

Konzeptuelles und prozedurales Wissen als latente Variablen: Ihre Interaktion beim Lernen mit Dezimalbrüchen

von Diplompsychologe
Michael Schneider

Von der Fakultät V – Verkehrs- und Maschinensysteme
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Philosophie
- Dr. phil. -

genehmigte Dissertation

Voritzender: Prof. Dr. D. Manzey
Gutachter: Prof. Dr. E. Stern
Gutachter: Prof. Dr. H. Jungermann

Tag der wissenschaftlichen Aussprache: 17. Februar 2006

Danksagung

Ich möchte mich bei allen Menschen bedanken, die während der Anfertigung dieser Dissertation in vielfältiger Weise für mich da waren.

Meiner Doktormutter und Forschungsgruppenleiterin Prof. Dr. Elsbeth Stern danke ich für die wertvollen inhaltlichen Impulse und den steten Ansporn. Sie hat mir große Freiräume gewährt, war aber immer für mich da, wenn ich sie brauchte, und war eine konstruktivistische Lehrerin im besten Sinne.

Meinen Kollegen im Enterprise-Projekt, Ilonca Hardy, Anja Felbrich, Henrik Saalbach und Roland Grabner, danke ich für konstruktive Kritik, Ratschläge und die Gespräche bei Café Latte, Kantinenessen und Schokomuffins.

Dem Direktor unserer Abteilung Erziehungswissenschaft und Bildungssysteme am Max-Planck-Institut für Bildungsforschung, Prof. Dr. Jürgen Baumert, danke ich für die intellektuell anregende Umgebung und die hervorragende logistische Ausstattung, die ich mitbenutzen durfte.

Dem Zweitbetreuer dieser Arbeit, Prof. Dr. Helmut Jungermann von der Technischen Universität Berlin, danke ich für seine hilfreichen und ermutigenden Hinweise und sein ehrliches Interesse.

Ich danke den aktuellen und ehemaligen studentischen Hilfskräften unseres Projekts Gabi Cierniak, Barbara Diffendaffer, Carmen Kardoes, Jenny Paetsch, Rainer Roth, Lennart Schalk, Henrieta Sokolova, Beate Stadelhofer und Ruth Wagner sowie unseren Forschungspraktikanten Frank Bültge, Hanna Dumont und Anna-Kristin Kalbacher für ihr Engagement in vielen kleinen und großen Dingen, für ihre Bereitschaft auch an Wochenenden zu arbeiten und für viele schöne Gespräche.

Hella Beister, Sylvia Weber und Manuela Meermann danke ich für ihre wichtige Hilfe in allen logistischen Dingen, für das Korrekturlesen und die Übersetzungshilfen.

Martin Brunner und Nele Julius-McElvany danke ich für den inhaltlichen Austausch und die Unterhaltungen in Berliner Teestuben, paduanischen Pizzerias und zypriotischen Tavernas.

Rainer Watermann, Oliver Lüdtke, Martin Brunner und Alexandra Shajek bin ich dankbar für Hinweise über den Umgang mit HLM und SEM.

Matthias Lippold und seinen Mitarbeitern danke ich für den EDV-Support und Gerda Fritsch von der Verwaltung für all das Geld.

Ich danke mehr als 400 Berliner Schülerinnen und Schülern für ihre Teilnahme an den Studien und Vorstudien, für ihre Geduld und Kooperationsbereitschaft. Ihren Eltern danke ich für das in uns gesetzte Vertrauen.

Den Schulleitern und Lehrern der folgenden Berliner Schulen sei dafür gedankt, dass wir in kurzen Unterrichtsbesuchen unsere Studien vorstellen und dadurch die StudienteilnehmerInnen gewinnen konnten: Eichendorf-Grundschule, Evangelische Schule Charlottenburg, Evangelische Schule Steglitz, Erich-Kästner-Grundschule, Freiherr-von-Hünefeld-Grundschule, Grundschule am Buschgraben, Grundschule am Rüdesheimer Platz, Grunewald-Grundschule, Halensee-Grundschule, Johann-Peter-Hebel-Grundschule, Judith-Kerr-Grundschule, Katharina-Heinroth-Grundschule, Kronach-Grundschule, Lietzensee-Grundschule, Mierendorff-Grundschule, Rothenburg-Grundschule, Sachsenwald-Grundschule, Süd-Grundschule.

Meinen Eltern, Ursula und Thomas Schneider, sowie meinen Geschwistern Matthias und Christine danke ich für die emotionale und kulinarische Unterstützung, Umzugshilfen und Unzähliges mehr. Mein Vater hat diese Arbeit überaus sorgfältig Korrektur gelesen und damit wesentlich zur Lesbarkeit des Textes beigetragen.

Während der Entstehung dieser Dissertation verstarb meine beste Freundin Laura Richter. Ihr ist diese Arbeit gewidmet.

Inhaltsverzeichnis

1	Einleitung	13
1.1	Theorie und Praxis	13
1.2	Gliederung der Arbeit	14
2	Befunde zum Erwerb mathematischer Konzepte und Prozeduren.....	17
2.1	Die Forschungstradition zu mathematischen Konzepten und Prozeduren ..	17
2.2	Die Entwicklung der Kompetenz des Zählens.....	17
2.2.1	<i>Gelmans Untersuchung der principles und skills des Zählens</i>	<i>17</i>
2.2.2	<i>Die Unterscheidung konzeptueller und prozeduraler Kompetenz</i>	<i>19</i>
2.3	Der Literaturüberblick von Rittle-Johnson und Siegler.....	21
2.3.1	<i>Untersuchte Inhaltsbereiche und Evidenztypen</i>	<i>21</i>
2.3.2	<i>Korrelative Relationen</i>	<i>22</i>
2.3.3	<i>Erwerbsreihenfolge in Querschnittsdesigns</i>	<i>23</i>
2.3.4	<i>Längsschnittliche Befunde</i>	<i>23</i>
2.3.5	<i>Trainingsstudien.....</i>	<i>24</i>
2.3.6	<i>Diskussion der Studien.....</i>	<i>26</i>
2.4	Neuere Befunde	27
2.4.1	<i>Korrelative Befunde und Erwerbsreihenfolge</i>	<i>27</i>
2.4.2	<i>Trainingsstudien.....</i>	<i>30</i>
2.4.3	<i>Diskussion der neueren Befunde.....</i>	<i>32</i>
2.5	Zusammenfassung	33
3	Das Iterative Modell.....	35
3.1	Modellannahmen	35
3.2	Die Äquivalenzproblemstudie	37
3.3	Die Dezimalbruchstudie	40
3.4	Diskussion des Modells	44
3.5	Zusammenfassung	46
4	Die Unterscheidung von konzeptuellem und prozeduralem Wissen.....	48
4.1	Verwandte Unterscheidungen.....	48
4.2	Definitorische Eingrenzung	53
4.3	Gründe gegen die Unterscheidung.....	56
4.3.1	<i>Kritik aus Sicht des Ansatzes der Situierten Kognition.....</i>	<i>56</i>
4.3.2	<i>Kritik aus didaktischer Sicht</i>	<i>57</i>
4.3.3	<i>Kritik an der Undifferenziertheit der Unterscheidung.....</i>	<i>58</i>
4.4	Gründe für die Unterscheidung.....	59
4.4.1	<i>Sprachlogische Beweise</i>	<i>59</i>
4.4.2	<i>Lernpsychologische Argumente</i>	<i>59</i>
4.4.3	<i>Befunde aus der Gedächtnisforschung.....</i>	<i>61</i>
4.4.4	<i>Neuropsychologische Evidenz.....</i>	<i>63</i>
4.4.5	<i>Evolutionpsychologische Überlegungen</i>	<i>66</i>
4.4.6	<i>Unterscheidung der Wissensarten: Fazit.....</i>	<i>67</i>
4.5	Hypothetische mentale Repräsentationen der Wissensarten	68
4.5.1	<i>Repräsentationen konzeptuellen Wissens.....</i>	<i>68</i>
4.5.2	<i>Repräsentationen prozeduralen Wissens</i>	<i>71</i>

4.5.3	<i>Beziehungen zwischen mentaler Repräsentationsform und Inhalt....</i>	73
4.6	Operationalisierungen der Wissensarten.....	74
4.6.1	<i>Die Messung konzeptuellen Wissens.....</i>	74
4.6.2	<i>Die Messung prozeduralen Wissens.....</i>	78
4.7	Beziehungen der Wissensarten zu anderen kognitiven Konstrukten.....	83
4.7.1	<i>Intelligenz als Prädiktor des Wissenserwerbs.....</i>	83
4.7.2	<i>Mentale Repräsentation von Zahlen.....</i>	84
4.8	Zusammenfassung.....	86
5	Erwerb und Abruf konzeptuellen und prozeduralen Wissens.....	87
5.1	Wege des Wissenserwerbs.....	87
5.2	Die ACT-Theorie.....	89
5.3	Die Theorie der representational redescription.....	92
5.4	Hypothetische Verarbeitungsmechanismen.....	94
5.4.1	<i>Direkter Erwerb konzeptuellen und prozeduralen Wissens.....</i>	94
5.4.2	<i>Umsetzung der Wissensarten ineinander.....</i>	97
5.4.3	<i>Abruf und Transfer der Wissensarten.....</i>	99
5.4.4	<i>Übersicht.....</i>	101
5.5	Erwerb der Wissensarten in schulischen Kontexten.....	102
5.5.1	<i>Zwei Fragestellungen.....</i>	102
5.5.2	<i>Unterrichtsziele.....</i>	102
5.5.3	<i>Unterrichtssequenzierung.....</i>	104
5.6	Zusammenfassung.....	105
6	Inhaltsspezifisches konzeptuelles und prozedurales Wissen.....	107
6.1	Domänenspezifität des Wissenserwerbs.....	107
6.2	Dezimalbruchwissen.....	107
6.3	Zahlenstrahlwissen.....	109
6.3.1	<i>Wissen über die externe Wissensrepräsentation.....</i>	109
6.3.2	<i>Passung von externer und interner Wissensrepräsentation.....</i>	111
6.4	Zusammenfassung.....	111
7	Validierung kognitiver Modelle mittels Strukturgleichungsmodellen ...	112
7.1	Pfadmodelle und Strukturgleichungsmodelle.....	112
7.2	Das Messmodell.....	112
7.3	Das Regressionsmodell.....	115
7.4	Modellierungsstrategien.....	117
7.5	Modellgütebeurteilung und Modellinterpretation.....	121
7.6	Zusammenfassung.....	123
8	Fragestellung der Dissertation.....	124
8.1	Herleitung der Forschungsfragen.....	124
8.2	Zusammenfassung.....	127
9	Studie 1: Interrelationen der Wissensarten.....	128
9.1	Fragestellungen und Hypothesen.....	128
9.2	Methode.....	129
9.2.1	<i>Stichprobe.....</i>	129
9.2.2	<i>Aufbau.....</i>	131

9.2.3	<i>Durchführung</i>	133
9.2.4	<i>Intervention</i>	134
9.2.5	<i>Maße konzeptuellen Wissens</i>	136
9.2.6	<i>Maße prozeduralen Wissens</i>	140
9.2.7	<i>Erhobene Kovariaten</i>	145
9.2.8	<i>Design und Auswertung</i>	147
9.3	<i>Ergebnisse bezüglich der Forschungsfragen</i>	151
9.3.1	<i>Beschreibung der Stichprobe</i>	151
9.3.2	<i>Eigenschaften der Wissensmaße</i>	152
9.3.3	<i>Konvergente Validitäten</i>	157
9.3.4	<i>Divergente Validitäten</i>	163
9.3.5	<i>Prädiktive Relationen</i>	167
9.4	<i>Ergebnisse explorativer Analysen</i>	170
9.4.1	<i>Intelligenz und die Wissensarten</i>	170
9.4.2	<i>Mentale Zahlenrepräsentation und die Wissensarten</i>	174
9.4.3	<i>Messinvarianz</i>	175
9.4.4	<i>Stabilität der Faktorinterkorrelationen</i>	178
9.4.5	<i>Einfluss von Personencharakteristika</i>	178
9.4.6	<i>Einfluss potenzieller Störvariablen</i>	179
9.4.7	<i>Explorative Faktoranalysen</i>	181
9.4.8	<i>Vergleichsmodelle mit manifesten Variablen</i>	183
9.5	<i>Diskussion</i>	184
9.5.1	<i>Operationalisierbarkeit der Wissensarten</i>	184
9.5.2	<i>Kausale Interrelationen</i>	190
9.5.3	<i>Weitere Einflussgrößen</i>	193
9.6	<i>Zusammenfassung</i>	194
10	Studie 2: Erwerb der Wissensarten	195
10.1	<i>Fragestellungen und Hypothesen</i>	195
10.2	<i>Methode</i>	197
10.2.1	<i>Stichprobe, Aufbau und Durchführung</i>	197
10.2.2	<i>Maße und Interventionen</i>	198
10.2.3	<i>Design und Auswertung</i>	202
10.3	<i>Ergebnisse</i>	205
10.3.1	<i>Beschreibung der Stichprobe</i>	205
10.3.2	<i>Charakteristika der Wissensscores</i>	207
10.3.3	<i>Einfluss der Treatments auf die Wissensscores</i>	209
10.3.4	<i>Einfluss der Treatments auf die Wissensscoreinterkorrelationen</i>	212
10.3.5	<i>Einfluss der Treatments auf die Einzelwissensmaße</i>	213
10.3.6	<i>Reliabilitäten der Wissensmaße</i>	216
10.3.7	<i>Einfluss des distance effects auf den Lernerfolg</i>	217
10.4	<i>Diskussion</i>	217
10.4.1	<i>Separate Vermittelbarkeit der Wissensarten</i>	217
10.4.2	<i>Kausale Interrelationen</i>	221
10.4.3	<i>Weitere Befunde</i>	223
10.5	<i>Zusammenfassung</i>	224
11	Diskussion	226
11.1	<i>Diskussion der Befunde zu den Forschungsfragen</i>	226
11.2	<i>Fazit und Ausblick</i>	234

11.2.1	<i>Psychologische Lerntheorien</i>	234
11.2.2	<i>Die Unterscheidung von Wissensarten</i>	238
11.2.3	<i>Methodische Implikationen der Befunde</i>	242
11.2.4	<i>Implikationen der Befunde für die Praxis</i>	245
11.3	<i>Zusammenfassung</i>	250
12	Literatur	252

Abbildungsverzeichnis

Abbildung 1: Pfaddiagramm eines cross-lagged panel designs.....	116
Abbildung 2: Screenshots der Grafikoberfläche der Intervention (links: vor der Antworteingabe; rechts: nach Eingabe einer inkorrekten Antwort).	135
Abbildung 3: Screenshot der Implementation des Maßes Übersetzung.	138
Abbildung 4: Screenshot der Implementation des Maßes Größenvergleich.	139
Abbildung 5: Screenshot der Implementation des Maßes Problemlösekorrektheit.....	141
Abbildung 6: Screenshot der Implementierung des Maßes Zugriffsasymmetrie (links: geübte Transformationsrichtung I, rechts: nicht geübte Transformationsrichtung II).	143
Abbildung 7: Screenshots der Implementation des Maßes Dual task-Kosten (links: Baseline-Bedingung I, rechts: Zweitaufgaben-Bedingung II).	143
Abbildung 8: Für die Auswertung verwendetes Grundmodell.	148
Abbildung 9: Pfaddiagramm von Modell R12.	169
Abbildung 10: Pfaddiagramm von Modell I12.....	171
Abbildung 11: Pfaddiagramme (links oben : Modell IK123, links unten: Modell K123, rechts oben: Modell IP123, rechts unten: Modell P123).....	173
Abbildung 12: Pfaddiagramm von Modell D12.	175
Abbildung 13: Cross-lagged panel-Modelle mit manifesten Variablen (links oben: Modell EP123, rechts oben: Modell EZ123, unten: Modell ÜP123).	184

Tabellenverzeichnis

Tabelle 1: Auswahl einiger von Lerntheoretikern bisher vorgeschlagener Wissensarten...	48
Tabelle 2: Zitationsanalyse der synonymen Verwendung von Namen für Wissensarten...	52
Tabelle 3: Definitionen konzeptuellen und prozeduralen Wissens.....	54
Tabelle 4: Einteilung von Theorien anhand der postulierten Relationen.....	88
Tabelle 5: Mögliche Verarbeitungsmechanismen für die Wissensarten.....	102
Tabelle 6: Verbreitete Misskonzepte über Dezimalbrüche.....	108
Tabelle 7: Übersicht über die Forschungsfragen.	126
Tabelle 8: Chronologische Reihenfolge der Datenerhebungsphasen.	134
Tabelle 9: In der Intervention verwendete Typen von Dezimalbrüchen.	135
Tabelle 10: Verbale Beschreibungen der zu evaluierenden Strategien.....	137
Tabelle 11: Fragen des Maßes Erklärung.	139
Tabelle 12: Übersicht über die Wissensmaße.	149
Tabelle 13: Mittelwerte, Standardabweichungen und Signifikanz des Kolmogorov-Smirnov-Anpassungstests der Wissensmaße für die drei Messzeitpunkte.	153
Tabelle 14: Veränderung der Wissensmaße (Mittelwertsdifferenz, Cohens <i>d</i> und Signifikanz) über die drei Messzeitpunkte.....	155
Tabelle 15: Indikatoren der Reliabilitäten der Wissensmaße.	156
Tabelle 16: Interraterreliabilitäten zu drei Messzeitpunkten.	157
Tabelle 17: Fit der Modelle zur Überprüfung der konvergenten Validitäten.	158
Tabelle 18: Faktorladungen in den Modellen K1EM bis P3EM.	160
Tabelle 19: Fit der Modelle zur Überprüfung der divergenten Validitäten.	163
Tabelle 20: Parameter der Chi-Quadrat-Differenz-Tests für sb-skalierte Werte.....	164
Tabelle 21: Fit der Modelle zur Exploration der divergenten Validitäten.....	166
Tabelle 22: Parameter der explorativen Chi-Quadrat-Differenz-Tests (sb-skaliert).	167
Tabelle 23: Parameter von Modell R12 mit Standardfehlern, z-Werten und Irrtumswahrscheinlichkeiten.	170
Tabelle 24: Koeffizienten von Modell R12 sowie von Modell I12 mit Standardfehlern, z-Werten und Irrtumswahrscheinlichkeiten.	172
Tabelle 25: Fitindizes der Modelle zur Analyse des KFT-Einflusses.	174
Tabelle 26: Fitindizes der Modelle zur Analyse der Messinvarianz.....	175
Tabelle 27: Parameter des Chi-Quadrat-Differenz-Tests auf Messinvarianz.	176
Tabelle 28: Koeffizienten von Modell R12F mit fixierten Faktorladungen.	177
Tabelle 29: Faktorinterkorrelationen bei Kontrolle von Intelligenz und Alter.....	178

Tabelle 30: Durch Personencharakteristika aufgeklärte Varianzanteile der latenten Faktoren.....	179
Tabelle 31: Varianzaufklärungen potenzieller Störvariablen mit ihren Signifikanzen.	180
Tabelle 32: Eigenwerte der drei Faktoren (eine Analyse pro Messzeitpunkt).	181
Tabelle 33: Standardisierte Ladungen der Wissensmaße auf drei Faktoren (eine Analyse pro Messzeitpunkt).....	181
Tabelle 34: Fitindizes der Modelle mit manifesten Variablen.	184
Tabelle 35: Erklärungsmöglichkeiten für die hohen Faktorinterkorrelationen.	190
Tabelle 36: Chronologische Reihenfolge der Datenerhebungsphasen.	198
Tabelle 37: Faktorladungen und Reliabilitäten von Modell R12F (Studie 1).	204
Tabelle 38: Deskriptive Charakteristika der drei Treatmentgruppen im Vergleich.	206
Tabelle 39: Ergebnisse der Varianzanalyse zur Dauer der Treatments.	207
Tabelle 40: Verteilungseigenschaften der Wissensscores.	207
Tabelle 41: Ladungen der explorativen Faktoranalyse für Mzp 1.....	209
Tabelle 42: Mittelwerte und Standardabweichungen der Wissensscores in den Treatmentgruppen.	209
Tabelle 43: Ergebnisse der Varianzanalysen zum ersten Messzeitpunkt.	210
Tabelle 44: Ergebnisse der Kovarianzanalysen bezüglich Treatmentgruppenunterschieden an Mzp 2.....	211
Tabelle 45: Ergebnisse der Kovarianzanalysen bezüglich Treatmentgruppenunterschieden an Mzp 3.....	212
Tabelle 46: Korrelationen zwischen den Scores der beiden Wissensarten.....	212
Tabelle 47: Mittelwerte und Standardabweichungen der Wissensmaße in den Treatmentgruppen.	213
Tabelle 48: Cohens d und Signifikanz der Veränderungen der Wissensmaße über die Mzp.	214
Tabelle 49: Aufgabenanzahlen und Cronbachs Alphas der Wissensmaße an Mzp 1.....	216
Tabelle 50: Interrater-Reliabilitäten für das Maß Erklärung.	216
Tabelle 51: Überblick über die erwarteten und beobachteten Befunde in Studie 2.	223
Tabelle 52: Vier mögliche Kombinationen von Repräsentationsform und -inhalt.....	240

„Zu handeln ist der Schlüssel zu jedem Erfolg.“

(Pablo Picasso zugeschrieben im Internet)

„Dem Anwenden muss das Erkennen vorausgehen.“

(Max Planck zugeschrieben im Internet)

*„Nature has contrived to have it both ways,
to get the best out of fast and dumb systems
and slow and contemplative ones,
by simply refusing to choose between them.“*

(Jerry Fodor, 1985, S. 4)

1 Einleitung

1.1 Theorie und Praxis

Wenn man Anfang März 2005 das Begriffspaar „Theorie Praxis“ auf der deutschen Seite der Internet-Suchmaschine Google eingab, erhielt man 1.370.000 Treffer. Die ersten zehn davon waren Webseiten über eine Theorie der Praxis des Philosophen und Soziologen Pierre Bourdieu, über Praxis und Theorie des Bankraubes, über eine Projektwerkstatt für Lehramtsstudierende der Uni Wien, über einen Studiengang Pflegeleitung der Hochschule Bremen, ein Onlinejournalartikel über das „Theorie-Praxis-Problem der Lehrerbildung“, Seiten über Leben und Opus des Malers Frank K. Opitz, über einen theoretisch und praktisch angelegten Segelkurs auf der Insel Elba, eine Seite mit Cartoons für Sportpädagogen, eine Seite auf dem Deutschen Bildungsserver und eine über eine Bildungsinitiative zur Verzahnung von Schulen, Hochschulen und Wirtschaft.

Dies verdeutlicht auf den ersten Blick, dass das Verhältnis von Theorie und Praxis in vielen verschiedenen Lebensbereichen, von der Philosophie über den Bankraub bis zum Segeln, eine wichtige Rolle spielt. Auf den zweiten Blick wird zusätzlich erkennbar, dass es im Themenbereich Bildung und Ausbildung eine besonders große Rolle zu spielen scheint. Sechs der zehn Webseiten können diesem Themenfeld zugeordnet werden – selbst wenn man den Segelkurs nicht dazuzählt.

Ein Auszug aus dem oben genannten Onlinejournalartikel mit dem vollständigen Titel *Das unstillbare Verlangen nach Praxisbezug – Zum Theorie-Praxis-Problem der Lehrerbildung am Exempel Schulpraktischer Studien* illustriert, wie spannungsvoll die Theorie-Praxis-Unterscheidung ist:

Die Einheit von Theorie und Praxis, von Hochschule und Schule, Ausbildung und Beruf, ist verloren - wenn es sie denn je gegeben hat. Theorie und Praxis sind getrennt. Das gilt beispielsweise für Volkswirtschaftslehre und die Tätigkeit von Volkswirten, Managementtheorie und Management, Bankbetriebslehre und Bankarbeit ebenso wie für Erziehungswissenschaft und pädagogische Praxis oder Fachdidaktikwissenschaft und Fachunterrichtspraxis. Wären Theorie und Praxis nicht getrennt, müssten sie nicht „vermittelt“ werden und die „Praxisferne“ der Theorie oder die „Theorieferne“ der Praxis wären kein Problem. (Hedtke, 2000)

Tatsächlich scheinen sich „Praktiker“ und „Theoretiker“ nicht selten mit Misstrauen zu begegnen. Während Theoretiker Praktikern tendenziell Konzeptlosigkeit vorwerfen, kritisieren Praktiker an Theoretikern gerne die Realitätsferne ihrer Theorien, die sie zur Lösung tatsächlicher Probleme nutzlos sein lässt. Auch die beiden Zitate von Pablo Picasso

und Max Planck am Anfang dieses Abschnitts passen in dieses Schema, indem sie hinterfragen, ob das Handeln ein Primat über das Erkennen besitzt oder umgekehrt.

Dieses Spannungsverhältnis herrscht jedoch nicht nur zwischen Menschen oder gesellschaftlichen Gruppen. Philosophen und Psychologen spekulieren darüber, dass die beiden Pole auch innerhalb einzelner Personen existieren könnten – in Form von abstrakt-konzeptuellem Verständnis einerseits und praktisch anwendbarem Handlungswissen andererseits. In der kognitiven Lehr-Lern-Psychologie werden diese beiden Konstrukte als *konzeptuelles* und *prozedurales Wissen* bezeichnet.

Die Wissensarten lassen sich noch schwerer beobachten als Konflikte zwischen gesellschaftlichen Gruppen, da erstere sich im Verhalten eines Menschen stets ineinander verwoben zeigen. Nur selten kann das Verhalten eines Menschen eindeutig auf die eine oder die andere Wissensart zurückgeführt werden.

Dass die Wissensarten nicht direkt beobachtbar sind, mag erklären, warum viele grundlegende Fragen über ihre Eigenschaften und ihre Interrelationen durch die Lehr-Lern-Psychologie noch immer nicht beantwortet worden sind. Man weiß zum Beispiel, dass Kinder, die bestimmte mathematische Probleme lösen können, häufig auch viel über die dahinter stehenden Prinzipien wissen und umgekehrt. Aber gibt es auch Fälle, in denen die eine Kompetenz ohne die andere vorhanden ist? Sind solche Fälle im Schulunterricht gezielt zu berücksichtigen? In welcher Reihenfolge sollten Kinder die Wissensarten erwerben: Müssen sie nicht erst die konzeptuellen Zusammenhänge kennen, bevor sie durch Übung neues Handlungswissen erwerben können? Aber können sie abstrakte konzeptuelle Erklärungen über ein Gebiet, das sie noch nie handelnd explorieren und erfahren konnten, wirklich verstehen? Reicht es vielleicht, ihnen nur eine der beiden Wissensarten im Unterricht zu vermitteln, weil sie die andere selbstständig daraus ableiten können?

Diesen Fragen wird in der vorliegenden Dissertation nachgegangen. Als Aufgabengebiet wurde die Verortung von Dezimalbrüchen auf einem Zahlenstrahl gewählt. Diagramme, wie der Zahlenstrahl, veranschaulichen abstrakte Konzepte und machen sie so dem konkreten Handeln zugänglich (Stern, Hardy, & Koerber, 2002), so dass beim Arbeiten mit ihnen das konzeptuelle und das prozedurale Dezimalbruchwissen der Kinder ineinandergreifen können.

1.2 Gliederung der Arbeit

Die Gründe, die eine Unterscheidung von konzeptuellem und prozeduralem Wissen nahelegen, und die Theorien, die ihre wechselseitige Interaktion zum Gegenstand haben,

sind teilweise sehr abstrakt. Damit die Leser diese abstrakten Gedanken mit konkreten Inhalten verbinden können, beginnt diese Arbeit mit einer Übersicht über die empirischen Befunde über konzeptuelles und prozedurales Wissen im Bereich des Mathematiklernens. In Kapitel 3 wird dann das theoretische Kernstück der Arbeit vorgestellt: das Iterative Modell. Es wird erläutert, auf welche Arbeiten es aufbaut, welche Modellannahmen getroffen werden, wie sie bisher empirisch überprüft wurden und welche Punkte noch ungeklärt sind.

Bei der Diskussion der konkreten empirischen Befunde und des theoretischen Modells werden einige Fragen aufgeworfen, die es sinnvoll erscheinen lassen, beide in einen umfassenderen Rahmen einzuordnen. So können Fragen nach der Interaktion von konzeptuellem und prozeduralem Wissen beim Lernen erst beantwortet werden, wenn geklärt ist, was überhaupt unter konzeptuellem und prozeduralem Wissen verstanden wird. Der Klärung dieses Punktes ist Kapitel 4 gewidmet, wobei vor allem auf kognitionswissenschaftliche Erkenntnisse zurückgegriffen wird.

Anschließend werden in Kapitel 5 grundlegende lerntheoretische Positionen und Befunde über Erwerb und Abruf der beiden Wissensarten besprochen. Ihre Implikationen für Themenbereiche wie optimale Wissenserwerbsreihenfolge und Transferförderung werden hervorgehoben und auch im Hinblick auf ihre Relevanz für die Gestaltung von Schulunterricht evaluiert.

Wissen ist immer Wissen über etwas, das heißt es ist notwendigerweise inhaltspezifisch. Bei der Untersuchung der Interaktion der beiden Wissensarten ist es darum wichtig zu reflektieren, in welchem Inhaltsbereich die Untersuchung stattfindet, welche spezifischen Konzepte und Prozeduren die Lernenden dabei zu erwerben haben und in welcher Weise solche inhaltspezifischen Charakteristika die zu untersuchenden Lernprozesse beeinflussen können. In Kapitel 6 wird deshalb besprochen, was über das Wissen von Kindern über Dezimalzahlen und Diagramme bekannt ist.

Kapitel 7 beschreibt Grundlagen der Analyse von Strukturgleichungsmodellen, die bei der Auswertung der empirischen Daten benötigt werden.

Nachdem in den vorherigen Kapiteln geklärt wurde, wie sich konzeptuelles und prozedurales Wissen unterscheiden, wie sie verschiedenen lerntheoretischen Positionen zufolge miteinander interagieren und wie konkretes konzeptuelles und prozedurales Wissen in den untersuchten Inhaltsbereichen aussehen kann, wird in Kapitel 8 erneut das Iterative Modell aufgegriffen. Aus einer Analyse der noch unzureichend geklärten Punkte des Modells werden die Forschungsfragen der Dissertation abgeleitet und es wird

aufgezeigt, in welcher Weise das in den vorherigen Kapiteln Besprochene zur Klärung dieser Forschungsfragen herangezogen werden kann. Auf dieser Grundlage werden Hypothesen aufgestellt und die Relevanz der Fragestellung der Dissertation für psychologische Lerntheorien und für die Unterrichtsgestaltung evaluiert.

In den darauffolgenden zwei Kapiteln werden die beiden empirischen Studien vorgestellt, die zur Beantwortung der Forschungsfragen durchgeführt wurden.

Schließlich werden im letzten Abschnitt die Ergebnisse der Studien zusammengefasst und interpretiert. Es wird besprochen, wie sie sich zu den schon bekannten empirischen Befunden verhalten und welche Implikationen sie für die zukünftige Forschung und die Unterrichtsgestaltung haben könnten.

Mit Begriffen, wie Schüler und Teilnehmer, die sich von ihrer grammatikalischen Form her lediglich auf männliche Personen beziehen, sind – sofern nicht ausdrücklich eingeschränkt – in der gesamten Arbeit inhaltlich immer auch die entsprechenden weiblichen Personen gemeint.

Die Anhänge, auf die im Text hingewiesen wird, befinden sich auf einer DVD in der Tasche am Ende dieser Arbeit. Der vorliegende Dissertationstext als PDF-Dokument ist in Anhang I zu finden. Die Datei kann zum Beispiel genutzt werden, um im Text nach Begriffen zu suchen.

2 Befunde zum Erwerb mathematischer Konzepte und Prozeduren

2.1 Die Forschungstradition zu mathematischen Konzepten und Prozeduren

Resnick und Ford (1981) konstatieren in ihrem Buch *The Psychology of Mathematics for Instruction*: „The relationship between computational skill and conceptual understanding is one of the oldest concerns in the psychology of mathematics“ (S. 246).

Die Entwicklung der empirischen Forschung zu *computational skill* und *conceptual understanding* im Bereich des Mathematiklernens wird in diesem Kapitel vorgestellt. Im folgenden Unterkapitel werden zunächst die Arbeiten von Gelman und Kollegen über die Entwicklung kindlicher *principles* und *skills* im Bereich des Zählens dargestellt. Diese Arbeiten zählen zu den ältesten und zugleich bekanntesten Studien über konzeptuelles Verständnis und prozedurale Fähigkeit im Bereich der Mathematik.

Im dritten Unterkapitel wird ein zentraler Überblicksartikel über nachfolgende Arbeiten vorgestellt, bevor im vierten die wichtigsten aktuellen Studien besprochen werden.

2.2 Die Entwicklung der Kompetenz des Zählens

2.2.1 Gelmans Untersuchung der *principles* und *skills* des Zählens

Rochel Gelman und Kollegen untersuchen seit fast drei Jahrzehnten, wie sich bei Kindern die Kompetenz des Zählens entwickelt. Sie unterscheiden dabei zwischen dem Verständnis der dem Zählen zugrunde liegenden Prinzipien einerseits und der tatsächlichen Fähigkeit zu zählen andererseits, wobei sie Ersteres als *principles* und Letztere als *skills* bezeichnen. Sie führen zwei Gründe für die Notwendigkeit dieser Unterscheidung an.

Zum einen können Kinder selbstständig üben zu zählen. Sie können sich dabei selbst korrigieren, wenn sie Fehler begehen. Das legt nahe, dass ihrem Verhalten zwei Quellen zugrunde liegen: eine, die das Zählverhalten generiert, und eine, die eine Überprüfung des Verhaltens auf seine Korrektheit hin ermöglicht. Zum anderen weisen falsche Zählweisen von Kindern ganz bestimmte Merkmale auf, die mit Merkmalen der korrekten Zählweise übereinstimmen. Drei- oder vierjährige Kinder zählen vier Objekte beispielsweise gelegentlich „1, 2, 5, 8“, aber fast nie „1, 8, 5, 2“, das heißt sie erhalten – auch beim Begehen von Fehlern – die ordinalen Relationen zwischen den Zahlen. Wenn man sie nach der Anzahl vorhandener Objekte fragt, so geben sie nach dem Zählen stets die Zahl an, bei der sie zu zählen aufgehört haben, unabhängig davon, ob sie richtig oder falsch gezählt haben. So könnte ein Kind vier Objekte so zählen: „1, 2, 8, 10“ und würde dann auf die

Frage nach der Anzahl der Objekte „zehn“ antworten, aber nicht „eins“, „zwei“ oder „acht“.

Durch qualitative Analysen solcher Verhaltensweisen von Kindern bei einer Reihe unterschiedlicher Zählaufgaben arbeiteten Gelman und Kollegen die wichtigsten Zählprinzipien heraus, die auch junge Kinder schon beachten (Überblick bei Gelman & Gallistel, 1978). Dieses sind: (1) das *one-one principle* – jedem Gegenstand wird beim Zählen genau eine Zahl zugeordnet; (2) das *stable order principle* – die Abfolge zwischen Zahlenwörtern bleibt über alle Trials hinweg konstant; (3) das *cardinal principle* – die letzte beim Zählen genannte Zahl benennt die Anzahl der vorhandenen Gegenstände; (4) das *abstraction principle* – es kann über beliebige Arten von Objekten hinweg gezählt werden; und (5) das *order-irrelevance principle* – die Objekte einer Menge können in beliebiger Reihenfolge gezählt werden, sofern dabei keine der anderen Zählprinzipien verletzt werden.

Gelman und Meck (1983) überprüften den Zusammenhang zwischen *principles* und *skills*, also das Verhältnis zwischen der Kenntnis der abstrakten Zählprinzipien einerseits und dem Zählverhalten andererseits. In drei Experimenten wiesen sie zunächst nach, dass Vier- und Fünfjährige über das *one-one principle*, das *stable order principle* und das *cardinal principle* verfügen, indem sie ihnen ein Stofftier vorführten, das unterschiedlich große Mengen von Gegenständen einige Male richtig zählte, einige Male jedoch gezielt gegen eines der drei Prinzipien verstieß. Die Kinder erkannten die Korrektheit oder Inkorrektheit des vorgespielten Zählverhaltens mit Wahrscheinlichkeiten, die jeweils weit über der Ratewahrscheinlichkeit lagen. Gleichzeitig waren sie jedoch nicht in der Lage, die Zählprinzipien selbst sprachlich auszudrücken. Gelman und Meck bezeichnen die Prinzipienkenntnis der Kinder daher als *implizites Wissen*.

Wenn schon junge Kinder über diese Prinzipien verfügen, wenn auch nur implizit, stellen sich zwei Fragen: (1) Warum begehen Kinder überhaupt Fehler beim Zählen? (2) Warum machen sie überproportional mehr Fehler beim Zählen großer Mengen als beim Zählen kleiner Mengen, wenn die Zählprinzipien doch unabhängig von der Mengengröße gelten? Letzteres erklären Gelman und Meck durch die Annahme, dass das Zählen großer Mengen eine anspruchsvollere Aufgabe ist als das Zählen kleiner Mengen. Die Fehler der Kinder werden durch die Aufgabenkomplexität hervorgerufen und überdecken ihr vorhandenes Wissen um die Zählprinzipien.

Einen solchen Einfluss der Aufgabenanforderungen auf das Zählverhalten von Kindern weisen sie in einem vierten Experiment nach. Dort ließen sie 36 Drei- und Vierjährige (unter anderem) Mengen von Gegenständen jeweils zweimal zählen, wobei die

Kinder jede Menge einmal beim Zählen anfassen durfte, während sie ein anderes Mal durch eine durchsichtige Plexiglasscheibe von ihnen getrennt war. Da Kinder normalerweise Gegenstände beim Zählen berühren und umherschieben, z.B. um schon gezählte von noch zu zählenden zu trennen, sollten die Aufgabenanforderungen in der zweiten Bedingung höher sein als in der ersten. Entsprechend sollten die Kinder in der ersten Bedingung, in der sie die Gegenstände berühren durften, weniger Fehler beim Zählen begehen als in der zweiten Bedingung. Zusätzlich hängen die Aufgabenanforderungen auch von der Anzahl der zu zählenden Objekte ab, so dass ein Einfluss der Objektanzahl auf die Korrektheit des Zählens erwartet wurde. Außerdem werden jüngere Kinder gewöhnlich stärker durch hohe Aufgabenanforderungen beeinträchtigt als ältere. Tatsächlich fanden Gelman und Meck hochsignifikante Haupteffekte der Einflussgrößen Versuchsbedingung (Berührung möglich vs. Berührung nicht möglich), Objektanzahl und Alter ebenso wie eine hochsignifikante Dreifachinteraktion.

Die Befunde der vier Experimente interpretieren sie dahingehend, dass bereits Drei- und Vierjährige die dem Zählen zugrunde liegenden Prinzipien kennen. Dass sie trotzdem nicht immer korrekt zählen, scheint an den sie teilweise überfordernden spezifischen Aufgabenanforderungen zu liegen.

2.2.2 Die Unterscheidung konzeptueller und prozeduraler Kompetenz

In nachfolgenden Publikationen ordnen Gelman und Kollegen diese Befunde in ein übergeordnetes Framework ein. So argumentieren Greeno, Riley und Gelman (1984) dafür, dass die beschriebenen Befunde über *principles* und *skills*, die in der Domäne des Zählens gewonnen wurden, keineswegs auf diese Domäne beschränkt sind, sondern generelle kognitive Strukturen und Prozesse exemplifizieren. Um dies zu verdeutlichen ziehen sie drei theoretische Positionen heran: die von Noam Chomsky, die von Jean Piaget und die von Newell und Simon.

Vom Psycholinguisten Chomsky übernehmen sie die Unterscheidung zwischen Kompetenz und Performanz und die Idee, die Beziehungen zwischen beiden durch theoretische Analysen zu untersuchen (z.B. Chomsky, 1965). Unter Berufung auf Piaget postulieren sie, dass die Grundlage der Kompetenz in innerpsychischen Wissensstrukturen besteht. Schließlich ziehen sie die Informationsverarbeitungstheorie des Problemlösens von Newell und Simon (1972) heran, um zu verdeutlichen, dass die Performanz, also das Verhalten einer Person beim Problemlösen, aus einem komplexen Zusammenspiel unterschiedlicher Wissenseinheiten bei der Handlungsplanung resultiert. Um dies zu illustrieren, entwickeln Greeno, Riley und Gelman in ihrem Artikel ein

Informationsverarbeitungsmodell der kognitiven Strukturen und Prozesse, die es Kindern ermöglichen zu zählen. Sie analysieren die Performanz von Kindern beim Zählen, um auf die dahinter stehenden Kompetenzen schließen zu können. Wie die oben besprochenen Befunde demonstrieren, besitzen Kinder gelegentlich Wissen um Zählprinzipien, ohne korrekt zählen zu können. Auch der umgekehrte Fall – jemand kann korrekt zählen, kennt aber die Prinzipien nicht – ist vorstellbar. Daher unterscheiden die Autoren zwischen einer konzeptuellen und einer prozeduralen Kompetenz der Kinder. Entsprechend den informationstheoretischen Annahmen in ihrem theoretischen Framework führen sie diese unterschiedlichen Kompetenzen auf zwei unterschiedliche dahinter liegende Wissensarten zurück: konzeptuelles und prozedurales Wissen. Aus verschiedenen Gründen führen sie zusätzlich eine dritte Kompetenz ein, die Anwendungskompetenz, die situationsangemessenes Handeln ermöglicht. Sie charakterisieren die konzeptuelle und die prozedurale Kompetenz und die dahinter stehenden Wissensrepräsentationen folgendermaßen:

The analysis has three components: conceptual, procedural and utilizational competence. Conceptual competence, which we have discussed in detail in this paper, is represented as a set of schemata that constitute conceptual structures in a task domain. [...] In analyzing counting, we chose a set of schemata that correspond to principles of cardinality, order and one-to-one correspondence, but a higher or lower level of abstraction could have been used.

The derivations also use procedural competence, in the form of heuristic planning rules [...]. The planning heuristics play the role of inference rules in the derivations and correspond to general competence underlying procedural knowledge. [...]

We view principles included in conceptual competence as constraints on procedural knowledge.
(Greeno et al., 1984, S. 136-138)

Die Autoren erläutern, dass die Beziehung zwischen inhaltlichen Konzepten und ihrer mentalen Repräsentation in Schemata noch unklar ist. Es kann nicht a priori von Eins-zu-Eins-Zuordnungen ausgegangen werden, sondern verschiedene Aspekte eines Prinzips könnten in unterschiedlichen Schemata gespeichert werden. Die Autoren konzentrieren sich in dem Artikel vor allem auf die Analyse der konzeptuellen Kompetenz und räumen ein, nur wenige Aussagen über die Natur prozeduralen Wissens machen zu können. Insgesamt gehen sie davon aus, dass die konzeptuelle Kompetenz die Performanz stärker leitet als die anderen Teilkompetenzen und auch Erwerb und Abruf prozeduralen Wissens beeinflusst.

Greeno und Kollegen diskutieren auch die forschungsmethodischen Schwierigkeiten bei der Untersuchung der konzeptuellen und prozeduralen Kompetenz, die sich daraus

ergeben, dass beide anhand des Verhaltens eines Menschen schlecht zu trennen sind. Diese Punkte werden in Abschnitt 4.5 näher erläutert.

Festzuhalten bleibt, dass Gelmans ursprüngliche Unterscheidung zwischen principles und skills verhaltensnah und damit – auch in ihren Implikationen für die Praxis – leicht nachzuvollziehen, jedoch kaum theoretisch unterfüttert war. Durch die Strategie, die Dichotomie von principles und skills aufzufassen als Unterscheidung zwischen Teilkompetenzen, denen innerpsychisch wiederum unterschiedliche Wissensarten zugrunde liegen, ergaben sich vielfältige Möglichkeiten, principles und skills mit Methoden und Erkenntnissen der kognitiven Lernforschung in Verbindung zu bringen und ihre Untersuchung als Weiterentwicklung kognitiver Lerntheorien aufzufassen. Jedoch wirft das beschriebene Vorgehen auch die Fragen auf, ob die Befunde zu den principles und skills des Zählens nicht einfacher interpretiert werden könnten. Insbesondere ist fraglich, ob die vorhandenen empirischen Befunde tatsächlich die aufwändige Annahme rechtfertigen, dass die Befunde zwei unterschiedliche Wissensarten mit unterschiedlichen mentalen Repräsentationsformen exemplifizieren, über denen unterschiedliche kognitive Mechanismen operieren. Da Gelman und Kollegen den Kindern unterschiedliche Aufgaben stellten, von denen die Kinder einige lösen konnten und andere nicht, ließen sich die Resultate beispielsweise durch die Annahme der Aufgabenspezifität von Wissensseinheiten einer einzigen Wissensart effizienter erklären als durch die Annahme zweier Wissensarten.

In jedem Fall haben die Arbeiten Gelmans und anderer zu einem verstärkten Interesse Pädagogischer Psychologen an den Wissensarten und ihrer Interaktion beim Lernen geführt. So entstanden nachfolgend zahlreiche Einzelarbeiten sowie der von James Hiebert (1986) herausgegebene Sammelband *Conceptual and Procedural Knowledge: The Case of Mathematics*, zu dem auch Gelman und Meck einen Beitrag liefern, in dem sie erneut den Zusammenhang zwischen principles und skills besprechen (Gelman & Meck, 1986). Die wichtigsten vor 1998 erschienenen Studien werden im folgenden Abschnitt besprochen.

2.3 Der Literaturüberblick von Rittle-Johnson und Siegler

2.3.1 Untersuchte Inhaltsbereiche und Evidenztypen

Rittle-Johnson und Siegler (1998) geben einen Überblick über die wichtigsten vor dem Erscheinungsjahr publizierten Studien über das Zusammenspiel von konzeptuellem und prozeduralem Wissen beim Mathematiklernen. Er stellt den bisher einzigen publizierten Literaturüberblick in diesem Forschungsfeld dar. Aufgrund seiner zentralen Bedeutung wird er hier in einem eigenen Unterkapitel vorgestellt.

Rittle-Johnson und Siegler thematisieren in ihrem Überblick nicht die unterschiedlichen Maße und Definitionen der Wissensarten, die in den Studien jeweils verwendet wurden. Darum werden diese im aktuellen Unterkapitel nicht einzeln vorgestellt. Stattdessen werden diese beiden Punkte in Kapitel 4 ausführlich analysiert. In einer vorläufigen Annäherung lässt sich sagen, dass die im aktuellen Kapitel besprochenen Arbeiten konzeptuelles Wissen als abstraktes Tiefenverständnis auffassen und es oft über Erklärungs-, Bewertungs- oder Transferaufgaben erheben. Prozedurales Wissen wird als effizient anwendbares Handlungswissen zur Lösung von Routineproblemen aufgefasst und häufig durch die Lösungsrate bei Routineproblemen operationalisiert.

Die von Rittle-Johnson und Siegler berichteten Befunde zu den Wissensarten stammen aus fünf verschiedenen Domänen: dem Zählen, der Addition einstelliger Zahlen, der Arithmetik mehrstelliger Zahlen, dem Bruchrechnen und dem proportionalen Denken. Die Studien untersuchten hauptsächlich Kinder im Grundschulalter, verwendeten dazu jedoch sehr unterschiedliche Designs. Rittle-Johnson und Siegler listen sieben verschiedene Typen erbrachter Evidenz auf: (1) Beobachtete Prozeduren verletzen keine Konzepte; (2) Kinder kennen gleichzeitig ein Konzept und eine korrespondierende Prozedur; (3) intervallskalierte Maße konzeptuellen und prozeduralen Wissens korrelieren positiv; (4) eine Prozedur ist vor einem Konzept vorhanden oder umgekehrt; (5) das Vorwissen in einer der beiden Wissensarten sagt den Zuwachs der anderen vorher; (6) Befunde aus Trainingsstudien ohne Kontrollgruppe; (7) Ergebnisse kontrollierter Experimente. In einer Übersicht der 34 wichtigsten Befunde wurden 29 von Rittle-Johnson und Siegler den Evidenzkategorien 1 bis 4 zugeordnet. Es handelt sich dabei also um korrelative Befunde aus Querschnittsstudien. Lediglich eine Studie lieferte längsschnittliche Befunde, vier untersuchten den Einfluss unterschiedlicher Trainings auf die Wissensarten. Es lag kein einziges Experiment mit Kontrollgruppe und Randomisierung der Studienteilnehmer auf die Gruppen vor. Bei der Interpretation der Studienanzahlen ist zu beachten, dass dieselbe Studie mehrere Arten von Evidenz liefern kann.

2.3.2 Korrelative Relationen

Sechzehn der vierunddreißig besprochenen Befunde zeigen durch Messung konzeptuellen und prozeduralen Wissens jeweils in einer Stichprobe, dass beide Wissensarten über Personen hinweg entweder qualitativ oder quantitativ zusammenhängen. Diese Befunde sind relativ einheitlich. Eine Ausnahme bilden drei Studien: Resnick fand in zwei Studien (Resnick, 1982; Resnick & Omanson, 1987) keine signifikante Korrelation zwischen beiden Wissensarten. Byrnes und Wasik (1991) fanden für eine Stichprobe von Viert- und

Sechstklässlern einen signifikanten Zusammenhang, der jedoch verschwand, wenn man das Alter der untersuchten Kinder auspartialisierte.

2.3.3 Erwerbsreihenfolge in Querschnittsdesigns

Uneinheitlich waren die Befunde von 13 querschnittlichen Studien zur Erwerbsreihenfolge der Wissensarten: Im Bereich des Zählens und der Multiplikation von Brüchen entwickelten sich Prozeduren anscheinend vor Konzepten. In den Bereichen proportionales Denken, Addition von Brüchen und Addition einstelliger Zahlen zeigten sich die Konzepte vor den Prozeduren. Der zeitliche Abstand war besonders groß beim proportionalen Denken, wo die Kinder die relevanten Konzepte mehrere Jahre vor den Vergleichs- und Rechenprozeduren erwarben (Dixon & Moore, 1996). Die Befunde im Bereich der Addition und Subtraktion mehrstelliger Zahlen sind widersprüchlich.

Rittle-Johnson und Siegler weisen darauf hin, dass in allen Domänen die Befunde eher uneinheitlich sind und auch innerhalb von Domänen große inter- und intrapersonale Unterschiede herrschen. Man beachte, dass die neueren Studien über die Erwerbsreihenfolge konzeptuellen und prozeduralen Wissens über das Zählen (z.B. Wynn, 1990) den oben berichteten Befunden von Gelman und Kollegen widersprechen, vermutlich aufgrund methodischer Unterschiede zwischen den Studien.

2.3.4 Längsschnittliche Befunde

Hiebert und Wearne (1996) erhoben in der einzigen bislang durchgeführten Längsschnittstudie das konzeptuelle und prozedurale Wissen von 70 Kindern über Addition und Subtraktion an mehreren Messzeitpunkten pro Jahr über die ersten drei Grundschuljahre hinweg.

In diesem Zeitraum wurden 17 Kinder traditionell unterrichtet, d.h. der Unterricht war lehrbuchbasiert und der Einführung neuer Prozeduren durch den Lehrer folgten Übungsphasen. Neun andere Kinder erhielten hingegen einen eher konstruktivistisch ausgerichteten Unterricht, der auf gemeinschaftlicher Reflektion über selbstentwickelten Prozeduren und dem Einsatz externer Repräsentationsformen, wie Dienes-Blöcken, beruhte. Hiebert und Wearne selbst sprechen jedoch nicht von konstruktivistischem, sondern, vorsichtiger, lediglich von *alternativem* Unterricht. Die übrigen 44 Kinder erhielten aufgrund logistischer Gründe verschiedene Mischungen beider Unterrichtsformen.

Die traditionell und die konstruktivistisch unterrichtete Gruppe unterschieden sich ausschließlich am letzten Messzeitpunkt voneinander, was angesichts des dreijährigen Untersuchungszeitraumes ein bemerkenswerter Befund ist. Die Unterschiede waren eher

qualitativer als quantitativer Art: Alle Kinder konnten etwa gleich viele konzeptuelle und prozedurale Aufgaben lösen. Die Kinder aus der eher konstruktivistisch unterrichteten Gruppe konnten jedoch Bedeutung und Funktionsweise der Prozeduren besser durch Rückgriff auf Konzepte erklären, was sich auch in höheren Korrelationen zwischen konzeptuellem und prozeduralem Wissen niederschlug. Über alle 70 Kinder hinweg zeigte sich, dass Kinder *nach* der Unterrichtseinheit zu einem Thema normalerweise über beide Wissensarten in dem Bereich verfügten. *Vor* der Unterrichtseinheit besaßen sie öfter konzeptuelles Wissen ohne prozedurales Wissen als prozedurales Wissen ohne konzeptuelles. Das lässt darauf schließen, dass das konzeptuelle Wissen sich vor dem prozeduralem entwickelte. Kinder, die schon vor einer Unterrichtseinheit konzeptuelles Wissen besaßen, entwickelten früher und öfter neue Prozeduren als Kinder ohne konzeptuelles Wissen. Das zeigt, dass konzeptuelles Wissen eine Quelle prozeduralen Wissens darstellen kann. Diese Zusammenhänge waren in den konstruktivistisch unterrichteten Klassen stärker ausgeprägt als in den konventionell unterrichteten. Ein weiterer Befund ist, dass Kinder mit hohem konzeptuellem Wissen mehr anspruchsvolle Transferprobleme lösten, das heißt Prozeduren besser auf neue Wissensgebiete übertragen konnten, und außerdem erworbene Prozeduren über die Zeit hinweg seltener vergaßen als Kinder mit niedrigem konzeptuellem Verständnis.

Leider wird die Validität dieser differenzierten und interessanten Befunde durch einige methodische Schwächen eingeschränkt, die für alle in diesem Abschnitt besprochenen Arbeiten typisch sind (s. Rittle-Johnson & Siegler, 1998), wie beispielsweise der Einsatz dichotomer Einzelitems statt intervallskaliierter Summenscores, unanalyisierte Reliabilitäten und – aufgrund einer nicht theoriegeleiteten Aufgabenauswahl – unbekannte Validitäten der Aufgaben, kleine Stichproben und mangelnde Randomisierung der Teilnehmer auf die Gruppen.

2.3.5 Trainingsstudien

Neben Hiebert und Wearne (1996) untersuchten drei weitere Forschergruppen konzeptuelles und prozedurales Wissen im Rahmen von Trainingsstudien. Im Gegensatz zu Erstgenannten nutzten sie jedoch Prätest-Posttest-Designs.

Byrnes und Wasik (1991) erhoben in ihrer Studie 2 konzeptuelles und prozedurales Wissen von Fünftklässlern über Bruchrechenaufgaben vor und nach drei kurzen Interventionen. Entgegen ihren Erwartungen wirkten sich die Interventionen, in denen jeweils unterschiedliche Informationen über das Rechnen mit Brüchen vermittelt wurden, nicht unterschiedlich auf den Erwerb beider Wissensarten aus. Über die Gruppen hinweg konnten die Autoren durch Vergleich der Messzeitpunkte zeigen, dass die Kinder

tendenziell konzeptuelles Wissen zeitlich früher erwerben als prozedurales. Das konzeptuelle Wissen im Prätest wirkte sich signifikant positiv auf das prozedurale Wissen, das heißt die Bruchrechenleistung, im Posttest aus. Inwieweit sich das prozedurale Vorwissen auf die konzeptuellen Lernerfolge auswirkt, wurde nicht untersucht.

Fuson und Briars (1990) maßen das konzeptuelle und prozedurale Wissen mehrerer hundert Erst- und Zweitklässler vor und nach einer speziellen Unterrichtseinheit, die – entgegen der damaligen schultypischen Vorgehensweise – unter Verwendung von Dienes-Blöcken schon so jungen Kindern das Addieren und Subtrahieren mehrstelliger Zahlen beibringen sollte. Die Relationen zwischen den beiden Wissensarten wurden jedoch weder eingehend theoretisch diskutiert noch empirisch untersucht. Die Wissensarten wurden lediglich getrennt gemessen, um die Breite der Lernerfolge besser abschätzen zu können. Die Autoren berichten einen signifikanten Anstieg beider Wissensarten von der Prätest- zur Posttestmessung. Da es keine Kontrollgruppe gab, können jedoch auch diese Erfolge nur unzureichend analysiert werden. Lediglich bei einigen Erstklässlern zeigten sich Unterschiede zwischen den Wissensarten: Die Kinder konnten die im Unterricht vermittelten Prozeduren zwar anwenden, sie jedoch konzeptuell nicht erklären. Fuson und Briars schließen daraus, dass man die Addition und Subtraktion von mehrstelligen Zahlen noch nicht in der ersten, wohl aber in der zweiten und dritten Klasse problemlos vermitteln kann. Dieses Themengebiet bis zur fünften Klasse zu unterrichten, wie es in den USA zum Zeitpunkt ihrer Studie geschah, halten sie für überflüssig.

Angesichts dieser positiven Befunde sind die negativen Resultate einer Studie von Resnick und Omanson (1987) erstaunlich: Sie vermittelten neun Viert-, Fünft- und Sechstklässlern, die bereits in einem Vortest falsche Subtraktionsprozeduren benutzt hatten, mit Hilfe von Dienes-Blöcken in jeweils 80-minütigen Einzelunterrichtseinheiten erfolgreich konzeptuelles Wissen über die Subtraktion mehrstelliger Zahlen. Acht der neun Kinder benutzten im Posttest immer noch fehlerhafte Prozeduren, obwohl ihr konzeptuelles Wissen und ihre Fähigkeit, Probleme mit Hilfe der Blöcke zu veranschaulichen, signifikant zugenommen hatten.

Das zeigt, dass sich konzeptuelles Wissen nicht unter allen Umständen positiv auf prozedurales auswirkt. Die Gründe könnten darin liegen, dass es den Kindern nicht gelang, ihre neu gewonnenen Konzepte zu den schon vorhandenen Prozeduren in Beziehung zu setzen, eventuell weil ihre fehlerhaften Prozeduren schon zu stark automatisiert waren. Dieser Zusammenhang ist jedoch nicht empirisch belegt.

2.3.6 Diskussion der Studien

Rittle-Johnson und Siegler (1998) heben im Diskussionsteil ihres Reviews die Wichtigkeit und die Interessantheit der vorliegenden Befunde hervor. Ihr gut strukturierter Literaturreview lässt erkennen, welche inhaltlichen Grundfragen für die Forschung zum Erwerb mathematischer Konzepte und Prozeduren von besonderem Interesse sind und wie diese Fragen methodisch bisher angegangen wurden: (1) Wie hängen konzeptuelles und prozedurales Wissen an einem Messzeitpunkt qualitativ zusammen und wie hoch sind sie quantitativ korreliert? (2) Welche deskriptive Erwerbsreihenfolge lässt sich für die beiden Wissensarten in Querschnitts- und Längsschnittstudien beobachten und welche Erwerbsreihenfolge ist Trainingsstudien zufolge optimal? (3) Welche Kausalrelationen fördern längsschnittliche Studien und kontrollierte Experimente zu Tage?

Rittle-Johnson und Siegler ziehen allerdings das Fazit, dass zur Zeit ihres Reviews keine dieser Fragen befriedigend beantwortet werden kann, da die vorliegenden Studien größeren methodischen Einschränkungen unterliegen.

Der gravierendste Mangel ist, dass – trotz des Vorliegens einiger interessanter Trainingsstudien – noch kein Experiment mit Treatment- und Kontrollgruppen sowie einer Randomisierung der Studienteilnehmer auf diese Gruppen durchgeführt wurde. Die prädiktiven Beziehungen zwischen konzeptuellem und prozeduralem Wissen in Längsschnittstudien könnten aufgrund der fehlenden Randomisierung mit anderen Einflüssen, zum Beispiel von Intelligenz und Motivation, konfundiert sein und liefern daher weniger valide Evidenz für Kausalbeziehungen zwischen den Wissensarten als Experimente liefern könnten. Außerdem liegen bislang lediglich die Befunde einer einzigen Längsschnittstudie vor.

Die Befunde zur Erwerbsreihenfolge der Wissensarten unterliegen darüber hinaus einer weiteren Einschränkung: Wenn Kinder einer Altersstufe eine Aufgabenart lösen können, die konzeptuelles Wissen operationalisieren soll, und eine Aufgabenart nicht lösen können, die prozedurales Wissen operationalisieren soll, so gibt das nicht zwangsläufig Aufschluss über die Verteilung ihres Wissens. Die Ergebnisse könnten alternativ auch durch unterschiedliche Schwierigkeiten der beiden Aufgabentypen erklärt werden: Leichtere Aufgaben können schon von jüngeren Kindern gelöst werden, schwierigere erst von älteren.

Die häufig gefundenen positiven Korrelationen zwischen konzeptuellem und prozeduralem Wissen geben wegen der potenziellen Konfundierung mit anderen Variablen kaum Auskunft über die tatsächlichen Beziehungen zwischen den Wissensarten. Intelligenter Kinder könnten zum Beispiel mehr Wissen beider Arten erworben haben als

weniger intelligente – bei gleichzeitiger Unabhängigkeit der Wissensarten voneinander. Dass solche Erklärungen nicht aus der Luft gegriffen sind, zeigt der oben berichtete Befund von Byrnes und Wasik (1991): Die ursprünglich gefundene Korrelation zwischen den Wissensarten verschwand, wenn sie das Alter der Kinder kontrollierten.

Rittle-Johnson und Siegler ziehen das Fazit, dass methodische Verbesserungen zukünftiger Studien, wie die Operationalisierung der Wissensarten durch jeweils sorgfältig konstruierte intervallskalierte Maße, Längsschnittstudien mit mehreren Messzeitpunkten und kontrollierte experimentelle Designs, die Forschung auch inhaltlich wesentlich voranbringen könnten. Als bisherigen Stand der Forschung fassen sie zusammen, dass konzeptuelles und prozedurales Wissen zwei unterscheidbare, aber zusammenhängende Konstrukte sind, die – unter bis dato nicht näher bekannten Umständen – differenziell durch Interventionen sowie durch die Alltagserfahrung von Kindern beeinflusst werden können.

2.4 Neuere Befunde

2.4.1 Korrelative Befunde und Erwerbsreihenfolge

Seit dem Literaturreview von Rittle-Johnson und Siegler (1998) sind eine Reihe weiterer Forschungsergebnisse veröffentlicht worden. Sie ergänzen das schon vorhandene Befundbild, verändern es aber nicht wesentlich.

An erster Stelle sind hier die Arbeiten von Canobi und Kollegen zu nennen. In vier korrelativen Studien untersuchten sie die Zusammenhänge zwischen dem konzeptuellen und dem prozeduralen Wissen über Addition und Subtraktion in verschiedenen Stichproben von fünf- bis achtjährigen Kindern. Sie fanden in Übereinstimmung mit älteren Arbeiten positive Zusammenhänge zwischen den beiden Wissensarten. Die Befunde von Canobi und Kollegen sind interessant, weil sie beide Wissensarten hochauflösender erhoben als es früher der Fall war. So maßen Canobi, Reeve und Pattison (1998) das prozedurale Wissen nicht mehr nur dichotom durch die Problemlösekorrektheit (Kind zählt korrekt vs. Kind zählt inkorrekt), wie z.B. Gelman und Meck es taten, sondern sie erhoben die Korrektheit der Problemlösungen in Form von (a) Summenscores der Lösungskorrektheiten über mehrere Aufgaben eines Typs, (b) der mittleren Problemlösegeschwindigkeit und (c) der Art der von den Kindern verwandten Problemlösestrategien, wie Retrievalstrategie, Dekompositionsstrategie usw., über 46 Additionsaufgaben hinweg. Das konzeptuelle Wissen der Kinder maßen sie, indem sie ihnen jeweils zwei Additionsaufgaben gleichzeitig auf einem Bildschirm präsentierten. Hinter der ersten Aufgabe stand schon ihre Lösung, hinter der zweiten Aufgabe noch nicht.

Die Kinder wurden dann jeweils gefragt, ob man die Lösung der ersten Aufgabe benutzen kann, um die zweite zu lösen. In einem Teil der Fälle war dies unter Anwendung bestimmter Umformungsregeln möglich, in einem anderen Teil der Fälle nicht. Beispielsweise hätten die Kinder die Aufgaben „ $2 + 7 = 9$ “ und „ $7 + 2 = ?$ “ auf dem Bildschirm sehen können. Wenn die Kinder erkannten, dass man die erste Aufgabe nutzen kann, um die Lösung der zweiten Aufgabe zu bestimmen, wurde das als Indiz dafür gewertet, dass sie konzeptuelles Wissen über das Kommutativitätsgesetz besitzen. In ähnlicher Weise wurde auch ihr konzeptuelles Wissen über die Zusammensetzbarkeit kleinerer Zahlen zu größeren (*numerische Komposition*) und ihr Wissen über das Assoziativitätsgesetz erhoben.

Canobi, Reeve und Pattison fanden, dass fast all ihre konzeptuellen und prozeduralen Maße mittelstark miteinander korrelierten. Außerdem gab es in ihrer Stichprobe mehr Personen mit hohem konzeptuellen und niedrigem prozeduralen Wissen als umgekehrt.

In einer Clusteranalyse über die verschiedenen Maße hinweg fanden Canobi und Kollegen sechs Personencluster, die sich durch unterschiedliche Kompetenzprofile beschreiben ließen. Die Autoren begründen jedoch nicht stichhaltig, warum sie gerade die Lösung mit sechs Clustern gewählt haben. Da ein Cluster nur eine Person beinhaltet, scheint ihre Wahl nicht optimal zu sein. Die Autoren verzichteten auch auf eine Benennung oder inhaltliche Interpretation der Cluster und berichten lediglich deskriptiv die Mittelwertsunterschiede zwischen den Clustern.

In drei nachfolgenden Arbeiten kommen Canobi und Kollegen im selben Inhaltsgebiet, jedoch mit unterschiedlichen Operationalisierungen zu ähnlichen Befunden: die beiden Wissensarten hängen eng zusammen, jedoch zeigen sich im Rahmen der Analyse interindividuelle Unterschiede auch teilweise Dissoziationen der Wissensarten (Canobi, 2004; Canobi, Reeve, & Pattison, 2002; Canobi, Reeve, & Pattison, 2003). Drei interessante zusätzliche Befunde dieser Querschnittsstudien sind: (1) Die Korrelation zwischen konzeptuellem und prozeduralem Wissen blieb auch dann bestehen, wenn man das Alter der Kinder kontrolliert (Canobi et al., 2003); (2) das Alter der Kinder (fünf bis acht Jahre) war positiv mit ihrem prozeduralen, nicht jedoch mit ihrem konzeptuellen Wissen korreliert (Canobi et al., 2003); (3) auch ihre Klassenstufe (eins bis drei) hing zwar positiv mit ihrem prozeduralen, nicht aber ihrem konzeptuellen Additions- und Subtraktionswissen zusammen. Da die Kinder in den Grundschuljahren das Addieren und Subtrahieren lernen, sind diese Befunde überraschend. Falls es sich nicht um methodische Artefakte handelt, hätten die Kinder in der Schule ausschließlich gelernt, mechanisch Rechenprozeduren zu befolgen, ohne ihre eigenen Handlungen inhaltlich zu verstehen.

In einer gerade abgeschlossenen Studie griffen Gilmore und Bryant (zur Publikation angenommen) Canobis Idee, clusteranalytisch Gruppen mit distinkten Profilen konzeptuellen und prozeduralen Wissens zu bilden, erneut auf und gelangten zu eindeutigeren Ergebnissen. In zwei Studien wurden insgesamt 127 Sechs- bis Neunjährige in Bezug auf ihr konzeptuelles und prozedurales Wissen über Inversionsprobleme getestet. In die Analysen gingen dann die Lösungsscores aller konzeptuellen und prozeduralen Einzelaufgaben ein. Die Clusteranalysen beider Studien führten zu ähnlichen Ergebnissen, nämlich jeweils einer Drei-Cluster-Lösung. Die Personen im ersten Cluster besaßen viel konzeptuelles und viel prozedurales Wissen, die Personen im zweiten Cluster besaßen wenig konzeptuelles und wenig prozedurales Wissen, die Personen im dritten Cluster besaßen viel konzeptuelles und wenig prozedurales Wissen. Die Clusterlösungen erklärten in beiden Studien jeweils ungefähr 60% der Varianz. Diese Befunde verweisen, so Gilmore und Bryant, nicht nur auf die partielle Unabhängigkeit der Wissensarten voneinander, sondern auch auf die Wichtigkeit der Berücksichtigung von interindividuellen Unterschieden, da die Mittelwerte der Gesamtgruppe für keine der drei gefundenen Subgruppen repräsentativ sind.

Hecht (1998) untersuchte durch den Vergleich verschiedener *structural equation models* (SEM) die Beziehungen zwischen Teilkompetenzen des Bruchrechnens. Er fand, dass verschiedene Maße konzeptuellen Wissens, wie Aufgaben zum Verständnis von Teil-Ganzem-Relationen, und ein Maß prozeduralen Wissens positiv korreliert waren und beide unabhängig voneinander einen Anteil der Gesamtvarianz der Bruchrechnenfähigkeit der Kinder erklärten. Im Gegensatz zu den meisten anderen oben besprochenen Arbeiten fasst er jedoch nicht die Korrektheit der Problemlösungen der Kinder als Maß ihres prozeduralen Wissens auf, sondern ließ sie jeweils vier Prozeduren daraufhin bewerten, ob sie zur Lösung eines Rechenproblems geeignet sind oder nicht. Diese Operationalisierung wurde jedoch von anderen Autoren zur Erhebung konzeptuellen Wissens und gerade nicht zur Erhebung prozeduralen Wissens benutzt (siehe Abschnitt 4.5), so dass die Ergebnisse der Studien kaum vergleichbar sind. Mittels konfirmatorischer Faktorenanalysen zeigte er, dass seine Maße konzeptuellen Wissens zwei hochkorrelierte, aber signifikant voneinander verschiedene Faktoren bilden. Konzeptuelles Wissen, so wie es in seiner Studie gemessen wurde, stellt also kein eindimensionales Konstrukt dar. Es ist jedoch zu erwarten, dass solche Befunde stark von der Art der verwendeten Aufgaben und Operationalisierungen abhängen. Hier wären Folgeuntersuchungen hilfreich. Da Hecht prozedurales Wissen nur über eine Aufgabenart operationalisiert hat, ging diese Wissensart nicht in seine konfirmatorischen Faktoranalysen mit ein und es lässt sich nicht sagen,

inwieweit konzeptuelles und prozedurales Wissen zwei unabhängig voneinander variierende Faktoren sind. Auch in dieser Richtung ließe sich sein Design sinnvoll erweitern.

2.4.2 Trainingsstudien

Howe, Tolmie, Duchak-Tanner und Rattray (2000) veröffentlichten nicht nur eine der wenigen Trainingsstudien, sondern auch eine der seltenen Untersuchungen zu den Wissensarten, die nicht im Bereich des Mathematiklernens, sondern des Erwerbs naturwissenschaftlicher Kompetenzen angesiedelt sind. Sie maßen das konzeptuelle und prozedurale Wissen von Grundschulern über wissenschaftliches Hypothesentesten im Bereich der physikalischen Optik vor und nach vier verschiedenen Interventionen. Dabei nahmen sie eine pädagogische Perspektive ein und verzichteten auf die Bezugnahme zu kognitiven Ansätzen. Ihre Operationalisierungen von konzeptuellem und prozeduralem Wissen stimmen darum nicht mit denen anderer Studien überein. Sie fassten konzeptuelles Wissen auf als Wissen über die Mechanismen und Ereignisse, die in den Experimenten untersucht wurden, und erhoben es durch einen multiple choice-Test. Prozedurales Wissen sahen sie an als verbalisierbares Wissen darüber, wie man Experimente gestaltet und auswertet. Sie erhoben es durch Einzelinterviews mit allen Kindern. Hier stellt sich die Frage, ob die Interviews wirklich prozedurales Wissen oder nicht vielmehr das konzeptuelle Verständnis der Kinder darüber, wie Experimente funktionieren, erfassten.

Die Kinder in den vier Treatmentgruppen entwarfen während der Treatmenterteilung in Kleingruppen ein Experiment, wobei die verschiedenen Gruppen unterschiedliche Mischungen von Gruppenarbeitsphasen und Lehrerfeedback erhielten. Anschließend unterschieden sich die Gruppen signifikant in ihrem konzeptuellen und prozeduralen Wissen voneinander, wobei die Gruppenarbeit sich eher positiv auf das konzeptuelle und ein lehrerzentrierter Unterricht sich eher positiv auf das prozedurale Wissen der Kinder auswirkte. Diese interessanten Ergebnisse sollten jedoch nicht überinterpretiert werden, da einerseits, wie oben beschrieben, unklar ist, welche Konstrukte durch die Tests erhoben wurden, und da andererseits das Treatment konfundiert war mit den spezifisch stattgefundenen Lehrereffekten und Gruppenprozessen, die sich in Unterrichtsstudien nur schwer kontrollieren lassen.

Porter und Masingila (2000) untersuchten bei College-Studenten durch Messung des konzeptuellen und prozeduralen Wissens vor und nach einem Collegesemester, ob eine Gruppe, die nach dem *writing to learn mathematics*-Ansatz (WTLM) unterrichtet wurde, signifikant besser abschneiden würde als eine vergleichbare, aber herkömmlich unterrichtete Gruppe. Dies war nicht der Fall und es zeigten sich auch keine differenziellen

Effekte für die Wissensarten. Eine Ursache dafür könnte in der eher ungewöhnlichen Art bestehen, in der die Autoren die Wissensarten erhoben: Nach einem selbstentwickelten Kategoriensystem, das sie in ihrem Artikel leider nicht näher charakterisiert haben, klassifizierten sie jeden in einem Rechentest gemachten Fehler entweder als konzeptuellen oder als prozeduralen Fehler. Je weniger Fehler eine Person machte, umso höher sollte dann ihr Wissen sein.

Eine weitere Studie zur Evaluation von Unterrichtsprogrammen, in denen konzeptuelles und prozedurales Wissen erhoben wurden, stammt von Blöte, Burg und Klein (2001). Sie erhoben das Wissen von 10 zweiten Klassen im Abstand von drei Monaten. Dazwischen nahm eine Hälfte der Klassen an einem Arithmetikunterricht nach dem Realistik Program Design (RPD) teil, die andere Hälfte an einem Unterricht nach dem Gradual Program Design (GPD). Das RPD legt den Schwerpunkt auf alltagsnah situierte Probleme sowie der Selbstentdeckung und gemeinschaftlichen Diskussion neuer Rechenprozeduren, beinhaltet jedoch auch Einheiten mit expliziten Instruktionen. Das GPD beinhaltet größere Anteile expliziter Instruktionen und erheblich mehr *paper and pencil*-Übungen als das RPD, dagegen treten Selbstentdeckung und Diskussionen zwischen den Schülerinnen und Schülern in den Hintergrund.

Blöte, van der Burg und Klein kamen nach sorgfältigen Auswertungen zu dem Schluss, dass die Schüler in der moderat-konstruktivistischen RPD-Bedingung mehr alternative Prozeduren zur Lösung der Rechenprobleme gelernt hatten. Darüber hinaus konnten sie ihre Prozeduren auch flexibler einsetzen als die Schüler in der GPD-Bedingung. Ein dritter Befund ist, dass die RPD-Schüler ihnen präsentierte Prozeduren besser daraufhin bewerten konnten, ob sie adäquat oder inadäquat zur Lösung einer Aufgabe sind. Interessanterweise hingen diese Bewertungen und der Einsatz von Prozeduren jedoch kaum zusammen. Einige von den Studienteilnehmern als wertvoll beurteilte Prozeduren wurden nur selten angewandt und andere, selten für gut befundene, häufig benutzt. Dies führt Blöte und Kollegen zu dem Schluss, dass ihre Aufgaben unterschiedliche Wissensarten messen: Während der tatsächliche Einsatz von Prozeduren das prozedurale Wissen der Kinder widerspiegelt, könnte die Bewertung von Prozeduren eher ihr konzeptuelles Verständnis indizieren. Folgt man dieser Annahme der Autoren ergibt sich folgendes Bild: In der eher konstruktivistischen RPD-Bedingung erwarben die Studienteilnehmer nicht nur mehr prozedurales, sondern auch mehr konzeptuelles Wissen. Da sich das konzeptuelle Wissen im Durchschnitt eher zeigte als das prozedurale, führen die Autoren in der Diskussion sogar den Anstieg des prozeduralen Wissens in der RPD-Bedingung auf das in der Bedingung zuvor erworbene konzeptuelle Wissen zurück. Der

Zusammenhang zwischen den beiden Wissensarten war in der RPD-Bedingung höher als in der GPD-Bedingung. Die Autoren weisen zu Recht auf die bemerkenswerte Übereinstimmung ihrer Befunde mit denen der oben zusammengefassten Längsschnittstudie von Hiebert und Wearne (1996) hin.

2.4.3 Diskussion der neueren Befunde

Inhaltlich unterscheiden sich die neueren kaum von den älteren Befunden: Konzeptuelles und prozedurales Wissen scheinen miteinander korreliert zu sein, sich aber auch teilweise unabhängig voneinander entwickeln zu können. Besonders interessant ist, dass Blöte, Burg und Klein (2001) in Übereinstimmung mit Hiebert und Wearne (1996) fanden, dass die Unterrichtsgestaltung die Relationen zwischen den Wissensarten beeinflusst. Kinder, die an eher konstruktivistischem Unterricht teilgenommen hatten, besaßen zwar weder mehr konzeptuelles noch mehr prozedurales Wissen als herkömmlich unterrichtete Kinder, konnten Prozeduren aber besser konzeptuell erklären und auf neue Inhaltsgebiete transferieren. Der Unterricht scheint sich also zumindest auf die Verknüpfung der beiden Wissensarten positiv ausgewirkt zu haben. Warum er in beiden Studien nicht auch ihre Niveaus differenziell beeinflusste, bleibt offen. Dass pädagogische Interventionen überhaupt zu unterschiedlich großen Zuwächsen der Wissensarten führen können, zeigte die Interventionsstudie von Howe, Tolmie, Duchak-Tanner und Ratray (2000). Sie fanden, dass lehrergeleitete Übungsphasen eher zu einem Anstieg des prozeduralen Wissens und Gruppendiskussionen eher zu einem Anstieg des konzeptuellen Wissens führen. Im Gegensatz dazu hatte die Intervention von Porter und Masingila (2000) keine differenziellen Auswirkungen auf die Wissensarten. Die Frage der tatsächlichen und der optimalen Erwerbsreihenfolge beider Wissensarten ist aufgrund des Fehlens mehrphasiger Interventions- und Längsschnittsstudien weitgehend ungeklärt geblieben, auch wenn einige ältere und neuere Querschnittsstudien nahe legen, dass sich das konzeptuelle Wissen entweder gleichzeitig oder früher zeigt als das prozedurale (z.B. Baroody & Gannon, 1984; Byrnes & Wasik, 1991; Cowan & Renton, 1996; Dixon & Moore, 1996; Gelman & Meck, 1983).

Zwei interessante methodische Innovationen wurden in neueren Arbeiten eingeführt: die Suche nach Personenclustern mit ähnlichen Wissensprofilen (Canobi, 2004; Canobi et al., 1998; Canobi et al., 2002; Canobi et al., 2003; Gilmore & Bryant, in Vorbereitung) und die Untersuchung der Faktorstruktur konzeptuellen Wissens durch Hecht (1998). Die Resultate dieser Methoden liefern bisher noch kein einheitliches Bild. So gelangten unterschiedliche Arbeiten zu unterschiedlichen Clusterlösungen und das Ergebnis der faktoranalytischen Studie – zwei voneinander relativ unabhängige konzeptuelle

Wissensfaktoren – passt schlecht zur Clusterlösung von Gilmore und Bryant, in denen die Probanden homogenes konzeptuelles Wissen aufwiesen. Es ist auch suboptimal, dass Hecht das prozedurale Wissen nicht in seine Faktoranalysen mit einbezog. Die Anwendung von Cluster- und Faktorenanalysen ist jedoch deswegen bemerkenswert, weil sich darin die inhaltliche Idee widerspiegelt, dass sich das Antwortverhalten von Probanden bei Einzelaufgaben stets auch durch inhaltspezifisches Wissen erklären lässt und grundsätzlich nicht die Postulierung unterschiedlicher Wissensarten erfordert. Die Wissensarten bieten nur dann einen explanatorischen Nutzen, wenn sie Verhaltensmuster von Probanden über mehrere verschiedene Aufgaben hinweg erklären.

Ob stabile Verhaltensmuster, so sie denn gefunden würden, wirklich auf konzeptuelles und prozedurales Wissen und nicht auf ganz andere Wissensarten hindeuten, kann freilich auch im Falle ihres Auftretens nur durch eine sorgfältige theoriegeleitete Aufgabenauswahl ermittelt werden. Gerade in diesem Punkt wird aber in den neueren Arbeiten eine große Verwirrung sichtbar: Es gibt kaum zwei Studien, die konzeptuelles und prozedurales Wissen auf die gleiche Weise operationalisieren. Das eindringlichste Beispiel hierfür liefern Blöte, Burg und Klein (2001) einerseits und Hecht (1998) andererseits. Erstere ließen Kinder die Adäquatheit vorgegebener Prozeduren evaluieren, um ihr konzeptuelles Wissen zu erheben, letzterer tat das Gleiche, um ihr prozedurales Wissen zu evaluieren. Beide Arbeiten wurden auch noch in derselben Fachzeitschrift, dem *Journal of Educational Psychology*, veröffentlicht.

Diese Verwirrung in Bezug auf die Operationalisierung der Wissensarten mag unter anderem erklären, warum einige Autoren (wie Canobi, 2004) Korrelationen zwischen den Wissensarten finden, andere (z.B. Resnick & Omanson, 1987) wiederum nicht, und warum manche (wie Gelman & Meck, 1983) finden, dass sich das konzeptuelle Wissen vor dem prozeduralen Wissen des Zählens entwickelt, während andere (vgl. Rittle-Johnson & Siegler, 1998) die umgekehrte Erwerbsreihenfolge beobachten.

Den reflektiertesten Umgang mit diesen Problemen und die darum wohl auch verlässlichsten empirischen Befunde stammen aus der aktuellen Forschung von Rittle-Johnson und Kollegen zum Iterativen Modell, die im nächsten Kapitel vorgestellt wird.

2.5 Zusammenfassung

Ein Überblick über die zentralen empirischen Studien zum Erwerb mathematischer Konzepte und Prozeduren ergab folgendes Bild: Die Befunde von Rochel Gelman und vielen nachfolgenden Autoren zeigen, dass das konzeptuelle Verständnis von Kindern und ihre Problemlösefähigkeiten nicht immer übereinstimmen, sondern meist nur mittelstark miteinander korreliert sind. In unterschiedlichen mathematischen Domänen wurden

unterschiedliche Erwerbsreihenfolgen beobachtet. Die optimale Erwerbsreihenfolge wurde bisher nicht ermittelt. In Trainingsstudien gelang es teilweise, die Verknüpfung der Wissensarten miteinander durch Treatments, insbesondere konstruktivistischen und traditionellen Unterricht, differenziell zu beeinflussen. Aufgrund des Fehlens von Experimenten sind die kausalen Beziehungen zwischen den Wissensarten bisher unbekannt. Das größte Hindernis bei der Untersuchung der Wissensarten ist die Unklarheit darüber, wie die Wissensarten valide und unabhängig voneinander gemessen werden können.

3 Das Iterative Modell

3.1 Modellannahmen

Rittle-Johnson und Alibali (1999) sowie Rittle-Johnson, Siegler und Alibali (2001) entwerfen ein Modell über die Beziehungen zwischen konzeptuellem und prozeduralem Wissen, wobei sie sich auf den oben zusammengefassten Literaturüberblick von Rittle-Johnson und Siegler (1998) beziehen. Das Modell basiert auf Annahmen zu vier verschiedenen Punkten: (1) den Relationen zwischen konzeptuellem und prozeduralem Wissen pro Messzeitpunkt, (2) den kausalen Relationen zwischen den Wissensarten über die Zeit hinweg, (3) den charakteristischen Eigenschaften der Wissensarten und (4) den Operationalisierungen der Wissensarten.

Zu Punkt 1, den Relationen zwischen den Wissensarten an einem Messzeitpunkt, nehmen sie in Übereinstimmung mit der bestehenden Literatur an, dass beide Wissensarten miteinander zusammenhängen, aber nicht identisch sind. Sie schreiben: „These two kinds of knowledge lie on a continuum and cannot always be separated; however, the two ends of the continuum represent two different types of knowledge“ (Rittle-Johnson & Alibali, 1999, 175). Die Aussage ist allerdings nur dann sinnvoll, wenn man den Autoren unterstellt, dass sie ein semantisches Kontinuum meinen. Quantitativ gesehen ist die Aussage falsch, weil man ja nicht mit logischer Notwendigkeit prozedurales Wissen verliert, wenn man konzeptuelles gewinnt, oder umgekehrt.

Bezüglich der kausalen Relationen (Punkt 2) gehen die Autoren im Gegensatz zu den Autoren aller früheren Arbeiten davon aus, dass sich nicht nur eine Wissensart positiv auf den Erwerb der anderen auswirkt, sondern dass die Wissensarten sich wechselseitig beeinflussen:

The majority of past research and theory on these relations has focussed on whether conceptual or procedural knowledge emerges first (Rittle-Johnson & Siegler, 1998). The developmental precedence of one type of knowledge over the other has been hotly debated (e.g., Gelman & Williams, 1998; Siegler, 1991; Siegler & Crowley, 1994; Sophian, 1997). In contrast to this past research and theory, we propose that throughout development, conceptual and procedural knowledge influence one another. Specifically, we propose that conceptual and procedural knowledge develop iteratively, with increases in one type of knowledge leading to increases in the other type of knowledge, which trigger new increases in the first. (Rittle-Johnson et al., 2001, S. 346)

Um diesen Punkt weiter zu verdeutlichen, bezeichnen Rittle-Johnson und Kollegen ihre Annahmen als *Iteratives Modell* und grenzen dieses gegen zwei andere Ansätze ab: *concepts-first-Theorien* und *procedures-first-Theorien*.

Erstere postulieren, dass Kinder gewöhnlich erst konzeptuelles Wissen erwerben, zum Beispiel durch Erklärungen von Eltern und Lehrern oder durch Bücher, und dann ihr prozedurales Wissen durch Anwendung und Übung daraus ableiten. Procedures-first-Theorien hingegen gehen vom umgekehrten Erwerbsweg aus: Prozedurales Wissen wird zuerst erworben, zum Beispiel durch trial-and-error-Lernen oder durch Imitation erfolgreicher Personen in der Umgebung. Das konzeptuelle Wissen wird dann, zum Beispiel durch Prozesse der *representational redescription*, langsam daraus abstrahiert. Für vertiefte Informationen über beide Ansätze siehe Unterkapitel 5.1.

Der Artikel von Rittle-Johnson et al. (2001) lässt deutlich erkennen, dass es den Autoren dabei nicht um die Kategorisierung einzelner Theorien oder Wissenschaftler geht. Sie deuten nur kurz und oberflächlich an, welche Publikationen welcher der beiden Kategorien zugerechnet werden können, nämlich Geary (1994), Gelman und Williams (1998) sowie Halford (1993) den concepts-first-Ansätzen und Fuson (1988), Karmiloff-Smith (1992) sowie Siegler und Stern (1998) den procedures-first-Ansätzen. Vermutlich würde sich kaum ein Autor tatsächlich explizit einer der beiden Positionen zuordnen, indem er oder sie ausschließt, dass bidirektionale kausale Beziehungen bestehen könnten. Trotzdem haben alle Ansätze außer dem Iterativen Modell stets auf einen der beiden Erwerbswege fokussiert, so dass es durchaus sinnvoll sein kann, veranschaulichend von concepts-first- und procedures-first-Ansätzen zu sprechen.

Den dritten und den vierten Punkt der Modellannahmen des Iterativen Modells behandeln Rittle-Johnson und Siegler integriert, indem sie in ihrer Definition der Wissensarten typische Operationalisierungen miterwähnen. Sie schreiben:

We define procedural knowledge as the ability to execute action sequences to solve problems. This type of knowledge is tied to specific problem types and therefore is not widely generalizable. To assess procedural knowledge researchers typically use routine tasks, such as counting a row of objects or solving standard arithmetic computations, because children are likely to use previously learned step-by-step solution methods to solve the problems (e.g., Briars & Siegler, 1984; Hiebert & Wearne, 1996). In contrast to procedural knowledge, we define conceptual knowledge as implicit or explicit understanding of the principles that govern a domain and of the interrelations between units of knowledge in a domain. This knowledge is flexible and not tied to specific problem types and is therefore generalizable. Furthermore, it may or may not be verbalizable. To assess conceptual knowledge, researchers use novel tasks, such as counting in nonstandard ways or evaluating unfamiliar procedures. Because children do not already know a procedure for solving the task, they must rely on their knowledge of relevant concepts to generate methods for solving the problems (e.g., Bisanz & LeFevre, 1992; Briars & Siegler, 1984; Gelman & Meck, 1983; Greeno et al., 1984; Hiebert & Wearne, 1996; Siegler & Crowley, 1994). (Rittle-Johnson et al., 2001, S. 346)

Dass Rittle-Johnson und Kollegen die angemessene Operationalisierung der Wissensarten mit in ihre Definition aufnehmen, ist kein Zufall. Sie attestieren früheren Studien in dieser Hinsicht methodische Unausgereiftheit:

As this review has shown, many studies provide suggestive evidence about the relations between conceptual and procedural knowledge. However, few studies have directly examined these relations. For several reasons, the methodologies commonly used in this area do not allow strong inferences to be drawn about how conceptual and procedural knowledge are related. First, in many studies conceptual and procedural knowledge are not assessed independently. Consequently it is impossible to ascertain how the two types of knowledge are related. Second, when independent assessments of conceptual knowledge are used, they are typically coarse-grained, categorical measures (i.e., children are classified as either having or not having the relevant conceptual understanding). Few studies have used fine-grained measures that can detect gradual changes in understanding. Third, in order to assess changes in knowledge, repeated assessments of children's knowledge are essential. Few studies in this area have assessed children's knowledge at more than one point in time. Finally, intervention studies are rare, and most do not include adequate controls that allow causal conclusions to be drawn. (Rittle-Johnson & Alibali, 1999, S. 177)

Neben diesen häufig vorkommenden, jedoch leicht vermeidbaren methodischen Schwächen beschreiben Rittle-Johnson et al. (2001) noch ein prinzipielles Problem: Zur Lösung jeder beliebigen Aufgabe ist prozedurales Wissen, also die Fähigkeit, zielgerichtet handeln zu können, erforderlich. Eine Testaufgabe, die konzeptuelles Wissen vollkommen unabhängig von prozeduralem erfasst, kann es also nicht geben.

Diese Überlegung bietet eine Alternativerklärung dafür, warum in der Literatur relativ häufig positive Korrelationen zwischen Maßen konzeptuellen und prozeduralen Wissens berichtet werden: Die Maße könnten teilweise dasselbe Konstrukt, nämlich prozedurales Wissen, erfasst haben und daher miteinander kovariieren – unabhängig davon, ob sich konzeptuelles und prozedurales Wissen innerpsychisch tatsächlich beeinflussen oder nicht. In ähnlicher Weise stellen die Überlegungen von Rittle-Johnson und Kollegen auch Befunde zur Erwerbsreihenfolge und zu prädiktiven Relationen zwischen den Wissensarten in Entwicklungsverläufen in Frage. Wie Rittle-Johnson und Kollegen mit dem Operationalisierungsproblem im Rahmen ihrer eigenen empirischen Studien umgegangen sind und zu welchen Befunden dies führte, wird im folgenden Abschnitt besprochen.

3.2 Die Äquivalenzproblemstudie

Um die Grundannahmen des Iterativen Modells zu überprüfen, führten Rittle-Johnson und Alibali (1999) eine Trainingsstudie mit Prätest-Posttest-Design durch, bei der sie das konzeptuelle und das prozedurale Wissen über Äquivalenzprobleme von 48 Viert- und

Fünftklässlern jeweils vor und nach verschiedenen Interventionen erhoben. Die Kinder bearbeiteten Prätest und Treatment an einem Tag und den Posttest am darauf folgenden.

Das konzeptuelle Wissen wurde durch zwei Maße erhoben: (a) Die Kinder sollten in Interviews sieben unterschiedliche Aufgaben lösen, wie „Definiere, was es bedeutet, wenn man sagt, dass zwei Mengen von Objekten gleich groß sind.“ und „Beurteile sechs Definitionen des Gleichheitszeichens als ‚sehr gut‘, ‚gut‘ oder ‚nicht so gut‘“ (frei übersetzt nach Rittle-Johnson & Alibali, 1999, S. 178). Diese Fragen bezogen sich auf insgesamt vier verschiedene Konzepte, die bei der Lösung von Äquivalenzaufgaben wichtig sind: die Bedeutung von Mengengleichheiten, die Bedeutung des Gleichheitszeichens, die Gleichungsstruktur der Äquivalenzaufgaben und den Zusammenhang zwischen der Bedeutung des Gleichheitszeichens und der Gleichungsstruktur. (b) Die zweite Aufgabenart, die konzeptuelles Wissen messen sollte, bestand darin, dass eine Versuchsleiterin den Kindern verbal und durch Gesten sechs verschiedene Prozeduren zur Lösung von Äquivalenzproblemen vorführte. Die Kinder sollten diese Prozeduren dann als „sehr gut“, „gut“ oder „nicht so gut“ (frei übersetzt nach Rittle-Johnson & Alibali, 1999, S. 178) bewerten.

Das prozedurale Wissen der Kinder erhoben Rittle-Johnson und Alibali, indem sie die Kinder (a) Standardäquivalenzprobleme der Form $a + b + c = a + \underline{\quad}$ bearbeiten ließen und (b) Transferprobleme, bei denen die Rechenoperation (Addition vs. Multiplikation), die Position des zu bestimmenden Wertes in der Gleichung und die Größen der gegebenen Summanden (äquivalente Summanden auf beiden Seiten vs. alle Summanden unterschiedlich) variiert wurden.

Es gab drei Treatmentgruppen. Die Kinder in der konzeptuellen Gruppe erhielten eine zwei Sätze lange Instruktion über das Prinzip, das beim Lösen von Äquivalenzaufgaben zu beachten ist (die Summen auf beiden Seiten des Gleichheitszeichens müssen gleich groß sein), ohne dass darauf eingegangen wurde, mittels welcher Prozeduren man dieses allgemeine Lösungsprinzip befolgen kann. In der prozeduralen Gruppe erhielten die Kinder eine zwei Sätze lange Instruktion darüber, mittels welcher Prozedur die gesuchte Zahl gefunden werden kann (nämlich durch Addieren der Summanden b und c). Die Kinder in beiden Treatmentgruppen bearbeiteten nach Erhalt der Instruktion ein Standardäquivalenzproblem ohne Feedback, erhielten ihre jeweilige Instruktion dann noch einmal und lösten schließlich ein weiteres Standardäquivalenzproblem ohne Feedback. Die Kinder in der Kontrollgruppe beendeten die Durchführung am ersten Tag nach dem Ausfüllen des Prätests und erhielten somit zwischen Prä- und Posttest keinerlei Instruktionen oder Übungen.

Die Prätestwerte zeigen in Übereinstimmung mit älteren Studien positive Zusammenhänge zwischen konzeptuellem und prozeduralem Wissen. Die Korrelation zwischen den beiden Prätestmaßen konzeptuellen Wissens beträgt jedoch nur $r = ,37$ und stellt die so genannte konvergente Validität der Maße, die ja eigentlich beide dasselbe Konstrukt erheben sollten, in Frage. Die konvergente Validität der beiden Maße prozeduralen Wissens wird von den Autoren nicht berichtet.

Auf beiden konzeptuellen Wissensskalen sind die Mittelwerte der konzeptuellen und der prozeduralen Gruppe höher als der Mittelwert der Kontrollgruppe. Jedoch zeigen sich keine Unterschiede zwischen der konzeptuellen und der prozeduralen Gruppe. Auch auf der prozeduralen Wissensskala Standardäquivalenzprobleme unterscheiden sich die konzeptuelle und die prozedurale Gruppe nicht in der Anzahl gelöster Probleme. Die konzeptuelle Gruppe nutzte jedoch mehrere unterschiedliche Prozeduren, während die prozedurale hauptsächlich die ihnen beigebrachte anwandte. Auf der prozeduralen Wissensskala Transferprobleme zeigt sich das erstaunliche Bild, dass die konzeptuelle Gruppe am besten abschnitt, gefolgt von der Kontrollgruppe und, noch dahinter, der prozeduralen Gruppe.

Darin, dass die Kinder in der konzeptuellen Gruppe auch prozedurales Wissen erwarben und die Kinder in der prozeduralen Gruppe auch konzeptuelles Wissen, sehen Rittle-Johnson und Alibali eine Bestätigung der Grundannahme des Iterativen Modells: Die Wissensarten entwickeln sich nicht unabhängig voneinander, sondern Zuwächse der einen Wissensart führen auch zu Zuwächsen der jeweils anderen. Dass die Kinder in der konzeptuellen Gruppe auf der prozeduralen Wissensskala Transferprobleme besser abschnitten als die prozedurale Gruppe, belegt den Autoren der Studie zufolge, dass das konzeptuelle Wissen sich stärker auf das prozedurale auswirkt als umgekehrt.

Eine wichtige Frage diskutieren die Autoren jedoch nicht: Wenn verschiedene Maße konzeptuellen Wissens niedrig miteinander korrelieren, Maße konzeptuellen und prozeduralen Wissens positiv miteinander korrelieren, wenn eine konzeptuell ausgerichtete Intervention nicht mehr konzeptuelles als prozedurales Wissen vermittelt und eine prozedural ausgerichtete Intervention nicht mehr prozedurales Wissen als konzeptuelles, welchen explanatorischen Nutzen hat dann überhaupt die Annahme zweier Wissensarten?

Dass das Treatment sich nicht in der zu erwartenden Weise auf die Maße konzeptuellen und prozeduralen Wissens ausgewirkt hat, kann mindestens vier verschiedene alternative Ursachen haben: (1) Es gibt innerpsychisch nur eine Art von Wissen oder mehrere Arten, die jedoch andere Charakteristika als konzeptuelles und prozedurales Wissen aufweisen und in der Untersuchung nicht erfasst wurden. (2) Es gibt

konzeptuelles und prozedurales Wissen. Sie können aber so schnell und effizient ineinander umgewandelt werden, dass experimentell induzierte Dissoziationen zwischen ihnen empirisch nicht nachgewiesen werden konnten. (3) Die Wissensarten existieren, lassen sich grundsätzlich differenziell beeinflussen, was im Experiment aufgrund einer ungeeigneten Treatmentsauswahl jedoch nicht geschah. (4) Die Wissensarten existieren, wurden durch das Treatment in die erwartungsgemäßen Richtungen beeinflusst, was jedoch aufgrund ungeeigneter Operationalisierungen nicht nachgewiesen werden konnte. Gründe 2 bis 4 könnten auch gemeinsam auftreten, Grund 1 hingegen nur alleine. Welche der Gründe zutreffen, lässt sich aus den in der Studie gewonnenen Daten nicht ableiten. Das deutet keineswegs auf methodische Defizite der Studie hin. Die Studie stellt vielmehr das einzige randomisierte Experiment mit Kontrollgruppe dar, das in der Literatur zu konzeptuellem und prozeduralem Wissen berichtet wird. Zudem werden Summenscores statt dichotomer Items zur Operationalisierung der Wissensarten benutzt und es werden zwei unterschiedliche Indikatoren für jede Wissensart herangezogen. Dass diese methodisch also sehr sorgfältig gestaltete Studie zu so schwer interpretierbaren Ergebnissen kommt, wirft die Frage auf, wie produktiv die Unterscheidung von konzeptuellem und prozeduralem Wissen als zwei verschiedenen Konstrukten wirklich ist.

Bei allen Einschränkungen hat die Studie jedoch auch einen wirklich neuen Befund geliefert: Die Kinder mit der konzeptuellen Intervention zeigten im Posttest einen besseren Transfer von Prozeduren auf neue Probleme als die Kinder in der prozeduralen Gruppe. Die Idee, dass konzeptuelles und prozedurales Wissen unterschiedlich transferfähig sind, wird schon in der oben wiedergegebenen Definition der Wissensarten erwähnt (siehe Seite 36f). Sie bildet auch die Grundidee hinter der Operationalisierung der Wissensarten in der zweiten und dritten von Rittle-Johnson und Kollegen zur Überprüfung des Iterativen Modells durchgeführten Studie.

3.3 Die Dezimalbruchstudie

Rittle-Johnson, Siegler und Alibali (2001) maßen zunächst das konzeptuelle und das prozedurale Wissen von 74 Fünft- und Sechstklässlern über Dezimalbrüche vor und nach vier zweiphasigen Interventionen, die sich in der ersten, nicht aber in der zweiten Phase unterschieden. Sie analysierten dann, wie die interindividuellen Vorwissensunterschiede der Kinder in Interaktion mit der Intervention die Zuwächse an konzeptuellem und prozeduralem Wissen determinieren.

Für die Intervention wurden die Kinder nach dem Prätest randomisiert einer von vier Treatmentgruppen zugewiesen. Die Mitglieder dieser Gruppen erhielten während der ersten Interventionsphase in Einzelsitzungen analog zu dem Treatment aus Rittle-Johnson

und Alibali (1999) (1) eine kurze konzeptuelle Instruktion, (2) eine kurze prozedurale Instruktion, (3) eine Instruktion, die beide Wissensarten vermitteln sollte, oder (4) keine Instruktion.

Anschließend spielte in der zweiten Interventionsphase jedes Kind das „Fang-das-Monster“-Spiel: In jeder der 12 Aufgaben des Spiels wurde am Computer ein Dezimalbruch zwischen null und eins präsentiert zusammen mit einem Zahlenstrahl, der ebenfalls von null bis eins reichte und bei dem lediglich die Endpunkte beschriftet waren. Dem Kind wurde gesagt, dass sich an der Position auf dem Zahlenstrahl, die der Dezimalbruch angibt, ein Monster versteckt und dass sie es fangen sollten, indem sie eine von vier vorgegebenen Positionen mit der Maus anklickten. Nach Eingabe dieser Antwort erschien als Feedback das Monster an der korrekten Position. Danach sollte das Kind mündlich erklären, warum die Dezimalzahl auf dem Zahlenstrahl an der durch das Monster indizierten Stelle liegt und nicht an einer anderen, und die nächste Aufgabe erschien auf dem Bildschirm. Die gesamte Intervention wurde auf Video aufgenommen. Direkt anschließend bearbeitete jedes Kind in einer Einzelsitzung den Posttest.

Die Prä- und Posttestmaße konzeptuellen und prozeduralen Wissens leiteten Rittle-Johnson und Kollegen aus ihren auf Seite 36 zitierten Definitionen der Wissensarten ab. Diese beschreiben konzeptuelles Wissen als flexibel auf neue Aufgabentypen übertragbar und prozedurales Wissen als an Routineprobleme gebunden. Da die Kinder in der Intervention lernen, Dezimalbrüche auf dem Zahlenstrahl zu positionieren, betrachten Rittle-Johnson und Kollegen die Aufgaben dieses Typs als Routineprobleme. Die mittlere Lösungsrate dieser Aufgaben vor, während und nach dem Training benutzten sie, um prozedurales Wissen zu erheben. Konzeptuelles Wissen erhoben sie über Transferprobleme, die zwar Wissen über Dezimalbrüche, nicht jedoch die tatsächliche Lokation eines Dezimalbruchs auf einem Zahlenstrahl erfordern. Dafür wählten sie fünf verschiedene Problemtypen: (a) den größeren von zwei gegebenen Dezimalbrüchen auswählen, (b) aus Antwortvorgaben einen Dezimalbruch wählen, der größer oder kleiner als ein gegebener Referenzwert ist, (c) einen Dezimalbruch aufschreiben, der zwischen zwei vorgegebenen liegt, (d) die Äquivalenz einiger Dezimalbrüche erkennen und (e) vorgegebene Lösungen für Additionen zweier Dezimalbrüche auf ihre Richtigkeit hin evaluieren. Von jedem Typ lösten die Kinder mehrere Aufgaben und der Summenscore der korrekten Lösungen wurde über die Aufgabentypen hinweg berechnet.

Die Korrelation zwischen dem konzeptuellen und dem prozeduralen Prätestscore beträgt $r = ,33$. Der Pre- und der Posttestscore des konzeptuellen Wissensmaßes korrelieren

mit $r = ,77$, die Scores des prozeduralen Wissensmaßes sind mit $r = ,31$ etwas weniger stabil.

Das konzeptuelle Wissen der Studienteilnehmer stieg vom Prä- zum Posttest von 33 auf 45 Prozent. Das prozedurale Wissen stieg von 28 Prozent im Prätest auf 60 Prozent in der Intervention und auf 68 Prozent im Posttest. Wie schon in der Studie von Rittle-Johnson und Alibali (1999), gab es keine signifikanten Unterschiede zwischen den vier Treatmentgruppen – weder in den Trainings- noch in den Posttestleistungen.

In Übereinstimmung mit dem Iterativen Modell und den Hypothesen der Autoren sagte der konzeptuelle Prätestwert die Zuwächse an prozeduralem Wissen während der Intervention voraus. Nachdem die Lösungskorrektheit bei den prozeduralen Trainings- und Posttestaufgaben um den prozeduralen Prätestwert bereinigt wurde, erklärte der konzeptuelle Prätestscore 23 Prozent der Trainingsleistung und 20 Prozent der Posttestleistung bei den prozeduralen Aufgaben. Nachdem die prozeduralen Leistungen der Kinder während der Intervention um den prozeduralen Prätestwissensscore bereinigt wurden, erklärten sie 22 Prozent der Varianz in den Prätest-Posttest-Zuwächsen des konzeptuellen Wissens.

Man beachte, dass bei der Vorhersage der prozeduralen Wissenszuwächse durch das konzeptuelle Wissen im Prätest die unabhängige Variable zeitlich früher gemessen wurde als die abhängige Variable. Bei der Vorhersage der konzeptuellen Wissenszuwächse während der Intervention durch die prozeduralen Wissenszuwächse während der Intervention besteht hingegen kein klarer zeitlicher Abstand zwischen den Messungen der unabhängigen und der abhängigen Variable.

Ob der prozedurale Prätestwissensscore die Zuwächse an konzeptuellem Wissen vorhersagt, wird von Rittle-Johnson und Kollegen nicht berichtet.

Einen zusätzlichen Befund lieferte die Auswertung der mündlichen Erklärungen, die die Kinder während des Monsterspiels gaben. Rittle-Johnson und Kollegen betrachteten die Korrektheit dieser Erklärungen als Indiz dafür, dass ein Kind versteht, wie man Dezimalbrüche mit Hilfe externer Repräsentationsformen darstellen kann. Weil externe Repräsentationsformen konzeptuelles Wissen veranschaulichen und dadurch dem konkreten Handeln zugänglich machen können, erwarteten Rittle-Johnson und Kollegen, dass die Fähigkeit ein Problem grafisch-visuell repräsentieren zu können als Mediator die Interaktion zwischen den Wissensarten beeinflusst. Genau dies zeigen auch ihre Daten: Das konzeptuelle Prätestwissen der Kinder sagt voraus, wie gut sie während der Intervention die Benutzung des Zahlenstrahls erklären können. Das Ausmaß dieser Kompetenz sagt wiederum ihre Zuwächse an prozeduralem Wissen voraus. Außerdem

sinkt der Zusammenhang zwischen konzeptuellem Vorwissen und prozeduralen Wissenszuwächsen substanziell, sobald die Fähigkeit die Zahlenstrahlbenutzung erklären zu können ausparialisiert wird.

Um diesen korrelativen Befund über eine mögliche Mediatorwirkung experimentell zu untermauern führten Rittle-Johnson et al. (2001) eine zweite Studie durch, die der ersten stark glich. Jedoch wurde auf die ursprünglichen vier Treatmentgruppen verzichtet. Stattdessen wurden zwei Faktoren variiert, die beide in unterschiedlicher Weise die korrekte externe Problemrepräsentation unterstützen sollten. Zum einen wurde variiert, ob die Probanden vor jeder Aufgabe einen verbalen Hinweis erhielten, der sie auf die Wichtigkeit der ersten Ziffer hinter dem Komma hinwies. Diese Ziffer wurde dann außerdem auf dem Bildschirm rot dargestellt. Zum anderen wurde variiert, ob auf dem Zahlenstrahl Ticks, also markierende Unterteilungen, eingezeichnet waren, die die Positionen der ganzen Zehntel anzeigten. Die Versuchspersonen wurden randomisiert einer der vier resultierenden Bedingungen zugeteilt.

Erwartungsgemäß zeigten sich erneut positive korrelative Beziehungen zwischen den Zuwächsen beider Wissensarten während des Trainings. Dass das Treatment diese Beziehungen als Mediator beeinflusst, wird jedoch nicht durch die Daten belegt. Einen signifikanten Einfluss übte das Treatment lediglich auf die Zuwächse an prozeduralem Wissen aus. Die Autoren schließen daher mit dem Fazit, dass die Wissensarten sich unzweifelhaft gegenseitig beeinflussen, während die Mechanismen, durch die das geschieht, vorläufig noch unklar sind.

Indem die Autoren aus den Daten beider Dezimalbruchstudien auf bidirektionale kausale Beziehungen zwischen den Wissensarten schließen, begehen sie jedoch einen Fehler. Denn experimentell abgesicherte Befunde können sie nicht vorlegen, weil die in Studie 1 eingesetzten Treatments zur separaten Vermittlung der Wissensarten nicht funktionierten. Ihre längsschnittlichen Befunde zeigen, dass konzeptuelles Wissen prädiktiv für spätere prozedurale Wissenszuwächse ist, was als Indiz für einen Kausaleinfluss konzeptuellen Wissens auf prozedurales betrachtet werden kann. Die umgekehrte Kausalrichtung vom prozeduralen Vorwissen auf spätere konzeptuelle Wissenszuwächse wurde von den Autoren jedoch nicht analysiert. Sie zeigen lediglich, dass die Zuwächse beider Wissensarten korreliert sind. Ihr Auswertungsdesign ist daher zur Überprüfung des innovativen Kerns ihres Modells, nämlich der Annahme bidirektionaler Kausalrelationen, nicht geeignet.

3.4 Diskussion des Modells

Vergleicht man die Arbeiten Rittle-Johnsons mit den älteren Untersuchungen, so fallen mindestens vier Punkte positiv auf: (1) Während viele ältere Studien sich zwar allgemein auf Theorien über Wissensarten berufen, leiten sie ihre Hypothesen jedoch fast nie aus einem konkreten theoretischen Modell ab, sondern gehen explorativ vor. Das Iterative Modell erlaubt hingegen die Ableitung falsifizierbarer Hypothesen. Es ist daher besonders gut dazu geeignet, empirische Studien theoretisch zu leiten. (2) In älteren Arbeiten werden die Wissensarten meist nur ungenau charakterisiert. Häufig werden Befunde über unterschiedliche Konstrukte, wie deklaratives und prozedurales Wissen, explizites und implizites Wissen oder semantisches und syntaktisches Wissen ohne jegliche Diskussion direkt auch konzeptuellem und prozeduralem Wissen zugeschrieben. In den Arbeiten zum Iterativen Modell werden konzeptuelles und prozedurales Wissen hingegen explizit definiert. Dies erlaubt es, Gemeinsamkeiten und Unterschiede zu anderen Begriffspaaren theoretisch und empirisch zu analysieren (vgl. Abschnitt 4.1). (3) Die relative Exaktheit des Iterativen Modells erlaubt, wie oben dargestellt, eine theoriegeleitete Auswahl der Operationalisierungen. Das Modell impliziert, dass konzeptuelles Wissen am besten durch Transferaufgaben und prozedurales Wissen am besten durch Routineaufgaben gemessen werden kann. Von dieser Idee könnte die Forschung zu den Wissensarten stark profitieren, weil es bisher keine einheitlichen Standards bezüglich der Operationalisierungen gab. (4) Die meisten Theorien auf dem Gebiet konzentrieren sich entweder auf die Ableitung von konzeptuellem Verständnis aus Handlungswissen oder auf die Ableitung von Handlungswissen aus konzeptuellem Verständnis und blenden die jeweils umgekehrte Erwerbsrichtung aus (siehe Kapitel 5). Rittle-Johnson und Kollegen weisen als erste darauf hin, dass bidirektionale Einflüsse zwischen den Wissensarten wahrscheinlich sind und gemeinsam untersucht und erklärt werden sollten. Ihre Arbeit beinhaltet somit auch eine stark integrative Komponente in dem Sinne, dass sie die Befunde mehrerer einander ursprünglich widersprechender theoretischer Positionen miteinander vereinbar macht.

Da das Iterative Modell auf einer sorgfältigen Analyse der bestehenden Studien basiert und in den genannten vier Punkten über diese Studien hinausgeht, wurde es als theoretischer Ausgangspunkt der vorliegenden Dissertation gewählt. Da es noch relativ neu und wenig erforscht ist, besteht zu einigen Modellannahmen Klärungsbedarf. Diese sollen hier angedeutet und im Rahmen der Vorstellung der Forschungsfragen der Arbeit (Kapitel 8) vertieft diskutiert werden.

Die Hauptschwäche des Iterativen Modells dürfte in den Operationalisierungen der Wissensarten liegen. Sie wurden aus der Grundannahme abgeleitet, dass konzeptuelles

Wissen flexibel auf neue Inhaltsbereiche transferiert werden kann, während prozedurales Wissen an Routineprobleme gebunden ist. Diese Annahme ist jedoch keineswegs unstrittig (Müller, 1999; Willingham, 1998) und es wäre gut, wenn man sie empirisch überprüfen könnte, bevor man sie jeder empirischen Forschung zu konzeptuellem und prozeduralem Wissen zugrunde legt. Um überprüfen zu können, ob tatsächlich konzeptuelles Wissen flexibel transferierbar und prozedurales Wissen an Routineprobleme gebunden ist, müsste man beide Wissensarten unabhängig voneinander messen können. Man muss also Eigenschaften der Wissensarten kennen, um die Wissensarten operationalisieren zu können, und man muss Operationalisierungen der Wissensarten kennen, um Eigenschaften der Wissensarten empirisch nachzuweisen. Aus diesem Teufelskreis lässt sich nur schwer ausbrechen. Es wird dennoch in der vorliegenden Arbeit versucht.

Ein weiteres Problem der vom Iterativen Modell implizierten Operationalisierungen besteht darin, dass es vorteilhaft wäre, wenn die Wissensarten in Prätest-Posttest-Designs gemessen werden könnten. Das ist jedoch nicht der Fall. Die Unterscheidung von Routine- und Transferaufgaben funktioniert immer nur *nach* einem Training, in dem bestimmte Aufgaben durch Übung zu Routineaufgaben wurden. Dasselbe Testitem erhebt nach Rittle-Johnson vor einem Treatment ein anderes Konstrukt als danach, weil es einmal ein Routineproblem darstellt und einmal nicht. Tatsächlich fanden Rittle-Johnson et al. (2001) empirische Hinweise auf das Vorliegen dieses Problems in ihrem Prätest-Posttest-Design. Sie stellten fest, dass die von den Kindern im Prätest verwendeten Prozeduren in keiner Weise mit den später von den Kindern verwendeten Konzepten und Prozeduren zusammenhängen und schreiben:

Because of their lack of previous experience with the number line problems, the procedural knowledge pretest likely did not tap prior knowledge of specific procedures for performing the task but rather a combination of procedures imported from other domains, conceptual knowledge, and guessing. [...] The procedural knowledge pretest was not a significant predictor of success on any assessment. The number line pretest may not have tapped procedural knowledge, because children did not have prior experience with the problems. Instead, scores on the procedural knowledge pretest were related to gains in procedural knowledge only to the extent that both were related to initial conceptual knowledge. (Rittle-Johnson et al., 2001, S. 352f)

Dieser Punkt, dass es offensichtlich nicht gelungen ist, die Wissensarten wirklich unabhängig voneinander zu erheben, schränkt die Validität der vorliegenden Befunde stark ein.

Solche Schwierigkeiten wären in experimentellen Designs noch größer, sofern die Gruppen in der Interventionsphase unterschiedliche Aufgaben bearbeiteten. Was dann im

Posttest für die eine Gruppe ein Routineproblem wäre, könnte für die andere ein bisher unbekanntes Transferproblem sein und umgekehrt. Dasselbe Testitem würde dann zum selben Zeitpunkt bei unterschiedlichen Gruppen unterschiedliche Wissensarten operationalisieren.

Wegen dieser noch ungelösten Probleme bei der Messung der Wissensarten ist unklar, inwieweit die vorhandenen Befunde zur Korrelation der Wissensarten, zu ihrer Erwerbsreihenfolge und zu ihren kausalen Interrelationen valide sind. Insbesondere ist auch unklar, ob wirklich von den bidirektionalen Kausaleinflüssen ausgegangen werden kann, die das Iterative Modell postuliert. Für diese wurde bedingt durch die benutzten Auswertungsdesigns bisher lediglich korrelative, aber weder experimentelle noch längsschnittliche Evidenz gefunden.

In der vorliegenden Arbeit werden konzeptuelles und prozedurales Wissen auf eine neue Weise operationalisiert, nämlich mittels Modellierung der beiden Wissensarten als latente Faktoren in structural equation models (SEM). Auf dieser Grundlage kann einerseits die Qualität des verwendeten Messmodells explizit getestet werden, zum anderen kann das Iterative Modell einer neuen, valideren Überprüfung unterzogen und mit konkurrierenden Ansätzen verglichen werden.

In den folgenden Kapiteln werden einige Fragen explizit besprochen, die in der bisherigen Übersicht über die empirische Forschung zum Erwerb mathematischer Konzepte und Prozeduren schon anklangen: Wie lassen sich die Wissensarten charakterisieren und gegeneinander und gegen andere Konstrukte abgrenzen (Kapitel 4)? Wie können die Wissensarten kognitiven Lerntheorien zufolge erworben werden und wie könnten sie während der Lernprozesse interagieren (Kapitel 5)? Was ist über konkretes konzeptuelles und prozedurales Wissen in spezifischen Inhaltsgebieten bekannt (Kapitel 6)? Wie lassen sich Strukturgleichungsmodelle zur Validierung kognitiver Modelle einsetzen (Kapitel 7)?

Nach der Besprechung dieser Fragen wird in Kapitel 8 erneut das Iterative Modell aufgegriffen und es werden die Forschungsfragen herausgearbeitet, auf denen die empirischen Untersuchungen basieren.

3.5 Zusammenfassung

Zu Illustrationszwecken lassen sich ältere Theorien über konzeptuelles und prozedurales Wissen je nach Erwerbsreihenfolge und -weg, auf die sie fokussieren, in concepts-first- und procedures-first-Ansätze einteilen. Das Iterative Modell von Rittle-Johnson et al. (2001) basiert hingegen auf der Annahme arbiträrer Erwerbsreihenfolgen und bidirektionaler kausaler Beziehungen zwischen den Wissensarten. Rittle-Johnson definiert

konzeptuelles Wissen als transferfähig und prozedurales als an Routineprobleme gebunden. Entsprechend nutzte sie Transferaufgaben und Routineaufgaben, um die Wissensarten zu messen und fand so in einer Studie positive, wenn auch lediglich korrelative Evidenz für das Iterative Modell. Was eine Routineaufgabe ist, kann jedoch über Personengruppen und Zeitpunkte hinweg variieren, so dass die Validität der Studie in Frage steht.

4 Die Unterscheidung von konzeptuellem und prozeduralem Wissen

4.1 Verwandte Unterscheidungen

In diesem Kapitel werden die Unterschiede zwischen konzeptuellem und prozeduralem Wissen besprochen. Im Vergleich zum vorherigen Kapitel wird das Thema dabei stärker aus einer theoretischen und weniger aus einer empirischen Perspektive betrachtet.

Im aktuellen Unterkapitel werden verwandte Unterscheidungen zwischen Wissensarten beschrieben, um das theoretische Umfeld zu verdeutlichen, in dem die Unterscheidung von konzeptuellem und prozeduralem Wissen angesiedelt ist. Im darauf folgenden werden die Wissensarten taxonomisch charakterisiert. Im vierten Unterkapitel werden Gründe gegen und für die Unterscheidung diskutiert. In beiden darauf folgenden werden hypothetische mentale Repräsentationen der Wissensarten und die verschiedenen typischen Operationalisierungen besprochen, bevor im letzten Unterkapitel mögliche Relationen zwischen konzeptuellem und prozeduralem Wissen einerseits und zusätzlichen psychologischen Konstrukten andererseits vorgestellt werden.

Tabelle 1: Auswahl einiger von Lerntheoretikern bisher vorgeschlagener Wissensarten.

Wissensarten	Autorenbeispiel
Faktenwissen & Anwendungswissen	(Oberauer, 1993)
focal & tacit knowledge	(Polanyi, 1962)
explicit & implicit knowledge	(Schacter, 1987)
knowing that & knowing how	(Ryle, 1949)
knowing why & knowing how-to	(Baroody, 2003)
competence & performance	(Chomsky, 1965)
understanding & success	(Piaget, 1978)
structures & procedures	(Inhelder & Piaget, 1980)
mathematical understanding & algorithmic performance	(Nesher, 1986)
principles & skills	(Gelman & Meck, 1983)
semantic & syntactic knowledge	(Resnick, 1982)
conceptual & procedural knowledge	(Hiebert, 1986)
meaningful & mechanic knowledge	(Baroody & Ginsburg, 1986)
schematic & teleologic knowledge	(VanLehn, 1986)
deductive & empirical knowledge	(Schoenfeld, 1986)
declarative & procedural knowledge	(Anderson, 1983)
declarative & nondeclarative knowledge	(Squire, 1992)
automatized & nonautomatized knowledge	(Anderson, 1983)

Schon im vorigen Kapitel wurde deutlich, dass die Unterscheidung zwischen konzeptuellem und prozeduralem Wissen nur eine von vielen Unterscheidungen ist, die in der Literatur zwischen verschiedenen Wissensarten getroffen werden. Tabelle 1 gibt einen Überblick über einige der wichtigsten Paare von Wissensarten, die im Zusammenhang mit konzeptuellem und prozeduralem Wissen erwähnt werden (vgl. auch die Aufstellung von

Haapsalo & Kadjievich, 2000). Ihnen gemeinsam ist, dass stets ein Paarling eher die Grundlage eines abstrakten Tiefenverständnisses zu sein scheint und der andere die Grundlage von Handlungskompetenz.

Weitere, jedoch nicht in dieses Schema passende Wissensarten sind zum Beispiel propositionales und analoges, sprachliches, bildliches, räumliches und numerisches, sowie situiertes und strategisches Wissen. Einen in der Literatur bisher kaum beachteten Versuch einer taxonomischen Einordnung möglichst vieler verschiedener Wissensarten liefern de Jong und Ferguson-Hessler (1996).

Eine exhaustive Charakterisierung der in Tabelle 1 dargebotenen Wissensarten kann hier nicht geleistet werden, unter anderem auch deshalb nicht, weil häufig die Schöpfer der Begriffe selbst auf konkrete Definitionen verzichten und ihre Begriffe eher heuristisch verwenden. Lediglich die wichtigsten Wissensarten sollen hier kurz inhaltlich beschrieben werden. Am bekanntesten von ihnen, unter anderem aus der ACT-Theorie (z.B. Anderson, 1983), ist wohl die Unterscheidung von deklarativem und prozeduralem Wissen. *Deklaratives Wissen* wird verstanden als schemagespeichertes, vernetztes und der bewussten Elaboration zugängliches Wissen über verbale, numerische oder bildhafte Informationen. *Prozedurales Wissen* wird verstanden als zielgerichtete, kompilierte und daher automatisiert anwendbare Produktionsregeln über die Anwendbarkeit mentaler oder physischer Operatoren zur Lösung von Problemen. Diese Unterscheidung ist eng mit einer weiteren verwandt: Unter *explizitem Wissen* wird dem Bewusstsein zugängliches und daher verbalisierbares, unter *implizitem Wissen* dem Bewusstsein unzugängliches und daher nicht verbalisierbares Wissen verstanden. Die Begriffe des *focal* und des *tacit knowledge* sind damit fast synonym. Resniks Unterscheidung zwischen *semantic* und *syntactic knowledge* über Mathematikaufgaben ist interessant, weil sie suggeriert, dass die Aufgaben, auch wenn sie in mathematischer Formelsprache notiert sind, analog zu natürlichen Sprachen verarbeitet werden, die ja ebenfalls eine Semantik und eine Syntax besitzen. Wie in Kapitel 2 geschildert, stammt auch Chomskys Unterscheidung zwischen *Kompetenz* und *Performanz* ursprünglich aus der Psycholinguistik und wurde dann auf allgemeinere lerntheoretische Fragen übertragen. Andersons Unterscheidung von *automatisiertem* und *nicht-automatisiertem* Wissen bezieht sich auf eine Teildimension der Unterscheidung von deklarativem und prozeduralem Wissen: Die Automatisierung ist eine von mehreren Eigenschaften, die deklaratives und prozedurales Wissen unterscheiden, und charakterisiert Handlungswissen, das schnell und ressourceneffizient ausgeführt werden kann, weil es kompiliert ist, also ein stabiles und durch Übung erworbenes Handlungsmuster darstellt (s. auch Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). VanLehns Unterscheidung

zwischen *schematic* und *teleologic knowledge* bezieht sich auf eine andere Subdimension der Unterscheidung zwischen deklarativem und prozeduralem Wissen, nämlich ihre kognitive Repräsentationsform: Es wird angenommen, das deklaratives Wissen durch symmetrische und untereinander vernetzte Schemata mental repräsentiert wird, während prozedurales Wissen in zielgerichteten, also teleologischen, und somit asymmetrischen Produktionsregeln gespeichert wird.

Eine andere Gruppe von Unterscheidungen bezieht sich weniger auf die mentalen Repräsentationen der Wissensarten als vielmehr auf ihren Inhalt, der einerseits Aussagen und andererseits Handlungen zum Gegenstand haben kann. Zu diesen Begriffspaaren gehören *knowing that* und *knowing how*, *knowing why* und *knowing how-to*, und auch *Faktenwissen* und *Anwendungswissen*.

Theorien, die Aussagen über die in Tabelle 1 genannten Wissensarten machen, werden gelegentlich unter den Oberbegriff der *dual component* oder *dual process theories of cognition* subsumiert (z.B. Maddox & Ashby, 2004; Sun, Slusarz, & Terry, 2005), da sie typischerweise zwei Wissensarten oder zwei Informationsverarbeitungsmodi einander gegenüberstellen. Die oben genannten Paare von Wissensarten entstammen den Bereichen der Kognitiven, der Pädagogischen und der Entwicklungspsychologie, der Psycholinguistik sowie der Philosophie. Entsprechend konstatieren Barrett und Mitarbeiter:

Dual process theories of the mind are ubiquitous in psychology.“ (Barrett, Tugade, & Engle, 2004, S. 553) Sie ergänzen: „*There are dual-process theories of attribution (e.g., Trope, 1986; Uleman, Newman, & Moskowitz, 1996), person perception (e.g., Brewer, 1988; Fiske & Neuberg, 1988; Gilbert, 1989; Zárate, Sanders, & Garza, 2000), stereotyping and prejudice (e.g., Devine, 1989), persuasion (e.g., Chaiken, 1980; Petty & Cacioppo, 1986), mental control (e.g., Wegner, 1994; Wenzlaff & Wegner, 2000; self-regulation (Baumeister & Heatherton, 1996; Metcalfe & Mischel, 1999), emotion (Teasdale, 1999; van Reekum & Scherer, 1997), and personality (Epstein, 1998). A central principle of these theories is that behavior is determined by the interplay of automatic and controlled processing.*“ (Barrett et al., 2004, S. 553)

Es stellt sich die Frage, wie sich die verschiedenen Paare von Wissensarten, die in so vielen Gebieten der Psychologie vorgeschlagen wurden, zu der Unterscheidung zwischen konzeptuellem und prozeduralem Wissen verhalten. Können Forschungsergebnisse über verschiedene Paare von Wissensarten hinweg generalisiert werden? Wenn ja, welche?

Leider gibt es bislang weder empirische Forschung noch theoretische Analysen zu diesem Thema. Nichts desto trotz verhalten sich viele Forscher, die sich mit konzeptuellem und prozeduralem Wissen auseinandersetzen, so als wären die meisten Paare von Wissensarten miteinander identisch.

Dies veranschaulicht Tabelle 2. Sie listet für einige wichtige lerntheoretische Artikel, die sich mit Wissensarten auseinandersetzen, auf, (a) welches Begriffspaar die Autoren laut Titel oder Abstrakt ihrer Publikation untersuchen, (b) welche Begriffspaare sie in ihrem Text als Synonyme für dieses Begriffspaar benutzen und (c) welche anderen Begriffspaare sie indirekt, nämlich durch die Zitation entsprechender anderer Arbeiten, mit ihrem Konstruktpaar gleichsetzen. Beispielsweise veröffentlichten Hiebert und Lefevre ein Buchkapitel mit dem Titel *Conceptual and procedural knowledge in mathematics: An introductory analysis*. In der Einleitung zu diesem Kapitel steht der Satz: „The distinction between concepts and procedures plays an important role in more general questions of knowledge acquisition“ (Hiebert & Lefevre, 1986, S. 1). Die Autoren verwenden die Begriffspaare conceptual knowledge und procedural knowledge (im Titel) und concepts und procedures (in der Einleitung) offensichtlich synonym zueinander und demonstrieren damit die plausible Auffassung, dass Erkenntnisse über das eine Paar ohne weiteres auch dem anderen zugeschrieben werden können. Diese direkten Gleichsetzungen sind in der dritten Spalte der Tabelle aufgelistet. An anderer Stelle schreiben dieselben Autoren: „Gelman and Gallistel (1978) describe another way in which procedures can bring conceptual knowledge into the problem solving arena“ (Hiebert & Lefevre, 1986, S. 15). Tatsächlich sprechen Gelman und Gallistel in dem zitierten Buch aber ausschließlich von principles und skills. Hier setzen Hiebert und Lefevre also erneut zwei unterschiedliche Begriffspaare miteinander gleich, dieses Mal jedoch indirekt, das heißt durch Zitation anderer Autoren und ohne das zweite Begriffspaar überhaupt zu nennen. Diese indirekten Gleichsetzungen sind in der vierten Spalte von Tabelle 2 aufgelistet.

Unschwer lässt sich anhand der Tabelle erkennen, dass es kaum Paare von Wissensarten gibt, die nicht mit konzeptuellem und prozeduralem Wissen gleichgesetzt worden sind. Deutlich wird auch, dass dieser Trend in der Mitte der achtziger Jahre begann. Mit der integrativen theoretischen Analyse von Greeno, Riley und Gelman (1984, vgl. Abschnitt 2.2.2) und dem Einleitungskapitel von Hiebert und Lefevre (1986) beginnt ein Trend, der in den aktuellen Publikationen (Rittle-Johnson & Alibali, 1999; Rittle-Johnson et al., 2001) zum Iterativen Modell seinen Höhepunkt findet. Plädoyers dafür, dass zwei Paare von Wissensarten nicht miteinander gleichgesetzt werden sollten, findet man in der Literatur hingegen nicht. Die einzige Ausnahme bildet die auf Seite 36 zitierte Definition von konzeptuellem und prozeduralem Wissen von Rittle-Johnson et al. (2001), die besagt, dass konzeptuelles Wissen sowohl explizit als auch implizit sein kann (vgl. auch Unterkapitel 0). Mandler (2004, S. 41-58) bekräftigt nach ausführlicher Diskussion

Tabelle 2: Zitationsanalyse der synonymen Verwendung von Namen für Wissensarten.

Autoren	Analysiertes Begriffspaar	Direkt verwendete Synonyme	Indirekt verwendete Synonyme
Gelman & Gallistel (1978)	principles & skill	–	–
Piaget & Inhelder (1980)	structures & processes	knowing-how & knowing why; action & meaning of action	–
Resnick (1982)	semantics & syntax	concepts & algorithms; concepts & procedures	–
Anderson (1983)	declarative & procedural knowledge	–	–
Gelman & Meck (1983)	principles & skill	–	–
Greeno, Riley & Gelman (1984)	conceptual & procedural competence	–	principles & skill (Gelman & Gallistel, 1978; Gelman & Meck, 1983); explicit & implicit procedures (Karmiloff-Smith, 1979)
Hiebert & Lefevre (1986)	conceptual & procedural knowledge	concepts & procedures; understanding & skill; semantics & syntax; principles & skills; meaningful & mechanic knowledge; schematic & teleologic knowledge	conceptual understanding & successful action (Piaget, 1978); declarative & procedural knowledge (Anderson, 1983); knowing that & knowing how to (Scheffler, 1965); semantic & episodic memory (Tulving, 1983);
Resnick & Omanson (1987)	–	–	semantic & syntax
Byrnes & Wasik (1991)	conceptual & procedural knowledge	–	declarative & procedural knowledge (Anderson, 1983); explicit & implicit knowledge (Karmiloff-Smith, 1986); structures & processes (Inhelder & Piaget, 1980)
Rittle-Johnson & Siegler (1998)	conceptual & procedural knowledge	concepts & procedures; conceptual understanding & procedural skill; understanding/principled knowledge & skill/algorithms/strategies;	principles & skills (Gelman & Gallistel, 1978);
Rittle-Johnson & Alibali (1999)	conceptual & procedural knowledge	concepts & procedures;	declarative & procedural knowledge (Singley & Anderson, 1989); principles & skill (Gelman & Gallistel, 1978; Gelman & Meck, 1983); implicit procedural & explicit conceptual knowledge (Karmiloff-Smith, 1986; Karmiloff-Smith, 1992)
Rittle-Johnson, Siegler & Alibali (2001)	conceptual understanding & procedural skill	concepts & procedures; conceptual & procedural knowledge	competence & performance (Sophian, 1997); explicit and implicit knowledge (Siegler & Stern, 1998); declarative & procedural knowledge (Anderson, 1993); principles & strategies (Baroody & Gannon, 1984; Dixon & Moore, 1996); principles & skills (Gelman & Gallistel, 1978; Gelman & Meck, 1983); conceptual & procedural competence (Greeno et al., 1984); explicit conceptual & implicit procedural knowledge (Karmiloff-Smith, 1992); conceptual understanding & successful action (Piaget, 1978)

die Richtigkeit dieser Annahme mit der Begründung, dass die Begriffe konzeptuelles und prozedurales Wissen psychische *Strukturen* beschreiben, die Begriffe explizit und implizit hingegen eher *Verarbeitungsmodi* bezeichnen, die beide Arten von Wissen (zumindest aber konzeptuelles) betreffen können.

Ob die Gleichsetzung der verbleibenden Paare von Wissensarten, die so weit verbreitet ist, gerechtfertigt ist, kann an dieser Stelle nicht entschieden werden, vor allem, weil sogar für die Paare einzeln betrachtet unklar ist, ob die Unterscheidung zweier Wissensarten sinnvoll ist und wie sie genau charakterisiert werden kann. Die vorliegende Arbeit konzentriert sich darum auf die Unterscheidung zwischen konzeptuellem und prozeduralem Wissen. Dort, wo Rückgriffe auf Befunde über andere Wissensarten stattfinden, wird das angemerkt.

Wie sich konzeptuelles und prozedurales Wissen definitorisch gegeneinander abgrenzen lassen, ist Gegenstand des folgenden Abschnitts.

Definitorische Eingrenzung

Tabelle 3 gibt die sechs Definitionen von konzeptuellem und prozeduralem Wissen wieder, die sich in der gängigen Literatur finden lassen. Die älteste stammt aus dem Einleitungskapitel von Hiebert und Lefevre, die neueste ist von Rittle-Johnson und Kollegen. Während sie sich in Wortwahl und Ausführlichkeit unterscheiden, sind sie sich inhaltlich ähnlich.

Konzeptuelles Wissen wird vor allem beschrieben als vernetztes Wissen, das in Schemata, semantischen Netzwerken oder Hierarchien gespeichert werden kann. Sein Inhalt besteht in den Prinzipien (principles) und Kernkonzepten (core concepts) einer Domäne. Es bildet die Grundlage inhaltlichen Verständnisses (understanding). Prozedurales Wissen wird vor allem beschrieben als Menge von Regeln, deren sequenzielle Anwendung die zielgerichtete Lösung von Problemen ermöglicht.

Im Gegensatz zu den anderen Definitionen greift die von Byrnes und Wasik zusätzlich noch auf die Unterscheidung von „Wissen, dass“ (knowing that) und „Wissen, wie“ (knowing how) zurück. Damit verweisen sie indirekt auf die philosophische Position Gilbert Ryles (1949), der das Begriffspaar prägte. Inhaltlich hat dieses Vorgehen jedoch einen Nachteil. Ein Schüler, der 30 Nachkommastellen der Zahl π auswendig lernt, erwirbt unzweifelhaft „Wissen, dass“. Worin sein Zuwachs an konzeptuellem Verständnis liegen soll, ist jedoch nicht zu sehen. Konzeptuelles Wissen wird daher besser durch die Formulierung „Wissen, warum“ (knowing why) beschrieben (Hiebert & Lefevre, 1986).

Tabelle 3: Definitionen konzeptuellen und prozeduralen Wissens.

Quelle	Konzeptuelles Wissen	Prozedurales Wissen
Hiebert & Lefevre (1986, S. 3ff)	“Conceptual knowledge is characterized most clearly as knowledge that is rich in relationships. It can be thought of as a connected web of knowledge, a network in which the linking relationships are as prominent as the discrete pieces of information. Relationships pervade the individual facts and propositions so that all pieces of information are linked to some network.”	“Procedural knowledge consists of rules, algorithms, or procedures used to solve mathematical tasks. A key feature of procedures is that they are executed in a predetermined linear sequence. It is the clearly sequential nature of procedures that probably sets them most apart from other forms of knowledge. [...] The procedures we are describing can be characterized as production systems (Anderson, 1983, Newell & Simon, 1972) in that they require some sort of recognizable input for firing.”
Byrnes & Wasik (1991, S. 777)	“Conceptual knowledge, which consists of the core concepts for a domain and their interrelations (i.e., ‘knowing that’), has been characterized using several different constructs, including semantic nets, hierarchies, and mental models.”	“Procedural knowledge, on the other hand, is ‘knowing how’, or the knowledge of the steps required to attain various goals. Procedures have been characterized using such constructs as skills, strategies, productions, and interiorized actions.”
Rittle-Johnson & Siegler (1998, S. 77)	“We define conceptual knowledge as understanding of the principles that govern the domain and of the interrelations between pieces of knowledge in a domain (although this knowledge does not need to be explicit). In the literature, this type of knowledge is also referred to as understanding or principled knowledge.”	“We define procedural knowledge as action sequences for solving problems. In the literature, this type of knowledge is sometimes referred to as skills, algorithms, or strategies.”
Rittle-Johnson & Alibali (1999, S. 175)	“We define conceptual knowledge as explicit or implicit understanding of the principles that govern a domain and of the interrelations between pieces of knowledge in a domain.”	“We define procedural knowledge as action sequences for solving problems.”
Haapsalo & Kadijevich (2000, S. 141)	“Conceptual knowledge denotes knowledge of and skilful ‘drive’ along particular networks, the elements of which can be concepts, rules (algorithms, procedures, etc.), and even problems [...] given in various representational forms.”	“Procedural knowledge denotes dynamic and successful utilization of particular rules, algorithms or procedures within relevant representational form(s). This usually requires not only the knowledge of the objects being utilized, but also knowledge of format and syntax for the representational system(s) expressing them.”
Rittle-Johnson et al. (2001, S. 346)	“We define conceptual knowledge as implicit or explicit understanding of the principles that govern a domain and of the interrelations between units of knowledge in a domain. This knowledge is flexible and not tied to specific problem types and is therefore generalizable. Furthermore, it may or may not be verbalizable. To assess conceptual knowledge, researchers use novel tasks, such as counting in nonstandard ways or evaluating unfamiliar procedures. Because children do not already know a procedure for solving the task, they must rely on their knowledge of relevant concepts to generate methods for solving the problems.”	“We define procedural knowledge as the ability to execute action sequences to solve problems. This type of knowledge is tied to specific problem types and therefore is not widely generalizable. To assess procedural knowledge researchers typically use routine tasks, such as counting a row of objects or solving standard arithmetic computations, because children are likely to use previously learned step-by-step solution methods to solve the problems.”

Vielleicht sollte man, wie die anderen fünf Autorengruppen, auf solche inhaltlich und grammatikalisch eher verwirrende Charakterisierungen ganz verzichten.

Die beiden Definitionen prozeduralen Wissens von Rittle-Johnson und Siegler (1998) sowie Rittle-Johnson und Alibali (1999) beschreiben prozedurales Wissen als Handlungsabfolgen (action sequences). Im Gegensatz zur in der Kognitionspsychologie weit verbreiteten Ansicht, dass Wissen Ursprung oder Grundlage von Handlungssequenzen ist und dass Wissen und Handlung somit zu unterscheiden sind, werden prozedurales Wissen und Handeln dabei gleichgesetzt. Dieser quasi behavioristische Ansatz hat eine unplausible Implikation: Eine Person, die gerade nicht handelt, besäße ihr zufolge kein prozedurales Wissen, so dass nicht erklärt werden kann, warum interindividuelle Wissensunterschiede meist über Situationen hinweg stabil sind. Die neuere Definition von Rittle-Johnson et al. (2001) umgeht dieses Problem, indem sie prozedurales Wissen als Fähigkeit zur Ausführung einer Handlungssequenz und nicht mehr als Handlungssequenz selbst bestimmt. Jemand der nicht handelt, kann trotzdem die Fähigkeit zum Handeln besitzen und mithin stabiles prozedurales Wissen.

Eine wichtige Eigenschaft prozeduralen Wissens wird in allen vorliegenden Definitionen nicht benannt: Prozedurales Wissen ermöglicht nicht nur die bloße Lösung von Problemen, sondern befähigt darüber hinaus auch zu einer schnellen und ressourceneffizienten Problemlösung (Anderson, 1983; Johnson, 2003; vgl. Abschnitt 4.4.2).

Das Definitionspaar von Rittle-Johnson et al. (2001) bezieht die unterschiedliche Transferierbarkeit der beiden Wissensarten auf neuartige Probleme mit ein und beschreibt konzeptuelles Wissen als flexibel auf neue Probleme transferierbar und prozedurales Wissen als an Routineprobleme gebunden. Haapsalo und Kadijevich (2000) beschreiben dieses Phänomen anstatt auf der Aufgabenebene auf der Ebene externer Repräsentationsformen. Ihnen zufolge kann konzeptuelles Wissen über externe Repräsentationsformen hinweg generalisiert werden, während prozedurales Wissen jeweils an bestimmte dem Problemlöser schon bekannte Repräsentationsformen gebunden ist.

Aus diesen unterschiedlichen Transferierbarkeiten leiten Rittle-Johnson et al. (2001) in ihrer Definition die typischen Operationalisierungen der Wissensarten ab, auf die ihre Definition als Einzige hinweist: Konzeptuelles Wissen wird häufig mittels neuartiger Aufgaben, prozedurales Wissen häufig mittels Routineaufgaben gemessen.

In diesen Aussagen der Definitionspaare von Haapsalo und Kadijevich sowie Rittle-Johnson und Kollegen liegt aber, wie schon bei der Diskussion des Iterativen Modells in Abschnitt 3.4 beschrieben, auch ein Nachteil: Wünschenswerterweise sollten Aussagen

über die unterschiedlichen Transferfähigkeiten der Wissensarten das Resultat empirischer Forschung und nicht a priori festgelegt sein. Empirische Befunde aus der Forschung zu prozeduralem Transfer (z.B. Novick, 1995, vgl. auch Abschnitt 4.3) belegen, dass Rittle-Johnsons Annahme, dass prozedurales Wissen immer an Routineprobleme gebunden ist, zumindest in dieser Generalität nicht richtig ist, so dass weitere empirische Forschung zu dieser Frage angebracht wäre.

Um der empirischen Forschung nicht vorzugreifen, soll an dieser Stelle auf eine Entscheidung für eines der fünf zitierten Definitionspaare oder die Aufstellung eigener Definitionen verzichtet werden. Stattdessen werden in den folgen Unterkapiteln die Wissensarten noch genauer inhaltlich beschrieben. In den Unterkapiteln 4.2 und 4.3 werden Argumente für und wider die Unterscheidung von konzeptuellem und prozeduralem Wissen diskutiert, Unterkapitel 4.4 gibt einen Review der vorliegenden Erkenntnisse über die mentalen Repräsentationsformen der Wissensarten und in Unterkapitel 4.5 wird das hier schon angerissene Thema der typischen Operationalisierungen beider Wissensarten vertieft.

4.2 Gründe gegen die Unterscheidung

4.2.1 Kritik aus Sicht des Ansatzes der Situierten Kognition

Die Unterscheidung von konzeptuellem und prozeduralem Wissen ist – wie Unterscheidungen zwischen Wissensarten generell – keinesfalls unumstritten. Die schärfsten Kritiker bestreiten aus der Perspektive der *situated cognition* beziehungsweise der *situated action*, dass die Postulierung mentaler Wissensrepräsentationen jeglicher Art sinnvoll ist. Die kognitive und die situierte Position beschreiben Cobb und Bowers so: „In the case of the cognitive perspective, a central organizing metaphor is that of knowledge as an entity that is acquired in one task setting and conveyed to other task settings. In contrast, a primary metaphor of the situated learning perspective is that of knowing as an activity that is situated with regard to an individual’s position in the world of social affairs“ (Cobb & Bowers, 1999, S. 5). Versteht man unter Wissen kognitive Entitäten, so erscheint es plausibel, diese Entitäten nach ähnlichen Eigenschaften in Kategorien, beispielsweise in konzeptuelle und prozedurale, einzuteilen. Versteht man unter Wissen hingegen eine situationsgebundene Aktivität, so erscheint die Annahme von Wissensarten deutlich weniger sinnvoll, denn: Im konkreten situationsbezogenen Verhalten zeigen sich die Wissensarten, falls es sie überhaupt gibt, immer nur ineinander verwoben und nie isoliert (Clancey, 1993; Greeno, 1997; Greeno, 1998; Säljö, 1999).

In der Auseinandersetzung zwischen dem Kognitionspsychologen John R. Anderson und dem Anhänger der Situierten Kognition James G. Greeno kristallisierte sich ein auch empirisch überprüfbarer Unterschied zwischen den Aussagen beider Positionen heraus: Wenn Wissen lediglich in situationsgebundenen Handlungen liegt, dann sollte ein Wissenstransfer über Situationen und Aufgabentypen hinweg, die unterschiedliche Handlungen oder unterschiedliche mentale Operationen erfordern, nicht möglich sein. Lediglich, wenn zwei Situationen dem Handelnden die gleichen Handlungsmöglichkeiten und -grenzen bieten – Greeno spricht hier von *affordances* und *constraints* im Sinne Gibsons ökologischer Theorie der Wahrnehmung (vgl. Greeno, 1994) – sollte Wissenstransfer möglich sein (Greeno, Moore, & Smith, 1993). Liegt Wissen hingegen in mentalen Repräsentationen, die durch eine kognitive Architektur jeweils mehr oder weniger flexibel transformiert und dadurch neuen Situationen angepasst werden können, so sollte Wissenstransfer auch zwischen Situationen stattfinden können, die unterschiedliche mentale oder motorische Operatoren erfordern und lediglich in abstrakt-relationalen Merkmalen übereinstimmen (Anderson, Reder, & Simon, 1996, 1997). Anderson (1990) bezeichnet diese Anpassungsleistung als den „adaptive character of thought“.

Den prinzipiell möglichen Untersuchungen darüber, ob Wissen über Situationen mit unterschiedlichen *affordances* und *constraints* transferiert werden kann, steht das Problem entgegen, dass niemand bestimmen kann, welche oder wie viele *affordances* und *constraints* zwei Situationen miteinander gemeinsam haben. Unter anderem Barnett und Ceci (2002) zeigen jedoch in einem beeindruckenden Literaturüberblick, dass es eine ganze Reihe von Befunden gibt, die sich zu Recht als Nachweis des Ferntransfers von Wissen bezeichnen lassen. Diese Befunde sind ohne die Annahme, dass es abstraktes Wissen gibt, schwer zu erklären. Das widerspricht nicht der Annahme, dass auch situative Faktoren beim Wissenserwerb und -abruf von großer Wichtigkeit sind.

Für Überlegungen zum konzeptuellen und prozeduralen Wissen ist in jedem Fall interessant, dass für Anderson und Greeno, ebenso wie für Rittle-Johnson und Kollegen, die Frage nach der Transferfähigkeit von Wissen zentral mit der Frage nach Wissensarten verbunden ist.

4.2.2 Kritik aus didaktischer Sicht

Direkt kritisiert wird die Unterscheidung zwischen konzeptuellem und prozeduralem Wissen von Claudia von Aufschnaiter (2005). Sie untersuchte anhand verschiedener Gruppen von Schülern und Studenten den Kompetenzerwerb im Bereich der Physik und versuchte, durch Auswertung von Videoaufzeichnungen von Unterrichtseinheiten Erkenntnisse über den Prozess des Kompetenzerwerbs zu gewinnen. In der Diskussion

beschreibt sie, dass Wissensarten nur schlecht geeignete Konstrukte sind, um die beobachteten Prozesse zu erklären, und beschreibt die Unterscheidung unterschiedlicher Wissensarten als unnötigerweise in Lernprozesse hineininterpretiert (von Aufschnaiter, in Vorbereitung).

Auch wenn Cobb und Greeno eine ausgesprochen theoretische Perspektive einnehmen und von Aufschnaiter eine angewandt-didaktische, ist ihr Grundargument doch das Gleiche: Die Wissensarten lassen sich anhand des Lernerverhaltens nicht differenzieren. Da wissenschaftstheoretisch gesehen nur solche theoretischen Konstrukte verwendet werden sollten, die zur Erklärung empirischer Phänomene unabdingbar sind (Poser, 2001, S. 48ff), sollte man in diesem Fall auf die Annahme unterschiedlicher Wissensarten verzichten.

Das Prinzip, nach dem eine wissenschaftliche Theorie so viele empirische Beobachtungen wie möglich durch so wenige theoretische Annahmen wie möglich erklären sollte, wird auch als Ockams Rasiermesser bezeichnet: Alle überflüssigen Annahmen einer Theorie sollen rigoros „weggeschnitten“ werden. Die Frage ist also: Gibt es empirische Phänomene, die ohne die Annahme zweier Wissensarten nicht erklärt werden können?

4.2.3 Kritik an der Undifferenziertheit der Unterscheidung

Einige Forscher, vor allem kognitiv orientierte, kritisieren die Unterscheidung aus einer entgegengesetzten Perspektive: Sie meinen nicht, dass die Unterscheidung zweier Wissensarten unnötig sei, sondern dass sie zu undifferenziert sei, weil es empirische Evidenz für eine Vielzahl unterschiedlicher Wissensarten gibt.

Lars Nyberg (1994) fand durch konfirmatorische Faktoranalysen, dass 14 Maße des deklarativen Gedächtnisses zwei verschiedene Faktoren bilden, die als semantisch-deklarative und episodisch-deklarative Gedächtnisleistung interpretiert werden können. Die Daten einer komplexen Folgeuntersuchung (Nyberg et al., 2003) legte sogar nahe, dass diese beiden Faktoren ihrerseits jeweils aus zwei Unterfaktoren bestehen.

Für prozedurales Wissen liegen ähnliche Befunde vor. Wie oben besprochen, ist die Automatisierung eine von mehreren wichtigen Eigenschaften prozeduralen Wissens. Verschiedene Autoren fanden jedoch sehr niedrige Korrelationen zwischen verschiedenen Maßen der Automatisierung (Besner, Stolz, & Boutilier, 1997; Kahneman & Chajczyk, 1983; Paap & Ogden, 1981; Regan, 1981; Strayer & Kramer, 1990). Da hierbei keine konfirmatorischen Faktoranalysen durchgeführt wurden, ist bisher unklar, ob prozedurales Wissen ein mehrdimensionales oder ein hierarchisches Konstrukt darstellt oder ob die

verschiedenen Maße des Konstrukts eventuell einfach aufgrund hoher Messfehler niedrig miteinander korrelierten (vgl. auch Jonides, Naveh-Benjamin, & Palmer, 1985; Pashler, Johnston, & Ruthruff, 2001).

Streng genommen spricht keine der in diesem Abschnitt vorgestellten Arbeiten gegen die Existenz konzeptuellen oder prozeduralen Wissens und keiner der genannten Autoren plädiert explizit dafür, die Unterscheidung zwischen den beiden Wissensarten aufzugeben. Sie machen aber doch deutlich, dass die Unterscheidung der Wissensarten, gerade auch aus einer messtheoretischen Perspektive, keineswegs unproblematisch ist und weiterer Elaboration bedarf.

4.3 Gründe für die Unterscheidung

4.3.1 Sprachlogische Beweise

Welche Gründe sprechen dafür, trotz der vorhandenen Zweifel konzeptuelles und prozedurales Wissen zu postulieren?

Byrnes und Wasik (1991) führen einen auf den ersten Blick interessanten Punkt an: Es gibt sprachlogische Beweise (z.B. Hintikka, 1975), die zeigen, dass die semantischen Inhalte prozedurales und konzeptuelles Wissen nicht (ohne Zusatzannahmen) ineinander umgewandelt werden können.

Bei näherer Betrachtung lässt sich gegen diesen Punkt jedoch einwenden, dass psychische Prozesse nicht zwangsläufig logischen Regeln folgen. Unter anderem die Experimente zum Hypothesentesten anhand der four card selection task (Wason, 1966) und zur Abweichung menschlicher Entscheider von mathematisch abgeleiteten Entscheidungsmodellen (z.B. Jungermann, 1983; Tversky & Kahneman, 1974) haben das deutlich gezeigt.

4.3.2 Lernpsychologische Argumente

Byrnes und Wasik führen noch drei weitere Gruppen von Argumenten an, die jedoch eher dem Bereich der Pädagogischen Psychologie zuzurechnen sind: (1) funktionale Argumente, die sich darauf beziehen, dass zur Ausführung unterschiedlicher kognitiver Funktionen unterschiedliche Wissensarten notwendig sind, zum Beispiel konzeptuelles Wissen, um Zusammenhänge erklären zu können, und prozedurales Wissen, um Probleme lösen zu können; (2) empirische Studien, die zeigen, dass Maße konzeptuellen und prozeduralen Wissens oft nicht perfekt korreliert sind; (3) Interventionsstudien, in denen dieselbe Intervention Maße konzeptuellen und prozeduralen Wissens unterschiedlich beeinflusste.

Punkt 1 ist wenig überzeugend. Aus der Beobachtung, dass die Psyche unterschiedliche Funktionen ausüben kann, wurde gefolgert, dass den Funktionen unterschiedliche Wissensarten zugrunde liegen müssen. Dieser Schluss ist jedoch nicht zwingend. So kann man sein Wissen um die Telefonnummern verschiedener Freunde sowohl benutzen, um sie anzurufen, als auch, um sie in einem Adressbuch zu notieren, als auch, um die Vorwahlen verschiedener Städte zu ermitteln. Ein und dieselbe Wissensseinheit kann also von unterschiedlichen psychischen Funktionen genutzt werden. Welches Wissen welchen Funktionen dienen kann, hängt dabei offensichtlich von Struktur und Funktionsweise der Psyche als Gesamtsystem ab. Die Gesamtarchitektur der Psyche ist jedoch noch nicht so weit bekannt, dass zwingende Annahmen über Wissensarten daraus abgeleitet werden könnten, im Gegenteil: Wenn Wissensarten bisher empirisch untersucht wurden, dann oft mit dem Fernziel, die Gesamtarchitektur der Psyche näher zu bestimmen (z.B. Anderson, 1983).

Punkt 2 ist etwas überzeugender. Wie in Kapitel 2 beschrieben, korrelieren Maße konzeptuellen und prozeduralen Wissens zwar oft positiv, nie jedoch perfekt miteinander und besaßen immer Korrelationskoeffizienten, die deutlich kleiner als eins waren. Doch hier liegt ein Zirkelschluss vor: Wenn man voraussetzt, dass es *Maße* konzeptuellen und prozeduralen Wissens gibt, dann folgt daraus auch ohne empirische Untersuchung, dass es konzeptuelles und prozedurales Wissen gibt. Was wäre aber, wenn die Maße in Wirklichkeit dasselbe Konstrukt erheben würden und nur aufgrund der Messungenauigkeit nicht vollständig korrelierten? Diese Möglichkeit kann für die vorliegenden Studien nicht ausgeschlossen werden. Beispielsweise fand Hecht (1998), wie oben beschrieben, für zwei Maße konzeptuellen Wissens, dass sie zwar positiv, aber nicht perfekt miteinander korrelieren.

Eine weitere Alternativerklärung für die teilweise Unabhängigkeit konzeptueller und prozeduraler Maße ist, dass die Maße zwar zwei unterschiedliche Konstrukte, nicht jedoch konzeptuelles und prozedurales Wissen erheben könnten. Schon in Kapitel 2.3.6 wurde dieses Problem im Zusammenhang mit Befunden zur Erwerbsreihenfolge der Wissensarten aufgeworfen. Wenn Kinder in einer Querschnittsstudie beispielsweise eine konzeptuelle Aufgabe lösen können, aber eine prozedurale nicht, könnte das zwar Aufschlüsse über die beiden Wissensarten und ihre Erwerbsreihenfolge geben. Genauso ließen sich die Ergebnisse jedoch dadurch erklären, dass die konzeptuelle Aufgabe leichter zu lösen ist als die prozedurale. Statt konzeptuellem und prozeduralem Wissen wären zwei Ausprägungen eines anderen Konstrukts, nämlich zwei Aufgabenschwierigkeiten, erhoben worden.

Diese Argumentation lässt sich auch auf Punkt 3 anwenden: Wenn man voraussetzt, dass in einer Studie ein Maß konzeptuellen und ein Maß prozeduralen Wissens verwendet wurden, geben unterschiedliche Reaktionen der Maße auf Interventionen sicherlich einen Hinweis auf unterschiedliche Eigenschaften von konzeptuellem und prozeduralem Wissen. Aber wenn man aus der Tatsache, dass verschiedene Maße unterschiedlich auf ein Treatment reagieren, auf die Vorhandenheit von konzeptuellem und prozeduralem Wissen schließen will, müsste man sie theoriegeleitet so auswählen, dass sie nicht nur anzeigen, dass zwei unterschiedliche Konstrukte vorliegen, sondern auch, dass sie zwei unterschiedliche Konstrukte messen, die genau die Eigenschaften haben, über die konzeptuelles und prozedurales Wissen definiert sind.

Wie oben erläutert, ist die einzige vorliegender Studie, bei der eine theoriegeleitete Auswahl der Maße benutzt wurde, die von Rittle-Johnson et al. (2001). Diese liefert jedoch keinen experimentellen Nachweis der Unterschiedlichkeit von konzeptuellem und prozeduralem Wissen. Im Gegenteil: Die eingesetzten Maße reagierten entgegen der Erwartungen nicht unterschiedlich auf das Treatment. Setzt man voraus, dass es die Wissensarten gibt und dass sie durch die Maße erhoben wurden, belegt die Studie, dass die Wissensarten miteinander korreliert sind und sich wechselseitig über die Zeit hinweg positiv beeinflussen. Dass es wirklich zwei Wissensarten gibt, belegt die Studie jedoch nicht, sondern sie setzt es – wie die anderen empirischen Studien in dem Forschungsfeld – voraus.

4.3.3 Befunde aus der Gedächtnisforschung

Schacter (1987) beschreibt in einem historischen Überblick, dass schon die Philosophen Descartes und Leibniz darauf hingewiesen haben, dass es neben dem bewussten Erinnern auch eine unbewusste Form des Erinnerns gibt, die sich im Verhalten zeigt. Tatsächlich wurden solche Prozesse ab der zweiten Hälfte des neunzehnten Jahrhunderts empirisch untersucht. Sie inspirierten psychoanalytische ebenso wie neurologische und experimentalpsychologische Arbeiten (z.B. Thorndike & Rock, 1934). Insbesondere die experimentalpsychologische Gedächtnisforschung hat dieses Phänomen und die Bedingungen, unter denen es auftritt, inzwischen ausgiebig untersucht und beispielsweise gezeigt, dass Probanden eine einmal auswendig gelernte Liste beim zweiten Mal schneller auswendig lernen, auch dann, wenn sie sich zwischendurch an kein einziges Element der Liste bewusst erinnern konnten (Nelson, 1978). Probanden können die syntaktische Korrektheit von Ausdrücken einer künstlichen Sprache nach einer Lernphase teilweise auch dann beurteilen, wenn ihnen die syntaktischen Regeln der Sprache nicht bewusst sind (Reber, 1989). Vergleichbare Effekte wurden in so unterschiedlichen Paradigmen wie

Priming und lexical decision tasks (vgl. Schacter, 1987) sowie komplexem Problemlösen (Berry & Broadbent, 1984; Berry & Broadbent, 1995) nachgewiesen.

Differenzen zwischen explizitem und implizitem Erinnern lassen sich durch drei theoretische Ansätze erklären (Foster & Jelicic, 1999): *Aktivierungstheorien* führen das unterschiedliche Verhalten von Probanden bei Tests des expliziten und impliziten Wissens darauf zurück, dass die Tests die vorhandenen Gedächtnisspuren unterschiedlich aktivieren. Die unterschiedliche Aktivierung der Gedächtnisspuren hat wiederum unterschiedliches Verhalten zur Folge. *Prozesstheorien* postulieren, dass im Kontext expliziten Lernens und Erinnerns andere Enkodierungs- und Abrufprozesse ablaufen als im Kontext impliziten Lernens. Bei der Bearbeitung impliziter Wissenstests dominieren ihnen zufolge datengetriebene *bottom-up*-Prozesse, bei der Bearbeitung expliziter Wissenstests konzeptgesteuerte *top-down*-Prozesse.

Beide Ansätze, Aktivierungstheorien und Prozesstheorien, basieren also auf der Annahme, dass es nur eine Art von Wissen gibt, die sich lediglich bei expliziten und impliziten Wissenstests unterschiedlich zeigt. Im Gegensatz dazu postulieren Anhänger einer dritten Position, den *multiple memory systems*-Ansätzen, dass explizites und implizites Lernen in unterschiedlichen Gedächtnissystemen stattfinden, die jeweils eine spezifische Art von Wissen verarbeiten.

Einer der ersten und wichtigsten Vertreter dieser Ansätze ist Endel Tulving. Gedächtnissysteme charakterisiert er wie folgt:

Memory systems constitute the major subdivisions of the overall organization of the memory complex. They are organizational structures of more elementary operating components. An operating component of a system consists of a neural substrate and its behavioral or cognitive correlates. Some components are shared by all systems, others are shared only by some, and still others are unique to individual systems. [...] Although there is no one-to-one correspondence between tasks and systems (e.g., Kinsbourne, 1976; Tulving, in press), they are nonetheless systematically related: A given memory system makes it possible for organisms to perform memory tasks that entail operating components unique to that systems. (Tulving, 1985, S. 386)

Tulving postuliert drei Gedächtnissysteme: das prozedurale, das semantische und das episodische Gedächtnis (z.B. Tulving, 1985; Tulving, 1972). Das prozedurale Gedächtnis ermöglicht die angemessene Reaktion auf äußere Reize, und sein Inhalt ist nicht bewusst. Das semantische Gedächtnis stellt Faktenwissen für die bewusste Verarbeitung bereit. Es ermöglicht die Reflexion über Zusammenhänge in der Welt. Das episodische Gedächtnis speichert Erinnerungen an erlebte Situationen. Es ermöglicht damit Selbstbewusstsein und die bewusste Reflexion der eigenen Lebens- und Lerngeschichte.

Nachfolgend entbrannte eine Debatte darüber, ob Prozesstheorien oder multiple memory systems-Ansätze Dissoziationen zwischen explizitem und implizitem Lernen besser erklären können. Mitchell (1989; Mitchell, Brown, & Murphy, 1990) fand in entwicklungspsychologischen Studien mit Erwachsenen, dass Maße der prozeduralen, der semantischen und der episodischen Gedächtnisleistung sich im Querschnitt unterschiedlich entwickeln. Mit Hilfe einer explorativen Faktoranalyse konnte er darüber hinaus zeigen, dass elf unterschiedliche Gedächtnismaße drei Faktoren bilden, die der Tulvingschen Theorie gut entsprechen.

Fast gleichzeitig wurden jedoch auch neue Befunde und Interpretationen älterer Daten berichtet, die darauf hindeuten, dass Dissoziationen zwischen prozeduralen, semantischen und episodischen Maßen besser durch prozesstheoretische Annahmen als durch die Annahme unterschiedlicher Gedächtnissysteme erklärt werden (z.B. Blaxton, 1989; Humphreys, Bain, & Pike, 1989).

Dass heute trotz der Uneindeutigkeit der experimentalpsychologischen Evidenz weitgehende Einigkeit über die Existenz multipler Gedächtnissysteme herrscht, liegt an überzeugenden Befunden aus der neurologischen Forschung.

4.3.4 Neuropsychologische Evidenz

In der neurologischen Forschung spricht man, wie auch in der Gedächtnisforschung, eher von Gedächtnissystemen als von Wissensarten. Der Begriff des Gedächtnissystems verdeutlicht, dass die Existenz unterschiedlicher Wissensarten überhaupt nur dann sinnvoll angenommen werden kann, wenn man zugleich unterschiedliche Verarbeitungsmechanismen für die Wissensarten postuliert (Anderson, 1978). In diesem Sinne geben Befunde über Wissensarten und ihre Eigenschaften immer auch Aufschluss über die Architektur der Psyche als informationsverarbeitendem System (Anderson, 1983).

Wissen hat außerdem *per definitionem* einen semantischen Gehalt, das heißt es ist bedeutungshaltig. Bedeutungen lassen sich, als immaterielle Konstrukte, nicht mit den Methoden der Neuroforschung untersuchen. Gedächtnissysteme können hingegen sowohl auf einer kognitiv-semantischen als auch auf einer neurologisch-materiellen Ebene untersucht und beschrieben werden. Der Begriff des Gedächtnissystems ist daher nicht nur umfassender als der Begriff der Wissensart, sondern er eröffnet zugleich eine interdisziplinäre Perspektive.

Squire beschreibt, wie innerhalb der Neurowissenschaften die Befunde aus drei verschiedenen Forschungssträngen zu der Einsicht, dass tatsächlich unterschiedliche Gedächtnissysteme existieren, konvergierten:

A central thesis is that work with rats, monkeys, and humans – which has sometimes seemed to proceed independently in 3 separate literatures – is now largely in agreement about the function of the hippocampus and related structures. A biological perspective is presented, which proposes multiple memory systems with different functions and distinct anatomical organizations. The hippocampus (together with anatomically related structures) is essential for a specific kind of memory, here termed declarative memory (similar terms include explicit and relational). Declarative memory is contrasted with a heterogeneous collection of nondeclarative (implicit) memory abilities that do not require the hippocampus (skills and habits, simple conditioning, and the phenomenon of priming). (Squire, 1992, S. 195)

Dass die Forschung mit Menschen, Affen und Ratten konvergierende Befunde liefert, ist vor allem deswegen bemerkenswert, weil die drei Forschungsstränge unterschiedlichen methodischen Einschränkungen unterliegen. Forschung zu Läsionen des menschlichen Gehirns folgt grundsätzlich korrelativen Designs, weil ausschließlich bei Unfällen oder Operationen entstandene unsystematisch über Hirnregionen verteilte Schäden analysiert werden können. Die Validität dieser Befunde wird nicht nur durch die mangelnde Randomisierung in Frage gestellt, sondern auch durch die Tatsache, dass die Schäden meist nicht auf einzelne Hirnstrukturen beschränkt sind, sondern jeweils verschiedene Strukturen unterschiedlich stark beeinträchtigen.

Bei Affen und Ratten können jedoch einzelne Hirnstrukturen durch Medikamente oder invasive Eingriffe gezielt manipuliert werden. Im Zuge dieser Forschung konnte gezeigt werden, dass die Tiere bei Gedächtnisaufgaben ähnliche Verhaltensauffälligkeiten wie amnestische Patienten zeigen, wenn man ihnen vergleichbare Hirnschäden zufügt. Durch gezielte Variation der Läsionen bei den Tieren konnte dann die Bedeutung einzelner Hirnstrukturen, insbesondere die des Hippocampus, für das deklarative Gedächtnis nachgewiesen werden (siehe Heindel, Salomon, Shults, Walicke, & Butters, 1989; Schacter, 1989; Squire & Zola, 1996).

Die Befunde zeigten zunächst allerdings nur, dass es neurologisch unterscheidbare Gedächtnissysteme gibt, die unabhängig voneinander bestehen oder ausfallen können. Ausführliche Überblicke über diese Befundlage geben Gabrieli (1998), Rolls (2000) und Willingham (1997). Es stellt sich jedoch die Anschlussfrage, welchen einzelnen psychischen Funktionen die Systeme dienen. So wird in der Literatur ausführlich diskutiert, welche psychologische Unterscheidung zwischen Wissensarten Unterschiede zwischen den neurologischen Gedächtnissystemen am besten charakterisiert: semantisches versus episodisches Wissen (vgl. Glisky, 1992), explizites versus implizites Wissen (Schacter, 1992), relationales versus isoliertes Wissen (Cohen, Poldrack, & Eichenbaum,

1997) oder flexibel transferierbares versus inflexibles kontextgebundenes Wissen (Reber, Knowlton, & Squire, 1996). Letztere Unterscheidung ist im Rahmen der vorliegenden Arbeit besonders interessant, weil sie den oben diskutierten lerntheoretischen Annahmen von Rittle-Johnson (siehe Kapitel 3) und Anderson (siehe Abschnitt 4.2.1) entspricht.

Reber et al. (1996) brachten zur Überprüfung der Annahme acht amnestischen Patienten mit beeinträchtigtem deklarativem Gedächtnis und 16 Probanden mit normaler Gedächtnisleistung in einer Trainingsstudie bei, Muster von Stimuli aufgrund komplexer probabilistischer Beziehungen zu klassifizieren, wobei beide Gruppen gleiche Leistungen erzielten. Bei einem anschließenden Transfertest waren die amnestischen Patienten der Kontrollgruppe jedoch unterlegen. Da die amnestischen Patienten das neu erworbene Wissen nur im nicht-deklarativen Gedächtnissystem abspeichern konnten, den Probanden mit normalem Gedächtnis jedoch deklaratives und nicht-deklaratives Gedächtnis zur Verfügung standen, belegen die Ergebnisse laut Reber et al., dass das deklarative Gedächtnis flexibel transferierbares Wissen, das nicht-deklarative Gedächtnis jedoch kontextgebundenes Wissen speichert.

Glisky (1992) konnte allerdings mit Hilfe eines leicht abgeänderten Designs nachweisen, dass amnestische Patienten mit beeinträchtigtem deklarativem Gedächtnis unter bestimmten Trainingsbedingungen doch Transferleistungen erbringen können. Ähnliche Probleme ergeben sich auch für die anderen vorgeschlagenen psychologisch-funktionalen Unterschiede zwischen den Gedächtnissystemen. Willingham (1998) weist daher alle bisher gemachten Vorschläge als unzureichend zurück:

The procedural/declarative distinction is rooted in neuroanatomy. Declarative memories depend on structures in the medial temporal lobe and diencephalon, whereas procedural memories do not (Squire, 1992). Procedural memories appear to depend on cortical networks, but also on various subcortical structures including the basal ganglia and the cerebellum (see Willingham, 1997, for a brief review). This anatomic evidence is quite complete and not in dispute here. Many researchers agree that there is a fundamental distinction to be made between declarative and procedural memories (although they do not always use those terms) but the distinction has remained largely neuroanatomic. One would expect that computational differences would go hand in hand with anatomic differences, but there has been a paucity of proposals of the computational differences between procedural and declarative memories. (Willingham, 1998, S. 689f)

Stark, Stark und Gordon (2005) bestätigen Willinghams Auffassung, dass die komputationalen beziehungsweise psychologischen Eigenschaften der Gedächtnissysteme nicht eindeutig feststehen, auf Grundlage eines aktuellen Literaturüberblicks und berichten die Ergebnisse einer Fallstudie, die erneut die Fähigkeit eines amnestischen Patienten

belegt, semantisches und generalisierbares Wissen zu erwerben – wenn auch nur unter speziellen Trainingsbedingungen.

Es bleibt also festzuhalten, dass die neurologische Forschung valide Evidenz für die Existenz unterschiedlicher Gedächtnissysteme liefert, dass bisher jedoch unklar ist, welche exakten psychologischen Eigenschaften diesen Systemen zugeschrieben werden können. Auch ist offen, inwieweit die an Amnesiepatienten sowie Affen und Ratten gewonnenen neurologischen Befunde auf die Normalbevölkerung generalisiert und beispielsweise zur Gestaltung von Lernsituationen genutzt werden können.

4.3.5 Evolutionspsychologische Überlegungen

Die bisher vorgestellten Untersuchungen setzten sich mit der Frage auseinander, wie viele Wissensarten es gibt. Evolutionspsychologische Ansätze beschäftigen sich aus einer anderen Perspektive mit dem Thema, indem sie diskutieren, warum überhaupt im Verlaufe der menschlichen Entstehungsgeschichte unterschiedliche Gedächtnissysteme beziehungsweise Wissensarten hätten entstehen sollen.

Sherry und Schacter (1987) stützen ihre Argumentation auf das Konstrukt der *adaptiven Spezialisierung*. Dieses stammt aus der Evolutionstheorie und ist sehr bekannt. Ein Beispiel für eine adaptive Spezialisierung ist der lange Hals der Giraffe. Durch die Ausbildung des langen Halses sind Giraffen darauf spezialisiert hochgelegenes Futter zu erreichen, das den meisten anderen Tieren verwehrt ist. Dieser Vorgang ist adaptiv, weil er eine Anpassung an ganz bestimmte Umweltbedingungen darstellt. Sherry und Schacter argumentieren dafür, dass solche adaptiven Spezialisierungen auch das Gedächtnis betreffen können. Die Anpassung des Gedächtnisses an spezifische Umweltbedingungen könnte sich in einer Spezies durchsetzen, wenn sie dieser Spezies einen Evolutionsvorteil bietet.

Das zweite Konstrukt, auf das Sherry und Schacter sich beziehen, ist das der *funktionalen Inkompatibilität*. Diese liegt dann vor, wenn angesichts der Umweltbedingungen zwei verschiedene adaptive Spezialisierungen einen Evolutionsvorteil bieten, diese jedoch nicht in ein und demselben System verwirklicht werden können. Man denke an die Zellen der menschlichen Netzhaut: Zapfen sind ideal zum hochauflösenden Sehen in Farbe, sind jedoch nicht sehr lichtempfindlich und funktionieren daher nur im Hellen; Stäbchen liefern Bilder, deren Farbqualität und Auflösung nicht so hoch ist. Aufgrund ihrer Lichtempfindlichkeit funktionieren sie jedoch auch im Dunkeln. Lichtrezeptorzellen, die die Vorteile von Stäbchen und Zapfen in sich vereinen, gibt es nicht, d.h. es liegt funktionelle Inkompatibilität vor. Welche Art von Zellen sollten also

optimalerweise die menschliche Netzhaut bilden? – Die „Antwort“ der Natur auf diese Frage ist bekannt: Statt *eines* Rezeptortyps sind *beide* auf der Netzhaut zu finden. Nach demselben Prinzip könnten auch zwei oder mehr unterschiedliche Gedächtnissysteme entstanden sein.

Diese Überlegungen beweisen nicht, dass es mehrere Wissensarten gibt. Sie legen lediglich dar, warum sie entstanden sein *könnten*. Da die menschliche Evolution ein singuläres historisches Ereignis darstellt, sind evolutionspsychologische Überlegungen dazu stets *post hoc*-Erklärungen, die – zumindest als Ganzes – nicht experimentell überprüft werden können. Dies schmälert jedoch nicht die Plausibilität der geschilderten Argumentation, die gut auf Konstrukte wie ein konzeptuelles und ein prozedurales Gedächtnissystem anwendbar ist: Ersteres könnte gut durchdachte, dafür aber langsame Handlungsentscheidungen ermöglichen, Letzteres schnelle, dafür aber unreflektierte Entscheidungen. Beide Gedächtnisfunktionen böten einen evolutionären Vorteil, sind aber funktional inkompatibel, weil sie Anpassungen in gegensätzliche Richtungen voraussetzen. Eine elegante Formulierung dieser Idee stellt das zu Beginn dieser Arbeit angeführte Zitat Jerry Fodors dar: „Nature has contrived to have it both ways, to get the best out of fast and dumb systems and slow and contemplative ones, by simply refusing to choose between them“ (Fodor, 1985, S. 4).

Klein, Cosmides, Tooby und Chance (2002) verfolgen den Gedanken, dass evolutionsspsychologische Überlegungen zu multiplen Gedächtnissystemen insbesondere durch entscheidungspsychologische Überlegungen untermauert werden können. Durch einen Literaturreview und eigene Daten belegen sie, dass – vorausgesetzt, es gibt unterschiedliche Gedächtnissysteme – es einen großen Anpassungsvorteil böte, wenn diese Systeme nicht isoliert arbeiteten, sondern sich interagierend ergänzten. In gewisser Weise kann die Evolutionstheorie also nicht nur erklären, wie verschiedene Gedächtnissysteme entstanden sein könnten, sondern auch, warum sie sich im Verhalten nicht isoliert zeigen und so schwer zu erforschen sind.

4.3.6 Unterscheidung der Wissensarten: Fazit

Die Befundlage zur Existenz unterschiedlicher Wissensarten ist unübersichtlich. Philosophische und evolutionspsychologische Überlegungen zeigen, dass ihre Annahme plausibel ist, beweisen jedoch nicht, dass sie notwendig ist. Ähnliches gilt für die Befunde aus der Lernpsychologie und der Gedächtnisforschung: Sie sind sowohl mit der Annahme vereinbar, dass es unterschiedliche Wissensarten gibt, als auch mit der Annahme, dass es sie nicht gibt. In der neurologischen Forschung herrscht weitgehend Einigkeit über die Existenz neurologisch voneinander unabhängig funktionierender Gedächtnissysteme bei

Menschen, Primaten und Ratten. Trotz ausgiebigerer Untersuchungen sind die genauen psychologischen Charakteristika dieser Systeme bisher jedoch unbekannt.

Aus der in dieser Arbeit eingenommenen Perspektive der Pädagogischen Psychologie ist dieses Gesamtbild ernüchternd. Es lässt jedoch auch erkennen, nach welchen Arten von Evidenz in der Pädagogischen Psychologie gesucht werden sollte: Konzeptuelles und prozedurales Wissen sollten seltener als bisher durch einzelne Aufgabentypen operationalisiert werden, weil Befunde zu einzelnen Aufgabentypen stets auch durch aufgabenspezifisches Wissen oder die spezifischen Aufgabenschwierigkeiten erklärt werden können. Die Annahme einer Wissensart ist nur dann ökonomisch, wenn sich über verschiedene Aufgabentypen hinweg kohärente Verhaltensmuster bei Lernenden zeigen. Man spricht hier in der Messtheorie auch von der *konvergenten Validität* verschiedener Maße. Erst wenn sie nachgewiesen wurde, kann die Annahme eines theoretischen Konstruktes, das die Maße beeinflusst, gerechtfertigt werden.

Um die Annahme zweier voneinander verschiedener Wissensarten zu rechtfertigen, müsste entsprechend gezeigt werden, dass Maße der Lernleistung in zwei verschiedene Gruppen zerfallen. Auf Maßen, die derselben Gruppe entstammen, sollten die Probanden sich möglichst ähnlich verhalten. Maße, die nicht zu derselben Gruppe gehören, sollten hingegen zumindest teilweise unabhängig voneinander variieren. Letzterer Sachverhalt wird in der Messtheorie auch als *divergente Validität* von Maßen bezeichnet.

4.4 Hypothetische mentale Repräsentationen der Wissensarten

4.4.1 Repräsentationen konzeptuellen Wissens

Wie der vorige Abschnitt zeigt, ist nicht unumstritten, ob konzeptuelles und prozedurales Wissen wirklich zwei distinkte und empirisch unterscheidbare Konstrukte darstellen. Theorien über unterschiedliche mentale Repräsentationen der Wissensarten haben daher immer einen spekulativen Charakter. Trotzdem wurden auf diesem Gebiet einflussreiche theoretische Ansätze entwickelt. Goldstone und Kersten (2003) sowie Medin (1989) geben Überblicke über die wichtigsten vier Positionen und ihre Entwicklungen.

Frühe Ansätze begriffen konzeptuelles Wissen vor allem als Wissen über Kategorien und gingen davon aus, dass Wissen über solche Kategorien in einer Art *mentalem Lexikon* gespeichert wird. Die Einträge sollten in Form von Regeln oder einer Definition angeben, welche Eigenschaften ein Gegenstand besitzen muss, um einer bestimmten Kategorie zugeordnet werden zu können. Wie jedoch beispielsweise Wittgensteins bekannte Untersuchung des Begriffs Spiel zeigt, gibt es Kategorien, die durch eine einzelne Definition nicht ausschließlich und vollständig erfasst werden. Außerdem fallen Menschen

überraschend häufig über Zeit und Kontexte hinweg inkonsistente Urteile über Kategoriezugehörigkeiten von Objekten und geben an, dass unterschiedliche Objekte einer Kategorie unterschiedlich typisch für diese Kategorie sind, was sich durch Definitionen in einem mentalen Lexikon nur schwer erklären lässt.

Diese Befunde führten zur Entwicklung von *Prototypen-Theorien* (z.B. Mervis & Rosch, 1981; Rosch, 1975), die postulieren, dass Kategorien jeweils um die mentale Repräsentation eines kategorietyptischen Prototyps herum organisiert sind. Fragt man Passanten nach den Eigenschaften eines Hauses, so nennen sie stets dieselben Merkmale: Wände, Boden, Dach, Fenster, Türen. Diese Merkmale beschreiben die prototypische Vorstellung, die Menschen im Allgemeinen von der Kategorie Haus besitzen. Entscheidend ist dabei, dass dieses prototypische Haus nicht real zu existieren braucht, da es eine Abstraktion, sozusagen eine Durchschnittsbildung, von den Häusern ist, die die Menschen im Laufe ihres Lebens kennen gelernt haben.

Auf diese Annahme verzichten die *Beispiel-Theorien*. Sie postulieren, dass alle Mitglieder einer Kategorie, die im Gedächtnis verfügbar sind, eine Kategorie festlegen. Ein neues Objekt würde demnach klassifiziert, indem es mit den im Gedächtnis gespeicherten früher schon erlebten Objekten und ihren Kategoriezugehörigkeiten verglichen wird.

Ein weiterer Ansatz postuliert, dass mental nur die *Grenzen* zwischen Kategorien, nicht jedoch ihr Inhalt repräsentiert wird (z.B. Ashby & Gott, 1988). Eine empirisch bestätigte Implikation dieser Annahme ist, dass Menschen einander ähnliche Stimuli besser diskriminieren können, wenn sie zwei unterschiedlichen Kategorien angehören, als wenn sie derselben Kategorie angehören.

Gemeinsam ist den beschriebenen Ansätzen, dass sie auf der Annahme basieren, dass stets einander ähnliche Objekte zu einer Kategorie zusammengefasst werden. Jedoch lassen sich leicht Kategorien finden, die nicht auf perzeptueller Ähnlichkeit basieren. Murphy und Medin (1985) nennen das Beispiel eines Mannes, der vollständig bekleidet in einen Swimmingpool springt. Nicht wenige Menschen, die eine solche Szene, beispielsweise auf einer Party, erleben, werden den Mann in die Kategorie *betrunken* einordnen. Dies jedoch nicht, weil er oder die Situation irgendeine perzeptuelle Ähnlichkeit mit früheren Begegnungen mit Betrunkenen aufweist, sondern weil sie eine Erklärung für das ungewöhnliche Verhalten des Mannes suchen. Sie bilden sich also eine Theorie über das Verhalten des Mannes und kategorisieren ihn auf Grundlage dieser Theorie. Zahlreiche nachfolgende Arbeiten belegen, dass Ähnlichkeitsurteile und konzeptuelle Einordnungen hoch miteinander korrelieren können, aber nicht müssen, und dass die Charakterisierung von Konzepten als theoriebasiert angemessen ist. Unter

anderem Goldstone und Kersten (2003) weisen darauf hin, dass die älteren Ansätze dadurch nicht obsolet geworden sind. Denn es stellt sich die Frage, wie die konzeptuellen Theorien innerpsychisch repräsentiert werden. Die vorliegende Evidenz lässt darauf schließen, dass Regeln, Prototypen und Beispiele dabei gemeinsam mit anderen Repräsentationsformen eine wichtige Rolle spielen können. Vermutlich lässt sich konzeptuelles Wissen also in mehr als einer Repräsentationsform speichern, die ihrerseits in übergeordnete Theorien eingebettet sind, wobei unter einer Theorie ein „organized system of knowledge“ (Goldstone & Kersten, 2003, S. 610) verstanden wird.

Zu den kognitiven Strukturen, die solche organisierten Wissenssysteme bilden könnten, zählen alle Arten relationaler Repräsentationsformen, wie Schemata (Bartlett, 1932), Konzepthierarchien (Collins & Quillian, 1969) und assoziationalistische oder semantische Netzwerktheorien (z.B. Anderson & Bower, 1973; Collins & Loftus, 1975; Doumas & Hummel, 2005; Norman & Rumelhart, 1975). Für die Existenz solcher Strukturen sprechen zahlreiche Befunde, wie systematische Erinnerungsverzerrungen und Reaktionszeiteffekte, insbesondere beim semantischen Priming. Letzteres kann gut durch Aktivationsausbreitung durch die Kanten assoziationalistischer Netzwerke erklärt werden.

Insgesamt sah man Konzepte also früher eher als isolierte Wissenseinheiten, die unabhängig von ihrem jeweiligen Inhalt verarbeitet werden, zum Beispiel als Einträge in einem mentalen Lexikon. Heutzutage wird angenommen, dass konzeptuelles Wissen aus Netzwerken verschiedener Wissenseinheiten besteht (vgl. auch diSessa & Sherin, 1998; Medin & Rips, 2005). Diese Netzwerke werden aktiv durch den Lernenden konstruiert, indem er neu zu speicherndes Wissen mit seinem Vorwissen verknüpft. Wegen der dabei ablaufenden Schlussfolgerungs- und Elaborationsprozesse können Einheiten konzeptuellen Wissens auch als subjektive Theorien über einen Inhaltsbereich charakterisiert werden.

Diese neueren kognitiven Theorien der Repräsentation konzeptuellen Wissens haben sich produktiv auf die Pädagogische Psychologie ausgewirkt. Zu den bekanntesten Beispielen zählt die Untersuchung von Vosniadou und Brewer (1992) zum *conceptual change*: Sie zeigt, dass Kinder im Grundschulalter das Konzept der Erde als Kugel nur schwer erwerben, wenn man es ihnen isoliert beizubringen versucht, weil das Konzept nicht mit den subjektiven Theorien der Kinder über die Schwerkraft sowie über die Beziehungen zwischen oben und unten vereinbar ist. Die Kinder verstehen nicht, warum man nicht über Kopf steht und von der Erde fällt, wenn man sich auf der unteren Hälfte der Kugel befindet. Vosniadou und Brewer zeichneten durch qualitative Untersuchungen nach, wie der Versuch der Kinder, das neu zu Lernende mit ihrer Denk- und Erfahrungswelt in

Einklang zu setzen zu spezifischen Misskonzepten über die Form der Erde führte und zeigen alternative Instruktionsansätze auf. Dies wird in Unterkapitel 5.4 näher beschrieben, wo der Erwerb konzeptuellen Wissens eingehender betrachtet wird.

4.4.2 Repräsentationen prozeduralen Wissens

Die zur Zeit einflussreichsten Theorien des Erwerbs prozeduralen Wissens (vgl. Johnson, 2003) sind die ACT-R-Theorie von John R. Anderson, die Instance-Theorie von Gordon D. Logan und die psychometrische Theorie des Fähigkeitserwerbs von Phillip L. Ackermann.

Die ACT-Theorie (Anderson, 1976, 1983, 1990, 1993; Anderson & Lebiere, 1998) basiert auf der Unterscheidung von deklarativem und prozeduralem Wissen, ist also eine dual components-Theorie im oben besprochenen Sinn (vgl. Unterkapitel 4.1). Während sie deklaratives Wissen als assoziationsistisches Netzwerk aus Chunks mit propositionalem, numerischem oder piktorialem Inhalt charakterisiert, beschreibt sie prozedurales Wissen als Ansammlung von unverbunden nebeneinander stehenden Produktionsregeln. Produktionsregeln (vgl. auch Newell, 1973; Newell & Simon, 1972) bestehen aus einem sog. Wenn-Teil und einem sog. Dann-Teil. Der Dann-Teil bezieht sich auf einen mentalen oder motorischen Operator, der Wenn-Teil nennt die Bedingungen unter denen er angewandt werden kann. In ACT besitzt jede Produktion einen *Stärkeparameter*, der immer dann erhöht wird, wenn eine Produktion erfolgreich angewandt wurde, und niedriger gesetzt, wenn sie erfolglos benutzt wurde (*production tuning*). Produktionen feuern immer dann, d.h. sie werden immer dann ausgeführt, wenn der Inhalt des Arbeitsgedächtnisses den Wenn-Teil der Produktion erfüllen, was laut ACT-R durch assoziationsistische pattern matching-Prozesse im Sinne von Rumelhart und McClelland (1982) geschieht. Weil Handlungsziele ebenso wie aktuelle Wahrnehmungsinhalte und Problemrepräsentationen im Arbeitsgedächtnis und im Bedingungsteil von Produktionen vorkommen können, greifen plangesteuerte top-down-Prozesse und datengetriebene bottom-up-Prozesse bei der Produktionsauswahl ineinander. Erfüllt der Arbeitsgedächtnisinhalt die Bedingungsteile mehrerer Produktionsregeln zugleich, entscheidet unter anderem die Stärke der entsprechenden Produktionen darüber, welche ausgewählt wird. Da die Stärke mit jeder erfolgreichen Ausführung einer Produktion steigt, erklärt dies die Festigung von prozeduralem Wissen durch Übung und Feedback.

Werden mehrere Produktionen häufig in derselben Reihenfolge nacheinander ausgeführt, so werden sie automatisch zu einer Gesamtproduktion verschmolzen (*production composition*). Den Prinzipien des Chunking folgend belegt die Gesamtproduktion weniger Arbeitsgedächtniskapazität als die Einzelproduktionen.

Außerdem setzt sie nur einen Patternmachtingprozess und nicht wie die Einzelproduktionen eine Serie von Matchings voraus und kann daher schneller ausgeführt werden als die Einzelproduktionen. Die Kosten dafür liegen auf der Hand: Die Gesamtproduktion ist wesentlich unflexibler als die frei miteinander kombinierbaren Einzelproduktionen. Weil sie die Bedingungsteile vieler Einzelproduktionen enthält, müssen viele Bedingungen erfüllt sein, damit die Gesamtproduktion angewendet werden kann.

Die Kodierung prozeduralen Wissens bezeichnet Anderson auch als *Kompilierung*. Dabei wird das Wissen mental in einem Zustand gespeichert, der eine besonders schnelle und effiziente Ausführbarkeit ermöglicht. Wissen in diesem Zustand lässt jedoch nicht oder kaum introspektiv einsehen, bewusst transformieren und verbalisieren. Der Prozess der Wissenskompilierung ist eng mit dem der *Automatisierung* (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977) verbunden.

Andersons Theorie erklärt also durch die Annahme von miteinander verschmelzbaren Produktionsregeln unter anderem, warum prozedurales Wissen einerseits schnell, effizient und zielgerichtet ausgeführt werden kann, andererseits jedoch oft implizit und unflexibel ist (vgl. auch die Darstellung der ACT-Theorie in Unterkapitel 5.2).

Einen alternativen Erklärungsansatz für den Erwerb prozeduralen automatisierten Wissens gibt die *instance theory* von Logan (1988; 1992; Logan, Taylor, & Etherton, 1999). Sie basiert auf der Annahme, dass jede bearbeitete Aufgabe im Langzeitgedächtnis gespeichert wird – bei der Erstbegegnung durch Anlegen einer neuen Gedächtnisspur, bei weiteren Begegnungen durch Stärkung dieser Spur. Bei der Lösung neuer Aufgaben greifen Menschen anfangs auf bewusste Rechenregeln zurück. Sobald ihnen jedoch eine Aufgabe präsentiert wird, für die schon eine Gedächtnisspur existiert, rufen sie die richtige Lösung direkt aus dem Langzeitgedächtnis ab. Logan kann mit diesen Annahmen einige Eigenschaften prozeduralen Wissens mit weniger Grundannahmen als die ACT-Theorie benötigt erklären. So erklären sowohl die ACT-Theorie als auch die *instance theory* das *power law of practice* (Newell & Rosenbloom, 1981), also das exponentielle Absinken von Lösungszeiten in Abhängigkeit vom Ausmaß der Übung. Rickard (1997) sowie Anderson und Fincham (1994) konnten jedoch durch gezielte Variationen von Trainingsaufgaben Lerneffekte erzielen, die nicht oder nur bei komplexen ad hoc-Annahmen mit der *instance theory* zu vereinbaren sind.

Die dritte bekannte Theorie des Fähigkeitserwerbs ist die psychometrische Theorie von Ackermann (1988; 1989). Da sie keine Annahmen zur mentalen Repräsentation

prozeduralen Wissens beinhaltet, sondern lediglich spezifiziert, wie wichtige psychometrische Konstrukte den Fähigkeitserwerb determinieren, wird sie in Abschnitt 4.6 beschrieben.

Johnson (2003) weist in seinem Überblicksartikel darauf hin, dass prozedurales Lernen auch erfolgreich durch computersimulierte neuronale Netze modelliert worden ist. Der Anwendungsbereich jedes Netzes war dabei aber nur klein, also zum Beispiel gebunden an eine spezifische Aufgabenart, Aufgabenoberfläche und einen bestimmten Wertebereich, und wurde der Flexibilität menschlicher prozeduraler Kompetenz nicht gerecht.

Als Fazit lässt sich festhalten, dass Produktionensysteme, Gedächtnisspuren mit Aufgabenlösungen und neuronale Netzwerke als mentale Repräsentationen von prozeduralem Wissen einen Teil der vorliegenden empirischen Evidenz erklären können, wobei Produktionensysteme bislang am besten untersucht sind und die größte explanatorische Kraft besitzen.

4.4.3 Beziehungen zwischen mentaler Repräsentationsform und Inhalt

Sun und Zhang (2003) haben darauf hingewiesen, dass bei Beschreibungen von Unterscheidungen zweier Wissensarten oft Annahmen über einen Zusammenhang zwischen der mentalen Wissensrepräsentation und dem Inhalt des Wissen gemacht werden, die nicht notwendig und nur schlecht empirisch belegt sind (vgl. auch Anderson, 1983; Hiebert & Wearne, 1996).

Häufig wird davon ausgegangen, dass Wissen über Fakten oder konzeptuelle Zusammenhänge eher in Netzwerken und Wissen über die Lösung von Problemen eher in Produktionsregeln gespeichert wird. Davon kann jedoch nicht a priori ausgegangen werden. Zum Beispiel könnte jemand die auf ein Konzept ausgerichtete Aufgabe „Bitte erkläre, was Dezimalbrüche sind.“ durch das Aufsagen einer Definition lösen. Diese Definition könnte den Inhalt seines semantischen Netzwerkes widerspiegeln. Genauso gut könnte sie jedoch auch ohne Verständnis des Inhalts auswendig gelernt und in einer Produktionsregel gespeichert worden sein.

Ebenso kann jemand bei der Beantwortung der auf eine Problemlösung ausgerichteten Frage „Wie addiert man zwei Brüche mit ungleichen Nennern?“ einerseits auf Produktionen, die die Problemlösung kodieren, zurückgreifen, andererseits auch auf den Inhalt seines semantischen Netzwerkes, der spezifiziert, warum man nicht einfach Zähler zu Zähler und Nenner zu Nenner addieren darf.

Während Repräsentationsform und Inhalt nicht a priori miteinander assoziiert sein müssen, besteht doch die Möglichkeit, dass sie es empirisch gesehen sind. Diese Frage kann bisher als offen angesehen werden.

4.5 Operationalisierungen der Wissensarten

4.5.1 Die Messung konzeptuellen Wissens

Dass Aussagen über die mentale Repräsentation von konzeptuellem und prozeduralem Wissen bisher einen eher spekulativen oder postulierenden Charakter haben, liegt vor allem auch an dem schon in den Unterabschnitten 2.4.3 und 3.4 beschriebenen Problem der Operationalisierung der Wissensarten. Im aktuellen Unterkapitel werden einige der wichtigsten Maße dargestellt, welche bisher in Studien zu konzeptuellem und prozeduralem Wissen benutzt wurden.

Bemerkungen zu diesem Thema, die einige der späteren Einsichten vorwegnehmen, machen Greeno et al. (1984). Wie auf Seite 19 und folgenden beschrieben, wird in ihrem Artikel Gelmans Unterscheidung zwischen den principles und skills des Zählens in den übergeordneten theoretischen Rahmen der dual components theories of cognition eingeordnet. Über die Messung konzeptuellen Wissens schreiben die Autoren dort:

The case for understanding of principles is strongest if a child is required to generate a new procedure or a modification of a known procedure, and the procedure that is formed is consistent with the principles. [...]

Evidence for understanding also is obtained if a child can evaluate performance as correct or incorrect with respect to a principle, as when children are shown examples of counting and can identify errors that are made, or when they spontaneously correct their own errors. Evaluation of examples is often used to test whether someone has acquired a categorical concept (e.g., Bruner, Goodnow, & Austin, 1956) [...]. A third form of evidence is provided by performance that is systematically consistent with a principle, especially when it occurs in a wide variety of contexts so that consistency would be unlikely in the absence of knowledge of the principle.

Evidence for understanding of principles always is problematic to some degree. Any single piece of evidence can be explained by without recourse to a hypothesis of understanding: performance consistent with the principle could be learned by rote, evaluation could involve simple comparison of example performance and covert performance of a rote procedure, and novel procedures could be generated by trial and error. Even so, a combination of evidence of these various kinds can constitute a compelling argument that principles are understood significantly, i.e. are understood as constraints on performance, even if the understanding is implicit, i.e., they cannot be stated by the child. (Greeno et al., 1984, S. 105f)

Greeno und Kollegen schlagen also drei unterschiedliche Maße konzeptuellen Wissens vor, schränken jedoch ein, dass es stets Alternativerklärungen für die erfolgreiche Lösung der jeweiligen Aufgabenart gibt. In Übereinstimmung mit den in Abschnitt 4.3.6 vorgestellten Überlegungen zur Unterscheidbarkeit der Wissensarten kommen sie zu dem Schluss, dass die überzeugendste Evidenz für das Vorhandensein konzeptuellen Wissens qualitativ vergleichbare Effekte auf mehreren unterschiedlichen Maßen wären. Im Folgenden wird darum ein kurzer Überblick über die verschiedenen Maße gegeben, die bisher zur Erhebung konzeptuellen Wissens benutzt wurden: verbale Befragung, concept maps, Flexibilität der Problemlösung, Nutzung von Analogien, Evaluationsaufgaben, Übersetzung in andere Zeichensysteme sowie Sortier- und Vergleichsaufgaben.

Verbale Befragung

Vosniadou (1994) beschreibt mündliche Interviews als Methode der Wahl, um konzeptuelles Wissen zu erheben. Grund dafür ist, dass sie in ihrem Ansatz des conceptual change konzeptuelles Wissen als mentale Modelle, zum Beispiel der Erde, konzeptualisiert, die in subjektive Theorien über einen Gegenstandsbereich, zum Beispiel Raumzeit und Schwerkraft, eingebettet sind. Dieses Informationsgefüge lässt sich nur durch möglichst offene Transferfragen erfassen, da Studienteilnehmer bei einfacheren Fragen dazu neigen, in Trainingskontexten erworbene Formulierungen auch dann zu replizieren, wenn sie ihren Inhalt nicht verstanden haben. Vosniadou schlägt vor, dass die Interviews nicht auf verbale Daten beschränkt sein sollten, und integrierte beispielsweise die Aufforderung bestimmte Sachverhalte in einer Zeichnung darzustellen in ihre Datenerhebung.

In der Forschung zu konzeptuellem und prozeduralem Wissen wurden Interviewfragen lediglich von Rittle-Johnson und Alibali (1999) benutzt. In ihrer Studie mit Äquivalenzproblemen fragten sie beispielsweise: „Define what it means for two sets of objects to be equal” (Rittle-Johnson & Alibali, 1999, S. 178).

Dass Interviews in der Forschung zu konzeptuellem und prozeduralem Wissen nur selten verwendet werden, liegt unter anderem daran, dass der Aufwand für die Durchführung, die qualitative Auswertung und die quantitative Kodierung der Antworten groß ist. Durch die direkte Interaktion zwischen Interviewer und Interviewtem sowie der notwendigen Kodierung der Antworten ist die Objektivität niedriger als beispielsweise bei Fragebogenitems. Der Hauptnachteil von Interviews besteht aber wohl darin, dass konzeptuelles Wissen auch implizit sein kann. Vor allem Goldin-Meadows Arbeiten zum *gesture-speech mismatch* haben eindrucksvoll gezeigt, dass Kinder in Phasen, in denen sie

sich neue Konzepte erschließen, diese durch Gesten schon anzeigen können, während sie sprachlich dazu noch nicht in der Lage sind, zum Beispiel einfach weil ihnen die notwendige Terminologie nicht geläufig ist (Church & Goldin-Meadow, 1986; Goldin-Meadow, Alibali, & Church, 1993). Wie oben aufgezeigt, beschrieb schon Gelman die principles des Zählens bei Kindern im Vorschulalter als implizites konzeptuelles Wissen.

Das Problem wird reduziert, wenn man die Interviewten auch Zeichnungen oder Diagramme anfertigen lässt. Regt man sie zu solchen Tätigkeiten an, stellt sich jedoch die Frage, inwieweit man ihr konzeptuelles Wissen unabhängig von ihrem prozeduralen operationalisiert.

Trotz dieser Einschränkungen gehören sorgfältig gestaltete Interviews sicherlich zu den validesten Untersuchungsinstrumenten, vor allem, wenn man ein Netzwerk sich auf einander beziehender Fragen verwendet, so dass die Konsistenz des Antwortverhaltens überprüft werden kann (Vosniadou, 1994).

Concept maps

Da konzeptuelles Wissen hypothetisch in Netzwerken gespeichert wird, liegt es nahe, es auf Weisen zu operationalisieren, die ihrerseits zu Netzwerkdarstellungen führen. Die einfachste Möglichkeit besteht darin, Probanden selbstständig ein konzeptuelles Netzwerk zeichnen oder ein teilweise vorgegebenes vervollständigen zu lassen (Hardy & Stadelhofer, unter Begutachtung; Lavigne, 2005; Ruiz-Primo & Shavelson, 1996). Aufwändigere Möglichkeiten, die jedoch auch Hinweise auf die Stärke der Verbindungen zwischen den Knoten des Netzwerks liefern, sind Wortassoziationsaufgaben und Ähnlichkeitsratings, die mittels geeigneter Auswertungsalgorithmen in Netzwerkdarstellungen transformiert werden können (Überblick bei Jonassen, Beissner, & Yacci, 1993).

Concept maps sind qualitativ zwar aufschlussreich, ihr Inhalt ist jedoch ausgesprochen schwierig valide zu quantifizieren und einer statistischen Auswertung zugänglich zu machen. Sie wurden in der Forschung zu konzeptuellem und prozeduralem Wissen bisher nicht benutzt.

Flexibilität der Problemlösung

Diese Operationalisierungsart wurde von Blöte et al. (2001) benutzt. Sie ließen Probanden mehrere Aufgaben jeweils zweimal lösen und forderten sie auf, beide Male unterschiedliche Problemlösestrategien zu benutzen. Die Anzahl der Aufgaben, die die Probanden tatsächlich durch zwei unterschiedliche Prozeduren lösen konnten, wurde dann

als Maß der prozeduralen Flexibilität betrachtet. Da prozedurales Wissen eher unflexibel, konzeptuelles Wissen jedoch eher flexibel ist, interpretierten Blöte et al. die Flexibilität als Maß des konzeptuellen Wissens.

Prozedurale Flexibilität ist ein wichtiges Lernziel. Wie Blöte et al. selber einräumen, ist prozedurale Flexibilität jedoch kein wirklich valides Maß für konzeptuelles Wissen (siehe Blöte et al., 2001, S. 628). Man kann sich leicht Probanden vorstellen, die zwei Prozeduren zur Lösung derselben Aufgabe gelernt haben, ohne das dazugehörige konzeptuelle Verständnis zu besitzen.

Nutzung von Analogien zwischen Aufgaben

Diese von Canobi et al. (1998) benutzte Aufgabenart wurde bereits auf Seite 27 beschrieben. Kindern wurden zwei Aufgaben gleichzeitig auf einem Bildschirm präsentiert. Je häufiger die Kinder korrekt erkannten, ob sich unter Anwendung geeigneter Umformungsregeln, beispielsweise dem Kommutativitätsgesetz, die Lösung der ersten Aufgabe auch die Lösung der zweiten Aufgabe ist, umso mehr konzeptuelles Wissen über die Umformungsregeln wurde ihnen zugesprochen. Unklar ist dabei allerdings, ob die Kinder nicht auch prozedurales Wissen anwenden können um zu sehen, ob die erste Aufgabe in die zweite transformiert werden kann.

Evaluationsaufgaben

Bei dieser Operationalisierung werden den Studienteilnehmern Lösungswege für Problemlöseaufgaben präsentiert, die sie auf ihre Richtigkeit oder Effektivität hin evaluieren sollen. Dies geschieht gewöhnlich auf einer Ratingskala. Verwendet wurden Evaluationsaufgaben unter anderem von Canobi, Reeve und Pattison (1998), Gelman und Meck (1983), Rittle-Johnson und Alibali (1999), sowie Siegler und Crowley (1994). Evaluationsaufgaben haben gegenüber Interviewfragen den Vorteil, dass der aktive Wortschatz der Befragten eine geringere Rolle spielt, weil sie keine sprachliche Antwort generieren, sondern bloß präsentierte Strategien nachvollziehen und bewerten müssen. Wegen des geschlossenen Antwortformates ist die Ratewahrscheinlichkeit bei ihnen jedoch höher.

Übersetzung in andere Zeichensysteme

Diese Operationalisierung wurde von Byrnes und Wasik (1991) verwendet, um das konzeptuelle Verständnis von Brüchen bei Kindern zu erheben. Den Versuchspersonen wurden zum Beispiel sechs Dreiecke gezeigt, von denen fünf schraffiert waren. Aus vier

ihnen in Bruchstrichnotation präsentierten Brüchen, z.B. $\frac{1}{4}$, sollten sie dann denjenigen auswählen, der dem numerischen Verhältnis der Anzahl schraffierter Dreiecke zu der Gesamtanzahl der Dreiecke entspricht. Die zugrunde liegende Annahme ist, dass man die Übersetzung von der einen in die andere Notation nur vornehmen kann, wenn man das inhaltliche Konzept, auf das sich beide Notationen beziehen, verstanden hat, so wie man auch einen Text nur dann in eine andere Sprache übersetzen kann, wenn man seinen Inhalt verstanden hat.

Leichte Abwandlungen dieser Aufgabenart, die ebenfalls von Byrnes und Wasik (1991) verwendet wurden, bestehen darin, die Studienteilnehmer das ursprüngliche Stimulus-Set nicht in Formelnotation, sondern in ein anderes Set teilweise schraffierter geometrischer Objekte übersetzen zu lassen, bzw. darin, den Teilnehmern einen geschriebenen Text vorzulegen, in dem die Aufteilung von Objekten, beispielsweise das Verteilen von Tortenstücken, beschrieben wird und sie dann nach der relativen Größe einzelner Anteile in Formelnotation zu befragen.

Der Nachteil dieser Operationalisierung besteht darin, dass das inhaltliche konzeptuelle Verständnis hier konfundiert mit dem Verständnis der jeweiligen Notationen erhoben wird.

Sortier- und Vergleichsaufgaben

Dieser Aufgabentyp wurde ebenfalls in der Studie von Byrnes und Wasik (1991) zum Verständnis von Brüchen verwendet und besteht darin, einige vorgegebene Brüche entsprechend ihrer numerischen Größe in eine Rangreihenfolge zu bringen. Diese Aufgabenart erfasst also das Verständnis der ordinalen Relationen zwischen Zahlen.

Im Rahmen von kognitionspsychologischen Studien, in denen Konzeptwissen oft als Kategoriewissen verstanden wird (s.o.), werden Versuchspersonen auch häufig gebeten, Objekte oder Bilder von Objekten anhand ihrer Eigenschaften in verschiedene Kategorien einzusortieren (vgl. Lavigne, 2005).

4.5.2 Die Messung prozeduralen Wissens

Jon R. Star (2000) stellte die Operationalisierungen für konzeptuelles und prozedurales Wissen einiger wichtiger Studien einander in einer Tabelle gegenüber. Während in der Spalte für konzeptuelles Wissen eine ganze Reihe unterschiedlicher Methoden aufgeführt wurden, stand in der Spalte für prozedurales Wissen pro Studie immer nur eine einzige Methode: Kinder sollten Routineprobleme lösen und es wurde der Prozentsatz korrekter

Antworten berechnet. Anhand dieser Beobachtung kritisiert Star, dass die Vielschichtigkeit und Komplexität prozeduralen Wissens in der Forschung häufig unterschätzt wird. Prozedurales Wissen wird oft als eindimensional und oberflächlich beschrieben, was Star für unangemessen hält.

Während diese Kritik auf die pädagogisch-psychologische Forschung zu konzeptuellem und prozeduralem Wissen sicherlich zutrifft, ignoriert sie die vielfältigen Erhebungsmethoden für prozedurales Wissen, die im Rahmen kognitionspsychologischer Experimente entwickelt wurden. Interessante Überblicke hierüber geben Jonides, Naveh-Benjamin und Palmer (1985) sowie Royer, Cisero und Carlo (1993). Vor allem Letztere verdeutlichen die potenzielle Mehrdimensionalität und Komplexität prozeduralen Wissens eindringlich, indem sie eine differenzierte Klassifikation von verschiedenen Operationalisierungsmöglichkeiten aufstellen.

Im Folgenden sollen nur vier der wichtigsten Operationalisierungen prozeduralen Wissens vorgestellt werden. Da in der Forschung zu konzeptuellem und prozeduralem Wissen vorwiegend eine einzige Operationalisierung (die Problemlöseleistung) benutzt wird, sind die drei anderen Maße (Problemlösedauer, Zugriffsasymmetrie und dual task-Kosten) kognitionspsychologischen Studien zum Erwerb prozeduralen Wissens entnommen. Bei den vier Maßen handelt es sich um die Problemlöseleistung, die Problemlösedauer, die Zugriffsasymmetrie und die dual task-Kosten.

Problemlöseleistung bei Routineproblemen

Da prozedurales Wissen die Voraussetzung für das erfolgreiche Lösen von Problemen ist, liegt es nahe, prozedurales Wissen über die Problemlöseleistung von Probanden zu operationalisieren. Dieser Weg stellt, wie oben beschrieben, in der Forschung zu konzeptuellem und prozeduralem Wissen den Standard dar, und wurde beispielsweise von Byrnes und Wasik (1991), Canobi et al. (1998), sowie Rittle-Johnson und Alibali (1999) gewählt. Letztere untersuchten prozedurales Wissen über mathematische Äquivalenzprobleme und legten ihren Vpn daher solche Probleme, wie „ $5 + 4 + 3 = 5 + __$ “, vor, wobei die Vpn jeweils die fehlende Zahl ergänzen sollten.

Wie in Unterkapitel 5.2 besprochen wird, könnte es möglich sein, dass man prozedurales Wissen ad hoc aus konzeptuellem ableiten kann, um ein neues Problem zu lösen. In der ACT-Theorie des Fähigkeitserwerbs wird dieses Phänomen als erste von drei Phasen des Fähigkeitserwerbs aufgefasst und als *deklarative Phase* bezeichnet. In der Praxis kann man das Phänomen beobachten, wenn Schüler vom Lehrer erst Instruktionen darüber erhalten, wie man eine neue Aufgabenart, z.B. quadratische Gleichungen, lösen

kann, und sie die Instruktionen anschließend erfolgreich auf Übungsaufgaben anwenden. Die Problemlöseleistung gibt also nur Aufschluss darüber, ob mindestens eine der beiden Wissensarten vorhanden ist, aber nicht welche es ist.

In kognitiven Experimenten wird die Problemlöseleistung daher gewöhnlich nicht als Maß prozeduralen Wissens verwandt.

Wie oben beschrieben benutzen Rittle-Johnson et al. (2001) das Maß zwar, setzen dabei aber nur Routineprobleme ein. Ihr Annahme dabei ist folgende: Falls die Vpn die Probleme bisher schon lösen konnten, ist unwahrscheinlich, dass sie ad hoc neue Prozeduren aus ihrem konzeptuellen Wissen ableiten, sondern sie werden sie vermutlich durch Anwendung ihres prozeduralen Wissens lösen. Falls die Vpn diese Probleme bisher nicht lösen konnten, so ist unwahrscheinlich, dass sie nun ad hoc neues prozedurales Wissen generieren.

Wie oben ausführlich besprochen, ist diese Begründung problematisch. Für unterschiedliche Personen und für einzelne Personen zu unterschiedlichen Zeitpunkten können ganz unterschiedliche Aufgaben Routineaufgaben darstellen. Insbesondere in Prätest-Posttest-Designs tritt dieses Problem auf. Außerdem könnten Personen wohl auch konzeptuelles Wissen benutzen, um Routineaufgaben zu lösen.

Problemlösedauer

Ein besseres Maß für die Vorhandenheit von prozeduralem Wissen könnte die Zeit sein, die zur Lösung eines Problems benötigt wird. Weil prozedurales Wissen aus automatisiert ablaufenden Handlungssequenzen besteht, kann es – sofern vorhanden – schnell und effizient zur Anwendung gebracht werden (z.B. Logan, 1985; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). Problemlöser, die ihre Prozeduren hingegen erst durch Elaboration und Transformation ihres konzeptuellen Wissens neu generieren müssen, brauchen relativ dazu gesehen wesentlich mehr Zeit für die Problemlösung (z.B. Anderson, 1982).

Die Problemlösedauer sinkt als Funktion der Übung exponentiell ab. Diese Gesetzmäßigkeit wurde in verschiedenen Bereichen der Psychologie unzählige Male repliziert (siehe Newell & Rosenbloom, 1981; Palmeri, 1999; Ritter & Schooler, 2001). Allerdings gilt dieses *power law of practice* nur, wenn Personen unter kontrollierten Bedingungen stets dieselbe Aufgabe mittels derselben Strategie lösen. Schon wenn sie stattdessen einen Mix aus unterschiedlichen Strategien benutzen, gilt es nicht mehr für das Gesamtverhalten (Rickard, 1997). Auch Änderungen des konzeptuellen Wissens können die Problemlösegeschwindigkeit beeinflussen, weil sie zu jeweils anderen Inferenz- und

Planungsprozessen führen können, die unterschiedlich viel Zeit benötigen. Werden Informationen, die anfangs stets mittels Inferenzprozessen erschlossen werden mussten, explizit im Gedächtnis verankert, kann dies die Problemlösedauer senken. Umgekehrt kann mehr Wissen über ein Problem komplexere Planungsprozesse ermöglichen (Gunzelmann & Anderson, 2003) und so die Problemlösedauer verlängern.

Betrachtet man die Zeiten über viele Problemlösungen hinweg, so deuten kontinuierliche, der exponentiell abfallenden Form folgende Zeitveränderungen eher auf einen Zuwachs an prozeduralem Wissen hin, während plötzliche starke Veränderungen in der Zeit eher eine Umstrukturierung deklarativen Wissens vermuten lassen (Anderson, Fincham, & Douglass, 1999).

Die Problemlösedauer wurde in unzähligen kognitiven Experimenten, jedoch kaum in pädagogisch-psychologischen Studien zu konzeptuellem und prozeduralem Wissen benutzt. Ausnahmen bilden die empirische Studie von Canobi et al. (1998) und die Forschung zum *strategy change model* von Robert S. Siegler (z.B. Shrager & Siegler, 1998; Torbeyns, Verschaffel, & Ghesquiére, 2005).

Zugriffsasymmetrie

Im Rahmen der Untersuchung der ACT-Theorie wurde ein weiterer interessanter Ansatz zur Erhebung prozeduralen Wissens entwickelt. Da Konzeptwissen mittels symmetrischer Relationen vernetzt ist, sollte es für die Abrufzeit nur einen kleinen Unterschied machen, ob man eine Information A vorgibt (z.B. eine Position auf einem Zahlenstrahl) und nach einer Information B (z.B. dem dargestellten Zahlenwert) fragt oder ob man Information B vorgibt und nach Information A fragt. Ist dasselbe Wissen jedoch in einer Produktionsregel statt in einem semantischen Netzwerk abgespeichert, so sind asymmetrische Effekte zu erwarten, weil, wie oben beschrieben, Produktionsregeln aufgrund ihres Wenn- und ihres Dann-Teils asymmetrisch funktionieren. Eine Produktionsregel könnte lauten: „Wenn Information A dargeboten wird, dann äußere Information B.“ Vpn, die über eine solche Produktion verfügen, müssten schneller reagieren können, wenn man ihnen Information A vorgibt und sie nach B fragt, als wenn man ihnen Information B vorgibt und sie nach A fragt. Dieser von Anderson (1983, S. 39) postulierte Effekt wurde inzwischen empirisch nachgewiesen und mehrfach repliziert (Anderson & Fincham, 1994; McKendree & Anderson, 1987; Pennington, Nicolich, & Rahm, 1995; Rabinowitz & Goldberg, 1995).

Empirisch wurde gefunden, dass auch die Lösungszeiten für den nicht geübten Aufgabentyp durch die Übung reduziert werden (z.B. Anderson & Fincham, 1994), jedoch

schwächer als in der geübten Bedingung. Dafür gibt es bislang jedoch keine gesicherte Erklärung.

Dieses Maß wurde in der Forschung zu konzeptuellem und prozeduralem Wissen noch nicht benutzt, ebenso wie das folgende, die dual task-Kosten.

Dual task-Kosten

Wenn eine Person eine Aufgabe einzeln ausführt, ist sie dabei gewöhnlich besser, als wenn sie gleichzeitig noch eine zweite Aufgabe ausführt. Dieses Phänomen wird auch als *dual task interference* bezeichnet. Mehrere Studien zeigten, dass diese dual task interference umso kleiner wird, je mehr Übung man mit den Aufgaben hat (z.B. Britton & Tesser, 1982; Levy & Pashler, 2001; Schumacher, Seymour, Glass, Kieras, & Meyer, 2001). Diverse Autoren, wie Ruthruff, Johnston und Selst (2001) oder Schneider (1985), führen dies auf die Automatisierung der Aufgabenbearbeitungsprozesse durch die Übung zurück.

Diese Erklärung steht im Einklang mit den theoretischen Konzeptionen von prozeduralem Wissen. Wie beispielsweise Anderson (1983, S. 231ff) beschreibt, bindet die Problemlösung durch Anwendung von deklarativem Wissen viele kognitive Ressourcen, weil dabei seriell ablaufende Prozesse der bewussten Wissenstransformation im Arbeitsgedächtnis ablaufen. Vorhandenes prozedurales Wissen kann hingegen ohne großen Ressourcenaufwand zur Anwendung gebracht werden, weil es aus automatisiert ablaufenden Verhaltensprogrammen besteht. Wird eine Aufgabe also auf Grundlage von prozeduralem Wissen gelöst, stehen dabei mehr Ressourcen zur gleichzeitigen Bearbeitung einer zweiten Aufgabe zur Verfügung, als wenn die Aufgabe durch konzeptuelles Wissen gelöst wird.

Weil die Forschungen zum *dual task*-Paradigma und zur Wissensprozeduralisierung zwei getrennte Forschungsstränge darstellen, wurde die dual task interference bisher noch nicht explizit als Maß prozeduralen Wissens herangezogen. Während die dual task interference von mehreren Autoren auf Automatisierungsprozesse und Übungseffekte zurückgeführt wird, wird dabei der Begriff des prozeduralen Wissens gewöhnlich nicht erwähnt, so dass die genauen Beziehungen zwischen den beiden Paradigmen unscharf bleiben (vgl. Royer et al., 1993).

Bei der methodischen Umsetzung des dual task-Paradigmas sind einige wichtige Punkte zu beachten. So gibt es verschiedene Typen von Zweitaufgaben, wie probe tasks, concurrent memory load tasks, motor production tasks und perceptual judgement tasks. Diese lassen sich teilweise für verschiedene Sinnesmodalitäten implementieren. So können perceptual judgement tasks die Beurteilung visueller, auditiver oder taktiler Reize

erfordern. Wird die Zweitaufgabe zur Untersuchung kognitiver Prozesse benutzt, sollte sie so gewählt sein, dass Erst- und Zweitaufgabe nicht schon auf der Ebene der Aufgabenwahrnehmung oder der Antwortäußerung interagieren, weil dies die kognitiven Effekte überdecken kann. Damit auf der kognitiven Ebene überhaupt dual task-Kosten entstehen, ist die Zweitaufgabe außerdem so auszuwählen, dass Erst- und Zweitaufgabe mit hoher Wahrscheinlichkeit dieselben kognitiven Ressourcen nutzen, z.B. beide die auditory loop oder beide den visuospatial sketchpad (z.B. Meyer & Kieras, 1997; Pashler, 1994).

Bei der Aufgabenauswahl ist ferner die ökologische Validität zu beachten. Vor allem dual task-Aufgaben im Rahmen des *psychological refractory period*-Paradigmas zur Erforschung der *central bottleneck*-Hypothese sind als ökologisch unvalide kritisiert worden, weil dabei einfache Aufgaben im Abstand von Sekundenbruchteilen versetzt präsentiert werden und möglichst schnell zu lösen sind, was im Alltag nur selten geschieht (vgl. Pashler, 1994). Andere Aufgabenkombinationen, wie zu schreiben, während man jemandem zuhört, oder schlussfolgernd zu denken, während man geht, sind jedoch normale Alltagstätigkeiten, deren Untersuchung durchaus ökologisch valide und daher generalisierbare Befunde liefern kann (z.B. Lindenberger, Marsiske, & Baltes, 2000).

Bei der Datenauswertung ist zu beachten, dass die dual task-Kosten sich einzeln oder kombiniert an vier verschiedenen Stellen zeigen können: in der Bearbeitungszeit der Erstaufgabe, der Lösungshäufigkeit der Erstaufgabe, der Bearbeitungszeit der Zweitaufgabe sowie der Lösungsrate der Zweitaufgabe. Zu dem aus anderen Paradigmen bekannten Problem des speed/accuracy trade-offs kommt bei dual task-Aufgaben also noch der mögliche trade-off zwischen Erst- und Zweitaufgabe hinzu. An welcher Stelle sich die dual task-Kosten zeigen werden, ist schwer vorherzusagen, weil dies, außer von den Instruktionen, auch von den spezifischen Lösungsstrategien der Probanden abhängt. Bei der Auswertung von dual task-Aufgaben sind daher sorgfältige post hoc-Analysen unabdingbar.

4.6 Beziehungen der Wissensarten zu anderen kognitiven Konstrukten

4.6.1 Intelligenz als Prädiktor des Wissenserwerbs

Über den Zusammenhang zwischen Wissen und allgemeiner Intelligenz schreibt Elsbeth Stern unter Verweis auf die *investment theory* von Cattell und darauf aufbauende aktuellere Ansätze: „Intelligenz beschreibt das Ausmaß, in dem es gelingt, die im jeweiligen kulturellen Kontext vorgefundenen Zeichensysteme zu erwerben und diese dann zur Aneignung und Anwendung von Wissen zu nutzen“ (Stern, 2001b, S. 200).

Während diese Aussage auf konzeptuelles Wissen, das ja gerade in intelligenten Verknüpfungen von Wissenseinheiten besteht, uneingeschränkt zutrifft, kann sie für prozedurales Wissen noch weiter ausdifferenziert werden. So besagt Ackermans (z.B. 1988; Ackerman & Cianciolo, 2000) psychometrische Theorie des Fähigkeitserwerbs, dass prozedurales Wissen in den verschiedenen Phasen des Fähigkeitserwerbs unterschiedlich stark von der Intelligenz abhängt. Genauer gesagt postuliert die Theorie in Übereinstimmung mit anderen, wie beispielsweise ACT-R, dass es drei Phasen des Fähigkeitserwerbs gibt. Ackerman weist durch empirische Untersuchungen nach, dass in den drei Phasen jeweils unterschiedliche Konstrukte unterschiedlich gute Prädiktoren des Fähigkeitsniveaus darstellen: In der initialen, der *deklarativen Phase*, sind die besten Prädiktoren Hintergrundwissen und Intelligenz, in der *assoziativen Phase* ist es vor allem *perceptual speed*, während die anderen Faktoren unwichtiger werden. In der dritten, der *autonomen Phase* sind psychomotorische Fähigkeiten die besten Prädiktoren.

Ackermans Theorie sagt also eine mit der Übung fortschreitende Entkopplung von Intelligenz und prozeduralem Kompetenzniveau vorher. Sie stimmt dabei überein mit neueren empirischen Befunden zur neuralen Effizienz von Probanden auf unterschiedlichen Expertiseniveaus (Grabner, Stern, & Neubauer, 2003). Es wurde gefunden, dass die Intelligenz die neurale Aktivierung nur bei niedrigem, nicht jedoch bei hohem Expertisegrad der Probanden beeinflusste. Stern (2001a) weist darauf hin, dass der Einfluss der Intelligenz auf den Lernerfolg grundsätzlich erheblich reduziert wird, wenn das bereichsspezifische Vorwissen mit in das Modell aufgenommen wird.

Insgesamt ist also zu vermuten, dass initial die Niveaus beider Wissensarten hoch mit der Intelligenz und Maßen des Vorwissens korrelieren. Die Korrelation zwischen konzeptuellem Wissen und Intelligenz sollte im Verlaufe der Übung auf hohem Niveau nur leicht absinken, während die zwischen prozeduralem Wissen und Intelligenz stark sinken sollte. Ein solches Muster könnte, würde es empirisch gefunden, auch als indirekte Evidenz für die Unterscheidung der beiden Wissensarten dienen.

4.6.2 Mentale Repräsentation von Zahlen

Untersucht man konzeptuelles und prozedurales Wissen im Bereich der Mathematik, so könnte man vermuten, dass neben der Intelligenz auch die mentale Repräsentation des wichtigsten Konstrukts der Schulmathematik eine Rolle spielen könnte: die mentale Repräsentation von Zahlen (vgl. Stern, Felbrich, & Schneider, zur Publikation angenommen).

Von zentraler Bedeutung ist hier das *triple code model* von Stanislas Dehaene (z.B. 1992; 1999). Ihm zufolge werden numerische Größen in drei verschiedenen Repräsentationsformaten gespeichert: dem visuell-arabischen Format („4“), dem auditiv-verbalen Format („vier“) und dem analogen Format, dem so genannten mentalen Zahlenstrahl. Während diese drei Repräsentationsformate beim Lösen mathematischer Probleme gewöhnlich eng ineinandergreifen, konnte Dehaene sie durch geschickte experimentelle Manipulationen isolieren und ihre jeweils spezifischen Eigenschaften nachweisen. Weitere Evidenz stammt aus klinischen Studien mit Läsionspatienten, bei denen dieses Repräsentationssystem teilweise beschädigt war (Überblick bei Dehaene, Dehaene-Lambertz, & Cohen, 1998).

Für die Existenz eines mentalen Zahlenstrahls sprechen vor allem der *distance effect* und der SNARC (*spatial-numerical association of response code*) *effect*. Ersterer zeigt sich, wenn Probanden mehrere nacheinander auf einem Bildschirm präsentierte Zahlen daraufhin beurteilen sollen, ob sie jeweils größer oder kleiner als ein Referenzwert sind. Die Reaktionszeiten der Probanden sind dabei umso kürzer, je weiter die jeweils zu beurteilende Zahl vom Referenzwert entfernt ist, wobei sie einer exponentiell abfallenden Funktion folgen. Dieser kontinuierlich verlaufende Effekt kann durch die anderen beiden diskret-symbolischen Repräsentationsformate nicht erklärt werden, sondern nur durch eine zusätzliche analoge Repräsentationsform (z.B. Dehaene, Dupoux, & Mehler, 1990).

Dass ein solcher mentaler Zahlenstrahl von links nach rechts verläuft, zeigt der SNARC effect. Probanden haben dabei Zahlen anhand vorgegebener Eigenschaften zu klassifizieren. Beispielsweise könnten sie aufgefordert sein, bei geraden Zahlen stets eine Taste am Computer zu drücken und bei ungeraden eine andere. Dehaene zeigte, dass Probanden unabhängig von der Aufgabe auf kleine Zahlen schneller mit der linken Hand und auf große Zahlen schneller mit der rechten reagieren. Dies lässt sich nur schwer durch diskret-symbolische Repräsentationsformate erklären, relativ zwanglos jedoch durch die Annahme der Links-Rechts-Orientierung eines mentalen Zahlenstrahls (Dehaene, 1993).

Dehaene und Akhavein (1995) konnten zeigen, dass bei den meisten Probanden im visuell-arabischen oder im auditiv-verbalen Format präsentierte Zahlen automatisiert in das analoge Format umgewandelt werden. Dieser Effekt war automatisiert in dem Sinne, dass er auch dann auftrat, wenn die semantische Größe der Zahl und somit die analoge Größenrepräsentation zur Lösung der Testaufgaben irrelevant war.

Es ist vorstellbar, dass die Fähigkeit, die semantische Größe numerischer Informationen aus verschiedenen Repräsentationsformaten automatisiert ableiten zu können, sich im Kontext des Mathematiklernens positiv auf den Erwerb konzeptuellen

Wissens auswirkt. Denn je stärker solche numerischen Prozesse automatisiert sind, umso weniger belasten sie das Arbeitsgedächtnis. Dadurch könnte einerseits der Erwerb konzeptuellen Wissens, das ja bewusste Elaborationen voraussetzt, gefördert werden. Zum anderen könnte durch eine Entlastung des Arbeitsgedächtnisses auch die Transformation von konzeptuellem in prozedurales Wissen und umgekehrt gefördert werden. Die Stärke von distance und SNARC effect müsste in diesem Fall positiv mit dem Ausmaß konzeptuellen Wissens korreliert sein und als Moderator auf etwaige Beziehungen zwischen den Wissensarten einwirken. Dies wurde bisher jedoch nicht überprüft.

4.7 Zusammenfassung

Theorien über konzeptuelles und prozedurales Wissen und viele ähnliche Paare von Wissensarten lassen sich unter die dual component theories of cognition subsumieren. Viele Begriffspaare für Wissensarten werden synonym verwendet, obwohl die empirischen Beziehungen zwischen den Paaren ungeklärt sind. Ein wichtiges Argument gegen eine Unterscheidung von Wissensarten ist ihre mangelnde Isolierbarkeit im Verhalten, ein wichtiges Argument für eine Unterscheidung sind neuropsychologische Befunde über distinkte Gedächtnissysteme, deren psychologisch-funktionale Charakteristika bisher jedoch ungeklärt sind. Auf theoretischer Ebene wird konzeptuelles Wissen als Netzwerk dem Bewusstsein zugänglicher konzeptueller Aussagen und Prinzipien gesehen, prozedurales Wissen als Sammlung automatisiert ausführbarer Produktionsregeln. Konzeptuelles Wissen wird gewöhnlich durch Aufgaben operationalisiert, die die flexible und reflektierte Anwendung von Konzepten auf neue Aufgabentypen verlangen. Prozedurales Wissen wird darüber gemessen, wie umfassend und effizient Probanden Routineprobleme lösen können. Beide Wissensarten könnten unterschiedliche Relationen zu Konstrukten wie der Intelligenz und der mentalen Repräsentation von Zahlen besitzen.

5 Erwerb und Abruf konzeptuellen und prozeduralen Wissens

5.1 Wege des Wissenserwerbs

Im letzten Kapitel wurde besprochen, welche Eigenschaften erworbenes konzeptuelles und prozedurales Wissen auszeichnen. Dieses Kapitel setzt sich nun mit den spezifischen Erwerbs- und Abrufmechanismen für die beiden Wissensarten auseinander.

Theorien hierüber werden, wie schon bei der Vorstellung des Iterativen Modells erwähnt, entsprechend der Erwerbsreihenfolge, die sie postulieren, häufig in verschiedene Kategorien unterteilt (Baroody, 2003; Rittle-Johnson & Siegler, 1998; Rittle-Johnson et al., 2001): Der *concepts-first view* basiert auf der Annahme, dass konzeptuelles Wissen vor prozeduralem erworben und Letzteres dann aus Ersterem abgeleitet wird. Der *skills-first view* postuliert umgekehrte Erwerbsreihenfolge und Kausalzusammenhänge. Dem Iterativen Modell liegt die Annahme zugrunde, dass die Erwerbsreihenfolge von arbiträren Umwelteinflüssen abhängt und beide Wissensarten sich gegenseitig positiv kausal beeinflussen. Dem *simultaneous development view* zufolge sind konzeptuelles und prozedurales Wissen untrennbar miteinander verbunden und entwickeln sich daher stets gleichzeitig.

Weitere Kategorien nennen Haapasalo und Kadjevich (2000) sowie Byrnes und Wasik (1991): Der *inactivation view* basiert auf der Annahme, dass die beiden Wissensarten unverbunden nebeneinander stehen. Der *simultaneous activation view* postuliert, dass konzeptuelles Wissen eine notwendige und hinreichende Bedingung für den Erwerb prozeduralen Wissens ist. Prozeduren könnten demnach dann, und nur dann, korrekt angewendet werden, wenn die korrekten Konzepte dahinterstehen. Der *dynamic interaction view* besagt, dass konzeptuelles Wissen eine notwendige, aber nicht hinreichende Bedingung für den Erwerb prozeduralen Wissens ist. Die Prozeduren müssen zwar durch Konzepte fundiert werden, können jedoch nicht automatisch aus ihnen abgeleitet werden und müssen zusätzlich unterrichtet werden. Der *genetic view* schließlich postuliert, dass konzeptuelles Wissen nur dann erworben werden kann (jedoch nicht automatisch erworben wird), wenn schon prozedurales vorhanden ist. Er postuliert also genau die umgekehrte Erwerbsrichtung wie der *dynamic interaction view*.

Haapasalo und Kadjevich (2000) weisen darauf hin, dass der *simultaneous activation view* und, vor allem, der *dynamic interaction view* in der pädagogischen Praxis viele Anhänger haben. Die Annahme, dass konzeptuelles Wissen notwendige aber nicht hinreichende Bedingung für den Erwerb prozeduralen Wissens ist, lässt sich gut zur

Rechtfertigung eines bestimmten Unterrichtsschemas nutzen: Zuerst werden in einem Lehrervortrag Konzepte erklärt, dann lösen die Schüler Übungsaufgaben und setzen die Konzepte in Prozeduren um.

Tabelle 4: Einteilung von Theorien anhand der postulierten Relationen.

Name	Hypothetische Relationen zwischen K und P
inactivation view	keine Relationen zwischen K und P
concepts-first view	K wird zuerst erworben, P daraus abgeleitet
procedures-first view	P wird zuerst erworben, K daraus abgeleitet
Iterative Model	arbiträre Erwerbsreihenfolge, bidirektionale Beziehungen
simultaneous development view	K und P untrennbar verbunden, daher gleichzeitig erworben und gleichzeitige Entwicklung
simultaneous activation view	K ist notwendige und hinreichende Bedingung für P, P jedoch weder notwendige noch hinreichende Bedingung für K
dynamic interaction view	K ist notwendige, aber nicht hinreichende Bedingung für P
genetic view	P ist notwendige, aber nicht hinreichende Bedingung für K

Anmerkungen. K: konzeptuelles Wissen; P: prozedurales Wissen.

Die unterschiedlichen Positionen sind in Tabelle 4 noch einmal zusammengestellt. Auffällig ist die unglückliche Namensähnlichkeit des simultaneous development view und des simultaneous activation view. Die Kategorien sind außerdem nicht exklusiv: Der dynamic interaction view kann als Unterpunkt des concepts-first view betrachtet werden, weil er besagt, dass Prozeduren erst dann erworben werden können, wenn schon die zugehörigen Konzepte vorhanden sind. In gleicher Weise ist der genetic view ein Unterpunkt des procedures-first view.

Während die den Ansätzen zugrunde liegenden Fragen nach der optimalen Wissenserwerbsreihenfolge und nach Relationen zwischen den Wissensarten berechtigt sind, ist zweifelhaft, ob eine so formalistisch-reduzierte Beschreibungsebene, wie die in Tabelle 4 dargestellte, alleine zur Optimierung von Lernprozessen beitragen kann. Detaillierte Informationen über die mentalen *Mechanismen* der Wissensverarbeitung könnten hier hilfreiche Ergänzungen sein.

Mindestens drei Typen solcher Mechanismen können unterschieden werden: (1) Mechanismen des Erwerbs konzeptuellen und prozeduralen Wissens, (2) Mechanismen der Umsetzung konzeptuellen in prozedurales Wissen und umgekehrt und (3) Mechanismen des Abrufs und des Transfers der Wissensarten auf neue Problemtypen und Inhaltsgebiete.

In Veröffentlichungen über konzeptuelles und prozedurales Wissen werden nur sporadisch Vermutungen über mögliche Wissensverarbeitungsmechanismen geäußert. Ein

vollständiger und systematischer Überblick fehlt. Theorien über ähnliche Paare von Wissensarten treffen hingegen detailliertere Annahmen. Darum werden in den folgenden beiden Abschnitten zunächst zwei der wichtigsten dieser dual component-Theorien dargestellt: die ACT-R-Theorie über deklaratives und prozedurales Wissen und die Theorie der representational redescription über implizites prozedurales und explizites konzeptuelles Wissen. Anschließend werden die darin und in den Publikationen über konzeptuelles und prozedurales Wissen beschriebenen Mechanismen in einem dritten Abschnitt zusammengefasst.

5.2 Die ACT-Theorie

Die ACT-Theorie von John R. Anderson (1976; 1983; 1990; 1993; Anderson & Lebiere, 1998) basiert auf der Unterscheidung von deklarativem und prozeduralem Wissen: „There are good reasons for wanting to have both declarative and procedural knowledge structures. Declarative knowledge is flexible and can be accessed in many ways. [...] Procedural knowledge is rigid but efficient“ (Anderson, 1983, S. 39). Vergleicht man diese Beschreibung mit den Definitionen konzeptuellen und prozeduralen Wissens von Rittle-Johnson et al. (siehe Seite 36), so ist die Ähnlichkeit der beiden Paare von Wissensarten unübersehbar. Da deklaratives Wissen neben konzeptuellen Prinzipien auch Wissen über spezifische Episoden, Bilder, Symbole und Handlungsziele (*goal states*) umfassen kann, wird es wohl am besten als Oberbegriff aufgefasst, unter den konzeptuelles Wissen subsumiert werden kann.

Anderson geht davon aus, dass deklaratives Wissen in einem konnektionistischen Netz von Chunks und prozedurales Wissen in Produktionsregeln gespeichert werden. Aktuelle Gedanken- oder Wahrnehmungsinhalte werden im deklarativen Wissensnetz aktiviert und aktivieren anschließend benachbarte Knoten im Netz (*spread of activation*), wodurch unter anderem semantische Primingeffekte gut erklärt werden können. Alle im Langzeitgedächtnis aktivierten Knoten gemeinsam stellen das Arbeitsgedächtnis dar. Indizieren *pattern matching*-Prozesse die Erfüllung des Bedingungsteils einer Produktion durch den aktuellen Arbeitsgedächtnisinhalt, so wird der Handlungsteil der Produktion automatisch ausgeführt (*production firing*). Werden die Bedingungsteile verschiedener Produktionen gleichzeitig erfüllt, entscheiden Konfliktlöseregeln darüber, welche ausgeführt wird. Die wichtigste Regel ist, dass die Produktionen jeweils eine Stärke besitzen und dass starke Produktionen schwachen gegenüber bevorzugt ausgeführt werden. Die Stärke einer Produktion steigt mit jeder erfolgreichen Ausführung und sinkt mit jeder erfolglosen, wodurch beispielsweise trial-and-error learning erklärt werden kann.

Der Erwerb deklarativen Wissens wird in ACT kaum problematisiert. Anderson (z.B. 1983, S. 171ff) geht davon aus, dass jeder wahrgenommene Stimulus – unabhängig von seinem Inhalt sowie Motivation und Intention des Lernalers – mit einer gewissen Wahrscheinlichkeit in das deklarative Langzeitgedächtnis übernommen wird. Auch in späteren Phasen der Informationsverarbeitung wird jeder Chunk, unabhängig von seinem Inhalt auf dieselbe Weise verarbeitet. Die Phänomene, dass bedeutungshaltiges Material besser erinnert wird als bedeutungsloses und dass eine Elaboration von Bedeutung gewöhnlich zu besseren Behaltensleistungen führt, wie sie zum Beispiel aus der Studie zum Erinnern von Doodles (Bower, Karlin, & Dueck, 1975) und zum levels-of-processing-Ansatz (Craik & Lockhart, 1972) bekannt sind, wird in ACT vor allem darüber erklärt, dass bedeutungshaltiges Material besser mit anderen Knoten im deklarativen Netz verknüpft werden kann als bedeutungsloses. Besser verknüpfte Knoten erreicht mehr sich im Netz ausbreitende Aktivierung, so dass die entsprechenden Knoteninhalte häufiger aktiviert und dadurch erinnert werden.

Über den Erwerb prozeduralen Wissens schreibt Anderson: „According to ACT*, all knowledge initially comes in declarative form and must be interpreted by general procedures“ (Anderson, 1983, S. 34). Das deklarative Wissen kann dabei durch Lehrer oder Lehrbücher vermittelt oder auch selbst erschlossen sein. Die allgemeinen Prozeduren, die in dieser Phase zur Anwendung kommen, sind aufgrund ihrer Vielseitigkeit allerdings langsam und umständlich auszuführen und binden dabei viele kognitive Ressourcen, wie z.B. Arbeitsgedächtniskapazität. Die Kodierung neuen prozeduralen Wissens (*production compilation*) im Zuge fortschreitender Übung besteht aus zwei Teilprozessen (z.B. Anderson, 1983, S. 235): Die sogenannte Prozeduralisierung (*proceduralization*, auch *production specialisation*) besteht darin, dass die verschiedenen Leerstellen (*slots*) beziehungsweise Variablen in Produktionsregeln mit Werten belegt werden, die bei der Lösung eines Problemtyps standardmäßig dort auftreten. Dadurch, dass diese Werte dann nicht mehr bei jeder Produktionsausführung neu bestimmt werden müssen, ist die Produktion zwar schneller ausführbar und effizienter geworden, jedoch wurde zugleich auch ihr Anwendungsbereich reduziert. Der zweite Teilprozess der Kompilierung neuen prozeduralen Wissen besteht in der so genannten *Komposition* von Prozeduren (*production composition*): Dabei werden Produktionen, die häufig nacheinander ausgeführt werden, zu einer einzigen Produktion zusammengefasst. Für die Ausführung bedeutet dies, dass statt mehreren Bedingungsteilen nur noch ein einziger auf den jeweiligen Arbeitsgedächtnisinhalt gematcht werden muss. Außerdem muss statt mehrerer

Produktionen nur noch eine einzige im Arbeitsgedächtnis gehalten werden. Dieselbe Problemlöseleistung kann dadurch schneller und ressourceneffizienter als zuvor erbracht werden. Da die Ausführungsreihenfolge der einzelnen Produktionsregeln nicht mehr geändert werden kann, sobald sie zu einer einzigen verschmolzen wurden, geht auch dies mit einem Verlust an Flexibilität einher.

Fertig kompilierte Prozeduren können im Verlaufe ihrer weiteren Anwendung noch weiter modifiziert werden (*production tuning*). Der Hauptmechanismus dabei ist die oben bereits erwähnte Veränderung der Produktionsstärke in Abhängigkeit vom Erfolg ihrer Anwendung. Möglicherweise können zu generell formulierte Produktionen auch spezialisiert (*production discrimination*) werden (durch Hinzufügen neuer Bedingungen zum Bedingungsteil) und zu restriktiv formulierte generalisiert werden (durch Elimination einzelner Bedingungen aus dem Bedingungsteil). Diese kognitiven Prozesse der Diskrimination und Generalisierung sind jedoch empirisch bisher nicht eindeutig belegt worden.

Im Zuge der Weiterentwicklung der ACT*-Theorie zu den aktuelleren Versionen der ACT-R-Theorie räumt Anderson ein, dass Fähigkeitserwerb nicht grundsätzlich mit dem Erwerb deklarativen Wissens über Instruktionen beginnen muss: „Originally, the emphasis had been on declarative memory for instructions, but now the emphasis is on declarative memory for examples“ (Anderson & Fincham, 1994, S. 1322). Anderson trägt damit empirischen Befunden Rechnung, die belegen, dass auch deklarative Erinnerungen an Lösungsbeispiele prozedurale Fähigkeiten verbessern können (vgl. die Ausführungen zu Logans instance theory of automatization in Abschnitt 4.4.2). Taatgen und Wallach (2002) stellen die interessante Hypothese auf, dass Personen bevorzugt Produktionserwerbsprozesse nutzen, weil diese besonders effizient sind, und lediglich wenn diese nicht möglich sind, beispielsweise weil in komplexen Problemlöseszenarien keine gültigen Handlungsregeln erkennbar sind, auf die gedächtnisintensiveren Prozesse des instance learning zurückgreifen.

Durch die Annahme, dass deklaratives und prozedurales Wissen unterschiedlichen Vergessensprozessen unterliegen, erklären Anderson und Fincham (1994) empirische Befunde über Personen, die anscheinend prozedurales Wissen unabhängig von deklarativem erwerben konnten (vgl. Abschnitt 4.3). Die ACT-Theorie in ihrer aktuellen Form kann also weder den concepts-first- noch den procedures-first-Ansätzen eindeutig zugeordnet werden. Sie postuliert die Prozesse, mittels derer deklaratives in prozedurales Wissen umgesetzt wird, jedoch mit beeindruckender Genauigkeit, während sie über den

umgekehrten Erwerbsweg wenig aussagt. Zur Ergänzung wird darum nun eine Theorie vorgestellt, die den Fokus auf den umgekehrten Erwerbsweg setzt.

5.3 Die Theorie der *representational redescription*

Jean Piagets entwicklungspsychologischer Ansatz beruht auf verschiedenen wichtigen Annahmen. Seine Annahme domänenübergreifender Entwicklungsstufen ist angesichts unzähliger empirischer Befunde zur Domänenspezifität des Wissenserwerbs und Fodors Überlegungen zur Modularität der Psyche nicht mehr haltbar.

Annette Karmiloff-Smith (1992) versucht durch ihre Theorie der *representational redescription* aufzuzeigen, dass eine andere wichtige Grundannahme Piagets unabhängig davon noch Gültigkeit besitzen kann, nämlich die Annahme, dass ein abstraktes Verständnis eines Inhaltsgebiets graduell aus Handlungserfahrungen abstrahiert wird. Während Piaget dies durch die Internalisierung von Handlungsschemata und nachfolgende Ausdifferenzierungen dieser Schemata durch Akkomodationsprozesse bei Verletzung des psychischen Äquilibrium erklärte, greift Karmiloff-Smith bewusst auf Konstrukte der zeitgenössischen Kognitionspsychologie, nämlich Wissensarten, zurück, um Piagets Ansatz an aktuellere Entwicklungen wieder anschlussfähig werden zu lassen. Sie spricht von implizitem und explizitem Wissen.

In der Literatur zum impliziten Lernen wird explizites Wissen vor allem als verbalisierbares und implizites Wissen als nichtverbalisierbares Wissen aufgefasst (Siegler & Stern, 1998). Aus dieser Perspektive können sowohl konzeptuelles als auch prozedurales Wissen explizit oder implizit sein (vgl. die Anmerkungen zu explizitem und implizitem Konzeptwissen in den Unterabschnitten 2.2.1 und 4.5.1). Karmiloff-Smith setzt jedoch einen etwas anderen Fokus, indem sie implizites Wissen mit prozeduralem Wissen gleichsetzt und explizites Wissen mit konzeptuellem. So schreibt sie über das initiale implizite Wissen von Kindern: „Information is encoded in procedural form. The procedure-like encodings are sequentially specified“ (Karmiloff-Smith, 1992, S. 20). Dieses implizite prozedurale Wissen ist Karmiloff-Smith zufolge enkapsuliert, d.h. es kann zwar effizient ausgeführt werden, höhere kognitive Prozesse können auf seinen Inhalt jedoch nicht zugreifen. Es kann daher weder bewusst verarbeitet, elaboriert noch verbalisiert werden. Karmiloff-Smith (1992, S. 20ff) bezeichnet dies als *Level I* des Wissenserwerbs, wobei das *I* für *implicit* steht..

Sobald sich das Wissen auf diesem Level stabilisiert hat und erfolgreich angewendet werden kann, wird aus ihm per *representational redescription* explizites Wissen des so genannten *Levels E1* abgeleitet. Das Wissen des Levels I bleibt dabei weiter bestehen und

wird nun um ein zusätzliches Repräsentationsformat ergänzt. Dieses explizite Wissen ist nicht mehr enkapsuliert, kann also elaboriert und somit auch zu anderen Wissenseinheiten in Bezug gesetzt werden. Das Wissen ist jedoch noch oberflächlich, wenig detailliert und muss nicht zwangsläufig bewusst und verbalisierbar sein.

Sobald sich das Wissen auf Level E1 stabilisiert hat, wird durch eine erneute representational redescription Wissen des Levels E2 daraus abgeleitet. Auf dieser neuen Ebene ist Wissen grundsätzlich bewusstseinsfähig, jedoch nicht verbalisierbar. Karmiloff-Smith geht davon aus, dass abstrakt-visuelle Repräsentationsformen, wie Diagramme, auf diesem Level von besonderer Bedeutung sind. Auch der in Abschnitt 4.5.1 erwähnte gesture-speech mismatch sollte sich auf dieser Stufe zeigen.

Den Endpunkt dieser Entwicklung, die auch als Wissensexplikation bezeichnet wird, stellt Level E3 dar. Auf diesem Level ist das Wissen sowohl explizit repräsentiert als auch bewusstseinsfähig und verbalisierbar. Es ist abstrakt, so dass seine Ursprünge in der Wahrnehmung und im Verhalten sich kaum noch darin widerspiegeln und es leicht zwischen verschiedenen Sinnesmodalitäten und Handlungskontexten transferiert werden kann.

Ein wichtiger Unterschied zwischen dem von Karmiloff-Smith und dem von Piaget vorgeschlagenen Erwerbsweg ist, dass nach Piaget abstrakteres Wissen erst bei Verlust des psychischen Äquilibriums, also bei kognitiven Konflikten, geschaffen wird, während Karmiloff-Smith davon ausgeht, dass eine stabile Beherrschung des Wissens auf einer Ebene, also kognitives Gleichgewicht, automatisch zur representational redescription führt. Des Weiteren postuliert Karmiloff-Smith nicht, dass die Phasen des Wissenserwerbs mit spezifischen Lebensaltern assoziiert sind, sondern lediglich, dass die Stufen – in jedem Inhaltsgebiet einzeln – nacheinander durchlaufen werden müssen.

Karmiloff-Smith räumt ein, dass sie diese Vorgänge nur auf einer abstrakten Ebene beschreiben kann und wenig zur Implementierung der Prozesse auf der komputationalen Ebene sagen kann. Die empirische Überprüfung des Modells leidet stark darunter, dass bis jetzt nicht abzusehen ist, wie die explizite Kodierung und die Bewusstseinsfähigkeit von Wissen unabhängig von der Verbalisierbarkeit dieses Wissens operationalisiert werden könnten.

Nichtsdestotrotz zeigt Karmiloff-Smith in ihrem Buch, dass die Theorie mit verschiedenen entwicklungspsychologischen Phänomenen in den Domänen des Spracherwerbs, der Physik, der Mathematik, der theory of mind und des Beherrschens von Notationsformen gut vereinbar ist.

Zusammengenommen bestätigen die ACT-Theorie und die Theorie der representational redescription Rittle-Johnsons Argumentationslinie bezüglich des Iterativen Modells: bisherige Theorien haben sich weitgehend auf die Analyse eines der beiden Erwerbswege beschränkt (concepts-first und procedures-first view). Die Prozesse, die beide Gruppen von Theorien vorschlagen, sind allerdings so plausibel, dass die Annahme bidirektionaler iterativer Beziehungen zwischen den Wissensarten gerechtfertigt scheint – auch wenn bisher kaum empirische Evidenz über solche Beziehungen vorliegt.

Auffällig ist, dass Karmiloff-Smith den initialen Erwerb des impliziten prozeduralen Wissens kaum thematisiert, sondern sich, ähnlich wie Anderson in seiner Theorie, stark auf Transformationsprozesse *zwischen* den Wissensarten konzentriert. Eingangs wurde jedoch die Frage nach drei verschiedenen Mechanismen der Verarbeitung konzeptuellen und prozeduralen Wissens aufgeworfen: (1) Mechanismen des direkten Erwerbs konzeptuellen und prozeduralen Wissens, (2) Mechanismen der Umsetzung konzeptuellen Wissens in prozedurales Wissen und umgekehrt, (3) Mechanismen des Abrufs und des Transfers der Wissensarten auf neue Problemtypen und Inhaltsgebiete. Da offensichtlich keine einzelne Lerntheorie, weder das Iterative Modell, die ACT-Theorie, die Theorie der representational redescription, noch andere, hier nicht besprochene, in der Lage ist, all diese Mechanismen hinreichend zu spezifizieren, wird im nächsten Abschnitt zu jedem der drei Mechanismen theorieübergreifend zusammengestellt, wie sie in der Literatur charakterisiert werden. Eine Übersicht darüber bietet Tabelle 5 am Ende des Abschnitts.

5.4 Hypothetische Verarbeitungsmechanismen

5.4.1 Direkter Erwerb konzeptuellen und prozeduralen Wissens

Konzeptuelles Wissen

Während die Theorie der representational redescription keine Angaben darüber macht, wie konzeptuelles Wissen direkt, d.h. ohne den vorherigen Erwerb prozeduralen Wissens, erworben werden kann, wird der ACT-Theorie zufolge konzeptuelles Wissen so erworben, wie auch alle anderen Arten deklarativen Wissens: Im Arbeitsgedächtnis memorierte Wissenschunks werden mit einer gewissen konstanten Wahrscheinlichkeit in das Langzeitgedächtnis übernommen und dort umso besser verknüpft, je stärker sie elaboriert werden.

Befunde zum conceptual change zeigen, dass dieses Bild übersimplifiziert sein könnte. So fanden Vosniadou und Brewer (1992) in der bereits oben erwähnten Arbeit zum Erwerb des Konzepts der Erde als Kugel, dass Kinder sich einige Fakten über Himmelskörper,

wie: „Auf dem Mond gibt es Krater, die durch Meteoriteneinschläge entstanden sind.“ leicht einprägen können, während ihnen dies bei anderen Fakten, wie: „Die Erde ist eine Kugel.“ nur sehr schwer gelingt. Offensichtlich ist Wissen um einige konzeptuelle Zusammenhänge leichter zu erwerben als um andere. Eine zentrale Rolle hierbei spielt das Vorwissen des Lerner. Was zunächst wie Konzepterwerb aussehen mag, ist in Wirklichkeit fast immer Konzeptwechsel, weil es kaum einen Lebensbereich gibt, dem ein Lerner nicht schon subjektiv befriedigende Präkonzepte oder Alltagstheorien entwickelt hat oder sogar auf angeborenes Prinzipienwissen zurückgreifen kann (Chi, Slotta, & Leeuw, 1994; Nesher, 1987; Spelke, Breinlinger, Macomber, & Jacobson, 1992).

Posner, Strike, Hewson und Gertzog (1982) geben in ihrer klassischen Arbeit zum *conceptual change* vier Bedingungen an, die erfüllt sein müssen, damit Konzeptwechsel stattfindet: (1) Der Lerner muss mit seinen vorhandenen Konzepten unzufrieden sein. (2) Das neue Konzept muss verständlich sein. (3) Das neue Konzept muss eine höhere Plausibilität besitzen als das vorherige. (4) Das neue Konzept sollte für die Wissensentwicklung des Lerner fruchtbar sein und ihm auch in anderen Kontexten als dem Erwerbskontext helfen.

Vosniadou, Ioannides, Dimitrakopoulou und Papademetriou (2001) geben sieben Prinzipien an, die beim Design von Lernumgebungen für konzeptuelles Wissen beachtet werden sollten: (1) Es sollte eher ein kleines Inhaltsgebiet vertieft behandelt werden als ein großes oberflächlich. (2) Die Reihenfolge, in der Konzepte präsentiert werden, sollte aus der sachlogischen Struktur des Gegenstandsbereiches abgeleitet sein. (3) Das Vorwissen der Lerner muss berücksichtigt werden. (4) Metakognitive und reflektive Prozesse sollen gefördert werden. (5) Stabile Präkonzepte der Lerner müssen abgebaut werden. (6) Beim Lerner muss eine Motivation für einen Konzeptwechsel geschaffen werden. (7) Kognitive Konflikte sollten dabei maßvoll und vorsichtig evoziert werden, weil sie zwar einerseits zur Aufgabe suboptimaler Präkonzepte, andererseits aber auch zur Verunsicherung der Lerner führen und alleine noch nicht den Aufbau neuer Konzepte hervorrufen.

Wie diSessa (1998) anmerkt, sind Prinzipien zur Förderung von *conceptual change* zwar pädagogisch wertvoll, geben jedoch kaum Aufschluss über die einzelnen kognitiven Prozesse, die sich während des Konzepterwerbs abspielen. DiSessa weist darauf hin, dass Wissenserwerb in semantischen Netzwerken aus zwei verschiedenen Prozessen bestehen könnte: durch Veränderung (Einfügen, Löschen, Ausdifferenzieren usw.) der Knoten des Netzwerkes und durch Veränderung der Kanten des Netzwerkes. Empirische Befunde hierüber liegen jedoch nicht vor, was auch daran liegt, dass Studien zum *cognitive change*

Konzeptwissen häufig über Interviewdaten erheben, die naturgemäß wenig Aufschluss über den Ablauf kognitiver Mechanismen geben.

Auf die zahlreichen kognitiven Experimente, zum Beispiel auch Säuglingsstudien, zum Kategorielernten kann an dieser Stelle nur hingewiesen werden, da die vorliegende Arbeit auf den Erwerb komplexerer Konzepte in Unterrichtskontexten fokussiert.

Prozedurales Wissen

Wie beschrieben, basiert der Ansatz von Karmiloff-Smiths auf der Idee, dass Kinder zuerst prozedurales Wissen direkt, das heißt ohne den „Umweg“ über konzeptuelles Wissen, erwerben. Sun, Merill und Peterson (2001) knüpfen beim Entwurf ihres *bottom-up models of skill learning* namens *Clarion* unter anderem auch an Karmiloff-Smiths Theorie an. In Clarion werden erfolgreiche Prozeduren mittels trial-and-error-Lernens identifiziert. Das so gewonnene prozedurale Wissen wird in einem neuronalen Netzwerk gespeichert. Da das Modell auf mathematischen Lernalgorithmen aus der Künstlichen Intelligenz-Forschung beruht, ist seine psychologische Validität unklar. Sun und Kollegen weisen jedoch darauf hin, dass ihr Modell gut geeignet ist, um einige empirisch beobachtete Phänomene des impliziten Lernens komputational zu erklären.

Auch die ACT-Theorie impliziert in ihrer aktuellen Form die Möglichkeit, dass konzeptuelles Wissen ohne prozedurales erworben werden kann. Insbesondere die Experimente von Anderson und Fincham (1994) sowie Anderson, Fincham und Douglass (1997) belegen detailliert, dass prozedurales Wissen aus deklarativ gespeicherten Lösungsbeispielen für eine Aufgabe abgeleitet werden kann. Durch Analogieschlüsse können diese Lösungsbeispiele dann auf neue Aufgaben desselben Typs übertragen werden. Lösungsbeispiele stellen zwar deklaratives Wissen dar, nicht jedoch konzeptuelles Wissen, da sie weder abstrakt, noch allgemeingültig, noch vernetzt sind. Der vorgeschlagene Mechanismus erklärt also, wie prozedurales Wissen unabhängig von konzeptuellem erworben werden kann.

Da Anderson und Kollegen unterschiedliche Speicher- und Vergessensprozesse für deklaratives und prozedurales Wissen postulieren, besteht zusätzlich sogar die Möglichkeit, dass permanentes prozedurales Wissen unabhängig von permanentem deklarativen Wissen erworben wird, nämlich dann, wenn neue Prozeduren wiederholt aus deklarativen Lösungsbeispielen oder deklarativen Instruktionen abgeleitet werden, die sich lediglich temporär im Arbeitsgedächtnis befinden. Wird eine solche Prozedur mehrfach geübt, kann sie in das prozedurale Langzeitgedächtnis übernommen werden, auch wenn

die zugehörige deklarative Information nicht in das deklarative Langzeitgedächtnis übergeht und vergessen wird.

Weitere Erwerbsprozesse für prozedurales Wissen postuliert Siegler's *strategy change model* (Shrager & Siegler, 1998). Es basiert auf der Annahme von konnektionistischer Verbindungen zwischen Problemtypen und Lösungen. Durch trial-and-error learning werden diese sukzessive optimiert. Dadurch lassen sich zwar Entscheidungen zwischen bekannten Strategien erklären, nicht jedoch die Neuentdeckung von Strategien. Sie basieren dem Modell zufolge auf zusätzlichen metakognitiven Prozessen, die dem Lerner helfen, Lösungswege zu beobachten, zu bewerten und zu vergleichen. Siegler's Modell fokussiert allerdings ganz auf Strategien. Diese definiert er als Untertyp prozeduralen Wissens. Über die Rolle anderer Wissensarten er macht keine Aussagen.

5.4.2 Umsetzung der Wissensarten ineinander

Umsetzung von konzeptuellem in prozedurales Wissen

Trotz der eben erläuterten Möglichkeiten prozedurales Wissen relativ unabhängig von deklarativem zu erwerben konstatiert Johnson (2003) in einem Überblicksartikel: „It is probably not too daring to say that all major models of skill acquisition, just as the acquisition of skill itself, begin with declarative memory“ (S. 499). Die Umsetzung konzeptuellen Wissens in prozedurales ist entsprechend ein gut untersuchter Mechanismus.

Die ACT-Theorie sieht, wie schon beschrieben, zwei Mechanismen vor, über die deklaratives Wissen zu einem Zuwachs an prozeduralem führen kann: Im Zuge der production specialization werden die Slots von Prozeduren mit aus dem deklarativen Wissen abgeleiteten permanenten Werten belegt. Im Verlaufe des production chunking werden Produktionen, die auf Basis deklarativer Handlungsanweisungen häufig hintereinander ausgeführt wurden, zu einer Gesamtproduktion verschmolzen.

Auch in der Literatur zu konzeptuellem und prozeduralem Wissen wurden einige mögliche Mechanismen vorgeschlagen. Rochel Gelman geht davon aus, dass Konzepte in Form von constraints die Auswahl von Prozeduren leiten, indem sie dem Lerner helfen, Prozeduren zu verstehen, zu bewerten und daher situationsangemessen auszuwählen (siehe Seite 17ff). Rittle-Johnson et al. (2001) äußern sich ähnlich. Blöte, Van der Burg und Klein (2001) ergänzen, dass die logische Folge von konzeptuellem Wissen eine erhöhte *prozedurale Flexibilität* ist. Wer prozedurales Wissen allein besitzt, sollte stets dieselben Routineprozeduren verwenden, wer konzeptuelles Wissen besitzt, sollte in der Lage sein, (a) neue Prozeduren zu konstruieren, falls notwendig, und (b) zwischen alternativen Prozeduren flexibel und adaptiv zu wählen. Außerdem vermuten Blöte und Kollegen, dass

konzeptuelles Wissen sich nicht nur auf die Auswahl, sondern auch auf die Ausführung von Prozeduren auswirken kann, weil es Informationen über die Plausibilität von Zwischenergebnissen und Endresultaten einer Problemlösung bereitstellt. Konzepte können Problemlösern also helfen, bereits gemachte Fehler zu sehen und zu korrigieren.

Konzeptuelles Wissen könnte sich neben diesen direkten Umwandlungsmechanismen auch über einen Mediator, nämlich über die Verbesserung der Problemrepräsentation, auf prozedurales Wissen auswirken (Hiebert & Lefevre, 1986). Novick und Hmelo (1994) schreiben „Understanding comes from having a good representation.“ (S. 1299) und weisen damit auf den Zusammenhang zwischen konzeptuellem Wissen und der Repräsentation von Problemen hin. Eine gute Problemrepräsentation kann die Auswahl und Ausführung effektiver Prozeduren erleichtern.

Umsetzung von prozeduralem in konzeptuelles Wissen

Rittle-Johnson et al. (2001) schlagen bei der Diskussion ihres Iterativen Modells vier verschiedene Mechanismen vor, mittels derer prozedurales Wissen zu Zuwächsen an konzeptuellem Wissen führen könnte.

Erstens könnte bei der Generierung, der Bewertung und der Ausführung von Prozeduren sowie bei der Ergebniskontrolle konzeptuelles Wissen verwendet werden. Einige kognitive Theorien, so auch die ACT-Theorie, besagen, dass schon die regelmäßige Aktivierung von Wissensseinheiten zur Verbesserung ihrer Abspeicherung und Verknüpfung im Gedächtnis führt. So könnte prozedurales Wissen konzeptuelles schon alleine dadurch stärken, dass es darauf zugreift und es dabei aktiviert.

Zweitens könnten Zuwächse an prozeduralem Wissen bewirken, dass während der Problemlösung mehr kognitive Ressourcen, wie Aufmerksamkeit und Arbeitsgedächtniskapazität, zur Verfügung stehen. Diese könnten wiederum zur Überprüfung, Elaboration und Abstraktion von Konzepten im Problemlösekontext genutzt werden.

Drittens kann die Beherrschung korrekter Prozeduren Lernenden helfen, ihre Misskonzepte zu bemerken und zu korrigieren. Wer korrekte Prozeduren beherrscht, produziert beispielsweise viele korrekte Lösungen, die das Falsifizieren von Misskonzepten oder das induktive Erschließen konzeptueller Gesetzmäßigkeiten erleichtern können.

Viertens könnte das Erlebnis, dass gewisse Prozeduren funktionieren, obwohl sie nicht mit den konzeptuell begründeten Erwartungen des Lernalters übereinstimmen, Lerner

neugierig machen und zur Reflektion anregen. Diese kann dann zu einer besseren Elaborierung des konzeptuellen Wissens führen.

Hiebert und Lefevre (1986) nennen zwei weitere Punkte: Da Konzepte statische Gedächtnisstrukturen darstellen, die erst durch prozedurales Wissen in Problemlösungen eingebracht werden, kann ein Zuwachs an prozeduralem Wissen auch zur Folge haben, dass danach mehr oder adäquatere Konzepte durch Prozeduren genutzt werden können als zuvor.

Einige Prozeduren, die man wohl auch als metakognitive Prozeduren bezeichnen könnte, ermöglichen die Elaboration, Verknüpfung und inferenzielle Erweiterung konzeptuellen Wissens. Ein Zuwachs an solchen Prozeduren könnte sich besonders günstig auf konzeptuelles Wissen auswirken.

Diese Aufstellung verdeutlicht, dass positive Auswirkungen prozeduralen Wissens auf konzeptuelles Wissen nicht automatisch das Vorliegen von representational redescription indizieren. Representational redescription könnte ein Prozess sein, der zusätzlich zu den hier beschriebenen Prozessen stattfindet oder er könnte auch ein Oberbegriff sein, unter den sie sich ganz oder teilweise subsumieren lassen. Fest steht, dass positive Einflüsse von prozeduralem auf konzeptuelles Wissen, sofern sie gefunden werden, sich kognitiv gut erklären lassen.

5.4.3 Abruf und Transfer der Wissensarten

Abruf und Transfer konzeptuellen Wissens

Die Transferforschung ist ein weites Forschungsfeld, das hier nicht vollständig dargestellt werden kann. Jedoch wurde in den vorgehenden Abschnitten dieser Arbeit deutlich, dass die Autoren des Iterativen Modells gemeinsam mit anderen davon ausgehen, dass konzeptuelles Wissen flexibel transferierbar und prozedurales Wissen an Routineprobleme gebunden ist (vgl. Kapitel 3, sowie die Unterkapitel 0, 4.4 und 4.5). Es soll kurz diskutiert werden, auf welche Verarbeitungsmechanismen dies zurückgeführt werden kann.

Mittels einer Fallstudie über die Arbeiten Johannes Keplers demonstrieren Gentner et al. (1997), dass Analogien häufig eine wichtige Rolle spielen, wenn Menschen neue Konzepte entwickeln. Dabei greifen sie einerseits auf Theorien des conceptual change zurück, andererseits auf Gentners *structure mapping theory of analogy* (Falkenhainer, Forbus, & Gentner, 1989; Gentner, 1983). Die structure mapping theory beschreibt den Prozess des analogen Schließens als Schemamappingprozess, bei dem die Elemente und ihre Interrelationen einer Sourcedomäne, zum Beispiel dem Sonnensystem, auf die Elemente und Interrelationen einer Targetdomäne, zum Beispiel ein Atom, gemappt wird.

Sofern ein Schlussfolgerer sich in der Sourcedomäne besser auskennt als in der Targetdomäne, kann er dann sein Wissen über Zusammenhänge in der Sourcedomäne nutzen, um Hypothesen (*candidate inferences*) über Zusammenhänge in der Targetdomäne zu entwickeln. In zahlreichen empirischen Arbeiten zeigen Gentner und Kollegen die Effektivität eines solchen so genannten *analogen Transfers*, nicht nur bei Wissenschaftlern, wie Kepler, sondern auch und vor allem bei lernenden Kindern, auf (z.B. Donnelly & McDaniel, 1993; Gentner, 1988; Gentner, Loewenstein, & Thomson, 2003; Gentner & Wolff, 2000). Die wichtigste Voraussetzung für analogen Transfer ist, dass das Wissen in der Targetdomäne in relationaler Form, also als Schema oder Wissensnetz, im Gedächtnis verfügbar ist, denn auf einander gemappt werden immer nur Systeme von Relationen, nie isolierte Elemente. Eine weitere Voraussetzung ist, dass das Wissen explizit verfügbar ist, da beispielsweise die Überprüfung von *candidate inferences* nur bewusst und reflektiert gelingen kann. Analoges Transfer im Sinne des *structure mappings* ist also nur mit konzeptuellem, nicht jedoch mit prozeduralem Wissen möglich (vgl. Holyoak, 2005).

Natürlich sind Analogieschlüsse nicht die einzige Möglichkeit, konzeptuelles Wissen zwischen Aufgabentypen oder Domänen zu transferieren. Das Beispiel der Analogieschlüsse lässt jedoch gut erkennen, was auch auf andere Arten von Transfer zutrifft: Konzeptuelles Wissen ist deswegen flexibler übertragbar als prozedurales, weil es relational und explizit repräsentiert wird. So ist es für bewusste Schlussfolgerungs- und Transformationsprozesse gut verfügbar, kann auch auf der verbalen Ebene verarbeitet und durch seine relationale Einbettung kontextsensitiv interpretiert werden.

Dass besonders der relationale Charakter von Wissen seine Transferfähigkeit fördert, zeigen verschiedene Arbeiten, die in Interventionen Wissen um abstrakte Relationen zwischen Wissensseinheiten vermitteln. So wurde gezeigt, dass eine relationale Sprache flexibles relationales Denken und Problemlösen fördert (z.B. Gentner & Loewenstein, 2002; Rattermann & Gentner, 1998) und dass Diagramme, die die abstrakte relationale Struktur von Problemen veranschaulichen, bei geeignetem Einsatz transferförderlich sind (z.B. Felbrich, 2005; Hardy, Schneider, Jonen, Stern, & Möller, zur Publikation angenommen; Novick & Hmelo, 1994; Stern, Aprea, & Ebner, 2003).

Novick und Hmelo (1994) sprechen dabei zwar von *prozeduralem Transfer*, weil es um die Übertragung von Problemlösewissen geht. Ihr Artikel lässt jedoch erkennen, dass das, was ihrer Auffassung nach transferiert wird, ein konzeptuelles Verständnis der abstrakten Problemstruktur ist und somit konzeptuelles Wissen darstellt. Dies verdeutlicht

ein weiteres Mal, wie eng die beiden Wissensarten innerpsychisch ineinander verwoben sind und wie schwer sie auch terminologisch und konzeptuell zu trennen sind.

Abruf und Transfer prozeduralen Wissens

Detaillierte Untersuchungen des Transfers prozeduralen Wissens über verschiedene Aufgabentypen hinweg werden von Singley und Anderson (1989) in ihrer Monografie „The Transfer of Cognitive Skill“ beschrieben. Sie greifen dabei auf die ACT-Theorie zurück, die ja besagt, dass prozedurales Wissen in Produktionen gespeichert wird, die immer dann ausgeführt werden, wenn der aktuelle Arbeitsgedächtnisinhalt ihrem Bedingungsteil entspricht.

Demnach sollte eine im Kontext einer Aufgabenart erworbene Produktion immer dann auch zur Lösung anderer Aufgabenarten genutzt werden können, wenn die beiden Aufgabenlösungen ganz oder teilweise gleiche Arbeitsgedächtnisinhalte implizieren. Diese Eigenschaft von Produktionsregeln wird als *use specificity* bezeichnet. Singley und Anderson (1989, S. 2ff) weisen darauf hin, dass dieses Prinzip eine große Ähnlichkeit mit Thorndikes *law of identical elements* besitzt. Während Thorndike jedoch vorschlug, dass die äußeren Strukturen von Aufgaben sich überlappen müssen, damit Transfer stattfinden kann, postulieren Singley und Anderson, dass die kognitiven Strukturen sich überlappen müssen. Die Richtigkeit dieser Annahme belegen sie in ihrem Buch durch mehrere Experimente. Einen weiteren Nachweis liefert Müller (1999).

Rittle-Johnson und Kollegen (2001) beschreiben in ihrer Definition prozeduralen Wissens, dass es an Routineprobleme gebunden ist. Das Prinzip der *use specificity of production rules* macht deutlich, was in diesem Zusammenhang als Routineproblem betrachtet werden kann, nämlich jede Aufgabe, die Arbeitsgedächtnisinhalte erzeugt, die ganz oder teilweise mit Arbeitsgedächtnisinhalten übereinstimmen, die aus Übungskontexten vertraut sind.

5.4.4 Übersicht

Tabelle 5 gibt einen Überblick über die in diesem Abschnitt vorgestellten hypothetischen Verarbeitungsmechanismen für konzeptuelles und prozedurales Wissen. Man beachte, dass die Kategorien nicht wechselseitig exklusiv sind, sondern lediglich eine Auflistung der von unterschiedlichen Autoren gemachten Vorschläge darstellen.

Die relativen Häufigkeiten und Wichtigkeiten dieser Mechanismen, zum Beispiel während des schulischen Lernens, sind bisher unbekannt. Dennoch sind die vorgeschlagenen Mechanismen theoretisch fundiert und empirisch belegt.

Tabelle 5: Mögliche Verarbeitungsmechanismen für die Wissensarten.

Verarbeitungsweg	Konzeptuelles Wissen	Prozedurales Wissen
Direkter Erwerb	• Memorieren • Elaborieren und Reflektieren • Netzknoten modifizieren • Netzkanten modifizieren	• Trial-and-error-Lernen • Analogien zu Lösungsbeispielen • Metakognition über Strategien
Umsetzung in jeweils andere Wissensart	• Production specialization • Production composition • Prozedurauswahl • Prozedurbewertung • Prozedurale Flexibilität • Ausführungsüberwachung • Ergebniskontrolle • Problemrepräsentation	• Aktivierung konzeptuellen Wissens • Reduzierter Ressourcenverbrauch • Falsifizierung von Misskonzepten • Konkrete Anregung zur Reflektion • Anwendung neuer Konzepte • Elaboration von Konzepten • Representational redescription
Abruf und Transfer	• Inferenzprozesse • Analoges Schließen • Diagrammatic Reasoning	• Produktionsanwendung

5.5 Erwerb der Wissensarten in schulischen Kontexten

5.5.1 Zwei Fragestellungen

Im vorangegangenen Abschnitt wurde darauf hingewiesen, dass in Experimenten zwar viele kognitive Mechanismen der Verarbeitung von konzeptuellem und prozeduralem Wissen nachgewiesen wurden, dass über das Zusammenspiel dieser Mechanismen im Schulunterricht jedoch wenig bekannt ist. Trotzdem werden die Relationen zwischen den beiden Wissensarten in der Literatur auch aus unterrichtswissenschaftlicher Perspektive diskutiert. Entsprechende Autoren verfolgen häufig das Ziel einer kognitiven Fundierung der Unterrichtswissenschaften (z.B. Bruer, 1994; MacGilly, 1994; Morgan, 1997). Konkret wird zum einen diskutiert, welche Gewichtung oder Verknüpfung von konzeptuellem und prozeduralem Wissen das optimale Unterrichtsziel darstellt, zum anderen wird versucht aus den kognitiven Modellen Aussagen über optimale Unterrichtssequenzierungen abzuleiten.

5.5.2 Unterrichtsziele

Baroody (2003) beschreibt eine sich über das gesamte zwanzigste Jahrhundert hinziehende Auseinandersetzung zwischen verschiedenen Gruppen von Didaktikern und Lerntheoretikern, die er als *skills versus concepts debate* bezeichnet. Vertreter behavioristischer oder assoziationsistischer Ideen vertraten die Ansicht, dass prozedurales Wissen das wichtigste Unterrichtsziel darstellt, weil es die schnelle und effiziente Lösung konkreter Probleme ermöglicht. Vertreter einer eher am Pädagogischen Konstruktivismus und conceptual change-Ansätzen orientierten Position sehen hingegen in konzeptuellem Verständnis das wichtigste Unterrichtsziel, weil es einem hilft, sich auch in neuen Lebenslagen orientieren zu können. Wie Baroody beschreibt, ist diese Debatte eng

verwoben mit den so genannten *math wars*, einer in den USA erbittert geführten öffentlichen Auseinandersetzung zwischen Politikern, Lehr-Lern-Forschern und Eltern um die Gestaltung von Unterrichtsreformen und insbesondere um die vom National Council of Teachers of Mathematics herausgegebenen Unterrichtsstandards (z.B. National Council of Teachers of Mathematics, 1989; National Council of Teachers of Mathematics, 2000). In den *math wars* werden vielfältige Interessen mit unterschiedlichen Arten von Argumenten verfolgt (für eine Analyse s. Schoenfeld, 2004). Die Frage nach der Gewichtung von konzeptuellem Verständnis und prozeduralen Fertigkeiten stellt nur einen Teilaspekt der Debatte dar. Dennoch lässt die Diskussion erkennen, dass auch ein öffentliches Interesse an Fragen des Verhältnisses zwischen konzeptueller und prozeduraler Kompetenz besteht und dass Aussagen über die Relationen zwischen konzeptuellem und prozeduralem Wissen direkt die Lehrplangestaltung beeinflussen können.

Baroody weist jedoch auch darauf hin, dass die *skills versus concepts debate*, also die Frage, ob Konzepte ein wichtigeres Lernziel darstellen als Prozeduren oder umgekehrt, aus wissenschaftlicher Sicht heute nicht mehr aktuell ist. Es hat sich die Erkenntnis durchgesetzt – auch durch schlechte Erfahrungen, die in den USA mit einseitigen Unterrichtsreformen gemacht wurden – dass Schüler im Leben beides brauchen und dass darum beides im Unterricht vermittelt werden muss.

Nach Baroody ist die bloße Vermittlung von beidem jedoch nicht ausreichend, weil Kinder oft nicht in der Lage sind, Konzepte und Prozeduren ausreichend gut miteinander zu verknüpfen, selbst wenn sie über beides verfügen. Diese Verknüpfung ist jedoch von besonderer Wichtigkeit, da sie *adaptive Expertise* verleiht: die Fähigkeit, Prozeduren durch Anwendung konzeptuellen Hintergrundwissens adaptiv und flexibel anwenden zu können und Konzepte durch prozedurales Hintergrundwissen schnell und effizient verwenden zu können. Das Konzept der adaptiven Expertise wurde von Hatano (1982) als Gegenentwurf zu den traditionell sehr auf prozedurale Fähigkeiten fokussierenden Expertise-Theorien (vgl. Johnson, 2003) vorgeschlagen. Es basiert auf der Idee, dass konzeptuelles und prozedurales Wissen jeweils spezifische Vor- und Nachteile haben, die eine einzelne Art von Wissen nicht in sich vereinen könnte (vgl. die evolutionspsychologischen Argumente in Unterabschnitt 4.3.5). Durch eine optimale Verknüpfung könnte, nach Hatano und Baroody, ein Lerner von den Vorteilen beider Wissensarten profitieren, ohne durch ihre Nachteile eingeschränkt zu werden. Wie eine solche Verknüpfung im Unterricht konkret erreicht werden kann und ob sie wirklich die vorgeschlagenen Auswirkungen hat und somit ein optimales Unterrichtsziel darstellt, ist jedoch noch nicht systematisch empirisch belegt und ein wichtiges zukünftiges Forschungsfeld.

5.5.3 Unterrichtssequenzierung

Die Frage nach der optimalen Verknüpfung von konzeptuellem und prozeduralem Wissen lenkt die Aufmerksamkeit weg von den Resultaten des Wissenserwerbs hin zum Prozess des Wissenserwerbs. Hier zeigt unter anderem die Theorie der Basismodelle von Fritz Oser (Oser & Patry, 1994; Oser & Sarasin, 1995; Oser & Baeriswyl, 2001), in welcher Weise kognitive Modelle der Interaktion von konzeptuellem und prozeduralem Wissen für die Unterrichtsgestaltung von Bedeutung sein könnten.

Oser unterscheidet zwischen der *Sichtstruktur* des Unterrichts und den *Basismodellen*. Das äußere Unterrichtsgeschehen bildet die Sichtstruktur. Dazu zählt das konkrete Unterrichtsmaterial, Unterrichtsumgebung, Aufteilung des Stoffs, Wahl von Einzel-, Gruppen- oder Partnerarbeit, Entscheidungen für darbietenden oder entwickelnden Unterricht und auch die Freiheitsgrade, mit denen die Kinder ihr Lernen gestalten können (Oser & Sarasin, 1995). Oser geht davon aus, dass auf dieser Ebene Lehrer viele Variationen implementieren sollten, um Abwechslung zu schaffen und um die Unterrichtsgestaltung an die Erfordernisse der jeweiligen Personengruppe und des jeweiligen Stoffs anzupassen.

Hinter der Ebene der Sichtstruktur verbirgt sich nach Oser jedoch eine zweite Ebene, nämlich die der psychologischen Lernprozesse. Oser identifiziert elf qualitativ unterschiedliche und aus unterrichtswissenschaftlicher Sicht relevante Lernprozesse. Diese bezeichnet er als Basismodelle des Unterrichts. Sie lauten: (1) Lernen durch Eigenerfahrung, (2) entwicklungsförderndes und strukturveränderndes Lernen, (3) Problemlösen (entdeckendes Lernen), (4a) Wissensaufbau (Begriffsbildung), (4b) Konzeptbildung, (5) betrachtendes Lernen, (6) Lernen von Strategien, (7) Routinebildung und Training von Fertigkeiten, (8) Motilitätsmodell, (9) Lernen dynamischer Beziehungen, Lernen gemeinsamer Normen durch Partizipation (Kooperationslernen), (10) Wert- und Identitätsaufbau, (11) Hypertextlernen. Oser weist darauf hin, dass Klassifizierungen grundlegender Unterrichtsprozesse nichts Neues sind. Man denke zum Beispiel an Aebli (1983) *Zwölf Grundformen des Lehrens*. Neu an Osers Ansatz ist hingegen die Hinwendung zur Frage der Unterrichtssequenzierung. Jedes der elf Basismodelle weist eine innere Logik auf, die seinen jeweiligen Ablauf bestimmt. Diese Logik ist jedoch nicht unterrichtswissenschaftlicher, sondern lernpsychologischer Natur. Beispielsweise ist für die Gestaltung von Modell 6, dem Strategieerwerb, ein kognitives Modell wichtig, das spezifiziert, woraus neue Strategien abgeleitet werden: aus verbal vermittelten Lösungsregeln oder aus memorierten Lösungsbeispielen. Modell 7 kann ein Lehrer nur effektiv in die Praxis umsetzen, wenn er die verschiedenen Phasen des kognitiven

Fähigkeitserwerbs kennt. Kognitiven Modellen fällt also bei der Unterrichtsgestaltung die Rolle zu, Spezifikationen kausaler Verbindungen und zeitlicher Abläufe zu liefern, die dann vom Lehrer ausgestaltet und in konkrete Unterrichtsabläufe umgesetzt werden.

Mindestens vier der elf Basismodelle (Begriffsbildung, Konzeptbildung, Strategielernen und Routinebildung) erfordern detailliertes Wissen über den Erwerb konzeptuellen oder prozeduralen Wissens.

Aber nicht nur für die innere Strukturierung einzelner Basismodelle kann solches Wissen hilfreich sein, auch bei der Sequenzierung verschiedener Basismodelle wird es benötigt. Soll zum Beispiel in einem Inhaltsgebiet erst konzeptuelles Wissen vermittelt werden und dann prozedurales oder umgekehrt? Hier könnten psychologische Erkenntnisse zur Richtigkeit von concepts-first oder procedures-first-Theorien (siehe Abschnitt 5.1) direkt zur Anwendung kommen.

Man beachte, dass Oser den Erwerb von konzeptuellem und prozeduralem Wissen in verschiedene Basismodelle aufgeteilt hat. Es stellt sich die Frage, wie sich dies mit Baroodys Forderung nach der Integration von konzeptuellem und prozeduralem Wissen zur Erzielung adaptiver Expertise verträgt. Ist ein zusätzliches Basismodell erforderlich, in dem solche Verknüpfungen erlernt werden, oder ergeben sie sich beim Erwerb der einen oder der anderen Wissensart quasi von selbst?

Die Diskussion der Relationen zwischen konzeptuellem und prozeduralem Wissen im Kontext der Unterrichtswissenschaften führt also zu einer ganzen Reihe produktiver Forschungsfragen. Diese sind selbstverständlich nicht an Osers Theorie gebunden, sondern wurden hier nur an ihrem Beispiel verdeutlicht. So hebt beispielsweise auch Siegler (2003) in seiner ausführlichen Diskussion der Implikationen der kognitiven Forschung für den Mathematikunterricht die Wichtigkeit von konzeptuellem und prozeduralem Wissen hervor. Da die kognitiven Relationen zwischen den Wissensarten aber noch nicht vollständig bekannt sind, ist weitere Forschung notwendig.

5.6 Zusammenfassung

Theorien über die Relationen zwischen konzeptuellem und prozeduralem Wissen lassen sich unter anderem einteilen in den concepts-first view, den procedures-first view, den inactivation view, der die Wissensarten als unverbunden darstellt, und das Iterative Modell, das bidirektionale Einflüsse postuliert. Aussagen über spezifische Verarbeitungsmechanismen treffen die ACT-Theorie, die Theorie der representational redescription, sowie zahlreiche weitere Arbeiten. Sie erklären Erwerb, Abruf und Transfer der Wissensarten sowie ihre Umsetzung ineinander. Ein integratives Prozessmodell über alle Befunde liegt nicht vor. Aus unterrichtswissenschaftlicher Perspektive stellt sich

adaptive Expertise, die auf der Integration beider Wissensarten beruht, als sinnvolles Unterrichtsziel dar. Prozessmodelle über den Erwerb konzeptuellen und prozeduralen Wissens könnten zur theoriegeleiteten Unterrichtssequenzierung genutzt werden.

6 Inhaltsspezifisches konzeptuelles und prozedurales Wissen

6.1 Domänenspezifität des Wissenserwerbs

Wie bei der Beschreibung der Theorie der representational redescription angeführt, kann heute entwicklungspsychologisch nicht mehr, wie noch zu Piagets Zeiten, von einer domänenunspezifischen Kompetenzentwicklung ausgegangen werden. Zwar entwickeln sich auch domänenübergreifend wirksame psychische Komponenten, wie die Arbeitsgedächtniskapazität und damit die allgemeine Intelligenz, der Einfluss domänenspezifischen Wissens scheint jedoch gleich wichtig oder noch wichtiger für den Verlauf der Kompetenzentwicklung zu sein (Stern, 2004; Stern & Schneider, in Vorbereitung).

Die Relationen zwischen konzeptuellem und prozeduralem Wissen könnten in unterschiedlichen Inhaltsdomänen durchaus unterschiedlich sein. Ein Beispiel dafür wurde in der Literaturübersicht in Abschnitt 2.3.3 vorgestellt: Die bisher vorliegenden Befunde lassen vermuten, dass die Erwerbsreihenfolge für beide Wissensarten über Domänen hinweg variiert.

Eine Untersuchung der Wissensarten sollte darum immer auch reflektieren, in welchen Inhaltsgebieten sie sich bewegt, welche spezifischen Konzepte und Prozeduren es dort gibt und wie sich die Eigenschaften dieser Konzepte und Prozeduren auf Lernprozesse auswirken könnten.

In den beiden später vorgestellten empirischen Studien wird der Vergleichbarkeit wegen der selbe Aufgabentyp benutzt, der auch von Rittle-Johnson und Kollegen in ihrer Studie zum Iterativen Modell benutzt wurde: Es sollen die korrekten Positionen von Dezimalbrüchen auf einem Zahlenstrahl gefunden werden. Daher werden im Folgenden zwei Inhaltsdomänen kurz besprochen: Dezimalbrüche und Zahlenstrahle.

6.2 Dezimalbruchwissen

Dezimalbrüche werden umgangssprachlich gewöhnlich als *Kommazahlen* bezeichnet. In didaktischen Kontexten (vgl. *Rahmenlehrplan Grundschule: Mathematik*, 2004) wird erstere Bezeichnung bevorzugt, weil sie sprachlich auf zwei Punkte hinweist: Erstens sind Dezimalbrüche eine von mehreren Darstellungsweisen für *Brüche*. Zweitens spielt bei der Interpretation von Dezimalbrüchen das Verständnis des *Dezimalsystems* der Zahlendarstellung eine wichtige Rolle. Denn so wie die Stellen vor dem Komma Hunderter, Zehner und Einer repräsentieren, repräsentieren die Stellen nach dem Komma Zehntel, Hundertstel und so weiter.

Ob es um Gewichts- und Längenmaße, um Preise in Euro und Cent, um das Volumen von Getränkeverpackungen, um die Anzeigen auf digitalen Fieberthermometern oder um Zapfsäulendisplays an Tankstellen geht – Dezimalbrüche sind im Alltagsleben ubiquitär. Nichtsdestotrotz verstehen Kinder und Erwachsene Dezimalbrüche nicht immer vollständig. Eine Studie mit einer großen, für die USA repräsentativen Stichprobe fand Dezimalbruchmisskonzepte bei der Hälfte aller Siebtklässler (Kouba, Carpenter, & Swafford, 1989). Andere Studien konnten selbst bei Erwachsenen noch Fehlinterpretationen nachweisen (vgl. Rittle-Johnson et al., 2001). Die häufigsten Misskonzepte sind in Tabelle 6 aufgeführt.

Tabelle 6: Verbreitete Misskonzepte über Dezimalbrüche.

Name	Bedeutung
Ganze-Zahlen-Misskonzept	„Das Komma ist bedeutungslos und kann ignoriert werden.“
Bruchstrich-Misskonzept	„Das Komma funktioniert wie ein Bruchstrich, trennt also Zähler und Nenner.“
Kürzer-ist-größer-Misskonzept	„Kürzere Dezimalbrüche geben stets einen größeren Wert an als längere.“
Länger-ist-kleiner-Misskonzept	„Längere Dezimalbrüche geben stets einen kleineren Wert an als kürzere.“

Hiebert (1992) analysiert, was ein umfassendes korrektes Verständnis von Dezimalbrüchen ausmacht. Dabei unterscheidet er zwei Aspekte: das Notationssystem und die repräsentierten Quantitäten.

Ein Verständnis des Notationssystems beruht auf der Kenntnis dreier Prinzipien: (1) Der Wert einer Ziffer ist eine Funktion ihrer Position relativ zum Komma. Beispielsweise stellt die Ziffer vor dem Komma Einer dar, die Ziffer nach dem Komma jedoch Zehntel. (2) Der Wert einer Ziffer, d.h. die durch sie repräsentierte Quantität, ist das Produkt aus dem Nominalwert und dem Stellenwert. Steht beispielsweise eine Vier direkt hinter dem Komma, so repräsentiert sie den Wert $4 * 1/10 = 0,4$. Die Vier wird in dem Zusammenhang als Nominalwert bezeichnet, das Zehntel als Stellenwert. (3) Der Wert eines Dezimalbruchs ist die Summe der Werte seiner einzelnen Ziffern. Der Gesamtwert von 52,06 entspricht also $5 * 10 + 2 * 1 + 0 * 1/10 + 6 * 1/100$.

Neben diesem Verständnis des Notationssystems ist nach Hiebert jedoch auch ein Verständnis der repräsentierten Quantitäten erforderlich, um kompetent mit Dezimalbrüchen umgehen zu können. Dezimalbrüche können real oder hypothetisch existierende Quantitäten repräsentieren. Dabei ist vor allem die Einsicht wichtig, dass Dezimalbrüche sowohl Quantitäten bezeichnen können, die sich nur in ganzzahligen

Einheiten messen lassen, als auch kontinuierliche Quantitäten. Dezimalbrüche können kontinuierliche Quantitäten mit einer beliebigen Genauigkeit angeben. Ihre jeweilige Genauigkeit entspricht der Anzahl ihrer Nachkommastellen. Weil diese Anzahl prinzipiell unbegrenzt ist, gilt: (1) Dezimalbrüche können Quantitäten mit beliebiger Genauigkeit angeben. (2) Zwischen zwei gegebenen Dezimalbrüchen liegen stets unendlich viele weitere Dezimalbrüche.

Da prozedurales Wissen stets aufgabenspezifisch ist und Dezimalbrüche Teil vieler verschiedener Aufgabentypen sein kann, lassen sich hier, im Gegensatz zum konzeptuellen Wissen, nicht einige wenige ihm zugrunde liegende Regeln angeben. Prozedurales Wissen über Dezimalbrüche wurde bisher lediglich in den Domänen der Grundrechenarten untersucht. In der vorliegenden Arbeit wurden jedoch, wie auch von Rittle-Johnson et al. (2001), Aufgaben benutzt, die die Lokalisierung eines Dezimalbruchs auf einem Zahlenstrahl erfordern. Über die prozeduralen Anforderungen dieses Aufgabentyps ist bisher nichts bekannt, da Rittle-Johnson et al. (2001) ihre Daten lediglich quantitativ, aber nicht qualitativ auswerteten.

Hiebert und Wearne (1986) benutzen das Beispiel von Dezimalbruchaufgaben, um zu illustrieren, dass konzeptuelles und prozedurales Wissen zu drei Zeitpunkten während des Problemlöseprozesses interagieren können: bei der Aufgabeninterpretation, der Prozedurausführung und der Antwortäußerung.

6.3 Zahlenstrahlwissen

6.3.1 Wissen über die externe Wissensrepräsentation

Diagramme, wie beispielsweise der Zahlenstrahl, können in Unterricht und Alltagsleben wertvolle Hilfsmittel sein (z.B. Novick, 2000; Shah & Hoeffner, 2002). Pikturale Darstellungen, wie Fotos, sind zwar anschaulich und konkret, jedoch ungeeignet zur Darstellung abstrakter Relationen. Texte können abstrakte und fiktive Aussagen zum Ausdruck bringen, sind jedoch wesentlich weniger anschaulich und konkret als pikturale Darstellungen. Diagramme kombinieren die Vorteile beider Wissensrepräsentationsformen, indem sie abstrakte Zusammenhänge anschaulich darstellen und strukturieren (Hardy & Stadelhofer, unter Begutachtung; Stern et al., 2003).

Larkin und Simon (1987) erklären in ihrem bekannten Artikel „Why a Diagram is (Sometimes) Worth Ten Thousand Words“ die Effektivität von Diagrammen unter Zuhilfenahme der Konstrukte der Informations- und der Komputationseffizienz. Zwei verschiedene externe Wissensrepräsentationen besitzen eine gleiche Informationseffizienz, wenn man aus ihnen dieselben Informationen ableiten kann. Sie besitzen die gleiche

Komputationseffizienz, wenn man diese Informationen mit demselben Rechenaufwand aus ihnen extrahieren kann. Durch ausführliche Fallstudien zeigen Larkin und Simon, dass Diagramme, die denselben Informationsgehalt wie Texte oder Bilder haben, trotzdem oft eine höhere Komputationseffizienz besitzen können, weil sie die Informationen besonders übersichtlich, leicht enkodierbar und dadurch auch leicht transformierbar zur Darstellung bringen.

Eine Analyse von Wainer (1992) ergab, dass sich drei Stufen der Diagrammkompetenz unterscheiden lassen. Auf der ersten Stufe (*read data*) können nur konkrete Datenpunkte abgelesen, eingetragen und verbalisiert werden. Auf der zweiten Stufe (*read between data*) kann der Diagrammleser verschiedene Datenpunkte zueinander in Beziehung setzen, über sie interpolieren und Trends identifizieren. Auf der dritten Kompetenzstufe (*read beyond data*) kann man die dargestellten Daten zu seinem Hintergrundwissen oder im Aufgabenkontext implizit enthaltenen Informationen in Beziehung setzen und sie aus dieser übergeordneten Perspektive kommentieren und bewerten.

Während Diagramme generell von großem Nutzen sein können, haben Kinder bei einigen Typen von Diagrammen Schwierigkeiten sie korrekt zu verstehen. Dies trifft vor allem auf Koordinatensysteme zu und auf andere Diagrammtypen, die mehrdimensionale Zusammenhänge darstellen (Guthrie, Weber, & Kimmerly, 1993; Leinhardt, Zaslavsky, & Stein, 1990). Der Zahlenstrahl als Veranschaulichung einer einzelnen Dimension wird hingegen auch von jungen Kindern gut verstanden und effizient genutzt. Sommerville und Bryant (1985) sowie Gattis (2002) untersuchten jeweils mit Kindern im Vorschulalter, unter welchen Bedingungen sie Koordinatenpaare oder Grafensteigungen in zweidimensionalen Koordinatensystemen finden und interpretieren können. Dabei zeigte sich, dass die Kinder mit einem Zahlenstrahl, also einer einzelnen Achse des Koordinatensystems, problemlos umgehen können. Bei geeigneter Aufgabenstellung waren sie sogar in der Lage, die Informationen der X- und der Y-Achse zu kombinieren und dadurch Wertepaare im Koordinatensystem aufzufinden.

Ein Vergleich von englischen Erst-, Zweit- und Drittklässlern zeigte, dass Kinder in der ersten Klasse beim Arbeiten mit dem Zahlenstrahl vor allem die ordinalen Beziehungen zwischen Zahlen zur Lokalisation von (natürlichen) Zahlen auf dem Strahl benutzen. Im Laufe der zweiten und der dritten Klasse nimmt dann graduell der Anteil an proportionalen Strategien zu, bei dem nicht nur die Reihenfolge der Zahlen, sondern auch ihre intervallskalierten Abstände mitberücksichtigt werden. Sind beispielsweise die Drei

und die Fünf auf einem Zahlenstrahl markiert und ein Kind sollte zusätzlich die Position der Vier eintragen, so würden Nutzer der ordinalen Strategien die Vier *irgendwo* rechts von der Drei und links von der Fünf eintragen. Nutzer von proportionalen Strategien würden hingegen die Position in der Mitte zwischen den beiden vorgegebenen Markierungen wählen, weil die Vier in der Mitte zwischen der Drei und der Fünf liegt (Petitto, 1990).

Rittle-Johnson et al. (2001) weisen darauf hin, dass Fünft- und Sechstklässler zwar mit dem Zahlenstrahl, nicht jedoch mit Dezimalbrüchen sicher umgehen können. Daher könnten Zahlenstrahle gut geeignet sein, um in dieser Altersgruppe Wissen über Dezimalbrüche zu veranschaulichen und zu vermitteln. Die von ihnen nachgewiesenen Wissenszuwächse während des „Fang-das-Monster“-Spiels bestätigen die Richtigkeit dieser Annahme.

6.3.2 Passung von externer und interner Wissensrepräsentation

Eine Erklärung, die Rittle-Johnson et al. (2001) für die Effektivität ihres „Fang-das-Monster“-Spiels geben, ist, dass Kinder Zahlen geistig unter anderem auf einem mentalen Zahlenstrahl repräsentieren (siehe Abschnitt 4.6.2). Das „Fang-das-Monster“-Spiel beinhaltet das Arbeiten mit dem externen Zahlenstrahl. Diese Übereinstimmung von externer und interner Wissensrepräsentation könnte sich förderlich auf den Wissenserwerbsprozess ausgewirkt haben. Empirische Belege für diese Hypothese fehlen jedoch bislang.

6.4 Zusammenfassung

Über den Ablauf kognitiver Prozesse beim Lernen mit Dezimalbrüchen und Zahlenstrahlen ist weniger bekannt als über die Kompetenzentwicklung auf diesen Gebieten. Schon im Vorschulalter können Kinder kompetent mit dem Zahlenstrahl umgehen und ihn zur Darstellung natürlicher Zahlen nutzen. Während der Grundschulzeit nutzen sie immer seltener ordinale und immer häufiger proportionale Strategien der Positionsbestimmung von Zahlen auf dem Strahl. Im Gegensatz dazu können Fünft- und Sechstklässler, wie sogar Erwachsene noch, nicht vollständig korrekt mit Dezimalbrüchen umgehen. Die Schwierigkeiten beruhen auf Fehlvorstellungen bezüglich der Notation, also der Bedeutung von Komma, Nominal- und Stellenwert, sowie Fehlvorstellungen bezüglich der Eigenschaften der repräsentierten Quantitäten, beispielsweise ihrer Ordinalrelationen, Kontinuität und Genauigkeit.

7 Validierung kognitiver Modelle mittels Strukturgleichungsmodellen

7.1 Pfadmodelle und Strukturgleichungsmodelle

Traditionelle Regressionsmodelle analysieren den Einfluss einer oder mehrerer unabhängiger Variablen auf eine abhängige Variable. Sie sind darum suboptimal, wenn Modelle mit mehreren abhängigen Variablen, mit erhobenen Mediatorvariablen oder gar Rückkopplungsschleifen zwischen Variablen auszuwerten sind. In diesen Fällen sind *Pfadmodelle* vorzuziehen. Pfadmodelle können Kombinationen mehrerer univariater Regressionen enthalten, die gleichzeitig analysiert werden, und sind so flexibel und effizient. Darüber hinaus lassen sich Pfadmodelle sehr übersichtlich in Pfaddiagrammen veranschaulichen.

Wesentlich erweitert wurde die Leistungsfähigkeit von Pfadmodellen durch ihre Kombination mit konfirmatorischen Faktoranalysen (KFA) zur Technik der *Strukturgleichungsmodelle* (*structural equation models*, SEM) durch Jöreskog (1973) und anderen. Im Gegensatz zu Pfadmodellen können Strukturgleichungsmodelle neben beobachteten so genannten *manifesten Variablen* auch *latente Variablen* enthalten, deren Ausprägungen nicht direkt gemessen wurden, sondern durch konfirmatorische Faktoranalysen über gemessene *Indikatoren* bestimmt werden.

Die Bestimmung der freien Parameter des Modells, also der Varianzen, Kovarianzen, Faktorladungen und Pfadkoeffizienten, auf Grundlage der gemessenen Daten zusammen mit der Bestimmung der Modellgüte, des so genannten *Modellfits*, wird als *Parameterschätzung* bezeichnet und unter Zuhilfenahme iterativer Näherungsalgorithmen durchgeführt.

Der Teil eines Strukturgleichungsmodells, der die konfirmatorischen Faktoranalysen spezifiziert, wird auch als *Messmodell* bezeichnet. Der Teil eines Gesamtmodells, der dem Pfadmodell entspricht, wird auch *Regressionsmodell* genannt. Die Charakteristika dieser Teilmodelle werden in den nächsten beiden Unterkapiteln besprochen. Nachfolgend werden Modellierungsstrategien und Modelinterpretation diskutiert (Gesamtüberblicke geben Bollen, 1989; Kaplan, 2000).

7.2 Das Messmodell

Die Faktoren des Messmodells werden im Zuge von SEM als latente Variablen konzeptualisiert. Im Gegensatz zur explorativen Faktoranalyse ist bei der konfirmatorischen a priori festzulegen, welche Indikatoren auf welchen Faktoren laden

dürfen. Die konkreten Stärken der Faktorladungen können dabei festgelegt oder alternativ dazu aus den empirischen Daten geschätzt werden. Weitere Unterschiede zur explorativen Faktoranalyse sind, dass in konfirmatorischen Analysen eine Reihe von *Modellfitindizes* berechnet werden können, die Aufschluss über die Güte des spezifizierten Modells geben. Da bei konfirmatorischen Faktoranalysen per definitionem Faktorladungen a priori festgelegt werden, sind keine Faktorrotationen, wie beispielsweise die Varimax-Rotation, durchführbar, denn diese würden die festgelegten Ladungen verändern.

Die Möglichkeit der Modellierung latenter Variablen ist aus mehreren Gründen attraktiv. So ist, gerade in der Psychologie, kaum ein einzelner Indikator ein perfektes Maß für ein Konstrukt, sondern jeder ist mit einem gewissen Maß an Fehlervarianz behaftet. Ein Faktorwert, der auf Grundlage mehrerer Indikatoren bestimmt wird, spiegelt nur den Varianzanteil wider, der all seinen Indikatoren gemeinsam ist. Er ist dadurch ein wesentlich reliableres und valideres Maß für ein Konstrukt als ein Einzelindikator es sein könnte.

Die Ladungen der Indikatoren auf ihren jeweiligen Faktoren geben darüber hinaus Aufschluss über die jeweilige Indikatorreliabilität. Die Reliabilität einer Messung ist definiert als Anteil der wahren Varianz an der gemessenen Varianz. Die quadrierten Faktorladungen geben, ähnlich wie der Determinationskoeffizient R^2 bei der Regression, den durch einen Faktor aufgeklärten Anteil der Varianz jedes Indikators an. Da die Indikatorvarianz die gemessene Varianz darstellt und die Faktorvarianz ein sinnvoller Schätzer der wahren Merkmalsvarianz ist, können die quadrierten Faktorladungen der Indikatoren als Reliabilitäten der Indikatoren betrachtet werden (Bollen, 1989, S. 222). Der Fall verkompliziert sich jedoch, wenn ein Indikator zur Bestimmung mehrerer interkorrelierter Faktoren genutzt wird.

Sofern die Indikatoren Summenscores darstellen, lässt sich die Untergrenze ihrer jeweiligen Reliabilität darüber hinaus über Cronbachs Alpha-Koeffizienten als Maß der internen Konsistenz abschätzen.

Das Messmodell liefert auch Informationen über die Validitäten der Indikatoren. Hierbei sind, neben der, wie stets, durch theoretische Abwägungen zu klärenden Inhaltsvalidität, vor allem die Konstrukte der *konvergenten* und der *divergenten* (auch: *diskriminanten*) *Validität* von großem Nutzen. Sie wurden von Campbell und Fiske (1959) im Rahmen des multitrait-multimethod-Ansatzes entwickelt, sind jedoch nicht auf diesen beschränkt. Konvergente Validität besteht immer dann, wenn verschiedene Indikatoren, die dasselbe Konstrukt operationalisieren sollen, hoch interkorreliert sind, sich also gut auf einen latenten Faktor abbilden lassen. Divergente Validität besteht dann, wenn Maße, die

nicht dasselbe Konstrukt operationalisieren, relativ niedrig interkorreliert sind. Natürlich dürfen sie korrelieren, weil ja auch die gemessenen Konstrukte zusammenhängen können. Wünschenswert ist aus messtheoretischer Sicht aber, dass verschiedene Maße desselben Konstrukts höher miteinander korrelieren als die Maße unterschiedlicher Konstrukte miteinander, das heißt konvergente und divergente Validität sollten in einem sinnvollen Verhältnis zueinander stehen (Bollen, 1989, S. 190ff).

Die konvergente Validität der Indikatoren eines latenten Faktors kann darüberhinaus durch Cronbachs Alpha abgeschätzt werden – wenn nämlich der Faktor als Skala und die Faktorindikatoren als Items dieser Skala aufgefasst werden. Die interne Konsistenz der Skala entspricht dann der konvergenten Validität der Indikatoren.

Sollen die Ausprägungen eines latenten Konstrukts über verschiedene Zeitpunkte oder verschiedene Subpopulationen hinweg verglichen werden, ist zuvor die *Messinvarianz* des Messmodells zu überprüfen. Sind die unstandardisierten Ladungen der Indikatoren auf ihren jeweiligen Faktoren über Messzeitpunkte oder Personengruppen hinweg nicht voneinander verschieden, so spricht man von *metrischer Messinvarianz*. Diese und weitere Typen der Messinvarianz werden beispielsweise von Cheung und Rensvold (2002) diskutiert. Liegt keine Messinvarianz vor, wirft das die Frage auf, inwieweit die Faktoren Operationalisierungen desselben Konstrukts darstellen und über Messzeitpunkte oder Subpopulationen hinweg miteinander verglichen werden können.

Die Modellierung latenter Variablen hat gegenüber der Analyse manifester Variablen also mindestens drei Vorteile: (1) Durch die Faktorisierung wird Fehlervarianz reduziert, wodurch Reliabilität und Validität der Messung steigen. (2) Durch Analyse der quadrierten Faktorladungen können die Reliabilitäten der manifesten Faktorindikatoren explizit bestimmt werden. (3) Durch konfirmatorische Analysen können die konvergente und die divergente Validität verschiedener Indikatoren überprüft und verglichen werden.

Das verdeutlicht, warum die Technik der konfirmatorischen Faktoranalyse unter anderem auch in der psychologischen Forschung zunehmend häufiger benutzt wird (Bollen, 2002). Carroll (1993) gibt in einer mehr als 700 Seiten starken Monografie einen beeindruckenden Überblick über die wichtigsten faktoranalytischen Befunde zur Struktur kognitiver Fähigkeiten, verbunden mit sorgfältigen vergleichenden Reanalysen der Originaldaten. Er geht ein auf die Domänen Sprache, Schlussfolgern, Gedächtnis und Lernen, visuelle Wahrnehmung, auditive Wahrnehmung, Ideengenerierung, cognitive speed, Wissen und Leistung sowie psychomotorische Fähigkeiten. Faktoren für unterschiedliche Wissensarten im oben besprochenen Sinne wurden dabei bisher nicht

gefunden. Der Grund dafür ist jedoch einfach: Es können nur Faktoren gefunden werden, für die Indikatoren erhoben wurden. Die von Carroll vorgestellten Studien untersuchten basale kognitive Fähigkeitsdimensionen und benutzten keine Indikatoren, die zur Messung übergreifender Wissensstrukturen geeignet gewesen wären.

7.3 Das Regressionsmodell

Das Regressionsmodell spezifiziert die Regressions- und Kovarianzbeziehungen zwischen den (manifesten und latenten) Variablen. Das Regressionsmodell darf beliebig komplex sein, solange es identifiziert ist (s.u.). Das Hauptproblem bei der Spezifikation des Regressionsmodells besteht darin, das Modell so zu gestalten, dass es die Überprüfung der Forschungshypothesen erlaubt, empirisch abgesicherte Hintergrundbedingungen mitberücksichtigt und keine unrealistischen Annahmen abbildet. Man unterscheidet zwischen *internen* und *externen Spezifikationsfehlern* (z.B. Kaplan, 1990). Bei internen Spezifikationsfehlern werden Beziehungen zwischen Variablen im Modell spezifiziert, die in der Realität nicht bestehen. Diese Fehler können im Laufe des Modellierungsprozesses durch Vergleiche der postulierten mit den empirisch gefundenen Beziehungen gefunden und behoben werden. Bei externen Spezifikationsfehlern werden Variablen nicht gemessen beziehungsweise nicht in das Modell aufgenommen, obwohl sie am zu untersuchenden Phänomen beteiligt sind. Solche Fehler können die Ergebnisse verzerren und sind statistisch nicht aufzuspüren. Sie können nur durch sorgfältige theoriegeleitete Modellkonstruktion vermieden werden (Kaplan, 2000, S. 96ff).

Der wohl häufigste Fehler im Umgang mit Regressionsmodellen ist die Interpretation von Regressionsbeziehungen als Kausalbeziehungen (Bollen, 1989, S. 40-79; Raykov, Tomer, & Nesselroade, 1991). Die Analyse von Strukturgleichungsmodellen wird häufig mit großen Datensätzen, zum Beispiel von large-scale-Schülerleistungsstudien, durchgeführt, die lediglich Einzelgruppen- oder Querschnittsdaten bereitstellen und keine Kausalanalysen erlauben. Die Gerichtetheit der Regressionspfade in den SEM führt jedoch immer wieder dazu, dass die Signifikanz eines solchen Pfades als Nachweis einer Kausalbeziehung interpretiert wird. Eine signifikante Regressionsbeziehung zeigt in Wirklichkeit zunächst jedoch nur, dass zwei Variablen miteinander korreliert sind, und von einer Korrelation kann bekanntlich nicht auf Kausalität geschlossen werden. Die Korrelation verrät nichts über die Kausalrichtung. Sie kann außerdem auch ohne jegliche Kausalbeziehung zwischen den Variablen auftreten, wenn nämlich beide durch mindestens eine dritte, unanalysierte, Variable beeinflusst werden (Bortz & Döring, 1995, S. 483-489), wenn also ein externer Spezifikationsfehler vorliegt, der grundsätzlich nie ganz ausgeschlossen werden kann.

Kritiker deuten dieses Problem oft als grundsätzliche Schwäche des SEM-Ansatzes, was jedoch nicht korrekt ist. In den meisten Studien mit Strukturgleichungsmodellen können keine Kausalbeziehungen analysiert werden, weil sie auf querschnittlichen Designs beruhen, die per se keinen Aufschluss über Kausalbeziehungen geben können. Werden experimentelle oder längsschnittliche Daten mit Strukturgleichungsmodellen analysiert, sind durchaus Kausalanalysen möglich (MacCallum & Austin, 2000; Muthén & Curran, 1997). Dies wird leicht einsichtig, wenn man sich vergegenwärtigt, dass Experimente normalerweise mittels Varianzanalysen ausgewertet werden. Diese stellen einen Spezialfall des Allgemeinen Linearen Modells dar, das auf Regressionen beruht und somit leicht als Strukturgleichungsmodell formuliert werden kann. Der Hauptgrund, aus dem die Varianzanalysen gewöhnlich den flexibleren und umfassenderen Strukturgleichungsmodellen zur Analyse experimentell gewonnener Daten vorgezogen werden, ist, dass Erstere wesentlich kleinere Stichprobenumfänge erfordern als Letztere.

Wenn Strukturgleichungsmodelle zu Kausalanalysen benutzt werden, dann oft im Rahmen der Auswertung von *cross-lagged panel designs*. Wie Bortz und Döring (1995, S. 485f) beschreiben, setzen sie längsschnittliche Erhebungen voraus, bei denen mindestens zwei Variablen an mindestens zwei Messzeitpunkten (Mzp) erhoben wurden.

Als Beispiel beschreiben Bortz und Döring eine Studie über die wechselseitigen Einflüsse zwischen Einkommen und Bildungsniveau. Um die Frage zu beantworten, ob das Einkommen das Bildungsniveau kausal beeinflusst oder das Bildungsniveau das Einkommen, wurden beide Variablen einmal gemessen, als die Studienteilnehmer 25 Jahre alt waren, und einmal, als sie 50 Jahre alt waren.

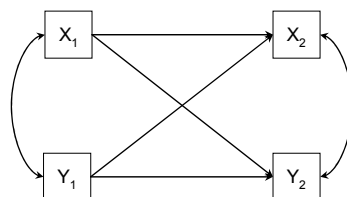


Abbildung 1: Pfaddiagramm eines cross-lagged panel designs.

Zur Auswertung kann dann das in Abbildung 1 dargestellte Modell verwendet werden. Die Variable x steht für das Bildungsniveau und die Variable y für das Einkommen, die Indizes geben den Messzeitpunkt an. Würde nun gefunden, dass der Einfluss von x_1 auf y_2 (bei Kontrolle des Einflusses von y_1 auf y_2) wesentlich größer ist als der Einfluss von y_1 auf x_2 (bei Kontrolle des Einflusses von x_1 auf x_2), so wäre das sehr gut mit der Hypothese

vereinbar, dass das Bildungsniveau das Einkommen kausal determiniert, aber nicht mit der Hypothese, dass das Einkommen das Bildungsniveau kausal determiniert. MacCallum und Austin (2000) weisen darauf hin, wie wichtig es ist, in diesem Design die autoregressiven Relationen (von x_1 auf x_2 und von y_1 auf y_2) zu kontrollieren. Würde man beispielsweise x_2 nur durch y_1 vorhersagen, anstatt durch x_1 und y_1 gemeinsam, könnte man den Einfluss von y_1 auf x_2 stark überschätzen, sofern y_1 und x_1 korreliert sind.

Bortz und Döring heben einerseits die Nützlichkeit dieses Designs, dessen Validität durch zusätzliche Messzeitpunkte noch weiter erhöht werden kann, hervor, betonen aber auch, dass es sich trotz allem nur um ein korrelatives und nicht um ein experimentelles Design handelt. Wenn die längsschnittlichen Veränderungen der Variablen im Modell durch weitere, nicht im Modell enthaltene verursacht werden, wenn also ein externer Spezifikationsfehler vorliegt, können signifikante Beziehungen von x_1 zu y_2 oder y_1 zu x_2 auftreten, auch wenn x und y sich nicht kausal beeinflussen. Das Design ist somit vor allem geeignet, die Plausibilität zweier gegensätzlicher Kausalhypothesen (x beeinflusst y versus y beeinflusst x) *relativ zueinander* zu bestimmen.

Burkholder und Harlow (2003) geben ein Beispiel für ein komplexes und mittels SEM ausgewertetes cross-lagged panel design mit drei Messzeitpunkten und diskutieren es ausführlich unter methodischen Gesichtspunkten.

Neben einfachen Regressionsdesigns und cross-lagged panel designs gibt es viele weitere Modelltypen, die sich in einem SEM-Framework verwirklichen lassen, unter anderem Wachstumskurvenmodelle, MIMIC-(multiple indicators multiple sources-) Modelle, item response-Modelle und verschiedene Klassen von Mehrebenenmodellen. Es sind auch Mischformen der Typen möglich (Muthén & Muthén, 1998-2004b).

7.4 Modellierungsstrategien

Die Analyse von Strukturgleichungsmodellen besteht aus einer Abfolge typischer Arbeitsschritte (Bentler, 1996; Bollen, 1989; Kaplan, 2000; Raykov et al., 1991). Zunächst sind die zu erhebenden Variablen und die zu analysierenden Modelle auszuwählen. MacCallum und Austin (2000) unterscheiden hier drei mögliche Vorgehensweisen: (1) die strikt konfirmatorische Vorgehensweise, bei der ein einzelnes a priori spezifiziertes Modell überprüft wird; (2) die modellgenerierende Vorgehensweise, bei der ein a priori-Modell überprüft und dann so lange modifiziert wird, bis es optimal an die Daten angepasst ist; und (3) die modellvergleichende Vorgehensweise, bei der theoriegeleitet mehrere a priori-Modelle spezifiziert und dann gegeneinander getestet werden. Die Vorgehensweisen 2 und 3 können auch miteinander kombiniert werden.

Da jedes Modell immer nur eine vereinfachende Annäherung an die Realität darstellt und nur Teilausschnitte der Realität abbilden kann, gibt es nicht das einzig wahre Modell, sondern unterschiedliche Modelle sind mit den empirisch gefundenen Daten mehr oder weniger gut vereinbar. Daher ist die modellvergleichende Vorgehensweise den anderen beiden Vorgehensweisen gewöhnlich vorzuziehen. Die Vorgehensweise erfordert nicht zwangsläufig explizite Vergleiche verschiedener Modelle, sondern kann auch darin bestehen theoriegeleitet zu testen, ob Pfadkoeffizienten in einem einzelnen Modell signifikant sind oder nicht.

Bei der Modellauswahl ist zu beachten, dass nur die Parameter von solchen Modellen geschätzt werden können, die empirisch identifiziert sind. Beispielsweise wird die Lage einer Geraden durch zwei Punkte festgelegt. Angesichts eines einzelnen Punktes ist ihre Lage unterdeterminiert. In gleicher Weise ist bei der Parameterschätzung eines Strukturgleichungsmodells das Verhältnis der Anzahl von den gegebenen zu den zu schätzenden Parametern entscheidend. Hier gibt es aufgrund der Komplexität des Schätzungsprozesses keine eindeutige Regel. Eine nützliche Heuristik, die so genannte *Abzählregel* (*counting rule*, *t-rule*), besagt, dass ein Modell identifiziert ist, wenn gilt:

$$t \leq \frac{s(s+1)}{2}$$

wobei s die Anzahl der gemessenen Variablen und t die Anzahl der zu schätzenden Parameter (Varianzen, Kovarianzen, Faktorladungen, Pfadkoeffizienten und Fehlerterme) im Modell ist (Kaplan, 2000, S. 56-58).

Wenn man latente Variablen modelliert, ist zusätzlich zu beachten, dass die Metrik dieser Variablen festgelegt werden muss, damit das Modell identifiziert ist. Wurden beispielsweise für einen Faktorindikator die Ausprägungen 1, 2 und 3 gemessen und für einen anderen Indikator desselben Faktors die Ausprägungen 2, 4 und 6, so könnten die Faktorwerte 1, 2, 3 lauten oder 2, 4, 6 oder 3, 6, 9 und so weiter. Weil die Beziehungen zwischen Indikатораusrprägungen und Faktorwerten korrelativ sind, Korrelationen aber streuungsinvariant sind, ist die Varianz der Faktorwerte angesichts der Indikatoren unterdeterminiert. Dieses Problem lässt sich auf zwei Weisen lösen: Man kann spezifizieren, dass der Faktor die gleiche Varianz wie einer der Indikatoren haben soll, indem man eine der unstandardisierten Faktorladungen auf den Wert 1 fixiert. Alternativ dazu kann man die Faktorvarianz auf den Wert 1 fixieren. Die Wahl der Alternative kann die Ergebnisse der Modellparameterschätzung beeinflussen und ist bei der Publikation der Ergebnisse anzugeben (z.B. Bollen, 2002). Selbiges gilt auch für die Mittelwerte der latenten Faktoren, sofern sie im Modell variieren können.

Nach der Datenerhebung sind die Daten zu bereinigen. Ausreißer können die Parameterschätzung verzerren und sollten eliminiert werden. Einige Schätzalgorithmen können mit *missing data* umgehen, andere nicht. Nach der Entscheidung für einen Schätzalgorithmus sind gegebenenfalls Fälle mit fehlenden Daten von der Analyse auszuschließen. In großen Datensätzen können die fehlenden Daten alternativ durch *multiple imputations* geschätzt werden (Raykov et al., 1991).

Im nächsten Arbeitsschritt sind die deskriptiven Eigenschaften der Daten zu überprüfen. Die Mittelwerte und Varianzen der manifesten Variablen sollten im Bereich derselben Größenordnung liegen, um Konvergenzprobleme zu vermeiden (Muthén & Muthén, 1998-2004b, S. 321).

Von großer Wichtigkeit ist die Überprüfung der Verteilungseigenschaften der manifesten Variablen. Der älteste und aufgrund mehrerer Vorzüge am häufigsten genutzte Schätzalgorithmus zur Parameterbestimmung ist der *maximum-likelihood estimator* (Rost, 2004, S. 303-308). Seine Anwendung setzt allerdings die multivariate Normalverteilung der gemessenen Variablen voraus. Notwendige, aber nicht hinreichende Bedingung für multivariate Normalverteilung ist die univariate Normalverteilung der einzelnen Variablen. Ist keine multivariate Normalverteilung gegeben, kann einer der anderen inzwischen entwickelten Schätzalgorithmen genutzt werden. Übersichten über sie und ihre Eigenschaften geben unter anderem Muthén und Muthén (1998-2004b, S. 398-402) sowie Bollen (1989, S. 104-123).

Eine positive Eigenschaft des maximum-likelihood estimators ist, dass er den Modellfit in Form eines Chi-Quadrat-Wertes bestimmt (Rost, 2004, S. 336). Auf Grundlage dieser Werte ist der Chi-Quadrat-Differenz-Test (Steiger, Shapiro, & Browne, 1985) möglich, der die Fits zweier Modelle miteinander vergleicht. Dabei muss das eine Modell durch Hinzufügen von Restriktionen, zum Beispiel der Fixierung von freien Parametern, aus dem anderen ableitbar sein. Das restriktivere der beiden Modelle wird dann als *genestetes (geschachteltes) Modell* bezeichnet, weil es einen Spezialfall darstellt, der unter das allgemeinere Modell fällt. Man kann dann mit Hilfe eines Chi-Quadrat-Differenz-Tests die Hypothese überprüfen, dass das genestete Modell (aufgrund der zusätzlichen Restriktionen) einen schlechteren Modellfit hat als das ursprüngliche Modell. Falls dies signifikant der Fall ist, sind die zusätzlichen Restriktionen zu verwerfen und das ursprüngliche Modell ist beizubehalten (Bollen, 1989, S. 289-291).

Der Chi-Quadrat-Differenz-Test ist eines der wichtigsten Mittel zum Modellvergleich und zur Modelloptimierung. Das ist einer der Gründe, warum der maximum-likelihood

estimator anderen Methoden gegenüber häufig vorgezogen wird, denn diese lassen gewöhnlich den Test nicht zu.

Vor diesem Hintergrund ist ein weiterer Schätzalgorithmus von besonderer Bedeutung, nämlich der maximum-likelihood estimator mit *Satorra-Bentler scaling correction* (*sb-Skalierung*; estimator MLM). Er hat gegenüber dem einfachen maximum-likelihood estimator den Vorteil, auch für nicht normalverteilte Daten geeignet zu sein. Gegenüber alternativen Schätzalgorithmen hat er den Vorteil, dass es Satorra und Bentler (1994; 1999) gelungen ist, einen auf der sb-skalierten Chi-Quadrat-Statistik beruhenden Differenztest zu entwickeln. Nevitt und Hancock (2004) zeigten in einer umfassenden Simulationsstudie, dass der maximum-likelihood estimator mit sb-Skalierung anderen Algorithmen in der Analyse nicht normalverteilter Daten überlegen ist – und zwar auch bei relativ kleinen Stichprobenumfängen.

Letzteres ist wichtig, da für die Analyse von Strukturgleichungsmodellen Stichprobenumfänge von minimal 200 und optimal über 500 empfohlen werden. Aufgrund der mathematischen Komplexität der Parameterschätzung lassen sich hierzu jedoch keine exakten Angaben analytisch herleiten. Der erforderliche Stichprobenumfang hängt stark von Effektstärken, Datenqualität und Modellkomplexität ab und müsste eigentlich durch Simulationsstudien für jedes Modell einzeln bestimmt werden, was in der Praxis jedoch selten geschieht.

Erst nach Abklärung all dieser Punkte, nach einer Entscheidung für die modellkonfirmatorische, modellgenerierende oder modellvergleichende Vorgehensweise, der Berücksichtigung von fehlenden Daten und Durchführung einer Ausreißerbereinigung, nach Überprüfung der Descriptives und der Verteilungseigenschaften der manifesten Variablen sowie der Entscheidung für einen Schätzalgorithmus kann mit der eigentlichen Parameterschätzung begonnen werden.

Die Schätzung wird stets durch Anwendung iterativer Näherungsalgorithmen durchgeführt. Diese werden beendet, wenn ein Konvergenzkriterium erfüllt ist, das das Erreichen einer durch den Algorithmus nicht weiter verbesserbaren Lösung anzeigt. Der Algorithmus findet eine solche Lösung jedoch nicht zwangsläufig. Es ist möglich, dass in endlicher Zeit keine Lösung gefunden wird, die das vorgegebene Konvergenzkriterium erfüllt. Das kann verschiedene Ursachen haben. Eine der häufigsten Ursachen ist, dass versucht wird, die Parameter eines Modells zu schätzen, das mit den empirischen Daten nur schlecht zu vereinbaren ist. Eine schlechte Datenqualität (zu kleiner Stichprobenumfang, schlechte Verteilungseigenschaften der Maße, viele missing values)

kann ebenfalls ursächlich sein. Bei Konvergenzproblemen sollte ein Modell verworfen werden, sofern konvergierende Alternativmodelle zur Verfügung stehen.

Insgesamt empfiehlt es sich, nie gleich ein komplexes Modell auf einmal zu schätzen, da dann ein eventueller schlechter Modellfit nicht eindeutig Teilen des Mess- oder des Regressionsmodells zugeschrieben werden kann. Es ist sinnvoll, zuerst das Messmodell alleine zu überprüfen und dann, wenn die Operationalisierbarkeit der latenten Variablen nachgewiesen wurde, das Regressions- beziehungsweise das Gesamtmodell zu testen. Weitere ergänzende Analysen, wie explorative Faktoranalysen über die manifesten Variablen, können sinnvoll sein, weil sie helfen modellinterne Spezifikationsfehler aufzudecken.

Der letzte Arbeitsschritt ist die Interpretation der Ergebnisse, der bei einer modellgenerierenden Vorgehensweise erneute Modellierungszyklen folgen können.

7.5 Modellgütebeurteilung und Modellinterpretation

Bei der Beurteilung und Interpretation der Ergebnisse der Parameterschätzung sind zwei Gruppen von Informationsquellen zu berücksichtigen: die geschätzten Ausprägungen der Modellkoeffizienten und die Modellfitindizes.

Die Modellkoeffizienten geben Aufschluss über spezifische Relationen zwischen den Variablen im Modell. Sie sind auf ihre Augenscheinvalidität und ihre theoretischen Implikationen hin zu interpretieren. Die meisten Programme geben sowohl standardisierte als auch unstandardisierte Koeffizienten aus. Die Unstandardisierten beziehen sich auf die manifesten Variablen in ihrer ursprünglichen Skalierung, die Standardisierten auf alle Variablen in ihrer z-standardisierten Form. Um eine Vergleichbarkeit von Koeffizienten zwischen Studien mit unterschiedlich skalierten Variablen zu gewährleisten, werden meistens die standardisierten Koeffizienten berichtet und interpretiert. Standardisierte Pfadkoeffizienten können wie standardisierte Regressionskoeffizienten interpretiert werden: Sie geben an, um wie viele Standardabweichungen sich die Ausprägungen der abhängigen Variable verändern, wenn die Ausprägungen der unabhängigen Variable um eine Standardabweichung zunehmen. Daher liegen standardisierte Pfadkoeffizienten normalerweise zwischen minus eins und plus eins, können jedoch auch größere und kleinere Werte annehmen.

Die Modellfitindizes (für einen ausführlichen Überblick siehe Yu, 2002) geben an, wie gut das vorgegebene Modell zu den empirisch gefundenen Daten passt. Allerdings ist gute Passung ein Begriff, der sich auf viele unterschiedliche Weisen konzeptualisieren und quantifizieren lässt. Dementsprechend gibt es nicht einen Fitindex, sondern eine ganze Reihe von ihnen. Jeder hat unterschiedliche Vor- und Nachteile und keiner kann als der

einzig korrekte angesehen werden. Im Idealfall deuten unterschiedliche Modellfitindizes bei der Modellevaluation in dieselbe Richtung. Das ist jedoch nicht zwangsläufig der Fall, weshalb bei der Publikation von SEM-Analysen stets mehrere Modellfitindizes berichtet werden sollten (z.B. Boomsma, 2000).

Ein Indikator des Modelfits ist die empirische Irrtumswahrscheinlichkeit des *Goodness-of-Fit-Chi-Quadrat-Tests*. Er leitet aus dem Chi-Quadrat-Anpassungswert die Wahrscheinlichkeit p ab, die empirisch gefundenen Daten zu erhalten unter der Voraussetzung, dass das spezifizierte Modell in der Population gilt (Rost, 2004, S. 303). Je größer diese Wahrscheinlichkeit, umso besser ist der Modellfit. Jedoch hängt die Wahrscheinlichkeit nicht nur vom Modellfit ab, sondern auch von Stichprobengröße und Modellkomplexität, so dass sie zwar standardmäßig berichtet, jedoch nur selten explizit interpretiert wird.

Wie oben beschrieben, sind der Chi-Quadrat-Wert und auf ihm basierende Signifikanztests nützlich, um zwei konkurrierende Modelle zu vergleichen. Um den Fit eines einzelnen Modells zu evaluieren, sind sie hingegen suboptimal, weil der Chi-Quadrat-Wert, also die Abweichung zwischen empirischen und erwarteten Daten, immer kleiner wird, je mehr zu schätzende Parameter in das Modell aufgenommen werden. Das komplexeste aller möglichen Modelle hat also immer den besten Modellfit. Außerdem wird der Test umso eher signifikant, je größer die Stichprobe ist. Bei großen Stichproben können darum schon sehr kleine Abweichungen zwischen erwarteten und gefundenen Werten zu einer Ablehnung des Modells führen.

Die *komparativen Fitindizes* vergleichen den Fit des geschätzten Modells mit dem Fit eines *baseline models*, das möglichst einfache Beziehungen zwischen den Variablen im Modell, zum Beispiel ihre wechselseitige Unabhängigkeit, spezifiziert. Ein wichtiger Vertreter der komparativen Fitindizes ist der *comparative fit index* (CFI), der auf der Weiterentwicklung älterer Indizes beruht (Kaplan, 2000, S. 107-109) und sich in empirischen Untersuchungen und Simulationsstudien als nützlicher Indikator des Fits erwiesen hat. Der CFI ist auf den Wertebereich zwischen null und eins normiert. Werte größer oder gleich ,95 indizieren einen guten Fit (Hu & Bentler, 1998, 1999).

Im Gegensatz zu komparativen Fitindizes geben absolute Fitindizes direkt an, wie gut das spezifizierte Modell die Daten reproduziert. Gewöhnlich werden zu diesem Zweck die Residuen der Variablen im Modell miteinander verrechnet. Die *weighted root mean square of residuals* (WRMR) gilt als einer der besten absoluten Fitindizes. Werte kleiner oder gleich eins indizieren einen guten Fit (Yu, 2002).

Der CFI und der WRMW haben anderen Indizes gegenüber mehrere Vorteile. Einer ihrer Nachteile ist jedoch, dass sie, so wie auch der Chi-Quadrat-Goodness-of-Fit-Wert, mit zunehmender Modellkomplexität einen zunehmenden Fit anzeigen. Ein Index, der dieses Problem vermeidet, ist der *root mean square error of approximation* (RMSEA), der wie der WRMR ein absoluter Fitindex ist. Zusätzlich zum absoluten Fit geht jedoch auch die Sparsamkeit der Modellannahmen (*model parsimony*) positiv in ihn ein. Ein RMSEA kleiner oder gleich ,06 indiziert eine gute Modellanpassung an die Daten (Hu & Bentler, 1999). Werte bis zu ,10 sind noch akzeptabel. Bei größeren sollte das Modell verworfen werden (Browne & Cudeck, 1993).

Die Berechnungsformel lautet nach Kenny und McCoach (2003, S. 334):

$$RMSEA = \sqrt{\frac{\chi^2 - df}{(N - 1) \cdot df}}$$

Dabei gibt χ^2 die Modellanpassung an, df die Freiheitsgrade des Modells und N die Stichprobengröße. Aus der Formel folgt, dass eine Wurzel aus einem negativen Wert gezogen werden müsste, wenn der Chi-Quadrat-Wert kleiner als die Anzahl der Freiheitsgrade ist. Einige Auswertungsprogramme, beispielsweise MPlus, geben in einem solchen Fall keinen RMSEA-Wert aus. Ein besonders kleiner Chi-Quadrat-Wert bedeutet jedoch, dass ein exzellenter Fit vorliegt. Daher ist es üblich und sinnvoll den RMSEA in einem solchen Fall manuell auf den Wert null zu setzen (Kenny & McCoach, 2003).

7.6 Zusammenfassung

Bei der Analyse von Strukturgleichungsmodellen (structural equation models; SEM) werden die Varianzen und Kovarianzen zwischen gemessenen Variablen benutzt, um die Anpassungsgüte und die Koeffizienten eines zuvor zu spezifizierenden Modells zu bestimmen. Das Modell kann sowohl manifeste als auch latente Variablen enthalten und besteht aus Messmodell (konfirmatorischen Faktoranalysen) und Regressionsmodell. Die Modellierung latenter Variablen erhöht gewöhnlich die Reliabilität und Validität der Operationalisierungen und erlaubt die explizite Überprüfung der konvergenten und der divergenten Validitäten der Faktorindikatoren. Cross-lagged panel designs ermöglichen im Gegensatz zu den meisten anderen Typen von Strukturgleichungsmodellen den Vergleich konkurrierender Kausalhypothesen.

8 Fragestellung der Dissertation

8.1 Herleitung der Forschungsfragen

Die folgenden drei Zitate fassen den aktuellen Forschungsstand zu konzeptuellem und prozeduralem Wissen treffend zusammen:

The issue of the developmental relation between understanding concepts and implementing procedures is of practical as well as theoretical importance. The widespread observation that many children perform poorly in school mathematics highlights the need for improved instruction. Historically, instructional reform has wavered back and forth between emphasizing procedures and emphasizing concepts (Hiebert & Lefevre, 1986). It has become clear that neither type of mathematical knowledge should be taught to the exclusion of the other. How to design instruction that will inculcate both kinds of knowledge is less clear, however. A better understanding of the interaction between knowledge of concepts and procedures may aid development of more effective teaching techniques. (Rittle-Johnson & Siegler, 1998, S. 77)

Relations between understanding and skill have been the target of many studies in psychology and education in general (Resnick & Ford, 1981) and in mathematics education in particular (Hiebert & Carpenter, 1992). Despite the attention, the question of how conceptual understanding and procedural skill interact remains unresolved and significant in the mathematics education community (Kilpatrick, 1988). (Hiebert & Wearne, 1996, S. 251)

Previous work has tended to focus on isolated aspects of the relationship between conceptual and procedural knowledge. Currently, the nature and significance of broad covariations in the patterns of conceptual-procedural understanding can only be the subject of speculation. (Canobi et al., 1998, S. 883)

Rittle-Johnson und Siegler weisen auf die theoretische Relevanz des Themas (vgl. die Kapitel 4 und 5 dieser Arbeit) und seine praktische Bedeutung, zum Beispiel für Unterrichtssequenzierung und Unterrichtsziele (vgl. Unterkapitel 5.5), hin. Sie erläutern auch, dass die Befunde noch nicht ausreichen, um Instruktionen zur Vermittlung konkreter Wissensinhalte (vgl. Kapitel 6) daraus ableiten zu können.

Hiebert und Wearne weisen auf die Vielzahl bisher durchgeführter Studien zu Relationen zwischen konzeptuellem und prozeduralem Wissen im Bereich des Mathematiklernens (s. Kapitel 2) und im Bereich genereller Lerntheorien (s. Kapitel 5) hin. Sie erwähnen auch, dass die Frage nach den Relationen zwischen den beiden Wissensarten trotzdem noch immer unbeantwortet ist. Wie oben geschildert, ist vor allem unklar, ob der concepts-first view, der procedures-first view, der inactivation view oder das Iterative Model zutreffend sind.

Canobi und Kollegen weisen schließlich darauf hin, dass insbesondere die Frage nach den umfassenden Kovariationsmustern verschiedener Maße der Wissensarten noch unbeantwortet ist. Dies ist deswegen wichtig, weil nur solche Muster Aufschluss darüber geben können, ob es sinnvoll ist, eine, zwei oder mehrere Wissensarten anzunehmen. Wie in Abschnitt 4.3.6 hergeleitet wurde, ist die Annahme einer Wissensart nur dann gerechtfertigt, wenn sie Kovariationen verschiedener Wissensmaße erklärt. Die Annahme zweier Wissensarten ist gerechtfertigt, wenn es zwei Gruppen von Maßen gibt, wobei die Maße derselben Gruppe kovariieren und die Maße unterschiedlicher Gruppen relativ unabhängig voneinander variieren.

Aus inhaltlicher Sicht stellen sich daher vier Fragen. Forschungsfrage 1 lautet: Kovariieren verschiedene Maße konzeptuellen beziehungsweise prozeduralen Wissens miteinander, so dass ihnen konvergente Validität zugesprochen werden kann? Forschungsfrage 2 lautet: Variieren die Maße konzeptuellen und prozeduralen Wissens (zumindest teilweise) unabhängig voneinander, so dass ihnen divergente Validität zugesprochen werden kann? Forschungsfrage 3 lautet: Wirkt sich konzeptuelles Wissen kausal auf prozedurales aus? Forschungsfrage 4 lautet: Wirkt sich prozedurales Wissen kausal auf konzeptuelles aus?

In Kapitel 7 wurde die Analysetechnik für Strukturgleichungsmodelle beschrieben. Alle vier Forschungsfragen können mit Hilfe dieser Technik untersucht werden, wenn konzeptuelles und prozedurales Wissen als latente Faktoren modelliert werden, die jeweils durch mehrere Indikatoren gemessen werden. Forschungsfrage 1 kann durch Analyse der konvergenten Validität verschiedener Indikatoren der latenten Faktoren im Messmodell untersucht werden. Zur Überprüfung der zweiten Frage können die divergenten Validitäten der Indikatoren der beiden latenten Faktoren pro Messzeitpunkt analysiert werden. Wenn die beiden Faktoren in einem Messwiederholungsdesign mindestens zweimal gemessen werden, liegt ein cross-lagged panel design vor. Die Pfadkoeffizienten des dazugehörigen Modells können dann Aufschluss über die Kausalrichtung der Relationen zwischen konzeptuellem und prozeduralem Wissen geben und so zur Beantwortung der dritten und der vierten Forschungsfrage beitragen.

Dieses Design wurde in der nachfolgend vorgestellten Studie 1 der vorliegenden Arbeit umgesetzt. Um die Validität der Befunde zu erhöhen, wurden die Wissensmaße zu drei Messzeitpunkten erhoben. Außerdem wurden die Intelligenz der Studienteilnehmer und andere Personencharakteristika als Kovariaten in das Modell aufgenommen.

Der größte Schwachpunkt des Designs von Studie 1 ist, dass es lediglich korrelative Befunde liefern kann, auch wenn diese aufgrund der Modellierung latenter Variablen und

des cross-lagged panel designs hochwertig sind. Um noch validere Antworten auf Forschungsfrage 3 und 4 zu gewinnen, wurde als Studie 2 eine Interventionsstudie mit Prätest-Posttest-Design und Randomisierung der Teilnehmer auf drei Treatmentgruppen durchgeführt. Das Design erlaubt die zusätzliche Untersuchung einer Forschungsfrage 5: Können konzeptuelles und prozedurales Wissen durch geeignete Interventionen unabhängig voneinander vermittelt werden?

Aus Gründen der Ökonomie soll in Studie 2 auf die Modellierung latenter Variablen in SEM verzichtet werden. Stattdessen sollen die in Studie 1 gewonnenen Befunde über die Operationalisierbarkeit der beiden Wissensarten genutzt werden, um sinnvolle manifeste Maße zu entwickeln.

Wie die Übersicht in Tabelle 7 illustriert, sind beide Studien auf die Beantwortung der Frage nach den kausalen Relationen ausgerichtet. Ihre Resultate können sich so gegenseitig validieren oder in Frage stellen. Studie 1 analysiert darüber hinaus die Operationalisierbarkeit der Wissensarten. Die Ergebnisse dazu sollen in Studie 2 verwendet werden, die zusätzlich untersucht, wie die Wissensarten unabhängig voneinander vermittelt werden können.

Tabelle 7: Übersicht über die Forschungsfragen.

Nr.	Forschungsfrage	Studie 1	Studie 2
1	Kovariieren verschiedene Maße konzeptuellen beziehungsweise prozeduralen Wissens miteinander, so dass ihnen konvergente Validität zugesprochen werden kann?	✓	✗
2	Variieren Maße konzeptuellen und prozeduralen Wissens (zumindest teilweise) unabhängig voneinander, so dass ihnen divergente Validität zugesprochen werden kann?	✓	✗
3	Wirkt sich konzeptuelles Wissen kausal auf prozedurales aus?	✓	✓
4	Wirkt sich prozedurales Wissen kausal auf konzeptuelles aus?	✓	✓
5	Können konzeptuelles und prozedurales Wissen durch geeignete Interventionen unabhängig voneinander vermittelt werden?	✗	✓

Anmerkungen. ✓ wird untersucht; ✗ wird nicht untersucht.

Der Überblick über die psychologisch-pädagogische Forschung zur Interaktion von konzeptuellem und prozeduralem Wissen beim Mathematiklernen ergab, dass das Iterative Modell von Rittle-Johnson et al. (2001) aus verschiedenen Gründen als den älteren Arbeiten überlegen angesehen werden kann (vgl. Kapitel 3). Es basiert auf einem umfassenden Literaturüberblick, exakten Definitionen der Wissensarten, aus denen sich hypothetische Operationalisierungen ableiten lassen, und ist durch die Annahme bidirektionaler Kausalrelationen ebenso innovativ wie integrativ. Seine bisherige

empirische Überprüfung ist dennoch ungenügend (siehe Unterkapitel 3.4), vor allem da sie lediglich korrelative Evidenz für die bidirektionalen Kausalrelationen zwischen den Wissensarten liefert. Da sie außerdem nur zwei Studien umfasst, erscheint es sinnvoll, die Modellannahmen unter Zuhilfenahme der in diesem Feld bisher nicht benutzten SEM-Technik zu überprüfen. Es wird erwartet, dass das Iterative Modell dem konkurrierenden concepts-first view und procedures-first view klar überlegen ist.

Aufgrund dieser engen inhaltlichen Anlehnung der unten berichteten Studien an das Iterative Modell wurden diese Studien auch methodisch so eng wie möglich an die schon bestehenden Studien zu dem Modell angelehnt. Wie in der Studie von Rittle-Johnson et al. (2001) wurde auch hier als Inhaltsdomäne das Arbeiten mit Dezimalbrüchen gewählt, als untersuchter Problemtyp die Verortung von Dezimalbrüchen auf dem Zahlenstrahl, als Intervention das „Fang-das-Monster“-Spiel und als Altersgruppe Fünft- und Sechstklässler. Die Konstanthaltung dieser Faktoren über die Studien hinweg erlaubt es, Unterschiede zwischen empirischen Befunden klarer einzelnen Ursachen zuordnen zu können als es sonst möglich wäre. Der Hauptunterschied zwischen den Studien besteht darin, dass Rittle-Johnson die Wissensarten über Summenscores operationalisierte, während sie hier als latente Faktoren modelliert werden. Weitere Unterschiede bestehen darin, dass einmal Schüler aus dem amerikanischen und einmal Schüler aus dem deutschen Schulsystem die Stichprobe bilden und dass unterschiedliche Instruktionen und grafische Aufgabengestaltungen benutzt werden.

Nach der Darstellung der beiden Studien werden in Kapitel 11 die Ergebnisse miteinander verglichen und diskutiert.

8.2 Zusammenfassung

Die beiden nachfolgend vorgestellten empirischen Studien dienen der Untersuchung der Fragen, ob verschiedene Maße konzeptuellen oder prozeduralen Wissens hoch miteinander korrelieren und ob Maße konzeptuellen und Maße prozeduralen Wissens teilweise unabhängig voneinander variieren. Nur dann ist die Annahme der erfolgreichen Messung zweier Wissensarten gerechtfertigt. Des Weiteren wird untersucht, ob die kausalen Beziehungen zwischen den Wissensarten unidirektional oder bidirektional sind und ob die Wissensarten sich unabhängig voneinander vermitteln lassen. Es wird erwartet, dass eine erfolgreiche Messung der Wissensarten möglich ist und dass die dem Iterativen Modell zugrundeliegenden Annahmen der unabhängigen Vermittelbarkeit und der bidirektionalen Kausalrelationen bestätigt werden können.

9 Studie 1: Interrelationen der Wissensarten

9.1 Fragestellungen und Hypothesen

Wie beschrieben, dient Studie 1 der Untersuchung der Forschungsfragen 1 bis 4 (siehe Tabelle 7 auf S. 126). Es soll überprüft werden, inwieweit verschiedene Maße konzeptuellen und prozeduralen Wissens konvergente und divergente Validität besitzen und was die Entwicklung der Maße über die Korrektheit des concepts-first views, des procedures-first views, des inactivation views und des Iterativen Modells aussagt.

Wie in der Studie von Rittle-Johnson et al. (2001) wird dazu ein Pretest-Posttest-Ein-Gruppen-Design verwandt, bei dem die interindividuellen Wissensunterschiede der Kinder vor einer Intervention dazu benutzt werden, die interindividuellen Wissensunterschiede nach der Intervention vorherzusagen. Im Gegensatz zu der bestehenden Studie, die ein suboptimales Auswertungsdesign verwendete, wird in der aktuellen Studie ein cross-lagged panel design verwendet, das Aufschluss über Kausalbeziehungen geben kann.

Die Wissensarten sollen als latente Faktoren modelliert werden, weil das erstens die Überprüfung der konvergenten und der divergenten Validität der Maße erlaubt und zweitens aufgrund der Ausfaktorisierung der maßspezifischen Fehlervarianzen zu präziseren Schätzungen der Wissensausprägungen führt als die Verwendung von Summenscores oder Einzelitems.

Die Modellierung der Wissensarten als latente Faktoren setzt voraus, dass sie jeweils über mehrere Maße operationalisiert werden. Gängige Maße für konzeptuelles und prozedurales Wissen wurden in Unterkapitel 4.5 beschrieben. Aus ökonomischen Gründen werden sie nicht alle implementiert, sondern es werden unter Gesichtspunkten der effizienten Durchführbarkeit vier Maße konzeptuellen Wissens und vier Maße prozeduralen Wissens ausgewählt. Den Wissensdefinitionen (siehe Unterkapitel 0) und der Methodik (siehe Unterkapitel 3.3) von Rittle-Johnson folgend, wird konzeptuelles Wissen durch vier verschiedene Arten von Transferaufgaben gemessen, da es flexibel transformierbar und transferierbar ist. Prozedurales Wissen wird durch vier verschiedene Verhaltensparameter beim Lösen von Routineaufgaben gemessen, da es an Problemtypen gebunden ist, die aus Übungskontexten vertraut sind. Als Routineproblem wird jedes Problem aufgefasst, das in der Verortung eines Dezimalbruchs auf einem Zahlenstrahl besteht, weil die Intervention aus Problemen dieses Typs besteht. Als Transferproblem wird jedes Problem aufgefasst, das Wissen über Dezimalbrüche, nicht jedoch die Verortung eines Dezimalbruchs auf einem Zahlenstrahl erfordert.

Alle acht Maße wurden bereits in publizierten Studien benutzt. Die Hypothese ist daher, dass die Maße konvergente und divergente Validitäten besitzen, die so groß sind, dass die latenten Faktoren gebildet und in Regressionsmodellen verwendet werden können. Die Hypothese bezüglich der Kausalrelationen ist dem Iterativen Modell entsprechend, dass bidirektionale Einflüsse vorliegen.

9.2 Methode

9.2.1 Stichprobe

An der Studie von Rittle-Johnson et al. (2001) nahmen amerikanische Fünft- und Sechstklässler teil. Dass nicht nur Schüler einer Klassenstufe eingeladen wurden, liegt daran, dass die Ergebnisse, die auf der Analyse interindividueller Unterschiede beruhen, umso valider sind, je mehr Varianz auf den verwendeten Maßen vorliegt. Die Wahl einer altersinhomogeneren Stichprobe dient der Steigerung dieser Varianz. Dieses Vorgehen erfordert sorgfältige Überprüfungen, ob die Normalverteilungen der Maße gegeben sind.

Das gewählte cross-lagged panel design erfordert, dass erstens schon zum Zeitpunkt des Prätests inhaltlich bedeutsame Vorwissensunterschiede bestehen und dass zweitens das Vorwissen so gering ist, dass während des gesamten Untersuchungszeitraumes Wissenszuwächse möglich sind. Insbesondere sind also Bodeneffekte im Prätest und Deckeneffekte im Posttest zu vermeiden.

Inhaltlich geht es in der Studie um die Verortung von Dezimalbrüchen auf dem Zahlenstrahl. Dezimalbrüche sind in Berlin, dem Durchführungsort der Studie, dem Lehrplan (*Rahmenlehrplan: Mathematik in der Grundschule*, 1986) zufolge im Laufe der sechsten Klasse zu unterrichten. In der Praxis werden sie gewöhnlich in der zweiten Hälfte des sechsten Schuljahres durchgenommen, was man im Lehrplan an der Stoffmenge erkennen kann, die in der sechsten Klassenstufe vor der Vermittlung von Dezimalbrüchen gelehrt werden soll. Fünft- und Sechstklässler in der ersten Schuljahreshälfte sind daher optimal zur Studienteilnahme geeignet, da der zu erwerbende Inhalt in ihrer Zone der nächsten Entwicklung liegt, so dass keine Bodeneffekte zu vermuten sind, andererseits überwiegend noch nicht in der Schule unterrichtet wurde, so dass keine Deckeneffekte zu befürchten sind.

Das Vorwissen der Schüler über Dezimalbrüche entstammt zwei unterschiedlichen Quellen: Zum einen sind ihnen im Alltag schon Mengen mit nicht ganzzahligen Quantitäten, die dazugehörigen Maßangaben sowie die Kommanotation begegnet (vgl. Unterkapitel 6.2). Zum anderen ist das Rechnen mit einfachen und leicht verständlichen

Alltagsbrüchen in Form von Maßangaben Teil des Mathematikunterrichts vor der fünften Klasse (vgl. *Rahmenlehrplan: Mathematik in der Grundschule*, 1986).

Wie in Abschnitt 6.3 berichtet, können Fünft- und Sechstklässler sicher mit dem (eindimensionalen) Zahlenstrahl umgehen. Falls die Kinder in der Stichprobe Dezimalbrüche nicht korrekt auf dem Zahlenstrahl abbilden können, liegt das mit großer Sicherheit an mangelhaftem Dezimalbruchwissen, nicht jedoch an mangelhaftem Zahlenstrahlwissen.

Um Studienteilnehmer zu gewinnen, wurden in den fünften und sechsten Klassen von zehn Berliner Grundschulen die geplanten Studien kurz kindgerecht dargestellt. Den Kindern wurde Informationsmaterial über die Studien (siehe Anhang A) für sie und für ihre Eltern mitgegeben. Die Schulleiter wurden vorab in einem separaten Brief (ebenfalls Anhang A) über die wissenschaftlichen Ziele der Studie informiert und um Erlaubnis für die Klassenbesuche gebeten. Die Schulen liegen in eher bürgerlichen Wohnvierteln (Steglitz, Hellersdorf und Charlottenburg), die nach ihrer räumlichen Nähe und somit nach ihrer guten Erreichbarkeit des Untersuchungsortes durch die Studienteilnehmer ausgewählt wurden.

Kinder und Eltern wurden durch das Informationsmaterial soweit wie möglich über die Ziele der Studie und die Einzelheiten der Durchführung aufgeklärt. Es wurde besonders darauf hingewiesen, dass jede Schülerin und jeder Schüler unabhängig von jeglichen Personeneigenschaften und Kompetenzen willkommen und kein besonders Vorwissen vonnöten ist.

Bei Interesse konnten die Eltern eine dem Material beiliegende bereits adressierte und frankierte Postkarte mit ihrer Telefonnummer und dem Namen des Kindes an unser Forschungsinstitut schicken. Das taten zirka 20% der Familien, was ungefähr 300 eingesandte Karten ergab. Anschließend wurden telefonisch mit 230 der Eltern Termine vereinbart, an denen ihre Kinder das Institut besuchen sollten. Es wurden Termine für jeweils anderthalbstündige Testsitzungen an zwei direkt aufeinanderfolgenden Tagen ausgemacht. Um so vielen Kindern wie möglich eine Teilnahme zu ermöglichen, konnten diese Termine an Werktagen zwischen 15 Uhr (frühester Beginn) und 19 Uhr (spätestes Ende) liegen. Während Schulferien und an Wochenenden waren Termine zwischen 10.30 und 19 Uhr möglich.

Bei diesen Terminen handelte es sich um den ersten und den zweiten Messzeitpunkt des Designs. Nach dem erfolgreichen Abschluss dieser Erhebung wurde eine Erweiterung des Designs um einen dritten Messzeitpunkt beschlossen, der ungefähr vier Monate nach

den ersten beiden lag, um die langfristige Wissensentwicklung außerhalb des Labors untersuchen zu können. Um die Termine für diesen Messzeitpunkt und einen zusätzlichen Intelligenztest auszumachen, wurden die Familien ein zweites Mal angerufen. Einige wenige Familien waren zu einer zweiten Teilname nicht oder nur teilweise bereit, so dass die Fallzahlen für den dritten Messzeitpunkt ($N = 213$) etwas niedriger sind als für die ersten beiden Messzeitpunkte ($N = 231$) und die Fallzahlen für den Intelligenztest ($N = 195$) niedriger als die Fallzahlen für den dritten Messzeitpunkt.

Um auch Familien mit niedriger Motivation zu einer Teilnahme zu bewegen, wurde den Eltern der Studienteilnehmer für die Teilnahme eines Kindes an den ersten beiden Messzeitpunkten zusammen 15,- Euro Aufwandsentschädigung gezahlt. Für eine Teilnahme am dritten Messzeitpunkt wurden 10,- Euro und für die Teilnahme am Intelligenztest (und weiteren Tests) 15,- Euro pro Kind gezahlt. Die Eltern gaben das Geld in den meisten Fällen an ihre Kinder weiter. Den Teilnehmern wurde vor oder nach der Sitzung oder in der Pause Getränke und Süßigkeiten gereicht.

9.2.2 Aufbau

Die Datenerhebungen wurden am Max-Planck-Institut für Bildungsforschung in Berlin-Dahlem durchgeführt. Sitzungen 1 bis 3 fanden im Enterprise-Lernlabor, Sitzung 4 in den beiden Sitzungssälen des Instituts statt.

Im Enterprise-Labor standen neun Einzelarbeitsplätze mit Laptops auf dafür geeigneten Tischen zur Verfügung. Sie waren durch Posterstellwände so voneinander getrennt, dass die Studienteilnehmer sich während der Datenerhebung nicht sehen konnten. Die Jalousien waren geschlossen, um am Fenster sitzende Personen nicht stärker als die übrigen durch visuelle Reize abzulenken. Da es sich um einen ruhig gelegenen Raum handelt und die Teilnehmer während der Sitzungen Kopfhörer trugen, traten auch auditive Störungen kaum auf.

Bei den Computern handelt es sich um schwarze Laptops der Marke Galaxy mit Betriebssystem Windows98, Intel-Pentium-Prozessoren und 32 MB Arbeitsspeicher. Angeschlossen waren eine einfache Computermouse von Mitsumi Electronics (Modell ECM-S2102) und ein Stereokopfhörer von Philips (Modell SBC HP 195). Es wurde eine Bildschirmauflösung von 800×600 Pixeln bei einer Farbtiefe von 16 Bit und einer Diagonale der integrierten LCD-Bildschirme von 30,5 cm gewählt.

Der Wissenstest und die Intervention waren als Computerprogramm (s. Anhänge B.2 und B.3) implementiert. Dieses wurde vor Beginn jeder Testsitzung durch den Versuchsleiter gestartet und eingestellt. Dabei wurden die Nummer des Arbeitsplatzes, die Versuchspersonennummer und der Name des jeweiligen Versuchsleiters eingegeben, so

dass sie zusammen mit Datum und Uhrzeit sowie den Eingaben der Studienteilnehmer in einer Protokolldatei gespeichert werden konnten. Neben dem Autor dieser Arbeit fungierten drei studentische Hilfskräfte und eine Forschungspraktikantin als Versuchsleiter.

Die grafische Programmoberfläche (s. z.B. Abbildung 2 auf Seite 135) wurde so gestaltet, dass sie den Bildschirm vollständig ausfüllt und den Nutzern keine Möglichkeit bietet, andere Programme oder Systemfunktionen zu erreichen. Der Screensaver wurde deaktiviert.

Die Studienteilnehmer erhielten ihre Instruktionen während der Durchführung über den Kopfhörer. Die Nutzung von Audioinstruktionen hat drei Vorteile: Erstens wurden dadurch auditive Interaktionen zwischen den Teilnehmern unterbunden. Zweitens konnte so jeder Teilnehmer die Aufgaben mit seiner eigenen Geschwindigkeit und in randomisierter Reihenfolge bearbeiten, ohne von den übrigen Teilnehmern abhängig zu sein. Drittens hat die Multimedia-Forschung gezeigt, dass Personen Erklärungen bezüglich eines Bilds am besten verarbeiten können, wenn diese auditiv mitgeteilt werden, während das Bild betrachtet wird, weil die Aufnahme der auditiven Instruktionen nicht mit der Wahrnehmung der visuellen Information interferiert (temporal contiguity principle und modality principle nach Mayer & Moreno, 2002). Die Studienteilnehmer konnten also die Programmoberfläche mit den jeweiligen Testaufgaben betrachten, während sie gleichzeitig die auditiven Instruktionen zur Aufgabenlösung hörten.

Das Programm und die Instruktionen sind so gestaltet, dass kein Vorwissen über Computer zum erfolgreichen Arbeiten notwendig ist. Alle Eingaben können über die Maus getätigt werden.

Die Benutzung der Tastatur war im Allgemeinen nicht notwendig. Neben den Computern lagen ein Schreibheft und Schreibzeug. Wo verbale Antworten notwendig waren, wurden die Kinder gebeten, diese handschriftlich in das Heft einzutragen, da dies den meisten Kindern leichter fällt als eine Tastatur zu benutzen.

Die obere Hälfte der Programmoberfläche enthält als Gedächtnisstütze jeweils eine kurze schriftliche Zusammenfassung der vorher als Audiodatei vorgespielten Aufgabeninstruktionen (für beides siehe Anhang B.1). Die untere Bildschirmhälfte dient der Präsentation der Aufgaben und Antwortalternativen.

Programmgestaltung und Instruktionen wurden in qualitativen Vorstudien mit 19 Fünft- und Sechstklässlern überprüft und weiterentwickelt. Diese bearbeiteten das Programm einzeln oder zu zweit und unterhielten sich währenddessen mit dem

Versuchsleiter über Vor- und Nachteile der Programmgestaltung und wiesen auf unverständliche Instruktionen hin. Im Anschluss an jede Durchführung wurden Programm und Instruktionen entsprechend der gewonnenen Erkenntnisse modifiziert und dann erneut erprobt.

9.2.3 Durchführung

Der erste und der zweite Messzeitpunkt (Sitzung 1 und 2) lagen für alle Studienteilnehmer zwischen dem 30.09. und dem 11.11.2005. Sie fanden pro Teilnehmer an zwei aufeinander folgenden Tagen statt. Der dritte Messzeitpunkt (Sitzung 3) lag zwischen dem 02.03. und dem 08.04.2005, also pro Teilnehmer etwa vier Monate nach den ersten beiden. Die Termine für den Intelligenztest (Sitzung 4) lagen zwischen dem 01.04. und dem 10.04.2005.

Die Studienteilnehmer wurden an den individuell vereinbarten Terminen im Eingangsbereich des Instituts empfangen, die Einverständniserklärungen der Erziehungsberechtigten wurden entgegengenommen und die Teilnehmer wurden in einen ruhigen Raum geführt, wo sie jeweils an einem Arbeitsplatz Platz nahmen. Bedingt durch die individuellen Terminabsprachen und die neun Computerarbeitsplätze waren pro Sitzung ein bis neun Teilnehmer gleichzeitig anwesend. Die Termine wurden so vereinbart, dass sie alle gleichzeitig begannen und endeten, so dass die Teilnehmer sich nicht gegenseitig beim Betreten oder Verlassen des Raums störten.

Nach einer kurzen Begrüßung erhielten die Teilnehmer eine allgemeine Einführung (siehe Anhang B.1) durch den Versuchsleiter. Ihnen wurde mitgeteilt, dass sie sich mit Fragen an den Versuchsleiter wenden könnten und dass sie dies am besten nach Beendigung der Aufgaben eines Maßes oder nach Erhalt der Instruktionen zu einem Maß und vor Beginn der Aufgabenbearbeitung tun sollten, damit die Zeitmessung, die Grundlage einiger Maße ist, nicht verzerrt würde. Sie wurden über die Funktionsweise eines Prätest-Posttest-Designs aufgeklärt und darauf hingewiesen, dass sie sich nicht schlecht zu fühlen bräuchten, wenn sie Aufgaben nicht lösen könnten, auch wenn dies im Prätest unter Umständen viele Aufgaben sein könnten.

Anschließend setzten die Teilnehmer die mit ihrem Computer verbundenen Kopfhörer auf, über die ihnen das Datenerhebungsprogramm die jeweils aufgabenrelevanten Instruktionen vermittelte.

Tabelle 8 gibt einen Überblick über die vier Datenerhebungssitzungen und ihre jeweiligen Phasen in chronologischer Reihenfolge. Sitzung 1 beginnt mit dem Prätest, gefolgt vom ersten Teil der Intervention. Sitzung 2 beginnt mit dem zweiten Teil der Intervention und endet mit dem Wissenstest. Diese Sitzungen zusammen entsprechen dem

Design von Rittle-Johnson et al. (2001). Die dritte Sitzung bestand aus einer Auffrischung des Wissens durch eine kurze Intervention und dem Wissenstest. Da die Kinder für die Bearbeitung von Wissenstest und Intervention in der zweiten und dritten Sitzung weniger Zeit als in der ersten benötigten, war in der zweiten und dritten Sitzung noch Zeit genug, um einige Kovariaten zu erheben (s.u.).

Tabelle 8: Chronologische Reihenfolge der Datenerhebungsphasen.

Inhalt	Aufgabenanzahl	Geschätzte Dauer in Minuten
Sitzung 1 (30.09.-11.11.05)		
Wissenstest	222	55
Intervention	80	30
Sitzung 2 (30.09.-11.11.05)		
Intervention	80	25
Wissenstest	222	50
Personenfragebogen	20	5
Sitzung 3 (02.03.-08.04.05)		
Distance effect	210	15
Intervention	12	20
Wissenstest	222	45
Weitere Tests	88	10
Sitzung 4 (01.04.-10.04.05)		
Intelligenztest KFT	Kurzform	90
Pause	-	15
Weitere Tests	31	40

Kinder, die vor den anderen fertig waren, wurden gebeten, sich bis zum Ende der Sitzung leise selbst an ihrem Arbeitsplatz zu beschäftigen. Dazu wurden ihnen ein Rätselheft und kleine Computerspiele zur Verfügung gestellt.

Die vierte Sitzung wurde zur Durchführung eines standardisierten Intelligenztests genutzt. Dafür konnten die Kinder das Institut an einem von acht angebotenen Terminen besuchen, an denen der Test in seiner Gruppenform mit jeweils zwischen 16 und 32 Personen durchgeführt wurde. Daran schloss sich eine Pause an, nach der weitere, hier irrelevante Variablen erhoben wurden.

9.2.4 Intervention

Die Intervention bestand aus einer Adaption des von Rittle-Johnson et al. (2001) entwickelten „Fang-das-Monster“-Spiels. Abbildung 2 zeigt zwei Screenshots der grafischen Benutzeroberfläche der Intervention.

Den Kindern wurde ein Dezimalbruch zwischen null und eins (exklusive) präsentiert. In Übereinstimmung mit dem Vorgehen von Rittle-Johnson und Kollegen entsprach die Zahl mit jeweils gleicher Wahrscheinlichkeit einem der in Tabelle 9 dargestellten Typen von Dezimalbrüchen.

Tabelle 9: In der Intervention verwendete Typen von Dezimalbrüchen.

Typ	Beschreibung	Beispiel
I	eine Nachkommastelle, keine Null hinter dem Komma	0,7
II	zwei Nachkommastellen, keine davon eine Null	0,31
III	drei Nachkommastellen, erste und dritte keine Null	0,462
IV	zwei Nachkommastellen, nur erste ist eine Null	0,09
V	drei Nachkommastellen, nur erste ist eine Null	0,085

Zusammen mit der Zahl wurde ein Zahlenstrahl präsentiert, dessen linkes Ende mit einer Null und dessen rechtes Ende mit einer Eins beschriftet war. Den Studienteilnehmern wurde über Kopfhörer mitgeteilt, dass der Dezimalbruch eine Position auf dem Zahlenstrahl angibt, an der sich ein Monster versteckt, und dass es ihre Aufgabe sei, das Monster zu fangen, indem sie den blauen Schieberegler mit der Maus auf diese Position bewegen (für die genauen Instruktionen s. Anhang B.1).

Nach der Eingabe und Bestätigung einer Position durch einen Teilnehmer erschien das Monster an der korrekten Position auf dem Zahlenstrahl. Hatte der Teilnehmer die Aufgabe korrekt gelöst, wurde sein Schieberegler grün. Hatte er sie falsch gelöst, wurde sein Schieberegler rot und ein zusätzlicher grüner Regler wurde an der korrekten Position eingeblendet (s. den rechten Screenshot in Abbildung 2). Bei der Bestimmung der Antwortkorrektheit benutzte das Programm eine Fehlertoleranz von $\pm 0,1$. Sollte beispielsweise die 0,4 abgebildet werden, durfte der Schieberegler sich noch an den Positionen 0,31 und 0,49 befinden, nicht jedoch an den Positionen 0,3 oder 0,55, um noch als richtige Antwort zu gelten. Diese Fehlertoleranzgrenze wurde auch von Rittle-Johnson verwendet und wurde eingeführt, da die Aufgabe das Dezimalbruchwissen der Teilnehmer, nicht jedoch ihre Wahrnehmungsgenauigkeit oder ihre psychomotorische Exaktheit erheben soll.

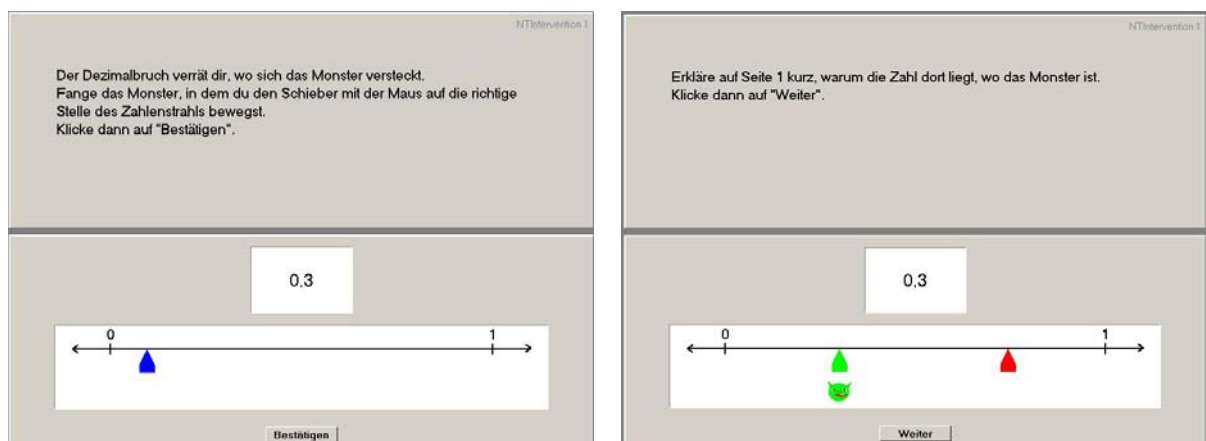


Abbildung 2: Screenshots der Grafikoberfläche der Intervention (links: vor der Antworteingabe; rechts: nach Eingabe einer inkorrekten Antwort).

Da die Übungsaufgaben eine kognitive und keine sensomotorische Fähigkeit vermitteln sollten, wurden vier verschiedene visuelle Aufgabenoberflächen entwickelt: (1) ein Zahlenstrahl, auf dem eine Position mit einem Schieberegler zu markieren ist; (2) ein Zahlenstrahl, auf dem eine Position durch einfaches Anklicken mit der Maus zu indizieren ist; (3) ein Zahlenstrahl, auf dem eine von vier durch Pfeile indizierten Positionen durch Anklicken mit der Maus auszuwählen ist; (4) eine Zufallsauswahl von einem bis fünf separaten Zahlenstrahlen, auf denen jeweils eine Position durch einen Pfeil angezeigt wird und von denen einer anzuklicken ist. Alle vier Oberflächentypen erfordern und trainieren dieselbe kognitive Fähigkeit, nämlich einen gegebenen Dezimalbruch in eine Position auf einem Zahlenstrahl umzusetzen.

In Sitzung 1 bearbeiteten die Probanden 80 Aufgaben (jeweils 20 von jedem der vier Oberflächentypen), in Sitzung 2 ebenfalls 80 (wieder jeweils 20 von jedem der vier Oberflächentypen) und in Sitzung 3 bearbeiteten sie 12 Aufgaben (jeweils 3 von jedem Oberflächentyp). In den Sitzungen 1 und 2 wurden sie bei 8 der Aufgaben und in Sitzung 3 bei 4 der Aufgaben nach Antworteingabe und Erhalt des Feedbacks gebeten, eine Erklärung für die korrekte Lösung der Aufgabe in das neben ihrem Rechner liegende Heft zu schreiben. Es kann erwartet werden, dass die Übung, das Feedback und die Generierung von Erklärungen zusammen zu Zuwächsen an konzeptuellem und prozeduralem Wissen führen (Newell & Rosenbloom, 1981; Rittle-Johnson et al., 2001; Siegler, 2002).

9.2.5 Maße konzeptuellen Wissens

Von den in Abschnitt 4.5.1 diskutierten gebräuchlichen Maßen konzeptuellen Wissens wurden unter Gesichtspunkten der Umsetzbarkeit in ein Computerprogramm und der Nutzbarkeit in Gruppenerhebungen vier Maße ausgewählt: (a) Evaluationsaufgaben, (b) Übersetzungen in andere Zeichensysteme, (c) Sortier- und Vergleichsaufgaben und (d) verbale Befragungen. Keine von ihnen erfordert die Platzierung eines Dezimalbruchs auf einem Zahlenstrahl. Sie stellen also Transferaufgaben im oben erläuterten Sinne dar.

Ob sie wirklich valide Maße konzeptuellen Wissens sind, steht noch nicht fest und ist Fragestellung dieser Studie. Wenn im Folgenden von Maßen konzeptuellen Wissens oder konzeptuellen Wissensmaßen gesprochen wird, so ist damit stets gemeint, dass diese Maße *hypothetisch* Maße konzeptuellen Wissens darstellen. Gleiches gilt für die Maße prozeduralen Wissens und die latenten Faktoren, die teilweise auch als Wissensfaktoren bezeichnet werden.

Die wörtlichen Instruktionen finden sich in Anhang B.1. Die Umsetzung der Maße in das Datenerhebungsprogramm und ihre Berechnung wird im Folgenden erläutert. Das

Programm wurde in der Sprache Borland Delphi Pascal in der dazugehörigen Programmierumgebung (Borland Software Corporation, 1983-2002) geschrieben. Der Quellcode des Programms befindet sich in B.2, eine lauffähige Programmversion in Anhang B.3. Eine Übersicht über alle verwendeten Wissensmaße wird in Abschnitt 9.2.8 gegeben.

Evaluation

Wie in Abschnitt 4.5.1 unter der Überschrift *Evaluation von Prozeduren* besprochen, sollten Evaluationsaufgaben konzeptuelles Wissen messen, weil zur Beurteilung der Korrektheit verbal präsentierter Problemlösestrategien vernetztes, explizites Wissen über allgemeine Zusammenhänge wichtig ist.

Den Probanden wurden auf dem Bildschirm die acht in Tabelle 10 dargestellten verbalen Beschreibungen von Problemlösestrategien präsentiert. Sie konnten jede durch das Anklicken von zwei Buttons mit der Maus als *eher gut* oder *eher schlecht* bewerten. Die Hälfte der Strategien war eher korrekt als inkorrekt.

Für jede korrekte Antwort wurde ein Punkt vergeben. Das Maß gibt an, wie viele Prozente der Maximalpunktzahl erreicht wurden.

Tabelle 10: Verbale Beschreibungen der zu evaluierenden Strategien.

Strategie	Darzu- stellende Zahl	Beschreibung der Strategie	Korrekt
1	0,1999	Weil die Zahl 0,1999 so viele Neunen enthält, ist sie groß. Große Zahlen sind auf dem Zahlenstrahl rechts. Darum habe ich dorthin geklickt.	eher nein
2	0,121536	Weil die Zahl 0,121536 so viele Stellen hat, ist sie sehr sehr groß. Darum habe ich so weit wie möglich rechts auf dem Zahlenstrahl geklickt.	eher nein
3	0,513	Weil die Zahl 0,513 nach dem Komma mit einer 5 beginnt, musste ich ungefähr in die Mitte des Zahlenstrahls klicken. Weil nach der 5 noch ein paar Ziffern folgen, habe ich etwas rechts neben die Mitte geklickt.	eher ja
4	0,35	0,35 liegt in der Mitte zwischen 0,3 und 0,4. Darum habe ich beide Stellen auf dem Zahlenstrahl gesucht und dann in die Mitte zwischen ihnen geklickt.	eher ja
5	0,6	Ich habe mir vorgestellt, dass der Zahlenstrahl aus 10 gleich großen Abschnitten besteht. Dann habe ich an das Ende des 6. Abschnitts geklickt.	eher ja
6	0,07	Ich habe mir vorgestellt, dass nach der Null 10 kleine Striche den Zahlenstrahl in gleich große Stücke einteilen. Dann habe ich die Striche abgezählt und auf den siebten geklickt.	eher nein
7	0,19	Erst habe ich mir überlegt, dass in der Mitte zwischen der 1 und der 9 die 5 liegt. Wegen der 5 habe ich dann auf die Mitte des Zahlenstrahls geklickt.	eher nein
8	0,201	Die Zahl 0,201 habe ich auf 0,2 abgerundet. 0,2 ist ein Fünftel. Darum habe ich auf das erste Fünftel des Zahlenstrahls geklickt.	eher ja

Übersetzung

Übersetzungen von einem Zeichensystem in ein anderes könnten konzeptuelles Wissen messen, weil man semantische Relationen nur übersetzen kann, wenn man sie verstanden hat.

Wie Abbildung 3 veranschaulicht, wurde den Teilnehmern pro Aufgabe ein Dezimalbruch zusammen mit vier Tortendiagrammen vorgelegt. Ihnen wurde mitgeteilt, dass die ganze Kreisfläche den Wert eins hat, und sie wurden aufgefordert auf die Kreisfläche zu klicken, bei der der Anteil der grau markierten Fläche an der Gesamtfläche dem Dezimalbruch entspricht. Es wurden 20 dieser Aufgaben präsentiert und hinterher wurde der Prozentsatz der korrekten Antworten berechnet.

Wie bei der Gestaltung der Intervention wurde auch bei der Gestaltung dieses und des folgenden Maßes darauf geachtet, dass die in Tabelle 9 dargestellten Typen von Dezimalbrüchen gleich häufig und in pseudozufälligen Abfolgen präsentiert wurden.



Abbildung 3: Screenshot der Implementation des Maßes Übersetzung.

Größenvergleich

Sortier- und Vergleichsaufgaben sollten konzeptuelles Wissen messen, weil ihre Lösung das Verständnis ordinaler oder nominaler Relationen zwischen Elementen einer Domäne erfordert. Bei Dezimalbrüchen sind Ordinalrelationen von größerer Wichtigkeit als Nominalrelationen. Daher wurden den Probanden nacheinander 20 Paare von Dezimalbrüchen gezeigt und sie sollten jeweils auf den größeren Paarling klicken. Anschließend wurde der Prozentsatz der korrekten Antworten berechnet. Abbildung 4 zeigt die zugehörige Programmoberfläche.

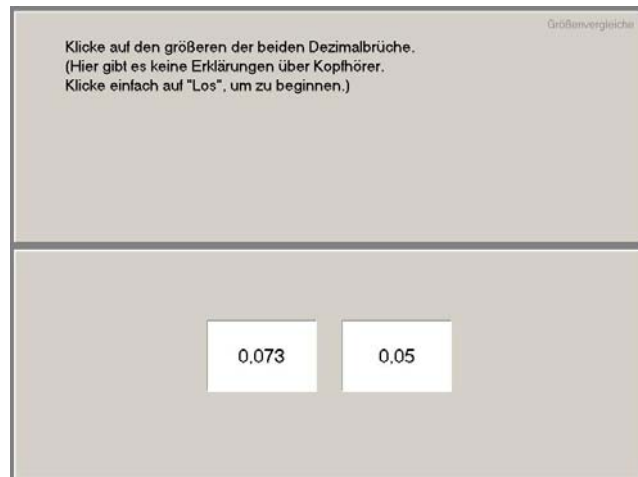


Abbildung 4: Screenshot der Implementation des Maßes Größenvergleich.

Erklärung

Antworten auf geeignete verbale Befragungen könnten gute Indikatoren für konzeptuelles Wissen sein, da sie Aufschluss darüber geben, ob explizites verbalisierbares Wissen über allgemeine Prinzipien in einer Domäne vorliegt.

Tabelle 11: Fragen des Maßes Erklärung.

Nr.	Frage	Scoringbeispiel		
		2 Punkte	1 Punkt	0 Punkte
1	Was bedeutet es, wenn in einer Zahl ein Komma vorkommt, wenn sie also ein Dezimalbruch ist?	Dass die Zahl zwischen z.B. 1 und 2 liegt und deshalb keine ganze Zahl ist.	Das bedeutet, dass die Zahl keine volle Zahl ist, also dass da eine Zahl ist und noch etwas Kleines.	Das sie eine Kommazahl ist.
2	Warum ist es sinnvoll, dass bei Tankstellen an den Zapfsäulen die verkaufte Benzinmenge als Dezimalbruch angegeben wird?	Es ist genauer, wenn ich nur einen Liter und 50 ml tanken würde müsste ich sonst 2 Liter oder nur 1 Liter bezahlen.	Weil man sonst mehr oder weniger bezahlt.	Damit man weiß wie viel Benzin ins Auto reinkommt, und damit man weiß wie viel man bezahlen muss.
3	In welchen Situationen im Leben benutzt man besser ganze Zahlen an Stelle von Dezimalbrüchen?	Wenn man einkauft und überschlagen will wie viel es kostet.	Wenn man telefoniert gibt es kein Komma.	Wenn man im Supermarkt etwas kauft.
4	Bei ganzen Zahlen hat eine lange Zahl immer einen größeren Wert als eine kurze. Ist das bei Dezimalbrüchen auch so? Begründe!	Nicht immer, z.B. die Zahl 0,182 ist kleiner als 0,82, obwohl 0,182 mehr Ziffern hat als 0,82.	Nein, weil bei Dezimalbrüchen nach dem Komma eine Null sein kann: z.B. 0,09 ist kleiner als 0,4.	Bei Dezimalbrüchen ist es genauso, weil z.B. die Zahl 111,574 ist größer als 0,423.

Aufgrund der Gruppentestungssituation wurde auf eine mündliche Befragung verzichtet. Stattdessen wurden nacheinander vier Fragen schriftlich auf dem Bildschirm eingeblendet und die Teilnehmer wurden gebeten, ihre Antworten in einem neben dem Computer liegenden Heft zu notieren. Nach dem Aufschreiben einer Antwort klickten sie einen Button an, woraufhin die jeweils nächste Frage eingeblendet wurde.

Die schriftlichen Antworten wurden von zwei unabhängigen und trainierten Ratern bewertet. Für vollständig korrekte Antworten wurden zwei Punkte vergeben, für teilweise korrekte ein Punkt und für falsche oder fehlende Antworten null Punkte. Insgesamt konnten auf dem Maß also null bis acht Punkte erzielt werden. Abschließend wurde der prozentuale Anteil der erzielten Punkte an der Maximalpunktzahl errechnet. Die Fragen und typische in der Hauptstudie gefundene Originalantworten, für die zwei, ein oder null Punkte vergeben wurden, sind in Tabelle 11 zusammengestellt.

9.2.6 Maße prozeduralen Wissens

Von den in Abschnitt 4.5.2 diskutierten gebräuchlichen Maßen prozeduralen Wissens wurden unter Gesichtspunkten der Umsetzbarkeit in ein Computerprogramm und der Nutzbarkeit in Gruppenerhebungen vier Maße ausgewählt: (a) Problemlösekorrektheit, (b) Problemlösedauer, (c) Zugriffsasymmetrie und (d) Dual task-Kosten. Sie erfordern alle die Platzierung eines Dezimalbruchs auf einem Zahlenstrahl, basieren also auf Routineaufgaben im oben erläuterten Sinn. Die wörtlichen Instruktionen finden sich in Anhang B.1. Die Umsetzung der Maße in das Datenerhebungsprogramm wird im Folgenden beschrieben.

Wie bei der Gestaltung der Intervention wurde auch bei der Gestaltung der Maße prozeduralen Wissens darauf geachtet, dass die in Tabelle 9 dargestellten Typen von Dezimalbrüchen gleich häufig und in pseudozufälligen Abfolgen benutzt wurden.

Problemlösekorrektheit

Die Problemlösekorrektheit sollte ein Indikator prozeduralen Wissens sein, weil prozedurales Wissen vornehmlich der Lösung von Problemen dient. Es wurde der aus der Intervention bekannte Problemtyp verwendet, bei dem die Position eines Dezimalbruchs auf einem Zahlenstrahl mittels eines Schiebereglers anzuzeigen ist (s. Abbildung 5). Die Studienteilnehmer wurden instruiert, ihre Antwortkorrektheiten zu maximieren, nicht jedoch ihre Lösungszeiten zu minimieren. Wie in der Intervention wurde die Korrektheit der Antworten mit einer Fehlertoleranz von $\pm 0,1$ bestimmt. Im Gegensatz zur Intervention erhielten die Kinder bei diesen Testaufgaben keine Rückmeldung über die Richtigkeit ihrer

Antworten. Nach Eingabe einer Antwort wurde sofort die jeweils nächste Aufgabe präsentiert. Die Kinder hatten 20 der Aufgaben zu lösen und es wurde der Prozentsatz ihrer korrekten Antworten berechnet.

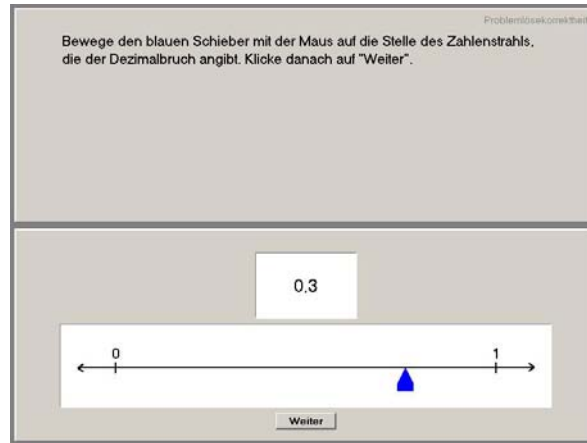


Abbildung 5: Screenshot der Implementation des Maßes Problemlösekorrektheit.

Problemlösedauer

Die Problemlösedauer könnte die Ausprägung prozeduralen Wissens anzeigen, weil zunehmendes prozedurales Wissen, unter anderem nach dem power law of practice, zu abnehmenden Lösungszeiten führt.

Den Kindern wurden nacheinander 20 Dezimalbrüche mit einem Zahlenstrahl präsentiert. Sie sollten die Position des Dezimalbruchs durch einmaliges Klicken auf den Zahlenstrahl angeben und wurden instruiert, dabei sowohl die Antwortkorrektheit als auch die Lösungsgeschwindigkeit zu maximieren. Das Programm maß die Zeiten von Beginn der Aufgabenpräsentation bis zur Antworteingabe durch den Mausklick in Millisekunden. Da nach dem power law of practice eine exponentielle Relation zwischen dem Ausmaß von Übung und den Lösungszeiten vorliegt, wurden die Lösungszeiten logarithmiert. Das hat auch den Vorteil, die typischerweise linksschiefe Verteilung von Reaktionszeiten in eine annähernd symmetrische Verteilung zu transformieren, und wird bei der psychometrischen Auswertung von Reaktionszeiten häufig durchgeführt. Man beachte, dass die ordinalen Relationen zwischen den Messwerten bei der Logarithmierung erhalten bleiben.

Nach Beendigung des Programms wurde pro Person die durchschnittliche logarithmierte Lösungszeit berechnet, wobei Ausreißerwerte, die mehr als drei Standardabweichungen über oder unter dem Mittelwert einer Person lagen, nicht mit in die

Berechnung eingingen. Solche Ausreißerwerte konnten beispielsweise entstehen, wenn ein Proband nieste und sich die Nase putzen musste.

Darüber hinaus wurden die ersten fünf Aufgaben als Übungstrials gewertet und nicht analysiert, so dass 15 Aufgaben in die Durchschnittsbildung eingingen. Dadurch sollte sichergestellt sein, dass nicht auch anfängliche Orientierungsprozesse, sondern möglichst nur die reine Aufgabenbearbeitungszeit gemessen wurde.

Zugriffsasymmetrie

Die Zugriffsasymmetrie kann vermutlich als Maß prozeduralen Wissens verwendet werden, da prozedurales Wissen eine asymmetrische, weil zielgerichtete Struktur aufweist. Eine zunehmende Asymmetrie zwischen einer geübten und einer nicht geübten Aufgabenlösungsrichtung deutet daher auf einen Zuwachs an prozeduralem Wissen hin.

Abbildung 6 zeigt die Aufgabenoberflächen. In der links dargestellten Bearbeitungsrichtung (Zugriffsasymmetrie I) war ein Dezimalbruch in eine Position auf einem Zahlenstrahl zu übersetzen. Dies wurde in der Intervention direkt geübt. In der rechts dargestellten Bearbeitungsrichtung (Zugriffsasymmetrie II) war eine Position auf einem Zahlenstrahl in einen Dezimalbruch zu übersetzen. Dies wurde in der Intervention nicht direkt geübt. Den Probanden wurden insgesamt 20 der Aufgaben jedes Typs in einem ABBA Design (10 mal Typ I, dann 20 mal Typ II, dann 10 mal Typ I) präsentiert. Das ABBA-Design dient dazu, Lernzuwächse während der Testung über die beiden Aufgabentypen hinweg auszubalancieren.

Gemessen wurde die jeweilige Lösungsdauer pro Aufgabe. Dann wurden wie bei der Problemlösedauer die durchschnittlichen ausreißerbereinigten logarithmierten Lösungszeiten pro Person und Aufgabentyp berechnet und es wurde analysiert, um wie viel Prozent die Lösungszeiten bei Typ II größer sind als bei Typ I. Der Prozentwert wurde durch Relativierung der absoluten Wertedifferenz an der Typ I-Lösungszeit berechnet (für die Berechnungsformel s. auch Tabelle 12). Er bildet so die Größe der Asymmetrie unabhängig von den absoluten Größen der Lösungszeiten ab und sollte bei zunehmender Übung größer werden, auch wenn die einzelnen absoluten Lösungszeiten absinken.

Um weitere, in dieser Arbeit nicht beschriebene Analysen zu ermöglichen, wurden nicht bei allen Aufgaben fünf Antwortalternativen präsentiert, wie in Abbildung 6 gezeigt, sondern zwischen eins und fünf. Die Anzahlen der Antwortalternativen, ihre Reihenfolge sowie die verwendeten Zahlen waren über beide Aufgabentypen hinweg jedoch identisch und wurden daher bei der Auswertung ignoriert.

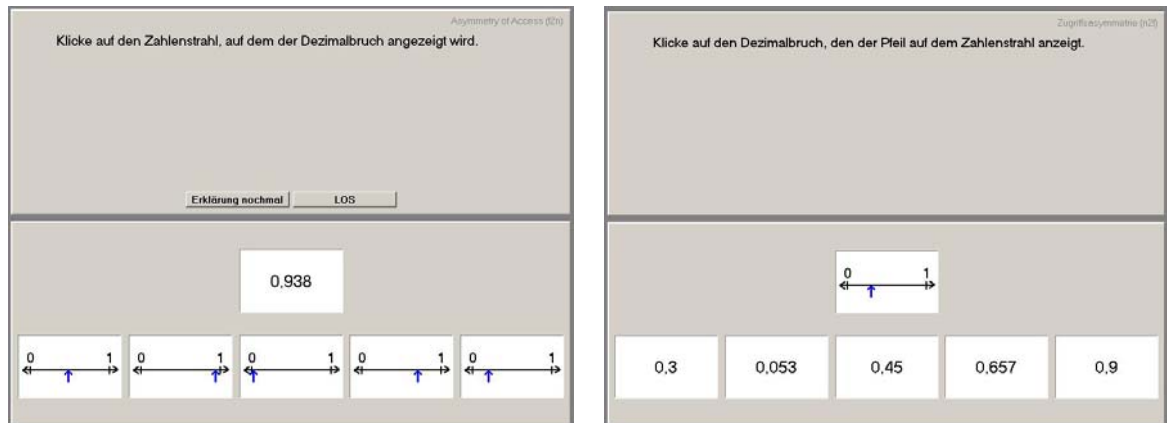


Abbildung 6: Screenshot der Implementierung des Maßes Zugriffsasymmetrie (links: geübte Transformationsrichtung I, rechts: nicht geübte Transformationsrichtung II).

Dual task-Kosten

Die Dual task-Kosten sollten prozedurales Wissen messen, weil prozedurales Wissen automatisierte und daher ressourceneffiziente Problemlösungen ermöglicht. Je mehr prozedurales Wissen jemand besitzt, umso mehr kognitive Ressourcen sollten ihm zur Lösung einer Zweitaufgabe zur Verfügung stehen und umso niedriger sollten die Dual task-Kosten ausfallen.

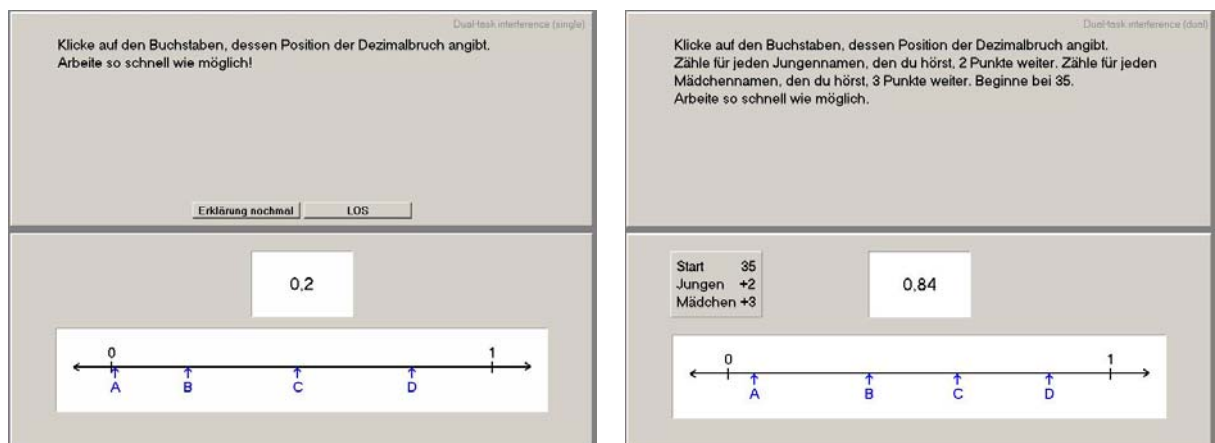


Abbildung 7: Screenshots der Implementation des Maßes Dual task-Kosten (links: Baseline-Bedingung I, rechts: Zweitaufgaben-Bedingung II).

Abbildung 7 zeigt die Grafikoberflächen in der Baseline-Bedingung (Dual task-Kosten I) und der Zweitaufgabenbedingung (Dual task-Kosten II). In der Baseline-Bedingung waren Routineaufgaben durch Anklicken einer von jeweils vier Antwortalternativen zu lösen. In der Zweitaufgabenbedingung hatten die Kinder zusätzlich zu diesen Aufgaben noch folgende zu lösen: Sie hatten anfangs an die Zahl 35 zu denken und hörten in pseudozufallsverteilten Zeitabständen von zwei bis vier Sekunden in Deutschland

gebräuchliche zweisilbige Vornamen (Felix, Sophie usw.) in einer Pseudozufallsreihenfolge über ihren Kopfhörer. Für jeden weiblichen Vornamen sollten die Probanden 3 Punkte zu dem memorierten Punktestand addieren, für jeden männlichen Vornamen 2 Punkte. Nach der Lösung von zehn Routinezahlenstrahlaufgaben wurden sie jeweils gefragt, bei welchem Punktestand sie angekommen waren, und hatten diesen in ein kleines Eingabefeld auf dem Bildschirm zu tippen.

Ein Proband sollte zu Beginn eines Aufgabenblocks also an die 35 denken und dann beginnen zehn der Zahlenstrahlaufgaben zu lösen. Während dessen hörte er als erstes zum Beispiel den Namen Peter und zählte auf Grund des männlichen Vornamens zwei Punkte zur 35 hinzu. Er merkte sich dann also die 37. Dann hörte er den Namen Anna, zählte zu der 37 drei Punkte hinzu und merkte sich den Punktestand 40. Nach der Lösung von zehn Routinezahlenstrahlaufgaben wurden keine Namen mehr eingespielt, sondern es öffnete sich das Fenster zur Eingabe des Punktestands.

Wie in Abschnitt 4.5.2 beschrieben, können Dual task-Kosten nur dann auftreten, wenn sichergestellt ist, dass die Probanden wirklich beide Aufgabentypen gleichzeitig lösen müssen und sich nicht auf einen alleine konzentrieren können. Das war hier der Fall. Die Probanden konnten die Zahlenstrahlaufgaben nicht ignorieren, weil erst nach der Lösung von zehn von ihnen die Eingabe des gezählten Punktestandes und ein Fortfahren mit dem Programm möglich waren. Die Probanden konnten die Namenszählaufgabe nicht ignorieren, weil sie den Punktestand in den Computer einzugeben hatten.

Um sicherzustellen, dass die Probanden die Aufgaben nicht nur oberflächlich bearbeiten, sondern sich bei beiden Aufgabentypen Mühe geben, werden ihnen in dual task-Designs typischerweise Rückmeldungen über ihre Leistungen gegeben und sie werden zu verstärkter Anstrengung aufgefordert, falls die Leistungen zu schlecht sind.

Das wurde auch hier getan: Nach jeder Eingabe des Punktestandes wurden die Probanden darüber informiert, wie viele der letzten 10 Zahlenstrahlaufgaben sie korrekt gelöst hatten und um wie viele Punkte der von ihnen eingegebene Punktestand von dem tatsächlichen abwich. Falls die Fehlerraten zu hoch waren, wurden die Probanden ermahnt, sich beim nächsten Zehnerblock mehr Mühe zu geben.

Da die Erstaufgabe visuell am Bildschirm und die Zweitaufgabe auditiv über den Kopfhörer gestellt wurden und da die Eingabe des Punktestandes nach und nicht während der Lösung eines Zehnerblocks erfolgte, war sichergestellt, dass Erst- und Zweitaufgabe lediglich auf kognitiver Ebene und nicht auf Ebene der Aufgabenwahrnehmung oder Antworteingabe interferieren konnten.

Die Namenszählaufgabe wurde auch als Zweitaufgabe gewählt, weil für Kinder das Zählen, das Memorieren von Zahlen und das Hören und Kategorisieren von Vornamen alltägliche schulische Anforderungen darstellen, denen ökologische Validität nicht abgesprochen werden kann. Auch dass Kinder Mathematikaufgaben lösen und gleichzeitig Vornamen anderer Kinder hören, ist im Unterricht nichts Ungewöhnliches.

Die Zweitaufgabe sollte mit der Erstaufgabe interferieren, weil sie sowohl starke verbale als auch starke numerische Anteile hat. Unabhängig davon, ob die Kinder die Erstaufgabe durch eher verbale oder eher numerische Strategien lösen, sollte die Zweitaufgabe also ihre Erstaufgabenleistung verringern.

Da die Probanden in der Zweitaufgabenbedingung nach jeder Lösung von zehn Zahlenstrahlaufgaben unterbrochen wurden, den Punktestand einzugeben hatten und Rückmeldungen über ihre Leistungen erhielten und da die Baseline-Bedingung der Zweitaufgabenbedingung – abgesehen von der Zweitaufgabe selbst – möglichst ähnlich sein soll, wurde auch in der Baseline-Bedingung nach jedem Zehnerblock von Aufgaben eine Rückmeldung über die Lösungskorrektheit erteilt.

Insgesamt bearbeiteten die Probanden 40 Aufgaben in der Baseline-Bedingung und 50 Aufgaben in der Zweitaufgaben-Bedingung in einem ABBA-Design (20 Baseline-Aufgaben ohne Zweitaufgaben, dann 50 Baseline-Aufgaben mit gleichzeitigen Zweitaufgaben, dann wieder 20 Baseline-Aufgaben ohne Zweitaufgaben), wobei der erste Zehnerblock in der Zweitaufgaben-Bedingung als Übungsteil für die komplexe Zweitaufgabentechnik betrachtet und zur Validitätssteigerung komplett von der Auswertung ausgeschlossen wurde.

Dann wurden pro Person und Bedingung die durchschnittlichen Lösungsraten berechnet. Die Dual task-Kosten wurden pro Person bestimmt, indem berechnet wurde, um wie viel Prozent die Lösungsrate in der Zweitaufgabenbedingung schlechter war als in der Baseline-Bedingung. Der Prozentwert wurde durch Relativierung an der absoluten Wertedifferenz an dem Durchschnittsmesswert der Baseline-Bedingung errechnet (für die Formel s. Tabelle 12). Er bildet also die Größe der Dual task-Kosten unabhängig von den absoluten Lösungsraten ab.

9.2.7 Erhobene Kovariaten

Um das Zustandekommen der interindividuellen Wissensunterschiede näher untersuchen zu können und um externe Spezifikationsfehler (siehe Seite 115) so weit wie möglich zu vermeiden, wurden einige Kovariaten über einen Personenfragebogen, einen Intelligenztest und ein Programm zur Erhebung der individuellen Ausprägung des distance effect erhoben.

Personencharakteristika

Der Fragebogen wurde am Ende von Sitzung 3 ausgefüllt und ist in Anhang B.4 einzusehen. Er enthält 20 Items zu Alter, Geschlecht, Klassenstufe, Mathematik- und Deutschnote im letzten Zeugnis, Vorerfahrung mit Computern und der Maus, Gründen für die Studienteilnahme, Spaß an der Untersuchung, der subjektiven Einschätzung des Lernerfolgs, der subjektiven Einschätzung der Testschwierigkeit, der subjektiven Einschätzung des bereichsspezifischen Vorwissens und zu einem retrospektiven Selbstbericht der dominanten Strategie zur Lösung der dual task-Aufgaben.

Intelligenz

Die Intelligenz wurde erhoben durch die Kurzform des Kognitiven Fähigkeitstests in seiner revidierten Form (KFT 4-12+ R) von Heller und Perleth (2000). Er ist für die Klassenstufen vier bis zwölf normiert, stellt die deutsche Fassung des Cognitive Ability Tests von Thorndike und Hagen dar und ist laut Manual zur differenziellen Intelligenzmessung geeignet. Die Kurzform umfasst jeweils zwei Skalen zur Erhebung verbaler Fähigkeiten, numerischer Fähigkeiten und nonverbal-figurativer Fähigkeiten. Sie liegt in einer Form A und einer Form B vor, ist für Gruppenuntersuchungen geeignet und dauert etwa 90 Minuten. Der Test ist auf T-Werte normiert. Er hat also in der Eichstichprobe einen Mittelwert von 50 und eine Streuung von 10. Die Normierung erfolgt auf Grundlage von Klassenstufe und Testform.

Die auf Basis der Konsistenzkoeffizienten bestimmte Reliabilität beträgt laut Manual ,95. Die Retestreliabilität über ein Jahr liegt für Fünft- und Sechstklässler je nach Substichprobe zwischen ,83 und ,87. Der Gesamttestwert korreliert bei Fünft- und Sechstklässlern zu ,09 bis ,46 mit der Deutschnote und zu ,37 bis ,51 mit der Mathematiknote. Im Testmanual (Heller & Perleth, 2000) werden ausführlich weitere Befunde berichtet, die die Validität und Reliabilität des Tests belegen.

Der Test wurde als Gruppentest unter Benutzung der standardisierten Instruktionen durchgeführt. Die Teilnehmer saßen in einem großen Tischkreis, wobei die Testformen A und B abwechselnd ausgelegt wurden.

Distance effect

Der distance effect (vgl. Abschnitt 4.6.2) wurde durch ein zusätzliches Computerprogramm erhoben. Dieses wurde ebenfalls in der Programmierumgebung für Borland Delphi Pascal (Borland Software Corporation, 1983-2002) entwickelt. In der Mitte eines ansonsten völlig schwarzen Bildschirms wurden den Probanden nacheinander 210

Zahlen zwischen 10 und 99 (mit Ausnahme der 55) in einer für alle Teilnehmer konstanten Pseudozufallsreihenfolge präsentiert. Das Erscheinen einer neuen Zahl wurde durch das Erscheinen des Buchstaben *x* in der Bildschirmmitte angekündigt. Nach 500 ms wurde der Bildschirm für 150 ms schwarz, woraufhin die neue Zahl erschien, die so lange auf dem Bildschirm blieb, bis der Benutzer seine Eingabe gemacht hatte. Dann verschwand die Zahl und der Bildschirm wurde für 2000 ms schwarz, bis ein neues *x* das Erscheinen der nächsten Zahl ankündigte.

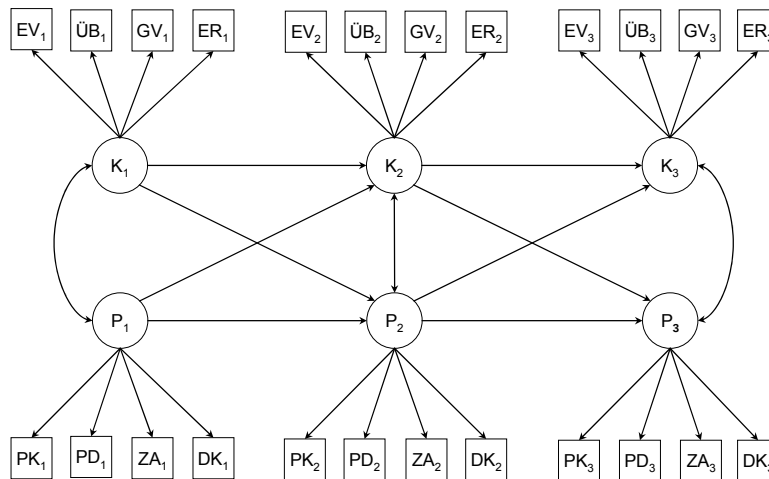
Die Aufgabe der Probanden bestand darin, auf der Laptoptastatur die Taste *v* zu drücken, falls die Zahl auf dem Bildschirm kleiner als 55 ist, und die Taste *m*, falls die Zahl größer als 55 ist. Die beiden Tasten waren der leichteren Auffindbarkeit wegen mit einem kleinen grünen (*v*) und einem kleinen blauen (*m*) Aufkleber versehen, auf die die Instruktionen Bezug nahmen.

Das Programm inklusive der Instruktionen, der Protokolldateien, der Auswertungssyntax und des ausgewerteten Datensatzes befindet sich in Anhang C. Die Gestaltung des Programms und die unten beschriebene Auswertung folgten dem bewährten Vorgehen von Dehaene und Kollegen (z.B. Dehaene et al., 1990).

Das Programm zeichnete die Lösungszeiten der Probanden zusammen mit dem jeweiligen Stimulus und der Antwortkorrektheit auf. Die ersten zehn Aufgaben wurden als Übungstrials gewertet und nicht ausgewertet. Für die übrigen 200 Aufgaben wurde zunächst die Differenz zwischen dem Wert 55 und dem auf dem Bildschirm präsentierten Wert berechnet. Die Lösungszeiten wurden logarithmiert, da nach Dehaene eine exponentielle Beziehung zwischen Wertedifferenz und Lösungszeit besteht. Dann wurde pro Proband eine lineare Regression mit unabhängiger Variable Wertedifferenz und abhängiger Variable logarithmierter Lösungszeit gerechnet. Der standardisierte Regressionskoeffizient wurde als Maß für die personenspezifische Stärke des distance effect betrachtet und für nachfolgende Analysen benutzt.

9.2.8 Design und Auswertung

Das zur Auswertung verwendete Grundmodell zeigt Abbildung 8. Es wird später auch als Modell R123 (R für prädiktive Relationen, 123 für die drei Messzeitpunkte) bezeichnet. Wie bei Pfaddiagrammen üblich, werden manifeste Variablen durch Rechtecke und latente Variablen durch Kreise dargestellt. Einköpfige Pfeile repräsentieren Regressionsbeziehungen, doppelköpfige Pfeile Kovarianz- oder Korrelationsbeziehungen. Pfeile von latenten zu manifesten Variablen indizieren die Faktorindikatoren und damit das Messmodell.



Anmerkungen: K: konzeptuelles Wissen; P: prozedurales Wissen; EV: Evaluation; ÜB: Übersetzung; GV: Größenvergleich; ER: Erklärung; PK: Problemlösekorrektheit; PD: Problemlösedauer; ZA: Zugriffsasymmetrie; DK: Dual task-Kosten; Indizes geben den Messzeitpunkt an; Fehlerterme nicht mit dargestellt.

Abbildung 8: Für die Auswertung verwendetes Grundmodell.

Das Diagramm lässt erkennen, dass zwei latente Variablen – konzeptuelles Wissen K und prozedurales Wissen P – an drei Messzeitpunkten durch jeweils vier Indikatoren gemessen wurden. Das Wissen zum zweiten und zum dritten Messzeitpunkt wird jeweils durch das vorherige konzeptuelle und prozedurale Wissen vorhergesagt.

Die Koeffizienten der Pfade von den latenten Faktoren zu den Indikatoren geben Aufschluss über die konvergenten Validitäten der Maße. Die Koeffizienten der doppelköpfigen Pfeile zwischen den latenten Variablen geben Aufschluss über die divergenten Validitäten der Wissensmaße zu einem Messzeitpunkt. Die Koeffizienten der Pfade zwischen den latenten Variablen geben Aufschluss über die prädiktiven Beziehungen zwischen den Wissensarten.

Die Skalierung der latenten Faktoren wurde festgelegt, indem die unstandardisierte Faktorladung des jeweils ersten Indikators jedes Faktors auf den Wert eins fixiert wurde. Das gilt auch für alle weiteren im Folgenden besprochenen Modelle.

Es sind $t = 59$ Modellkoeffizienten (4 Indikatorresiduen, 3 Faktorladungen, 1 Faktorvarianz für jeden der 6 latenten Faktoren zuzüglich 8 Regressionspfadkoeffizienten und 3 Kovarianzen zwischen den Faktoren) auf Basis von $s = 24$ gemessenen Variablen zu schätzen. Durch Einsetzen der Werte in die Abzählregel:

$$t \leq \frac{s(s+1)}{2}$$

ergibt sich die wahre Aussage $59 \leq 300$. Die Abzählregel ist somit erfüllt und deutet darauf hin, dass Modell R123 empirisch identifiziert ist.

Entsprechend der üblichen Praxis wurde das Modell nicht gleich als Ganzes überprüft, weil dann unklar wäre, welche der zahlreichen Interrelationen für das Zustandekommen eines guten oder schlechten Modellfits verantwortlich ist. Stattdessen wurden die einzelnen Modellteile nacheinander evaluiert. Zuerst wurden die Wissensmaße, ihre Verteilungseigenschaften und ihre Entwicklungen über die Messzeitpunkte jeweils einzeln analysiert. Dann wurden für jeden Faktor einzeln die konvergenten Validitäten seiner Indikatoren und anschließend für jeden Messzeitpunkt einzeln die divergenten Validitäten der Indikatoren überprüft. Danach wurden die prädiktiven Beziehungen zwischen den Messzeitpunkten getestet. Schließlich wurden nacheinander die Kovariaten KFT und Distance effect in das Modell eingefügt und die daraus resultierenden Veränderungen analysiert.

Tabelle 12 gibt eine Übersicht über die acht Wissensmaße und die Teilmaße, aus denen sich die Werte für die Zugriffsasymmetrie und die Dual task-Kosten zusammensetzen.

Tabelle 12: Übersicht über die Wissensmaße.

	Anforderung	Maßeinheit	Aufgaben- anzahl
Maße konzeptuellen Wissens (Transferaufgaben)			
Evaluation	Bewertung verbal beschriebener Problemlöseprozeduren	%	8
Übersetzung	Umsetzung von Dezimalbrüchen in Tortendiagramme	%	20
Größenvergleich	Paarweise Größenvergleiche von Dezimalbrüchen	%	20
Erklärung	Niederschrift von Erklärungen allgemeiner Eigenschaften von Dezimalbrüchen	%	4
Maße prozeduralen Wissens (Routineaufgaben)			
Problemlösekorrektheit	Lösungsrate bei Routineproblemen	%	20
Problemlösedauer	Lösungsdauer bei Routineproblemen	10·ln(s)	15
Zugriffsasymmetrie I ^a	Lösungsdauer bei Routineproblemen	10·ln(s)	20
Zugriffsasymmetrie II ^a	Lösungsdauer bei reversen Routineproblemen	10·ln(s)	20
Zugriffsasymmetrie	100 mal (Zugriffsasymmetrie II minus Zugriffsasymmetrie I) geteilt durch Zugriffsasymmetrie I	Δ%	-
Dual task-Kosten I ^a	Lösungsrate bei Routineproblemen (Baseline-Bedingung)	%	40
Dual task-Kosten II ^a	Lösungsrate bei Routineproblemen und gleichzeitiger Zweitaufgabe (Dual task-Bedingung)	%	40
Dual task-Kosten	100 mal (Dual task-Kosten I minus Dual task-Kosten II) geteilt durch Dual task-Kosten I	Δ%	-

Anmerkung. ^a Kein eigenständiges Wissensmaß, sondern Bestandteil eines anderen Maßes.

Zur Auswertung der Einzelmaße wurde der Datensatz *FinalData1.sav* in Anhang D.3 benutzt. Um ihn zu erhalten, wurden die Protokolldateien (s. Anhang D.1) von den Untersuchungslaptops in ein gemeinsames Verzeichnis kopiert. Durch die Syntaxdateien und Programme in Anhang D.2 wurden sie zusammengefügt, aggregiert, und die Einzelmaße wurden entsprechend oben stehender Beschreibungen berechnet. Anschließend wurden die von Hand eingegebenen Personenfragebogen- und Intelligenztestdaten (s. Anhang D.1) an diesen Datensatz angehängt. Der fertig aufbereitete Datensatz mit den individuellen Ausprägungen des distance effects (Anhang C) wurde ebenfalls an den Gesamtdatensatz angehängt.

Zur Analyse der Strukturgleichungsmodelle wurde das Programm Mplus (Muthén & Muthén, 1998-2004a) benutzt. Da zur Auswertung kein SPSS-Datensatz verwendet werden kann, wurden aus dem Datensatz *FinalData1.sav* verschiedene SEM-Datensätze abgeleitet, die sich in Anhang D.3 befinden.

In MPlus eingelesen wurden die Individualdaten, analysiert wurde die Kovarianzmatrix. Da eine multivariate Normalverteilung der Maße nicht gegeben war (siehe Seite 154), wurde der sb-skalierte maximum-likelihood Schätzalgorithmus (MLM; siehe Unterkapitel 7.4) verwendet. Da er die Analyse von missing values nicht erlaubt, wurden Fälle mit fehlenden Daten von der Analyse ausgeschlossen. Dies traf pro Analyse nur auf wenige Personen zu (s.u.).

Wie oben beschrieben, wurden die Daten auf intrapersonaler Ebene ausreißerbereinigt, das heißt es wurden einzelne Datenpunkte entfernt, die drei Standardabweichungen über oder unter dem intrapersonalen Mittelwert lagen. Da die Wissensmaße pro Person Summen über mehrere Einzelaufgaben darstellen, entstehen auf Ebene der Summenscores durch die intrapersonale Ausreißerbereinigung keine zusätzlichen missing values. Bei einer interpersonalen Ausreißerbereinigung wäre dies anders: Würden Summenscores von Personen, die weit über oder unter dem Stichprobenmittelwert liegen, aus dem Datensatz entfernt, führte dies zu zusätzlichen missing values und die Personen müssten ganz aus dem Datensatz ausgeschlossen werden. Dadurch würde nicht nur die Anzahl der auswertbaren Personen, sondern auch die interpersonale Varianz unnötig reduziert. Auf diese Art der Ausreißerbereinigung wurde daher verzichtet. Eine Verzerrung der Ergebnisse durch die Ausreißer stellt wegen der Robustheit des MLM-Schätzalgorithmusses keine Gefahr dar.

Die Aufgabenreihenfolge wurde pro Person über die Messzeitpunkte hinweg konstant gehalten, um keine unnötige Fehlervarianz zu erzeugen. Über die Personen hinweg wurde

die Aufgabenreihenfolge hingegen blockrandomisiert, damit die Reihenfolge die Ergebnisse der Faktoranalysen nicht verzerrt. Mit blockrandomisiert ist dabei gemeint, dass die Aufgaben eines Wissensmaßes stets direkt hintereinander präsentiert wurden, die Erhebungsreihenfolge der einzelnen Wissensmaße sich jedoch von Person zu Person unterschied. Angesichts der Stichprobengröße kann angenommen werden, dass sich etwaige Reihenfolgeeffekte dabei ausgleichen.

Für alle folgenden Ergebnisberichte gilt: Das Symbol *** steht für eine Signifikanz mit einer Alphafehlerwahrscheinlichkeit von $p \leq ,001$. ** symbolisiert eine Signifikanz mit $p \leq ,010$. * steht für eine Signifikanz mit $p \leq ,050$. Das Symbol + zeigt einen nicht signifikanten Trend mit $p \leq 0,100$ an.

9.3 Ergebnisse bezüglich der Forschungsfragen

9.3.1 Beschreibung der Stichprobe

Von den 231 Teilnehmern des ersten Messzeitpunktes nahmen alle auch am zweiten Messzeitpunkt teil. Zum dritten Messzeitpunkt erschienen 213 der Teilnehmer. Neun dieser 213 Teilnehmer wurden post hoc von der Auswertung ausgeschlossen, da sie die Tests nicht vollständig bearbeiteten, so dass auszuschließende missing values vorlagen, da sie in gravierender Weise nicht die Instruktionen befolgten oder da sie aufgrund schwerwiegender Sprachschwierigkeiten nicht in der Lage waren, die Instruktionen überhaupt zu verstehen. Die einzelnen Gründe für die neun Personen sind in Anhang D.2.3 aufgeführt. Die Auswertung erfolgte für 204 Personen. Das entspricht 88% der Ursprungsstichprobe und ist angesichts des mehrmonatigen Untersuchungszeitraumes eine gute Quote. Die 204 Personen weisen an allen drei Messzeitpunkten auf keinem der acht Wissensmaße missing values auf, sind also für die geplante Auswertung geeignet. Die nachfolgenden Angaben beziehen sich auf diese Personengruppe. Die Auswertungssyntax und die SPSS-Ausgabedatei für den aktuellen Abschnitt sind in Anhang D.4 zu finden.

Die Teilnehmer sind zu 51% weiblich und zu 49% männlich. Sie besuchten zu 47% die fünfte und zu 53% die sechste Klassenstufe. Die jüngsten Teilnehmer waren 9,3 die ältesten 13,8 Jahre alt. Median und Mittelwert der Altersverteilung liegen bei 11,3 Jahren ($SD = 0,7$). Der Mittelwert der Mathematiknote im letzten Schulzeugnis betrug 2,5 ($SD = 0,9$), der Mittelwert der letzten Deutschnote 2,3 ($SD = 0,8$). Der Wertebereich beider Noten reichte von der Eins bis zur Fünf.

Der mittlere KFT-Gesamtwert der Stichprobe liegt mit $m = 54$ (umgerechnet in IQ-Punkte: $m = 106$) nur leicht über dem normierten Mittelwert von 50 (umgerechnet in IQ-Punkte: 100). Die Streuung in der Stichprobe unterschätzt mit $SD = 8$ die normierte

Streuung von $SD = 10$ nur unwesentlich. Der geringste KFT-Gesamtwert in der Stichprobe beträgt 26 (umgerechnet in IQ-Punkte: 64), der höchste 76 (umgerechnet in IQ-Punkte: 139).

Von den Teilnehmern gaben 23% an, in der Schule noch nie einen Computer benutzt zu haben; 6% äußerten, zu Hause noch nie einen Computer benutzt zu haben. Nur eine einzige Person gab an, dass sie bisher noch nie eine Computermaus bedient hatte.

Von den drei Aussagen „Ich nehme an dieser Untersuchung teil, weil ich etwas lernen möchte.“, „Ich nehme an dieser Untersuchung teil, weil ich mir das interessant vorstelle.“ und „Ich nehme an dieser Untersuchung teil, weil ich dafür etwas Geld bekomme.“ erhielt die erste am meisten Zustimmung ($m = 4,7$ mit $SD = 0,9$ auf einer fünfstufigen Ratingskala von 1 = *stimmt überhaupt nicht* bis 5 = *stimmt völlig*), die zweite etwas weniger ($m = 4,3$; $SD = 0,9$) und die dritte am wenigsten ($m = 2,7$; $SD = 1,3$). Natürlich ist nicht auszuschließen, dass die Teilnehmer teilweise im Sinne der sozialen Erwünschtheit antworteten.

Die Aussage „Die Teilnahme an der Untersuchung hat mir Spaß gemacht.“ wurde auf oben beschriebener Ratingskala mit $m = 4,7$ und $SD = 0,6$ bewertet, fand also breite Zustimmung. Die Bewertungen 1 oder 2 wurden von keiner Person vergeben. Auch die Aussage „Ich glaube, ich habe etwas über Dezimalbrüche gelernt.“ wurde überwiegend positiv bewertet ($m = 4,5$; $SD = 0,8$). Die Aussage „Ich fand die Aufgaben schwierig.“ erzielte ein Durchschnittsrating von $m = 2,8$ ($SD = 1,0$), welches ungefähr in der Mitte der Skala liegt.

Wie oben beschrieben, wurden diese Daten zu Ende der zweiten Testsitzung erhoben. Sie verdeutlichen, dass es sich um eine geschlechts- und altersmäßig ausgewogene Stichprobe handelt, die durch die Wissenstests und das „Fang-das-Monster“-Spiel weder über- noch unterfordert war. Die Schulnoten lassen erkennen, dass auch Schülerinnen und Schüler mit durchschnittlichen oder unterdurchschnittlichen Schulleistungen zur Studienteilnahme bereit waren. Auch die Intelligenzdaten zeigen an, dass die Teilnehmer in Bezug auf ihre intellektuelle Leistungsfähigkeit weitgehend repräsentativ für deutsche Schülerinnen und Schüler sind.

9.3.2 Eigenschaften der Wissensmaße

Die Analysedateien zu den hier vorgestellten Ergebnissen befinden sich in Anhang D.5. Im Folgenden werden die Verteilungseigenschaften der Maße, die Veränderungen über die Messzeitpunkte hinweg und die Reliabilitäten der Maße berichtet.

Vor den Analysen wurde die Ausreißerbereinigung der Reaktionszeitmaße Problemlösedauer, Zugriffsasymmetrie I und Zugriffsasymmetrie II durchgeführt. Pro Maß und Messzeitpunkt wurden zwischen 24 und 49 Werte von nachfolgenden Analysen ausgeschlossen. Angesichts der pro Maß und Messzeitpunkt erhobenen $204 \times 20 = 4080$ Werte lagen nur wenige Ausreißer vor. Pro Person, Maß und Messzeitpunkt wurde höchstens ein Ausreißerwert gefunden (s. das Protokoll der Ausreißerbereinigung in Anhang D.2.3).

Verteilungseigenschaften

Tabelle 13 zeigt die Mittelwerte, Standardabweichungen und Signifikanzen des Kolmogorov-Smirnov-Anpassungstests an die Normalverteilung für alle Wissensmaße zu den drei Messzeitpunkten. Die Maße konzeptuellen Wissens befinden sich im oberen Teil der Tabelle, die Maße prozeduralen Wissens im unteren.

Tabelle 13: Mittelwerte, Standardabweichungen und Signifikanz des Kolmogorov-Smirnov-Anpassungstests der Wissensmaße für die drei Messzeitpunkte.

Wissensmaß	Mzp 1			Mzp 2			Mzp 3		
	<i>m</i>	<i>SD</i>	Sign.	<i>m</i>	<i>SD</i>	Sign.	<i>m</i>	<i>SD</i>	Sign.
Maße konzeptuellen Wissens									
Evaluation ^a	60,23	21,10	**	66,73	21,66	***	76,04	18,81	***
Übersetzung ^a	51,18	16,93	***	69,49	19,15	***	69,85	17,82	*
Größenvergleiche ^a	75,07	15,75	***	83,63	12,69	***	86,79	13,61	***
Erklärungen ^a	16,54	20,63	***	21,69	23,17	***	31,80	25,98	**
Maße prozeduralen Wissens									
Problemlösekorrektheit ^a	56,72	21,32	**	88,06	16,27	***	86,20	16,27	***
Problemlösedauer ^b	79,56	4,49	*	78,17	3,09	n.s.	77,71	3,40	+
Zugriffsasymmetrie I ^b	80,91	2,92	n.s.	77,76	2,37	n.s.	77,30	2,41	n.s.
Zugriffsasymmetrie II ^b	80,01	2,79	n.s.	78,17	2,53	n.s.	77,44	2,53	+
Zugriffsasymmetrie ^c	-1,15	2,61	n.s.	0,50	1,86	n.s.	0,15	2,38	+
Dual task-Kosten I ^a	60,10	22,40	***	90,88	13,58	***	89,27	15,80	***
Dual task-Kosten II ^a	51,70	24,75	***	85,18	19,49	***	85,31	19,68	***
Dual task-Kosten ^d	15,43	19,33	*	7,04	14,78	***	5,39	12,63	***

Anmerkungen. ^a Werte in %; ^b Werte in $10 \times \ln(s)$; ^c Werte in $\Delta 10 \times \ln(s)$; ^d Werte in $\Delta \%$.

Die Mittelwerte der prozentualen Maße liegen alle über 15% und unter 90%, die Maße der Lösungszeiten liegen umgerechnet zwischen 2,3 und 3,3 Sekunden, so dass keines der Maße starke Decken- oder Bodeneffekte aufweist, auch wenn einige Mittelwerte in Verbindung mit den Standardabweichungen erkennen lassen, dass einige Verteilungen deutlich an einen Skalenanfang oder ein Skalenende stoßen. Das trifft vor allem zu auf die relativ niedrigen Werte des Maßes Erklärung zum ersten Messzeitpunkt und die relativ hohen Werte der Maße Größenvergleich, Problemlösekorrektheit sowie Dual task-Kosten I und II beim zweiten oder dritten Messzeitpunkt.

Die Mittelwerte lassen weiterhin erkennen, dass zum ersten Messzeitpunkt schon moderat ausgeprägtes konzeptuelles und prozedurales Vorwissen vorhanden war. Das sei anhand des Maßes Problemlösekorrektheit veranschaulicht: Die Ratewahrscheinlichkeit bei den Problemlöseaufgaben beträgt angesichts der relativen Länge des Zahlenstrahls von einem Punkt und der Fehlertoleranzgrenze von $\pm 0,1$ Punkten (siehe Seite 135) genau 20%, da 0,2 Punkte und damit ein Fünftel des Zahlenstrahls als korrekte Eingabeposition eines Dezimalbruchs gelten. Die durchschnittliche Lösungsrate beträgt hingegen mehr als 60%, ist also fast dreimal so groß wie die Ratewahrscheinlichkeit, was nur durch Vorwissen erklärt werden kann. Ähnliches gilt auch für die übrigen Maße.

Da die Maße vermutlich unterschiedliche Schwierigkeiten besitzen, geben relative Unterschiede der Maße zueinander keinen Aufschluss darüber, ob die Kinder über mehr konzeptuelles als prozedurales Vorwissen verfügten oder umgekehrt. Sicher ist nur, dass bereits Vorwissen vorhanden war.

Die Standardabweichungen lassen erkennen, dass es auf jedem Maß hinreichend große interpersonale Varianz gibt. Das ist wichtig, da es gerade die interindividuellen Unterschiede sind, die mittels SEM analysiert werden sollen. Die prozentualen Maße weisen untereinander ähnlich große Varianzen auf. Gleiches gilt für die Reaktionszeitmaße. Diese beiden Gruppen von Maßen können hingegen nicht sinnvoll miteinander verglichen werden, da sie auf unterschiedlichen Skalen angesiedelt sind.

Mittels Kolmogorov-Smirnov-Anpassungstest wurde geprüft, inwieweit sich die empirisch vorgefundene Verteilung eines Maßes zu einem Messzeitpunkt signifikant von der Normalverteilung unterschied. Die Signifikanzen in

Tabelle 13 zeigen, dass diese Abweichungen von der Normalverteilung an allen drei Messzeitpunkten und für ganz unterschiedliche Maße teilweise erheblich sind. Daher kann auch von einer multivariaten Normalverteilung nicht ausgegangen werden. Histogramme der Verteilungen aller Wissensmaße können in Anhang D.5 eingesehen werden.

Veränderungen über die Messzeitpunkte

Die Veränderungen der Wissensmaße über die drei Messzeitpunkte werden in Tabelle 14 wiedergegeben. Von Messzeitpunkt 1 zu Messzeitpunkt 2 entwickelten sich alle Maße erwartungsgemäß. Alle prozentualen Lösungsraten nehmen hochsignifikant zu und indizieren damit Wissenszuwächse. Die Lösungsratenzuwächse bei den Routineaufgaben (Problemlösekorrektheit, Dual task-Kosten I, Dual task-Kosten II) sind größer als bei den vier Transferaufgabentypen im oberen Teil der Tabelle. Dies ist plausibel, da die

Routineaufgaben direkt in der Intervention geübt wurden, während die Transferaufgaben nur indirekt mit den Interventionsinhalten zusammenhängen.

Tabelle 14: Veränderung der Wissensmaße (Mittelwertsdifferenz, Cohens *d* und Signifikanz) über die drei Messzeitpunkte.

	Mzp 1 zu Mzp 2			Mzp 2 zu Mzp 3		
	$m_2 - m_1$	<i>d</i>	Sign.	$m_3 - m_2$	<i>d</i>	Sign.
Maße konzeptuellen Wissens						
Evaluation ^a	6,50	0,30	***	9,31	0,42	***
Übersetzung ^a	18,31	1,05	***	0,37	0,02	n.s.
Größenvergleiche ^a	8,56	0,53	***	3,16	0,24	***
Erklärungen ^a	5,15	0,24	***	10,11	0,43	***
Maße prozeduralen Wissens						
Problemlösekorrektheit ^a	31,35	1,44	***	-1,86	-0,11	*
Problemlösedauer ^b	-1,39	-0,29	***	-0,46	-0,13	*
Zugriffsasymmetrie I ^b	-3,15	-0,95	***	-0,46	-0,16	**
Zugriffsasymmetrie II ^b	-1,84	-0,57	***	-0,73	-0,24	***
Zugriffsasymmetrie ^c	1,66	0,56	***	-0,35	-0,14	+
Dual task-Kosten I ^a	30,79	1,36	***	-1,62	-0,11	*
Dual task-Kosten II ^a	33,49	1,33	***	0,13	0,01	n.s.
Dual task-Kosten ^d	-8,39	-0,43	***	-1,65	-0,11	+

Anmerkungen. ^a Werte in $\Delta\%$; ^b Werte in $\Delta 10 \times \ln(s)$; ^c Werte in $\Delta \Delta 10 \times \ln(s)$; ^d Werte in $\Delta \Delta\%$.

Auch alle Lösungszeiten wurden hochsignifikant geringer. Da alle vier Maße konzeptuellen Wissens und alle vier Maße prozeduralen Wissens hochsignifikante Wissenszuwächse indizieren, hat sich die Intervention also wie erwartet auf beide Wissensarten positiv ausgewirkt.

Beim Maß Zugriffsasymmetrie verringerten sich sowohl die Zeiten für die Lösung des geübten Aufgabentyps I als auch des ungeübten Aufgabentyps II. Der Lerneffekt war bei Aufgabentyp I jedoch stärker ausgeprägt als bei Aufgabentyp II, so dass die Asymmetrie zwischen beiden erwartungsgemäß größer wird.

Auch die Dual task-Kosten entwickelten sich erwartungsgemäß. In beiden Bedingungen wurden die Kinder besser. Jedoch sind die Wissenszuwächse in der Dual task-Bedingung (Typ II) größer als in der Baseline-Bedingung (Typ I), so dass die prozentualen Kosten vom ersten zum zweiten Messzeitpunkt absinken.

Vergleicht man die Effektstärken, die in Tabelle 14 in Form von Cohens *d* angegeben werden, stellt man fest, dass die Auswirkungen der Intervention auf konzeptuelles Wissen am sensitivsten durch das Maß Übersetzung und die Auswirkungen der Intervention auf prozedurales Wissen am sensitivsten durch das Maß Problemlösekorrektheit abgebildet werden.

Die Veränderungen über die vier Monate vom zweiten zum dritten Messzeitpunkt weisen ein komplexeres Muster auf. Die Maße Evaluation, Größenvergleiche, Erklärungen und Problemlösedauer zeigen signifikante Wissenszuwächse an. Die Maße Übersetzung, Zugriffsasymmetrie und Dual task-Kosten veränderten sich nicht signifikant. Die

Problemlösekorrektheit indiziert schwache, aber angesichts der großen Stichprobe signifikante Wissensverluste. Auffällig ist, dass sich nicht alle vier Maße konzeptuellen oder alle vier Maße prozeduralen Wissens jeweils in die selbe Richtung entwickeln.

Reliabilitäten der Maße

Vor den SEM-Analysen stehen schon drei verschiedene Indikatoren zur Verfügung, die etwas über die Reliabilitäten der verwendeten Maße aussagen. Diese sind in Tabelle 15 zusammengestellt. Für die Analysedateien siehe Anhang D.6. Zunächst liefert die Itemanzahl eine a priori-Information über die Reliabilitäten der Maße: Je mehr Messwerte in einen Summenscore eingehen, umso reliabler ist er, weil sich die Fehler ausmitteln. Cronbachs Alpha-Koeffizienten geben darüber hinaus Aufschluss über die internen Konsistenzen der Maße. Sie stellen jeweils eine untere Schranke dar, über der die wahre Reliabilität liegen muss.

Tabelle 15: Indikatoren der Reliabilitäten der Wissensmaße.

	Itemanzahl	Cronbachs Alpha			Stabilität	
		Mzp 1	Mzp 2	Mzp 3	Mzp 1 zu 2	Mzp 2 zu 3
Maße konzeptuellen Wissens						
Evaluation	8	,44	,54	,47	,40	,49
Übersetzung	20	,70	,80	,76	,47	,63
Größenvergleich	20	,77	,75	,80	,63	,65
Erklärung	4	,52	,59	,54	,73	,56
Maße prozeduralen Wissens						
Problemlösekorrektheit	20	,82	,85	,85	,39	,74
Problemlösedauer	15	,94	,93	,94	,40	,63
Zugriffsasymmetrie I	20	,88	,88	,90	,47	,61
Zugriffsasymmetrie II	20	,87	,87	,89	,50	,60
Zugriffsasymmetrie	-	- ^a	- ^a	- ^a	,04	,16
Dual task-Kosten I	40	,92	,92	,93	,42	,82
Dual task-Kosten II	40	,93	,94	,95	,50	,82
Dual task-Kosten	-	- ^a	- ^a	- ^a	,23	,61

Anmerkung. ^a für Differenzwerte nicht berechenbar.

Die internen Konsistenzen der Wissensmaße sind über die Messzeitpunkte hinweg sehr stabil. Die Konsistenzen der Maße prozeduralen Wissens liegen alle über ,80 und sind daher als gut oder sehr gut anzusehen. Die internen Konsistenzen der Maße konzeptuellen Wissens sind insgesamt geringer. Die beiden verbalen Aufgabentypen Evaluation und Erklärungen haben niedrigere interne Konsistenzen als die beiden nonverbalen Aufgabentypen, allerdings gehen in sie auch weniger Items ein als in die anderen beiden Summenscores.

Den dritten Indikator der Reliabilitäten stellen die Stabilitätskoeffizienten dar. Sie entsprechen der Korrelation der Ausprägungen eines Wissensmaßes zu einem und dem darauf folgenden Messzeitpunkt. Die Stabilitäten hängen mit den Retestreliabilität der

Maße positiv und mit individuell verschiedenen Lernprozessen zwischen den Messungen negativ zusammen. Die Stabilitätskoeffizienten sind also keine perfekten Maße der Reliabilitäten, können aber dazu dienen die unteren Schranken der Reliabilität abzuschätzen.

Als niedrig sind vor allem die Stabilitäten der Maße Zugriffsasymmetrie und Dual task-Kosten vom ersten zum zweiten Messzeitpunkt anzusehen. Bei der Zugriffsasymmetrie ist auch die Stabilität vom zweiten zum dritten Messzeitpunkt niedrig. Entweder weisen diese Maße also niedrige Retestreliabilitäten auf oder sie bilden interindividuell unterschiedliche Lernerfahrungen sensibler ab als die übrigen Maße, die durchweg höhere Stabilitäten besitzen.

Beim Maß Erklärung stellt sich darüber hinaus die Frage nach der Beurteilerübereinstimmung bei der Bewertung der schriftlichen Antworten. Da die beiden Beurteiler zwei Punkte für vollständig korrekte, einen Punkt für teilweise korrekte und null Punkte für fehlende oder inkorrekte Antworten vergaben, wurde für jede der vier Einzelaufgaben Kendalls *tau* berechnet. Die Summenscores werden in nachfolgenden Strukturgleichungsmodellen als intervallskalierte Variablen modelliert und daher hier durch Produkt-Moment-Korrelationen ausgewertet. Tabelle 16 zeigt die entsprechenden Koeffizienten. Die Interraterreliabilitäten sind zufriedenstellend und, insbesondere für die Summenscores, stabil über die drei Messzeitpunkte.

Tabelle 16: Interraterreliabilitäten zu drei Messzeitpunkten.

	Mzp 1	Mzp 2	Mzp 3
Aufgabe 1 ^a	,79	,79	,86
Aufgabe 2 ^a	,67	,68	,82
Aufgabe 3 ^a	,84	,83	,84
Aufgabe 4 ^a	,65	,76	,82
Summenscore ^b	,88	,87	,92

Anmerkungen. ^a Kendalls τ ; ^b Pearsons r .

9.3.3 Konvergente Validitäten

Die vorangehenden Analysen stützen die Annahme, dass die Daten zur Auswertung mittels Strukturgleichungsmodellen geeignet sind. Diese werde im Folgenden zunächst benutzt, um die konvergenten Validitäten der Maße zu überprüfen. Es wurde der Datensatz SEMData.dat aus Anhang D.3 benutzt. Die Syntax- und Outputdateien des Analyseprogramms MPlus befinden sich in Anhang D.7.

Analyse der einfachen Modelle

Für das konzeptuelle Wissen zum ersten Messzeitpunkt liegen vier verschiedene Indikatoren vor. Es wurde ein im Folgenden auch Einfaktormodell genanntes Modell K1E (K für konzeptuelles Wissen, 1 für ersten Messzeitpunkt und E für Einfaktormodell) spezifiziert, bei dem die vier Maße auf einem latenten Faktor laden. Die Fitindizes des Modells geben darüber Aufschluss, ob die Annahme des latenten Faktors und damit einer einzigen den Maßen unterliegenden Wissensdimension gerechtfertigt ist. Entsprechende Modelle wurden auch für die anderen beiden Messzeitpunkte erstellt (Modell K2E und Modell K3E) sowie für die Maße prozeduralen Wissens (Modelle P1E, P2E und P3E).

Die Fitindizes in den oberen sechs Zeilen von Tabelle 17 zeichnen ein gemischtes Bild. Der WRMR liegt durchgängig unter eins und zeigt damit gute Anpassungen der Einfaktormodelle an die Daten an. Dem CFI zufolge haben alle Modelle bis auf Modell P2E einen guten Fit. Der RMSEA hingegen bescheinigt nur den Modellen K1E, P1E und P3E einen guten Fit.

Tabelle 17: Fit der Modelle zur Überprüfung der konvergenten Validitäten.

	Name	χ^2	df	p	CFI	WRMR	RMSEA	R ²	α
Originalmodelle									
Konzeptuelles Wissen									
Mzp 1	K1E	0,781	2	,675	1,000	0,204	0,000 ^a	,38	,69
Mzp 2	K2E	6,658	2	,035	0,972	0,862	0,107	,39	,67
Mzp 3	K3E	5,151	2	,075	0,988	0,438	0,088	,49	,76
Prozedurales Wissen									
Mzp 1	P1E	2,772	2	,247	0,972	0,541	0,044	,22	,32
Mzp 2	P2E	3,915	2	,139	0,926	0,625	0,069	,28	,24
Mzp 3	P3E	3,398	2	,180	0,934	0,603	0,059	,25	,44
Modifizierte Modelle									
Konzeptuelles Wissen									
Mzp 1	K1EM	0,096	1	,757	1,000	0,069	0,000 ^a	,38	,69
Mzp 2	K2EM	0,078	1	,781	1,000	0,064	0,000 ^a	,38	,67
Mzp 3	K3EM	1,679	1	,780	0,997	0,197	0,058	,49	,76
Prozedurales Wissen									
Mzp 1	P1EM	0,010	1	,919	1,000	0,038	0,000 ^a	,21	,32
Mzp 2	P2EM	0,558	1	,455	1,000	0,189	0,000 ^a	,27	,24
Mzp 3	P3EM	0,928	1	,336	1,000	0,280	0,000 ^a	,26	,44

Anmerkung. ^a Der Koeffizient wurde manuell auf Null gesetzt, da $df > \chi^2$ (s. S. 123).

Analyse der modifizierten Modelle

Der Schätzalgorithmus basiert bei Strukturgleichungsmodellen auf der Annahme, dass die Koeffizienten aller Pfade, die nicht Teil des Modells sind, den Wert null haben. Für die spezifizierten Modelle bedeutet dies, dass versucht wird, die Kovarianzen der Variablen ausschließlich durch den Einfluss des latenten Faktors und durch keine einzige zusätzliche Relation zu erklären. In zwei Fällen ist diese Annahme jedoch unplausibel. Von den vier

Maßen konzeptuellen Wissens erfordern zwei zusätzlich zum konzeptuellen Wissen über Dezimalbrüche noch sprachliche Kompetenzen, während die anderen beiden unterschiedliche nonverbale Fähigkeiten voraussetzen. Wenn man das konzeptuelle Wissen aus den vier Maßen auspartialisiert, sollten daher nicht alle Maße zu null korrelieren, sondern die Residuen der Maße Evaluation und Erklärung sollten noch Kovarianz aufweisen, die erst bei zusätzlicher Kontrolle der Sprachkompetenz verschwinden würde. Dies kann man bei der Modellierung berücksichtigen, indem man in die Einfaktormodelle eine Kovarianzbeziehung zwischen den Residuen (d.h. den Varianzanteilen, die nicht durch einen latenten Faktor aufgeklärt werden) der Maße Evaluation und Erklärung aufnimmt. Ähnliches gilt für die prozeduralen Wissensmaße Problemlösedauer und Asymmetrie. Während alle anderen sechs Maße Lösungsraten darstellen, basieren diese beiden auf Zeitmessungen. Auch hier sollte eine Kovarianz der Residuen zugelassen werden. Durch die Berücksichtigung der zusätzlichen Kovarianzen entstehen sechs neue Modelle, die hier als Modell K1EM bis Modell P3EM bezeichnet werden, wobei das M für modifiziert steht.

Die Fitindizes dieser Modelle stehen in der unteren Hälfte von Tabelle 17. Der Vergleich der ursprünglichen mit den modifizierten Modellen zeigt, dass der durch den Faktor aufgeklärte Varianzanteil durch die Modifikation abgesehen von Rundungsfehlern nicht beeinflusst wurde. Dies ist eine Folge davon, dass die Relationen zwischen dem Faktor und seinen Indikatoren in den ursprünglichen und den modifizierten Modellen genau gleich spezifiziert wurden.

Die übrigen Indizes belegen, dass die Modifikation zu beinahe perfekt an die Daten angepassten Modellen geführt hat. Alle CFI-Werte stoßen an die normierte Obergrenze von 1,000, alle WRMR-Indizes liegen weit unter der Cut-Off-Grenze von 1,000 und alle RMSEA-Werte bis auf einen stoßen an die normierte Untergrenze von 0,000.

Diese exzellenten Modellfits sprechen für die Richtigkeit der oben erläuterten Zusatzannahmen bezüglich der Kovarianzen.

Obwohl sie nicht Teil des a priori spezifizierten Modells (s. Abbildung 8 auf S. 148) waren, wurden die Kovarianzbeziehungen in alle nachfolgenden Modelle mitaufgenommen. Das hat den Sinn, dass ein etwaiger später gefundener schlechter Fit eines Modells direkt auf die dann zu überprüfenden Annahmen über die Wissensarten und nicht alternativ auch auf Schwächen des Messmodells attribuiert werden kann.

Für die nachfolgenden Analysen kann festgehalten werden, dass die Indizes der Modelle K1EM bis P3EM in Tabelle 17 ausgezeichnete konvergente Validitäten der Maße anzeigen.

Alle latenten Wissensfaktoren klären zwischen 20 und 50 Prozent der Gesamtvarianz ihrer jeweiligen Indikatoren auf. Die Maße konzeptuellen Wissens erklären deutlich mehr Varianz als die Maße prozeduralen Wissens. Außerdem weisen die konzeptuellen Wissensfaktoren höhere interne Konsistenzen auf als die prozeduralen Wissensfaktoren.

Die interne Konsistenz der Faktoren ist am dritten Messzeitpunkt jeweils höher als an den beiden vorherigen Messzeitpunkten.

Analyse des Vorzeichens der Faktorladungen

Die Faktorladungen in Tabelle 18 geben den Beitrag der einzelnen Indikatoren zu einem Faktor an. Die Ladungen des Indikators Dual task-Kosten weisen ein negatives Vorzeichen auf. Das ist erwartungsgemäß, da die Kosten bei zunehmendem prozeduralen Wissen abnehmen sollten. Auch die positiven Ladungen aller prozeduralen Lösungsraten sind erwartungsgemäß. Einzig das Maß Problemlösedauer weist ein unerwartetes Vorzeichen auf. Da ein Zuwachs an prozeduralem Wissen zu einem Absinken der Lösungsdauer führen sollte, wäre hier ein negatives Vorzeichen zu erwarten. Das positive Vorzeichen belegt jedoch, dass die Lösungsdauer mit zunehmendem prozeduralen Wissen zunimmt.

Tabelle 18: Faktorladungen in den Modellen K1EM bis P3EM.

Wissensmaß	Messzeitpunkt		
	1	2	3
Standardisierte Koeffizienten			
Maße konzeptuellen Wissens			
Evaluation	0,636	0,494	0,601
Übersetzung	0,674	0,745	0,826
Größenvergleich	0,682	0,773	0,764
Erklärung	0,423	0,352	0,565
Maße prozeduralen Wissens			
Problemlösekorrektheit	0,763	0,626	0,603
Problemlösedauer	0,059	0,434	0,359
Zugriffsasymmetrie	0,372	0,458	0,416
Dual task-Kosten	-0,334	-0,538	-0,598
Unstandardisierte Koeffizienten			
Maße konzeptuellen Wissens			
Evaluation	1,000	1,000	1,000
Übersetzung	0,850	1,333	1,303
Größenvergleich	0,800	0,916	0,920
Erklärung	0,650	0,761	1,300
Maße prozeduralen Wissens			
Problemlösekorrektheit	1,000	1,000	1,000
Problemlösedauer	0,016	0,131	0,116
Zugriffsasymmetrie	0,060	0,084	0,094
Dual task-Kosten	-0,397	-0,781	-0,715

Dieser Befund ist auch darum erstaunlich, weil die Lösungsdauer von Messzeitpunkt zu Messzeitpunkt, also mit zunehmender Übung, signifikant kleiner wurde (s. Tabelle 14 auf S. 155), während beispielsweise die Problemlösekorrektheit mit zunehmender Übung

größer wurde. Wie können beide Maße trotzdem zu allen drei Messzeitpunkten positiv korreliert sein (die Korrelationen betragen $r_1 = ,04$, $r_2 = ,30$ und $r_3 = ,26$)?

Die Beobachtungen der Versuchsleiter ergaben, dass die Studienteilnehmer stets bemüht waren sich an die Instruktionen zu halten. Die Instruktion für das Maß Problemlösedauer lautet: „Jetzt siehst Du einen Dezimalbruch und einen Zahlenstrahl. Deine Aufgabe lautet: ‚Klicke auf die Stelle des Zahlenstrahls, die zu dem Dezimalbruch gehört. Bearbeite die Aufgabe so schnell wie möglich. Mache dabei keine Pausen.‘ Hier ist es also wichtig, dass Du die Aufgaben richtig, aber auch schnell löst.“ Kinder, die einzelne Aufgaben nicht lösen konnten, waren bemüht, wenigstens den Teil der Instruktion, der sie zu zügigem Arbeiten aufforderte, zu befolgen. Sie klickten dann einfach möglichst schnell auf eine beliebige Stelle des Zahlenstrahls, um sofort mit der nächsten Aufgabe fortzufahren. Das geschah bei Kindern mit wenig prozeduralem Wissen häufiger als bei Kindern mit viel prozeduralem Wissen. Letztere konnten viele Lösungen herleiten und hatten daher auch viele Lösungen akkurat einzugeben, was sie Zeit kostete.

Das erklärt, warum die Kinder über die Messzeitpunkte hinweg immer kürzere Lösungsdauern aufweisen: Wer die Aufgaben korrekt löste, wurde dabei immer schneller und wer sich mangels Wissen einfach durchklickte, wurde dabei auch immer schneller. Diese Geschwindigkeitszuwächse sind anscheinend so stark, dass sie die Lösungsdauerzuwächse aufgrund der Wissenszuwächse über die Messzeitpunkte hinweg verdecken.

Trotz der sinkenden Gesamtlösungsdauer können nach dieser Erklärung an allen drei Messzeitpunkten prozedurales Wissen und Problemlösedauer positiv korreliert sein.

Die Erklärung ist natürlich eine Interpretation der Resultate und nicht selbst ein empirisches Resultat. Die Interpretation wird hier im Ergebnisteil schon dargelegt, da es wichtig ist, die konvergenten und die divergenten Validitäten der Maße vollständig geklärt zu haben, bevor das Regressionsmodell analysiert wird. Die teilweise Vermischung von Ergebnisdarstellung und Ergebnisinterpretation ist bei der Analyse von Strukturgleichungsmodellen daher üblich und sinnvoll (Kaplan, 2000, S. 7-10; Raykov et al., 1991).

Eine Möglichkeit mit dem Problem der positiven Faktorladung umzugehen, bestünde darin, alle Aufgaben, die nicht korrekt gelöst wurden, von der Berechnung des Maßes Problemlösedauer auszuschließen. Dies würde jedoch dazu führen, dass das Maß für Kinder, die alle Aufgaben falsch gelöst haben, nicht berechnet werden kann. Da der Datensatz keine missing values enthalten darf (vgl. Kapitel 7), müssten diese Kinder vollständig von der Analyse ausgeschlossen werden. Bei den Kindern mit wenigen

verbleibenden Werten käme es darüberhinaus zu Problemen bei der Ausreißerbereinigung und sie müssten ebenfalls von den Analysen ausgeschlossen werden. Weil zur Analyse von Strukturgleichungsmodellen große Stichproben ($N \geq 200$) erforderlich sind, würde dies die Schätzung der übrigen Modellparameter beeinträchtigen. Das Vorgehen wurde zwar parallel zu den unten beschriebenen Hauptanalysen explorativ durchgeführt. Dabei traten jedoch Konvergenzprobleme bei der Parameterschätzung auf, so dass die Ergebnisse hier nicht vorgestellt werden.

Weil die konfirmatorischen Faktoranalysen einen exzellenten Fit der Messmodelle aller drei prozeduralen Wissensfaktoren belegen und das unerwartete Vorzeichen des Maßes Problemlösedauer durch die Beobachtungen der Versuchsleiter erklärt werden kann, wird das Maß in den nachfolgenden Analysen verwendet. Da es einen systematischen Bezug zum prozeduralen Wissensfaktor aufweist, würde es die Validität und Reliabilität des Faktors eher schwächen als stärken, wenn das Maß ausgeschlossen würde.

Analyse der Höhen der Faktorladungen

Betrachtet man zusätzlich zu den Vorzeichen auch die Höhe der Faktorladungen, so zeigt sich, dass die Ladungen auf den konzeptuellen Wissensfaktoren im Schnitt etwas höher und homogener sind als die Ladungen auf den prozeduralen Wissensfaktoren. Dies steht in Übereinstimmung mit dem Befund, dass die konzeptuellen Faktoren mehr Varianz aufklären und höhere interne Konsistenzen besitzen als die prozeduralen (vgl. Tabelle 17).

Die Absolutwerte der Ladungen einiger Maße, wie der Übersetzung, der Dual task-Kosten und der Problemlösedauer, nehmen über die Zeit hinweg zu.

Der beste einzelne Indikator für konzeptuelles Wissen ist das Maß Übersetzung mit einer durchschnittlichen standardisierten Faktorladung von 0,751. Der beste einzelne Indikator für prozedurales Wissen ist das Maß Problemlösekorrektheit mit einer durchschnittlichen standardisierten Faktorladung von 0,668. Der schlechteste einzelne Indikator für konzeptuelles Wissen ist die Erklärung, der schlechteste Indikator für prozedurales Wissen die Problemlösedauer.

Weil die Maße Zugriffsasymmetrie und Dual task-Kosten sich über die Zeit hinweg entwickeln wie erwartet und auch die erwarteten Beziehungen zu den latenten Wissensfaktoren aufweisen, wurde auf eine Überprüfung des Vorliegens von speed-accuracy trade-offs und trade-offs zwischen Erst- und Zweitaufgabe verzichtet.

9.3.4 Divergente Validitäten

Hypothesentestende Analysen

Zur Überprüfung der divergenten Validitäten der Maße wurde zunächst pro Messzeitpunkt getestet, ob ein Zweifaktormodell mit einem konzeptuellen und einem prozeduralen Wissensfaktor (Modell Z1, Z2 und Z3 mit Z für Zweifaktormodell) die Daten signifikant besser fittet als ein Einfaktormodell (Modell E1, E2 und E3 mit E für Einfaktormodell).

Die Ergebnisse der Analysen geben die ersten sechs Zeilen von Tabelle 19 wieder. Die zugehörigen Dateien befinden sich in Anhang D.8.1. Gemessen an CFI, WRMR und RMSEA haben alle sechs Modelle einen guten oder akzeptablen Fit. Die Koeffizienten lassen keine klare Entscheidung zwischen den Ein- und den Zweifaktormodellen zu.

Tabelle 19: Fit der Modelle zur Überprüfung der divergenten Validitäten.

Faktoren-anzahl	Modell	χ^2	df	p	CFI	WRMR	RMSEA
Mzp 1							
1 Faktor	E1	16,976	18	,525	1,000	0,680	0,000 ^a
2 Faktoren	Z1	13,406	17	,709	1,000	0,633	0,000 ^a
Mzp 2							
1 Faktor	E2	29,526	18	,042	0,960	0,899	0,056
2 Faktoren	Z2	29,095	17	,034	0,958	0,910	0,059
Mzp 3							
1 Faktor	E3	31,265	18	,027	0,962	0,768	0,060
2 Faktoren	Z3	30,738	17	,022	0,961	0,797	0,063

Anmerkung. ^a Koeffizient manuell auf null gesetzt, da $df > \chi^2$ (s. S. 123).

Um zu testen, ob die Unterschiede im Modellfit überhaupt von Zufallsschwankungen zu unterscheiden sind, wurden für die drei Messzeitpunkte Chi-Quadrat-Differenz-Tests für sb-skalierte Werte gerechnet. Diese Tests wurden in Unterkapitel 7.4 vorgestellt. Die Testparameter sind in Tabelle 20 aufgeführt. Die erste wertgefüllte Spalte in der Tabelle gibt die unskalierten, also durch den normalen maximum-likelihood-Schätzalgorithmus bestimmten, Chi-Quadrat-Werte an (siehe Anhang D.8.2), die darauf folgende Spalte die sb-skalierten Chi-Quadrat-Werte (siehe Anhang D.8.1). Die Spalte Skalierungsfaktor gibt das Verhältnis von unskalierten und skalierten Werten zueinander an. Dieser Wert kann direkt aus den MPlus-Output-Dateien abgelesen werden.

In den mit Differenz beschrifteten Zeilen sind die Differenz der unskalierten Chi-Quadrat-Werte sowie die Differenz der Freiheitsgrade der beiden gegeneinander zu testenden Modelle aufgeführt. Weiter rechts steht der Korrekturfaktor, der aus den Freiheitsgraden und den Skalierungsfaktoren der beiden Modelle errechnet wurde. Die entsprechende Formel ist dem Internet-Tutorial der offiziellen Homepage des SEM-Programms MPlus entnommen (MPlus-Homepage, 2005) und stellt eine vereinfachende Umformulierung der von Satorra und Bentler (1999, S. 7) entwickelten Formel dar. Die

korrigierte Chi-Quadrat-Differenz in der vorletzten Spalte der Tabelle ergibt sich durch Division der Differenz der unskalierten Chi-Quadrat-Werte durch den Korrekturfaktor. Im Gegensatz zur Differenz der unskalierten Werte ist die korrigierte Wertedifferenz Chi-Quadrat-verteilt mit einer Anzahl von Freiheitsgraden, die der Differenz der Freiheitsgrade der beiden zu vergleichenden Modelle entspricht. Auf Basis der korrigierten Chi-Quadrat-Differenz und der Freiheitsgraddifferenz kann daher die Irrtumswahrscheinlichkeit in der letzten Spalte der Tabelle berechnet werden. Sie bezieht sich auf die Hypothese, dass das restriktivere der beiden Modelle einen überzufällig schlechteren Fit besitzt als das andere Modell.

Tabelle 20: Parameter der Chi-Quadrat-Differenz-Tests für sb-skalierte Werte.

	unskaliertes χ^2	skaliertes χ^2	df	Skalierungs- faktor	Korrektur- faktor	korrigierte χ^2 -Differenz	p
Mzp 1							
Modell E1	16,133	16,976	18	0,949	-	-	-
Modell Z1	12,731	13,406	17	0,950	-	-	-
Differenz	3,402	-	1	-	0,932	3,650	,056
Mzp 2							
Modell E2	35,123	29,526	18	1,190	-	-	-
Modell Z2	33,672	29,095	17	1,157	-	-	-
Differenz	1,451	-	1	-	1,751	0,829	,363
Mzp 3							
Modell E3	42,748	31,265	18	1,367	-	-	-
Modell Z3	41,103	30,738	17	1,337	-	-	-
Differenz	1,645	-	1	-	1,877	0,876	,349

Anmerkung. N = 204 in allen Analysen.

Die genauen Berechnungen der Werte in der Tabelle sind in Anhang D.8.3 einsehbar. Die Irrtumswahrscheinlichkeiten zeigen, dass die Fitdifferenzen zwischen allen drei Paaren von Ein- und Zweifaktormodellen im Bereich der Zufallsschwankungen liegen.

Die Ursache für den gleich guten Fit der Einfaktor- und der Zweifaktormodelle liegt in den hohen Korrelationen des konzeptuellen und des prozeduralen Faktors in den Zweifaktormodellen. Die Korrelationskoeffizienten betragen ,85 für den ersten, ,94 für den zweiten und ,94 für den dritten Messzeitpunkt. Angesichts dieser extrem hohen und doch nicht perfekten Korrelationen lässt sich leicht verstehen, warum die Modellfitindizes keine eindeutige Entscheidung zwischen den Ein- und den Zweifaktormodellen ermöglichen. Für die weitere Auswertung bleibt also festzuhalten, dass die divergente Validität der Maße empirisch weder eindeutig bestätigt noch widerlegt werden konnte.

In einem solchen Fall ist eine Entscheidung auf Grundlage theoretischer Abwägungen zu fällen. Für die weitere Verwendung der Einfaktormodelle spricht, dass sie die Daten sparsamer, das heißt durch weniger Modellparameter, erklären können als die

Zweifaktormodelle. Für die Zweifaktormodelle spricht, dass die Unterscheidung von konzeptuellem und prozeduralem Wissen von großer theoretischer Bedeutung und von forschungsheuristischem Nutzen ist. Weitere Analysen der Beziehungen zwischen konzeptuellem und prozeduralem Wissen sind nur möglich, wenn die Zweifaktormodelle akzeptiert werden, da die Wissensarten nur in diesem Fall operationalisiert werden können. Daher werden die Zweifaktormodelle für die weiteren Analysen verwendet, jedoch immer unter dem Vorbehalt, dass die divergente Validität der Maße empirisch weder be- noch widerlegt werden konnte. Ein weiteres Argument für die Verwendung der Zweifaktormodelle liefern die im Folgenden beschriebenen explorativen Analysen.

Explorative Analysen der divergenten Validitäten

Die drei hohen Korrelationen zwischen den konzeptuellen und den prozeduralen Faktoren verdeutlichen, dass ein Teststärkeproblem vorliegt. Ein Zweifaktormodell mit zwei perfekt (d.h. mit $r = 1,00$) korrelierenden Faktoren ist äquivalent zu einem Einfaktormodell. Die empirisch gefundenen Korrelationen sind jedoch kleiner als eins und sprechen somit für die Richtigkeit des Zweifaktormodells – nur ist die Verschiedenheit der Koeffizienten vom Wert eins nicht statistisch absicherbar. Bei einer Erhöhung der Teststärke könnte sich das ändern.

Eine solche Erhöhung der Teststärke kann erzielt werden, indem mehrere Messzeitpunkte gemeinsam analysiert werden. In den Zweifaktormodellen, die sich nur auf einen Messzeitpunkt bezogen, waren empirisch gegeben die Varianzen der 8 manifesten Variablen sowie die 28 Kovarianzen dieser Variablen. Es konnten also 36 empirisch gegebene Größen zur Schätzung der Modellparameter verwendet werden. Analysiert man hingegen zwei Messzeitpunkte gemeinsam in einem Modell, so sind 16 Varianzen und 120 Kovarianzen empirisch gegeben. Bei drei Messzeitpunkten sind 24 Varianzen und 276 Kovarianzen gegeben. Dieser nichtlineare Zuwachs an Kovarianzen ergibt sich aus der quadratischen Form der Varianz-Kovarianz-Matrix. Er führt dazu, dass das Verhältnis zwischen gegebenen und zu schätzenden Parametern bei Modellen über mehrere Messzeitpunkte hinweg besser ist als bei Modellen über einen Messzeitpunkt.

Daher wurden Modell E1 und Modell E2 zu einem Modell E12 zusammengefasst und Modell Z1 und Modell Z2 zu einem Modell Z12. Modell E12 umfasst 16 Indikatoren und zwei latente Faktoren (einen Wissensfaktor pro Messzeitpunkt), Modell Z12 dieselben 16 Indikatoren, jedoch vier latente Faktoren (einen konzeptuellen und einen prozeduralen Faktor pro Messzeitpunkt).

Da nur das Mess-, nicht jedoch das Regressionsmodell getestet werden soll, wurden keine Regressionsbeziehungen zwischen den Messzeitpunkten spezifiziert.

Das Messwiederholungsdesign führt dazu, dass die Indikatorwerte und die Faktorwerte über die Messzeitpunkte hinweg natürlich nicht unabhängig voneinander sind. Daher wurden in Modell E12 und Modell Z12 jeweils Kovariationen zwischen den latenten Faktoren einer Wissensart über die Messzeitpunkte hinweg zugelassen. Auch die Residuen jedes Wissensmaßes dürfen über die Messzeitpunkte hinweg kovariieren.

In gleicher Weise wie die Modelle E12 und Z12 wurden dann auch die über alle drei Messzeitpunkte gehenden Modelle E123 und Z123 gebildet.

Tabelle 21 zeigt die Fitindizes dieser Modelle (für die Dateien siehe Anhang D.8.1). Der Übersichtlichkeit wegen werden sie noch einmal zusammen mit den Fitindizes der schon analysierten Modelle aufgelistet.

Tabelle 21: Fit der Modelle zur Exploration der divergenten Validitäten.

Faktoren pro Mzp	Modell	χ^2	df	p	CFI	WRMR	RMSEA
Mzp 1							
1 Faktor	E1	16,976	18	,525	1,000	0,680	0,000 ^a
2 Faktoren	Z1	13,406	17	,709	1,000	0,633	0,000 ^a
Mzp 2							
1 Faktor	E2	29,526	18	,042	0,960	0,899	0,056
2 Faktoren	Z2	29,095	17	,034	0,958	0,910	0,059
Mzp 3							
1 Faktor	E3	31,265	18	,027	0,962	0,768	0,060
2 Faktoren	Z3	30,738	17	,022	0,961	0,797	0,063
Mzp 1 & 2							
1 Faktor	E12	146,407	91	,000	0,941	1,439	0,055
2 Faktoren	Z12	127,277	86	,003	0,956	1,356	0,049
Mzp 1, 2 & 3							
1 Faktor	E123	306,754	219	,000	0,951	1,389	0,044
2 Faktoren	Z123 ^b	280,095	207	,000	0,959	1,324	0,042

Anmerkungen. ^a Der Koeffizient wurde manuell auf null gesetzt, da $df > \chi^2$;

^b Angaben zu diesem Modell unter Vorbehalt, Erläuterung im Text.

CFI, WRMR und RMSEA der Ein- und der Zweifaktormodelle liegen noch immer nah beieinander. Die in Tabelle 22 aufgeführten Parameter der Chi-Quadrat-Differenz-Tests für sb-skalierte Werte (siehe die Anhänge D.8.1, D.8.2 und D.8.3) zeigen jedoch, dass die Einfaktormodelle E12 und E123 jeweils einen signifikant schlechteren Fit besitzen als das zugehörige Zweifaktormodell Z12 oder Z123.

Die Modelle E123 und Z123 unterscheiden sich mit einer größeren Irrtumswahrscheinlichkeit als die Modelle E12 und Z12, obwohl der Test über alle drei Messzeitpunkte eine höhere Teststärke aufweist als der Test über zwei Messzeitpunkte.

Tabelle 22: Parameter der explorativen Chi-Quadrat-Differenz-Tests (sb-skaliert).

Modell	unskaliertes χ^2	skaliertes χ^2	df	Skalierungsfaktor	Korrekturfaktor	korrigierte χ^2 -Differenz	p
Mzp 1 & 2							
Modell E12	150,517	146,407	91	1,028	-	-	-
Modell Z12	130,687	127,277	86	1,027	-	-	-
Differenz	19,830	-	5	-	1,045	18,972	,002
Mzp 1, 2 & 3							
Modell E123	330,330	306,754	219	1,077	-	-	-
Modell Z123 ^a	301,051	280,095	207	1,075	-	-	-
Differenz	29,279	-	12	-	1,112	26,342	,010

Anmerkungen. ^a Angaben unter Vorbehalt, Erläuterung im Text; N = 204 in allen Analysen.

Diese Befunde gelten jedoch nur unter Vorbehalt. Bei der Schätzung der Parameter von Modell Z123 gab das Analyseprogramm eine Warnmeldung aus, die auf unspezifische Konvergenzprobleme des Schätzalgorithmusses hinweist. Die Modellfitindizes können daher nicht als gesichert angesehen werden. Außerdem ist die divergente Validität ein Konstrukt, das sich auf einzelne Tests bezieht, nicht auf komplexe Messwiederholungsdesigns als Ganzes. Schließlich ändern die Befunde auch nichts an den hohen Interkorrelationen zwischen den latenten Wissensfaktoren. Dennoch zeigen sie, dass die oben theoretisch-argumentativ gerechtfertigte Entscheidung für die weitere Verwendung der Zweifaktormodelle nicht gänzlich an der empirisch erfassbaren Wirklichkeit vorbeigeht.

9.3.5 Prädiktive Relationen

Nachdem nun also das Messmodell für die Wissensarten etabliert wurde, können die prädiktiven Beziehungen zwischen den Wissensarten über die Messzeitpunkte hinweg analysiert werden. Alle dazugehörigen Syntax- und Outputdateien sind in Anhang D.9 zusammengestellt.

Zuerst wurde ein Modell R123 spezifiziert, das dem in Abbildung 8 dargestellten Grundmodell entspricht. Pro Messzeitpunkt gibt es einen konzeptuellen und einen prozeduralen Wissensfaktor, die korreliert sein dürfen, und die Ausprägungen der Werte beider Faktoren zu einem Messzeitpunkt werden durch die Ausprägungen der Werte beider Faktoren zum jeweils vorgehenden Messzeitpunkt vorhergesagt. Es handelt sich also um das oben diskutierte cross-lagged panel design.

Das Analyseprogramm lieferte bei der Schätzung der Parameter von Modell R123 eine Warnmeldung, die auf unbestimmte Probleme bei der Schätzung hinwies. Das Modell ist darum für eine hypothesentestende Auswertung nicht geeignet. Da die Natur der Probleme unbekannt ist, sollten auch explorative Analysen nur unter sehr starken Vorbehalten durchgeführt werden. Daher wurde hier auf eine Präsentation und Diskussion des Modells verzichtet. Die Ergebnisse der Parameterschätzung sind dennoch in Anhang D.9 einsehbar.

Der wahrscheinlichste Grund für die Probleme ist, dass die Faktoren K2 und P2 jeweils beide benutzt wurden, um die Faktoren K3 und P3 vorherzusagen. Da sowohl K2 und P2 als auch K3 und P3 untereinander sehr hoch korreliert sind, war eine Bestimmung der Regressionspfadkoeffizienten vermutlich nicht möglich. In diesem Fall hätten sich die niedrigen divergenten Validitäten der Maße hinderlich auf die Analyse des Regressionsmodells ausgewirkt.

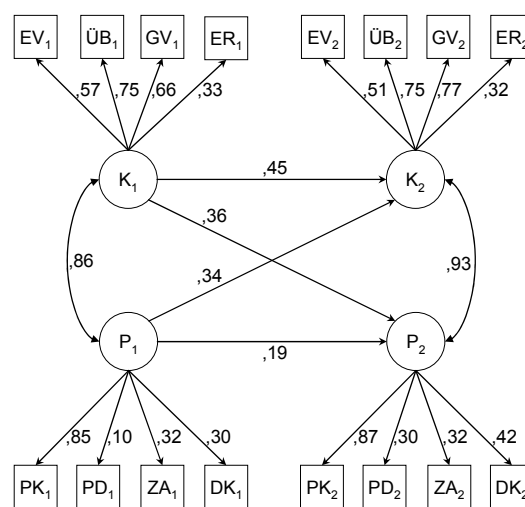
Um dieser Annahme nachzugehen, wurde ein Modell R12 spezifiziert, das nur den Ausschnitt von Modell R123 umfasst, der sich auf die ersten beiden Messzeitpunkte bezieht. Wenn die Regressionsbeziehungen zwischen den Faktoren des zweiten und dritten Messzeitpunktes für die Probleme verantwortlich waren, sollten die Probleme bei Modell P12 nicht auftreten.

Tatsächlich ließen sich die Parameter von Modell R12 problemlos schätzen. In den Outputdateien in Anhang D.9 finden sich die vollständige Interkorrelationsmatrix sowie die Varianz-Kovarianz-Matrix aller manifesten Wissensmaße. Der Chi-Quadrat-Wert beträgt $\chi^2 = 127,277$ bei $df = 86$ Freiheitsgraden und einer Wahrscheinlichkeit von $p = 0,003$. Der CFI beträgt 0,956, der WRMR 1,356 und der RMSEA 0,049. Der CFI und der RMSEA zeigen einen guten Modellfit an. Der WRMR liegt leicht über dem cut-off-Punkt von 1,000 und zeigt dadurch, dass das Modell zwar gut, aber nicht perfekt an die Daten angepasst ist.

Das Pfaddiagramm von Modell R12 in Abbildung 9 beinhaltet die standardisierten Koeffizienten der Faktorladungen, der Faktorinterkorrelationen und der Regressionspfade. Häufig werden in Pfaddiagrammen statt der Interkorrelationen Kovarianzen dargestellt. Für die Fragestellung dieser Arbeit erscheinen die Korrelationen jedoch relevanter (vgl. den Abschnitt über die divergenten Validitäten). Man beachte, dass die Kovarianzen zwischen den Residuen der Maßen gleicher und unterschiedlicher Messzeitpunkte nicht mitdargestellt werden. Wie oben beschrieben, wurden Kovarianzen zwischen den Maßen Evaluation und Erklärung (wegen des Einflusses der Sprachkompetenz), den Maßen Problemlösedauer und Zugriffsasymmetrie (wegen des Einflusses der Zeitmessung) während eines Messzeitpunktes und zwischen allen drei Messungen jedes Maßes über die drei Messzeitpunkte hinweg (wegen der Abhängigkeiten aufgrund der Messwiederholungen) zugelassen.

Die standardisierten Koeffizienten von Modell R12, die unstandardisierten Koeffizienten mit ihren Standardfehlern, die Quotienten z von unstandardisierten Koeffizienten und Standardfehlern sowie die Irrtumswahrscheinlichkeiten p und die

Signifikanzen der z-Werte sind in Tabelle 23 zusammengestellt. Die Werte in den ersten vier Spalten wurden mit MPlus berechnet, die Irrtumswahrscheinlichkeiten der z-Werte mittels einer Excel-Tabelle (s. Anhang D.9). Die Irrtumswahrscheinlichkeiten beziehen sich auf die Hypothese, dass der jeweilige Koeffizient überzufällig größer als null ist. Für die Faktorinterkorrelationen gibt MPlus leider keine Signifikanzen aus (Muthén & Muthén, 1998-2004b). Gleiches gilt für die Faktorindikatoren mit den auf eins fixierten Ladungen. Beides stellt hier keinen Nachteil dar, weil konvergente und divergente Validitäten ja schon ausführlich analysiert wurden.



Anmerkungen. Fehlerterme und Kovarianzen zwischen Residuen nicht mitdargestellt.
 K: konzeptuelles Wissen; P: prozedurales Wissen; EV: Evaluation; ÜB: Übersetzung;
 GV: Größenvergleich; ER: Erklärung; PK: Problemlösekorrektheit; PD: Problemlösedauer;
 ZA: Zugriffsasymmetrie; DK: Dual task-Kosten.

Abbildung 9: Pfaddiagramm von Modell R12.

Dass die in Tabelle 18 (s. S. 48) und die hier berichteten Faktorladungen und Faktorinterkorrelationen nicht vollständig übereinstimmen, erklärt sich daraus, dass sie vorher einzeln und für Modell P12 alle gemeinsam geschätzt wurden. Die Unterschiede sind jedoch minimal und rein quantitativer Natur. An der Qualität der oben diskutierten Befunde ändert sich nichts.

Bisher noch nicht analysiert wurden die Regressionspfade zwischen den latenten Faktoren. Am stärksten ausgeprägt ist der Koeffizient des Pfades von K1 zu K2 mit 0,452. Die Koeffizienten der Pfade von K1 zu P2 und von P1 zu K2 sind mit Werten von 0,341 und 0,363 etwa gleich groß. Am schwächsten ist der Koeffizient des Pfades von P1 zu P2 mit einem Wert von 0,193 ausgeprägt. P1 und K1 klären zusammen 58,3 Prozent der Varianz von Faktor K2 und 28,9 Prozent der Varianz von Faktor P2 auf.

Tabelle 23: Parameter von Modell R12 mit Standardfehlern, z-Werten und Irrtumswahrscheinlichkeiten.

Relation	Standardisierter Koeffizient	Unstandardisierter Koeffizient	S.E.	z	p	Sign.
Faktorladungen						
Mzp 1						
Konzeptuelles Wissen (K1)						
Evaluation	0,574	1,000	0,000	-	-	-
Übersetzung	0,752	1,065	0,130	8,222	0,000	***
Größenvergleich	0,663	0,878	0,103	8,557	0,000	***
Erklärung	0,331	0,553	0,117	4,739	0,000	***
Prozedurales Wissen (P1)						
Problemlösekorrektheit	0,852	1,000	0,000	-	-	-
Problemlösedauer	0,100	0,025	0,019	1,270	0,102	n.s.
Zugriffsasymmetrie	0,324	0,047	0,011	4,142	0,000	***
Dual task-Kosten	-0,303	-0,321	0,073	-4,376	0,000	***
Mzp 2						
Konzeptuelles Wissen (K2)						
Evaluation	0,509	1,000	0,000	-	-	-
Übersetzung	0,746	1,291	0,179	7,213	0,000	***
Größenvergleich	0,767	0,888	0,116	7,648	0,000	***
Erklärung	0,320	0,651	0,131	4,954	0,000	***
Prozedurales Wissen (P2)						
Problemlösekorrektheit	0,871	1,000	0,000	-	-	-
Problemlösedauer	0,296	0,064	0,023	2,733	0,003	**
Zugriffsasymmetrie	0,322	0,042	0,014	2,983	0,001	**
Dual task-Kosten	-0,421	-0,439	0,089	-4,941	0,000	***
Korrelation						
K1 und P1	0,855	– ^a	– ^a	– ^a	-	-
K2 und P2	0,934	– ^a	– ^a	– ^a	-	-
Regressionspfade						
K1 zu K2	0,452	0,416	0,223	1,869	0,031	*
P1 zu K2	0,341	0,208	0,158	1,316	0,094	+
K1 zu P2	0,363	0,429	0,344	1,249	0,106	ns.
P1 zu P2	0,193	0,151	0,249	0,605	0,273	ns.

Anmerkung. ^a Diese Parameter werden von MPlus nicht berechnet.

Den aus dem Iterativen Modell abgeleiteten Hypothesen zufolge sollten die Koeffizienten der Pfade von K1 zu P2 und P1 zu K2 etwa gleich stark ausgeprägte Koeffizienten besitzen. Das ist hier der Fall. Jedoch ist nur der Koeffizient des Pfads von K1 zu K2 signifikant. Wie schon bei der Untersuchung der divergenten Validitäten wäre hier eine größere Teststärke hilfreich.

Vor der Interpretation dieser Befunde wird nun untersucht, wie die Kovariaten die Wissensentwicklung beeinflussen.

9.4 Ergebnisse explorativer Analysen

9.4.1 Intelligenz und die Wissensarten

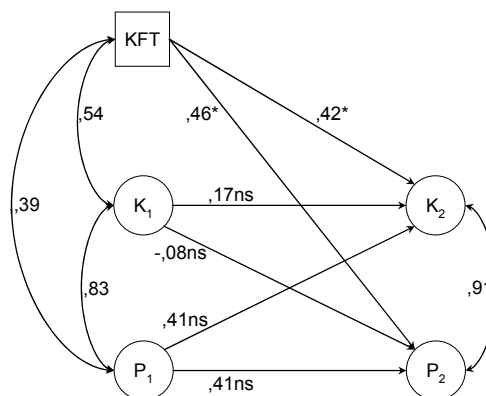
Für die Analysen wurde der Datensatz SEMData-KFT.dat aus Anhang D.3 verwendet. Die Syntax- und Outputdateien sind in Anhang D.10 einsehbar.

Intelligenz als Kovariate im Grundmodell

Um zu untersuchen, wie Intelligenz und bereichsspezifisches konzeptuelles und prozedurales Vorwissen beim Wissenserwerb ineinandergreifen und welches der wichtigste Prädiktor des Lernerfolgs ist, wurde ein Modell I12 (mit I für Intelligenz) aus Modell R12 abgeleitet. Die KFT-Gesamtscores wurden als Kovariate in das Modell eingebracht und gemeinsam mit dem konzeptuellen und prozeduralen Wissen vor der Intervention als Prädiktor des Wissens nach der Intervention verwendet.

Da missing values in der Analyse nicht zulässig waren, wurden von den 204 bisher analysierten Personen 9 von den weiteren Analysen ausgeschlossen, weil sie den Intelligenztest nicht oder nicht vollständig bearbeitet hatten. Die folgenden Analysen beziehen sich also auf eine Stichprobengröße von $N = 195$.

Modell I12 hat einen Modellfit von $\chi^2 = 137,926$ mit $df = 98$ Freiheitsgraden und einer Wahrscheinlichkeit von $p = 0,005$ (CFI = 0,962; WRMR = 1,290; RMSEA = 0,046) und ist somit, wie schon Modell R12, gut, jedoch nicht perfekt an die Daten angepasst. Abbildung 10 zeigt die Struktur des Modells. Die genauen Koeffizienten und ihre Signifikanzen können Tabelle 24 entnommen werden. Die erste Spalte der Tabelle enthält zum Vergleich noch einmal die standardisierten Koeffizienten von Modell R12.



Anmerkungen. Indikatoren, Residuen, Kovarianzbeziehungen zwischen Residuen und Signifikanz der Korrelationen nicht mitdargestellt; K: konzeptuelles Wissen; P: prozedurales Wissen.

Abbildung 10: Pfaddiagramm von Modell I12.

Der KFT-Wert korreliert am ersten Messzeitpunkt hochsignifikant mit beiden Wissensarten. Die Korrelation zwischen KFT und konzeptuellem Wissen fällt höher aus als die zwischen KFT und prozeduralem. Beide Korrelationen sind deutlich niedriger als die Korrelation der Wissensfaktoren untereinander.

Tabelle 24: Koeffizienten von Modell R12 sowie von Modell I12 mit Standardfehlern, z-Werten und Irrtumswahrscheinlichkeiten.

	Modell R12	Modell I12					
	Standardisierter Koeffizient	Standardisierter Koeffizient	Unstandardisierter Koeffizient	S.E.	z	p	Sign.
Korrelationen							
K1 und P1	0,855	0,830	– ^a	– ^a	– ^a	–	–
KFT und K1	–	0,544	– ^a	– ^a	– ^a	–	–
KFT und P1	–	0,389	– ^a	– ^a	– ^a	–	–
K2 und P2	0,934	0,911	– ^a	– ^a	– ^a	–	–
Regressionspfade							
K1 zu K2	0,452	0,166	0,158	0,246	0,644	0,260	n.s.
P1 zu K2	0,341	0,411	0,254	0,164	1,547	0,061	+
KFT zu K2	–	0,424	0,604	0,118	5,120	0,000	***
K1 zu P2	0,363	–0,083	–0,098	0,338	–0,289	0,614	n.s.
P1 zu P2	0,193	0,409	0,313	0,238	1,316	0,094	+
KFT zu P2	–	0,462	0,815	0,144	5,659	0,000	***

Anmerkung. ^a Diese Parameter werden von MPlus nicht berechnet.

Die Koeffizienten der Regressionspfade zeigen, dass die Intelligenz sich etwa gleich stark und hochsignifikant auf das konzeptuelle und das prozedurale Wissen im Posttest auswirkte.

Ein Vergleich der Pfadkoeffizienten von Modell R12 und Modell I12 zeigt, dass sich bei Einführung der Kovariate KFT in das Modell die Beziehungen zwischen den latenten Wissensfaktoren ändern. Der Einfluss des konzeptuellen Wissens im Prätest (K1) auf das konzeptuelle und prozedurale Wissen im Posttest (K2 und P2) sinkt stark ab. Das prozedurale Posttestwissen entwickelt sich nach Modell I12 völlig unbeeinflusst durch das konzeptuelle Vorwissen.

Im Vergleich zu Modell R12 steigt in Modell I12 der Einfluss des prozeduralen Vorwissens (P1) auf das konzeptuelle Posttestwissen (K2) und vor allem auf das prozedurale Posttestwissen (P2) an.

Während der gesteigerte Einfluss des prozeduralen Prätestwissens bei Kontrolle der Intelligenz noch erklärt werden kann, ist das völlige Verschwinden des Einflusses von K1 auf P2 nicht mehr sinnvoll erklärbar. Dass konzeptuelles Vorwissen sich günstig auf den Erwerb prozeduralen Wissens auswirkt, ist ein plausibler und oft replizierter Befund. Viele Theorien des Fähigkeitserwerbs bauen auf ihn auf (vgl. Abschnitt 5.4.2). Angesichts der hohen Interkorrelationen der drei unabhängigen Variablen (K1, P1 und KFT) deutet dieser kontraintuitive Befund auf das Vorliegen von suppressionsähnlichen Effekten (Bortz, 1993, S. 423-426) hin. Die Modellparameter lassen keinesfalls den sicheren Schluss zu, dass konzeptuelles Vorwissen sich bei Kontrolle der Intelligenz nicht auf prozedurale Wissenszuwächse auswirkt.

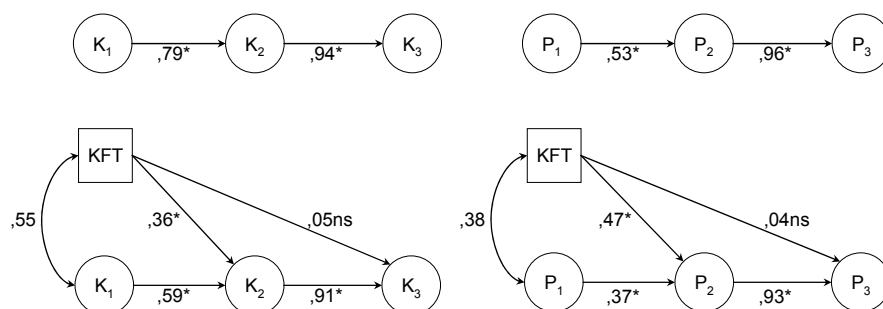
Einzelmodelle für Intelligenz und Wissensarten

Um sicherere Befunde über den Einfluss der Intelligenz auf die Wissensentwicklung zu gewinnen und das Problem der hohen Interkorrelationen der Prädiktoren zu reduzieren, wurden im Folgenden Einzelmodelle analysiert, die jeweils auf die Entwicklung einer Wissensart fokussieren.

Die hohen Interkorrelationen zwischen konzeptuellem und prozeduralem Wissen sind der Grund dafür, warum der dritte Messzeitpunkt nicht mitanalysiert werden konnte. Sie führten zu Konvergenzproblemen. In den folgenden Einzelanalysen entfällt dieses Problem der hohen Interkorrelationen. Die Modelle spezifizieren daher die Relationen über alle drei Messzeitpunkte.

Es wurden ein Modell K123 mit den drei latenten Faktoren K1, K2 und K3 ohne Kovariate und ein entsprechendes Modell IK123 mit Kovariate KFT spezifiziert. In gleicher Weise wurden für das prozedurale Wissen ein Modell P123 ohne und ein Modell IP123 mit dem KFT als Kovariate spezifiziert.

Die vier Modelle mit den geschätzten Koeffizienten sind in Abbildung 11 dargestellt. Die zugehörigen Fitindizes gibt Tabelle 25 wieder. CFI, WRMR und RMSEA zeigen für alle vier Modelle einen sehr guten Fit an. Die Wissensarten lassen sich einzeln betrachtet also über alle drei Messzeitpunkte hinweg modellieren. Dies belegt die oben getroffene Annahme, dass die Konvergenzprobleme von Modell R123 (zur Analyse der Relationen beider Wissensarten über alle drei Messzeitpunkte hinweg) durch die niedrigen divergenten Validitäten der Wissensmaße verursacht wurden.



Anmerkungen. Indikatoren, Fehlerterme, Kovarianzen zwischen Residuen sowie Signifikanz der Korrelationen nicht mitdargestellt; K: konzeptuelles Wissen; P: prozedurales Wissen.

Abbildung 11: Pfaddiagramme (links oben : Modell IK123, links unten: Modell K123, rechts oben: Modell IP123, rechts unten: Modell P123).

Durch Vergleich der beiden oberen mit den beiden unteren Modellen in Abbildung 11 lässt sich erkennen, wie die Kontrolle der Intelligenz die Regressionsbeziehungen zwischen den Wissensfaktoren verändert. Das Wissen zum zweiten Messzeitpunkt wird sowohl durch

das jeweilige Vorwissen als auch durch die Intelligenz beeinflusst. Entsprechend wird der Einfluss von K1 auf K2 und von P1 auf P2 kleiner, wenn die Intelligenz mit in das jeweilige Modell aufgenommen wird.

Vom zweiten zum dritten Messzeitpunkt bleiben die interindividuellen Wissensunterschiede sehr stabil und die Intelligenz hat keinen signifikanten Einfluss auf die Zuwächse an konzeptuellem oder prozeduralem Wissen.

Tabelle 25: Fitindizes der Modelle zur Analyse des KFT-Einflusses.

	χ^2	df	p	CFI	WRMR	RMSEA
Konzeptuelles Wissen						
Modell K123	45,447	37	,161	0,992	0,918	0,034
Modell IK123	57,121	46	,126	0,991	0,921	0,035
Prozedurales Wissen						
Modell P123	50,607	37	,076	0,961	0,816	0,043
Modell IP123	61,059	46	,068	0,965	0,795	0,041

In Modell IK123 ist das Vorwissen (K1) ein besserer Prädiktor für das Wissen zum zweiten Messzeitpunkt (K2) als die Intelligenz. In Modell IP123 ist es genau umgekehrt: Die Intelligenz ist vom ersten zum zweiten Messzeitpunkt ein besserer Prädiktor des Lernerfolgs als das Vorwissen.

9.4.2 Mentale Zahlenrepräsentation und die Wissensarten

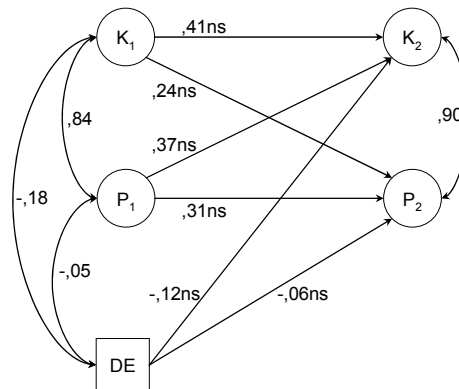
Es wurden $N = 194$ Fälle verwendet, da einige Personen aus Zeitgründen nicht an der Erhebung des distance effect teilnehmen konnten. Für die Analysen wurde Datensatz SEMData-DE.dat aus Anhang D.3 verwendet. Die Syntax- und Outputdateien sind in Anhang D.11 zusammengestellt.

Die Berechnung der Stärke des distance effects erfolgte wie in Abschnitt 9.2.7 beschrieben. Die semantische Distanz zwischen zu beurteilender Zahl und Referenzzahl hatte bei 61% der Teilnehmer einen signifikanten Einfluss auf die zur Beurteilung benötigte Zeit. Sie erklärte dabei pro Person zwischen 0% und 48% der Varianz der Beurteilungszeiten. Die gemittelte Varianzaufklärung über alle Personen beträgt 13%.

Zur Auswertung in MPlus wurde ein Modell D12 verwendet, das Modell I12 gleicht, abgesehen davon, dass statt des KFT-Werts die Ausprägung des distance effects als Kovariate benutzt wurde.

Das Modell und seine geschätzten Koeffizienten zeigt Abbildung 12. Das Modell hat einen zufriedenstellenden Fit von $\chi^2 = 130,674$ mit $df = 98$ Freiheitsgraden und einer Wahrscheinlichkeit von $p = ,015$ (CFI = 0,966; WRMR = 1,277; RMSEA = 0,041). Die

interindividuellen Unterschiede im distance effect weisen keinerlei signifikante Relationen zu den latenten Faktoren konzeptuellen oder prozeduralen Wissens auf.



Anmerkungen. Indikatoren, Residuen, Kovarianzbeziehungen zwischen Residuen sowie Signifikanz der Korrelationen nicht mitdargestellt; K: konzeptuelles Wissen; P: prozedurales Wissen; DE: Distance effect.

Abbildung 12: Pfaddiagramm von Modell D12.

9.4.3 Messinvarianz

Um zu überprüfen, ob das Messmodell für den ersten und das für den zweiten Messzeitpunkt sich signifikant voneinander unterscheiden, wurde Modell R12V (V für variierende Koeffizienten) mit einem Modell R12F (mit F für fixierte Koeffizienten) verglichen. Bei ersterem durften die Faktorladungen frei variieren. Bei Letzterem wurden die Faktorladungen jedes Maßes über die beiden Messzeitpunkte hinweg gezwungen, denselben Wert anzunehmen. Dieses Vorgehen wird auch als Parameterfixierung bezeichnet. Beispielsweise wurde in der Syntax angegeben, dass die Ladung des Maßes Evaluation vom ersten Messzeitpunkt auf dem Faktor K1 und die Ladung des Maßes Evaluation vom zweiten Messzeitpunkt auf Faktor K2 genau gleich groß sein sollen. Entsprechend wurde für alle Maße vorgegangen. Falls die Faktorladungen vom ersten und vom zweiten Messzeitpunkt sich überzufällig stark unterscheiden, sollte der Fit des Modells mit gleichgesetzten Koeffizienten signifikant schlechter sein als der Fit des Modells mit frei variierbaren Parametern.

Tabelle 26: Fitindizes der Modelle zur Analyse der Messinvarianz.

	χ^2	df	p	CFI	WRMR	RMSEA
Konzeptuelles und prozedurales Wissen an Mzp 1 & 2						
Modell R12V	127,277	86	,003	0,956	1,356	0,049
Modell R12F	133,614	92	,003	0,956	1,441	0,047
Konzeptuelles Wissen an Mzp 1, 2 & 3						
Modell K123V	51,115	37	,061	0,987	0,988	0,043
Modell K123F	65,344	43	,016	0,979	1,126	0,050
Prozedurales Wissen an Mzp 1, 2 & 3						
Modell P123V	51,008	37	,063	0,962	0,864	0,043
Modell P123F	54,036	43	,121	0,970	0,999	0,035

Tabelle 27: Parameter des Chi-Quadrat-Differenz-Tests auf Messinvarianz.

Modell	unskaliertes χ^2	skaliertes χ^2	df	Skalierungsfaktor	Korrekturfaktor	korrigierte χ^2 -Differenz	p
Konzeptuelles und prozedurales Wissen an Mzp 1 & 2							
Modell R12F	137,454	133,614	92	1,029	-	-	-
Modell R12V	130,687	127,277	86	1,027	-	-	-
Differenz	6,767	-	6	-	1,058	6,398	0,380
Konzeptuelles Wissen an Mzp 1, 2 & 3							
Modell K123F	65,380	65,344	43	1,001	-	-	-
Modell K123V	52,178	51,115	37	1,021	-	-	-
Differenz	13,202	-	6	-	0,878	15,042	0,020
Prozedurales Wissen an Mzp 1, 2 & 3							
Modell P123F	80,080	54,036	43	1,482	-	-	-
Modell P123V	72,198	51,008	37	1,415	-	-	-
Differenz	7,882	-	6	-	1,895	4,159	0,655

Anmerkung. N = 204 in allen Analysen.

Zur Analyse benutzt wurde Datensatz SEMData.dat aus Anhang D.3. Die Analysedateien befinden sich in Anhang D.12. Die oberen beiden Zeilen von Tabelle 26 geben die Fitindizes von Modell R12V und Modell R12F wieder. In Tabelle 27 sind die Parameter des zugehörigen Chi-Quadrat-Differenz-Tests für sb-skalierte Daten aufgelistet.

Der Test zeigt, dass der Modellfit durch die Gleichsetzung der Faktorladungen über die ersten beiden Messzeitpunkte hinweg nicht signifikant schlechter wird. Für diese beiden Messzeitpunkte liegt also metrische Messinvarianz vor. Die entsprechenden Faktorladungen gibt Tabelle 28 wieder. Die unstandardisierten Faktorladungen sind über die Messzeitpunkte hinweg pro Wissensmaß identisch. Die standardisierten Koeffizienten unterscheiden sich trotzdem, da die Wissensmaße unterschiedliche Varianzen aufweisen. Durch die Gleichsetzung der Faktorladungen werden die Regressionskoeffizienten kaum verzerrt. Sie bleiben jeweils im selben Wertebereich und sind nach wie vor zwar der Effektstärke nach, nicht aber der Signifikanz nach von null verschieden.

Die Prüfung der Messinvarianz über alle drei Messzeitpunkte kann wegen der Konvergenzprobleme von Modell R123 nur für jede Wissensart separat durchgeführt werden. Für konzeptuelles Wissen wurden dazu das Modell K123V mit variierenden Faktorladungen und das Modell K123F mit fixierten Faktorladungen spezifiziert, für prozedurales Wissen die Modelle P123V und P123F.

Die Fitindizes dieser Modelle sind in den unteren Zeilen von Tabelle 26 (s. S. 175) aufgeführt, die Chi-Quadrat-Differenz-Tests für die sb-skalierten Werte im unteren Teil von Tabelle 27 (s. S. 176). Die Tests belegen, dass eine Messinvarianz im engeren Sinne über alle drei Messzeitpunkte zwar für prozedurales, nicht jedoch für konzeptuelles Wissen gegeben ist. Da die Messinvarianz für die ersten beiden Messzeitpunkte für beide Wissensarten nachgewiesen wurde, kann das nur bedeuten, dass sich die Faktorladungen

der Ladungen des konzeptuellen Wissensfaktors vom zweiten zum dritten Messzeitpunkt signifikant verändern. Die entsprechenden Faktorladungen in Tabelle 18 sind relativ homogen verteilt. Sie geben keinen Hinweis darauf, dass ein einzelnes Maß für die signifikanten Unterschiede verantwortlich sein könnte, sondern machen es durch ihre relativ homogene Verteilung eher wahrscheinlich, dass es das Gesamtmuster ist, das sich graduell geändert hat. An den Faktorladungen lässt sich auch erkennen, dass die Effektstärke der Schwankungen eher gering sein dürfte. Für diese Annahme spricht auch, dass Modell K123F trotz der fixierten Parameter einen sehr guten CFI- Wert von 0,979 und einen guten RMSEA-Wert von 0,050 besitzt und lediglich der WRMR mit einem Wert von 1,126 einen leichten Missfit indiziert. Da Modell M123F angesichts dieser Indizes nicht verworfen zu werden braucht, liegt metrische Messinvarianz im weiteren Sinne für beide Wissensarten und alle drei Messzeitpunkte vor.

Tabelle 28: Koeffizienten von Modell R12F mit fixierten Faktorladungen.

Relation	Standardisierter Koeffizient	Unstandardisierter Koeffizient	S.E.	z	p	Sign.
Faktorladungen						
Mzp 1						
Konzeptuelles Wissen (K1)						
Evaluation	0,547	1,000	0,000	-	-	-
Übersetzung	0,781	1,191	0,125	9,502	0,000	***
Größenvergleich	0,628	0,864	0,088	0,815	0,208	n.s.
Erklärung	0,336	0,595	0,110	5,386	0,000	***
Prozedurales Wissen (P1)						
Problemlösekorrektheit	0,813	1,000	0,000	-	-	-
Problemlösedauer	0,194	0,052	0,018	2,873	0,002	**
Zugriffsasymmetrie	0,292	0,044	0,009	4,713	0,000	***
Dual task-Kosten	-0,347	-0,394	0,059	6,712	0,000	***
Mzp 2						
Konzeptuelles Wissen (K2)						
Evaluation	0,522	1,000	0,000	-	-	-
Übersetzung	0,728	1,191	0,125	9,502	0,000	***
Größenvergleich	0,770	0,864	0,088	9,815	0,000	***
Erklärung	0,303	0,595	0,110	5,386	0,000	***
Prozedurales Wissen (P2)						
Problemlösekorrektheit	0,833	1,000	0,000	-	-	-
Problemlösedauer	0,247	0,052	0,018	2,873	0,002	**
Zugriffsasymmetrie	0,340	0,044	0,009	4,713	0,000	***
Dual task-Kosten	-0,390	-0,394	0,059	-6,712	0,000	***
Korrelation						
K1 und P1	0,882	– ^a	– ^a	– ^a	-	-
K2 und P2	0,930	– ^a	– ^a	– ^a	-	-
Regressionspfade						
K1 zu K2	0,368	0,372	0,315	1,179	0,119	n.s.
P1 zu K2	0,423	0,283	0,220	1,284	0,100	+
K1 zu P2	0,355	0,427	0,452	0,944	0,173	n.s.
P1 zu P2	0,221	0,186	0,318	0,585	0,279	n.s.

Anmerkung. ^a Diese Parameter werden von MPlus nicht berechnet.

9.4.4 Stabilität der Faktorinterkorrelationen

Es wurde bereits berichtet, dass die latenten Faktoren hohe Interkorrelationen aufweisen (vgl. die Befunde zu den Modellen Z1, Z2 und Z3 auf Seite 164).

Im Folgenden wird untersucht, ob diese Korrelationen niedriger werden, wenn man die Intelligenz und das Alter der Kinder kontrolliert. Dazu wurden die durch Mplus ermittelten Faktorwerte für die jeweils zwei latenten Faktoren in den Modellen Z1, Z2 und Z3 an den Datensatz FinalData1.sav angehängt. Der resultierende Datensatz FinalData3.sav (s. Anhang D.3) erlaubt die Untersuchung der Zusammenhänge zwischen den Faktorwerten und den Hintergrundmerkmalen in SPSS. Die Analysedateien befinden sich in Anhang D.13.

Die Werte in Tabelle 29 zeigen, dass die Korrelationen zwischen dem konzeptuellen und dem prozeduralen Faktor nicht nennenswert niedriger werden, wenn Alter und Intelligenz kontrolliert werden. Die mit SPSS bestimmten Werte in der ersten Zeile der Tabelle sind höher als die auf Seite 164 berichteten mit MPlus bestimmten Faktorinterkorrelationen. Der Grund hierfür ist, dass SPSS die Korrelation analytisch aus den beiden Vektoren von Faktorwerten ableitet, während MPlus erstens einen iterativen Schätzalgorithmus und zweitens die Varianz-Kovarianz-Matrix zur Parameterschätzung verwendet.

Tabelle 29: Faktorinterkorrelationen bei Kontrolle von Intelligenz und Alter.

Kontrollierte Variable	Mzp 1	Mzp 2	Mzp 3
-	,94	,99	,99
KFT	,93	,98	,98
Alter	,94	,99	,99
KFT & Alter	,93	,98	,98

9.4.5 Einfluss von Personencharakteristika

Datensatz FinalData3.sav (Anhang D.3) wurde im Folgenden benutzt, um den Einfluss verschiedener Personencharakteristika auf die Wissensarten zu untersuchen. Hierbei wurde auf die Verwendung von MPlus verzichtet, da die Personencharakteristika teilweise missing values aufweisen, die man in SPSS leichter berücksichtigen kann als in MPlus.

Die zuvor mittels MPlus bestimmten Faktorwerte der latenten Faktoren wurden in univariaten Regressionen beziehungsweise einfaktoriellen Varianzanalysen auf einen einzelnen Prädiktor pro Analyse zurückgeführt. Da die interindividuellen Personenunterschiede über den zweiten und den dritten Messzeitpunkt außerordentlich stabil sind (vgl. die Modelle K123 und P123 in Abbildung 11 auf Seite 173), wurden die Analysen nur für die ersten beiden Messzeitpunkte, also für die Faktoren K1, P1, K2 und

P2, durchgeführt. In Tabelle 30 werden die Anteile der durch jeden Prädiktor aufgeklärten Varianz R^2 , die Irrtumswahrscheinlichkeiten p für die Hypothese, dass der erklärte Varianzanteil überzufällig größer als null ist, und die Signifikanzen dieser Hypothese dargestellt. Die Syntax- und Outputdateien befinden sich in Anhang D.14.

Die besten Prädiktoren für das Wissen an beiden Messzeitpunkten sind die Mathematiknote und die Intelligenz (KFT). Die aktuell besuchte Schule und das selbsteingeschätzte Vorwissen klären keinen signifikanten Varianzanteil auf. Alter, Geschlecht und Klassenstufe erklären jeweils zum ersten Messzeitpunkt mehr Varianz als zum zweiten. KFT und Schulnoten erklären beim zweiten Messzeitpunkt mehr Varianz als beim ersten. Die Klassenstufe erklärt am ersten Messzeitpunkt nur halb so viel Varianz wie die Intelligenz oder die Mathematiknote am zweiten Messzeitpunkt sogar noch weniger. Erwartungsgemäß weisen Fünftklässler an beiden Messzeitpunkten weniger konzeptuelles und prozedurales Wissen auf als Sechstklässler. Jungen besitzen im Durchschnitt mehr Wissen beider Arten als Mädchen.

Tabelle 30: Durch Personencharakteristika aufgeklärte Varianzanteile der latenten Faktoren.

Prädiktor	Konzeptuelle Wissensfaktoren			Prozedurale Wissensfaktoren		
	R^2	p	Sign.	R^2	p	Sign.
Mzp 1						
Alter	,020	,049	*	,015	,092	+
Geschlecht	,076	,000	***	,108	,000	***
Klassenstufe	,083	,000	***	,082	,000	***
Schule	,036	,918	n.s.	,048	,778	n.s.
KFT	,166	,000	***	,211	,000	***
Mathematiknote	,169	,000	***	,202	,000	***
Deutschnote	,050	,002	**	,042	,004	**
Selbsteingeschätztes Vorwissen	,015	,117	n.s.	,013	,137	n.s.
Mzp 2						
Alter	,005	,313	n.s.	,005	,325	n.s.
Geschlecht	,037	,006	**	,044	,003	**
Klassenstufe	,022	,035	*	,025	,024	*
Schule	,031	,958	n.s.	,032	,952	n.s.
KFT	,343	,000	***	,363	,000	n.s.
Mathematiknote	,261	,000	***	,276	,000	***
Deutschnote	,136	,000	***	,135	,000	***
Selbsteingeschätztes Vorwissen	,023	,052	+	,021	,059	+

9.4.6 Einfluss potenzieller Störvariablen

Von den potenziellen Störvariablen, die die Varianzen der Wissensmaße beeinflussen könnten, kann der Einfluss dreier Variablen hier quantitativ analysiert werden: der Einfluss des Versuchsleiters, des Arbeitsplatzes und der Position eines Maßes in der Erhebungsreihenfolge aller acht Wissensmaße.

Es gab fünf Versuchsleiter und es standen neun Computerarbeitsplätze zur Verfügung. Innerhalb der Erhebungsreihenfolge der Wissensmaße kann ein Maß, je nach randomisierter Erhebungsreihenfolge, eine Position von eins bis acht besitzen.

Pro Wissensmaß, Messzeitpunkt und potenzieller Störvariable wurden drei einfaktorielle Varianzanalysen mit jeweils $N = 204$ Fällen gerechnet. Tabelle 31 zeigt die Varianzaufklärungen η^2 , die Irrtumswahrscheinlichkeiten p und die Signifikanzen. Auf eine Korrektur zur Verhinderung der Alphafehlerakkumulation wurde verzichtet, da es sich lediglich um explorative Analysen handelt. Die Analysedateien befinden sich in Anhang D.15.

Der Versuchsleiter hatte keinen signifikanten Einfluss auf die Wissensmaße. Der Arbeitsplatz hatte einen signifikanten Einfluss auf lediglich ein einziges Maß, die Evaluation am zweiten Messzeitpunkt. Die Testreihenfolge beeinflusste nachweisbar nur die Maße Evaluation und Problemlösekorrektheit am ersten Messzeitpunkt und das Maß Zugriffsasymmetrie am zweiten Messzeitpunkt.

Tabelle 31: Varianzaufklärungen potenzieller Störvariablen mit ihren Signifikanzen.

Prädiktor	Versuchsleiter			Arbeitsplatz			Testreihenfolge		
	η^2	p	Sign.	η^2	p	Sign.	η^2	p	Sign.
Mzp 1									
Maße konzeptuellen Wissens									
Evaluation	,018	,469	n.s.	,050	,251	n.s.	,084	,015	*
Übersetzung	,024	,308	n.s.	,036	,512	n.s.	,064	,069	+
Größenvergleich	,030	,192	n.s.	,064	,106	n.s.	,014	,901	n.s.
Erklärung	,012	,656	n.s.	,018	,890	n.s.	,052	,158	n.s.
Faktorwerte (K1)	,024	,301	n.s.	,038	,463	n.s.	-	-	-
Maße prozeduralen Wissens									
Problemlösekorrektheit	,019	,439	n.s.	,036	,520	n.s.	,086	,013	*
Problemlösedauer	,012	,659	n.s.	,033	,569	n.s.	,031	,520	n.s.
Zugriffsasymmetrie	,021	,367	n.s.	,049	,275	n.s.	,046	,226	n.s.
Dual task-Kosten	,015	,551	n.s.	,042	,382	n.s.	,033	,472	n.s.
Faktorwerte (P1)	,028	,232	n.s.	,048	,282	n.s.	-	-	-
Mzp 2									
Maße konzeptuellen Wissens									
Evaluation	,036	,116	n.s.	,077	,045	*	,056	,121	n.s.
Übersetzung	,009	,757	n.s.	,041	,404	n.s.	,053	,143	n.s.
Größenvergleich	,031	,179	n.s.	,041	,402	n.s.	,037	,379	n.s.
Erklärung	,016	,530	n.s.	,033	,580	n.s.	,021	,758	n.s.
Faktorwerte (K2)	,008	,823	n.s.	,039	,437	n.s.	-	-	-
Maße prozeduralen Wissens									
Problemlösekorrektheit	,005	,895	n.s.	,034	,554	n.s.	,020	,777	n.s.
Problemlösedauer	,012	,677	n.s.	,032	,588	n.s.	,027	,598	n.s.
Zugriffsasymmetrie	,006	,873	n.s.	,048	,282	n.s.	,075	,030	*
Dual task-Kosten	,013	,640	n.s.	,052	,224	n.s.	,030	,524	n.s.
Faktorwerte (P2)	,011	,681	n.s.	,041	,402	n.s.	-	-	-

Anmerkungen. $N = 204$; Signifikanzbestimmung ohne Alphafehlerkorrektur.

Würde eine Bonferoni-Korrektur durchgeführt, betrüge das kritische Alphafehlerniveau für die 56 in Tabelle 31 zusammengestellten Signifikanztests $\alpha_{\text{crit}} = 0,05 / 56 = 0,00089$ (berechnet nach Formel 7.63 in Bortz, 1993, S. 249). Gemessen an diesem

Niveau hat keine der potenziellen Störvariablen einen signifikanten Einfluss auf die Wissensmaße.

9.4.7 Explorative Faktoranalysen

Für jeden Messzeitpunkt einzeln wurde eine explorative Faktoranalyse (principal components analysis mit Varimax-Rotation) durchgeführt, um zusätzliche Informationen über die Faktorstruktur der Wissensmaße zu erhalten.

Wie Tabelle 32 wiedergibt, haben zu jedem der drei Messzeitpunkte jeweils drei der explorativ gefundenen Faktoren einen Eigenwert größer eins. Tabelle 33 zeigt pro Messzeitpunkt die standardisierten Faktorladungen der acht Wissensmaße auf den drei explorativ gefundenen Faktoren. Dass zu den drei Messzeitpunkten jeweils drei Faktoren gefunden wurden, bedeutet nicht, dass diese Faktorentriplets über die Messzeitpunkte hinweg ähnliche Konstrukte in ähnlicher Weise abbilden. Auch wenn das Gesamtbild der Faktorladungen über die Messzeitpunkte hinweg bemerkenswert gleich bleibt, was sich mit obigem Befund weitgehender metrischer Messinvarianz deckt, ändern sich einzelne Faktorladungen doch. Beispielsweise lädt das Maß Erklärungen am zweiten Messzeitpunkt zu -,51 auf dem dritten Faktor. Am dritten Messzeitpunkt lädt es jedoch zu ,09 auf dem dritten Faktor. Die Faktoren sollten daher nicht über die Messzeitpunkte hinweg inhaltlich miteinander gleichgesetzt werden. Entsprechend wurden die Faktoren in Tabelle 33 alle unterschiedlich benannt. Der erste Index hinter dem F gibt jeweils den Messzeitpunkt, der zweite Index die Nummer des Faktors an.

Tabelle 32: Eigenwerte der drei Faktoren (eine Analyse pro Messzeitpunkt).

	Mzp 1	Mzp 2	Mzp 3
1. Faktor	2,81	3,11	3,48
2. Faktor	1,09	1,14	1,04
3. Faktor	1,01	1,02	1,02

Tabelle 33: Standardisierte Ladungen der Wissensmaße auf drei Faktoren (eine Analyse pro Messzeitpunkt).

	Mzp 1			Mzp 2			Mzp 3		
	F ₁₁	F ₁₂	F ₁₃	F ₂₁	F ₂₂	F ₂₃	F ₃₁	F ₃₂	F ₃₃
Maße konzeptuellen Wissens									
Evaluation	,75	,05	-,06	,74	-,08	-,30	,69	,29	-,09
Übersetzung	,70	,33	,12	,71	,35	,12	,82	,10	,16
Größenvergleiche	,73	,14	,07	,80	,11	,15	,85	,01	-,09
Erklärungen	,65	-,17	-,28	,46	,29	-,51	,64	,26	,09
Maße prozeduralen Wissens									
Problemlösekorrektheit	,67	,43	,11	,76	,29	,25	,82	,15	,25
Problemlösedauer	-,03	-,03	,89	,22	,09	,81	,07	,03	,96
Zugriffsasymmetrie	,04	,79	-,33	,01	,89	-,12	,09	,85	-,11
Dual task-Kosten	,17	,59	,26	,31	,60	,16	,28	,67	,26

Anmerkung. Ladungen > ,40 als Interpretationshilfe fettgedruckt.

Die Eigenwerte zeigen für alle drei Messzeitpunkte, dass eine Einfaktorlösung schon viel Varianz erklärt und ein zweiter und dritter Faktor nur relativ geringe Beiträge zur Varianzaufklärung liefern. Der Befund entspricht inhaltlich den oben berichteten hohen Faktorinterkorrelationen.

Die Maße konzeptuellen Wissens laden alle hoch auf dem ersten zu jedem Messzeitpunkt gefundenen Faktor und relativ niedrig auf den anderen Faktoren. Die Maße prozeduralen Wissens sind mit den drei Faktoren jeweils in unsystematischerer Weise verbunden. Dies stimmt damit überein, dass oben für die konzeptuellen Wissensfaktoren höhere interne Konsistenzen gefunden wurden als für die prozeduralen Wissensfaktoren.

Betrachtet man die Maße prozeduralen Wissens näher, so zeigt sich, dass das Maß Problemlösekorrektheit stets hoch auf dem Faktor lädt, auf dem auch die vier konzeptuellen Maße hoch laden. Darüber hinaus lädt es jedoch noch auf ein oder zwei anderen Faktoren, auf denen die prozeduralen Maße höher laden als die konzeptuellen. Das Maß Problemlösekorrektheit scheint also stärker als die anderen Maße beide Wissensarten zugleich zu messen und die divergenten Validitäten der Maße zu beeinträchtigen.

Die Maße Zugriffsasymmetrie und Dual task-Kosten laden an allen drei Messzeitpunkten relativ hoch auf dem jeweils zweiten gefundenen Faktor. Sie weisen also systematische Zusammenhänge auf, die unabhängig von den konzeptuellen Wissensmaßen sind.

Das Maß Problemlösedauer lädt als einziges systematisch hoch auf dem dritten Faktor. Dies passt zu dem Befund, dass seine Faktorladungen in den konfirmatorischen Analysen niedrig waren und ein unerwartetes Vorzeichen aufwiesen.

Da die Befunde Anlass zu der Vermutung geben, dass die niedrigen divergenten Validitäten vor allem, jedoch nicht nur durch das Maß Problemlösekorrektheit hervorgerufen werden, wurde aus Modell R123 (vgl. Abbildung 8 sowie Abschnitt 9.3.5) durch Entfernung des Maßes Problemlösekorrektheit Modell R123M (M für modifiziert) abgeleitet. Die Metrik der prozeduralen Wissensfaktoren wurde dabei über das Maß Zugriffsasymmetrie festgelegt. Falls die Parameterschätzung von Modell R123 an den niedrigen konvergenten Validitäten scheiterte und falls das Maß Problemlösekorrektheit für die niedrigen konvergenten Validitäten mitverantwortlich ist, so könnte die Schätzung von Modell R123M erfolgreicher verlaufen.

Die entsprechenden Analysedateien befinden sich in Anhang D.16. Die Resultate werden hier jedoch nicht berichtet, weil eine Warnmeldung in MPlus unspezifische Probleme bei der Parameterschätzung anzeigte. Das Problem könnte darin bestehen, dass

das Maß Problemlösekorrektheit der stärkste Indikator der prozeduralen Wissensfaktoren war. Entfernt man es aus den Analysen, könnten die Kovarianzen zwischen den verbleibenden drei Indikatoren zu niedrig sein, um die Extraktion eines Faktors zu erlauben, vor allem da das Maß Problemlösedauer niedrig mit den anderen beiden interkorreliert ist.

9.4.8 Vergleichsmodelle mit manifesten Variablen

Oben wurde die Annahme getroffen, dass die Modellierung von konzeptuellem und prozeduralem Wissen als latente Variablen wesentlich validere Resultate über ihre Interrelationen liefert als die Analyse manifester Einzelmaße.

Um die Richtigkeit dieser Annahme zu illustrieren, wurden drei cross-lagged panel designs über drei Messzeitpunkte gerechnet, wobei anstelle des konzeptuellen Faktors ein Einzelmaß konzeptuellen Wissens und anstelle des prozeduralen Faktors ein Einzelmaß prozeduralen Wissens benutzt wurde.

In der pädagogisch-psychologischen Forschung ist das Maß Evaluation von Prozeduren einer der am häufigsten verwendeten Indikatoren für konzeptuelles Wissen. Die Problemlösekorrektheit ist einer der am häufigsten verwendeten Indikatoren für prozedurales Wissen. Die Relationen zwischen diesen Maßen wurden daher in einem Modell EP123 (mit E für Evaluation und P für Problemlösekorrektheit) spezifiziert.

In der kognitiven Forschung werden verbale Erklärungen häufig benutzt um konzeptuelles oder deklaratives Wissen zu erheben. Die Problemlösedauer ist das am häufigsten benutzte Maß zur Operationalisierung prozeduralen Wissens. Da das Maß in der vorliegenden Untersuchung nicht optimal funktioniert zu haben scheint, wurde stattdessen das Maß Zugriffsasymmetrie verwendet, das ebenfalls auf der Analyse von Reaktionszeiten beruht. Die Relationen zwischen dem Maß Erklärungen und dem Maß Zugriffsasymmetrie beschreibt Modell EZ123 (mit E für Erklärungen und Z für Zugriffsasymmetrie).

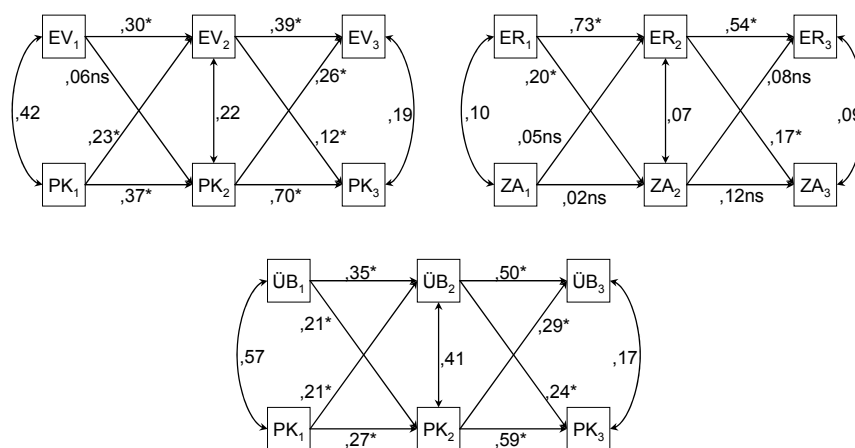
Die besten Einzelmaße für konzeptuelles und prozedurales Wissen in der vorliegenden Untersuchung sind – gemessen an den Faktorladungen – das Maß Übersetzung und das Maß Problemlösekorrektheit. Die Beziehungen zwischen ihnen beschreibt Modell ÜP123 (mit Ü für Übersetzung und P für Problemlösekorrektheit).

Den Fit der drei Modelle gibt Tabelle 34 wieder. Das Entscheidende an den Modellen ist jedoch nicht ihre Modellanpassung, sondern ihre in Abbildung 13 gezeigten Pfadkoeffizienten, die sich deutlich untereinander und ebenso deutlich von den Koeffizienten zwischen den latenten Faktoren in den Modellen R12, K123 und P123 unterscheiden. Die meisten Koeffizienten sind in den hier analysierten Modellen niedriger

als in den Modellen mit latenten Faktoren. Besonders auffällig ist das bei den Faktorinterkorrelationen, die zwischen den latenten Faktoren extrem hoch und zwischen den manifesten Variablen teilweise nicht einmal signifikant sind. Auch in Bezug auf die kausalen Relationen zwischen den Wissensarten implizieren alle hier dargestellten Modelle unterschiedliche Annahmen, da die Koeffizienten der Pfade von K1 zu P2 und von P1 zu K2 sehr unterschiedlich große Werte besitzen. Tendenziell sind die Pfadkoeffizienten von Modell EP123 am besten mit dem procedures-first view, die Koeffizienten von Modell EZ123 am besten mit dem concepts-first view und die Koeffizienten von Modell ÜP123 am besten mit dem Iterativen Modell vereinbar.

Tabelle 34: Fitindizes der Modelle mit manifesten Variablen.

Indikator konzeptuellen Wissens	Indikator prozeduralen Wissens	Modell	χ^2	df	p	CFI	WRMR	RMSEA
Evaluation	Problemlösekorrektheit	Modell EP123	19,848	4	,001	0,954	1,136	0,139
Erklärung	Zugriffsasymmetrie	Modell EZ123	9,830	4	,043	0,975	0,711	0,085
Übersetzung	Problemlösekorrektheit	Modell ÜP123	32,972	4	,000	0,944	1,136	0,188



Anmerkungen. Signifikanz der Korrelationen nicht mitdargestellt; EV: Evaluation; PK: Problemlösekorrektheit; ER: Erklärung; ZA: Zugriffsasymmetrie; ÜB: Übersetzung.

Abbildung 13: Cross-lagged panel-Modelle mit manifesten Variablen (links oben: Modell EP123, rechts oben: Modell EZ123, unten: Modell ÜP123).

9.5 Diskussion

9.5.1 Operationalisierbarkeit der Wissensarten

Forschungsfragen 1 und 2 beziehen sich auf die konvergente und die divergente Validität der verwendeten Wissensmaße. Sie werden zuerst diskutiert.

Konvergente Validitäten

Bezüglich der Frage nach den konvergenten Validitäten kann eine positive Antwort gegeben werden. Die vier Maße einer Wissensart pro Messzeitpunkt sind substantiell interkorreliert und lassen sich daher gut auf einen ihnen gemeinsamen latenten Wissensfaktor zurückführen.

Die konvergente Validität der Maße konzeptuellen Wissens ist an allen drei Messzeitpunkten höher als die konvergente Validität der Maße prozeduralen Wissens. Die konzeptuellen Wissensfaktoren besitzen höhere interne Konsistenzen, klären mehr Indikatorvarianz auf und ihre Indikatoren laden – im Gegensatz zu den Indikatoren der prozeduralen Wissensfaktoren – bei einer explorativen Faktoranalyse hoch auf einem gemeinsamen Faktor.

Die niedrige interne Konsistenz des prozeduralen Wissensfaktors steht im Einklang mit den in Abschnitt 4.2.3 berichteten Befunden, dass verschiedene Maße prozedurales Wissen niedrig interkorreliert sind.

Die niedrige interne Konsistenz des prozeduralen Wissensfaktors im Vergleich zum konzeptuellen kann nicht auf unterschiedliche Reliabilitäten der Faktorindikatoren zurückgeführt werden. Hier zeigt sich nämlich genau das umgekehrte Bild: Die internen Konsistenzen der Einzelmaße prozeduralen Wissens sind durchgängig höher als die internen Konsistenzen der Einzelmaße konzeptuellen Wissens. Dass die Maße prozeduralen Wissens zwar substantielle, nicht jedoch optimale konvergente Validitäten aufweisen, ist also anscheinend eher ein Problem der Validität als der Reliabilität. Die oben berichteten konfirmatorischen Faktoranalysen zeigen allerdings klar, dass die konvergente Validität der Maße prozeduralen Wissens gegeben ist.

Die plausibelste Erklärung für das Gesamtbild der Befunde ist, dass prozedurales Wissen einen Faktor zweiter Ordnung darstellt, der mehrere teilweise voneinander unabhängige Subdimensionen (Faktoren erster Ordnung) beeinflusst, die ihrerseits die manifesten Maße prozeduralen Wissens beeinflussen. Die Angemessenheit eines solchen hierarchischen Faktormodells prozeduralen Wissens kann jedoch nicht mit den in dieser Arbeit benutzten vier Indikatoren allein überprüft werden. Hierzu sind weitere Studien notwendig, die mehrere Indikatoren pro Faktor erster Ordnung erheben.

Besondere Aufmerksamkeit sollte dabei dem Maß Problemlösedauer geschenkt werden. Es ist als Maß der Prozeduralisierung oder Automatisierung von Wissen in Experimenten zum Fähigkeitserwerb ubiquitär. Die hier vorgelegten Analysen zeigen jedoch, dass eventuell seine generelle Konstruktvalidität als Maß prozeduralen Wissens, zumindest jedoch seine ökologische Validität als Maß prozeduralen Wissens in einer

Computerlernumgebung für Kinder äußerst gering ist. Angesichts der vielen kognitiven, motivationalen und situativen Faktoren, die die Arbeitsgeschwindigkeit von Kindern beeinflussen können, ist das ein plausibles Resultat. Ein spezielles Problem, das sich in der Untersuchung zeigte, war, dass Kinder mit wenig prozeduralem Wissen sich durch die Aufgaben einfach durchklickten, was entgegen der theoretischen Annahmen dazu führte, dass Kinder umso geringere Lösungsdauern aufwiesen, je weniger prozedurales Wissen sie besaßen. Dieser Effekt zeigt sich nicht nur in den Daten, sondern wird auch durch Beobachtungen der Versuchsleiter gestützt.

Die Maße Zugriffsasymmetrie und Dual task-Kosten stellen bessere Operationalisierungen prozeduralen Wissens dar. Ein Grund dafür könnte darin bestehen, dass es sich bei ihnen um Differenzwerte handelt, bei denen die Ausprägung eines Maßes relativ zu einer Baseline-Bedingung angegeben wird. Allgemeine Faktoren, beispielsweise Motivation und Ablenkung, beeinflussen die Maße beider Bedingungen. Unterschiede zwischen Baseline- und Manipulationsbedingung können daher nichts mit solch allgemein wirkenden Faktoren zu tun haben.

Gelegentlich wurde vermutet, dass Differenzwerte – unbeschadet ihrer fraglos guten Konstruktvaliditäten – grundsätzlich niedrige Reliabilitäten aufweisen könnten (Paul, Churchill, & Brown, 1993; Rost, 2004, S. 274-276). Neuere Befunde weisen jedoch darauf hin, dass das Konzept der Reliabilität von Differenzwerten inhaltlich schlecht zu greifen und daher auch schlecht durch eine Formel anzugeben ist. Die früher benutzte Formel zur Reliabilitätsbestimmung von Differenzwerten beruht auf unplausiblen Annahmen (unkorrelierten Prätest- und Posttestwerten) und führt dadurch zu starken Unterschätzungen der tatsächlichen Reliabilitäten. Das Problem besteht also weniger auf der Ebene der tatsächlichen Reliabilität als vielmehr auf der Ebene ihrer Berechnung (Chiou & Spreng, 1996). Auch die Faktorladungen in der vorliegenden Untersuchung geben keinen Anlass, die Reliabilität oder Validität der beiden auf Differenzwerten beruhenden Indikatoren in Frage zu stellen.

Gemessen an den Faktorladungen in den konfirmatorischen Analysen ist das beste manifeste Maß für konzeptuelles Wissen die Übersetzung von Dezimalbrüchen in Tortendiagramme. Das beste manifeste Maß für prozedurales Wissen ist die Problemlösekorrektheit bei Routineproblemen.

Trotz der oben erläuterten Einschränkungen sind die Befunde zur konvergenten Validität der Maße insgesamt ausgesprochen positiv. Extreme Anhänger des situated cognition und des situated action views bestreiten, dass die Annahme situations- und

aufgabenübergreifender und in diesem Sinne abstrakter Wissensrepräsentationen sinnvoll ist (vgl. Abschnitt 4.2.1). Die hier vorgelegten Befunde zeigen jedoch, dass das Gegenteil der Fall ist. Obwohl inhaltliche Aufgabenanforderungen (Erklären, Vergleichen, Transformieren, Bewerten usw.), grafische Oberflächen (Zahlenstrahle, Tortendiagramme, Schrift, Zahlen) und Antwortformate (multiple choice, halb offen und offen) der Implementierungen aller acht Maße sich substantiell voneinander unterscheiden, klären die latenten Wissensfaktoren statistisch signifikante und praktisch bedeutsame Varianzanteile der acht Maße auf. Dies ist ein wichtiger Befund, der ohne die Annahme mindestens einer Form von abstrakter Wissensrepräsentation nicht erklärt werden kann. Er wird zusätzlich dadurch gestützt, dass die Kinder in der Intervention lediglich Routineaufgaben (ergänzt durch aufgabenbezogene Selbsterklärungen) übten, sich jedoch auch auf allen vier Arten von Transferaufgaben signifikant verbesserten. Aufgrund des Fehlens einer Kontrollgruppe ist dieser Befund zwar weniger valide, da es sich rein theoretisch auch um interventionsunabhängige Retesteffekte handeln könnte. Jedoch wurde gefunden, dass die Zuwächse der Lösungsraten bei den Routineaufgaben erheblich größer sind als bei den Transferaufgaben. Dies kann leicht durch kausale Effekte der Intervention, jedoch nur schwer durch reine Retesteffekte erklärt werden.

Divergente Validitäten

Dass eine optimale Messung der Wissensarten trotz guter konvergenter Validitäten nicht möglich ist, liegt an den suboptimalen divergenten Validitäten der Wissensmaße. Diese zeigen sich zum einen darin, dass für alle drei Messzeitpunkte das jeweilige Zweifaktormodell trotz weniger sparsamer Modellannahmen nicht signifikant besser an die Daten angepasst ist als das jeweilige Einfaktormodell. Zum anderen zeigen sich die niedrigen divergenten Validitäten in den ausgesprochen hohen Faktorinterkorrelationen, die je nach Messzeitpunkt und Berechnungsart zwischen ,83 und ,99 liegen. Diese Korrelationen sprechen zwar für die Sauberkeit von Design und Durchführung, weil sie nur bei extrem hoher Reliabilität der Faktorwerte auftreten können. Sie zeigen jedoch auch, dass es nur bedingt gelungen ist, zwei teilweise voneinander unabhängige Wissensarten zu operationalisieren.

Die vier Maße konzeptuellen Wissens beruhen auf Transferaufgaben. Keine der Aufgaben war Teil der Intervention. Keine erfordert die Verortung eines Dezimalbruchs auf einem Zahlenstrahl. Alle vier basieren auf durchschnittlichen Lösungsraten. Die Maße prozeduralen Wissens wurden alle mittels Routineaufgaben erhoben, die Teil der Intervention waren und die Verortung von Dezimalbrüchen auf dem Zahlenstrahl

erfordern. Die prozeduralen Wissensmaße stellen einen Mix aus absoluten Werten und prozentualen Veränderungsraten, aus Lösungsraten und Lösungszeiten dar. Offensichtlich haben all diese Unterschiede zwischen konzeptuellen und prozeduralen Maßen ihre Kovarianzen nur minimal reduziert.

Der Zusammenhang zwischen dem konzeptuellen und dem prozeduralen Wissensfaktor steigt deskriptiv gesehen vom ersten zum zweiten Messzeitpunkt an. Am ersten Messzeitpunkt beträgt die Faktorinterkorrelation ,85. Am zweiten und dritten Messzeitpunkt beträgt die Interkorrelation ,94. Dies gibt Anlass zu der Vermutung, dass konzeptuelles und prozedurales Wissen am Anfang des Wissenserwerbs etwas unverbundener nebeneinander stehen könnten als in späteren Phasen, beispielsweise weil sie durch unterschiedliche Transfermechanismen aus bekannten Inhaltsgebieten in das neue übertragen werden, aber umso stärker integriert werden, je mehr Lernerfahrungen man in dem neuen Inhaltsgebiet sammelt.

Falls wirklich bidirektionale Kausalbeziehungen zwischen den Wissensarten herrschen, wie es das Iterative Modell postuliert, wäre das eine plausible Annahme. Man kann nur hoffen, dass sie zutrifft, denn nach Baroody (2003) stellt die Integration von konzeptuellem und prozeduralem Wissen die Grundlage einer umfassenden adaptiven Expertise dar (s. Abschnitt 5.5.2). Die hier analysierten Daten lassen vermuten, dass die Integration von konzeptuellem und prozeduralem Wissen ein leicht erreichbarer, geradezu natürlich auftretender Zustand ist. Allerdings ist nicht auszuschließen, dass dieser Befund auf das Lernen mit dem „Fang-das-Monster“-Spiel beschränkt ist. Das Spiel wurde so gestaltet, dass es sowohl zu Zuwächsen an konzeptuellem als auch an prozeduralem Wissen führt. Werden Kindern die Wissensarten unabhängig voneinander vermittelt, könnten vielleicht stärkere Dissoziationen auftreten.

Außerdem stammen die Befunde lediglich aus einem korrelativen Design. Dass konzeptuelles und prozedurales Wissen so hoch interkorreliert sind, belegt nicht zwangsläufig, dass sie innerpsychisch gut integriert, also in irgendeiner Weise miteinander vernetzt, gespeichert wurden. Es kann nicht ausgeschlossen werden, dass sie bei den untersuchten Kindern innerpsychisch keine direkte Verbindung aufweisen und lediglich gemeinsam durch eine dritte Variable beeinflusst wurden. Bei dieser Variablen kann es sich jedoch nicht um das Alter oder die Intelligenz handeln, weil die Faktorinterkorrelationen bei Auspartialisierung dieser beiden Variablen nicht absinken.

Eine weitere mögliche Erklärung für die die niedrigen divergenten Validitäten ist, dass sie lediglich auf eine suboptimale Auswahl der Maße zurückzuführen sind. In explorativen

Faktoranalysen wurde gefunden, dass nur zwei Maße prozeduralen Wissens (Zugriffsasymmetrie und Dual task-Kosten) hoch auf einem gemeinsamen Faktor laden. Die Problemlösekorrektheit lädt auf einem Faktor hoch, auf dem kein weiteres Maß prozeduralen Wissens, jedoch alle vier Maße konzeptuellen Wissens hoch laden (s. Tabelle 33 auf Seite 181). Wie die Faktorladungen aus den konfirmatorischen Analysen zeigen (s. Tabelle 18 auf Seite 160), trägt jedoch ausgerechnet das Maß Problemlösekorrektheit am meisten zur Bestimmung der prozeduralen Wissensfaktoren bei. Es kann daher angenommen werden, dass die Faktorinterkorrelationen sanken, wenn das Maß Problemlösekorrektheit nicht als Maß prozeduralen Wissens verwendet würde. Eine explorative Überprüfung dieser Vermutung scheiterte an Konvergenzproblemen, vermutlich, weil das Fehlen des Indikators mit der höchsten Ladung auf dem prozeduralen Faktor (in Verbindung mit den Problemen, die für das Maß Problemlösedauer gefunden wurden) dazu führte, dass die Werte der prozeduralen Wissensfaktoren nicht mehr bestimmt werden konnten.

Messbarkeit der Wissensarten: Fazit

Insgesamt können die Ergebnisse auf einer inhaltlichen und einer methodischen Ebene interpretiert werden. Auf der inhaltlichen Ebene gibt es drei Möglichkeiten: (1) Konzeptuelles und prozedurales Wissen können prinzipiell nicht unabhängig voneinander variieren, stellen also zusammen ein einziges homogenes oder hierarchisch strukturiertes Konstrukt dar. (2) Konzeptuelles und prozedurales Wissen können prinzipiell unabhängig voneinander variieren. Sie taten das hier jedoch aufgrund der besonderen Lernsituation nicht. (3) Konzeptuelles und prozedurales Wissen können prinzipiell unabhängig voneinander variieren und taten das in der Untersuchungssituation. Auf der methodischen Ebene gibt es die Möglichkeiten: (1) Die Maße konzeptuellen und prozeduralen Wissens wurden optimal ausgewählt, so dass sie die Wissensarten valide erfassen. (2) Die Maße wurden suboptimal ausgewählt und erfassen die Wissensarten daher nicht valide.

Diese Möglichkeiten sind in Tabelle 35 zusammengestellt. Die Annahme, dass die Maße hinreichend valide waren und die Wissensarten in der Untersuchungssituation unabhängig voneinander variierten, ist logisch nicht mit den hohen Faktorinterkorrelationen vereinbar und kann ausgeschlossen werden. Ob die Probleme bei der Messung prozeduralen Wissens durch die Problemlösekorrektheit und Problemlösedauer so gravierend waren, dass auch die anderen beiden Möglichkeiten in der linken Spalte der Tabelle ausgeschlossen werden können, ist unklar.

Tabelle 35: Erklärungsmöglichkeiten für die hohen Faktorinterkorrelationen.

Wissensarten	Wissensmaße	
	hinreichend valide	zu unvalide
miteinander identisch	✓	✓
prinzipiell unabhängig; in Untersuchungssituation hoch interkorreliert	✓	✓
prinzipiell und in Untersuchungssituation unabhängig voneinander variierend	✗	✓

Anmerkungen. ✓ Erklärung möglich, ✗ Erklärung logisch nicht möglich.

Trotz der niedrigen divergenten Validitäten der Maße wurden nach der Analyse des Messmodells Kausalanalysen mit getrennten Faktoren für konzeptuelles und prozedurales Wissen durchgeführt. Dies lässt sich durch die große theoretische Wichtigkeit der Unterscheidung der Wissensarten, durch die gleich guten Fitindizes der Ein- und der Zweifaktormodelle bei Betrachtung einzelner Messzeitpunkte, durch den signifikant schlechteren Fit der Einfaktormodelle im Vergleich zu den Zweifaktormodellen bei gleichzeitiger Betrachtung mehrerer Messzeitpunkte, durch die deskriptiv unter eins liegenden Faktorinterkorrelationen rechtfertigen und durch den Befund, dass die Maße konzeptuellen und prozeduralen Wissens in explorativen Faktoranalysen überwiegend auf unterschiedlichen Faktoren hoch laden.

9.5.2 Kausale Interrelationen

Die Parameter eines cross-lagged panel designs über alle drei Messzeitpunkte ließen sich aufgrund von Konvergenzproblemen nicht schätzen. Das ist nicht überraschend, wenn man bedenkt, dass die Wissensarten einerseits, vor allem am zweiten und am dritten Messzeitpunkt, hoch interkorreliert sind, andererseits in dem entsprechenden Modell jedoch beide als Prädiktoren der späteren Entwicklung der Wissensarten benutzt wurden. Es wurde versucht zwei zu ,94 korrelierte Variablen zu benutzen, um zwei andere zu ,94 korrelierte Variablen vorherzusagen.

Verzichtet man auf diesen Teil des Modells und untersucht entweder die Entwicklung einzelner Wissensarten über die drei Messzeitpunkte oder die Entwicklung beider Wissensarten über die ersten beiden Messzeitpunkte, ergeben sich keine Konvergenzprobleme.

Jedoch verbleiben Interpretationsprobleme, die dadurch entstehen, dass weder der Einfluss des konzeptuellen Wissens auf späteres prozedurales Wissen noch der Einfluss prozeduralen Wissens auf späteres konzeptuelles Wissen statistisch signifikant ist.

Streng genommen belegen die Daten daher weder einen kausalen Einfluss von konzeptuellem Wissen auf prozedurales noch umgekehrt, sondern sind am besten mit dem inactivation view vereinbar. Wie man Tabelle 23 auf Seite 170 entnehmen kann, liegt das

nicht an den absoluten Größen der Pfadkoeffizienten, die standardisiert ,34 und ,36 betragen, sondern daran, dass die Standardfehler groß sind im Vergleich zu den unstandardisierten Pfadkoeffizienten. Hier liegt also ein Teststärkeproblem vor.

Die Frage ist jedoch, ob die mangelnde Teststärke wirklich etwas über die Wissensarten aussagt oder eher mehr über das hier verwendete Forschungsdesign. Zwei Probleme erschwerten die Schätzung der Parameter. Zum einen wurde versucht, zwei zu ,85 korrelierte Variablen zu benutzen, um zwei zu ,94 korrelierte Variablen vorherzusagen. Zum anderen beträgt zwar der für SEM minimal erforderliche Stichprobenumfang 200 Personen. Optimal sind jedoch Stichprobengrößen ab $N = 500$. Beides erschwert eine akkurate Parameterschätzung und kann zu den großen Standardfehlern beigetragen haben, sagt jedoch natürlich nichts über die Wissensarten selbst aus.

Methodiker wie Gigerenzer (1998) empfehlen aufgrund der bekannten Nachteile von Signifikanztests, Hypothesen nicht anhand der Signifikanz, sondern anhand der Effektstärke zu überprüfen. Das ist auch im vorliegenden Fall angebracht.

Die Effektstärken der Koeffizienten liegen im selben Wertebereich wie die von Rittle-Johnson et al. (2001) gefundenen. Auffällig ist, dass die Werte der beiden Koeffizienten mit ,34 und ,36 fast exakt gleich groß sind. Geht man davon aus, dass die Hauptfrage der Untersuchung nicht ist, ob es überhaupt kausale Beziehungen zwischen den Wissensarten gibt, sondern ob die Beziehungen uni- oder bidirektional sind, so kann eine klare Antwort gegeben werden: Die Effektstärken der Koeffizienten sind weder mit dem concepts-first noch mit dem procedures-first view zu vereinbaren, sondern belegen hypothesenkonform die Richtigkeit des Iterativen Modells, das bidirektionale Kausalbeziehungen postuliert.

Dieser Schluss ist auch theoretisch plausibel. Sowohl der concepts-first view als auch der procedures-first view sind schon lange theoretisch etabliert. Wäre einer offensichtlich falsch und durch empirische Daten leicht zu widerlegen, wäre das sicherlich schon geschehen. Die Uneindeutigkeit der hier vorgelegten Befunde erklärt, warum beide Sichtweisen so lange nebeneinander bestehen konnten und noch immer bestehen. Sie sind beide ausgesprochen schwierig zu widerlegen.

Im Gegensatz zu früheren Arbeiten, die beträchtliche methodische Defizite aufweisen (Messung nur einer Wissensart, Verwendung dichotomer Items, Verzicht auf Längsschnittdaten und so weiter; vgl. die Abschnitte 2.3.6 und 2.4.3 sowie Kapitel 8), zeigt die vorliegende Studie mit Hilfe von Strukturgleichungsmodellen über acht Wissensmaße und drei Messzeitpunkte klar auf, wo die Probleme bei der Bestimmung der Relationen zwischen konzeptuellem und prozeduralem Wissen liegen und wie die ambivalenten Befunde zustande kommen: Es wurde belegt, dass die internen Konsistenzen

der einzelnen Wissensmaße kein Problem darstellen. Auch Anzeichen für Motivations- oder Überforderungsprobleme wurden nicht gefunden. Die Maße wurden nicht praktisch bedeutsam durch die Störvariablen Versuchsleiter, Computerarbeitsplatz und Testreihenfolge verzerrt. Die Maße konzeptuellen und prozeduralen Wissens weisen ausreichend große konvergente Validitäten auf, obwohl prozedurales Wissen anscheinend eher ein hierarchisches als ein homogenes Konstrukt darstellt.

Die Schwierigkeiten bei der Hypothesentestung entstehen vor allem dadurch, dass die divergenten Validitäten der Maße problematisch hoch sind, was viele der hypothesentestenden Analysen behindert.

Für die generelle Validität der hier vorgestellten Ergebnisse spricht auch, dass eine Stichprobe verwendet wurde, die mit $N = 204$ größer ist als in allen zuvor publizierten Laborstudien zu konzeptuellem und prozeduralem Wissen.

Die Ambivalenz der aktuellen Befunde bezüglich der Kausalrelationen geht also nicht einfach auf abstellbare methodische Defizite der Studie zurück, sondern zeigt substanzielle Probleme und Problemlösechancen des Forschungsfeldes auf.

Besonders offensichtlich werden die Defizite der bisherigen Forschung bei der Analyse der drei cross-lagged panel-Modelle, die statt der latenten Faktoren manifeste Wissensmaße verwendeten (s. Abbildung 13 auf S. 184). Je nach den verwendeten Maßen bestätigen die Modelle entweder den concepts-first view, den procedures-first view oder das Iterative Modell.

Das verdeutlicht zweierlei: Erstens können die großen inhaltlichen Unterschiede zwischen den drei Modellen bisher nicht post hoc begründet und noch viel weniger a priori vorhergesagt werden. Hier sind weitere theoretische und empirische Analysen notwendig, die maßspezifische Verzerrungseffekte aufklären. Zweitens sind Einzelmaße offensichtlich grundsätzlich ungenügend, um allgemeingültige Befunde über konzeptuelles und prozedurales Wissen herzuleiten. Wie in Unterkapitel 4.5 schon auf der theoretischen Ebene erläutert wurde, weisen Einzelmaße zwar Zusammenhänge zu den Wissensarten, stets jedoch auch spezifische Fehlervarianzanteile auf, die zwangsläufig zu systematischen Verzerrungen der Messung der Wissensarten führen müssen. Bei der Verwendung latenter Faktoren werden diese maßspezifischen Varianzanteile ausfaktoriert und nur die gemeinsame Varianz aller Maße bildet die Grundlage darauf aufbauender Analysen. Dieser Weg wurde in der vorliegenden Arbeit zum ersten Mal beschritten. Die hier dargestellten Ergebnisse können daher als die bisher validesten zu den Relationen zwischen konzeptuellem und prozeduralem Wissen betrachtet werden.

Für die weitere Forschung ist interessant, dass eine weitgehende metrische Messinvarianz der verwendeten Wissensmaße nachgewiesen wurde. Sie können problemlos in Designs mit mehreren Messzeitpunkten eingesetzt werden, sofern die Zeitspannen zwischen den Messzeitpunkten eher im Bereich von Tagen als im Bereich von Monaten liegen.

9.5.3 Weitere Einflussgrößen

Konzeptuelles und prozedurales Wissen stellen keine isolierten Konstrukte dar. Sie weisen Zusammenhänge mit Geschlecht, Alter, Klassenstufe, der Intelligenz, der Mathematik- und der Deutschnote auf, wobei der Einfluss der ersten drei Faktoren während des Trainings abnahm, während der Einfluss der letzten drei Faktoren zunahm. Dies kann so interpretiert werden, dass die ersten drei Faktoren lediglich vor dem Training, die letzten drei jedoch auch während des Trainings wirksam waren.

Wie schon die Personeneigenschaften so weist auch die Intelligenz keine substanziell unterschiedlichen Relationen zu konzeptuellem und prozeduralem Wissen auf. Auch hier zeigt sich wieder das Problem der niedrigen divergenten Validitäten der Maße. Weil konzeptuelles und prozedurales Wissen so hoch interkorreliert sind, werden sie entweder beide oder beide nicht durch eine Kovariate beeinflusst. Da auch die Intelligenz mit den beiden Wissensarten korreliert ist, zeigten sich bei gleichzeitiger Verwendung aller drei Maße suppressionsähnliche Verzerrungseffekte.

Wie Einzelanalysen für die Wissensarten zeigten, trug die Intelligenz während der Intervention gemeinsam mit dem Vorwissen zu den Wissenszuwächsen bei. Die Einflüsse von Intelligenz und Vorwissen waren ungefähr gleich stark. Die Zuwächse an konzeptuellem Wissen wurden etwas besser durch das Vorwissen, die Zuwächse an prozeduralem Wissen etwas besser durch die Intelligenz vorausgesagt.

Während der viermonatigen Pause zwischen dem zweiten und dem dritten Messzeitpunkt hatte die Intelligenz hingegen keinen Einfluss auf die Wissenszuwächse zusätzlich zum Vorwissen. Zusammen mit dem Befund, dass die Wissenszuwächse klein und unsystematisch über die Wissensmaße verteilt waren, ja teilweise sogar Wissensverluste deutlich wurden, deutet das darauf hin, dass während der viermonatigen Phase Lern- oder Elaborationsprozesse kaum stattfanden. Die Intelligenz hatte daher in dieser Phase nicht die Möglichkeit wirksam zu werden. Dies verdeutlicht, dass die Wichtigkeit stabiler Wissensrepräsentationen für eine langfristige Expertiseentwicklung und Kompetenzakkumulierung nicht unterschätzt werden sollte.

Der distance effect wurde entsprechend dem triple code model von Dehaene als Indiz für die Repräsentation von Zahlen auf einem mentalen Zahlenstrahl interpretiert. Die

Vermutung von Rittle-Johnson et al. (2001), dass die Intervention auch deswegen so erfolgreich wirkt, weil das Arbeiten mit dem externen Zahlenstrahl die Aktivierung und Anwendung des internen Zahlenstrahls fördern würde, konnte auf Ebene interindividueller Unterschiede nicht nachgewiesen werden. Der distance effect hatte keinen Einfluss auf die Wissenszuwächse.

9.6 Zusammenfassung

Das konzeptuelle und das prozedurale Wissen von 204 Fünft- und Sechstklässlern über Dezimalbrüche wurden an drei Messzeitpunkten durch je vier Maße erhoben und jeweils als latenter Faktor modelliert. Die konvergenten Validitäten der Maße waren gut. Das Maß Problemlösedauer wies jedoch Faktorladungen mit unerwartetem Vorzeichen auf. Die latenten Faktoren konzeptuellen und prozeduralen Wissens waren zu allen drei Messzeitpunkten sehr hoch interkorreliert. Eine mögliche Ursache ist, dass die Problemlösekorrektheit als Maß prozeduralen Wissens verwendet wurde, jedoch hoch mit den Maßen konzeptuellen Wissens korreliert ist. Die prädiktiven Relationen zwischen dem ersten und dem zweiten Messzeitpunkt deuten auf die Korrektheit des Iterativen Modells hin, belegen sie aber aufgrund mangelnder Teststärke nicht eindeutig. Die Wissenszuwächse während der Intervention wurden durch Vorwissen und Intelligenz gleichermaßen beeinflusst, die Wissenszuwächse während der nachfolgenden viermonatigen Phase nur durch das Vorwissen.

10 Studie 2: Erwerb der Wissensarten

10.1 Fragestellungen und Hypothesen

Studie 1 lieferte wichtige Indizien für die Beantwortung der Forschungsfragen 3 und 4, die sich auf die kausalen Relationen zwischen den Wissensarten beziehen. Endgültige Antworten konnten jedoch noch nicht gegeben werden – zum einen, weil Studie 1 ein korrelatives statt eines experimentellen Designs zugrunde liegt, zum anderen, weil die divergente Validität der verwendeten Maße niedrig ist.

Trotz dieser methodischen Einschränkungen führte Studie 1 zu wichtigen neuen Befunden, weil im Gegensatz zu früheren Studien konzeptuelles und prozedurales Wissen als latente Faktoren modelliert wurden. Es wurde gezeigt, dass dies zu hochwertigeren und zu qualitativ anderen Resultaten führt als die Verwendung einzelner Wissensmaße.

Um in Studie 2 die methodischen Einschränkungen von Studie 1 zu kompensieren, ohne ganz auf die methodischen Vorzüge von Studie 1 verzichten zu müssen, soll folgendermaßen vorgegangen werden: Die Teilnehmer von Studie 2 werden auf drei Gruppen randomisiert. Eine Gruppe erhält in einer ersten Lernphase vorwiegend konzeptuelles Wissen, eine Gruppe vorwiegend prozedurales Wissen und eine Gruppe keine der beiden Wissensarten. Um den Erfolg dieses Vorgehens nachzuweisen, bearbeiten die Teilnehmer vor und nach der Treatmenterteilung die Wissenstests von Studie 1. In einer zweiten Lernphase spielen alle Teilnehmer das „Fang-das-Monster“-Spiel und bearbeiten anschließend ein drittes Mal den Wissenstest.

Wenn die Gruppe, die in der ersten Lernphase vorwiegend konzeptuelles Wissen erhielt, am Ende der zweiten Lernphase mehr prozedurales Wissen besitzt als die Kontrollgruppe, würde das eindeutig einen Kausaleinfluss konzeptuellen Wissens auf prozedurales belegen. Denn die Treatment- und die Kontrollgruppe unterscheiden sich lediglich in einem einzigen Faktor: Die Treatmentgruppe erhielt in der ersten Phase konzeptuelles Wissen, die Kontrollgruppe nicht. Unterscheiden sich die Gruppen nach dem „Fang-das-Monster“-Spiel signifikant in ihrem prozeduralen Wissen, so können nur die zuvor erzeugten Gruppenunterschiede im konzeptuellen Wissen dafür ursächlich sein.

Gleiches gilt für den umgekehrten Erwerbsweg: Wenn die Gruppe, die in der ersten Lernphase vorwiegend prozedurales Wissen erhielt, am Ende der zweiten Lernphase mehr konzeptuelles Wissen besitzt als die Kontrollgruppe, so kann das nur durch einen positiven Kausaleinfluss prozeduralen Wissens auf konzeptuelles erklärt werden.

Für die Untersuchung der Forschungsfragen 3 und 4 fungieren die ersten beiden Messzeitpunkte also lediglich als manipulation check und als Kontrolle des Vorwissens.

Dennoch kann in dieser Phase auch eine wichtige inhaltliche Frage untersucht werden, die Forschungsfrage 5 dieser Arbeit darstellt, nämlich: Können konzeptuelles und prozedurales Wissen überhaupt unabhängig voneinander vermittelt werden? Rittle-Johnson und Alibali (1999) ist es in ihrer Äquivalenzproblemstudie nicht gelungen, einer Gruppe überwiegend konzeptuelles und einer Gruppe überwiegend prozedurales Wissen zu vermitteln (s. Unterkapitel 3.2). Auch in der Dezimalbruchstudie von Rittle-Johnson et al. (2001) gelang das nicht (s. Unterkapitel 3.3). In der Trainingsstudie von Byrnes und Wasik (1991) wirkten sich drei unterschiedliche Instruktionsformen entgegen der Hypothesen der Autoren nicht unterschiedlich auf konzeptuelles und prozedurales Wissen aus (s. Abschnitt 2.3.5). In einer dreijährigen längsschnittlichen Untersuchung einer eher traditionell und einer eher konstruktivistisch unterrichteten Schülergruppe fanden Hiebert und Wearne (1996) erst am letzten Messzeitpunkt Unterschiede zwischen dem konzeptuellen und prozeduralen Wissen der beiden Gruppen. Diese bestanden jedoch nicht in Mittelwertsdifferenzen, sondern darin, dass die Wissensarten in der konstruktivistisch unterrichteten Gruppe höher interkorreliert waren als in der traditionell unterrichteten Gruppe (s. Abschnitt 2.3.4).

Lediglich Resnick und Omanson (1987) berichten, dass eine Intervention, die zu einem signifikanten Anstieg konzeptuellen Wissens führte, nicht auch zu einem Anstieg prozeduralen Wissens führte (s. Abschnitt 2.3.5). Allerdings umfasste ihr Design keine Kontrollgruppe und ihre Stichprobe bestand lediglich aus neun Kindern mit Lernproblemen.

Gemeinsam mit dem Befund der hohen Faktorinterkorrelationen in Studie 1 werfen diese Ergebnisse die Frage auf, ob konzeptuelles und prozedurales Wissen überhaupt unabhängig voneinander verarbeitet und erinnert werden können. Da die bisherigen Studien jedoch nur Einzelmaße für die Wissensarten verwendeten, deren Nachteile oben belegt wurden, ist es sinnvoll die Frage nach der separaten Vermittelbarkeit der Wissensarten noch einmal unter Anwendung des in Studie 1 verwendeten acht Wissensmaße zu untersuchen. Im Vergleich zu älteren Studien erhöhen sich dadurch die Reliabilitäten und Validitäten der Messungen und damit auch die Teststärke des Designs.

Das Mehrgruppendedesign von Studie 2 ist nicht gut zur Durchführung erneuter konfirmatorischer Faktoranalysen geeignet, vor allem dann nicht, wenn aus Gründen der Ökonomie die Stichprobe nicht mehrere Hundert Teilnehmer umfassen soll, was hier der Fall ist. Es stellt sich also die Frage, wie die in Studie 1 gefundenen Faktorstrukturen

benutzt werden können, um manifeste Maße für konzeptuelles und prozedurales Wissen in Studie 2 zu entwickeln. Diese Frage wird in Abschnitt 10.2.3 weiter verfolgt.

Entsprechend dem Iterativen Modell wird angenommen, dass konzeptuelles und prozedurales Wissen teilweise unabhängig voneinander vermittelt werden können und dass bidirektionale kausale Relationen zwischen den Wissensarten nachweisbar sind.

10.2 Methode

10.2.1 Stichprobe, Aufbau und Durchführung

Die in Studie 1 gewählte Altersgruppe der Fünft- und Sechstklässler erwies sich als gut für die Studiendurchführung geeignet, jedoch wurden am dritten Messzeitpunkt leichte Deckeneffekte beobachtet. Daher wurde Studie 2 ausschließlich mit Fünftklässlern durchgeführt.

Diese wurden in gleicher Weise und aus denselben Stadtteilen rekrutiert wie die Teilnehmer von Studie 1, wobei sechs zusätzliche Schulen besucht wurden. Für die Teilnahme an allen drei Messzeitpunkten wurde den Eltern pro Kind eine Aufwandsentschädigung von 20 Euro gezahlt. Das ausgeteilte Informationsmaterial für Schulleiter, Eltern und Schüler befindet sich in Anhang E.

Die Erhebung erfolgte vom 16.01. bis zum 30.04.2005. Aufbau und Durchführung entsprachen im Wesentlichen Aufbau und Durchführung der ersten drei Sitzungen von Studie 1. An dieser Stelle wird lediglich auf die Unterschiede eingegangen. Diese bestehen darin, dass Sitzung 1 fünf oder sechs Tage vor Sitzung 2 stattfand und Sitzung 3 genau einen Tag nach Sitzung 2. Diese Aufteilung wurde gewählt, da gewöhnlich nur wenige Kinder drei Tage direkt hintereinander Zeit zur Studienteilnahme haben.

Wie Tabelle 36 zeigt, wurden während Sitzung 1 der Wissenstest und eine Erhebung des distance effects durchgeführt. Während Sitzung 2 wurden die Treatments erteilt, der Wissenstest durchgeführt und erneut der distance effect erhoben. Da die Kinder individuell an Computern arbeiteten und auch das Treatment über den Computer erhielten, konnten Kinder verschiedener Treatmentgruppen gleichzeitig an der Datenerhebung teilnehmen. In Sitzung 3 waren 80 Aufgaben des „Fang-das-Monster“-Spiels zu lösen, bevor der Wissenstest und ein Fragebogen zu Personeneigenschaften (s. Anhang F.2) folgten. Letzterer stellt eine leicht modifizierte Version des Personenfragebogens von Studie 1 dar.

Der Wissenstest, die Treatments und das „Fang-das-Monster“-Spiel wurden in einem Computerprogramm zusammengefasst, das sie bei vorheriger Eingabe von Treatmentgruppe und Sitzungsnummer automatisch in der richtigen Reihenfolge präsentierte.

Tabelle 36: Chronologische Reihenfolge der Datenerhebungsphasen.

Inhalte der Sitzungen	Aufgabenanzahl	Geschätzte Dauer in Minuten
Sitzung 1		
Wissenstest	222	55
Distance effect	210	15
Sitzung 2		
Treatment (TK / TP / TX)	6 / 80 / 7	20
Wissenstest	222	50
Distance effect	210	15
Sitzung 3		
„Fang-das-Monster“-Spiel	12	20
Wissenstest	222	45
Personenfragebogen	20	5

10.2.2 Maße und Interventionen

Der an den drei Messzeitpunkten eingesetzte Wissenstest ist identisch mit dem in Studie 1 eingesetzten (vgl. Unterkapitel 9.2). Auf die Kovariate Intelligenz wurde im Gegensatz zu Studie 1 verzichtet. Die Teilnehmer von Studie 2 erhielten in der ersten Lernphase (Sitzung 2) eines von drei Treatments, die nicht Teil von Studie 1 waren. Das „Fang-das-Monster“-Spiel wurde unverändert von Studie 1 nach Studie 2 übernommen. Jedoch bearbeiteten die Teilnehmer statt der 172 Aufgaben in Studie 1 in Studie 2 lediglich 80 Aufgaben im Rahmen des Spiels, 8 davon mit zusätzlichen schriftlichen Selbsterklärungen.

In Studie 1 zeigten sich am dritten Messzeitpunkt leichte Deckeneffekte. Diese könnten in Studie 2 verstärkt auftreten, weil der Untersuchungszeitraum statt vier Monaten eine Woche beträgt und weil das selbstentdeckende Lernen mit Feedback beim „Fang-das-Monster“-Spiel in zwei Treatmentgruppen durch eine Intervention zur gezielten Wissensvermittlung ergänzt wird. Damit in Studie 2 trotzdem keine Deckeneffekte auftreten, wurden in Studie 2, wie schon berichtet, lediglich Fünftklässler untersucht. Außerdem wurden die beiden nachfolgend beschriebenen Interventionen nicht so gestaltet, dass sie die größtmögliche Menge an Wissen vermitteln, sondern lediglich so, dass angenommen werden kann, dass sie eine nachweisbare Menge an Wissen vermitteln.

Intervention TK zur Vermittlung konzeptuellen Wissens

Das Treatment von Gruppe TK wurde so gestaltet, dass die Gruppe mit hoher Wahrscheinlichkeit mehr konzeptuelles als prozedurales Wissen erwirbt. Lerntheorien, insbesondere der ACT-Theorie, zufolge kann konzeptuelles Wissen durch die Rezeption von Erklärungen und durch eigene Reflektion erworben werden (vgl. Abschnitt 5.4.1). Beides wurde in der Intervention TK implementiert. Die Erklärung wurde den Teilnehmern über den Kopfhörer vorgespielt, während sie gleichzeitig in der Erklärung benutzte Beispieldezimalbrüche auf dem Bildschirm sehen konnten.

Der Erklärungsinhalt ist an Hieberts (1992) Analyse der konzeptuellen Wissensgrundlagen des Dezimalbruchverständnisses angelehnt. Den Teilnehmern wurde unter Verwendung konkreter anschaulicher Beispiele erläutert, dass bei Dezimalbrüchen die Stelle direkt vor dem Komma die Einer angibt, die Ziffer direkt nach dem Komma die Zehntel, die darauf folgende Ziffer die Hundertstel und die darauf folgende die Tausendstel. Die Erklärung enthielt damit Informationen über alle fünf von Hiebert herausgearbeiteten Komponenten des Dezimalbruchverständnisses (s. Unterkapitel 6.2).

Am Ende der etwa vier Minuten dauernden Erklärung wurde auch kurz der von null bis eins reichende Zahlenstrahl des Wissenstests gezeigt und es wurde auf die Beschriftung des Zahlenstrahls hingewiesen und erklärt, dass alle Zahlen zwischen null und eins auf ihm abgebildet werden können. Genaue Prozeduren dafür wurden jedoch nicht vorgestellt, weil dies eventuell zu einem Anstieg des prozeduralen Wissens führen könnte.

Um den Teilnehmern Gelegenheit zu geben, über die Erklärungsinhalte zu reflektieren und sie zu ihrem Vorwissen in Bezug zu setzen, wurden ihnen dann nacheinander jeweils (1) ein Dezimalbruch und ein Balkendiagramm, (2) ein Dezimalbruch und ein Foto eines angeschnittenen Apfels und (3) ein Dezimalbruch und ein Bruch in Bruchstrichnotation auf dem Bildschirm gezeigt. Sie wurden darauf hingewiesen, dass alle dargestellten Entitäten (Dezimalbrüche, Balkendiagramme, Fotos und Brüche mit Bruchstrich) zur Repräsentation von Quantitäten geeignet sind, und gebeten, zu jedem der drei Paare eine wichtige Gemeinsamkeit und einen wichtigen Unterschied zwischen den beiden Repräsentationsformen aufzuschreiben.

Dies bot ihnen zum Beispiel die Gelegenheit, darüber nachzudenken, dass alle vier Repräsentationsformen dazu geeignet sind, nicht ganzzahlige Quantitäten zu repräsentieren, dass diese durch Brüche genauer dargestellt werden können als durch Fotos und Balkendiagramme. Auf Preisschildern werden Dezimalbrüche statt Brüche mit Bruchstrich verwendet, weil man erstere leichter vergleichen kann. Mit Dezimalbrüchen kann man auch leichter rechnen als mit Balkendiagrammen. Die Reflexion über diese und ähnliche Punkte sollte den Kindern helfen, generelle Eigenschaften und den generellen Nutzen von Dezimalbrüchen besser zu verstehen und die zuvor gehörte Erklärung zu ihrem Vorwissen in Bezug zu setzen. Nachdem die Kinder versucht hatten, alle sechs Fragen (eine Gemeinsamkeit und einen Unterschied für jedes der drei Paare) zu beantworten, wurde ihnen für jede Frage eine korrekte Beispielantwort gezeigt, damit auch Kinder von der Reflexionsphase profitieren konnten, denen die eigene Reflexion schwer fiel.

Da die Teilnehmer von Intervention TK die Zahlenstrahlaufgaben weder präsentiert, noch erklärt bekommen, noch ausprobieren können, sollten sie kein prozedurales Wissen über diesen Aufgabentyp erwerben können.

Der gesamte Text und das Bildmaterial des Treatments TK befinden sich in Anhang F.1.2. Die Umsetzung dieser und der folgenden beiden Interventionen in das Computerprogramm kann in den Anhängen B.2 und B.3 besichtigt werden.

Intervention TP zur Vermittlung prozeduralen Wissens

Um den Teilnehmern der Intervention TP prozedurales Wissen über die Zahlenstrahlaufgaben möglichst unabhängig von konzeptuellem Wissen zu vermitteln, wurden ihnen zuerst die korrekten Positionen von fünf verschiedene Dezimalbrüchen auf dem Zahlenstrahl gezeigt und sie wurden gebeten, den Schieberegler auf diese Stelle zu bewegen und sie sich zu merken. Anschließend wurden ihnen in einem kleinen Test die Dezimalbrüche wieder dargeboten, und sie sollten ihre Positionen auf dem Zahlenstrahl markieren. Geling das nicht, wurden ihnen die fünf Beispiele erneut präsentiert so lange, bis sie den Test ohne Fehler absolvieren konnten. Je einer der fünf Dezimalbrüche entsprach einem der fünf Typen in Tabelle 9 auf Seite 135.

Anschließend wurden den Probanden insgesamt 80 Routineaufgaben dargeboten, jeweils 20 davon entsprachen einem der vier in Studie 1 verwendeten Oberflächentypen für die Routineaufgaben (Zahlenstrahl mit Schieberegler, frei anklickbarer Zahlenstrahl, multiple choice mit mehreren Zahlenstrahlen und multiple choice mit einem Zahlenstrahl). Den Probanden wurden dabei keine Rückmeldungen über die Korrektheit ihrer Eingaben gegeben.

Die Intervention basiert auf der Annahme, dass prozedurales Wissen nicht nur aus abstrakt-deklarativem Wissen, sondern auch aus konkreten Lösungsbeispielen (instances) abgeleitet werden kann (vgl. Unterkapitel 5.4). Dies wurde unter anderem von Anderson und Fincham (1994) sowie Anderson, Fincham und Douglass (1997) empirisch nachgewiesen und wird beispielsweise von der instance theory of automatization von Logan (1988) erklärt.

Die genannten Quellen belegen, dass prozedurales Wissen nicht sofort und automatisch aus Lösungsbeispielen abgeleitet wird, sondern dass dies nur geschieht, wenn die Aufgabenlösung geübt wird. Daher werden den Probanden in der Intervention erst fünf Lösungsbeispiele vermittelt, bevor sie diese in einer Trainingsphase anhand eines breiteren Zahlenmaterials in generelles prozedurales Wissen über die Zahlenstrahlaufgabe umsetzen können. Die Zahlen wurden dabei so wie in Studie 1 pseudozufällig ausgewählt.

Die Erteilung von Feedback in der Trainingsphase hätte den Erwerb prozeduralen Wissens sicherlich zusätzlich begünstigt. Sie hätte von den Teilnehmern jedoch auch zur Testung von Hypothesen über konzeptuelle Zusammenhänge und zur Falsifizierung von Misskonzepten genutzt werden können. Daher wurde auf Feedback in der Übungsphase verzichtet.

Es kann nicht ausgeschlossen werden, dass die Probanden während der Übung selbstständig aus den Lösungsbeispielen oder aus ihrem Vorwissen neues konzeptuelles Wissen ableiten können. Allerdings kann erwartet werden, dass in Gruppe TP die Zuwächse an prozeduralem Wissen wesentlich größer sind als die Zuwächse an konzeptuellem Wissen.

Der genaue Ablauf von Intervention TP und die wörtlichen Instruktionen befinden sich in Anhang F.1.3.

Intervention TX für die Kontrollgruppe

Die Kontrollgruppe sollte für etwa dieselbe Zeit am Computer beschäftigt werden wie die anderen beiden Gruppen. Aus Studie 1 ist bekannt, dass die Auseinandersetzung mit den Dezimalbrüchen den Kindern Spaß macht und als mittelschwierig erlebt wird. Daher wurde eine ebensolche Aufgabe für die Intervention TX gesucht. Damit die Teilnehmer von Intervention TX als Kontrollgruppe fungieren können, dürfen sie während der Intervention weder konzeptuelles noch prozedurales Wissen über Dezimalbrüche erhalten.

Um dieses Ziel zu erreichen, wurden den Kindern am Bildschirm nacheinander sieben Absätze aus einem Text über Elefanten gezeigt. Der Text war inhaltlich und sprachlich für Fünftklässler geeignet. Pro Absatz waren drei unter dem Text angegebene Worte durch ein Sternchen ersetzt worden. Außerdem enthielt jeder Absatz zwei Rechtschreibfehler. Die Aufgabe der Kinder bestand darin, die Sterne wieder durch die Worte zu ersetzen und die Rechtschreibfehler zu korrigieren. Die Wortersetzungen und einige der Fehlerkorrekturen ließen sich nur durchführen, wenn man den Text wirklich las und seine Aussagen inhaltlich nachvollzog. Man konnte sich den jeweils nächsten Absatz nur anzeigen lassen, wenn man alle Lücken und Fehler auf dem Bildschirm beseitigt hatte. Um Kinder mit schlechten Rechtschreibkenntnissen nicht zu überfordern, wurden die Positionen von noch vorhandenen Fehlern schwarz markiert, wenn man versuchte sich trotz noch vorhandener Fehler den jeweiligen Folgeabsatz anzeigen zu lassen. Den Inhalt der markierten Fehler mussten die Kinder jedoch selbst herausbekommen. Während dieser Intervention und den anderen beiden durften sie sich bei Fragen jederzeit an den Versuchsleiter wenden.

Da die Intervention TX nicht mathematischen Inhalts ist, wird erwartet, dass ihre Bearbeitung weder positive noch negative Effekte auf die Bearbeitung des Wissenstests haben wird, so dass Prätest-Posttest-Wissenszuwächse von Gruppe TX als Retesteffekte interpretiert werden können.

Das Maß Distance effect wurde erhoben wie in Studie 1 (vgl. S. 146). Damit die Retestreliabilität bestimmt werden kann, wurde es in Sitzung 1 und Sitzung 2 gemessen. Um zu überprüfen, ob die mentale Repräsentation von Zahlen einen Einfluss auf den Erwerb konzeptuellen und prozeduralen Wissens hat, wurde es als Kovariate in den hypothesentestenden Analysen verwendet.

10.2.3 Design und Auswertung

Es stellt sich die Frage, in welcher Weise die Befunde über die Faktorstruktur der Wissensmaße aus Studie 1 in Studie 2 genutzt werden können, ohne dass erneut konfirmatorische Faktoranalysen durchgeführt werden müssen.

Wichtig ist der in Abschnitt 9.4.3 berichtete Befund einer weitgehenden metrischen Messinvarianz der Wissensmaße. Zwar zeigte sich eine leichtere Verschlechterung der Messinvarianz bezüglich konzeptuellen Wissens, wenn die Messzeitpunkte mehrere Monate auseinanderliegen. Für prozedurales Wissen lag jedoch zu allen drei Messzeitpunkten und für beide Wissensarten zu den ersten beiden Messzeitpunkten uneingeschränkte Messinvarianz vor. Für die aktuelle Studie, deren Messzeitpunkte ja nur Tage, aber nicht Monate auseinanderliegen, bedeutet das, dass an allen drei Messzeitpunkten dieselben Operationalisierungen der Wissensarten verwendet werden können.

Ideal wäre für Studie 2 eine aus den Faktorladungen in Studie 1 abgeleitete Formel, in die man pro Messzeitpunkt und Wissensart die vier Indikatorwerte einer Person einsetzen und dann in den Faktorwert umrechnen kann. Eine solche Formel kann jedoch nicht hergeleitet werden. Die Pfeile in den Pfaddiagrammen in Kapitel 9 veranschaulichen das, indem sie von den latenten Faktoren zu ihren jeweiligen Indikatoren und nicht in die umgekehrte Richtung weisen. Alle in Studie 1 gerechneten Modelle analysieren also nicht, wie aus den Indikatorwerten die Faktorwerte abgeleitet werden können, sondern wie die Faktorwerte die Indikatorwerte determinieren. Dies ist die in SEM übliche Analyserichtung. Sie begründet sich aus der Annahme, dass es die Ausprägung des latenten Faktors (hier: einer Wissensart) ist, die die gemessenen Indikatorwerte (hier: eines Wissensmaßes) hervorruft. Eine umgekehrte Verursachungsrichtung ist zwar prinzipiell vorstellbar und analysierbar (Bollen & Lennox, 1991; Borsboom, Mellenbergh, &

Heerden, 2003), führt aber zu konzeptuellen und methodischen Problemen. Vor allem im vorliegenden Fall ist sie inhaltlich unangemessen und auch methodisch problematisch, weil zwar für die Faktorladungen der Indikatoren (vgl. Abschnitt 9.4.3), offensichtlich jedoch nicht für die Intercepts der Indikatoren Messinvarianz vorliegt (vgl. die Outputdatei von Modell R12F in Anhang D.12.1 und die Übersicht über die Mittelwertsunterschiede zwischen den Indikatoren in Tabelle 14 auf S. 155).

Als Ausweg bietet sich zunächst an, die vier Indikatorwerte pro Person, Wissensart und Messzeitpunkt einfach zu einem Summenscore aufzuaddieren, um die Ausprägung der Wissensart abzuschätzen. Dabei besteht jedoch das Problem, dass die Maße auf unterschiedlichen Skalen mit verschiedenen Mittelwerten und Varianzen liegen und daher nicht gleichberechtigt in den Summenscore eingehen würden. Um alle Maße auf dieselbe Skala zu bringen, können sie z-standardisiert werden. Da sie pro Messzeitpunkt gleich gewichtet in die Berechnung der Summenscores eingehen sollen, müssten sie an ihren Mittelwerten und Standardabweichungen pro Messzeitpunkt standardisiert werden. Dabei bleiben alle interindividuellen Unterschiede, also auch die Treatmentgruppenunterschiede, pro Messzeitpunkt vollständig erhalten. Die Ausprägungen der Summenscores für eine Wissensart können hingegen nicht über die Messzeitpunkte hinweg verglichen werden, weil sie unterschiedlichen Transformationen unterzogen worden sind. Beispielsweise kann vermutet werden, dass der wahre Gesamtstichprobenmittelwert des Summenscores für konzeptuelles Wissen über die Messzeitpunkte hinweg ansteigen wird. Berechnet man die Gesamtstichprobenmittelwerte hingegen auf Grundlage der z-standardisierten Werte, kann sich prinzipiell kein Anstieg mehr zeigen, weil die Gruppenmittelwerte zu jedem Messzeitpunkt null betragen.

Die Treatmentgruppenunterschiede sollen, da es sich um eine nominalskalierte unabhängige und eine intervallskalierte abhängige Variable handelt, varianzanalytisch ausgewertet werden.

Aus obigen Überlegungen zur mangelnden Vergleichbarkeit der z-standardisierten Messwerte und Summenscores folgt, dass der Messzeitpunkt nicht als zweiter Faktor in das varianzanalytische Design eingebracht werden kann. Denn Mittelwertsunterschiede über die Zeit hinweg können nicht sinnvoll interpretiert werden. Jedoch bleiben bei der z-Standardisierung die interindividuellen Unterschiede pro Messzeitpunkt auf Intervallskalenniveau erhalten. Es bietet sich daher an, die Vorwissensunterschiede im varianzanalytischen Design zu kontrollieren, indem das konzeptuelle und prozedurale Vorwissen der Teilnehmer in Form von Kovariaten miteingebracht wird. So können die

interindividuellen Unterschiede in der Kovariate benutzt werden, um die abhängige Variable zu bereinigen.

Das vorgeschlagene Design bietet die Möglichkeit, Befunde über die Faktorstruktur der Wissensmaße in Studie 1 doch noch indirekt zur Optimierung der Wissensmessung in Studie 2 zu verwenden. Das bisher geschilderte Design sieht nämlich vor, dass die vier Wissensmaße gleichberechtigt in ihren jeweiligen Summenscore eingehen. Aus Studie 1 ist jedoch bekannt, dass die Maße unterschiedliche Reliabilitäten besitzen und die wahre Ausprägung einer Wissensart unterschiedlich gut abschätzen. Maße für diese Reliabilitätsunterschiede stellen die quadrierten Faktorladungen dar, da sie den Anteil der durch den latenten Faktor aufgeklärten Varianz an der gemessenen Indikatorvarianz angeben (vgl. Unterkapitel 7.2). Reliablere Indikatoren sollten auf die Abschätzung der Ausprägung einer Wissensart jedoch einen größeren Einfluss haben als unreliablere. Um dies zu erreichen, wurden die z-standardisierten Wissensmaße bei ihrer Aufaddierung zu dem jeweiligen Summenscore mit ihren Reliabilitätskoeffizienten gewichtet.

Die Reliabilitätskoeffizienten zusammen mit den Faktorladungen gibt Tabelle 37 wieder. Die Faktorladungen wurden Tabelle 28 auf Seite 177 entnommen. Da sie sich pro Messzeitpunkt leicht unterscheiden, wurde die mittlere Reliabilität durch Addition der Einzelreliabilitäten und Division durch zwei berechnet. Da Reliabilitäten als Varianzanteile definiert sind und Varianzanteile intervallskaliert sind, ist diese Durchschnittsbildung zulässig.

Tabelle 37: Faktorladungen und Reliabilitäten von Modell R12F (Studie 1).

	Standardisierte Ladungen		Reliabilitäten		Mittlere Reliabilität
	Mzp 1	Mzp 2	Mzp 1	Mzp 2	
Maße konzeptuellen Wissens					
Evaluation	,547	,522	,299	,272	,286
Übersetzung	,781	,728	,610	,530	,570
Größenvergleich	,628	,770	,394	,593	,494
Erklärung	,336	,303	,113	,092	,102
Maße prozeduralen Wissens					
Problemlösekorrektheit	,813	,833	,661	,694	,677
Problemlösedauer	,194	,247	,038	,061	,049
Zugriffsasymmetrie	,292	,340	,085	,116	,100
Dual task-Kosten	-,347	-,390	,120	,152	,136

Im Folgenden werden zur besseren Unterscheidbarkeit die einzelnen Maße für konzeptuelles oder prozedurales Wissen als *Wissensmaße* und die gewichteten Summen über jeweils vier Wissensmaße als *Wissensscores* bezeichnet. Wie bisher wird auch weiterhin die programmierte Form der acht Wissensmaße als Wissenstest bezeichnet.

Zur besseren Darstellbarkeit wurden die z-standardisierten Werte der Wissensmaße vor der Berechnung der Wissensscores mit 10 multipliziert und zu 50 addiert. Die durch Aufsummierung von je vier Wissensmaßen entstandenen Wissensscores wurden anschließend durch die Summe der Gewichtungen geteilt, weil so die Mittelwerte aller z-standardisierten Wissensmaße und aller Wissensscores bei 50 liegen und leichter zu vergleichen sind.

Die schriftlichen Antworten der Kinder im Rahmen des Maßes Erklärung wurden von denselben beiden Ratern und in derselben Weise wie in Studie 1 kodiert.

Die Entwicklung und Erprobung des Trainingsmaterials erfolgte in einer qualitativen Vorstudie mit 38 Teilnehmern. Während ein oder zwei Teilnehmer Tests und Treatment absolvierten, wurden sie jeweils zur Gestaltung und zum Verständnis befragt. Außerdem wurden ihre Daten daraufhin ausgewertet, ob die Treatments wirksam sind und ungefähr gleich lange dauern. Anschließend wurde das Programm jeweils modifiziert und dann an den nächsten Teilnehmern erprobt.

Die Datenaufbereitung erfolgte analog zum Vorgehen in Studie 1. Anhang H.2 enthält die zugehörigen SPSS-Syntaxdateien und Programme.

10.3 Ergebnisse

10.3.1 Beschreibung der Stichprobe

Von den 93 Kindern, mit denen Termine zur Studiendurchführung vereinbart wurden, mussten 9 von der Auswertung ausgeschlossen werden, weil sie nicht vollständig an Sitzung 1 oder Sitzung 2 teilnehmen konnten. Bei zweien dieser Kinder stürzte ohne ihr Verschulden der Computer ab, vier fehlten entschuldigt wegen Krankheit und drei fehlten ohne Angabe von Gründen. In Sitzung 3 stürzte bei einem Teilnehmer der Rechner ab und ein Teilnehmer nahm den Termin nicht wahr. Die Daten dieser beiden Personen (aus den ersten beiden Sitzungen) wurden mitausgewertet. Eine Übersicht über die einzelnen Fälle mit unvollständigen Daten und die jeweilige Ursache befindet sich in Anhang H.4.

Die Durchführung erfolgte während der Wintermonate und die erscheinenden Teilnehmer gaben an, dass eine Erkältungswelle an ihren Schulen mitverantwortlich für die Zahl der Studienabbrecher sei.

Alle nachfolgenden Analysen beziehen sich auf 84 Personen, von denen für 2 keine Daten vom dritten Messzeitpunkt vorliegen. Beide entstammen Gruppe TK. In Gruppe TK befinden sich insgesamt 31 Personen (18 weiblich; 13 männlich), in Gruppe TP 27 Personen (10 weiblich; 17 männlich) und in Gruppe TX 26 Personen (11 weiblich; 15 männlich). Alle Rohdaten befinden sich in Anhang H.1. Die Ergebnisse im aktuellen

Abschnitt wurden unter Verwendung der Syntax in Anhang H.4 berechnet. In allen Analysen dieses Kapitels wird, sofern keine anderslautenden Informationen gegeben werden, der Datensatz FinalData1.sav aus Anhang H.3 benutzt.

Tabelle 38 zeigt die wichtigsten erhobenen Personeneigenschaften für die drei Treatmentgruppen und die Gesamtstichprobe. Die Notenmittelwerte der Gesamtstichprobe sind fast identisch mit denen aus Studie 1 und zeigen deutlich, dass nicht nur Kinder mit guten Schulleistungen an der Studie teilnahmen.

Tabelle 38: Deskriptive Charakteristika der drei Treatmentgruppen im Vergleich.

	Gruppe TK		Gruppe TP		Gruppe TX		Gesamtstichprobe			
	<i>m</i>	<i>SD</i>	<i>m</i>	<i>SD</i>	<i>m</i>	<i>SD</i>	<i>m</i>	<i>SD</i>	<i>min</i>	<i>max</i>
Alter in Jahren	11,2	0,4	11,4	0,9	11,3	0,5	11,3	0,6	10,4	15,0
Mathematiknote	2,4	0,7	2,4	0,7	2,5	0,8	2,4	0,7	4	1
Deutschnote	2,5	0,7	2,6	0,7	2,5	0,8	2,5	0,7	4	1
„Ich kann gut mit dem Computer umgehen.“ ^a	3,9	0,9	4,0	1,0	3,9	0,8	3,9	0,9	2	5
„Ich habe Übung im Umgang mit der Maus.“ ^a	4,3	0,8	4,6	0,6	4,4	0,9	4,4	0,8	2	5
„Ich nehme an dieser Untersuchung teil, weil ich etwas lernen möchte.“ ^a	4,5	0,8	4,5	0,8	4,5	0,9	4,6	0,6	1	5
„Ich nehme an dieser Untersuchung teil, weil ich dafür etwas Geld bekomme.“ ^a	2,3	1,2	2,6	1,1	2,4	1,2	2,4	1,2	1	5
„Ich weiß über Dezimalbrüche gut Bescheid.“ ^a	2,0	1,0	2,1	1,0	2,2	1,2	2,1	1,0	1	5
Dauer der Treatments	20,6	4,4	13,5	6,3	22,4	10,4	18,8	8,1	7	63

Anmerkung. ^a Angaben auf einer fünfstufigen Ratingskala von 1 = *stimmt überhaupt nicht* bis 5 = *stimmt völlig*.

Zwischen den Treatmentgruppen bestehen keine inhaltlich bedeutsamen Unterschiede hinsichtlich Alter, Schulnoten, Computererfahrung, Motivation der Studienteilnahme und Selbsteinschätzung des Vorwissens.

Pro Treatmentgruppe war lediglich die Anzahl der zu bearbeitenden Aufgaben vorgegeben und jedes Kind konnte mit seiner eigenen Geschwindigkeit arbeiten. Daher werden nun die Treatmentgruppen auf ihre Bearbeitungszeiten hin verglichen.

Die Mittelwerte betragen $m = 20,6$ Minuten ($SD = 4,4$) für Gruppe TK, $m = 13,5$ Minuten ($SD = 6,2$) für Gruppe TP und $m = 22,4$ Minuten ($SD = 10,4$) für Gruppe TX.

Tabelle 39 zeigt die Ergebnisse der Varianzanalyse mit unabhängiger Variable Treatmentgruppe (TK, TP, TX) und abhängiger Variable Treatmentdauer. Die Gruppen unterscheiden sich hochsignifikant voneinander. Wie Scheffé-Tests belegen, unterscheiden sich die Gruppe TK und TX nicht signifikant voneinander ($\Delta m = 1,8\text{min}$; $p = ,650$). Gruppe TP braucht jedoch weniger Zeit als Gruppe TK ($\Delta m = 7,1\text{min}$; $p = ,002$) und Gruppe TX ($\Delta m = 8,9\text{min}$; $p = ,000$).

Tabelle 39: Ergebnisse der Varianzanalyse zur Dauer der Treatments.

	QS	df	$\hat{\sigma}^2$	F	p	Sign.	η^2
Treatment	1187,376	2	593,688	11,180	,000	***	,216
Fehler	4301,450	81	53,104	-	-	-	-
Total	5488,826	83	-	-	-	-	-

10.3.2 Charakteristika der Wissensscores

Im Rahmen der Ausreißerbereinigung wurden bei den Maßen Problemlösedauer, Zugriffsasymmetrie I und Zugriffsasymmetrie II pro Messzeitpunkt zwischen 6 und 15 Werte entfernt, jedoch nie mehr als ein Wert pro Person, Maß und Messzeitpunkt (s. Anhang H.2.3). Aus den bereinigten Wissensmaßen wurden dann die Wissensscores abgeleitet.

Tabelle 40 gibt für jeden Wissensscore seinen Mittelwert an, seine Standardabweichung, Irrtumswahrscheinlichkeit p und Signifikanz des Kolmogorov-Smirnov-Anpassungstests an die Normalverteilung, die Pearson-Korrelation r zwischen den gewichtet aufsummierten Wissensmaßen und den ungewichtet aufsummierten Wissensmaßen und Cronbachs Alphakoeffizienten als Maß der internen Konsistenz der vier jeweiligen Wissensmaße als Operationalisierungen desselben Konstrukts (für die Syntax- und Outputdateien s. Anhang H.4). Wie erwartet, hat jeder Mittelwert aufgrund der Standardisierung den Wert 50. Die Standardabweichungen der Wissensscores sind wegen des gewichteten Aufsummierens etwas kleiner als die auf den Wert 10 gebrachten Standardabweichungen der Wissensmaße. Die Kolmogorov-Smirnov-Anpassungstests belegen, dass nur zwei der sechs Maße signifikant von der Normalverteilung abweichen. Die beiden Irrtumswahrscheinlichkeiten liegen jedoch nicht weit unter der kritischen Signifikanzgrenze und deuten somit auf eher geringe Effektstärken hin.

Tabelle 40: Verteilungseigenschaften der Wissensscores.

	Verteilung		KS-Anpassungstest an Normalverteilung		r der gew. mit den ungew. Scores	α
	m	SD	p	Sign.		
Mzp 1						
Wissensscore K_1	50,0	8,2	,216	n.s.	,509	,750
Wissensscore P_1	50,0	8,0	,459	n.s.	,699	,473
Mzp 2						
Wissensscore K_2	50,0	8,4	,665	n.s.	,589	,713
Wissensscore P_2	50,0	7,9	,044	*	,684	,569
Mzp 3						
Wissensscore K_3	50,0	8,4	,279	n.s.	,513	,783
Wissensscore P_3	50,0	7,9	,027	*	,606	,527

Anmerkungen. $N = 84$ für Mzp 1 und Mzp 2; $N = 82$ für Mzp 3.

Die Korrelationen der Summen der mit den Reliabilitäten gewichteten Wissensmaße und der Summen der ungewichteten Wissensmaße sind alle deutlich kleiner als eins und

verdeutlichen, dass die Gewichtung die Schätzung der Wissensausprägungen substanziell beeinflusst.

Cronbachs Alphawerte liegen im selben Bereich wie die entsprechenden internen Konsistenzen der Faktoren in Studie 1 (vgl. Tabelle 17 auf S. 158). Erneut weisen die Messungen konzeptuellen Wissens höhere interne Konsistenzen als die Messungen prozeduralen Wissen auf. Bei der Interpretation ist zu beachten, dass die Alphakoeffizienten für den zweiten und den dritten Messzeitpunkt nur unter Vorbehalten interpretiert werden dürfen, weil aufgrund der Treatmenterteilung nicht mehr alle Personenwerte als voneinander unabhängige Messungen betrachtet werden können.

Am ersten Messzeitpunkt sind die Wissensscores für konzeptuelles und prozedurales Wissen zu $r_1 = ,699$ korreliert, am zweiten Messzeitpunkt zu $r_2 = ,748$ und am dritten zu $r_3 = ,650$. Sie sind damit etwas niedriger als in Studie 1, doch noch immer hoch. Weil am zweiten und am dritten Messzeitpunkt keine unabhängigen Messungen vorliegen, werden die Korrelationen in dieser Arbeit jedoch nicht inhaltlich interpretiert, sondern lediglich als explorativ gewonnene Indizien präsentiert, die die Annahme stützen, dass die in Studie 1 gewonnenen Befunde über die Verwendbarkeit der Wissensmaße auf Studie 2 generalisierbar sind.

Weitere Indizien, die darauf hindeuten, liefert eine explorative Faktoranalyse (principal components analysis mit Varimax-Rotation), die wegen des Problems der abhängigen Messungen nach der Treatmenterteilung nur für die Wissensmaße des ersten Messzeitpunkts durchgeführt wurde. Dabei wurden zwei Faktoren mit einem Eigenwert größer eins gefunden. Der erste erklärt 39,3% der Varianz, der zweite 15,4%. Die Faktorladungen werden in Tabelle 41 dargestellt. In Übereinstimmung mit den Befunden von Studie 1 laden alle vier Maße konzeptuellen Wissens hoch auf einem gemeinsamen Faktor, auf dem auch die Problemlösekorrektheit als Maß prozeduralen Wissens hoch lädt. Die drei übrigen Maße prozeduralen Wissens bilden dieses Mal ein etwas homogeneres Muster, indem sie alle hoch auf dem zweiten Faktor laden. Wie erwartet, werden die Dual task-Kosten umso kleiner, je größer die Zugriffsasymmetrie wird. Wie in Studie 1 hat jedoch auch hier die Problemlösedauer eine Faktorladung mit unerwartetem Vorzeichen. Sie sollte gemeinsam mit den Dual task-Kosten kleiner werden, wenn das prozedurale Wissen zunimmt, wird tatsächlich jedoch eher größer. All diese Parallelen zwischen den Befunden zu den Funktionsweisen der Wissensmaße in Studie 1 und Studie 2 lassen darauf schließen, dass es gerechtfertigt ist, die Maße konzeptuellen und prozeduralen Wissens

jeweils zu einem Wissensscore zusammenzufassen. Die Einwirkungen der Treatments auf diese Scores werden im Folgenden berichtet.

Tabelle 41: Ladungen der explorativen Faktoranalyse für Mzp 1.

	Faktor 1	Faktor 2
Maße konzeptuellen Wissens		
Evaluation	,77	,11
Übersetzung	,81	,10
Größenvergleich	,82	,05
Erklärung	,49	,15
Maße prozeduralen Wissens		
Problemlösekorrektheit	,83	,08
Problemlösedauer	-,03	,66
Zugriffsasymmetrie	,28	,60
Dual task-Kosten	-,10	-,77

Anmerkungen. Ladungen > .40 als Interpretationshilfe fettgedruckt; N = 84.

10.3.3 Einfluss der Treatments auf die Wissensscores

Alle Syntax- und Outputdateien der Analysen dieses Abschnitts befinden sich in Anhang H.6. Die Mittelwerte und Standardabweichungen der Wissensscores in der Gesamtstichprobe und den drei Treatmentgruppen zu den drei Messzeitpunkten gibt Tabelle 42 an. Man beachte, dass die Werte nur zeilen-, nicht jedoch spaltenweise verglichen werden können, weil sie über die Messzeitpunkte hinweg an unterschiedlichen Werten standardisiert worden sind. Gruppe TP hat zu allen drei Messzeitpunkten die höchsten Ausprägungen sowohl auf dem Score für konzeptuelles als auch auf dem Score für prozedurales Wissen. Gemessen an den Standardabweichungen sind jedoch die meisten Mittelwertsunterschiede in der Tabelle relativ gering.

Tabelle 42: Mittelwerte und Standardabweichungen der Wissensscores in den Treatmentgruppen.

	Gesamt		Gruppe TK		Gruppe TP		Gruppe TX	
	m	SD	m	SD	m	SD	m	SD
Mzp 1								
Wissensscore K ₁	50,0	8,2	49,5	7,1	51,5	9,2	49,1	8,4
Wissensscore P ₁	50,0	8,0	50,1	8,9	51,0	7,8	48,9	6,8
Mzp 2								
Wissensscore K ₂	50,0	8,4	48,7	8,4	52,8	7,7	48,6	7,4
Wissensscore P ₂	50,0	7,9	48,1	7,9	54,5	7,1	47,6	8,5
Mzp 3								
Wissensscore K ₃	50,0	8,4	50,0	8,9	51,2	7,9	48,8	8,4
Wissensscore P ₃	50,0	7,9	50,5	8,1	51,9	5,7	47,6	9,3

Gruppenunterschiede vor der ersten Lernphase

In zwei einfaktoriellen Varianzanalysen mit jeweils dreifach gestuftem Faktor Treatmentgruppe (TL, TP, TX) wurde zunächst für den ersten Messzeitpunkt getestet, ob die Gruppen sich bezüglich ihres konzeptuellen oder ihres prozeduralen Wissensscores

überzufällig voneinander unterscheiden. Zur Voraussetzungsüberprüfung wurde für beide Analysen Levenes Test auf Gleichheit der Fehlervarianzen durchgeführt. Dieser zeigte weder für den konzeptuellen Wissensscore ($F_{(2; 81)} = 1,679$; $p = ,193$) noch für den prozeduralen ($F_{(2; 81)} = 1,496$; $p = ,230$) eine Verletzung an.

Die Ergebnisse der beiden Varianzanalysen in Tabelle 43 verdeutlichen, dass es zum ersten Messzeitpunkt bezüglich des konzeptuellen und prozeduralen Vorwissens keine überzufällig großen Gruppenunterschiede gibt.

Tabelle 43: Ergebnisse der Varianzanalysen zum ersten Messzeitpunkt.

	QS	df	$\hat{\sigma}^2$	F	p	Sign.	η^2
Konzeptueller Wissensscore K ₁							
Treatment	92,864	2	46,432	0,686	,506	n.s.	,017
Fehler	5482,137	81	67,681	-	-	-	-
Total	5575,001	83	-	-	-	-	-
Prozeduraler Wissensscore P ₁							
Treatment	60,642	2	30,321	0,483	,619	n.s.	,012
Fehler	5088,379	81	62,819	-	-	-	-
Total	5149,021	83	-	-	-	-	-

Treatmentgruppenunterschiede nach der ersten Lernphase

Im Folgenden werden die Einflüsse der Treatments auf die Wissenszuwächse während der ersten Lernphase (Sitzung 2) überprüft. In dieser Lernphase wurden die drei unterschiedlichen Treatments erteilt. Hypothesengemäß sollte Gruppe TK bis zum zweiten Messzeitpunkt mehr konzeptuelles Wissen erwerben als die anderen beiden und Gruppe TP mehr prozedurales Wissen als die anderen beiden. Wenn hier und im Folgenden von Wissenszuwächsen gesprochen wird, so sind damit stets für die Kovariate Vorwissen kontrollierte Wissensunterschiede zu einem Messzeitpunkt gemeint, nicht Wertedifferenzen zwischen Messzeitpunkten, die aufgrund der z-Standardisierungen der Wissensmaße pro Messzeitpunkt nicht interpretierbar sind.

Um die Gruppenunterschiede zu testen, wurde für jede Wissensart einzeln eine einfaktorielle Kovarianzanalyse mit dem Faktor Treatmentgruppe (TK, TP, TX) gerechnet. Um Vorwissensunterschiede auszugleichen, wurde als Kovariate der jeweilige Wissensscore vom ersten Messzeitpunkt benutzt. Levenes Test auf Gleichheit der Fehlervarianzen indizierte weder für den konzeptuellen Wissensscore ($F_{(2; 81)} = 0,490$; $p = ,615$) noch für den prozeduralen ($F_{(2; 81)} = 1,195$; $p = ,308$) signifikante Voraussetzungsverletzungen.

Tabelle 44 zeigt die Ergebnisse der beiden Varianzanalysen. In beiden Fällen leistete das jeweilige Vorwissen einen großen und statistisch sehr signifikanten Beitrag zur

Varianzaufklärung. Dieser Beitrag ist für die abhängige Variable K_2 größer als für die abhängige Variable P_2 .

Tabelle 44: Ergebnisse der Kovarianzanalysen bezüglich Treatmentgruppenunterschieden an Mzp 2.

	QS	df	$\hat{\sigma}^2$	F	p	Sign.	Part. η^2
Konzeptueller Wissensscore K_2							
Kovariate K_1	2657,779	1	2657,779	90,224	,000	***	,530
Treatment	115,938	2	57,969	1,968	,146	n.s.	,047
Fehler	2356,610	80	29,458	-	-	-	-
Total	5319,938	83	-	-	-	-	-
Prozeduraler Wissensscore P_2							
Kovariate P_1	1938,401	1	1938,401	50,683	,000	***	,388
Treatment	593,132	2	296,566	7,754	,001	**	,162
Fehler	3059,643	80	38,246	-	-	-	-
Total	5805,188	83	-	-	-	-	-

Die Treatments wirkten sich den aktuellen Analysen zufolge nicht signifikant auf die abhängige Variable konzeptuelles Wissen (K_2) aus, hatten jedoch einen hochsignifikanten Einfluss auf die abhängige Variable prozedurales Wissen (P_2). Geplante Kontraste belegen, dass Gruppe TP hypothesenkonform größere Zuwächse an prozeduralem Wissen erzielte als Gruppe TK (*Cohens* $d = 0,60$; $p = ,001$) und als Gruppe TX (*Cohens* $d = 0,88$; $p = ,002$).

Treatmentgruppenunterschiede nach der zweiten Lernphase

Um die Lernzuwächse während der zweiten Lernphase, des „Fang-das-Monster“-Spiels, das für alle drei Treatmentgruppen gleich war, zu vergleichen, wurde pro Wissensart eine einfaktorielle Kovarianzanalyse mit Faktor Treatment (TK, TP, TX) gerechnet. Um das Vorwissen zu kontrollieren, wurde der Score der jeweiligen Wissensart am zweiten Messzeitpunkt als Kovariate benutzt. Levenes Test auf Gleichheit der Fehlervarianzen indizierte für die abhängige Variable K_3 ($F_{(2; 79)} = 1,352$; $p = ,265$) keine Voraussetzungsverletzungen. Für die abhängige Variable P_3 wurden hingegen signifikante Gruppenunterschiede der Fehlervarianzen gefunden ($F_{(2; 79)} = 4,189$; $p = ,019$).

Die Ergebnisse der Kovarianzanalysen zeigt Tabelle 45. Erneut sind das konzeptuelle und das prozedurale Vorwissen Kovariaten mit hochsignifikanten und sehr bedeutsamen Einflüssen, wobei die interindividuellen Differenzen des konzeptuellen Wissens stabiler sind als die des prozeduralen Wissens. Jedoch hat das Treatment aus der ersten Lernphase keinerlei nachweisbare Einflüsse auf Unterschiede in den Zuwächsen konzeptuellen oder prozeduralen Wissens in der zweiten Lernphase. Die entsprechenden Effektstärken sind so klein (vgl. auch die Gruppenmittelwerte in Tabelle 42 auf S. 209), dass hier trotz

Voraussetzungsverletzungen auf die Durchführung voraussetzungsärmerer Verfahren, wie beispielsweise Rangvarianzanalysen, verzichtet werden kann.

Die Wissenszuwächse über beide Lernphasen hinweg wurden explorativ analysiert (s. Anhang H.6), führten jedoch zu keinen zusätzlichen Befunden und werden darum hier nicht berichtet. Weitere explorative Analysen werden in den folgenden Abschnitten vorgestellt.

Tabelle 45: Ergebnisse der Kovarianzanalysen bezüglich Treatmentgruppenunterschieden an Mzp 3.

	QS	df	$\hat{\sigma}^2$	F	p	Sign.	Part. η^2
Konzeptueller Wissensscore K ₃							
Kovariate K ₂	3318,666	1	3318,666	112,359	,000	***	,590
Treatment	62,808	2	31,404	1,063	,350	n.s.	,027
Fehler	2303,833	78	29,536	-	-	-	-
Total	5695,492	81	-	-	-	-	-
Prozeduraler Wissensscore P ₃							
Kovariate P ₂	1900,619	1	1900,619	50,987	,000	***	,395
Treatment	104,104	2	52,052	1,396	,254	n.s.	,035
Fehler	2907,552	78	37,276	-	-	-	-
Total	5061,294	81	-	-	-	-	-

10.3.4 Einfluss der Treatments auf die Wissensscoreinterkorrelationen

Wie in Abschnitt 2.3.4 und Unterkapitel 10.1 berichtet, wurde in Studien gefunden, dass Treatments sich auch dann unterschiedlich auf die Interkorrelationen zwischen Maßen konzeptuellen und prozeduralen Wissens auswirken können, wenn sie keine differenziellen Mittelwertsunterschiede hervorrufen. Um das Vorliegen solcher Effekte zu überprüfen, wurden die Korrelationen zwischen dem konzeptuellen und dem prozeduralen Wissensscore getrennt nach Treatmentgruppe und Messzeitpunkt berechnet.

Tabelle 46: Korrelationen zwischen den Scores der beiden Wissensarten.

	Mzp 1	Mzp 2	Mzp 3
Gruppe TK	,73	,72	,65
Gruppe TP	,68	,71	,56
Gruppe TX	,72	,78	,72

Die Resultate (für die Analysen s. Anhang H.7) sind in Tabelle 46 aufgeführt. Die Korrelationen sind alle hochsignifikant und liegen zwischen ,56 und ,73. Sie geben keine Hinweise auf systematische und praktisch bedeutsame Einflüsse der Treatments auf die Zusammenhänge zwischen den Wissensarten. Auffällig ist, dass sie sämtlich kleiner sind als die in Studie 1 gefundenen Faktorinterkorrelationen. Bei der Interpretation dieses Befundes wird mit einzubeziehen sein, dass die Stichprobengröße und dadurch auch die

interindividuelle Varianz in Studie 2 kleiner war als in Studie 1, was sich korrelationsreduzierend auswirken kann (Bortz, 1993, S. 198).

10.3.5 Einfluss der Treatments auf die Einzelwissensmaße

Die hypothesentestenden Analysen in Studie 2 stellen multivariate Tests dar in dem Sinne, dass mehrere abhängige Variablen zugleich in die Berechnung der Wissensscores eingeflossen sind. Wenn die multivariaten Analysen keine signifikanten Gruppenunterschiede aufweisen, können auch signifikante Gruppenunterschiede der Einzelmaße nicht als Bestätigung der Hypothesen gewertet werden. Trotzdem werden im Folgenden die Gruppenunterschiede bezüglich aller Einzelmaße explorativ analysiert, weil das wertvolle Hintergrundinformationen liefern kann. Die entsprechenden Auswertungsdateien befinden sich in Anhang H.8.

Tabelle 47: Mittelwerte und Standardabweichungen der Wissensmaße in den Treatmentgruppen.

	Mzp 1		Mzp 2		Mzp 3	
	<i>m</i>	<i>SD</i>	<i>m</i>	<i>SD</i>	<i>m</i>	<i>SD</i>
Maße konzeptuellen Wissens						
Evaluation						
Gruppe TK ^a	61,69	14,77	69,76	20,36	71,98	23,03
Gruppe TP ^a	56,48	20,91	68,98	20,03	66,67	22,20
Gruppe TX ^a	55,77	20,69	66,35	21,44	68,27	19,76
Übersetzung						
Gruppe TK ^a	49,19	14,95	56,45	17,90	70,52	18,00
Gruppe TP ^a	55,37	21,57	66,67	20,52	73,70	15,66
Gruppe TX ^a	49,62	15,93	56,54	16,66	70,77	16,11
Größenvergleich						
Gruppe TK ^a	74,03	14,52	77,10	18,38	85,00	13,43
Gruppe TP ^a	78,70	14,58	86,85	11,86	89,07	10,75
Gruppe TX ^a	75,38	14,62	79,81	13,30	82,50	12,67
Erklärung						
Gruppe TK ^a	16,53	21,50	27,02	18,85	25,00	20,41
Gruppe TP ^a	20,37	21,13	21,76	26,99	20,83	22,47
Gruppe TX ^a	12,02	15,61	15,39	15,54	20,67	18,01
Maße prozeduralen Wissens						
Problemlösekorrektheit						
Gruppe TK ^a	54,84	20,31	66,13	20,44	87,93	15,78
Gruppe TP ^a	56,48	21,02	83,52	18,39	91,30	10,88
Gruppe TX ^a	51,16	17,28	65,00	21,40	80,00	22,41
Problemlösedauer						
Gruppe TK ^b	81,27	3,71	79,43	3,40	79,17	3,62
Gruppe TP ^b	80,30	4,22	78,17	2,57	78,14	2,27
Gruppe TX ^b	79,16	4,65	78,16	1,79	77,57	2,36
Zugriffsasymmetrie						
Gruppe TK ^c	-1,64	1,98	-0,30	2,01	0,22	1,86
Gruppe TP ^c	-1,03	2,31	0,79	1,88	0,01	1,33
Gruppe TX ^c	-1,03	2,42	0,18	1,98	0,05	1,93
Dual task-Kosten						
Gruppe TK ^d	11,54	17,48	11,88	17,89	7,63	12,30
Gruppe TP ^d	9,05	25,79	7,89	10,84	5,36	7,56
Gruppe TX ^d	10,20	17,36	13,55	13,98	2,86	7,51

Anmerkungen. ^a Werte in %; ^b Werte in 10×ln(s); ^c Werte in Δ10×ln(s); ^d Werte in Δ%.

Tabelle 47 gibt die Mittelwerte und Standardabweichungen der acht Wissensmaße an den drei Messzeitpunkten für die drei Treatmentgruppen an. Die Werte belegen, dass keine starken Decken- oder Bodeneffekte vorliegen und alle Maße interpersonale Varianz aufweisen. Das Gesamtbild der Werte ist bei diesem Grad an Detailliertheit jedoch schwer zu interpretieren.

Tabelle 48: Cohens d und Signifikanz der Veränderungen der Wissensmaße über die Mzp.

	Mzp 1 zu Mzp 2			Mzp 2 zu Mzp 3		
	d	p	Sign.	d	p	Sign.
Maße konzeptuellen Wissens						
Evaluation						
Gruppe TK	0,32	,039	*	0,07	,468	n.s.
Gruppe TP	0,43	,002	**	-0,08	,476	n.s.
Gruppe TX	0,36	,019	*	0,07	,503	n.s.
Gruppe $\times$ Mzp	-	,705	n.s.	-	,519	n.s.
Übersetzung						
Gruppe TK	0,31	0,26	n.s.	0,55	,000	***
Gruppe TP	0,38	,006	**	0,27	,026	*
Gruppe TX	0,30	,037	*	0,61	,000	***
Gruppe $\times$ Mzp	-	,598	n.s.	-	,141	n.s.
Größenvergleich						
Gruppe TK	0,13	,277	n.s.	0,35	,008	**
Gruppe TP	0,43	,000	***	0,14	,103	n.s.
Gruppe TX	0,22	,035	*	0,15	,161	n.s.
Gruppe $\times$ Mzp	-	,287	n.s.	-	,101	n.s.
Erklärung						
Gruppe TK	0,37	,001	**	-0,07	,444	n.s.
Gruppe TP	0,04	,542	n.s.	-0,03	,161	n.s.
Gruppe TX	0,15	,258	n.s.	0,22	,054	+
Gruppe $\times$ Mzp	-	,045	*	-	,149	n.s.
Maße prozeduralen Wissens						
Problemlösekorrektheit						
Gruppe TK	0,39	,000	***	0,84	,000	***
Gruppe TP	0,97	,000	***	0,36	,006	**
Gruppe TX	0,50	,002	**	0,48	,001	**
Gruppe $\times$ Mzp	-	,003	**	-	,030	*
Problemlösedauer						
Gruppe TK	-0,37	,007	**	-0,05	,474	n.s.
Gruppe TP	-0,43	,024	*	-0,01	,953	n.s.
Gruppe TX	-0,20	,291	n.s.	-0,20	,281	n.s.
Gruppe $\times$ Mzp	-	,610	n.s.	-	,736	n.s.
Zugriffsasymmetrie						
Gruppe TK	0,47	,008	**	0,19	,038	*
Gruppe TP	0,61	,004	**	-0,34	,097	+
Gruppe TX	0,39	,031	*	-0,05	,815	n.s.
Gruppe $\times$ Mzp	-	,699	n.s.	-	,058	+
Dual task-Kosten						
Gruppe TK	0,01	,914	n.s.	-0,20	,088	+
Gruppe TP	-0,04	,813	n.s.	-0,19	,157	n.s.
Gruppe TX	0,15	,418	n.s.	-0,67	,001	**
Gruppe $\times$ Mzp	-	,373	n.s.	-	,067	+

Um die Interpretation zu erleichtern, wurde für jedes Maß, jede Lernphase und jede Treatmentgruppe getrennt jeweils mittels eines t -Tests für abhängige Stichproben geprüft, ob eine signifikante Veränderung vorliegt. Die Effektstärken dieser Veränderungen (Cohens d) und die Signifikanzen der t -Tests können Tabelle 48 entnommen werden. Pro

Maß und Lernphase ist dort außerdem aufgeführt, ob ein signifikanter Interaktionseffekt zwischen Treatment und Messzeitpunkt vorliegt. Dies wurde jeweils mittels Varianzanalyse mit dreifach gestuftem between-subjects-Faktor Treatment und dem zweifach gestuften within-persons-Faktor Messzeitpunkt geprüft. Da es sich um explorative Analysen handelt, wurden die Voraussetzungen nicht überprüft und die Parameter der Signifikanztests werden hier nicht aufgelistet. Sie können in Anhang H.8 eingesehen werden. Auf eine Alphafehlerkorrektur wurde verzichtet.

Alle signifikanten Effekte indizieren Wissenszuwächse, das heißt keine der Variablen entwickelte sich über die drei Messzeitpunkte hinweg überzufällig stark in eine unerwartete Richtung. Jedoch sind 35 der 72 Effekte nicht signifikant und zeigen an, dass nicht alle drei Gruppen auf jedem der Maße Wissenszuwächse erzielten.

Wäre konzeptuelles Wissen ein homogenes Konstrukt, das durch alle vier Maße konzeptuellen Wissens erhoben wird und das in der ersten Lernphase hauptsächlich in Intervention TK vermittelt wurde, so sollte Gruppe TK bei diesen vier Maßen größere Effektstärken aufweisen als die anderen beiden Gruppen. Dies ist bei den Maßen Evaluation, Übersetzung und Größenvergleich jedoch nicht der Fall. Die Effektstärken der drei Gruppen unterscheiden sich dort jeweils nicht signifikant voneinander. Lediglich das Maß Erklärung weist das erwartete Muster von Effektstärken auf: Gruppe TK hat signifikant größere Wissenszuwächse über die beiden Messzeitpunkte als die übrigen beiden Gruppen.

Bezüglich der Maße prozeduralen Wissens wurde für die erste Lernphase vermutet, dass Gruppe TP die größten Effektstärken aufweist. Dies ist bei den drei Maßen Problemlösedauer, Zugriffsasymmetrie und Dual task-Kosten jedoch nicht der Fall. Bei ihnen liegen keine signifikanten Unterschiede zwischen den Effektstärken vor. Einzig das Maß Problemlösedauer weist das erwartete Muster auf: Gruppe TP hat bei weitem die größten Wissenszuwächse und die Interaktion von Gruppe und Messzeitpunkt ist hoch signifikant.

Für die zweite Lernphase wurde erwartet, dass Gruppe TK auf den Maßen prozeduralen Wissens größere Lernzuwächse aufweist als Gruppe TX und dass Gruppe TP auf den Maßen konzeptuellen Wissens größere Lernzuwächse aufweist als Gruppe TX. Betrachtet man die einzelnen Effektstärken, so zeigt sich ein inhomogenes Muster, das diesen Erwartungen nicht entspricht. Da keine der Interaktionen zwischen Treatmentgruppe und Messzeitpunkt signifikant ist, sollte dieses Muster uninterpretiert bleiben, weil es im Bereich der Zufallsschwankungen liegt.

10.3.6 Reliabilitäten der Wissensmaße

In Abschnitt 10.3.2 wurden schon die internen Konsistenzen der Wissensscores untersucht, um zu zeigen, dass die Wissensmaße in Studie 1 und Studie 2 aus messtheoretischer Sicht sehr ähnlich funktionierten. Um diesen Befund weiter zu belegen, wurden die internen Konsistenzen der einzelnen Wissensmaße in Form von Cronbachs Alpha und die Stabilitäten berechnet. Da aufgrund der Treatmenterteilung in den Sitzungen 2 und 3 nicht mehr alle Messungen voneinander unabhängig sind, werden hier nur die Koeffizienten für den ersten Messzeitpunkt berichtet. Jedoch wurden die Werte für alle drei Messzeitpunkte berechnet und sind in Anhang H.9 einsehbar.

Tabelle 49 gibt die Alpha-Koeffizienten zusammen mit der Anzahl der jeweils eingegangenen Items wieder. Wie in Studie 1 sind die internen Konsistenzen der Maße prozeduralen Wissens im Schnitt höher als die der Maße konzeptuellen Wissens. Wiederum weisen die beiden Maße Evaluation und Erklärung die geringsten Werte auf.

Tabelle 49: Aufgabenanzahlen und Cronbachs Alphas der Wissensmaße an Mzp 1.

	<i>n</i>	α
Maße konzeptuellen Wissens		
Evaluation	8	,27
Übersetzung	20	,72
Größenvergleich	20	,73
Erklärung	4	,49
Maße prozeduralen Wissens		
Problemlösekorrektheit	20	,77
Problemlösedauer	15	,93
Zugriffsasymmetrie I	20	,76
Zugriffsasymmetrie II	20	,63
Zugriffsasymmetrie	-	-
Dual task-Kosten I	40	,92
Dual task-Kosten II	40	,93
Dual task-Kosten	-	-

Die Interrater-Reliabilitäten für das Maß Erklärung zeigt Tabelle 50. Sie zeigen für alle Aufgaben und insbesondere die Summenscores gute Urteilerübereinstimmungen an.

Tabelle 50: Interrater-Reliabilitäten für das Maß Erklärung.

	Mzp 1	Mzp 2	Mzp 3
Aufgabe 1 ^a	,72	,79	,68
Aufgabe 2 ^a	,73	,75	,81
Aufgabe 3 ^a	,78	,70	,80
Aufgabe 4 ^a	,83	,89	,77
Summenscore ^b	,90	,85	,85

Anmerkungen. ^a Kendalls τ ; ^b Pearsons r .

10.3.7 Einfluss des distance effects auf den Lernerfolg

Der Distance effect wurde berechnet wie in Studie 1. Die entsprechenden Rohdaten, Syntaxen und Outputdateien befinden sich in Anhang G. Das Datenerhebungsprogramm befindet sich in Anhang C.

Der distance effect trat in Sitzung 1 bei 61% und in Sitzung 3 bei 51% der Teilnehmer in statistisch signifikanter Weise ($ps < .05$) auf. In Sitzung 1 erklärte er maximal 46% und durchschnittlich 13% der Varianz der Lösungszeiten. In Sitzung 3 erklärte er maximal 52% und durchschnittlich 11% der Varianz.

Die Werte der Variablen Distance effect gemessen in Sitzung 1 und Sitzung 2 korrelieren nicht signifikant miteinander ($r = ,16$; $N = 82$; $p = ,158$) und lassen darauf schließen, dass entweder die Retestreliabilität oder die intrapersonale Stabilität der Variable sehr gering ist.

Um zu prüfen, ob die interindividuellen Unterschiede im distance effect den Lernerfolg beeinflussen, wurde zunächst pro Person die mittlere Stärke des Effekts über beide Messzeitpunkte hinweg berechnet. Dann wurden die acht Varianz- und Kovarianzanalysen zum Einfluss der Treatments auf die Wissensscores aus Abschnitt 10.3.4 erneut gerechnet. Dieses Mal jedoch mit dem mittleren Distance effect als zusätzlicher Kovariate in jeder Analyse.

Die Syntaxen und Ergebnisse können in Anhang H.10 eingesehen werden. Der Distance effect hatte in keiner der Analysen einen signifikanten Einfluss. Da dieses Ergebnis angesichts der Befunde aus Studie 1 (s. Abschnitt 9.4.2) und angesichts der extrem niedrigen Retestreliabilität in Studie 2 weder überraschend noch interessant ist, werden die Koeffizienten der Analysen hier nicht aufgelistet.

10.4 Diskussion

10.4.1 Separate Vermittelbarkeit der Wissensarten

Die Studie wurde durchgeführt, um zwei Punkte zu untersuchen: die separate Vermittelbarkeit von konzeptuellem und prozeduralem Wissen und die kausalen Relationen zwischen den Wissensarten.

Eine Gruppe von Kindern erhielt verbale Erklärungen genereller Eigenschaften von Dezimalbrüchen, an die sich eine Reflektionsphase anschloss. Eine andere Gruppe lernte Lösungsbeispiele auswendig und löste dann Übungsaufgaben ohne Feedback. Eine dritte Gruppe, die Kontrollgruppe, las und bearbeitete einen Lückentext, der keine Informationen über Dezimalbrüche beinhaltet.

Entgegen den Hypothesen erwarb die Gruppe, die die verbalen Erklärungen erhielt, kein konzeptuelles Wissen, was sich darin zeigt, dass ihre entsprechenden Leistungen nach der Intervention nicht über denen der Kontrollgruppe liegen.

In Übereinstimmung mit den Hypothesen gewann die Gruppe, die die Lösungsbeispiele auswendig lernte, mehr prozedurales Wissen als die anderen beiden Gruppen. Da die drei Gruppen sich danach nicht überzufällig in ihrem konzeptuellen Wissen unterschieden, ist das Auswendiglernen von Lösungsbeispielen mit anschließender Übung ohne Feedback also geeignet, um prozedurales Wissen ohne konzeptuelles zu vermitteln.

Die Vermittlung prozeduralen Wissens

Letzterer Befund ist aus zwei Gründen wichtig. Zum einen belegt er, dass die Unterscheidung von konzeptuellem und prozeduralem Wissen sinnvoll ist. Ohne die Unterscheidung kann nur schwer erklärt werden, warum die Maße prozeduralen Wissens systematisch anders auf das Treatment reagierten als die Maße konzeptuellen Wissens. Zum anderen stellt der Besitz prozeduralen Wissens ohne das dazugehörige konzeptuelle Wissen kein optimales Unterrichtsziel dar (vgl. Abschnitt 5.5.2). Die Befunde zeigen, dass dieser Zustand tatsächlich auftreten kann, was angesichts der hohen Korrelationen zwischen den Wissensarten in Studie 1 keine Selbstverständlichkeit darstellt. Bei der Unterrichtsgestaltung ist also darauf zu achten, diesen Zustand zu vermeiden.

Die Vermittlung konzeptuellen Wissens

Die Intervention, die den Hypothesen zufolge konzeptuelles Wissen hätte vermitteln sollen, dauerte im Gruppendurchschnitt signifikant länger (21 statt 14 Minuten) als die Intervention, die prozedurales Wissen vermitteln sollte. Das war nicht beabsichtigt und wurde in der Vorstudie so nicht beobachtet. Eine Ursache für die Unterschiede zwischen Vor- und Hauptstudie könnte sein, dass Kinder mit wenig Vorwissen in der Vorstudie sich wegen der durchgängigen individuellen Betreuung durch den Versuchsleiter bemühten über alle Aufgaben nachzudenken, während Kinder mit wenig Vorwissen sich in der Hauptstudie, die als Gruppenuntersuchung durchgeführt wurde, öfter einfach durch die Aufgaben durchklickten, ohne überhaupt nach den richtigen Antworten zu suchen. Dass Kinder mit geringem Vorwissen diese Bearbeitungsstrategie benutzten, ist aus Studie 1 bekannt und zeigt sich auch in Studie 2 an den unerwarteterweise positiven Korrelationen zwischen der Lösungsrate und Lösungsdauer bei den Zahlenstrahlaufgaben in den Wissenstests.

Angesichts der Richtung der Gruppenunterschiede in den Wissenszuwächsen behindern die unterschiedlichen Interventionsdauern die Interpretation jedoch nicht, sondern fördern sie sogar. Bemerkenswerterweise hat nämlich die Gruppe, die konzeptuelles Wissen erwerben sollte, das aber nicht tat, eine längere Intervention erhalten, als die Gruppe, die prozedurales Wissen erhalten sollte und auch tatsächlich erwarb. Dass die zeitlich kürzere Intervention effektiv wirkte, während die zeitlich längere ohne messbare Auswirkungen blieb, deutet darauf hin, dass prozedurales Wissen wesentlich leichter zu vermitteln und zu erwerben ist als konzeptuelles.

Dieser Befund steht im Einklang damit, dass die interindividuellen Unterschiede im konzeptuellen Wissen stabiler waren als die interindividuellen Unterschiede im prozeduralen Wissen. Das zeigt sich in den Analysen dadurch, dass das konzeptuelle Vorwissen einen größeren Einfluss auf das spätere konzeptuelle Wissen hat als das prozedurale Vorwissen auf das spätere prozedurale Wissen, was sowohl für die erste als auch für die zweite Lernphase gilt.

Konzeptuelles Wissen ist also gleich auf zwei verschiedene Weisen stabiler als prozedurales: die interindividuellen Unterschiede verändern sich über die Zeit hinweg weniger und der Gruppenmittelwert ist schwieriger durch Interventionen zu beeinflussen als es beim prozeduralen Wissen der Fall ist.

Eine mögliche Alternativerklärung dafür, weshalb die Gruppe, die konzeptuelles Wissen erwerben sollte, nach der ersten Lernphase nicht mehr konzeptuelles Wissen zeigte als die Kontrollgruppe, ist schwer zu sehen. Beide Gruppen unterscheiden sich weder in der zeitlichen Länge des jeweils erteilten Treatments noch in wichtigen Personeneigenschaften, wie Alter, Vorwissen oder Schulnoten. Eine Konfundierung des Treatments mit Störvariablen kann mit großer Sicherheit ausgeschlossen werden.

Dass das gewählte Treatment grundsätzlich inhaltlich ungeeignet zur Vermittlung konzeptuellen Wissens ist, kann durch eine einfache Überlegung ausgeschlossen werden. In Unterkapitel 6.2 wurde berichtet, dass die häufigsten Misskonzepte in Bezug auf Dezimalbrüche das Ganze-Zahlen-Misskonzept und das Bruchstrich-Misskonzept sind. Der folgende Textausschnitt aus der Intervention macht die Unangemessenheit beider Misskonzepte deutlich:

Du siehst jetzt die Zahl 2,5 auf dem Bildschirm. Die 2 steht vor dem Komma, die 5 hinter dem Komma. Die Ziffern vor dem Komma sagen Dir immer, wie viele ganze Einheiten es von etwas gibt. Wenn Du also 2,5 Schokoriegel hast, so hast Du mindestens 2 ganze Schokoriegel. Die Ziffern nach dem Komma sagen Dir immer, wie viele Teile von etwas es gibt. Wenn Du 2,5 Schokoriegel hast, hast Du also nicht nur 2 ganze Schokoriegel, sondern auch noch einen Teil von einem Schokoriegel

dazu. Aber wie groß ist dieser Teil? Die erste Stelle nach dem Komma gibt immer an, wie viele Zehntel man hat. Zusätzlich zu den 2 ganzen Riegeln hast Du also noch 5 Zehntel von einem ganzen Riegel. Um Dir zu überlegen wie viel 5 Zehntel sind, kannst Du Dir vorstellen, dass Du einen Schokoriegel in zehn gleich große Teile schneidest und 5 von diesen Teilen nimmst. Das sind 5 Zehntel. (vgl. das Programm in Anhang B.2 und B.3 sowie den Interventionstext in Anhang F.1.2)

Der Inhalt dieses Textauszugs wird in der Intervention noch weiter ausgeführt und erklärt. Auf der inhaltslogischen Ebene widerlegt er beide Misskonzepte und sollte, so er verstanden und erinnert wird, direkt zu deutlichen Leistungssteigerungen bei der Bearbeitung der konzeptuellen Tests führen.

Allerdings reichte ein einmaliges Hören der Erklärung vielleicht nicht aus, um ihren Inhalt wirklich zu verstehen und fest im Gedächtnis zu verankern. In Abschnitt 5.4.1 wurden Kriterien vorgestellt, die eine Intervention erfüllen muss, um conceptual change fördern zu können. Die hier verwendete Intervention erfüllt diese Kriterien nicht vollständig. Beispielsweise konfrontiert sie die Teilnehmer nicht mit den suboptimalen Eigenschaften ihrer Misskonzepte und demonstriert ihnen, dass die zu vermittelnden Konzepte besser funktionieren, sondern stellt die korrekten Konzepte vor, ohne mögliche Misskonzepte überhaupt zu thematisieren.

Die wahrscheinlichste Erklärung für das Fehlschlagen der Vermittlung konzeptuellen Wissens ist also die generelle Stabilität konzeptuellen Wissens in Verbindung mit der zu geringen Treatmentintensität aufgrund einer suboptimalen Interventionsgestaltung.

Explorative Betrachtung der einzelnen Wissensmaße

Das Bild der Lernprozesse zwischen dem ersten und dem zweiten Messzeitpunkt verkompliziert sich leicht, wenn man statt der Wissensscores die einzelnen Wissensmaße betrachtet. Hier zeigen drei der Maße konzeptuellen Wissens und drei der Maße prozeduralen Wissens vergleichbare Wissenszuwächse in allen drei Gruppen. Da die Wissenszuwächse in der Kontrollgruppe lediglich Retesteffekte widerspiegeln, haben die beiden anderen Treatments offensichtlich zu keinen zusätzlichen Wissenszuwächsen geführt.

Ein Maß konzeptuellen Wissens, die Erklärung, und ein Maß prozeduralen Wissens, die Problemlösekorrektheit, verhielten sich jedoch genau hypothesenkonform: Die Erklärung wurde stärker durch die Intervention zur Vermittlung konzeptuellen Wissens beeinflusst als durch die anderen beiden Interventionen. Die Problemlösekorrektheit wurde stärker durch die Intervention zur Vermittlung prozeduralen Wissens beeinflusst als durch die anderen beiden Interventionen. Während der Effekt bei der Problemlösekorrektheit

offensichtlich so stark war, dass er zu hypothesenkonformen Befunden bezüglich des prozeduralen Wissensscores führte, war er bei der Erklärung etwas schwächer und zeigte sich im konzeptuellen Wissensscore nicht mehr.

Dass die Maße Erklärung und Problemlösekorrektheit am stärksten auf die Interventionen reagierten, könnte daran liegen, dass sie am besten auf das jeweilige Treatment abgestimmt waren. In der Reflektionsphase der Intervention zur Vermittlung konzeptuellen Wissens wurde das Erklären geübt und in der Übungsphase der Intervention zur Vermittlung prozeduralen Wissens das Lösen von Zahlenstrahlaufgaben.

Beide Übungsaktivitäten beinhalteten jedoch keine Feedbacks. Es erfolgte somit kein *teaching to the test*, also kein Training, das lediglich auf eine optimale Testbewältigung ohne Erwerb nachhaltigen Wissens abzielte. In der Intervention und im Test waren Erklärungen unterschiedlichen Inhalts gefordert und die in der Intervention und im Test präsentierten Dezimalbrüche bei den Zahlenstrahlaufgaben stimmten nicht überein. Die beiden Maße Erklärung und Problemlösekorrektheit zeigen somit echte und hypothesenkonforme Wissenszuwächse an. Dass diese sich nicht auch auf den anderen sechs Maßen, die den Interventionsinhalten etwas unähnlicher waren, zeigten, lässt sich gut durch den Schluss erklären, der schon oben gezogen wurde: Die Treatments wurden möglicherweise nicht intensiv genug vermittelt. Sie erzielten daher schwache und instabile Effekte, die auf Maße, die Transfer- oder Umsetzungsleistungen erfordern, nicht mehr durchschlagen. Wäre die konzeptuelle Erklärung in der Intervention mehrfach wiederholt oder besser mit der Reflexionsphase verknüpft gewesen und wären die Anzahl der präsentierten Lösungsbeispiele größer oder die Übungsphase länger gewesen, so hätten sich die erwarteten Effekte eventuell auf allen Maßen gezeigt. Das kann im Rahmen der vorliegenden Daten aber natürlich nicht nachgewiesen werden.

Vermittlung der Wissensarten: Fazit

Als Ergebnis kann festgehalten werden, dass es nicht gelungen ist konzeptuelles Wissen zu vermitteln. Die Vermittlung prozeduralen Wissens unabhängig von konzeptuellem gelang hingegen, zeigte sich jedoch vor allem in den teststarken Analysen der zusammengesetzten Wissensscores, nicht aber in allen Analysen der einzelnen Wissensmaße.

10.4.2 Kausale Interrelationen

Ein Nachteil des Designs von Studie 2 besteht darin, dass die Forschungsfragen nach den kausalen Interrelationen zwischen den Wissensarten nur dann untersucht werden können, wenn die ebenfalls untersuchte Forschungsfrage nach der separaten Vermittelbarkeit von konzeptuellem und prozeduralem Wissen zuvor positiv beantwortet wurde.

Da sich prozedurales Wissen unabhängig von konzeptuellem vermitteln ließ, konnte untersucht werden, ob dieses in der ersten Lernphase vermittelte prozedurale Wissen in der zweiten Lernphase zu Zuwächsen an konzeptuellem Wissen führte. Das war nicht der Fall. Die Gruppe, die das prozedurale Wissen erhalten hatte, besaß am dritten Messzeitpunkt nicht mehr konzeptuelles Wissen als die Kontrollgruppe. Eine mögliche Erklärung dafür ist, dass prozedurales Wissen nicht kausal zu Zuwächsen an konzeptuellem Wissen führt und dass somit der concepts-first view die beobachteten Lernvorgänge adäquat beschreibt. Eine mögliche Alternativerklärung ist, dass die Gruppenunterschiede im prozeduralen Wissen am zweiten Messzeitpunkt zu gering und instabil waren, um nachfolgend nachweisbare Auswirkungen zu haben. Schließlich konnten die Gruppenunterschiede im prozeduralen Wissen am zweiten Messzeitpunkt nur bei einem der vier Maße prozeduralen Wissens nachgewiesen werden.

Da alle drei Gruppen zum zweiten Messzeitpunkt gleich viel konzeptuelles Wissen besaßen, konnte leider nicht untersucht werden, ob in der ersten Lernphase produzierte Unterschiede im konzeptuellen Wissen in der zweiten Lernphase zu Zuwächsen an prozeduralem Wissen führten.

Da die Treatmentgruppen einander in Bezug auf alle gemessenen Personeneigenschaften sehr ähnlich waren, scheiden Samplingeffekte als Erklärung für die unerwarteten Probleme bei der Hypothesentestung aus. Die zeitlichen Längen der Interventionen unterscheiden sich um einige Minuten. Plausible Gründe zur Annahme einer dadurch verursachten Ergebnisverzerrung sind jedoch nicht zu sehen. Die Stichprobengröße war mit knapp 30 Personen pro Gruppe nicht zum Nachweis schwacher Effekte geeignet, entspricht jedoch der Stichprobengröße in vergleichbaren Trainingsstudien und kann sinnvollerweise nicht als Hauptgrund für die Ambiguität der Resultate betrachtet werden. Angesichts der Wichtigkeit, die konzeptuellem und prozeduralem Wissen in theoretischen Diskussionen zugesprochen wird, ist die in dieser Studie offenbar gewordene Fragilität der zu beobachtenden Effekte überraschend.

Jedoch wurden, wie in Abschnitt 10.2.2 beschrieben, die Treatments absichtlich mit abgeschwächten Intensitäten erteilt, um Deckeneffekte zu vermeiden und weil der Erfolg des „Fang-das-Monster“-Spiels in Studie 1 nahelegte, dass die Vermittlung von stabilem konzeptuellen und prozeduralen Wissen durch kleine, einfache Interventionen relativ leicht möglich sei. Diese Einschätzung war in ihrer Allgemeinheit offenbar falsch.

Es empfiehlt sich eine Wiederholung von Studie 2, bei der das Treatment zur Vermittlung konzeptuellen Wissens stärker den Kriterien von conceptual change-Ansätzen

entspricht, und bei der das Treatment zur Vermittlung prozeduralen Wissens intensiviert wird, um stärkere Effekte hervorzurufen. Falls eine breite und stabile Wissensvermittlung in der ersten Lernphase gelänge, könnten in der zweiten Lernphase die Kausalhypothesen zum Iterativen Modell einer gültigen Prüfung unterzogen werden.

Tabelle 51 gibt einen Überblick über die erwarteten Prozesse und Befunde in Studie 2 (normal gedruckt) und die empirisch gefundenen (fettgedruckt).

10.4.3 Weitere Befunde

Die internen Konsistenzen der einzelnen Wissensmaße sind weitgehend zufriedenstellend, lediglich für die Maße Evaluation und Erklärung niedrig und im Schnitt

Tabelle 51: Überblick über die erwarteten und beobachteten Befunde in Studie 2.

	Test 1	Lernphase 1	Test 2	Lernphase 2	Test 3
Gruppe TK	Gruppe besitzt etwa gleich viel konzeptuelles und prozedurales Wissen wie Gruppen TP und TX	Erwerb konzeptuellen Wissens	Gruppe besitzt mehr konzeptuelles Wissen als Gruppe TX; Gruppe besitzt so viel prozedurales Wissen wie Gruppe TX	Neuerwerb konzeptuellen und prozeduralen Wissens; Umsetzung konzeptuellen Wissens in prozedurales	Gruppe besitzt mehr prozedurales Wissen als Gruppe TX
Gruppe TP	Gruppe besitzt etwa gleich viel konzeptuelles und prozedurales Wissen wie Gruppen TK und TX	Erwerb prozeduralen Wissens	Gruppe besitzt mehr prozedurales Wissen als Gruppe TX; Gruppe besitzt so viel konzeptuelles Wissen wie Gruppe TX	Neuerwerb konzeptuellen und prozeduralen Wissens; Umsetzung prozeduralen Wissens in konzeptuelles	Gruppe besitzt mehr konzeptuelles Wissen als Gruppe TX
Gruppe TX	–	Kein Erwerb testrelevanten Wissens	–	Neuerwerb konzeptuellen und prozeduralen Wissens	–

Anmerkung. Empirisch bestätigte Erwartungen fettgedruckt.

für die Maße prozeduralen Wissens etwas höher als für die Maße konzeptuellen Wissens. Die Interrater-Reliabilitäten des Maßes Erklärung sind gut.

Die explorative Faktoranalyse für die acht Wissensmaße am ersten Messzeitpunkt indiziert, dass die acht Maße tatsächlich zwei Faktoren bilden und rechtfertigen das nachfolgende Vorgehen, die acht Maße zu zwei Wissensscores zusammenzufassen. Die Faktoranalyse zeigt aber auch zwei Probleme der Wissensmessungen an: Erstens brauchten die Kinder entgegen der Erwartungen umso mehr Zeit zur Lösung der Zahlenstrahlprobleme, je mehr prozedurales Wissen (gemessen an den anderen drei Maßen) sie besaßen. Zweitens lud das Maß Problemlösekorrektheit, das eigentlich prozedurales Wissen messen sollte, nicht hoch auf dem Faktor, auf dem die übrigen drei

Maße prozeduralen Wissens hoch luden, sondern auf dem Faktor, auf dem die vier Maße konzeptuellen Wissens hoch luden.

Die Ergebnisse der Faktoranalyse sollten vorsichtig interpretiert werden, da es sich bei ihr zum einen um ein exploratives Verfahren handelt und sie zum anderen voneinander unabhängige Faktoren extrahiert. Mit großer Sicherheit sind konzeptuelles und prozedurales Wissen in Wirklichkeit jedoch interkorreliert, so dass das Verfahren nicht optimal geeignet ist und lediglich explorative Hintergrundinformationen liefern kann.

Es wurden relativ hohe Interkorrelationen zwischen den Scores konzeptuellen und prozeduralen Wissens pro Messzeitpunkt und Treatmentgruppe gefunden. Die Treatments wirkten sich nicht nachweisbar auf diese Korrelationen aus, was darauf hindeutet, dass die drei Treatments nicht zu unterschiedlich starken Integrationen der Wissensarten im Sinne Baroodys führten. Die Befunde geben also keinen Aufschluss darüber, welche Unterrichtstechnik adaptive Expertise (vgl. Abschnitt 5.5.2) behindern oder fördern kann.

Die verhältnismäßig hohen Interkorrelationen zwischen den Faktoren können teilweise dadurch erklärt werden, dass das Maß Problemlösekorrektheit hoch mit den konzeptuellen Maßen korreliert ist, jedoch in den Score prozeduralen Wissens einfluss.

Der distance effect als Maß dafür, zu welchem Ausmaß Kinder Zahlen auf einem mentalen Zahlenstrahl repräsentieren, hatte keinen Einfluss auf die Wissenszuwächse und interagierte auch nicht mit den Treatments. Die zweimalige Erhebung der Stärke des Effekts pro Person ermöglichte die Aufdeckung einer wichtigen Ursache dafür: Die Messungen korrelieren sehr niedrig und nicht signifikant miteinander. Der Grund hierfür ist unklar. Entweder handelt es sich beim distance effect um kein stabiles Personenmerkmal oder die Retestreliabilität war sehr niedrig, zum Beispiel weil Studienteilnehmer die Instruktionen nicht vollständig befolgten oder die Zeitmessung sehr ungenau war. Die letzten beiden Gründe sind jedoch unwahrscheinlich, da der distance effect selbst an beiden Messzeitpunkten in signifikanter Weise bei der Mehrzahl der Teilnehmer nachgewiesen werden konnte.

10.5 Zusammenfassung

Es wurden dieselben Wissensmaße wie in Studie 1 benutzt, um in einer experimentellen Trainingsstudie mit 84 Fünftklässlern zu überprüfen, ob sich konzeptuelles und prozedurales Wissen in einer ersten Lernphase unabhängig voneinander vermitteln lassen und ob das vermittelte Wissen in einer zweiten, für alle Gruppen gleichen Lernphase in die jeweils andere Wissensart umgesetzt wird. Es gelang, in der ersten Lernphase prozedurales Wissen unabhängig von konzeptuellem zu vermitteln. Konzeptuelles Wissen zu vermitteln

gelang in dieser Phase nicht. Es wurde keine Evidenz dafür gefunden, dass die Zuwächse an prozeduralem Wissen sich in der zweiten Lernphase kausal auf die Zuwächse an konzeptuellem Wissen auswirkten. Ob konzeptuelles Wissen sich kausal auf prozedurales Wissen auswirkt, konnte nicht untersucht werden. Eine plausible Ursache für die wenigen und schwachen gefundenen Effekte sind zu geringe Treatmentintensitäten.

11 Diskussion

11.1 Diskussion der Befunde zu den Forschungsfragen

Studie 1 und Studie 2 lassen sich einordnen in die große Reihe von Studien, die Beziehungen zwischen konzeptuellem und prozeduralem Wissen im Bereich des Mathematiklernens untersuchen. Im Gegensatz zu älteren Arbeiten wird in den vorliegenden beiden Studien dem Problem der Messung der Wissensarten große Aufmerksamkeit geschenkt. Im Folgenden wird diskutiert, was die gewonnenen Befunde bezüglich der fünf in Kapitel 8 vorgestellten Forschungsfragen aussagen. Dazu werden die Forschungsfragen einzeln betrachtet.

Forschungsfrage 1: Kovariieren verschiedene Maße konzeptuellen beziehungsweise prozeduralen Wissens miteinander, so dass ihnen konvergente Validität zugesprochen werden kann?

Im Theorieteil wurde herausgearbeitet, dass die Postulierung mindestens einer Wissensart dann produktiv ist, wenn sich durch sie konsistente Verhaltensmuster von Personen über verschiedene Maße hinweg erklären lassen (vgl. Abschnitt 4.3.6). Dieser Punkt wurde in Studie 1 untersucht, indem konzeptuelles und prozedurales Wissen jeweils durch vier verschiedene Maße erhoben und als unterliegende latente Faktoren modelliert wurden. Dieses Vorgehen ist inhaltlich sinnvoll, weil sich Wissensarten im Verhalten nie isoliert und ungefiltert zeigen, sondern weil sie ihm höchstens latent zugrunde liegen. Das Vorgehen ist auch methodisch sinnvoll, weil die Analysetechnik für Strukturgleichungsmodelle effiziente Werkzeuge zur Untersuchung latenter Variablen bereitstellt.

Die Analysen ergaben, dass die Maße einer Wissensart pro Messzeitpunkt tatsächlich so stark interkorreliert sind, dass jeweils die Extraktion eines latenten Faktors gerechtfertigt ist. Wie schon in der Diskussion von Studie 1 angemerkt wurde, ist das ein bemerkenswerter Befund, weil er belegt, dass das Dezimalbruchwissen der Kinder zu einem großen Teil nicht an Aufgabenanforderungen, Aufgabenpräsentation oder Antwortformate gebunden ist. Die Faktoren spiegeln in diesem Sinne abstraktes Wissen wider.

Die Befunde zu den konvergenten Validitäten sind über die drei Messzeitpunkte hinweg stabil. Selbst der Beitrag, den jeweils ein Einzelmaß zu einem Faktor leistet, unterscheidet sich zwischen den ersten beiden Messzeitpunkten nicht und zwischen dem zweiten und dem dritten nur geringfügig, so dass weitgehend metrische Messinvarianz

vorliegt. Die explorativen Faktoranalysen von Studie 1 und Studie 2 lassen vermuten, dass die konvergenten Validitäten der Maße auch in Studie 2 gegeben waren. Dort konnte sie jedoch nicht konfirmatorisch analysiert werden, weil die Stichprobe aus ökonomischen Gründen auf einen Umfang begrenzt war, der solche Analysen nicht zulässt.

Rittle-Johnson et al. (2001) äußerten die Vermutung, dass die Maße prozeduralen Wissens vor dem „Fang-das-Monster“-Spiel andere Konstrukte messen würden als danach, weil die Zahlenstrahlaufgaben, die zur Messung prozeduralen Wissens benutzt wurden, lediglich nach dem Training, nicht aber davor Routineaufgaben darstellen. Die Analysen von Studie 1 bestätigen diese Vermutung nicht. Die Befunde zur metrischen Messinvarianz zeigen, dass sich die Faktorstruktur der Indikatoren weder in qualitativer noch in quantitativer Weise über die Messzeitpunkte hinweg substantiell verändert.

Die Messinvarianz der Maße konzeptuellen Wissens ist darüberhinaus ein Indiz gegen die Richtigkeit der Annahme diSessas (z.B. diSessa, Gillespie, & Esterly, 2004), dass konzeptuelles Wissen von Kindern in einem neuen Inhaltsgebiet stets fragmentiert ist und erst durch Lernprozesse vereinheitlicht wird. Denn in diesem Fall hätten die Korrelationen zwischen den einzelnen Maßen, die ja teilweise unterschiedliches Lösungswissen voraussetzen, über die Zeit hinweg ansteigen müssen, was nicht der Fall war.

Die Befunde sind leichter mit der Idee vereinbar, dass das konzeptuelle Wissen von Novizen aus kohärenten naiven Theorien besteht, die durch Lernprozesse weiterentwickelt werden (z.B. Vosniadou & Brewer, 1992). Die subjektiven Theorien der Kinder wurden in Studie 1 und Studie 2 nicht gezielt gemessen, so dass die vorliegenden Befunde nicht sehr aussagekräftig sind. Zukünftige Studien sollten sich diesem Thema stärker widmen, weil die qualitativ unterschiedlichen subjektiven Theorien der Kinder hilfreich sein könnten, um interindividuelle Unterschiede auf der quantitativen Ebene der Wissensmaße zu erklären.

Aus methodischer Perspektive könnte auch die Untersuchung weiterer Typen von Messinvarianz, beispielsweise der skalaren Invarianz im Sinne Cheung und Rensvolds (2002), interessant sein, weil diese direkte Vergleiche der Faktorwerte über die Messzeitpunkte hinweg erlauben würde. Dann könnte zum Beispiel getestet werden, ob die Faktorwerte über die Zeit hinweg signifikant zunehmen.

Ein einschränkender Befund bezüglich der Validität der Einzelmaße ist, dass die Probanden Routineprobleme mit zunehmendem prozeduralem Wissen nicht, wie erwartet, schneller, sondern langsamer lösten. Dieser Befund war über Studie 1 und 2 hinweg stabil. Die hohe interne Konsistenz des Maßes Problemlösedauer zeigt, dass kein zufälliges Verhalten der Probanden oder Messfehler verantwortlich sein können. Das Phänomen kann hingegen durch die Annahme erklärt werden, dass Probanden mit hinreichend großem

Wissen mehr Zeit zur Lösungsfindung und akkuraten Lösungseingabe brauchen als Probanden, die aufgrund eines Mangels an Lösungswissen nicht viel Zeit sinnvoll mit Lösungsfindung und -eingabe verbringen können. Die Beobachtungen der Versuchsleiter bestätigen diese Vermutung.

Weil die Problemlösedauer eine hohe interne Konsistenz aufweist, das beobachtete Phänomen über alle Messzeitpunkte von Studie 1 (und Studie 2) hinweg stabil auftrat und sich gut erklären lässt, wurde das Maß nicht von den nachfolgenden Analysen ausgeschlossen und auch in Studie 2 benutzt. Diese Entscheidung hat auch negative Implikationen, weil der Ausschluss nicht vollständig valider Maße von weiteren Analysen gewöhnlich die Interpretation der Ergebnisse erleichtert. Jedoch ist zu bedenken, dass die konvergente Validität aller vier Maße prozeduralen Wissens, also auch der Problemlösedauer, nachgewiesen wurde und dass die Parameterschätzung der Strukturgleichungsmodelle umso besser konvergiert und umso validere Lösungen liefert, je mehr Indikatoren pro Faktor zur Verfügung stehen. Die weitere Nutzung des Maßes in den Analysen von Studie 1 und Studie 2 stellt somit das kleinere von zwei Übeln dar. Weitere Untersuchungen sollten aber, soweit möglich, auf das Maß verzichten und ökologisch validere Alternativen nutzen.

Forschungsfrage 2: Variieren Maße konzeptuellen und prozeduralen Wissens (zumindest teilweise) unabhängig voneinander, so dass ihnen divergente Validität zugesprochen werden kann?

Wenn sich zwei hypothetische Konstrukte unter keinen Umständen teilweise unabhängig voneinander messen lassen, sollte die theoretische Unterscheidung zwischen ihnen nicht aufrechterhalten werden. Denn es gibt dann keine empirische Beobachtung, zu deren Erklärung die Unterscheidung der Konstrukte notwendig wäre (vgl. Abschnitt 4.3.6).

Die Befunde aus Studie 1 und Studie 2 sind in dieser Hinsicht nicht eindeutig. Die in Studie 1 gewonnenen Daten sind etwa gleich gut mit der Annahme, dass zwei Wissensarten gemessen wurden, und der Annahme, dass nur eine Wissensart gemessen wurde, vereinbar. Modelliert man konzeptuelles und prozedurales Wissen pro Messzeitpunkt als zwei latente Faktoren, so sind die beiden jeweils sehr hoch, jedoch nicht perfekt interkorreliert. Wegen der großen theoretischen Wichtigkeit der Unterscheidung der Wissensarten und um die Analyse der Kausalbeziehungen zwischen ihnen zu ermöglichen, wurden konzeptuelles und prozedurales Wissen trotzdem in Studie 1 und Studie 2 als zwei getrennte Konstrukte modelliert.

Es stellt sich die Frage, wie die hohen Interkorrelationen zwischen den beiden Konstrukten zu interpretieren sind. In Studie 1 wurde gezeigt, dass es sich nicht um Scheinkorrelationen handelt, die lediglich durch den Einfluss des Alters oder der Intelligenz hervorgerufen werden. In Studie 2 wurde gezeigt, dass sich auch die drei gewählten Treatments nicht auf die Höhe der Interkorrelationen auswirken.

Dass die Interkorrelationen in Studie 2 etwas niedriger waren als in Studie 1, lässt sich leicht dadurch erklären, dass in Studie 1 die Wissensarten als latente Faktoren modelliert wurden, in deren Bestimmung die maßspezifischen Fehlervarianzen nicht miteinflussen. In Studie 2 wurden die Wissensarten als Summenscores modelliert, in die zwar reliablere Maße stärker gewichtet einfließen als unreliablere Maße, in die jedoch die maßspezifischen Fehlervarianzen trotzdem miteingingen. In Studie 2 wurden die Wissensarten also weniger reliabel gemessen als in Studie 1. Diese zusätzliche Fehlervarianz erklärt die reduzierten Interkorrelationen zwischen den Wissensarten in Studie 2. Wäre der Einfluss des Treatments auf die Wissensmaße in multivariaten Varianzanalysen untersucht worden, hätten die Maße nicht mit ihren Reliabilitäten gewichtet werden können und der Anteil der Fehlervarianz an der Gesamtvarianz wäre noch größer gewesen. Auf dieses Vorgehen wurde darum verzichtet.

Es sind verschiedene Untersuchungsbedingungen vorstellbar, unter denen der quantitative Zusammenhang zwischen konzeptuellem und prozeduralem Wissen geringer sein könnte als in Studie 1 und Studie 2 beobachtet. Ein wichtiger Punkt ist die zeitliche Auflösung der Messungen der Wissensarten. Falls konzeptuelles und prozedurales Wissen innerhalb von Sekunden oder wenigen Minuten ineinander umgesetzt werden können, so können Dissoziationen zwischen ihnen nur durch Tests aufgedeckt werden, die weniger Zeit in Anspruch nehmen. Die in Studie 1 und Studie 2 verwendeten Tests dauerten jedoch durchschnittlich 30 bis 45 Minuten. Kürzere Tests könnten hilfreich zur Untersuchung der Fragestellung sein, sind jedoch schwierig zu konstruieren, vor allem wenn die zeitliche Auflösung Sekunden oder Sekundenbruchteile betragen soll. Insbesondere die Modellierung der Wissensarten als latente Variablen wäre dann nicht mehr möglich, weil dazu ja mehrere Indikatoren pro Wissensart gemessen werden müssten, so dass der Zeitaufwand noch größer wäre als bei der Verwendung manifester Maße.

Eine andere Möglichkeit zur Reduzierung der Korrelation zwischen den Wissensarten könnte in der langfristigen Anwendung geeigneter Treatments bestehen. In Studie 2 wurde gezeigt, dass prozedurales Wissen unabhängig von konzeptuellem vermittelt werden kann. Jedoch war der Effekt so instabil, dass er nach einer weiteren Lernphase nicht mehr nachweisbar war. Hiebert und Wearne (1996) zeigten, dass konstruktivistischer Unterricht

sich positiv auf die Verknüpfung von Konzepten und Prozeduren auswirken kann, wobei der Effekt jedoch erst nach drei Unterrichtsjahren auftrat. Angesichts der Befunde von Hiebert und Wearne zur Wirkungsdauer ihres Treatments ist die Instabilität der Wissenszuwächse nach der 20-minütigen Intervention in Studie 2 nicht überraschend. Wenn Kinder ein Treatment wie in Studie 2 über einen längeren Zeitraum erhalten würden, könnten sich möglicherweise deutliche Dissoziationen zwischen konzeptuellem und prozeduralem Wissen entwickeln.

Fasst man die letzten beiden Punkte zusammen so lässt sich sagen, dass Dissoziationen zwischen konzeptuellem und prozeduralem Wissen umso wahrscheinlicher beobachtet werden können, je kürzer die Tests und je intensiver die Interventionen zwischen den Tests sind.

Eine weitere Variable, die die Korrelation zwischen konzeptuellem und prozeduralem Wissen moderieren könnte, ist die Komplexität des Lerngegenstands. Im Zuge seiner umfangreichen Untersuchungen zum expliziten und impliziten Lernen befand Reber (1989): „A rich and complex stimulus domain is a prerequisite for the occurrence of implicit learning. If the system in use is too simple, or if the code can be broken by conscious effort, then one will not see implicit processes” (S. 220). Die Unterscheidung von explizitem und implizitem Wissen ist nicht identisch, aber ähnlich der Unterscheidung von konzeptuellem und prozeduralem Wissen. Rebers Beobachtung kann also nicht direkt auf konzeptuelles und prozedurales Wissen übertragen werden. Sie kann und sollte aber doch Vermutungen und Untersuchungen dazu anregen, ob sich Dissoziationen zwischen konzeptuellem und prozeduralem Wissen stärker in komplexen als in einfach strukturierten Inhaltsgebieten zeigen.

Für Studie 1 und Studie 2 wurde ein relativ einfach strukturiertes Inhaltsgebiet gewählt, um die Interpretation der Ergebnisse zu vereinfachen. Das könnte die hohen Interkorrelationen der Wissensarten mitverursacht haben.

Bei Untersuchungen zu diesem Thema wird zu beachten sein, dass das konzeptuelle und das prozedurale Wissen in einem Inhaltsgebiet unterschiedlich komplex sein können. Die Regeln des Schachspiels sind zum Beispiel relativ leicht konzeptuell zu verstehen, während die erfolgreiche Anwendung dieser Regeln ausgesprochen umfangreiches prozedurales Wissen über Zugkombinationen, Kurzzeit-, Langzeit-, Angriffs- und Verteidigungsstrategien erfordern kann. Die Steuerung einer komplexen technischen Anlage, beispielsweise eines Kernkraftwerks, von einem elektronischen Überwachungsstand aus setzt hingegen ein umfangreiches konzeptuelles Verständnis des

zu steuernden Systems voraus, während die Bedienung der Steuerungssoftware im Normalfall prozedural sehr einfach sein kann. Es könnte interessant sein, empirisch zu überprüfen, wie sich diese unterschiedlichen Komplexitäten auf Prozesse des Erwerbs und der Ineinanderumsetzung von konzeptuellem und prozeduralem Wissen auswirken.

Damit gibt es mindestens drei Faktoren, die die hohen Interkorrelationen zwischen konzeptuellem und prozeduralem Wissen reduzieren könnten: die Testdauer, die Treatmentintensität und die konzeptuelle und prozedurale Komplexität der Untersuchungsdomäne. Erst nachdem diese hypothetischen Möglichkeiten zur Reduzierung der hohen Korrelationen zwischen konzeptuellem und prozeduralem Wissen erfolglos erprobt wurden kann sinnvollerweise darüber nachgedacht werden, ob die Unterscheidung zwischen den Wissensarten aufgegeben werden sollte, weil sie empirisch nicht belegt werden kann. Diesen Schluss aus den Ergebnissen von Studie 1 und Studie 2 zu ziehen, wäre verfrüht.

Aufgrund der in Studie 1 und Studie 2 gefundenen hohen Interkorrelationen zwischen den Wissensarten wiesen konzeptuelles und prozedurales Wissen dort keine substanziell unterschiedlichen Relationen zu Variablen wie Intelligenz oder Nutzung des mentalen Zahlenstrahls auf.

Forschungsfragen 3 und 4: Wirkt sich konzeptuelles Wissen kausal auf prozedurales aus? Wirkt sich prozedurales Wissen kausal auf konzeptuelles aus?

Das Iterative Modell postuliert im Gegensatz zu älteren Ansätzen bidirektionale kausale Beziehungen zwischen konzeptuellem und prozeduralem Wissen. Entgegen der Erwartungen konnten diese in Studie 1 und Studie 2 nicht eindeutig belegt werden. Die Befunde belegen jedoch auch nicht die Inkorrektheit der Hypothesen.

In Studie 1 wurde geprüft, ob das konzeptuelle Wissen vor der Intervention das prozedurale Wissen nach der Intervention voraussagt und das prozedurale Wissen vor der Intervention das konzeptuelle Wissen nach der Intervention. Trotz praktisch bedeutsamer Stärken beider Effekte war keiner von ihnen signifikant größer als null. Mögliche Ursachen dafür sind die zur Analyse von Strukturgleichungsmodellen relativ geringe Stichprobengröße von $N = 204$ und die niedrigen divergenten Validitäten der Wissensmaße, die zu hohen Interkorrelationen der Prädiktoren führten.

In Studie 2 sollte überprüft werden, ob in einer ersten Lernphase vermitteltes konzeptuelles Wissen sich in einer zweiten Lernphase kausal auf prozedurales auswirkt und ob in einer ersten Lernphase vermitteltes prozedurales Wissen sich in einer zweiten Lernphase kausal auf konzeptuelles auswirkt. Weil die Vermittlung konzeptuellen Wissens

in der ersten Lernphase nicht gelang, konnte die Umsetzung konzeptuellen Wissens in prozedurales nicht untersucht werden. Die Vermittlung prozeduralen Wissens in der ersten Lernphase gelang, führte jedoch zu keinen nachweisbaren Zuwächsen an konzeptuellem Wissen in der zweiten Lernphase.

Während der Nachweis der erwarteten Effekte in der zweiten Lernphase bidirektionale Kausalrelationen belegt hätte, kann aus der Abwesenheit der erwarteten Effekte nicht auf die Abwesenheit der bidirektionalen Relationen geschlossen werden. Die in der ersten Lernphase vermittelten Wissensunterschiede waren sehr gering. Bei Betrachtung der Einzelmaße zeigten sie sich lediglich bei einem Maß konzeptuellen und einem Maß prozeduralen Wissens. Dass sich am Ende der zweiten Lernphase keine Gruppenunterschiede zeigen, ist also vermutlich darauf zurückzuführen, dass schon zu Beginn der Phase nur schwache und inkonsistente Wissensunterschiede bestanden.

In gewisser Weise wurde der Nachweis kausaler Beziehungen zwischen den Wissensarten in Studie 1 und Studie 2 durch das gleiche Problem beeinträchtigt: Die Wissensarten hingen in beiden Fällen so eng zusammen, dass ihre individuellen Beiträge zu späteren Wissenszuwächsen nicht bestimmt werden konnten.

Im vorherigen Abschnitt wurden mögliche Moderatorvariablen diskutiert, deren Einflüsse die Interkorrelationen zwischen den Wissensarten senken könnten. Ähnliche Moderatorvariablen könnten auch die kausalen Relationen zwischen konzeptuellem und prozeduralem Wissen beeinflussen.

Die zeitliche Auflösung der Untersuchung spielt auch hier eine wichtige Rolle. Studie 1 und Studie 2 sind geeignet, um Prozesse nachzuweisen, die Stunden oder Tage dauern. Dieses Zeitfenster ist interessant, weil auch Schulstunden zu einem Thema jeweils 45 bis 90 Minuten dauern und über mehrere Tage verteilt sein können. Es könnte aber sein, dass die Kausalbeziehungen auf unterschiedlichen zeitlichen Ebenen unterschiedlich ausgeprägt sind. So fokussieren kognitionspsychologische Untersuchungen im Rahmen der ACT-Theorie vor allem auf die Ableitung prozeduralen Wissens aus deklarativem Wissen in einem Zeitraum von Sekundenbruchteilen bis Stunden. Entwicklungspsychologische Untersuchungen zu Karmiloff-Smiths Theorie der representational redescription fokussieren auf die Umsetzung impliziten prozeduralen Wissens in explizites konzeptuelles in einem Zeitraum von mehreren Stunden bis hin zu mehreren Jahren.

Die Existenzen beider postulierten Prozesse schließen sich gegenseitig nicht aus. Sie könnten durchaus nebeneinander bestehen und sich trotzdem je nach zeitlicher Auflösung einer Untersuchung nur einzeln zeigen. Sie wären dann in Designs, wie denen von Studie 1

und Studie 2 oder den von Rittle-Johnson benutzten, nicht gemeinsam nachweisbar, weil diese jeweils nur ein einziges Zeitfenster untersuchen.

Neben der zeitlichen Auflösung einer Untersuchung kann auch die Untersuchungsdomäne die Ergebnisse beeinflussen. Im vorherigen Abschnitt wurde beschrieben, dass das konzeptuelle und das prozedurale Wissen in verschiedenen Domänen unterschiedlich komplex sein können. Das könnte neben ihren Interkorrelationen auch ihre Kausalrelationen beeinflussen. Schon in Abschnitt 2.3.6 wurde darauf hingewiesen, dass unterschiedliche Aufgabenschwierigkeiten, die psychologisch gesehen mit der Aufgabenkomplexität eng zusammenhängen, die Befunde mitbeeinflussen können. Neben solchen auf der quantitativen Betrachtungsebene beobachtbaren Prozessen können natürlich auch die spezifischen Inhalte unterschiedlicher Domänen die Kausalrelationen beeinflussen, indem sie jeweils unterschiedliche Überlegungen und Assoziationen in den Lernenden auslösen. So berichtet der Literaturüberblick von Rittle-Johnson und Siegler (1998) große Domänenunterschiede zwischen den Befunden zur Erwerbsreihenfolge der Wissensarten. Er verdeutlichte jedoch auch, dass sich mangels kontrollierter Prätest-Posttest-Studien über die Auswirkung von Domänenunterschieden auf kausale Relationen bis jetzt nichts empirisch Gesichertes aussagen lässt.

Diese Überlegungen verdeutlichen, dass die Befunde von Studie 1 und Studie 2 nicht auf andere Zeitfenster und andere Domänen als die verwendeten generalisiert werden können. Sie machen zugleich auf sinnvolle Folgeuntersuchungen aufmerksam.

Forschungsfrage 5: Können konzeptuelles und prozedurales Wissen durch geeignete Interventionen unabhängig voneinander vermittelt werden?

Die Vermittelbarkeit konzeptuellen und prozeduralen Wissens durch kurze Intervention wurde in der ersten Phase von Studie 2 überprüft. Wie berichtet gelang die Vermittlung konzeptuellen Wissens durch das Vorlesen einer detaillierten Erklärung wichtiger allgemeiner Eigenschaften von Dezimalbrüchen mit anschließender Reflexion nicht. Diese Intervention vermittelte auch kein prozedurales Wissen. Die Vermittlung prozeduralen Wissens durch das Auswendiglernen von Lösungsbeispielen mit anschließender Problemlöseübung gelang. Erwartungsgemäß wurde dabei kein konzeptuelles Wissen vermittelt.

Dass konzeptuelles und prozedurales Wissen überhaupt vermittelbar sind, steht außer Frage (s. Abschnitt 5.4.1). Die Vermittelbarkeit durch relativ einfache Interventionen zeigte sich auch in den Zuwächsen beider Wissensarten vom ersten zum zweiten Messzeitpunkt in Studie 1.

Wie in der Einleitung zu Studie 2 beschrieben, versuchten die Autoren von mindestens vier Studien (nämlich Byrnes & Wasik, 1991; Hiebert & Wearne, 1996; Rittle-Johnson & Alibali, 1999; Rittle-Johnson et al., 2001) erfolglos, konzeptuelles Wissen ohne prozedurales und umgekehrt zu vermitteln. Diesen Studien steht lediglich eine einzige (nämlich Resnick & Omanson, 1987) mit neun Teilnehmern mit Lernschwierigkeiten gegenüber, in der konzeptuelles Wissen vermittelt wurde, ohne dass gleichzeitig das prozedurale anstieg.

Die aktuelle Studie zeigt erstmals, dass es möglich ist, prozedurales Wissen ohne konzeptuelles zu vermitteln. Sie belegt dadurch auch, dass die Umsetzung prozeduralen Wissens in konzeptuelles Wissen nicht so schnell und effizient sein kann, dass sie schon während der Intervention stattfand.

Dass es nicht gelang konzeptuelles Wissen ohne prozedurales zu vermitteln, wurde in der Diskussion von Studie 2 zunächst auf eine suboptimale Gestaltung des Treatment zurückgeführt. Das Ergebnis passt jedoch so gut in das Gesamtbild der eben genannten fünf Studien zu derselben Fragestellung, dass es nicht auf studienspezifische Ursachen alleine zurückgeführt werden sollte. Dass fünf von sechs Studien erfolglos versuchten, konzeptuelles Wissen ohne prozedurales zu vermitteln, obwohl alle Studien auf unterschiedlichen Treatments, Maßen, Stichproben und Inhaltsgebieten basieren, ist ein Befund, der zeigt, dass in diesem Bereich noch zu bewältigende Herausforderungen für kognitive Lerntheorien bestehen.

Die Implikationen der Befunde für die zukünftige Forschung werden im folgenden Unterkapitel diskutiert.

11.2 Fazit und Ausblick

11.2.1 Psychologische Lerntheorien

Das Iterative Modell und die Überprüfung von Kausalhypothesen

Eines der Hauptziele von Studie 1 und Studie 2 bestand darin, eindeutigere Evidenz für die Korrektheit des Iterativen Modells und die Inkorrektheit des concepts-first views, des procedures-first views und des inactivation views zu liefern. Das ist nicht gelungen. Klare, unzweideutige Belege für bidirektionale Einflüsse zwischen konzeptuellem und prozeduralem Wissen konnten weder durch das längsschnittliche Design von Studie 1 noch durch das experimentelle Design von Studie 2 erbracht werden. Zusammen mit der Äquivalenzproblemstudie und den beiden Teilstudien zur Dezimalbruchkompetenz haben damit nun schon vier Studien erfolglos versucht, die Überlegenheit des Iterativen Modells

gegenüber den anderen drei zu belegen. Auch keiner der drei konkurrierenden Ansätze war den anderen klar überlegen. Die bestehenden Theorien miteinander zu vergleichen ist offensichtlich ausgesprochen schwierig. Insbesondere ist die experimentelle Überprüfung der Kausalhypothesen problematisch, weil nicht klar ist, wie die Wissensarten gezielt und unabhängig voneinander vermittelt werden können, was zur Treatmenterteilung notwendig ist.

Für die zukünftige Forschung stellt sich daher die Aufgabe, Treatments zu finden, die einzelne Wissensarten gezielt beeinflussen können. Solche Treatments würden einerseits belegen, dass es unterschiedliche Wissensarten gibt, und andererseits eine zielgerichtete Gestaltung von Lernumgebungen zur Wissensvermittlung ermöglichen. Vor allem stellen solche Treatments aber die Voraussetzung dafür dar, dass die Kausalrelationen zwischen den Wissensarten in Experimenten überprüft werden können. Studie 2, die von ihrem Design her sehr valide Befunde über die Kausalrelationen hätte erbringen können, war deswegen nicht erfolgreich, weil die Vermittlung der Wissensarten in der ersten Lernphase nur unzureichend gelang. Sobald das Problem der einzelnen Vermittlung der Wissensarten gelöst ist, sollten die Kausalrelationen leicht zu untersuchen sein.

Die Ambiguität der bisher vorliegenden Befunde zu Interrelationen der Wissensarten verdeutlicht, warum die Debatten um die Unterscheidbarkeit und um die relative Wichtigkeit von konzeptuellem und prozeduralem Wissen, um ihre Erwerbsreihenfolge und um ihre kausalen Interrelationen so lange andauerten und teilweise noch immer unentschieden sind.

Spekulative Anteile von Lerntheorien

Dass die bisherigen Befunde keine eindeutigen Schlüsse über die Relationen zwischen konzeptuellem und prozeduralem Wissen zulassen, muss nicht unbedingt ein grundsätzliches, unüberwindbares Problem darstellen, sondern könnte einfach den aktuellen Stand der Forschung widerspiegeln.

Die Diskrepanz zwischen der Detailliertheit der Theorien und der Uneindeutigkeit der Befunde kann sogar als Motor wirken, der wichtige Impulse für die weitere Forschung gibt. Voraussetzung dafür ist allerdings, dass die Theorien so formuliert werden, dass sie prinzipiell empirisch überprüfbar sind.

Das ist oft nur teilweise der Fall. Zwei Beispiele wurden in Kapitel 5 besprochen. Wenn Karmiloff-Smith verschiedene Ebenen expliziten Wissens postuliert, die sich angeblich in der Expliztheit der Repräsentation, der Bewusstseinsfähigkeit und der Verbalisierbarkeit des Wissens unterscheiden, so ist schwer zu sehen, wie sie empirisch

untersuchbar sein können, da explizites Wissen doch stets über sprachliche Aufgaben gemessen wird und subjektive Bewusstseinsinhalte unabhängig von verbalen Berichten empirisch nicht erfasst werden können. Ähnliches gilt für die von der ACT-Theorie vorgeschlagenen Mechanismen der Umsetzung deklarativen Wissens in prozedurales. Diese werden mit einem sehr hohen Grad an Detailliertheit beschrieben, während andererseits unklar ist, ob und wie die Wissensarten überhaupt unabhängig voneinander gemessen werden können – ein Ungleichgewicht, dass nun schon seit zwanzig Jahren besteht. Eine klarere Trennung zwischen empirisch überprüfbaren und rein spekulativen Theorieteilen sollte stärker als bisher angestrebt werden.

Die Wichtigkeit von Prozessmodellen

Die beiden genannten Theorien bieten aber auch Vorzüge, indem sie die Aufmerksamkeit weg von Wissen als statischer Repräsentation hin auf die Verarbeitungsprozesse von Wissen lenken. Das Iterative Modell beinhaltet keine Aussagen über solche Prozesse, was seine größte theoretische Schwäche darstellt. Wie in Unterkapitel 5.1 dargestellt, sind Aussagen über Erwerbsreihenfolge und Richtung der Kausaleinflüsse zwischen den Wissensarten gute Orientierungspunkte für den Entwurf von Lerntheorien und die Gestaltung von Unterricht. Die Forschung darf jedoch nicht bei ihnen stehen bleiben, sondern muss auch die Mechanismen aufdecken, die Erwerbsreihenfolgen und Kausaleinflüssen zugrunde liegen. Erst diese können intra- und interindividuelle Unterschiede sowie Domänendifferenzen erklären und die Entwicklung wirkungsvoller Interventionen ermöglichen.

Hier zeigt sich jedoch ein großes methodisches Problem: Wenn schon solch einfache und globale Fragen, wie die nach den Kausalrelationen, so schwierig zu beantworten sind – was nicht zuletzt auch die vorliegende Arbeit verdeutlicht – um wie viel schwieriger müssen dann Fragen nach spezifischen Verarbeitungsmechanismen der Wissensarten zu untersuchen und zu beantworten sein? Wie schon im vorherigen Abschnitt beschrieben wurde, ist es hier wichtig, nicht in reine Spekulationen über solche Prozesse abzugleiten, sondern Methoden zu entwickeln, die ihre empirische Untersuchung zulassen.

Die zeitliche Struktur kognitiver Prozesse

Die Schwierigkeit bei der Untersuchung von Wissensverarbeitungsprozessen besteht nicht nur in der Entwicklung ihrer Operationalisierungen, sondern auch darin, eine angemessene theoretische Beschreibungsebene für sie zu finden. In Unterkapitel 11.1 wurde angemerkt, dass in der vorliegenden Arbeit lediglich Lernprozesse untersucht wurden, die Stunden

oder Tage dauerten. Anderson (2002) weist jedoch darauf hin, dass einige kognitive Prozesse, wie die Aktivierung eines Neurons, nur eine Millisekunde dauern, während andere, wie die Konstruktion und Vermittlung von Wissen in sozialen Kontexten, Monate und Jahre in Anspruch nehmen können. Die Entwicklung von kognitiven Modellen, die die Interaktion so unterschiedlicher Mechanismen mit so unterschiedlichen zeitlichen Skalierungen integriert beschreiben können, stellt Anderson zu Recht als eine der großen bisher unbewältigten Herausforderungen der kognitiven Psychologie dar.

Eine Möglichkeit, verschiedene Prozesse, die auf unterschiedlichen Zeitebenen ablaufen, überhaupt systematisch gleichzeitig zu untersuchen, bieten mikrogenetische Designs (Miller & Coyle, 1999; Siegler & Crowley, 1991). In mikrogenetischen Designs bearbeiten die Probanden viele einander ähnliche Aufgaben, gewöhnlich mit Feedback. Die Aufgabenbearbeitung fungiert dabei gleichzeitig als Treatment, weil mit zunehmender Bearbeitungsdauer die Aufgabenkorrektheiten zunehmen und die verwendeten Lösungsstrategien effizienter werden, und als Test, weil die jeweilige Lösungsqualität aufgezeichnet wird. Die zeitliche Auflösung der gewonnenen Daten ist dadurch sehr hoch. Nehmen dieselben Personen an mehreren mikrogenetischen Datenerhebungssitzungen teil, können die Daten pro Sitzung aggregiert werden. Werden zwei Variablen in solchen Designs gemessen, so können ihre Relationen sowohl auf der Ebene kurzfristiger Interaktionen von Aufgabe zu Aufgabe als auch auf der Ebene langfristiger Interaktionen von Sitzung zu Sitzung ausgewertet werden.

Die einzige bisher veröffentlichte mikrogenetische Studie, die kurzfristige und langfristige Lerneffekte derselben Personen vergleicht, stammt von Siegler und Svetina (2002). Sie untersuchten allerdings nicht die Beziehungen zwischen konzeptuellem und prozeduralem Wissen, sondern Strategieentwicklungen.

Die mikrogenetische Untersuchung der Relationen zwischen den Wissensarten besitzt somit großes Potenzial, auch wenn ihr zeitlicher Fokus natürlich nicht die gesamte von Anderson beschriebene zeitliche Bandbreite kognitiver Prozesse abdeckt. Einer der Hauptgründe dafür, warum eine solche Studie noch nicht durchgeführt wurde, ist, dass die valide Messung von konzeptuellem und prozeduralem Wissen in mikrogenetischen Designs noch schwieriger ist als in anderen Designs, weil sie pro Messzeitpunkt durch jeweils eine einzige Aufgabe geleistet werden muss, da in mikrogenetischen Designs jede Aufgabe einen eigenen Messzeitpunkt darstellt.

Wissen, Intelligenz und Arbeitsgedächtnis

Bessere kognitive Prozessmodelle der Verarbeitung von konzeptuellem und prozeduralem Wissen könnten helfen, die Beziehungen der Wissensarten zu anderen zentralen psychologischen Konstrukten, insbesondere der Intelligenz, zu klären. In Studie 1 dieser Arbeit hatten Intelligenz und Vorwissen ungefähr gleich starke Einflüsse auf spätere Wissenszuwächse. Noch interessanter als die Ergebnisse solcher Regressionsanalysen wären freilich Befunde darüber, an welchen Stellen und zu welchen Zeiten die Intelligenz welche Rolle bei der Verarbeitung von konzeptuellem und prozeduralem Wissen spielt. Da die Intelligenz eng mit der Arbeitsgedächtniskapazität zusammenhängt, könnten dazu Überlegungen hilfreich sein, wie die Wissensarten im Arbeitsgedächtnis beziehungsweise in seinen Unterkomponenten verarbeitet werden. Wie Abschnitt 4.3.3 und Unterkapitel 4.4 erkennen lassen, greifen Überlegungen zu Wissensarten zwar häufig Befunde aus der Forschung zum Langzeitgedächtnis auf, ignorieren jedoch Arbeitsgedächtnismodelle.

11.2.2 Die Unterscheidung von Wissensarten

Hierarchische Wissensstrukturen

Schon in der Diskussion von Studie 1 wurde darauf hingewiesen, dass die relativ niedrige interne Konsistenz des prozeduralen Wissensfaktors in Verbindung mit der nachgewiesenen konvergenten Validität der Maße prozeduralen Wissens dafür spricht, dass prozedurales Wissen ein hierarchisches Konstrukt darstellt, das aus mehreren Subdimensionen besteht, die durch einen gemeinsamen Faktor zweiter Ordnung beeinflusst werden. Die Angemessenheit eines solchen Modells konnte in Studie 1 und Studie 2 nicht überprüft werden, da eine solche Überprüfung voraussetzt, dass mehrere Maße pro Subdimension erhoben werden. Jedoch kann die Automatisierung als eine Eigenschaft prozeduralen Wissens betrachtet werden und mehrere Studien zeigten, dass das Konstrukt der Automatisierung aus verschiedenen Subdimensionen besteht (vgl. Abschnitt 4.2.3). Allerdings fanden diese Studien lediglich niedrige Korrelationen zwischen zwei oder mehr Maßen der Automatisierung. Die konvergente Validität verschiedener Maße einer möglichen Subdimension und eine hierarchische Gesamtstruktur der Automatisierung überprüften sie nicht.

Eine hierarchische Struktur konzeptuellen Wissens ist angesichts der hohen internen Konsistenz der Faktoren konzeptuellen Wissens nicht wahrscheinlich. Allerdings könnte sich die Befundlage ändern, wenn das Wissen älterer Kinder über komplexere Inhaltsgebiete, wie die Evolution oder Wirtschaftssysteme, erhoben würde.

Lösungsbeispiele als mögliche Mediatoren

In Abschnitt 4.2.3 wurden berichtet, dass Nyberg und Kollegen (Nyberg, 1994; Nyberg et al., 2003) mittels konfirmatorischer Faktoranalysen zwei Subfaktoren deklarativen Wissens fanden, die als semantisch-deklarative und episodisch-deklarative Gedächtnisleistung interpretiert werden können. Die semantisch-deklarative Gedächtnisleistung bezieht sich auf das Erinnern allgemeingültiger inhaltlicher Zusammenhänge, die semantisch-episodische auf das Erinnern konkreter Situationen und Beispiele.

Die semantisch-deklarative Gedächtniskomponente weist damit eine große Ähnlichkeit zu konzeptuellem Wissen auf. Doch auch die episodisch-deklarative Komponente könnte bei der Verarbeitung konzeptuellen und prozeduralen Wissens eine Rolle spielen. Sie kann nämlich auch Lösungsbeispiele umfassen. In Studie 2 wurde nachgewiesen, dass Lösungsbeispiele eine Quelle prozeduralen Wissens darstellen. Das episodisch-deklarative Gedächtnis könnte somit eine wichtige Mittlerrolle zwischen verschiedenen Wissensarten einnehmen, weil es deklaratives Wissen umfasst, aber gleichzeitig Quelle prozeduralen Wissens ist.

In zukünftigen Studien zu deklarativem oder konzeptuellem und prozeduralem Wissen sollte darum das Wissen um konkrete Lösungsbeispiele, im Bereich von Multiplikationsaufgaben beispielsweise das auswendig gelernte Ein-mal-Eins, miterhoben und als Mediator zwischen den Wissensarten modelliert werden. Dabei könnte konzeptuelles Wissen als semantisch-deklarative Gedächtnisleistung und die Lösungsbeispiele als episodisch-deklarative Gedächtnisleistung konzeptualisiert werden, so dass als Messmodell für beide das hierarchische Modell deklarativen Wissens von Nyberg mit den entsprechenden beiden Unterfaktoren herangezogen werden kann.

Die Unterscheidung von mentaler Repräsentation und Inhalt

In den meisten Studien zur Untersuchung von konzeptuellem und prozeduralem Wissen, so auch in den beiden hier vorgelegten, wird nicht zwischen der mentalen Repräsentationsform von Wissen und dem Inhalt des Wissens unterschieden. Wie in Abschnitt 4.4.3 beschrieben wurde, kann diese Unterscheidung aber in manchen Fällen sinnvoll sein.

Beispielsweise ist vorstellbar, dass manche Kinder quadratische Gleichungen lösen, ohne das Konzept der quadratischen Relationen zwischen Variablen erklären zu können. Andere Kinder können vielleicht das Konzept erklären, aber die Gleichungen nicht lösen. Beides kann man durch die Annahme erklären, dass die einen Kinder nur prozedurales Wissen besitzen und die anderen nur konzeptuelles. Dabei ist jedoch die Frage, wie die

Wissensarten kognitiv gespeichert werden, unerheblich. Selbst wenn beide Wissensarten in derselben Weise kognitiv repräsentiert würden, bliebe die Unterscheidung zwischen ihnen sinnvoll, weil sie auch auf der Ebene der Wissensinhalte liegt.

Man kann ein Kind bitten, das allgemeine Konzept quadratischer Relationen zu erklären, und man kann ein Kind bitten, quadratische Gleichungen zu lösen. Dabei weiß man ohne weitere Untersuchung, dass die eine Aufgabe (inhaltlich gesehen) konzeptuelles Wissen erhebt und die andere Aufgabe (inhaltlich gesehen) prozedurales, weil die Aufgabentexte sich auf unterschiedliche Inhalte beziehen. Wenn im Folgenden Wissensarten entlang dieser inhaltlichen Dimension unterschieden werden, wird von „Konzepten“ und „Prozeduren“ gesprochen.

Unabhängig von dieser Unterscheidung kann man jedoch auch fragen, wie das Wissen, das zur Lösung der beiden Aufgaben notwendig ist, kognitiv repräsentiert wird. Es ist vorstellbar, dass Wissen über das Konzept in explizit kodierten und daher verbalisierbaren Knoten von Wissensnetzen gespeichert werden. Diese Art der Repräsentation wird konzeptuellem Wissen, nicht jedoch prozeduralem Wissen zugesprochen (vgl. Abschnitt 4.4.1). Dieses Wissen wird durch Zelle A in Tabelle 52 symbolisiert. Es ist jedoch auch vorstellbar, dass Wissen über das Konzept in automatisiert abrufbaren, implizit kodierten und daher nicht verbalisierbaren Produktionsregeln gespeichert wurde. Diese Produktionsregeln würden immer dann ausgeführt werden, wenn man gebeten würde, das Konzept zu erklären. Je öfter man diese Aufgabe übte, umso stärker würde das Wissen automatisiert. Produktionsregeln gelten als Repräsentationsform prozeduralen Wissens. Solches prozedurales Wissen über Konzepte wird durch Zelle C in Tabelle 52 symbolisiert.

Tabelle 52: Vier mögliche Kombinationen von Repräsentationsform und -inhalt.

Mentale Repräsentationsform	Repräsentationsinhalt	
	Konzept	Prozedur
konzeptuell	A	B
prozedural	C	D

Ähnliches gilt für das Wissen über Prozeduren. Es kann in automatisiert ausführbaren Produktionsregeln (s. Zelle D) genauso gut repräsentiert werden wie in einem Netz explizit repräsentierter Wissenseinheiten (s. Zelle B). Letzteres erkennt man daran, dass Kinder Prozeduren ja auch verbal beschreiben, ihre Funktionsweise erklären und bewerten können, was nicht möglich sein dürfte, wenn Prozeduren nur als implizite Produktionsregeln gespeichert würden.

Die Unterscheidung der Wissensarten auf der Inhaltsebene wird in vielen eher pädagogisch-psychologisch ausgerichteten Arbeiten benutzt, zum Beispiel in den meisten in Kapitel 2 vorgestellten. Die Unterscheidung auf der Ebene kognitiver Repräsentationsformate wird vor allem in kognitiven Arbeiten, zum Beispiel über Gedächtnisstrukturen oder Automatisierungsprozesse, benutzt (vgl. Unterkapitel 4.4).

Es wurde beschrieben, dass viele Autoren nicht zwischen diesen beiden Dimensionen der Unterscheidung von konzeptuellem und prozeduralem Wissen unterscheiden und nicht wahrnehmen, dass kognitive und pädagogisch-psychologische Arbeiten die Konstrukte teilweise unterschiedlich interpretieren. Das belegen unter anderem die Zitationsanalyse in Tabelle 2 auf Seite 52 und auch einige der Definitionen in

Tabelle 3 auf Seite 54, die die Wissensarten gleichzeitig durch inhaltliche Charakteristika und durch eine kognitive Repräsentationsform definieren.

Greeno et al. (1984) weisen darauf hin, dass es sehr produktiv sein kann, die eher pädagogischen und die eher kognitiven Befunde zu integrieren (vgl. Abschnitt 2.2.2). Auch im Rahmen der vorliegenden Arbeit wurde deutlich, dass es viele interessante Berührungspunkte zwischen den beiden Forschungssträngen gibt. In Studie 1 wurde auch empirisch gezeigt, dass verschiedene Maße, die aus ganz unterschiedlichen kognitiven und pädagogischen Studien adaptiert wurden, konvergente Validitäten besitzen.

Die Fokussierung auf die Gemeinsamkeiten der eher pädagogischen und der eher kognitiven Ansätze zur Unterscheidung von konzeptuellem und prozeduralem Wissen hat allerdings dazu geführt, dass die allermeisten Studien in diesem Gebiet lediglich die durch die Zellen A und D in Tabelle 52 beschriebenen Konstrukte untersuchten. Gerade wenn von Interesse ist, wie konzeptuelles und prozedurales Wissen miteinander interagieren, könnten aber die durch die Zellen B und C symbolisierten Konstrukte konzeptuell repräsentierten Wissens über Prozeduren und prozedural repräsentierten Wissens über Konzepte von Bedeutung sein.

Folgeuntersuchungen zu dieser Fragestellung sind sinnvoll. Die Unterscheidung der beiden Dimensionen könnte auch auf der theoretischen Ebene zu mehr Klarheit beitragen und Auswahl und Vergleich von Operationalisierung der Wissensarten vereinfachen. Beispielsweise wurde in Abschnitt 2.4.3 beschrieben, dass Blöte et al. (2001) die Evaluation von Prozeduren benutzten, um konzeptuelles Wissen zu erheben, und Hecht (1998) dasselbe Maß, um prozedurales Wissen zu erheben. Diese Verwirrung konnte dadurch entstehen, dass das Maß Evaluation vermutlich das Wissen aus Zelle B in Tabelle 52 operationalisiert. Inhaltlich gesehen bezieht sich die Aufgabe auf Prozeduren, nicht auf Konzepte. Doch die Bewertung der Prozeduren auf der sprachlichen Ebene setzt voraus,

dass explizites Wissen über sie vorhanden ist, das Relationen zu anderen Wissensseinheiten aufweist, also konzeptuell repräsentiert wird. Hecht hat sich bei seiner Interpretation des Maßes offensichtlich eher an der inhaltlichen Dimension orientiert, Blöte und Kollegen eher an der Dimension der Repräsentationsformate.

11.2.3 Methodische Implikationen der Befunde

Die Verwendung einzelner Wissensmaße

Angesichts der Befunde von Studie 1 und Studie 2 sollte in der weiteren Forschung die quantitative Ausprägung konzeptuellen Wissens vor allem durch einfache Transferaufgaben gemessen werden, weil diese die höchsten Reliabilitäten besitzen. Prozedurales Wissen sollte weder über die nicht valide Problemlösedauer noch über die auch konzeptuelles Wissen messende Problemlösekorrektheit erhoben werden, sondern möglichst durch Zugriffsasymmetrie oder Dual task-Kosten.

Die meisten internen Konsistenzen der in Studie 1 und Studie 2 verwendeten Einzelmaße waren zufriedenstellend. Einzig die beiden Maße, in denen verbale Erklärungen oder Bewertungen verbal beschriebener Problemlösestrategien verlangt waren, wiesen niedrige interne Konsistenzen auf. Bei der Konstruktion von Wissensmaßen, in denen verbales Material eine große Rolle spielt, sollte daher zukünftig auf die Homogenität der verwendeten verbalen Aufgaben besonderer Wert gelegt werden. Ferner sollten verbale Maße auf mehr Einzelaufgaben basieren als nonverbale Maße, um die Reliabilitätsunterschiede zu kompensieren.

Maße, die erhoben werden, indem die Probanden eine Reihe von Aufgaben lösen, die sich lediglich im numerischen Aufgabenmaterial, aber nicht in den sonstigen Aufgabenanforderungen unterscheiden, besitzen, wie Studie 1 und Studie 2 zeigten, hohe interne Konsistenzen und gute Reliabilitäten. Man muss bei ihrer Verwendung allerdings sehr darauf achten Deckeneffekte zu vermeiden, denn wenn ein Proband das aufgabenrelevante Lösungswissen erwirbt, kann er damit wegen der Homogenität der Aufgaben oft auf einen Schlag viele von ihnen lösen. Vor allem in Untersuchungen mit mehreren Messzeitpunkten könnte die Konstruktion von inhomogeneren Maßen sinnvoll sein, die zwar niedrigere interne Konsistenzen besitzen, aber aus unterschiedlich schwierigen Aufgaben bestehen, so dass der Summenscore mit zunehmendem Lösungswissen nur graduell ansteigt.

Die Verwendung von Strukturgleichungsmodellen in der Wissenspsychologie

Die systematischen Vergleiche jeweils mehrerer Maße von konzeptuellem und prozeduralem Wissen, die zu den hier diskutierten Ergebnissen führten, waren möglich, weil in der vorliegenden Arbeit konzeptuelles und prozedurales Wissen als latente Variablen in Strukturgleichungsmodellen konzeptualisiert wurden. Dieser Ansatz hat sich außerordentlich bewährt. Er hat nicht nur Probleme in der Messung der Wissensarten explizit aufgezeigt, die den früheren Forschungsarbeiten implizit innewohnten, sondern hat auch eine wesentlich validere Untersuchung der längsschnittlichen Relationen zwischen den Wissensarten im Vergleich zu früheren Studien ermöglicht.

Besonders anschaulich wurde das im Abschnitt 9.4.8. Dort wurden drei verschiedene Pfadmodelle vorgestellt, in denen konzeptuelles und prozedurales Wissen jeweils durch eine manifeste Variable erhoben wurden, und mit dem Strukturgleichungsmodell verglichen, in dem die Wissensarten als latente Faktoren modelliert wurden. Alle drei Pfadmodelle mit manifesten Variablen implizieren qualitativ unterschiedliche, ja sogar einander widersprechende Aussagen über die Beziehungen zwischen konzeptuellem und prozeduralem Wissen und keine dieser Aussagen stimmt mit denen überein, die aus den Koeffizienten des Strukturgleichungsmodells mit den latenten Variablen folgen.

Dieses Ergebnis ist der vielleicht wichtigste Befund der vorliegenden Arbeit. Er stellt die Validität aller bestehenden Studien zu konzeptuellem und prozeduralem Wissen in Frage. Schon Rittle-Johnson und Siegler (1998) wiesen in ihrem Literaturüberblick darauf hin, dass die Befunde zu konzeptuellem und prozeduralem Wissen inhomogen sind. Während sie das interpretativ hauptsächlich auf Domänenunterschiede zurückführten, legen die in Studie 1 und Studie 2 präsentierten Befunde nahe, dass alleine die Verwendung unterschiedlicher Wissensmaße ausreichen kann, um ein solch inhomogenes Bild zu erzeugen. Nachfolgende Studien werden sehr vorsichtig und reflektiert mit diesem Problem umgehen müssen.

Die vorliegende Arbeit legt nahe, dass drei Strategien beim Umgang mit dem Problem hilfreich sein könnten: (a) Es sollten vor allem solche manifesten Wissensmaße benutzt werden, deren Validität in konfirmatorischen Faktoranalysen nachgewiesen wurde. (b) Es sollten in jeder Studie mehrere Maße pro Wissensart benutzt werden, die dann miteinander kombiniert und verglichen werden können. (c) So oft wie möglich sollten die Wissensarten als latente Faktoren modelliert werden.

Messtheoretisch gesehen ist die Modellierung der Wissensarten als latente Faktoren (Punkt c) stets der Verwendung uni- oder multivariater Wissensmaße (Punkte a und b) vorzuziehen. Jedoch verdeutlichte Studie 1 unter anderem auch die pragmatischen

Nachteile der Arbeit mit Strukturgleichungsmodellen: Es werden große Stichproben, zwei bis vier Indikatoren pro Faktor und möglichst normalverteilte Maße mit wenigen oder keinen fehlenden Werten benötigt und es ist immer möglich, dass einige Forschungshypothesen wegen Konvergenzproblemen nicht explizit getestet werden können.

Für die Entwicklung des Forschungsfeldes könnten darum wenige große faktoranalytische Studien hilfreich sein, in denen optimale manifeste Einzelmaße oder Summenscores der Wissensarten validiert werden. Dabei könnten einerseits theoriegeleitet verschiedene Typen von Testaufgaben, beispielsweise verbale und nonverbale, verglichen werden, andererseits könnten auch direkt standardisierte Tests der Wissensarten in einem Inhaltsgebiet entwickelt werden. Nachfolgende Studien zu den Interrelationen der Wissensarten könnten dann auf den großen methodischen Aufwand von SEM-Studien verzichten und die zuvor gewonnenen Erkenntnisse oder die zuvor entwickelten standardisierten Tests zur Messung der Wissensarten verwenden.

Bildgebende Verfahren

Da es wichtig ist, kognitive Prozesse der Wissensverarbeitung zu untersuchen, diese jedoch anhand des Verhaltens schlecht beobachtbar sind, könnte man denken, dass neuropsychologische Methoden, insbesondere bildgebende Verfahren, zur Untersuchung der Wissensarten besonders geeignet sind, weil sie den „direkten Blick“ ins Gehirn erlauben.

Tatsächlich können neuropsychologische Methoden jedoch nur wenig zur Erforschung der Wissensarten beitragen. Würde ein bildgebendes Verfahren beispielsweise indizieren, dass bei der Bearbeitung einer Aufgabe zwei Bereiche im Hirn überdurchschnittlich stark aktiviert werden, so könnte das bedeuten, dass zwei verschiedene Wissensarten zur Aufgabenlösung benutzt werden. Es könnte aber auch bedeuten, dass eine Wissensart in zwei verschiedenen Hirnarealen gespeichert wurde und von dort abgerufen wird. Es könnte ebenfalls bedeuten, dass in einem Hirnareal die eigentliche Aufgabenlösung durchgeführt wird, während das andere die Antwortformulierung übernimmt. Diese Reihe von Vermutungen lässt sich beliebig fortsetzen. Natürlich gibt es Hintergrundinformationen über die Funktionsweise einzelner Hirnteile, die einige dieser Hypothesen wahrscheinlicher und andere unwahrscheinlicher machen. Trotzdem können neuropsychologische Verfahren alleine nicht belegen, ob es unterschiedliche Wissensarten gibt.

Ursache für diese Unterbestimmtheit neuropsychologischer Evidenz ist, dass Wissensarten durch funktionale Attribute, wie Flexibilität, Transferfähigkeit, Verbalisierbarkeit und so weiter, definiert werden, bildgebende Verfahren jedoch lediglich räumliche und zeitliche Aktivierungsmuster aufzeichnen (Schumacher, 2005).

Wenn man schon weiß, welche Wissensarten zur Lösung einer Aufgabe erforderlich sind, ist es durchaus möglich, mit bildgebenden Verfahren nach *neurologischen Korrelaten* dieser Wissensart zu suchen. Die Suche nach neurologischen Korrelaten der Wissensarten kann aber erst erfolgen, wenn die Wissensarten durch psychologische Tests auf der Verhaltensebene nachweisbar sind.

Diese Grenzen bildgebender Verfahren zeigten sich schon in Abschnitt 4.3.4, indem berichtet wurde, dass sich mit neurologischen Studien durchaus unabhängig voneinander funktionsfähige Gedächtnissysteme nachweisen lassen, dass deren psychologische Charakteristika, also ihre Eigenschaften auf der funktionalen Beschreibungsebene, trotz intensiver Untersuchungen jedoch nicht nachgewiesen werden konnten. Die neuropsychologischen Befunde besitzen daher für Fragen der Gestaltung von Schulunterricht für Kinder ohne Hirnschäden bisher so gut wie keine Implikationen.

Der besondere Wert neuropsychologischer Methoden als Methoden der Grundlagenforschung, die innovative Befunde über räumliche und zeitliche Aktivationsmuster im Hirn und ihre Korrelation mit psychologischen Variablen liefern, wird dadurch natürlich nicht geschmälert.

11.2.4 Implikationen der Befunde für die Praxis

Wie in Unterkapitel 5.5 beschrieben wurde, könnten Befunde über konzeptuelles und prozedurales Wissen sowohl zur Auswahl optimaler Unterrichtsziele als auch zur Auswahl optimaler Unterrichtssequenzierungen genutzt werden.

Adaptive Expertise als Unterrichtsziel

Baroodys (2003) Plädoyer dafür, dass die Integration von konzeptuellem und prozeduralem Wissen ein wichtiges Unterrichtsziel darstellt, weil sie die Grundlage von adaptiver Expertise ist, setzt voraus, dass starke Dissoziationen zwischen konzeptuellem und prozeduralem Wissen auftreten können und in der Praxis tatsächlich auftreten. Denn sonst müsste der Zustand der Integration nicht extra als Unterrichtsziel angestrebt werden. Baroodys Annahme steht entgegen, dass viele frühere Studien schon positive Korrelationen zwischen konzeptuellem und prozeduralem Wissen fanden. Darüber hinaus legen die Befunde von Studie 1 und Studie 2 nahe, dass die tatsächliche Stärke des Zusammenhangs bisher eher unter- als überschätzt wurde. Die Integration von

konzeptuellem und prozeduralem Wissen scheint grundsätzlich nicht besonders schwierig zu erreichen zu sein.

Interessant ist allerdings, dass Studienteilnehmer, die die Zahlenstrahlaufgaben in einer Intervention in Studie 2 ohne Feedback übten, prozedurales Wissen unabhängig von konzeptuellem erwarben. Studienteilnehmer, die in Studie 1 dieselben Aufgaben im Rahmen des „Fang-das-Monster“-Spiels mit Feedback übten, erwarben nämlich beide Wissensarten. Dissoziationen zwischen Wissensarten könnten also vor allem bei ungenügendem Feedback auftreten. Die Rolle des Feedback in diesem Zusammenhang sollte in zukünftigen Studien eingehender erforscht werden. Insbesondere sollte auch überprüft werden, ob das Lösen von Übungsaufgaben ohne direktes Feedback, was beispielsweise bei Hausaufgaben der Fall ist, auch langfristig zu rein prozeduralem Wissen ohne tieferes konzeptuelles Verständnis führt.

Die optimale Wissenserwerbsreihenfolge

Weil die Wissensarten in den bisherigen Untersuchungen so eng zusammenhingen, lassen sich bis jetzt keine Aussagen über die optimale Wissenserwerbsreihenfolge treffen. Falls die Wissensarten stets gemeinsam erworben würden, wäre die Frage nach der optimalen Erwerbsreihenfolge sinnlos. Auch auf diesem Forschungsgebiet können erst substanzielle Fortschritte erzielt werden, wenn Treatments identifiziert wurden, die die Vermittlung einer Wissensart ohne die jeweils andere erlauben, weil erst dann Hypothesen getestet werden können.

Ökologische Validität der Designs

Die Beziehungen zwischen konzeptuellem und prozeduralem Wissen können sich zwischen Domänen und sogar zwischen verschiedenen Zeitskalen der Beobachtung unterscheiden. Umso wichtiger ist die Untersuchung der Wissensarten in Designs mit hoher ökologischer Validität. Während die zeitliche Strukturierung und das gewählte Inhaltsgebiet der beiden empirischen Studien dieser Arbeit eine große Nähe zu schulischem Lernen aufweisen, besitzen das konkrete Studienmaterial und das Lernsetting nur eine eingeschränkte ökologische Validität.

Dass konzeptuelles und prozedurales Wissen hoch interkorreliert waren, nachdem Kinder am Computer mit dem „Fang-das-Monster“-Spiel lernten, bedeutet nicht automatisch, dass die Korrelation auch hoch ist, nachdem Kinder in der Schule eine Unterrichtseinheit über Dezimalbrüche absolvierten. Neben den schon vorgeschlagenen Experimenten, die die Wirkung einzelner Treatments vergleichen, wären darum auch

Feldstudien interessant, in denen Wissenstests in mehreren Klassen jeweils vor und nach einer Unterrichtseinheit zu einem Thema durchgeführt würden. Ein solches Design könnte direkten Aufschluss darüber geben, ob die Wissensarten zur Erklärung schulischer Lernprozesse nützlich sind oder nicht.

Als vorbildlich in dieser Hinsicht kann die Längsschnittstudie von Hiebert und Wearne (1996) gelten. Sie verfolgten die Entwicklung konzeptuellen und prozeduralen Wissens von Schülergruppen, die an traditionellem und konstruktivistischem Unterricht teilnahmen. Ihre Tests konzeptuellen und prozeduralen Wissens bezogen sich direkt auf die Unterrichtsinhalte. Wie berichtet, fanden sie, dass die Unterrichtsart sich nicht unterschiedlich auf die Mittelwerte der Wissensarten auswirkte, wohl aber auf ihre Interkorrelationen.

Würde man die SEM-Methodik der vorliegenden Arbeit mit dem Design von Hiebert und Wearne verbinden, so könnten direkte konfirmatorische Tests darüber durchgeführt werden, ob verschiedene Arten von Unterricht sich unterschiedlich auswirken (a) auf die Mittelwerte der Wissensarten, (b) auf die Faktorstruktur mehrerer Maße derselben Wissensart, (c) auf die Interkorrelationen zwischen den Wissensarten und (d) auf die prädiktiven Relationen zwischen den Wissensarten vor und nach einer Unterrichtseinheit.

Falls es nicht möglich ist, Schüler innerhalb des normalen Schulunterrichts randomisiert verschiedenen Arten von Unterricht zuzuordnen, so könnten auch bestehende Unterschiede zwischen Staaten oder Bundesländern für quasiexperimentelle Designs genutzt werden. Beispielsweise könnte man die Wissenstests mit 200 Schülern aus einem Bundesland, das bei den Pisa-Mathematik-Tests (Prenzel et al., 2005) besonders schlecht abschnitt, und 200 Schülern aus einem Bundesland, das bei den PISA-Mathematik-Tests besonders gut abschnitt, durchführen und überprüfen, ob die Bundesländer sich nicht nur auf Ebene der Gesamtleistung, sondern auch auf Ebene der eben genannten vier Parameter unterscheiden. Die Arbeitshypothese wäre dabei, dass die in dem Bundesland mit höherer Pisa-Leistung vorgefundene Wissensstruktur eine eher gute Grundlage und die in dem Bundesland mit niedrigerer Pisa-Leistung vorgefundene Wissensstruktur eine eher ungünstige Grundlage komplexer Problemlöseprozesse darstellt. Das könnte in nachfolgenden Laborexperimenten weiter validiert werden.

Die Gestaltung von Lernsoftware

Ein konkretes Beispiel dafür, wie Erkenntnisse über die Vermittelbarkeit von Wissensarten in der Praxis genutzt werden können, stellen die kognitiven Tutoren dar, die auf Grundlage der ACT-Theorie entwickelt wurden (z.B. Anderson, Boyle, Farrell, & Reiser, 1987;

Anderson, Corbett, Koedinger, & Pelletier, 1995). Das sind Lernprogramme für Kinder, die für verschiedene Domänen implementiert wurden, beispielsweise Algebra, Geometrie und Computerprogrammierung. Kinder können mit ihnen am Bildschirm Aufgaben lösen. Das Programm vergleicht dabei die vom Kind benutzten Produktionsregeln mit abgespeicherten Produktionsregeln, die Experten zur Lösung desselben Problems benutzen. Dadurch kann das Programm nicht nur feststellen, welche Aufgaben ein Kind lösen kann und welche nicht, sondern es analysiert auch, welche Produktionsregeln dem Kind fehlen und welche es fehlerhaft anwendet. Dadurch kann das Programm hoch individualisierte Rückmeldungen und Unterstützungen geben, die genau auf das bestehende Wissen eines Kindes zugeschnitten sind.

Die Tutoren sind bis jetzt auf Problembereiche beschränkt, die von relativ hoher prozeduraler Komplexität, jedoch niedrigerer konzeptueller Komplexität sind. Sie können bis jetzt zwar prozedurales, nicht jedoch konzeptuelles Wissen gezielt vermitteln, unter anderem weil sprachliche Reflexions- und Schlussfolgerungsprozesse sich wesentlich schwieriger auf einem Computer modellieren lassen als Problemlöseprozesse. Weitere Informationen über die Vermittelbarkeit einzelner Wissensarten könnten hier also hilfreich sein.

Anknüpfungspunkte zu anderen Forschungsfeldern

Die Forschung zu konzeptuellem und prozeduralem Wissen beschäftigt sich mit der Frage, welche Wissensstrukturen die Grundlage von Expertise darstellen und wie sie erworben werden können. Diese Fragestellung ist nicht nur in Bezug auf schulisches Lernen interessant, sondern auch in Bezug auf viele andere Lebenssituationen, in denen Expertise vonnöten ist.

Wichtige Erkenntnisse liefert unter anderem die Forschung zum *naturalistic decision making* (z.B. Zsombok & Klein, 1997). In diesem Paradigma werden reale Entscheidungen außerhalb des Labors untersucht und zwar vor allem solche, die mit Zeitdruck, Stress, Unsicherheit, hohen Kosten bei Fehlentscheidungen und schlecht definierten Zielzuständen einhergehen. Solche Entscheidungssituationen können beispielsweise im Rahmen von Brandbekämpfungseinsätzen, Notfallsituationen bei der Flugzeugsteuerung, der Anästhesie, Kernkraftwerksunglücken und militärischen Operationen entstehen. Wie vor allem die Forschung zum *recognition-primed decision making* (Überblick bei Lipshitz, Klein, Orasanu, & Salas, 2001) ergab, vergleichen erfahrene Entscheider in solchen Situationen oft nicht explizit verschiedene Handlungsmöglichkeiten miteinander. Stattdessen wird ihnen sofort nach der Situationswahrnehmung die adäquate

Handlungsstrategie bewusst. Das wird dadurch erklärt, dass die Entscheider aufgrund ihrer Vorerfahrung Muster von Situationen im Langzeitgedächtnis gespeichert haben, in denen bestimmte Handlungsstrategien adäquat sind. Durch automatisiert ablaufende pattern matching-Prozesse werden diese Handlungsstrategien aktiviert, sofern die aktuelle Situation eine ausreichende Ähnlichkeit mit früher erlebten Situationen besitzt. Hier zeigen sich deutlich Parallelen einerseits zu der Aussage der ACT-Theorie, dass eine Produktion dann aktiviert wird, wenn der aktuelle Arbeitsgedächtnisinhalt den Bedingungsteil der Produktion matcht, und andererseits zur im Rahmen des Paradigmas der situated cognition aufgestellten These, dass Wissenstransfer zwischen zwei Situationen nur dann stattfindet, wenn in beiden Situationen hinreichend ähnliche Handlungsmöglichkeiten und -grenzen wahrgenommen werden. Das macht deutlich, wie wichtig prozedurales Wissen im Rahmen des naturalistic decision making ist.

Die empirische Forschung zeigte aber auch die Wichtigkeit konzeptuellen Wissens, das nämlich immer dann zur Entscheidungsfindung herangezogen wird, wenn eine Situation nicht genügend Ähnlichkeit zu schon bewältigten Situationen besitzt. In diesem Fall ist einerseits genaues Wissen über die Beschaffenheit der aktuellen Situation, andererseits aber auch ein konzeptuelles Tiefenverständnis des zu steuernden Systems, beispielsweise des Flugzeugs mit all seinen technischen Komponenten und ihren Interaktionsmöglichkeiten, notwendig.

Die Grundidee hinter dem recognition-primed decision making weist eine große Nähe zu der evolutionpsychologischen These auf, dass prozedurales Wissen schnelle und effiziente Routineentscheidungen ermöglicht und konzeptuelles Wissen langsame, aber wohldurchdachte Reaktionen auf neue Situationen. Keine einzelne Wissensart kann beides leisten, daher bietet der Besitz beider gemeinsam einen Evolutionsvorteil (s. Abschnitt 4.3.5). Dieser Gedanke stützt wiederum Baroodys Idee von verknüpftem konzeptuellen und prozeduralen Wissen als optimalem Lernziel (s. Abschnitt 5.5.2).

Der Forschung zu konzeptuellem und prozeduralem Wissen und zum naturalistic decision making sind offensichtlich einige Grundfragen gemeinsam, beispielsweise: Wie können die Wissensarten optimal erworben und optimal miteinander verknüpft werden? Wann findet der Transfer von prozeduralem Wissen von alten auf ähnliche neue Problemsituationen statt und wie kann er gefördert werden? Wie kann erreicht werden, dass neue Problemtypen nicht durch inadäquates übergeneralisiertes prozedurales Wissen, sondern durch das notwendige konzeptuelle Wissen gelöst werden? Parallelen zwischen den Forschungsfeldern sollten gezielt weiterverfolgt werden.

Auch in anderen Bereichen der Entscheidungsforschung zeigen sich interessante Berührungspunkte. So diskutieren beispielsweise Kahneman und Frederick (2005) ausführlich, wie die Befunde des *heuristics and biases approach* durch dual process theories of cognition erklärt werden können.

Ergänzend sei auf ein besonders eindrückliches Beispiel dafür hingewiesen, wie gut Befunde zu den Wissensarten auch außerhalb der Pädagogischen Psychologie genutzt werden können. Wiener et al. (2004) diskutieren in ihrem Artikel: „Guided jury discretion in capital murder cases: The role of declarative and procedural knowledge“ die genauen Umstände, unter denen in den USA Jurys aus juristischen Laien über die Schuld von Angeklagten befinden, die des vorsätzlichen Mordes angeklagt sind. Die Entscheidungen der Jury haben in einigen Bundesstaaten Einfluss darauf, ob für schuldig befundene Angeklagte die Todesstrafe erhalten. Jedes Jurymitglied erhält zum Prozess genaue schriftliche Anweisung über seine Aufgabe, die Art, wie es die Aufgabe lösen soll, und welche juristischen Umstände dabei zu berücksichtigen sind. Wiener und Kollegen nutzten die Unterscheidung von deklarativem und prozeduralem Wissen, um anhand konkreter Fallbeispiele zu analysieren, wie sich das deklarative Wissen, das die Juroren schon vor dem Prozess über Verbrechen und Rechtsprechung besitzen, auf ihre Verarbeitung der deklarativen Anweisungen zu ihren Aufgaben im Prozess auswirkt. Ferner untersuchten sie, wie sich die Abwesenheit prozeduralen Wissens, das heißt jeglicher Vorerfahrung mit ihrer Aufgabe als Juror in diesem Prozess bemerkbar macht. Dass diese beiden Umstände die Urteilsfindung negativ beeinflussen können, wird jedem Leser des Artikels einleuchten, obwohl die Autoren Fragen nach kognitiven Verarbeitungsmechanismen, Repräsentationen und Operationalisierungen der Wissensarten nicht berühren. Dieses Beispiel zeigt, wie produktiv Theorien über konzeptuelles und prozedurales Wissen und ähnliche Wissensarten schon jetzt in der Praxis genutzt werden können, obwohl noch so viele ihrer Einzelheiten zu erforschen bleiben.

11.3 Zusammenfassung

Die beiden empirischen Studien dieser Arbeit zeigen, dass verschiedene Maße konzeptuellen oder prozeduralen Wissens gute konvergente Validitäten sowie weitgehend metrische Messinvarianz besitzen. Die beiden Wissensarten hingen jedoch korrelativ sehr eng zusammen und ließen sich auch durch Treatments kaum differenziell beeinflussen. Ihre hohe Interkorrelation behinderte in beiden Studien die Verwendung der Wissensarten als Prädiktoren für spätere Wissenszuwächse, so dass keine gesicherten Aussagen über Kausalrelationen ableitbar waren. Zukünftige Studien sollten Treatments entwickeln, die

die Wissensarten einzeln beeinflussen können, und Strukturgleichungsmodelle verwenden, um standardisierte Tests der Wissensarten und ihrer Subdimensionen zu entwickeln. Mit Hilfe solcher Instrumente sollten durch Laborexperimente, längsschnittliche Feldstudien und mikrogenetische Studien auf verschiedenen Zeitebenen Prozessmodelle der Verarbeitung der Wissensarten empirisch validiert werden.

12 Literatur

- Ackerman, P. L. (1988). Components of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General*, 117, 288-318.
- Ackerman, P. L. (1989). Individual differences and skill acquisition. In P. L. Ackerman, Sternberg, R. J., & Glaser, R. (Ed.), *Learning and individual differences* (pp. 165-217). New York: Freeman.
- Ackerman, P. L., & Cianciolo, A. T. (2000). Cognitive, perceptual-speed, and psychomotor determinants of individual differences during skill acquisition. *Journal of Experimental Psychology: Applied*, 6(4), 259-290.
- Aebli, H. (1983). *Zwölf Grundformen des Lehrens*. Stuttgart: Klett-Cotta.
- Anderson, J. R. (1976). *Language, memory and thought*. Hillsdale, NJ: Erlbaum.
- Anderson, J. R. (1978). Arguments concerning representations for mental imagery. *Psychological Review*, 85, 249-277.
- Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, 89, 369-406.
- Anderson, J. R. (1983). *The architecture of cognition* (10th ed.). Cambridge, MA.: Harvard University Press.
- Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.
- Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.
- Anderson, J. R. (2002). Spanning seven orders of magnitude: A challenge for cognitive modeling. *Cognitive Science*, 26, 85-112.
- Anderson, J. R., & Bower, G. H. (1973). *Human associative memory*. Washington: Winston.
- Anderson, J. R., Boyle, C. F., Farrell, R., & Reiser, B. J. (1987). Cognitive principles in the design of computer tutors. In P. Morris (Ed.), *Modeling cognition* (pp. 93-133). New York: Wiley.
- Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: lessons learned. *The Journal of the Learning Sciences*, 4(2), 167-207.
- Anderson, J. R., & Fincham, J. M. (1994). Acquisition of procedural skills from examples. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(6), 1322-1340.
- Anderson, J. R., Fincham, J. M., & Douglass, S. (1997). The role of examples and rules in the acquisition of a cognitive skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(4), 932-945.
- Anderson, J. R., Fincham, J. M., & Douglass, S. (1999). Practice and retention: A unifying analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(5), 1120-1136.
- Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Hillsdale, NJ: Erlbaum.
- Anderson, J. R., Reder, L. M., & Simon, H. A. (1996). Situated learning and education. *Educational Researcher*, 25(4), 5-11.
- Anderson, J. R., Reder, L. M., & Simon, H. A. (1997). Situative versus cognitive perspectives: Form versus substance. *Educational Researcher*, 18(1), 18-21.
- Ashby, F. G., & Gott, R. (1988). Decision rules in perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 33-53.
- Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, 128(4), 612-637.

- Baroody, A. J. (2003). The development of adaptive expertise and flexibility: The integration of conceptual and procedural knowledge. In A. J. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Constructing adaptive expertise* (pp. 1-33). Mahwah, NJ: Erlbaum.
- Baroody, A. J., & Gannon, K. E. (1984). The development of the commutativity principle and economical addition strategies. *Cognition and Instruction*, 1, 321-339.
- Baroody, A. J., & Ginsburg, H. P. (1986). The relationship between initial meaningful and mechanical knowledge of arithmetic. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 75-112). Hillsdale, NJ: Erlbaum.
- Barrett, L. F., Tugade, M. M., & Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin*, 130(4), 553-573.
- Bartlett, F. C. (1932). *Remembering: An experimental and social study*. Cambridge: Cambridge University Press.
- Bentler, P. M. (1996). Covariance structure analysis: Statistical practice, theory, and directions. *Annual Review of Psychology*, 47, 563-592.
- Berry, D. C., & Broadbent, D. E. (1984). On the relationship between task performance and associated verbalizable knowledge. *Quarterly Journal of Experimental Psychology*, 36A, 209-231.
- Berry, D. C., & Broadbent, D. E. (1995). Implicit learning in the control of complex systems. In P. A. Frensch & J. Funke (Eds.), *Complex Problem Solving: The European Perspective* (pp. 131-150). Hillsdale, NJ: Erlbaum.
- Besner, D., Stolz, J. A., & Boutilier, C. (1997). The stroop effect and the myth of automaticity. *Psychonomic Bulletin & Review*, 4(2), 221-225.
- Blaxton, T. A. (1989). Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(4), 657-668.
- Blöte, A. W., Burg, E. v. d., & Klein, A. S. (2001). Students' flexibility in solving two-digit addition and subtraction problems: Instruction effects. *Journal of Educational Psychology*, 93(3), 627-638.
- Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, 110(2), 305-314.
- Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: Wiley.
- Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, 53, 605-634.
- Boomsma, A. (2000). Reporting analyses of covariance structures. *Structural Equation Modeling*, 7(3), 461-483.
- Borland Software Corporation. (1983-2002). Borland Delphi Professional (Version 7.0) [Programmentwicklungsumgebung]. Scotts Valley, CA: Borland Software Corporation.
- Borsboom, D., Mellenbergh, G. J., & Heerden, J. v. (2003). The theoretical status of latent variables. *Psychological Review*, 110(2), 203-219.
- Bortz, J. (1993). *Statistik für Sozialwissenschaftler* (4th ed.). Berlin: Springer.
- Bortz, J., & Döring, N. (1995). *Forschungsmethoden und Evaluation für Sozialwissenschaftler*. Berlin: Springer.
- Bower, G. H., Karlin, M. B., & Dueck, A. (1975). Comprehension and memory for pictures. *Memory and Cognition*, 2, 216-220.
- Britton, B. K., & Tesser, A. (1982). Effects of prior knowledge on use of cognitive capacity in three complex cognitive tasks. *Journal of Verbal Learning and Verbal Behavior*, 21, 421-436.

- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models* (pp. 136-162). Newbury Park: Sage.
- Bruer, J. T. (1994). *Schools for thought: A science of learning in the classroom*. Cambridge, MA: MIT Press.
- Burkholder, G. J., & Harlow, L. L. (2003). An illustration of a longitudinal cross-lagged panel design for larger structural equation models. *Structural Equation Modeling*, 10(3), 465-486.
- Byrnes, J. P., & Wasik, B. A. (1991). Role of conceptual knowledge in mathematical procedural learning. *Developmental Psychology*, 27(5), 777-786.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Canobi, K. H. (2004). Individual differences in children's addition and subtraction knowledge. *Cognitive Development*, 19, 81-93.
- Canobi, K. H., Reeve, R. A., & Pattison, P. E. (1998). The role of conceptual understanding in children's addition problem solving. *Developmental Psychology*, 34(5), 882-891.
- Canobi, K. H., Reeve, R. A., & Pattison, P. E. (2002). Young children's understanding of addition concepts. *Educational Psychology*, 22(5), 513-532.
- Canobi, K. H., Reeve, R. A., & Pattison, P. E. (2003). Patterns of knowledge in children's addition. *Developmental Psychology*, 39(3), 521-534.
- Caroll, J. B. (1993). *Human cognitive abilities. A survey of factor-analytic studies*. New York: Cambridge University Press.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233-255.
- Chi, M. T. H., Slotta, J. D., & Leeuw, N. d. (1994). From things to processes: A theory of conceptual change for learning science concepts. *Learning and Instruction*, 4, 27-43.
- Chiou, J.-s., & Spreng, R. A. (1996). The reliability of difference scores: A re-examination. *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, 9, 158-167.
- Chomsky, N. (1965). *Aspects of a theory of syntax*. Cambridge, MA: MIT Press.
- Church, R. B., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition*, 23, 43-71.
- Clancey, W. J. (1993). Situated action: A neuropsychological interpretation. *Cognitive Science*, 17, 87-107.
- Cobb, P., & Bowers, J. (1999). Cognitive and situated learning perspectives in theory and practice. *Educational Researcher*, 28(2), 4-15.
- Cohen, N. J., Poldrack, R. A., & Eichenbaum, H. (1997). Memory for items and memory for relations in the procedural/declarative memory framework. *Memory*, 5, 131-178.
- Collins, A. M., & Loftus, E. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82, 407-428.
- Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8, 240-247.
- Cowan, R., & Renton, M. (1996). Do they know what they are doing? Children's use of economical addition strategies and knowledge of commutativity. *Educational Psychology*, 16(4), 407-420.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671-684.
- Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, 44(1-2), 1-42.

- Dehaene, S. (1999). *The number sense: How the mind creates mathematics*: Oxford University Press.
- Dehaene, S., & Akhavein, R. (1995). Attention, automaticity, and levels of representation in number processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(2), 314-326.
- Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, 122(3), 371-396.
- Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neurosciences*, 21(8), 355-361.
- Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance*, 16(3), 626-641.
- diSessa, A. A., Gillespie, N. M., & Esterly, J. B. (2004). Coherence versus fragmentation in the development of the concept of force. *Cognitive Science*, 28, 843-900.
- diSessa, A. A., & Sherin, B. L. (1998). What changes in conceptual change? *International Journal of Science Education*, 20(10), 1155-1191.
- Dixon, J. A., & Moore, C. F. (1996). The development of intuitive principles in choosing mathematical strategies. *Developmental Psychology*, 32, 241-253.
- Donnelly, C. M., & McDaniel, M. A. (1993). Use of analogy in learning scientific concepts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(4), 975-987.
- Doumas, L. A. A., & Hummel, J. E. (2005). Approaches to modeling human mental representations: What works, what doesn't, and why. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 73-94). Cambridge: Cambridge University Press.
- Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence*, 41, 1-63.
- Felbrich, A. (2005). *Kontrastierungen als effektive Lerngelegenheiten zur Vermittlung von Wissen über Repräsentationsformen am Beispiel des Graphen einer linearen Funktion*. Unpublished doctoral dissertation, Technische Universität Berlin, Berlin.
- Fodor, J. (1985). Précis of *The Modularity of Mind*. *Behavioral and Brain Sciences*, 8, 1-42.
- Foster, J. K., & Jelicic, M. (1999). *Memory: Systems, process or function?* New York: Oxford University Press.
- Fuson, K. C. (1988). *Children's counting and concept of number*. New York: Springer.
- Fuson, K. C., & Briars, D. J. (1990). Using a base-ten blocks learning/teaching approach for first- and second grade place-value and multidigit addition and subtraction. *Journal for Research in Mathematics Education*, 21(3), 180-206.
- Gabrieli, J. D. E. (1998). Cognitive neuroscience of human memory. *Annual Review of Psychology*, 49, 87-115.
- Gattis, M. (2002). Structure mapping in spatial reasoning. *Cognitive Development*, 17, 1157-1183.
- Geary, D. C. (1994). *Children's mathematical development: Research and practical applications*. Washington, DC: American Psychological Association.
- Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number*. Cambridge, MA: Harvard University Press.
- Gelman, R., & Meck, E. (1983). Preschoolers' counting: Principles before skill. *Cognition*, 13, 343-359.
- Gelman, R., & Meck, E. (1986). The notion of principle: The case of counting. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 29-57). Hillsdale, NJ: Erlbaum.

- Gelman, R., & Williams, E. M. (1998). Enabling constraints for cognitive development and learning: Domain specificity and epigenesis. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology: Cognition, perception, and language* (5th ed., Vol. 2, pp. 575-630). New York: Wiley.
- Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7, 155-170.
- Gentner, D. (1988). Metaphor as structure-mapping: The relational shift. *Child Development*, 59, 47-59.
- Gentner, D., Brem, S., Ferguson, R. W., Markman, A. B., Levidow, B. B., Wolff, P., & Forbus, K. D. (1997). Analogical reasoning and conceptual change: A case study of Johannes Kepler. *Journal of the Learning Sciences*, 6(1), 3-40.
- Gentner, D., & Loewenstein, J. (2002). Relational language and relational thought. In E. Amsel & J. P. Byrnes (Eds.), *Language, literacy, and cognitive development: The development and consequences of symbolic communication* (pp. 87-120). 2002: Erlbaum.
- Gentner, D., Loewenstein, J., & Thomson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology*, 95(2), 393-408.
- Gentner, D., & Wolff, P. (2000). Metaphor and knowledge change. In E. Dietrich & A. Markman (Eds.), *Cognitive dynamics: Conceptual change in humans and machines* (pp. 295-342). Mahwah, NJ: Erlbaum.
- Gigerenzer, G. (1998). We need statistical thinking, not statistical rituals. *Behavioral and Brain Sciences*, 21(2), 199-200.
- Gilmore, C. K., & Bryant, P. (zur Publikation angenommen). Individual differences in children's understanding of inversion and arithmetic skill. *British Journal of Psychology*.
- Glisky, E. L. (1992). Acquisition and transfer of declarative and procedural knowledge by memory-impaired patients: A computer data-entry task. *Neuropsychologia*, 30(10), 899-910.
- Goldin-Meadow, S., Alibali, M. W., & Church, R. B. (1993). Transitions in concept acquisition: Using the hand to read the mind. *Psychological Review*, 100(2), 279-297.
- Goldstone, R. L., & Kersten, A. (2003). Concepts and categorization. In I. B. Weiner (Ed.), *Handbook of psychology* (Vol. 4, pp. 599-621). Hoboken, NJ: Wiley.
- Grabner, R., Stern, E., & Neubauer, A. (2003). When intelligence loses its impact: Neural efficiency during reasoning in a highly familiar area. *International Journal of Psychophysiology*, 49, 89-98.
- Greeno, J. G. (1994). Gibson's affordances. *Psychological Review*, 101(2), 336-342.
- Greeno, J. G. (1997). On claims that answer the wrong questions. *Educational Researcher*, 5-21.
- Greeno, J. G. (1998). The situativity of knowing, learning, and research. *American Psychologist*, 53(1), 5-26.
- Greeno, J. G., Moore, J. L., & Smith, D. R. (1993). Transfer of situated learning. In D. K. Detterman & R. J. Sternberg (Eds.), *Transfer on trial: Intelligence, cognition, and instruction*. Norwood, NJ: Ablex.
- Greeno, J. G., Riley, M. S., & Gelman, R. (1984). Conceptual competence and children's counting. *Cognitive Psychology*, 16, 94-143.
- Gunzelmann, G., & Anderson, J. R. (2003). Problem solving: Increased planning with practice. *Cognitive Systems Research*, 4, 57-76.
- Guthrie, J. T., Weber, S., & Kimmerly, N. (1993). Searching documents: Cognitive processes and deficits in understanding graphs, tables, and illustrations. *Contemporary Educational Psychology*, 18, 186-221.

- Haapsalo, L., & Kadjievich, D. (2000). Two types of mathematical knowledge and their relation. *Journal für Mathematik-Didaktik*, 21(2), 139-157.
- Halford, G. S. (1993). *Children's understanding: The development of mental models*. Hillsdale, NJ: Erlbaum.
- Hardy, I., Schneider, M., Jonen, A., Stern, E., & Möller, K. (zur Publikation angenommen). Fostering diagrammatic reasoning in science education. *Swiss Journal of Psychology*.
- Hardy, I., & Stadelhofer, B. (unter Begutachtung). Concept Maps wirkungsvoll als Strukturierungshilfe einsetzen: Welche Rolle spielt die aktive Konstruktion? *Zeitschrift für Pädagogische Psychologie*.
- Hatano, G. (1982). Cognitive consequences of practice in culture specific procedural skills. *The Quarterly Newsletter of the Laboratory of Comparative Human Cognition*, 4, 15-18.
- Hecht, S. A. (1998). Toward an information-processing account of individual differences in fraction skills. *Journal of Educational Psychology*, 90(3), 545-559.
- Hedtke, R. (2000). *Das unstillbare Verlangen nach Praxisbezug: Zum Theorie-Praxis-Problem der Lehrerbildung am Beispiel Schulpraktischer Studien* [Onlinejournalartikel]. Retrieved 11. März, 2005, from the World Wide Web: <http://www.sowi-onlinejournal.de/lehrerbildung/hedtke.htm>
- Heindel, W. C., Salomon, D. P., Shults, C. W., Walicke, P. A., & Butters, N. (1989). Neuropsychological evidence for multiple implicit memory systems: A comparison of Alzheimer's, Huntington's, and Parkinson's Disease patients. *The Journal of Neuroscience*, 9(2), 582-587.
- Heller, K. A., & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision: Manual*. Göttingen: Hogrefe.
- Hiebert, J. (1992). Mathematical, cognitive, and instructional analyses of decimal fractions. In G. Leinhardt & R. Putnam & R. Hattup (Eds.), *Analysis of mathematics for mathematics teaching* (pp. 283-322). Hillsdale, NJ: Erlbaum.
- Hiebert, J. (Ed.). (1986). *Conceptual and procedural knowledge: The case of mathematics*. Hillsdale, NJ: Erlbaum.
- Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 1-27). Hillsdale, NJ: Erlbaum.
- Hiebert, J., & Wearne, D. (1986). Procedures over concepts: The acquisition of decimal number knowledge. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 199-223). Hillsdale, NJ: Erlbaum.
- Hiebert, J., & Wearne, D. (1996). Instruction, understanding, and skill in multidigit addition and subtraction. *Cognition and Instruction*, 14(3), 251-283.
- Hintikka, J. K. J. (1975). *The intentions of intentionality and other new models for modality*. Boston, MA: Reidel.
- Holyoak, K. J. (2005). Analogy. In K. J. Holyoak (Ed.), *The Cambridge handbook of thinking and reasoning* (pp. 117-142). Cambridge: Cambridge University Press.
- Howe, C., Tolmie, A., Duchak-Tanner, V., & Rattray, C. (2000). Hypothesis testing in science: group consensus and the acquisition of conceptual and procedural knowledge. *Learning and Instruction*, 10, 361-391.
- Hu, L.-T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3(4), 424-453.
- Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1-55.

- Humphreys, M. S., Bain, J. D., & Pike, R. (1989). Different ways to cue a coherent memory system: A theory for episodic, semantic, and procedural tasks. *Psychological Review*, 96(2), 208-233.
- Inhelder, B., & Piaget, J. (1980). Procedures and structures. In D. R. Olson (Ed.), *The social foundations of language and thought* (pp. 19-27). New York: Norton.
- Johnson, A. (2003). Procedural memory and skill acquisition. In I. B. Weiner (Ed.), *Handbook of Psychology* (pp. 499-523). Hoboken, NJ: Wiley.
- Jonassen, D. H., Beissner, K., & Yacci, M. (1993). *Structural knowledge: Techniques for representing, conveying, and acquiring structural knowledge*. Hillsdale, NJ: Erlbaum.
- Jong, T. d., & Ferguson-Hessler, M. G. M. (1996). Types and qualities of knowledge. *Educational Psychologist*, 31(2), 105-113.
- Jonides, J., Naveh-Benjamin, M., & Palmer, J. (1985). Assessing automaticity. *Acta Psychologica*, 60, 157-171.
- Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural Equation Models in the Social Sciences* (pp. 85-112). New York: Academic Press.
- Jungermann, H. (1983). The two camps on rationality. In R. W. Scholz (Ed.), *Decision making under uncertainty*. Amsterdam: North-Holland.
- Kahneman, D., & Chajczyk, D. (1983). Tests of the automaticity of reading: Dilution of Stroop effects by color-irrelevant stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, 9(4), 497-509.
- Kahneman, D., & Frederick, S. (2005). A model of heuristic judgement. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 267-293). Cambridge: Cambridge University Press.
- Kaplan, D. (1990). Evaluating and modifying covariance structure models: A review and recommendation. *Multivariate Behavioral Research*, 25(2), 137-155.
- Kaplan, D. (2000). *Structural Equation Modelling: Foundations and Extensions*. Thousand Oaks: Sage.
- Karmiloff-Smith, A. (1979). Micro- and macrodevelopmental changes in language acquisition and other representational systems. *Cognitive Science*, 3, 91-118.
- Karmiloff-Smith, A. (1986). From meta-processes to conscious access: Evidence from children's metalinguistic and repair data. *Cognition*, 23, 95-147.
- Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science* (2nd ed.). Cambridge, MA: MIT Press.
- Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling*, 10(3), 333-351.
- Klein, S. B., Cosmides, L., Tooby, J., & Chance, S. (2002). Decisions and the evolution of memory: Multiple systems, multiple functions. *Psychological Review*, 109(2), 306-329.
- Kouba, V. L., Carpenter, P. T., & Swafford, J. O. (1989). Number and operations. In M. M. Lindquist (Ed.), *Results from the fourth mathematics assessment of the National Assessment of Educational Progress* (pp. 64-93). Reston, VA: National Council of Teachers of Mathematics.
- Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11, 65-99.
- Lavigne, N. C. (2005). Mutually informative measures of knowledge: Concept maps plus problem sorts in statistics. *Educational Assessment*, 10(1), 39-71.
- Leinhardt, G., Zaslavsky, O., & Stein, M. K. (1990). Functions, graphs, and graphing: Tasks, learning, and teaching. *Review of Educational Research*, 60(1), 1-64.

- Levy, J., & Pashler, H. (2001). Is dual-task slowing instruction dependent? *Journal of Experimental Psychology: Human Perception and Performance*, 27, 862-869.
- Lindenberger, U., Marsiske, M., & Baltes, P. B. (2000). Memorizing while walking: Increase in dual-task costs from young adulthood to old age. *Psychology and Aging*, 15(3), 417-436.
- Lipshitz, R., Klein, G., Orasanu, J., & Salas, E. (2001). Taking stock of naturalistic decision making. *Journal of Behavioral Decision Making*, 14, 331-352.
- Logan, G. D. (1985). Skill and automaticity: Relations, implications, and future directions. *Canadian Journal of Psychology*, 39(2), 367-386.
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, 95(4), 492-527.
- Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 883-914.
- Logan, G. D., Taylor, S. E., & Etherton, J. L. (1999). Attention and automaticity: Toward a theoretical integration. *Psychological Research*, 62, 165-181.
- MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, 31, 201-226.
- MacGilly, K. (1994). *Classroom lessons: Integrating cognitive theory and classroom practice*. Cambridge, MA: MIT Press.
- Maddox, W. T., & Ashby, F. G. (2004). Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behavioral Processes*, 66, 309-332.
- Mandler, J. M. (2004). *The foundations of mind*. New York: Oxford University Press.
- Mayer, R. E., & Moreno, R. (2002). Animation as an aid to multimedia learning. *Educational Psychology Review*, 14(1), 87-99.
- McKendree, J. E., & Anderson, J. R. (1987). Frequency and practice effects on the composition of knowledge in LISP evaluation. In J. M. Carroll (Ed.), *Cognitive aspects of human-computer interaction* (pp. 236-259). Cambridge, MA: MIT Press.
- Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist*, 44(12), 1469-1481.
- Medin, D. L., & Rips, L. J. (2005). Concepts and categories: Memory, meaning, and metaphysics. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning*. Cambridge: Cambridge University Press.
- Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology*, 32, 89-115.
- Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review*, 104(1), 3-65.
- Miller, P. H., & Coyle, T. R. (1999). Developmental change: Lessons from microgenesis. In E. K. Scholnick & K. Nelson & S. A. Gelman & P. H. Miller (Eds.), *Conceptual development: Piaget's legacy* (pp. 209-239). Mahwah, NJ: Erlbaum.
- Mitchell, D. B. (1989). How many memory systems? Evidence from aging. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(1), 31-49.
- Mitchell, D. B., Brown, A. S., & Murphy, D. R. (1990). Dissociations between procedural and episodic memory: Effects of time and aging. *Psychology and Aging*, 5(2), 264-276.
- Morgan, H. (1997). *Cognitive styles and classroom learning*. Westport, CO: Praeger.
- MPlus-Homepage. (2005). *Chi-square difference testing using the Satorra-Bentler scaled chi-square* [internet tutorial]. Retrieved October 04, 2005, from the World Wide Web: <http://www.statmodel.com/chidiff.shtml>
- Müller, B. (1999). Use specificity of cognitive skills: Evidence for production rules? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(1), 191-207.

- Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, 92, 289-316.
- Muthén, B. O., & Curran, P. J. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods*, 2(4), 371-402.
- Muthén, B. O., & Muthén, L. (1998-2004a). MPlus (Version 3.01) [Software]: Muthén & Muthén.
- Muthén, B. O., & Muthén, L. (1998-2004b). *MPlus User's Guide* (3rd ed.). Los Angeles, CA: Muthén & Muthén.
- National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics: Standards 2000*. Reston, VA: National Council of Teachers of Mathematics.
- Nelson, T. O. (1978). Detecting small amounts of information in memory: Savings for nonrecognized items. *Journal of Experimental Psychology: Human Learning and Memory*, 4, 453-468.
- Nesher, P. (1986). Are mathematical understanding and algorithmic performance related? *For the Learning of Mathematics*, 6(3), 2-9.
- Nesher, P. (1987). Towards an instructional theory: The role of students' misconceptions. *For the Learning of Mathematics*, 7, 33-40.
- Nevitt, J., & Hancock, G. R. (2004). Evaluating small sample approaches for model test statistics in structural equation modeling. *Multivariate Behavioral Research*, 39(3), 439-478.
- Newell, A. (1973). Production systems: models of control structures. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press.
- Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1-55). Hillsdale, NJ: Erlbaum.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Norman, D. A., & Rumelhart, D. E. (1975). *Explorations of cognition*. San Francisco: Freeman.
- Novick, L. R. (1995). Some determinants of successful analogical transfer in the solution of algebra word problems. *Thinking and Reasoning*, 1, 5-30.
- Novick, L. R. (2001). Spatial diagrams: Key instruments in the toolbox for thought. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 40, pp. 279-325). San Diego, CA: Academic Press.
- Novick, L. R., & Hmelo, C. E. (1994). Transferring symbolic representations across nonisomorphic problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(6), 1296-1321.
- Nyberg, L. (1994). A structural equation modeling approach to the multiple memory systems question. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(2), 485-491.
- Nyberg, L., Maitland, S. B., Rönnlund, M., Bäckman, L., Dixon, R. A., Wahlin, A., & Nilsson, L.-G. (2003). Selective adult age differences in an age-invariant multifactor model of declarative memory. *Psychology and Aging*, 18(1), 149-160.
- Oberauer, K. (1993). Prozedurales und deklaratives Wissen und das Paradigma der Informationsverarbeitung. *Sprache & Kognition*, 12(1), 30-43.

- Oser, F., & Patry, J.-L. (1994). Sichtstruktur und Basismodelle des Unterrichts: Über den Zusammenhang von Lehren und Lernen unter dem Gesichtspunkt psychologischer Lernverläufe. In R. Olechowski & B. Rollett (Eds.), *Theorie und Praxis: Aspekte empirisch-pädagogischer Forschung - qualitative und quantitative Methoden* (pp. 138-168). Frankfurt: Lang.
- Oser, F., & Sarasin, S. (1995). *Basismodelle des Unterrichts: Von der Sequenzierung als Lernerleichterung* [LLF-Berichte, Vol. 11]. Interdisziplinäres Zentrum für Lern- und Lehrforschung der Universität Potsdam. Retrieved August 10, 2005, from the World Wide Web: <http://pub.ub.uni-potsdam.de/zsr/llf/llf0.htm>
- Oser, F. K., & Baeriswyl, F. J. (2001). Choreographies of teaching: Bridging instruction to learning. In V. Richardson (Ed.), *AERA's handbook of research on teaching* (4th ed.). New York: MacMillan.
- Paap, K. R., & Ogden, W. C. (1981). Letter encoding is an obligatory but capacity demanding operation. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 518-527.
- Palmeri, T. J. (1999). Theories of automaticity and the Power Law of Practice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(2), 543-551.
- Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, 116(2), 220-244.
- Pashler, H., Johnston, J. C., & Ruthruff, E. (2001). Attention and performance. *Annual Review of Psychology*, 52, 629-651.
- Paul, P. J., Churchill, G. A., Jr., & Brown, T. J. (1993). Caution in the use of difference scores in consumer research. *Journal of Consumer Research*, 19, 655-622.
- Pennington, N., Nicolich, R., & Rahm, J. (1995). Transfer of training between cognitive subskills: Is knowledge really use specific? *Cognitive Psychology*, 28, 175-224.
- Petitto, A. L. (1990). Development of numberline and measurement concepts. *Cognition and Instruction*, 7(1), 55-78.
- Piaget, J. (1978). *Success and understanding*. Cambridge, MA: Harvard University Press.
- Polanyi, M. (1962). *Personal Knowledge: Towards a post-critical philosophy*. London: Routledge and Kegan Paul.
- Porter, M. K., & Masingila, J. O. (2000). Examining the effects of writing on conceptual and procedural knowledge in calculus. *Educational Studies in Mathematics*, 42(2), 165-177.
- Poser, H. (2001). *Wissenschaftstheorie: Eine philosophische Einführung*. Stuttgart: Reclam.
- Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accomodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, 66(2), 211-227.
- Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rost, J., & Schiefele, U. (2005). *Vorinformation zu PISA 2003: Zentrale Ergebnisse des zweiten Vergleichs der Länder in Deutschland* [PDF-Broschüre]. Retrieved October 21, 2005, from the World Wide Web: http://pisa.ipn.uni-kiel.de/Vorinformation_E.pdf
- Rabinowitz, M., & Goldberg, N. (1995). Evaluating the structure-process hypothesis. In F. Weinert & W. Schneider (Eds.), *Memory performance and competencies: Issues in growth and development* (pp. 225-242). Hillsdale, NJ: Erlbaum.
- Rahmenlehrplan Grundschule: Mathematik*. (2004). Berlin: Wissenschaft und Technik.
- Rahmenlehrplan: Mathematik in der Grundschule*. (1986). Berlin: Stenatsverwaltung für Schule, Bildung und Sport.
- Rattermann, M. J., & Gentner, D. (1998). The effect of language on similarity: The use of relational labels improves young children's performance in a mapping task. In D. G. K. Holyoak, & B. Kokinov (Ed.), *Advances in analogy research: Integration of*

- theory & data from the cognitive, computational, and neural sciences* (pp. 274-282). Sophia, Bulgaria: New Bulgarian University.
- Raykov, T., Tomer, A., & Nesselroade, J. R. (1991). Reporting structural equation modeling results in *Psychology and Aging*: Some proposed guidelines. *Psychology and Aging*, 6(4), 499-503.
- Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, 118(3), 219-235.
- Reber, P. J., Knowlton, B. J., & Squire, L. R. (1996). Dissociable properties of memory systems: Differences in the flexibility of declarative and nondeclarative knowledge. *Behavioral Neuroscience*, 110(5), 861-871.
- Regan, J. E. (1981). Automaticity and learning: Effects of familiarity on naming letters. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 180-195.
- Resnick, L. B. (1982). Syntax and semantics in learning to subtract. In T. P. Carpenter & J. M. Moser & T. A. Romburg (Eds.), *Addition and subtraction: A cognitive perspective* (pp. 136-155). Hillsdale, NJ: Erlbaum.
- Resnick, L. B., & Ford, W. W. (1981). *The psychology of mathematics for instruction*. Hillsdale, NJ: Erlbaum.
- Resnick, L. B., & Omanson, S. F. (1987). Learning to understand arithmetic. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 3, pp. 41-95). Hillsdale, NJ: Erlbaum.
- Rickard, T. C. (1997). Bending the power law: A CMPL theory of strategy shifts and the automatization of cognitive skills. *Journal of Experimental Psychology: General*, 126(3), 288-311.
- Ritter, F. E., & Schooler, L. J. (2001). Learning curve, the, *International Encyclopedia of the Social and Behavioral Sciences*: Elsevier.
- Rittle-Johnson, B., & Alibali, M. W. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other? *Journal of Educational Psychology*, 91(1), 175-189.
- Rittle-Johnson, B., & Siegler, R. S. (1998). The relation between conceptual and procedural knowledge in learning mathematics: A review. In C. Donlan (Ed.), *The development of mathematical skills* (pp. 75-328). Hove, UK: Psychology Press.
- Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology*, 93(2), 346-362.
- Rolls, E. T. (2000). Memory systems in the brain. *Annual Review of Psychology*, 51, 599-630.
- Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104, 192-232.
- Rost, J. (2004). *Lehrbuch Testtheorie - Testkonstruktion* (2nd ed.). Bern: Huber.
- Royer, J. M., Cisero, C. A., & Carlo, M. S. (1993). Techniques and procedures for assessing cognitive skills. *Review of Educational Research*, 63(2), 201-243.
- Ruiz-Primo, M. A., & Shavelson, R. J. (1996). Problems and issues in the use of concept maps in science assessment. *Journal of Research in Science Teaching*, 33(6), 569-600.
- Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: pt. 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, 89, 60-94.
- Ruthruff, E., Johnston, J. C., & Selst, M. A. v. (2001). Why practice reduces dual-task interference. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 3-21.

- Ryle, G. (1949). *The concept of mind*. London: Hutchinson House.
- Säljö, R. (1999). Concepts, cognition and discourse: From mental structures to discursive tools. In W. Schnotz & S. Vosniadou & M. Carretero (Eds.), *New Perspectives on Conceptual Change*. Amsterdam: Pergamon.
- Satorra, A., & Bentler, P. M. (1994). Correction to test statistics and standard errors in covariance structure analysis. In A. v. Eye & C. C. Clogg (Eds.), *Latent variable analysis: Applications for developmental research* (pp. 399-419). Thousand Oaks, CA: Sage.
- Satorra, A., & Bentler, P. M. (1999, August 3). *A scaled difference chi-square test statistic for moment structure analysis* [working paper]. University Pompeu Fabra, Department of Economics. Retrieved February 1, 2005, from the World Wide Web: <http://ssrn.com/abstract=199064>
- Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(3), 501-518.
- Schacter, D. L. (1989). Memory. In M. I. Posner (Ed.), *Foundations of Cognitive Science* (pp. 683-725). Cambridge, MA: MIT Press.
- Schacter, D. L. (1992). Understanding implicit memory: A cognitive neuroscience approach. *American Psychologist*, 47(4), 559-569.
- Scheffler, I. (1965). *Conditions of knowledge: An introduction to epistemology and education*. Chicago: University of Chicago Press.
- Schneider, W. (1985). Toward a model of attention and the development of automaticity. In M. Posner & O. S. Martin (Eds.), *Attention and performance* (Vol. XI). Arlington, VA: U.S. Army Research Institute.
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic information processing: I. Detection, search, and attention. *Psychological Review*, 84, 1-66.
- Schoenfeld, A. (1986). On having and using geometric knowledge. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics*. Hillsdale, NJ: Erlbaum.
- Schoenfeld, A. H. (2004). The math wars. *Educational Policy*, 18(1), 253-286.
- Schumacher, E. H., Seymour, T. L., Glass, J. M., Kieras, D. E., & Meyer, D. E. (2001). Virtually perfect time sharing in dual-task performance: Uncorking the central attentional bottleneck. *Psychological Science*, 121, 101-108.
- Schumacher, R. (2005). Die prinzipielle Unterbestimmtheit der Hirnforschung im Hinblick auf die Gestaltung schulischen Lernens. In D. Struma (Ed.), *Ist das psychophysische Problem gelöst? Bewusstsein, Willensfreiheit und die Neurowissenschaften*. Frankfurt a. M.: Suhrkamp.
- Shah, P., & Hoeffner, J. (2002). Review of graph comprehension research: Implications for instruction. *Educational Psychology Review*, 14(1), 47-69.
- Sherry, D. F., & Schacter, D. L. (1987). The evolution of multiple memory systems. *Psychological Review*, 94(4), 439-454.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, 84, 127-190.
- Shrager, J., & Siegler, R. S. (1998). SCADS: A model of children's strategy choices and strategy discoveries. *Psychological Science*, 9(5), 405-410.
- Siegler, R. S. (2002). Microgenetic studies of self-explanation. In N. Granott, & Parziale, J. (Eds.), *Microdevelopment: Transition processes in Development and Learning* (pp. 31-58). Cambridge: Cambridge University Press.
- Siegler, R. S. (2003). Implications of cognitive science research for mathematics education. In J. Kilpatrick & W. B. Martin & D. E. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 219-233). Reston, VA: National Council of Teachers of Mathematics.

- Siegler, R. S., & Crowley, K. (1991). The microgenetic method: A direct means for studying cognitive development. *American Psychologist*, 46(6), 606-620.
- Siegler, R. S., & Crowley, K. (1994). Constraints on learning in nonprivileged domains. *Cognitive Psychology*, 27, 194-226.
- Siegler, R. S., & Stern, E. (1998). Conscious and unconscious strategy discoveries: A microgenetic analysis. *Journal of Experimental Psychology: General*, 127(4), 377-397.
- Siegler, R. S., & Svetina, M. (2002). A microgenetic/cross-sectional study of matrix completion: Comparing short-term and long-term change. *Child Development*, 73(3), 793-809.
- Singley, M. K., & Anderson, J. R. (1989). *The transfer of cognitive skill*. Cambridge, MA: Harvard University Press.
- Sommerville, S., & Bryant, P. (1985). Young children's use of spatial coordinates. *Child Development*, 604-613.
- Sophian, C. (1997). Beyond competence: The significance of performance for conceptual development. *Cognitive Development*, 12, 281-303.
- Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, 4, 605-632.
- Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, 99(2), 195-231.
- Squire, L. R., & Zola, S. M. (1996, November). *Structure and function of declarative and nondeclarative memory systems*. Paper presented at the Memory: Recoding Experience in Cells and Circuits, Irvine, CA.
- Star, J. R. (2000). *On the relationship between knowing and doing in procedural learning*. Paper presented at the Fourth International Conference of the Learning Sciences.
- Stark, C., Stark, S., & Gordon, B. (2005). New semantic learnign and generalization in a patient with amnesia. *Neuropsychology*, 19(2), 139-151.
- Steiger, J. H., Shapiro, A., & Browne, M. W. (1985). On the multivariate asymptotic distribution of sequaential chi-square statistics. *Psychometrika*, 50, 253-264.
- Stern, E. (2001a). Intelligence, prior knowledge, and learning. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (Vol. 11, pp. 7670-7674). Oxford: Elsevier Science.
- Stern, E. (2001b). Intelligenz, Wissen, Transfer und der Umgang mit Zeichensystemen. In E. Stern & J. Guthke (Eds.), *Perspektiven der Intelligenzforschung* (pp. 163-203). Lengerich: Pabst.
- Stern, E. (2004). Wie viel Hirn braucht die Schule? Chancen und Grenzen einer neuropsychologischen Lehr-Lern-Forschung. *Zeitschrift für Pädagogik*, 50(4), 531-538.
- Stern, E., Aprea, C., & Ebner, H. G. (2003). Improving cross-content transfer in text processing by means of active graphical representation. *Learning and Instruction*, 13, 191-203.
- Stern, E., Felbrich, A., & Schneider, M. (zur Publikation angenommen). Mathematik lernen. In D. Rost (Ed.), *Handwörterbuch Pädagogische Psychologie* (3rd ed.). Weinheim: Beltz.
- Stern, E., Hardy, I., & Koerber, S. (2002). Die Nutzung grafisch-visueller Repräsentationsformen im Sachunterricht. In K. Speckelsen, Hartinger, A., Möller, K. (Ed.), *Ansätze und Methoden empirischer Forschung zum Sachunterricht* (pp. 119-131). Bad Heilbrunn: Klinkhard.
- Stern, E., & Schneider, M. (in Vorbereitung). Informationsverarbeitung. In M. Hasselhorn & W. Schneider (Eds.), *Handbuch der Entwicklungspsychologie*. Göttingen: Hogrefe.

- Strayer, D. L., & Kramer, A. F. (1990). Attentional requirements of automatic and controlled processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(1), 67-82.
- Sun, R., Merrill, E., & Peterson, T. (2001). From implicit skill to explicit knowledge: a bottom-up model of skill learning. *Cognitive Science*, 25, 203-244.
- Sun, R., Slusarz, P., & Terry, C. (2005). The interaction of the explicit and the implicit in skill learning: A dual-process approach. *Psychological Review*, 112(1), 159-192.
- Sun, R., & Zhang, X. (2003, April 10-12). *Accessibility versus action-centeredness in the representation of cognitive skills*. Paper presented at the Fifth International Conference on Cognitive Modeling, Bamberg, Germany.
- Taatgen, N. A., & Wallach, D. (2002). Whether skill acquisition is rule or instance based is determined by the structure of the task. *Cognitive Science Quarterly*, 2(2), 163-204.
- Thorndike, E. L., & Rock, R. T., Jr. (1934). Learning without awareness of what is being learned or intended to learn it. *Journal of Experimental Psychology*, 17, 1-19.
- Torbeyns, J., Verschaffel, L., & Ghesqui re, P. (2005). Simple addition strategies in a first-grade class with multiple strategy instruction. *Cognition and Instruction*, 23(1), 1-21.
- Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381-403). San Diego, CA: Academic Press.
- Tulving, E. (1983). *Elements of episodic memory*. New York: Oxford University Press.
- Tulving, E. (1985). How many memory systems are there? *American Psychologist*, 40(4), 385-398.
- Tversky, A., & Kahneman, D. (1974). Judgement and uncertainty: Heuristics and biases. *Science*, 1124-1131.
- VanLehn, K. (1986). Arithmetic procedures are induced from examples. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics*. Hillsdale, NJ: Erlbaum.
- von Aufschnaiter, C. (in Vorbereitung). How students develop and use an explicit understanding of physics concepts. In S. Vosniadou & A. Baltas & X. Vamvakoussi (Eds.), *Advances in conceptual change research*.
- Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction*, 4, 45-69.
- Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, 24, 535-585.
- Vosniadou, S., Ioannides, C., Dimitrakopoulou, A., & Papademetriou, E. (2001). Designing learning environments to promote conceptual change in science. *Learning and Instruction*, 11, 381-419.
- Wainer, H. (1992). Understanding graphs and tables. *Educational Researcher*, 21, 14-23.
- Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New Horizons in Psychology*. Harmondsworth: Penguin.
- Wiener, R. L., Rogers, M., Winter, R., Hurt, L., Hackney, A., Kadela, K., Seib, H., Rauch, S., Warren, L., & Morasco, B. (2004). Guided jury discretion in capital murder cases. *Psychology, Public Policy, and Law*, 10(4), 516-576.
- Willingham, D. B. (1997). Systems of memory in the human brain. *Neuron*, 18, 5-8.
- Willingham, D. B. (1998). What differentiates declarative and procedural memories: Reply to Cohen, Poldrack, and Eichenbaum (1997). *Memory*, 6(6), 689-699.
- Wynn, K. (1990). Children's understanding of counting. *Cognition*, 36, 155-193.
- Yu, C.-Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. Unpublished doctoral dissertation, University of California, Los Angeles, CA.
- Zsombok, C. E., & Klein, G. (Eds.). (1997). *Naturalistic decision making*. Mahwah, NJ: Erlbaum.

