	1.43	0.33	0.98	-0.3
14	272	16.54	2.04	0.27	0.99	-0.1
15	272	18.01	1.93	0.16	1.05	0.5
16	271	81.18	-1.29	0.33	1.00	0.0
17	272	74.63	-0.88	0.25	1.05	0.7
18	272	36.40	0.91	0.23	1.07	1.5
19	272	55.88	0.04	0.51	0.93	-1.8
20	272	51.84	0.22	0.42	0.98	-0.5
21	272	58.82	-0.09	0.39	0.97	-0.8
22	272	30.88	1.18	0.34	0.97	-0.5
23	272	55.51	0.06	0.34	1.02	0.4
24	272	23.16	1.60	0.31	1.00	0.0
25	272	21.69	1.69	0.14	1.07	0.8
26	272	2.21	4.24	-0.01	0.99	0.1
	EAP/PV-Reliabilität: 0.62					
	Varianz: 0.42					
Skala	Trennschärfe: $M = 0.31$, $SD = 0.11$, $MIN = -0.01$, $MAX = 0.51$					
	Schwierigkeit: $M = 0$, $SD = 1.64$, $MIN = -3.09$, $MAX = 4.24$					
	Infit: $M = 1.00$, $SD = 0.05$, $MIN = 0.93$, $MAX = 1.11$					
Stufe 2-3	EAP/PV-Reliabilität: 0.63, Varianz: 0.44					
Stufe 4	EAP/PV-Reliabilität: 0.55, Varianz: 0.31					

Tabelle B.3

Turm von London: Statistische Kennwerte der eindimensionalen Rasch-Skalierung

Item	N	Lösungs- häufigkeit (%)	Schwierig- keit	Trenn- schärfe	Infit MNSQ	T-Wert
1	270	59.63	-2.44	0.62	1.15	1.6
2	270	56.30	-2.20	0.70	0.94	-0.7
3	270	47.78	-1.60	0.74	0.9	-1.3
4	270	30.37	-0.31	0.73	0.98	-0.2
5	270	20.00	0.62	0.66	0.95	-0.4
6	270	17.78	0.86	0.70	0.85	-1.4
7	270	10.37	1.83	0.56	0.97	-0.2
8	270	4.07	3.24	0.49	0.91	-0.3
	EAP/PV-Reliabilität: 0.79					
	Varianz: 4.60					
Skala	Trennschärfe: $M = 0.65$, $SD = 0.09$, $MIN = 0.49$, $MAX = 0.74$					
	Schwierigkeit: $M = 0$, $SD = 2.02$, $MIN = -2.44$, $MAX = 3.24$					
	Infit: $M = 0.96$, $SD = 0.09$, $MIN = 0.85$, $MAX = 1.15$					
Stufe 2-3	EAP/PV-Reliabilität: 0.69					
	Varianz: 2.85					
Stufe 4	EAP/PV-Reliabilität: 0.81					
	Varianz: 5.45					

Tabelle B.4

Stroop-Test: Statistische Kennwerte der eindimensionalen Rasch-Skalierung

Item	N	Lösungs- häufigkeit (%)	Schwierig- keit	Trenn- schärfe	Infit MNSQ	T-Wert
1	276	94.57	-2.38	0.37	1.16	0.8
2	276	93.12	-2.01	0.42	1.12	0.7
3	276	90.94	-1.55	0.32	1.49	2.6
4	276	95.65	-2.75	0.50	0.76	-1.1
5	276	96.01	-2.90	0.49	0.73	-1.1
6	276	99.28	-5.09	0.24	0.88	0.0
7	276	91.30	-1.68	0.37	1.33	1.8
8	276	92.03	-1.84	0.41	1.23	1.3
9	276	92.03	-1.84	0.41	1.23	1.2
10	276	95.65	-2.86	0.48	0.76	-1.0
11	276	91.30	-1.69	0.49	1.07	0.5
12	276	94.20	-2.40	0.55	0.74	-1.3
13	276	93.84	-2.29	0.48	0.93	-0.3
14	276	93.12	-2.10	0.53	0.86	-0.7
15	276	93.12	-2.10	0.52	0.87	-0.6
16	276	93.48	-2.19	0.56	0.74	-1.4
17	276	93.84	-2.28	0.60	0.61	-2.2
18	276	91.67	-1.75	0.60	0.74	-1.6
19	276	89.13	-1.27	0.64	0.74	-1.9
20	276	90.22	-1.45	0.67	0.65	-2.5
21	276	88.04	-1.08	0.69	0.68	-2.4
22	276	85.14	-0.65	0.68	0.72	-2.3
23	276	83.70	-0.46	0.71	0.71	-2.6
24	276	84.42	-0.55	0.73	0.62	-3.5
25	276	80.80	-0.10	0.69	0.78	-2.1
26	276	76.45	0.37	0.71	0.79	-2.1
27	276	73.91	0.62	0.71	0.78	-2.4
28	276	69.20	1.06	0.77	0.63	-4.6
29	276	63.77	1.53	0.76	0.64	-4.6
30	276	58.33	1.98	0.74	0.70	-3.7
31	276	54.71	2.28	0.75	0.62	-5.0
32	276	50.00	2.66	0.73	0.61	-5.0
33	276	40.94	3.39	0.69	0.67	-4.0
34	276	38.04	3.64	0.68	0.58	-5.1

Item	N	Lösungs- häufigkeit (%)	Schwierig- keit	Trenn- schärfe	Infit MNSQ	T-Wert
35	276	33.70	4.00	0.68	0.57	-5.0
36	276	30.80	4.26	0.66	0.56	-5.0
37	276	24.28	4.85	0.61	0.59	-4.2
38	276	21.74	5.10	0.55	0.62	-3.8
39	276	17.03	5.58	0.49	0.62	-3.6
40	276	14.13	5.92	0.49	0.54	-4.5
	EAP/PV-Reliabilität: 1.16					
	Varianz: 6.21					
Skala	Trennschärfe: $M = 0.60$, $SD = 0.137$, $MIN = 0.24$, $MAX = 0.77$					
	Schwierigkeit: $M = 0$, $SD = 2.810$, $MIN = -5.089$, $MAX = 5.916$					
	Infit: $M = 0.79$, $SD = 0.23$, $MIN = 0.54$, $MAX = 1.49$					
Stufe 2-3	EAP/PV-Reliabilität: 1.05					
	Varianz: 7.46					
Stufe 4	EAP/PV-Reliabilität: 1.19					
	Varianz: 4.54					

Tabelle B.5

Räumlicher Test: Statistische Kennwerte der eindimensionalen Rasch-Skalierung

Item	N	Lösungs- häufigkeit (%)	Schwierig- keit	Trenn- schärfe	Infit MNSQ	T-Wert
1	271	90.04	-1.81	0.29	1.03	0.2
2	271	89.30	-1.72	0.22	1.09	0.7
3	271	47.23	0.99	0.62	0.95	-0.9
4	271	35.79	1.60	0.67	0.89	-1.9
5	271	69.74	-0.20	0.44	1.07	1.0
6	271	69.00	-0.16	0.54	0.99	-0.1
7	271	95.57	-2.76	0.11	1.04	0.3
8	271	67.53	-0.08	0.49	1.06	0.9
9	271	65.68	0.03	0.50	1.07	1.0
10	271	32.84	1.77	0.63	0.90	-1.5
11	271	23.62	2.35	0.54	0.96	-0.5
	EAP/PV-Reliabilität: 0.64					
	Varianz: 1.39					
Skala	Trennschärfe: $M = 0.46$, $SD = 0.180$, $MIN = 0.11$, $MAX = 0.67$					
	Schwierigkeit: $M = 0$, $SD = 1.62$, $MIN = -2.76$, $MAX = 2.35$					
	Infit: $M = 1.00$, $SD = 0.071$, $MIN = 0.89$, $MAX = 1.09$					
Stufe 2-3	EAP/PV-Reliabilität: 0.59					
	Varianz: 1.03					
Stufe 4	EAP/PV-Reliabilität: 0.66					
	Varianz: 1.57					

Tabelle B.6

Arlin-Test: Statistische Kennwerte der eindimensionalen Rasch-Skalierung

Item	N	Lösungs- häufigkeit (%)	Schwierig- keit	Trenn- schärfe	Infit MNSQ	T-Wert
1	141	36.88	-0.06	0.23	1.02	0.3
2	141	15.60	1.11	0.27	1.00	0.1
3	141	31.91	0.16	0.35	0.99	-0.1
4	141	21.28	0.72	0.37	0.97	-0.2
5	141	36.88	-0.06	0.10	1.04	0.9
6	141	29.08	0.30	0.20	1.02	0.3
7	141	37.59	-0.09	0.39	0.99	-0.2
8	141	32.62	0.13	0.32	1.00	0.0
9	141	27.66	0.37	0.16	1.01	0.1
10	141	58.16	-0.95	0.29	1.00	0.1
11	141	53.90	-0.77	0.30	1.00	-0.1
12	141	44.68	-0.39	0.36	0.98	-0.6
13	141	40.43	-0.22	0.38	0.99	-0.2
14	141	53.90	-0.77	0.34	0.99	-0.3
15	141	37.59	-0.09	0.30	1.01	0.3
16	141	22.70	0.64	0.20	1.01	0.1

EAP/PV-Reliabilität: 0.22
 Varianz: 0.10
 Skala Stufe 4 Trennschärfe: $M = 0.29$, $SD = 0.09$, $MIN = 0.10$, $MAX = 0.39$
 Schwierigkeit: $M = 0$, $SD = 0.56$, $MIN = -0.95$, $MAX = 1.11$
 Infit: $M = 1.00$, $SD = 0.02$, $MIN = 0.97$, $MAX = 1.04$

Tabelle B.7

Skala naturwissenschaftsbezogene Einstellungen (TIMSS): Statistische Kennwerte des eindimensionalen Partial-Credit-Modells

Item	N	Lösungs- häufigkeit (%)	Schwierig- keit	Trenn- schärfe	Infit MNSQ	T-Wert
1	272	37.50	0.39	0.83	1.08	0.9
2	271	58.30	-0.41	0.89	0.82	-1.8
3	267	59.93	0.03	0.86	1.01	0.1
	EAP/PV-Reliabilität: 0.73 Varianz: 3.28					
Skala	Trennschärfe: $M = 0.86$, $SD = 0.03$, $MIN = 0.83$, $MAX = 0.89$ Schwierigkeit: $M = 0$, $SD = 0.400$, $MIN = -0.412$, $MAX = 0.387$ Infit: $M = 0.98$, $SD = 0.134$, $MIN = 0.82$, $MAX = 1.08$					
Stufe 2-3	EAP/PV-Reliabilität: 0.75 Varianz: 3.54					
Stufe 4	EAP/PV-Reliabilität: 0.71 Varianz: 3.09					

Tabelle B.8

Skala naturwissenschaftsbezogenes Fähigkeitsselbstkonzept (TIMSS): Statistische Kennwerte des eindimensionalen Partial-Credit-Modells

Item	N	Lösungs- häufigkeit (%)	Schwierig- keit	Trenn- schärfe	Infit MNSQ	T-Wert
1	272	37.87	0.21	0.76	0.95	-0.5
2	269	37.92	0.29	0.72	1.08	1.0
3	268	60.07	-0.06	0.75	1.05	0.5
4	269	68.77	-0.44	0.77	0.88	-1.1
	EAP/PV-Reliabilität: 0.66 Varianz: 1.66					
Skala	Trennschärfe: $M = 0.75$, $SD = 0.02$, $MIN = 0.72$, $MAX = 0.77$ Schwierigkeit: $M = 0$, $SD = 0.33$, $MIN = -0.44$, $MAX = 0.29$ Infit: $M = 0.99$, $SD = 0.09$, $MIN = 0.88$, $MAX = 1.08$					
Stufe 2-3	EAP/PV-Reliabilität: 0.67 Varianz: 1.46					
Stufe 4	EAP/PV-Reliabilität: 0.67 Varianz: 1.98					

Tabelle B.9

Skala Interesse an forschenden Aktivitäten (ICA-R): Statistische Kennwerte des eindimensionalen Partial-Credit-Modells

Item	N	Lösungshäufigkeit (%)	Schwierigkeit	Trennschärfe	Infit MNSQ	T-Wert
1	260	28.08	0.05	0.71	0.91	-1.0
2	260	26.92	0.31	0.63	1.07	0.9
3	265	25.28	0.51	0.69	1.04	0.5
4	264	42.05	-0.46	0.55	1.07	0.7
5	263	45.63	-0.41	0.64	0.98	-0.2
	EAP/PV-Reliabilität: 0.60					
	Varianz: 0.69					
Skala	Trennschärfe: $M = 0.64$, $SD = 0.06$, $MIN = 0.55$, $MAX = 0.71$					
	Schwierigkeit: $M = 0$, $SD = 0.43$, $MIN = -0.46$, $MAX = 0.51$					
	Infit: $M = 1.01$, $SD = 0.07$, $MIN = 0.91$, $MAX = 1.07$					
Stufe 2-3	EAP/PV-Reliabilität: 0.61					
	Varianz: 0.73					
Stufe 4	EAP/PV-Reliabilität: 0.60					
	Varianz: 0.68					

B.2 Latente einfache Regressionsanalysen

Tabelle B.10

Pearson-Korrelationen der WLE-Personenparameter und Summenscores der Prädiktoren

Leseverständnis (N = 271)	0.98
Intelligenz (N = 272)	0.99
Problemlösen (N = 270)	1.00
Inhibition (N = 276)	0.96
Räumliches Denken (N = 271)	1.00
Formal-operatorische Fähigkeiten (N = 141)	1.00
Einstellungen (N = 267)	0.97
Fähigkeitsselbstkonzept (N = 266)	0.95
Interesse an forschenden Aktivitäten (N = 242)	0.96

Tabelle B.11

Ergebnisse latenter einfacher Regressionen zur Vorhersage der Kompetenz im wissenschaftlichen Denken bezogen auf die Gesamtstichprobe

Prädiktor	<i>B</i>	<i>SE B</i>	<i>KI</i>
Klassenstufe	0.60*	0.05	[0.50;0.70]
Leseverständnis	0.11*	0.01	[0.09;0.13]
Intelligenz	0.12*	0.01	[0.09;0.15]
Problemlösen	0.25*	0.02	[0.21;0.29]
Inhibition	0.04*	0.01	[0.03;0.05]
Räumliches Denken	0.18*	0.02	[0.14;0.23]
Einstellungen	0.03	0.02	[-0.01;0.08]
Fähigkeitsselbstkonzept	0.09*	0.02	[0.05;0.14]
Interesse an forschenden Aktivitäten	0.04*	0.02	[0.00;0.08]

Tabelle B.12

Ergebnisse latenter einfacher Regressionen zur Vorhersage der Kompetenz im wissenschaftlichen Denken bezogen auf das frühe Grundschulalter

Prädiktor	<i>B</i>	<i>SE B</i>	<i>KI</i>
Leseverständnis	0.07*	0.01	[0.05;0.09]
Intelligenz	0.06*	0.02	[0.02;0.09]
Problemlösen	0.21*	0.03	[0.14;0.27]
Inhibition	0.02*	0.01	[0.00;0.03]
Räumliches Denken	0.12*	0.03	[0.06;0.17]
Einstellungen	0.08*	0.02	[0.03;0.12]
Fähigkeitsselbstkonzept	0.09*	0.02	[0.04;0.13]
Interesse an forschenden Aktivitäten	0.06*	0.02	[0.02;0.11]

Tabelle B.13

Ergebnisse latenter einfacher Regressionen zur Vorhersage der Kompetenz im wissenschaftlichen Denken bezogen auf das späte Grundschulalter

Prädiktor	<i>B</i>	<i>SE B</i>	<i>KI</i>
Leseverständnis	0.09*	0.01	[0.07;0.12]
Intelligenz	0.12*	0.02	[0.08;0.15]
Problemlösen	0.19*	0.02	[0.14;0.23]
Inhibition	0.01	0.01	[-0.01;0.03]
Räumliches Denken	0.15*	0.03	[0.10;0.20]
Formal-operatorische Fähigkeiten	0.12*	0.03	[0.06;0.18]
Einstellungen	-0.04	0.03	[-0.10;0.02]
Fähigkeitsselbstkonzept	0.09*	0.03	[0.04;0.14]
Interesse an forschenden Aktivitäten	0.04	0.03	[-0.01;0.09]