Modellierungskompetenzen beim mathematischen Problemlösen

Inventarisierung von Modellierungsprozessen beim Lösen mathematischer Textaufgaben und Entwicklung eines diagnostischen Instrumentariums

Julia Riebel

Vom Promotionsausschuss des Fachbereichs Psychologie der Universität Koblenz-Landau zur Verleihung des akademischen Grades Doktor der Philosophie (Dr. phil.) genehmigte Dissertation

Datum der Disputation: 5. Februar 2010

Vorsitzender des Promotionsausschusses:	Prof. Dr. Ingmar Hosenfeld
Berichterstatter	Prof. Dr. Reinhold S. Jäger PD Dr. Gabriele E. Dlugosch

Inhalt

Zusammenfassung

Mathematisches Modellieren bezeichnet die verschiedenen Prozesse, die Menschen durchlaufen, wenn sie versuchen, reale Probleme mathematisch zu lösen oder Textaufgaben zu bearbeiten.

In der Literatur werden im sogenannten Modellierungszyklus fünf aufeinanderfolgende Teilprozesse genannt, die den Ablauf des Problemlöseprozesses beschreiben (Blum, 2003).

Beim *Strukturieren* wird zunächst eine Problemsituation im kognitionspsychologischen Sinne verstanden, die wesentlichen Merkmale werden abstrahiert um dann im zweiten Schritt (*Mathematisieren*) in ein mathematisches Modell (bestehend aus Gleichungen, Symbolen, Operatoren, Ziffern etc.) übersetzt zu werden. Beim *Verarbeiten* dieses mathematischen Modells mithilfe mathematischen Wissens wird das Problem innermathematisch gelöst, woraufhin das Ergebnis im vierten Schritt (*Interpretieren*) rückübertragen wird auf die ursprüngliche Problemsituation. Die dort aufgeworfene Frage wird beim Interpretieren beantwortet. Der letzte Schritt (*Validieren*) sieht eine Überprüfung des gesamten Problemlöseprozesses vor: Kommt der Modellierer zu dem Schluss, dass sein Ergebnis eventuell nicht korrekt bzw. nicht optimal ist, wird der Modellierungszyklus von vorne durchlaufen.

Der Modellierungszyklus dient dabei sowohl als *deskriptives* als auch als *präskriptives (normatives)* Modell, da er einerseits *be*schreibt, wie Modellieren abläuft und andererseits *vor*schreibt, wie Schüler z.B. beim Lösen von mathematischen Textaufgaben vorgehen sollen. Dennoch liegen für beide Zielsetzungen keine empirischen Studien vor, die belegen könnten, dass der Modellierungszyklus sich tatsächlich als deskriptives oder als präskriptives Modell eignet. In der vorliegenden Arbeit konnte die Eignung als deskriptives Modell nur mit Einschränkungen bestätigt werden. So zeigte sich in einer Voruntersuchung, dass das Stadium *Validieren* bei Schülern der Klassenstufe 4 nicht zu beobachten ist und dass der Ablauf der Stadien deutlich von der vorgegebenen Reihenfolge abweichen kann. Eine revidierte Version des Modellierungszyklus als rekursives Modell wird daher vorgeschlagen.

In einer zweiten, größeren Untersuchung wurde die Eignung als normatives Modell untersucht, indem die vier Stadien *Strukturieren*, *Mathematisieren*, *Verarbeiten* und *Interpretieren* isoliert voneinander erfasst wurden, wofür zuvor eigens Items konstruiert worden waren. Die vier Unterskalen wurden an mehreren Kriterien - darunter die Fähigkeit zum Lösen kompletter Modellierungsaufgaben - validiert, was den normativen Charakter des Modells bestätigt.

Um Modellierungskompetenz frühzeitig diagnostizieren und dementsprechend fördern zu können, ist die Entwicklung von Methoden angezeigt, die vier Teilkompetenzen *Strukturieren*, *Mathematisieren*, *Verarbeiten* und *Interpretieren* bereits im Grundschulalter unabhängig voneinander objektiv, reliabel und valide zu erfassen. Mit der Entwicklung und Analyse von Items zu den vier Teilbereichen wurde hierfür mit der vorliegenden Arbeit ebenfalls eine Grundlage geschaffen.

Einleitung

„Hallo, wie kann man beim Verständnis von Textaufgaben helfen? Oder das Verständnis im Alltag unauffällig fördern? Rechnen (plus und minus, sowie einfache Mal-Aufgaben) ist kein Problem, aber er (7) schafft es einfach nicht, die Aufgaben im Text zu erfassen. So kann er mit der folgenden Aufgabe nichts anfangen: Nina hat 10 Euro. Sie soll Brote kaufen. Ein Brot kostet zwei Euro. Ich muß allerdings dazu sagen, daß er sich nicht gut konzentrieren kann und nicht gerne liest - allerdings das Gelesene in der Regel erfaßt. Bin für jede Hilfe dankbar.“, schreibt die Mutter eines Zweitklässlers am 30.05.2003 in einem Internetforum für Eltern[1].

Was hier berichtet wird, ist kein Einzelfall. Dass Schülern das Bearbeiten von Textaufgaben schwer fällt ist gemeinhin bekannt (Hegarty, Mayer & Monk, 1995). Zahlreiche Untersuchungen beschäftigen sich mit der Frage, woran dies liegt (Arendasy, Sommer & Glück, 2004; Fuchs et al., 2006; Kim, 2007; Rhymer & Cates, 2007; Stern, 1993) und wie Textaufgaben gestaltet bzw. wie das Lösen von Textaufgaben trainiert werden kann, um dieses Problem zu beheben (Hasemann & Stern, 2002; Schmid, 2008; Sommer, Arendasy & Glück, 2004).

Grundlage für solche Betrachtungen ist aber zunächst die Frage, was beim Lösen von mathematischen Textaufgaben eigentlich passiert, m. a. W. welche Anforderungen bewältigt werden müssen, um vom präsentierten Aufgabentext zu einem Lösungssatz zu kommen. Das Lösen von alltagsrelevanten Problemen (wie sie z.B. in Textaufgaben formuliert werden) mithilfe mathematischen Wissens wird als mathematisches Modellieren bezeichnet (Blum, 2003).

Die systematische Untersuchung der hierbei ablaufenden Einzelprozesse begann 1976, als Henry Pollak auf der dritten „International Conference on Mathematics Education“ in Karlsruhe den „Circle of Modelling“ vorstellte, der postuliert, dass beim mathematischen Problemlösen die folgenden fünf Teilschritte nacheinander ablaufen (Pollak, 1979):

- *Analysis of the Problem*: Die wesentlichen Merkmale der Aufgabenstellung werden vom Modellierenden erfasst, das Problem wird verstanden.
- *Mathematising*: Die zuvor als relevant identifizierten Merkmale werden „mathematisiert“, das heißt es wird ein Lösungsansatz (z.B. in Form einer zu lö-

[1] http://www.mysnip.de/forum-archiv/thema/8737/972959/Probleme+mit+Textaufgaben+-+2.+Klasse.html

senden Gleichung) aufgestellt, der in mathematischer Sprache, also in Ziffern, Zeichen, Symbolen und Operatoren ausgedrückt ist.

- *Solution of the Model*: Aus diesem Lösungsansatz wird nun mithilfe mathematischen Wissens eine mathematische Lösung des Problems gefunden (z.B. indem die zuvor aufgestellte Gleichung gelöst wird).
- *Interpretation*: Diese Lösung wird dann wieder auf die ursprüngliche Aufgabenstellung rückübertragen und vor deren Hintergrund interpretiert.
- *Validation*: Im letzten Schritt wird die Lösung auf ihre Richtigkeit überprüft.

Es ließe sich nun die Vermutung aufstellen, dass es bei dem im Eingangsbeispiel beschriebenen Schüler an der Fähigkeit zum *Mathematising* mangelt. Denn auch wenn er laut seiner Mutter nicht gerne liest, so hat er doch keine Probleme damit, den Sinn gelesener Texte zu erfassen (*Analysis of the Problem*) und auch das eigentliche Rechnen (*Solution of the Model*) scheint eher unproblematisch zu sein. Wie es mit der Fähigkeit zu *Interpretation* und *Validation* aussieht, bleibt fraglich. Dies liegt daran, dass Modellierungsprozesse hierarchisch angeordnet sind. Scheitert ein Schüler bereits in den ersten Stadien, so wird er die weiteren Stadien vermutlich gar nicht erst erreichen, weshalb sich dann keine differenzierten Aussagen über seine Kompetenzen ableiten lassen.

Gerade aus der Tatsache heraus, dass so viele Schüler Probleme beim Modellieren haben ergibt sich jedoch in Kombination mit der großen Bedeutung, welche der Modellierungskompetenz beigemessen wird (siehe Kapitel 1.2) ein Bedarf, Schüler schon möglichst früh angemessen zu fördern. Förderung ohne eine vorhergehende gründliche Diagnostik ist aber wiederum sinnlos und unökonomisch. Ziel der vorliegenden Arbeit ist daher die Erforschung von Wegen zur unabhängigen Diagnostik der beim Modellieren relevanten Einzelprozesse bereits im Grundschulalter.

Kapitel 1 gibt zunächst einen Überblick über die relevanten Arbeiten aus dem Bereich der Forschung zu Modellieren und Modellierungskompetenz. An dieser Stelle sei angemerkt, dass diese Arbeiten in erster Linie aus der Fachdidaktik stammen, weshalb den dort thematisierten Prozessen wie Textverstehen, Bildung mentaler Modelle etc. häufig die kognitionspsychologische Fundierung fehlt. Diese kann allerdings auch im Rahmen der vorliegenden Arbeit nicht aufgearbeitet werden, da die Zielsetzung hier vorrangig auf der Untersuchung des Modellierungsprozesses auf Makroebene liegt. Wie viel Komplexität sich hinter den scheinbar einfachsten Pro-

zessen auf der Mikroebene verbergen kann, soll hier nur anhand eines Beispiels verdeutlicht werden:
In der Grundschule beinhalten viele Aufgaben einfache Multiplikationsaufgaben, die den Abruf von Informationen zum Einmaleins aus dem Langzeitgedächtnis erfordern. Die Tatsache allein, dass es Schülern schwerer fällt, den Eintrag 7 x 8 abzurufen, als den für 3 x 7 führte zu intensiven Forschungsbemühungen, die noch längst nicht an ihrem Ende angelangt sind (Domahs, Delazer &Nuerk, 2006; Greer, 1992).

Aufbauend auf der im ersten Kapitel dargestellten vorhandenen Wissensbasis werden in Kapitel 2 die noch bestehenden Wissenslücken aufgezeigt, die Hauptfragestellungen der Arbeit abgeleitet und Möglichkeiten zu deren Operationalisierung diskutiert.

In den Kapitel 3 und 4 werden die Ergebnisse der zur Beantwortung der Fragestellung durchgeführten Studien dargestellt und diskutiert, eine zusammenfassende Diskussion und Einordnung erfolgt in Kapitel 5.

1. Theoretischer und empirischer Hintergrund

Wie in der Einleitung bereits angedeutet geht es in der vorliegenden Arbeit um die Diagnostik von Modellierungskompetenzen. Der theoretische Teil der Arbeit befasst sich daher mit den wichtigsten theoretischen und empirischen Grundlagen aus der Literatur zu Modellieren. In Abschnitt 1.1 wird zunächst eine Arbeitsdefinition von Modellieren herausgearbeitet; in Abschnitt 1.2 wird dargelegt, weshalb sich die Beschäftigung mit Modellieren überhaupt lohnt; und in Abschnitt 1.3 wird der Ablauf von Modellierungsprozessen beim mathematischen Problemlösen dargestellt.

In der Literatur wird häufig von Modellierungs*kompetenz* gesprochen, wenn es darum geht, das Fähigkeitsniveau beim Modellieren zu beschreiben. Abschnitt 1.4 widmet sich daher ausführlich der Modellierungskompetenz, worauf in Abschnitt 1.5 herausgearbeitet wird, warum ein Bedarf an zuverlässigen Diagnoseinstrumenten zur Erfassung von Modellierungskompetenzen bei Grundschulabgängern besteht. Von einem solchen Bedarf ausgehend ist es angezeigt, Wege zur Operationalisierung von Modellierungskompetenzen zu finden. Hierauf wird in Abschnitt 1.6 eingegangen.

Bei der Sichtung der Literatur fällt auf, dass sich vor allem Mathematikdidaktiker, z.T. auch Bildungsforscher, mit Modellieren und Modellierungskompetenzen beschäftigt haben. Auf der psychologisch-diagnostischen Seite fehlt bisher das Interesse an diesem Themenkreis. Vielleicht ist dies der Grund dafür, dass die wenigen empirischen Untersuchungen zumeist eher qualitativ-deskriptiv und fallstudienartig angelegt sind. Zur empirischen Überprüfung der gängigen Theorien sind sie daher nur bedingt geeignet. Eine Zielsetzung der vorliegenden Arbeit ist es daher, die von der didaktischen Seite formulierten Theorien empirisch zu überprüfen und somit einen ersten Schritt zur Schließung dieser Lücken zu unternehmen.

1.1. Modellieren und Modellbildung

In den Bildungsstandards der Kultusministerkonferenz (KMK, 2005) wird unter anderem gefordert, jeder Schüler müsse über die Fähigkeit zum Modellieren verfügen. Was aber ist darunter zu verstehen?

In der Fachdidaktik definiert man Modellieren oft dergestalt, dass die einzelnen Schritte, die beim Modellieren vonstatten gehen, aufgelistet und ausführlich erläutert werden (s.u.). Verfolgt man diesen Ansatz, so lässt sich Modellieren auffassen als der Weg, den Personen beschreiten, wenn sie ein reales Problem auf mathematischem Weg zu lösen versuchen. Besondere Bedeutung haben hierbei Übersetzungsprozesse, bei denen einerseits die reale Problemsituation in eine mathematische Sprache übersetzt und andererseits eine mathematische Lösung wiederum vor dem Hintergrund des realen Problems interpretiert wird. In der Literatur finden sich Definitionen mit unterschiedlichem Abstraktionsgrad. Barbosa (2003) betont lediglich den pädagogischen Hintergrund und schreibt: „Modelling is a learning milieu where students are invited to take a problem and investigate a situation with reference to reality in mathematics (S.230).” Modellieren aber nur als ein bestimmtes didaktisches Setting zu bezeichnen, greift sicherlich zu kurz, weshalb diese Aussage als definitorische Grundlage als unzureichend betrachtet werden muss. Von besonderer Wichtigkeit ist die Betonung der Tatsache, dass es beim Modellieren um die Verbindung zweier „Welten“ geht, dies beinhaltet z.B. die Definition von English (2003, S. 5): „Mathematical modelling is frequently viewed as the construction of a link or bridge between mathematics as a way of making sense of our physical and social world, and mathematics as a set of abstract, formal structures.”

Umstritten ist, ob man unter Modellieren neben dem Übersetzungsvorgang von der Problemsituation zum mathematischen Modell auch denjenigen Übersetzungsvorgang von der mathematischen Lösung zurück zur Problemsituation sowie die zwischengeschalteten Verarbeitungsschritte subsumieren soll. Blum (2003) unterscheidet daher zwischen Modellieren im engeren und im weiteren Sinne. Im *engeren Sinne* würde man lediglich den Schritt von einer Problemsituation zu einem mathematischen Modell als Modellieren bezeichnen, im *weiteren Sinne* hingegen den gesamten Problemlöseprozess, wie er in Abschnitt 1.3.2 beschrieben wird.

Mathematiker unterscheiden mitunter auch zwischen „Modelling“ und „Applications“. Streng genommen bezeichnet Modellieren dann lediglich (wie auch beim engeren Sinn von Blums Definition) die Verbindung von der Realität in die Mathematik. Der

Weg von der Mathematik zur realen Welt fällt unter den Begriff „Application", also die Anwendung der Mathematik auf die Wirklichkeit. Dennoch kann man auch hier Modellieren als Oberbegriff für beide Richtungen verwenden (Pollak, 2007).

Wann immer im Folgenden der Begriff Modellieren verwendet wird, ist der gesamte – noch näher zu beschreibende – Prozessablauf gemeint.

Auch die Kognitionspsychologie beschäftigt sich mit Modellen und Modellbildungen als Heurismen beim Lösen von Problemen. Modelle sind demnach (vgl. Arbinger, 1997, S. 79) Abbildungen der Wirklichkeit, die diese vereinfacht darstellen, um die Problemsituation auf wenige wichtige, lösungsrelevante Aspekte zu reduzieren. Mathematische Modelle unterscheiden sich von anderen Modellen hinsichtlich der Art, wie solche Abbildungen gestaltet sind. Bei mathematischen Modellen bestehen diese aus einer Beschreibung der Problemsituation in mathematischer Sprache, also z.B. durch Zahlen, Variablen, Operatoren und / oder Symbole.

Unter Berücksichtigung sowohl der fachdidaktischen als auch der kognitionspsychologischen Perspektive, können mathematische Modelle daher wie folgt definiert werden:

> **Mathematische Modelle sind Abbildungen der Wirklichkeit, in denen Problemsituationen abstrahiert und in eine mathematische Sprache übersetzt werden.**

Mathematisches Modellieren beschreibt demzufolge den Prozess, den Personen beschreiten, wenn sie Problemsituationen mithilfe solcher mathematischer Modelle lösen. Dies beinhaltet das Verstehen der Situation; das Übersetzen in mathematische Zeichen, Symbole und Operatoren; das Lösen mithilfe von innermathematischen Algorithmen; die Interpretation der mathematischen Lösung vor dem Hintergrund der ursprünglichen Problemsituation und letztlich die kritische Prüfung.

1.2 Die Bedeutung mathematischen Modellierens

Warum sollte man Schülern überhaupt mathematisches Modellieren lehren? Barbosa (2003) fasst die folgenden aus der Diskussion innerhalb der mathematischen Fachdidaktik hervorgegangenen Argumente hierfür zusammen: Modellierungsaufgaben

- motivieren den Lerner,
- erleichtern das Erlernen mathematischer Sachverhalte
- bereiten den Lerner darauf vor, mathematisches Wissen in unterschiedlichen Gebieten anzuwenden,
- tragen zur Entwicklung allgemeiner Kompetenzen bei und
- führen dazu, dass Lerner die soziokulturelle Rolle der Mathematik verstehen lernen.

Diese Argumente sind durchaus nachvollziehbar, es wird hierin jedoch nicht deutlich, wo diese Vorteile spezifisch für Modellierungaufgaben sind, insbesondere weil weder theoretische Überlegungen noch empirische Belege angeführt werden, um diese Argumente entsprechend zu untermauern.

Im Folgenden soll sich daher an der Rechtfertigung von Blum (2003) orientiert werden, der argumentiert, mathematisches Modellieren stelle zum einen den Grund dar, *warum* Menschen Mathematik lernen, zum anderen aber auch die Art und Weise *wie* Menschen Mathematik lernen. Oder anders ausgedrückt: Mathematik kann hilfreich dabei sein, Modellieren zu erlernen, gleichzeitig kann aber auch Modellieren hilfreich dabei sein, Mathematik zu erlernen (Ottesen, 2001). Bevor auf diese Begründungen näher eingegangen wird, soll ihnen eine Begründung aus historischer Perspektive vorangestellt werden.

1.2.1 Die historische Perspektive

Modellieren im Sinne von Modellbildung spielt innerhalb der Mathematik vor allem bei der angewandten Mathematik eine Rolle. Die angewandte Mathematik, welche seit jeher von großer Bedeutung ist, war von Anfang an verbunden mit ihren Nachbardisziplinen Physik, Astronomie und Ingenieurwissenschaften (Blum, Galbraith, Henn & Niss, 2007). Die reine Mathematik wurde ab dem 19. Jahrhundert gemeinsam mit der angewandten Mathematik gelehrt. Von nun an gab es immer wieder Trends, zeitweise zugunsten der reinen, dann wieder zugunsten der angewandten Mathematik (Kilpatrick, 1992). Einer dieser Trends war in den 60er und 70er Jahren die Bewegung „Neue Mathematik". Laut den Verfechtern dieser Bewegung sollten Kinder bereits möglichst früh lernen, logisch und abstrakt zu denken. Hieraus resultierte zum Beispiel die Einführung der Mengenlehre in den Grundschulen. Im tertiären Bildungsbereich häuften sich ungefähr zur gleichen Zeit Klagen, die im Gegenteil dazu *mehr* Anwendungsbezug im Mathematikunterricht forderten:

Im Vereinigten Königreich kamen in den späten Fünfzigern immer wieder Beschwerden aus den Führungsriegen der Industrie, dass Universitätsabsolventen trotz hervorragender Abschlüsse in Mathematik nicht dazu fähig seien, ihr Wissen einzusetzen, um anstehende reale Probleme zu lösen.

Unabhängig davon, in welche Richtung das Pendel gerade schwingt: Modellieren war immer – mal mehr, mal weniger – ein Bestandteil von Mathematik und von Mathematikunterricht (Für eine ausführliche Geschichte des mathematischen Modellierens vgl. Pollak, 2003, 2007).

Nun lässt sich zwar allein aus dieser Tatsache heraus noch keine Rechtfertigung dafür ableiten, Modellieren auch weiterhin als zentralen Part des Curriculums im Fach Mathematik beizubehalten, jedoch gilt es, Folgendes zu bedenken:

- Erstens hat der Mathematikunterricht – trotz aller Modeerscheinungen und zeitweiliger Konzentration auf die reine Mathematik – den Anwendungsaspekt mathematischen Wissens nie komplett aus den Augen verloren. Dies zeigt, dass die Fachdidaktik - und somit die auf diesem Gebiet tätigen Experten - die Wichtigkeit von Modellierungskompetenz zu jeder Zeit anerkannt haben.
- Zweitens zeigt gerade das oben beschriebene Beispiel um die britischen Studierenden die Konsequenzen einer Vernachlässigung von Anwendungsbezügen während der Ausbildung.

1.2.2 Modellieren als Mittel zum Zweck

Man kann sich nun die Frage stellen, warum Modellieren von alters her eine so große Rolle im Mathematikunterricht spielt. Um dies zu beantworten, kann man auf die von Blum (2003) angeführten Argumente zurückgreifen.

Betrachten wir zunächst die Perspektive von Modellieren als Mittel zum Zweck – als die Art und Weise wie Menschen Mathematik lernen. Was ist hiermit gemeint? Modellieren bezeichnet einen Weg, Probleme mathematisch zu lösen. Hierbei wird mathematisches Wissen angewandt und somit vertieft. Sofern bei einem Problem nicht auf bereits vorhandene Lösungsalgorithmen zurückgegriffen werden kann, sondern neue Lösungswege gefunden werden müssen, kann so auch neues Wissen erworben werden – ganz im Sinne des situierten Lernens (Greeno, Moore & Smith, 1993). Dass man mathematisches Wissen auch durch die Beschäftigung mit Alltagsproblemen erwerben kann, wird an zwei Beispielen deutlich:

- Erstens sei auf die Entwicklung von Basisfähigkeiten im Kleinkind- und Vorschulalter verwiesen, wie sie in Kapitel 1.4.2.4 dargestellt werden. Kinder erwerben durch die aktive Auseinandersetzung mit ihrer konkreten Umwelt erste Zähl-, Mengen und Rechenkenntnisse, lange bevor sie mit der abstrakten Sprache der Mathematik in Berührung kommen (Gelman, 2000). Es ist durchaus angebracht davon auszugehen, dass jedes Kind bestimmte mathematische Prinzipien nicht vermittelt bekommt, sondern für sich selbst neu erfindet (Bryant & Nuñes, 2002). Dies geschieht aber nur deshalb, weil Kinder sich realen Problemen gegenübergestellt sehen, für die sie Lösungsmöglichkeiten entwickeln müssen. Spontan greifen sie zu diesem Zwecke auf mathematische Konzepte zurück, lange bevor sie wissen, was sich hinter dem Begriff „Mathematik" überhaupt verbirgt.

- Zweitens kann man ähnliche Vorgänge beobachten, wenn man Menschen ohne - oder mit nur geringer - Schulbildung untersucht. Auch diese müssen täglich Probleme lösen, für deren Bearbeitung sie auf selbstentwickelte Algorithmen zurückgreifen müssen. Auf diese Art und Weise lernen sie zu rechnen – ohne in der Schule diesbezüglich unterrichtet worden zu sein. Dies belegen zum Beispiel die Studien von Carraher, Carraher & Schliemann (1985), in denen Straßenverkäufer untersucht wurden, die kaum formal in Mathematik unterrichtet worden waren, aber dennoch in der Lage waren, arithmetische Probleme zu lösen. In die gleiche Kategorie fallen die von Arbinger (1997) zitierten Untersuchungen an brasilianischen Buchmachern, die durch die Konfrontation mit mathematischen Problemen von sich aus eine Reihe recht komplexer Lösungsstrategien erwarben.

Beide Beispiele befassen sich zwar nicht explizit mit mathematischem Modellieren, trotzdem beschreiben sie Prozesse, wie sie auch beim mathematischen Modellieren zu beobachten sind, nämlich die Anwendung bzw. Entwicklung von mathematischen Vorgehensweisen zur Lösung realer Probleme.

Diese Prozesse scheinen bei Menschen also spontan aufzutreten (Brase, 2002). Sie auch im formalen Unterricht zu fördern, um das Lernen von Mathematik voranzubringen, ist daher naheliegend. Dass dies eine vielversprechende Strategie sein kann, zeigen auch Untersuchungen im Mathematikunterricht selbst: Dunne & Galbraith (2003) untersuchten zwei Parallelklassen, von denen eine über ein Unterrichtsjahr hinweg vorwiegend mit Modellierungsproblemen arbeitete, die andere hingegen nicht. Am Ende des Jahres zeigte die Modellierungsklasse ein deutlich besseres Verständnis für mathematische Inhalte und Prinzipien als die Parallelklasse. Zwar schränken die Autoren selbst ein, dass dies kein echtes Kontrollgruppendesign darstelle und man die Untersuchung lediglich als Fallstudie betrachten könne, dennoch können die Ergebnisse zumindest als ein Hinweis auf die positiven Auswirkungen von Modellierungsaufgaben gewertet werden.

An dieser Stelle sei angemerkt, dass in der Mathematikdidaktik ein Unterschied gemacht wird zwischen Modellierungsaufgaben und Textaufgaben, wie man sie klassischerweise aus dem Mathematikunterricht kennt, und zwar dergestalt, dass Modellierungsaufgaben als eine Teilmenge von Textaufgaben aufgefasst werden. Auch bei Modellierungsaufgaben handelt es sich um verbal präsentierte Problemstellungen, allerdings mit dem Unterschied, dass sie Verbindungen zur realen Welt enthalten, d.h. dass hier Probleme vorgegeben werden, die einem auch im wahren Leben begegnen können.

Das in 1.3.1 beschriebene Beispiel „Berechnen der Wandfläche eines Zimmers beim Renovieren“ ist daher eine Textaufgabe, die auch als Modellierungsaufgabe aufgefasst werden kann. Hingegen ist die Aufgabe

Peter ist acht, wenn er drei Jahre
älter wäre, dann wäre seine Großmutter
genau fünfmal so alt wie er.

zwar eine typische Textaufgabe, keinesfalls aber eine Modellierungsaufgabe, da sie nicht den geringsten Lebensweltbezug aufweist (Galbraith, 2007). Aus kognitionspsychologischer Sicht macht eine solche Unterscheidung hingegen keinen Sinn, da auch im Beispiel ohne Lebensweltbezug eine Modellbildung, wie sie oben definiert wurde, erfolgen muss. Für die vorliegende Arbeit soll daher der Einfachheit halber keine Unterscheidung getroffen werden zwischen Textaufgaben und Modellierungsaufgaben.

1.2.3 Modellieren als Bildungsziel

Mathematisches Modellieren kann aber statt als reines Mittel auch als Bildungs**ziel** angestrebt werden. Modellieren ist als Kompetenz erstrebenswert, da die Fähigkeit, real gegebene Probleme mathematisch zu lösen, eine bedeutende Kompetenz zur Bewältigung von solchen Problemen ist, wie sie Schülern sowohl in ihrem beruflichen als auch in ihrem späteren Leben immer wieder begegnen werden. Modellieren ist der Grund, warum Menschen Mathematik lernen sollen, ganz im Sinne von „Non scolae sed vitae discimus!“

Henry Pollak – ein Pionier auf dem Gebiet der Erforschung und didaktischen Implementierung von Modellieren – drückt dies so aus: „Society provides the time for mathematics to be taught in schools, colleges and universities not because it is beautiful, which it is, or because it provides great training for the mind, which it does, but because it is so useful.” (Crouch & Haines, 2003, S.3).

Aus diesem Grund wird dem Modellieren von Bildungsforschern eine so große Bedeutung zugeschrieben: Sie ist Bestandteil der sogenannten „Mathematical Literacy“, einem Begriff aus der Bildungsforschung, der auch den PISA-Untersuchungen zugrunde liegt. Die OECD/PISA Definition von Mathematical Literacy lautet: „Mathematische Grundbildung ist die Fähigkeit einer Person, die Rolle zu erkennen und zu verstehen, die Mathematik in der Welt spielt, fundierte mathematische Urteile abzugeben und sich auf eine Weise mit der Mathematik zu befassen, die den Anforderungen des gegenwärtigen und künftigen Lebens dieser Person als konstruktivem, engagiertem und reflektierendem Bürger entspricht“ (Baumert, Klieme, Neubrand, Prenzel, Schiefele, Schneider, Tillmann & Weiß, 2000, S.3). Weiterhin heißt es im Rahmenkonzept zu PISA: „Der Schwerpunkt [liegt] auf der funktionalen Anwendung von mathematischen Kenntnissen in ganz unterschiedlichen Kontexten und auf ganz unterschiedliche, Reflexion und Einsicht erfordernde

Weise [S.2]“. Laut Neubrand (2003) steckt hinter dieser Formulierung nichts anderes als Modellieren, etwas vorsichtiger ausgedrückt ließe sich zumindest festhalten, dass Modellieren als ein Teilbereich von mathematischer Grundbildung angesehen werden kann.

Auch im nationalen Kontext wird der Modellierungsfähigkeit eine bedeutende Rolle zugeschrieben: Die Bildungsstandards der Kultusministerkonferenz (KMK, 2005) nennen für die Grundschule fünf, für den mittleren Bildungsabschluss sechs mathematische Kompetenzen, die als erklärtes Bildungsziel eines jeden deutschen Schülers am Ende seiner Schullaufbahn zu verstehen sind. Eine dieser mathematischen Kompetenzen ist das Modellieren. Die Autoren der IGLU-Studie (Walther, Geiser, Langeheine & Lobemeier, 2003) räumen der Modellierungskompetenz ebenfalls eine tragende Rolle ein: Ihrer Meinung nach muss sich die Effektivität des Mathematikunterrichts an der Fähigkeit zu Modellieren messen lassen. Klieme, Neubrand & Lüdtke (2001) ziehen aus der Mathematikstudie von PISA 2000 das Fazit: „Im Zentrum der mathematischen Grundbildung steht die Modellierungsfähigkeit [S.186].“ Aus all dem lässt sich folgern, dass Modellieren einerseits als wichtige Fähigkeit zu betrachten ist und andererseits allgemein anerkannt wird, wie wichtig es ist, diese Fähigkeit zu besitzen.

1.3 Modellieren und Modellierungszyklus

Im vorhergehenden Teilkapitel wurde Modellieren definiert als der Ablauf einer Reihe von Teilschritten beim Lösen eines Problems. Im folgenden Abschnitt soll nun dargelegt werden, worin diese Teilschritte im Einzelnen bestehen und in welcher Reihenfolge sie ablaufen. Das Ablaufschema, welches diese Informationen liefert, wird als Modellierungszyklus bezeichnet. Dieser so beschriebene Modellierungszyklus hat sowohl *deskriptiven* als auch *präskriptiven* Charakter: er liefert zum einen eine Beschreibung des Modellierens und zum anderen eine Vorschrift, wie man beim Modellieren vorgehen soll.

In der Literatur ist eine Reihe verschiedener Varianten des Modellierungszyklus zu finden. Im letzten Abschnitt dieses Teilkapitels werden diese Varianten kurz dargestellt und diskutiert. Der vorliegenden Arbeit wird die Variante nach Blum et al. (2004) zugrunde gelegt, die auch in den PISA – Studien Verwendung fand.

1.3.1 Prozessablauf

Der Modellierungszyklus besteht aus fünf Stadien (Realsituation, Realmodell, mathematisches Modell, mathematische Ergebnisse und reale Ergebnisse) sowie aus fünf Übergangsprozessen (*Strukturieren*, *Mathematisieren*, *Verarbeiten*, *Interpretieren* und letztlich *Validieren*), die diese Stadien miteinander verbinden. Eine schematische Darstellung ist Abbildung 1 zu entnehmen, Tabelle 1 illustriert den Prozess darüber hinaus an einem einfachen kommentierten Beispiel.

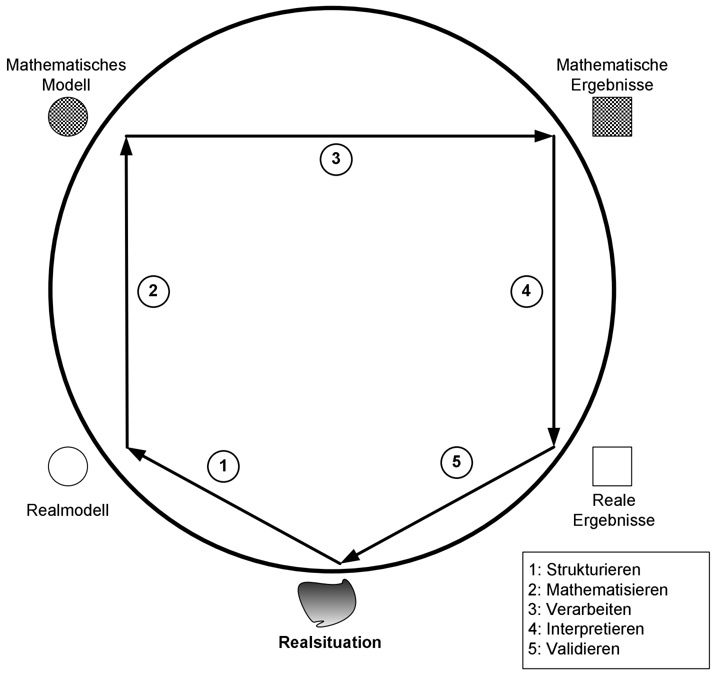


Abbildung 1: Modellierungskreislauf (angelehnt an Blum et al. 2004)

Nachfolgend werden die Stadien und die Prozesse, die diese Stadien miteinander verbinden, kurz vorgestellt. Eine ausführlichere Darstellung der Inhalte erfolgt im Kontext der Operationalisierung (siehe Kapitel 2.1.2).

<u>Das Stadium „Realsituation“:</u> Der Kreislauf beginnt typischerweise mit einer Realsituation, also einer konkret gegebenen Problemstellung. Diese kann einem im täglichen Leben begegnen, wenn man beispielsweise ein Zimmer streichen will und - um die Menge der Farbe zu bestimmen - die Wandfläche ausrechnen muss. Sie kann aber auch in schriftlicher Form, z.B. als Textaufgabe vorliegen. Die Realsituation enthält oftmals auch eine ganze Reihe an Informationen, die für die Lösung des Problems entweder irrelevant ist oder der Einfachheit halber vernachlässigt werden muss. Diese Informationen werden im nächsten Schritt „aussortiert“.

<u>Der Vorgang „Strukturieren“:</u> *Strukturieren* nennt man den Vorgang, der den Übergang von der Realsituation zum Realmodell beschreibt. Die für die Lösung des Problems relevanten Informationen werden selektiert, die wichtigsten Größen und Variablen werden identifiziert, und es wird bestimmt, wo das Problem liegt und welche Informationen noch fehlen. Das Problem wird im kognitionspsychologischen Sinne „verstanden“, d.h. beim Modellierer entsteht eine mentale Repräsentation der Situation, das Realmodell.

Das Stadium „Realmodell“: Das Realmodell als kompakte Version der Realsituation enthält nun nur noch die für das Problem relevanten Informationen. Basiert die Realsituation nicht auf einem konkreten Erlebnis des Modellierers selbst, sondern auf einer Textaufgabe, dann kann angenommen werden, dass dieser sich zu diesem Zeitpunkt ein „Bild der Situation“ gemacht hat und sich nun vorstellen kann, wie eine reale Problemsituation aussähe.

Der Vorgang „Mathematisieren“: Während beim *Strukturieren* lediglich eine Informationsreduktion realisiert wird, wird beim *Mathematisieren* zum ersten Mal ein Übersetzungsprozess in Gang gesetzt. Das Realmodell – basierend entweder auf einer tatsächlich erlebten Situation oder einer schriftlich formulierten Problemschilderung – wird übersetzt in eine symbolische Sprache, das formale mathematische Modell. Dieser Prozess ist gleichbedeutend mit Modellieren im engeren Sinn, nämlich mit der Darstellung eines realen, analogen Sachverhalts durch Zahlen, Operatoren, Symbole und Graphiken.

Das Stadium „Mathematisches Modell“: Das mathematische Modell drückt das zu lösende Problem in mathematischen Begriffen aus: Das zu streichende Zimmer im Renovierungsbeispiel ist als ein Würfel (ohne Boden) abgebildet worden, Sachverhalte die mit Abnahmen und Zunahmen, Vermehrungen und Aufteilungen zu tun haben sind durch mathematische Operatoren (+, - , · , :) dargestellt, Anzahlen sind in arabischen Ziffern angegeben und Zusammenhänge sind in Funktionen ausgedrückt (z.B. die Vermehrung eines Kapitals bei der Zinsrechnung mit der Wachstumsfunktion ($K_n = K_0 \cdot (1 + p/100)^n$). Je nach gewünschtem Detailliertheitsgrad und abhängig von der mathematischen Kompetenz des Modellierers kann ein mathematisches Modell mehr oder weniger komplex sein. Dies lässt sich wiederum am o. g. Renovierungsbeispiel illustrieren:

Beispiel: Renovieren eines Raumes – Berechnung der benötigten Wandfarbe

Schritt 3: Mathematisieren

Ein Raum ist 5 m lang, 2.80 m breit und 2.50 m hoch. Er hat auf einer der langen Seiten zwei Fenster, die jeweils 1.50 m hoch und 90 cm breit sind. Außerdem hat der Raum auf einer der schmalen Seiten eine Tür, die 1 m breit und 2 m hoch ist. Dem Prospekt für die Farbe ist zu entnehmen, dass pro Quadratmeter 0.20 l Farbe benötigt werden.

Ein einfaches Modell würde Fenster und Türen ignorieren und die Menge der Farbe (V) einfach nur anhand der Größe der Wandflächen berechnen:

$V\,[l] = 0.2\,(2 \cdot 5.00 \cdot 2.50 + 2 \cdot 2.80 \cdot 2.50)$

Wollte man genauer vorgehen, um möglichst wenig Farbe übrig zu haben, so würde man ein detaillierteres Modell aufstellen, das die Flächen von Türen und Fenstern entsprechend berücksichtigt:

$V[l] = 0.2\,(2 \cdot 5.00 \cdot 2.50 + 2 \cdot 2.80 \cdot 2.50 – 2 \cdot 1.50 \cdot 0.90 – 1.00 \cdot 2.00)$

Das Aufstellen des mathematischen Modells wird oft als der kritische Punkt beim Modellieren betrachtet. Dies mag insbesondere daran liegen, dass an diesem Punkt des Prozessmodells ein größerer kognitiver Aufwand vermutet wird als an anderen Stellen (Crouch & Haines, 2004).

<u>Der Vorgang „Verarbeiten“:</u> Beim *Verarbeiten* geht es darum, vom mathematischen Modell zu einer mathematischen Lösung zu gelangen, also um das Ausführen von mathematischen Operationen, Algorithmen, etc. Im einfachsten Falle bedeutet dies lediglich, etwas auszurechnen, wie im Renovierungsbeispiel den Term „$0.2\,(2 \cdot 5.00 \cdot 2.50 + 2 \cdot 2.80 \cdot 2.50)$“, und die Lösung zu notieren. Unter Umständen erfordert dieser Vorgang einige Zwischenschritte. Betrachtet man hingegen komplexere Probleme, so müssen an dieser Stelle Gleichungssysteme gelöst oder Approximationen gefunden, Funktionen diskutiert, Computersimulationen durchgeführt werden etc.

Für die in der hier anvisierten Zielgruppe von Grundschulabsolventen relevanten Aufgaben und auch für die meisten alltagsrelevanten Situationen beschränkt sich das *Verarbeiten* auf das Anwenden von bekannten Regeln und Algorithmen. Die

Ausführung dieses Schrittes verlangt keinen Bezug zur Realität, es erfolg eine rein innermathematische Bearbeitung des Problems. Am Ende des innermathematischen Problemlöseprozesses steht das mathematische Ergebnis.

Das Stadium „Mathematische Ergebnisse“: Mathematische Ergebnisse sind die Endprodukte des innermathematischen Problemlöseprozesses. Diese sind nach wie vor in der formalen mathematischen Sprache ausgedrückt, bei einfachen Aufgaben also in Form von arabischen Ziffern. Im Renovierungsbeispiel wären dies die Zahl 7.8 für das einfachere Modell und die Zahl 6.86 für das Modell, welches Fenster und Tür berücksichtigt. Diesen Ergebnissen wird zunächst noch keine Bedeutung zugeschrieben, diese erhalten sie erst durch das *Interpretieren*.

Der Vorgang „Interpretieren“: Das *Interpretieren* der mathematischen Ergebnisse übersetzt die mathematischen in reale Ergebnisse. Den reinen Zahlen, die am Ende des mathematischen Lösungsprozesses stehen, wird wieder ein Sinn verliehen, indem sie in einen Zusammenhang mit dem Ursprungsproblem gesetzt werden. Die Zahl 7.8 aus dem Beispiel wird gelesen als „7.8 Liter“. *Interpretieren* kann als der spiegelbildliche Prozess zu *Mathematisieren* betrachtet werden: Von der symbolischen, mathematischen Sprache erfolgt eine Rückübersetzung in die anschauliche Alltagssprache. Somit ist dies der zweite Übersetzungsvorgang, den das mathematische Modellieren erfordert. Der Kreis im Prozessmodell beginnt sich zu schließen, denn man ist zurückgelangt in die reale Welt und zu den realen Ergebnissen.

Das Stadium „Reale Ergebnisse“: Im letzen Stadium des Modellierungsprozesses liegen dann die realen Ergebnisse des Problemlöseprozesses vor. Im Beispiel: Es wurde ausgerechnet, dass – nimmt man wieder das einfachere Modell – fast 8 Liter Farbe benötigt werden. Dem realen Ergebnis kann der Modellierer nun wieder ein klares Vorstellungsbild zuordnen: 8 Liter Farbe sind etwas weniger als der Inhalt eines 10-Liter-Eimers.

An sich könnte man das Problem an dieser Stelle als gelöst betrachten und – auf das Beispiel bezogen – einen Eimer Farbe kaufen gehen. Was aber, wenn dem Modellierer während des Prozesses ein Fehler unterlaufen ist? Um zumindest grobe Fehler ausschließen zu können, muss noch ein weiterer Schritt folgen, der den Kreis zur Realsituation endgültig schließt: Das *Validieren*.

Der Prozess „Validieren“: *Validieren* heißt, die realen Ergebnisse in den Kontext der Realsituation zurückzuübertragen und sich die Frage zu stellen, ob die Ergebnisse im Zusammenhang mit der Realsituation eine adäquate und realistische Lösung für das ursprünglich formulierte Problem darstellen. Dies lässt sich wiederum am Renovierungsbeispiel verdeutlichen.

Beispiel: Renovieren eines Raumes – Berechnung der benötigten Wandfarbe

Schritt 5: Validieren

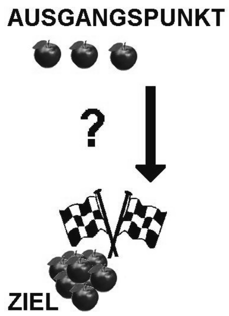
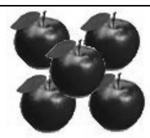
Angenommen, dem Modellierer sei in Schritt 2 ein Fehler unterlaufen: Irritiert von der Aussage „0.2 Liter Farbe pro Quadratmeter“ begeht er den Fehler, nicht die Fläche der Wände zu berechnen, sondern stattdessen die Grundfläche des Zimmers. Diese beträgt 14 m². Daraus ergäbe sich ein Bedarf an 2.8 Litern Farbe.

Dieses Ergebnis nimmt der Modellierer idealerweise aber nicht einfach hin, sondern stellt sich die Menge von 2.8 Litern Farbe vor. Vielleicht vergleicht er sie im Geiste mit dem großen 10-Liter-Eimer, den er für das Nebenzimmer benötigt hat, das nur unwesentlich größer ist. Daraufhin treten Zweifel bei ihm auf, und er überprüft noch einmal sein Modell.

Nicht immer allerdings ist die Validierung erfolgreich in dem Sinne, dass vollzogene Fehler auch tatsächlich aufgedeckt werden. Ein Verzicht auf diesen Kontrollprozess bedeutet jedoch gleichzeitig ein Risiko für suboptimale oder falsche Lösungen. Dies ist insbesondere dann ein gravierender Nachteil für den Problemlöser, wenn er nicht mangels Kompetenz, sondern aufgrund von „Flüchtigkeitsfehlern“ zu einem falschen Ergebnis gelangt ist.

Treten Zweifel an der Richtigkeit der Ergebnisse auf, so muss der Kreislauf von neuem durchlaufen werden. Auch wenn dies bei einfachen und alltäglichen Problemen wohl eher die Ausnahme darstellen dürfte, sind bei komplexen Vorgängen der angewandten Mathematik mehrere Durchgänge durchaus üblich. Dies liegt darin begründet, dass es bei komplizierteren Problemen selten nur eine richtige Lösung gibt. Oftmals gibt es sogar überhaupt keine eindeutige Lösung und die mathematischen Modelle sind nur bessere oder schlechtere Annäherungen an eine Ideallösung. Halten diese Annäherungen dem Vergleich mit der Realsituation nicht stand, müssen neue Lösungswege gesucht werden. Aufgrund dieser sich wiederholenden Durchläufe spricht man vom Modellierungs*zyklus*.

Tabelle 1: Beispielhafter Prozessablauf

Stadien *Übergänge*	Beispiel	Erläuterung
Realsituation	Anna und Peter möchten einen Apfelkuchen backen. Neben den Zutaten, die sie bereits zuhause haben, benötigen sie noch 500g Mehl und 8 Äpfel. Drei Äpfel und eine Tüte Mehl leihen sie sich bei der Nachbarin. Wie viele Äpfel müssen sie noch organisieren, um den Kuchen backen zu können?	Gegeben ist eine Situation aus dem täglichen Leben.
	Strukturieren	
Realmodell		Für das Problem unbedeutende Informationen (z.B. 500g Mehl) wurden ignoriert. Es wurde konkretisiert, was die eigentliche Aufgabe darstellt: Dass drei Äpfel gegeben sind, man am Ende acht Äpfel haben möchte und nun vor der Frage steht, wie viele weitere noch hinzukommen müssen.
	Mathematisieren	
Mathematisches Modell	3 + ? = 8 oder 8 – 3 = ?	Das mathematische Modell drückt den im Realmodell vereinfacht dargestellten Sachverhalt in einer mathematischen Sprache aus. Zahlwörter sind durch Zahlenwerte ersetzt, die Problemstellungen „wie viele […] organisieren“ ist durch Operatoren dargestellt (plus wie viel bzw. minus).
	Verarbeiten	
Mathematische Ergebnisse	5	Am Ende des innermathematischen Lösungsprozesses des Problems steht ein in mathematischer Sprache formuliertes Ergebnis.
	Interpretieren	
Reale Ergebnisse	 ODER „Anna und Peter müssen noch fünf Äpfel zusammentragen.“	Das reale Ergebnis ist wieder als Konkretisierung (siehe Bild) vorstellbar oder als Aussage formulierbar.
	Validieren	

1.3.2 Der Modellierungszyklus als Präskription und Deskription von Modellierungsprozessen

Modelle – so wurde oben definiert – sind vereinfachte, auf wesentliche Merkmale reduzierte Darstellungen der Wirklichkeit, wobei zentrale Merkmale und Relationen, die in der Wirklichkeit gelten, auch im Modell enthalten sind. In diesem Sinn ist auch der Modellierungszyklus selbst ein Modell, nämlich ein Modell der Vorgehensweise beim Lösen von Modellierungsaufhaben.

Theoretisch kann der Modellierungszyklus einerseits als präskriptives, andererseits auch als deskriptives Modell von Modellierungsprozessen aufgefasst werden. Fraglich bleibt jedoch, ob diese Sichtweise auch einer empirischen Überprüfung standhält.

Ein präskriptives oder auch normatives Modell postuliert, wie ein Prozess ablaufen muss. Es macht Aussagen darüber, unter welchen Bedingungen die Ergebnisse des Prozesses optimal sind. Diese Bedingungen sind konkrete Regeln, die beschreiben, welche Aktionen in welcher Reihenfolge auszuführen sind, um eine möglichst hohe Qualität des Endproduktes zu gewährleisten. Dies heißt nicht, dass diese Regeln in der Praxis auch eingehalten werden.

Eignet sich der Modellierungszyklus als präskriptives Modell? In der Fachliteratur wird dies nicht in Frage gestellt: Ein (ggf. mehrfaches) Durchlaufen der Stadien wird als die beste Methode verstanden, um reale Probleme mathematisch zu lösen.

Ob ein am theoretisch postulierten Modellierungszyklus orientiertes Abschreiten der Einzelphasen aber tatsächlich der beste Weg ist, um Modellierungsprobleme erfolgreich zu lösen, muss jedoch in der Praxis erst noch erwiesen werden bevor der Modellierungszyklus als normatives Modell akzeptiert werden kann.

Ebenso wichtig ist aber die Frage, ob der Modellierungszyklus auch als deskriptives Modell fungieren kann. Ein deskriptives Prozessmodell beschreibt einen Ablauf so, wie er tatsächlich bei allen oder den meisten Personen abläuft. Wäre der Modellierungszyklus ein rein theoretisches Modell, welches das Vorgehen von Problemlösern nicht adäquat beschreiben kann, so wäre sein praktischer Nutzen in Frage zu stellen.

Borromeo Ferri und ihre Kollegen beschäftigten sich daher in ihrem Projekt „KOM²: „Kognitionspsychologische Analysen von Modellierungsprozessen im Mathematikunterricht“ (Borromeo Ferri, 2006; Borromeo Ferri & Kaiser, 2006; Borromeo Ferri, Leiss & Blum, 2006) mit der Frage, ob sich der Modellierungskreislauf mit seinen Unterteilungen, wie sie in der Literatur beschrieben werden, auch tatsächlich in der Herangehensweise beim Lösen von Modellierungsaufgaben widerspiegelt. Die Autoren filmten 38 Schülergruppen à zwei Personen, um anschließend die individuellen Modellierungsrouten zu beschreiben. Diese Beschreibung konzentrierte sich jedoch primär auf eine detaillierte Darstellung dessen, was in den jeweiligen Einzelphasen passiert. Informationen beispielsweise über Abweichungen von der theoretisch postulierten Reihenfolge sind den Berichten nicht zu entnehmen. Trotzdem kommen die Autoren zu dem Fazit: Der Modellierungszyklus stellt ein gutes Modell der kognitiven Vorgänge beim Modellieren dar, auch wenn nicht alle Schritte dem Modellierer bewusst sind.

Eine weitere Erkenntnis der Untersuchung ist, dass es individuelle Unterschiede bei der Art und Weise gibt, wie Modellierungsaufgaben bearbeitet werden. Die Unterschiede werden von den Autoren auf unterschiedliche mathematische Denkstile zurückgeführt, welche die Art der mentalen Repräsentation (verbal vs. bildhaft) von Informationen betreffen. Darüber hinaus sind jedoch auch weitere – bisher eventuell noch unbekannte – Einflussgrößen denkbar.

Festzuhalten bleibt, dass die in der Theorie zu Modellieren und Modellierungskompetenzen implizierten Grundgedanken vom Modellierungszyklus als deskriptives und präskriptives Modell bisher nicht ausreichend untersucht wurden. Bedenkt man die vielen praktischen Implikationen, die aus dem Modellierungszyklus abgeleitet werden, ist der Prüfung beider theoretischer Konzepte daher höchste Priorität einzuräumen.

1.3.3 Varianten des Modellierungszyklus

Neben der oben diskutierten Version des Modellierungszyklus finden sich in der Literatur weitere Versionen. Unterschiede zum oben beschriebenen Modell beziehen sich hierbei entweder auf den Beginn oder auf das Ende des Kreislaufs. Houston (2007) verwendet beispielsweise ein Modell, das auf Penrose (1978) zurückgeht und am Ende des Modellierungszyklus nach dem Validierungsschritt eine Verzweigung vorsieht zwischen einerseits „revise“ (falls bei der Validierung Zweifel an der Richtigkeit der Lösung aufgekommen sind) und andererseits „report“ (sobald eine Lösung als endgültig akzeptiert worden ist und die Berichtlegung eingeleitet werden kann).

Viel interessanter sind jedoch Varianten bezüglich des Kreislaufbeginns, wie Borromeo Ferri (2006) sie in einem Überblicksartikel zusammenfasst. Abbildung 2 gibt einen Überblick über die dort referierten Modellgruppen, auf die im Folgenden kurz eingegangen werden soll. Wie aus der Darstellung ersichtlich wird, fällt das oben beschriebene Modell nach Blum et al. (2004) in die Gruppe 3.

Gruppe 1						
RS	→	SM/MRS	→	RM	→	MM
Gruppe 2						
		SM/MRS	→	RM	→	MM
Gruppe 3						
RS	→			RM	→	MM
Gruppe 4						
RS	→					MM

RS = Realsituation, SM/MRS = Situationsmodell, RM = Realmodell, MM = mathematisches Modell

Abbildung 2: Gruppen von Modellierungsmodellen (Borromeo Ferri, 2006)

Modelle der Gruppe 1 postulieren drei Zwischenschritte von der Realsituation zum mathematischen Modell: Die Realsituation (RS) wird zunächst in einem Situationsmodell (SM), also in einer mentalen Repräsentation der Situation, vom Modellierer verinnerlicht und verstanden. Dieses Situationsmodell ist noch nicht – wie dies später beim Realmodell (RM) der Fall sein wird – von überflüssigen Informationen befreit und vereinfacht. Es handelt sich hierbei lediglich um eine Art mentaler Kopie der Situation. Ein derart differenziertes Modell ist laut Borromeo Ferri vor allem dann indiziert, wenn man sich besonders für die kognitiven Prozesse während des Modellierens interessiert. Dann kann es sinnvoll sein, den Schritt, den wir bisher als „*Strukturieren*" bezeichnet haben, in zwei Einzelkomponenten zu zerlegen und diese getrennt zu untersuchen. Blum & Leiss (2007) bezeichnen das Situationsmodell als einen besonders zentralen Schritt, da an dieser Stelle deutlich wird, ob Individuen das ihnen gestellte Problem auch tatsächlich verstanden haben.

Modelle der dritten Gruppe ignorieren hingegen den Zwischenschritt und gehen von der Realsituation direkt über zum Realmodell. Diese Modellvariante liegt auch den PISA-Studien zugrunde (PISA-Konsortium-Deutschland, 2001, 2004) und ist als eine Vorgängervariante zu den Modellen der Gruppe 1 zu betrachten, da in der Fachdidaktik das Interesse an kognitiven Vorgängen beim Modellieren noch verhältnismäßig neu ist.

Die Vereinfachung in Gestalt von Modellen der Gruppe 2 hingegen begründet sich wie folgt: Diese Modelle sind dann besonders praktikabel, wenn man sich mit Textaufgaben befasst. Die Argumente Borromeo Ferris (2006) lassen sich nachvollziehen, dass Modellierer beim Umgang mit Textaufgaben an sich keiner wirklichen „Realsituation" ausgesetzt sind. Textaufgaben beinhalten oftmals bereits eine so stark vereinfachte Situationsdarstellung, dass diese direkt in ein mentales Modell übernommen werden kann. Die Autorin weist zurecht darauf hin, dass hier sogar zeitweise eine umgekehrte Reihenfolge im Modellierungskreislauf heraufbeschworen werden kann – die Modellierer sehen sich einem Situationsmodell gegenüber, zu dem sie sich erst eine zugehörige Realsituation vorstellen müssen.

Modelle der vierten Gruppe sind zu stark vereinfacht, um den Anforderungen an den normativen und deskriptiven Charakter eines Modellierungsmodells gerecht zu werden. Dies ist aber auch nicht ihre Zielsetzung, denn sie haben in erster Linie eine didaktische Funktion: Einige Autoren (vgl. z.B. Maaß, 2004) argumentieren nämlich, zur Modellierungskompetenz gehöre auch Wissen über den idealtypischen Ablauf des Modellierungszyklus. Solches Wissen erleichtere dem Lernenden metakognitive Prozesse der Planung während des Modellierens. Um jedoch gerade auch jüngeren Schülern die Prinzipien und Abläufe beim Modellieren verständlich zu machen, ist auf komplizierte Ausdifferenzierungen möglichst zu verzichten und der Modellierungszyklus kompakt und verständlich darzustellen.

Welche Variante ist nun zu bevorzugen? Borromeo Ferri (2006) zieht die Möglichkeit in Betracht, je nach Zielsetzung verschiedene Prozessmodelle heranzuziehen. Sie unterscheidet zwischen zwei Betrachtungsweisen: Eine aus Sicht der Forschung und eine aus Sicht der Didaktik. Dies ist durchaus überzeugend. Geht es darum, die Prozesse beim mathematischen Modellieren genau zu untersuchen, so sollten möglichst alle Subprozesse in Betracht gezogen werden. Für die Forschung ist daher ein möglichst differenziertes Modell der Gruppe 1 indiziert. Anders für pädagogisch-didaktische Zwecke, wenn es darum geht, Wissen über Modellieren zu vermitteln. Dann sind einfache und prägnante Modelle der Gruppe 4 zu wählen.

Von anderen Autoren wird aber eher die gegensätzliche Sichtweise vertreten, gerade für didaktische Zwecke sehr ausgefeilte Modellierungsmodelle zu verwenden, die dann im Sinne eines normativen Modells den Lernenden konkrete Handlungsanweisungen geben. Ein solches Modell ist das von Dunne & Galbraith (2003), welches neben den fünf bereits bekannten Prozessen *Strukturieren*, *Mathematisieren*, *Verarbeiten*, *Interpretieren* und *Validieren* einen weiteren Schritt enthält, der zwischen *Strukturieren* und *Mathematisieren* anzusiedeln ist: „List all data and assumptions being made.“ Dies ist aber an sich nicht als ein substanziell neuer Schritt zu betrachten, vielmehr handelt es sich hier um eine Hilfestellung, die dem Modellierer die Aufstellung des mathematischen Modells erleichtern soll. Noch einen Schritt weiter geht die „How to Model Mathematically Table“ von Geng (2003): Das dort beschriebene Flussdiagramm richtet sich an Lernende und gibt für jede Stufe im Modellierungszyklus eine Checkliste mit Aktivitäten vor, die jeweils abzuarbeiten sind.

Nun gibt es aber neben der Forschungsperspektive und dem didaktischen Blickwinkel noch eine dritte Betrachtungsweise, die in der Literatur kaum angesprochen wird: Nämlich die diagnostische Sichtweise, die für die vorliegende Arbeit im Fokus steht. Diagnostik von Modellierungskompetenz sollte möglichst genau sein, damit Stärken und Schwächen der Lernenden präzise aufgedeckt und in der Folge gezielte Fördermaßnahmen eingeleitet werden können. Andererseits muss Diagnostik aber auch sparsam sein, um den Probanden nicht zu überfordern, und sollte sich nur auf solche Prozesse konzentrieren, die im Hinblick auf eine valide Diagnose auch operationalisierbar sind.

Als Grundlage für die Entwicklung von diagnostischen Instrumenten werden Modelle der dritten Gruppe vorgeschlagen. Sicherlich wären Modelle der ersten Gruppe wegen ihres hohen Differenzierungsgrades noch wünschenswerter, da eine Aufspaltung des Strukturierungsschrittes weitere wertvolle Informationen über die

Verstehensprozesse beim Modellierer liefern könnte. Auch sind durchaus Wege denkbar, mentale Repräsentationen der Beobachtung zugänglich zu machen. Derartige Untersuchungen sind jedoch im Rahmen eines ökonomischen Testverfahrens, das auch alle anderen Stadien im Modellierungszyklus erfassen soll, nicht praktikabel. Dies gilt insbesondere im Hinblick auf eine schnelle, standardisierte Auswertung. Außerdem kommt bei Instrumenten, die mit Textaufgaben arbeiten, die bereits angedeutete Verwischung der Grenzen zwischen Realsituation und Situationsmodell hinzu, die eine künstliche Trennung der beiden Stadien zusätzlich erschwert. Im Übrigen gilt für alle dargestellten Varianten, dass es sich hierbei keineswegs um verschiedene Theorien des Modellierens handelt. Über den Ablauf der verschiedenen Teilprozesse herrscht Einigkeit, es bestehen lediglich Unterschiede in der Darstellungsweise derselben Theorie. Auch Autoren, die selbst einfache Modelle verwenden, leugnen nicht die Existenz der von ihnen „ausgeblendeten“ Phasen.

Modellieren und Modellierungszyklus wurden nun dargestellt und erläutert. Der Modellierungszyklus bildet eine Orientierungsgrundlage bei der Beobachtung und Bewertung von Modellierungsprozessen. Die Frage, inwieweit er sich zur Präskription und Deskription tatsächlich auftretender Stadien, Prozesse und deren Reihenfolge eignet, wird an anderer Stelle noch einmal aufgegriffen.

1.4 Modellierungskompetenz

Setzt man sich zum Ziel, Modellierungskompetenz zu untersuchen, so muss man sich zunächst darüber klar werden, was Modellierungskompetenz ist. Zerlegt man den Ausdruck in seine Bestandteile, so müsste die Antwort lauten: Modellierungskompetenz bezeichnet die Kompetenz zu Modellieren. Was aber ist Kompetenz? Betrachtet man Modellieren als einen Teil mathematischer Kompetenz, so stellt sich hingehen die Frage, was wiederum mathematische Kompetenz ausmacht. Im Folgenden soll daher ein Verständnis von Modellierungskompetenz hergeleitet werden, indem zunächst Definitionen von Kompetenz, dann solche von mathematischer Kompetenz und letztlich Modellierungskompetenz betrachtet werden.

1.4.1 Der Kompetenzbegriff

Der Begriff „Kompetenz“ ist ein in den Bildungswissenschaften häufig benutzter Terminus. Leider ist der Begriff jedoch ebenso unscharf wie populär. Oft wird er verwendet, ohne dass deutlich wird, was unter Kompetenz überhaupt zu verstehen ist.

Dies liegt nicht zuletzt darin begründet, dass mindest drei verschiedene Sichtweisen auf den Kompetenzbegriff unterschieden werden können, wie sie bei Klieme, Hartig & Rauch (2008) diskutiert werden.

- Aus der Sprachpsychologie (Chomsky, 1968) kennt man den Begriff Kompetenz für das den sprachlichen Fähigkeiten zugrunde liegende kognitive System.
  Hierbei wird ebenso wie auch in der Lernpsychologie (Spada, 2005) unterschieden zwischen Kompetenz und Performanz. Kompetenz ist der Auffassung Chomskys zufolge nicht direkt beobachtbar: „…linguistic competence […] underlies behavior but […] is not realized in any direct or simple way in behaviour.” (Chomsky, 1986; S. 4). Ähnlich wie bei anderen Konstrukten (bspw. Intelligenz) wird Kompetenz äquivalent zu latenten Eigenschaften aufgefasst, wohingegen die Performanz die Manifestation der Kompetenz auf der Verhaltensebene darstellt. Die Messung der Kompetenz (s.u.) kann dann realisiert werden, indem mithilfe psychometrischer Modelle von der Performanz beim Lösen von Items Rückschlüsse auf den Ausprägungsgrad der Kompetenz gezogen werden (Beck & Klieme, 2007).

- Eine andere Perspektive ist diejenige von Kompetenzen als Kompromiss zwischen Bildung (verstanden als Persönlichkeitsentwicklung und Befähigung zur Teilnahme an der menschlichen Kultur) und Qualifizierung (verstanden als Aufbau von Wissen und Fähigkeiten, die benötigt werden um bestimmte Tätigkeiten auszuführen).

  Die Notwendigkeit zu einem solchen Kompromiss liegt im Wandel der Bildungsziele begründet. Während diese früher über Inhalte definiert waren (materiale Bildungsziele), geht der Ruf immer mehr nach allgemeinen Fähigkeiten, nach sog. „life skills“ wie z.B. allgemeine Problemlösefähigkeit (formale Bildungsziele). Die Vermittlung reiner Wissensinhalte ist nicht mehr praktikabel, da heute Gelerntes morgen bereits überholt sein kann. „Life skills“ (Lebenskompetenzen) und „Bildung“ sind jedoch derart abstrakte Begriffe, dass sich nur schwer konkrete Unterrichtsinhalte daraus ableiten lassen. Einen Mittelweg bilden Kompetenzen: Sie werden betrachtet (z.B. Klieme, 2004) als inhaltsübergreifend, zugleich aber immer noch anforderungs- und situationsbezogen. Zwar sind sie auf bestimmte Kontexte und Situationen bezogen, innerhalb dieser aber dennoch verallgemeinerbar und daher durchaus als allgemeine Dispositionen zu betrachten.

- Die funktionale Sichtweise der Kompetenz (McClelland, 1973) betont vor allem die Verbindung zum „richtigen Leben“. Kompetenzen werden betrachtet als „realized abilities“, also als die in die Tat umgesetzten Fähigkeiten, mit Anforderungen des Lebens in bestimmten Situationen fertig zu werden.

Welcher Sichtweise soll man sich nun anzuschließen? Von der funktionalen Betrachtungsweise soll im Folgenden Abstand genommen werden, da hier eine Vermischung der Konzepte „Kompetenz“ und „Performanz“ erfolgt, welche der begrifflichen Klarheit abträglich ist.

Ähnliches gilt für die Definition von Weinert (2001, S.27), der Kompetenz definiert als „die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können.“

Motivationale, volitionale und soziale Aspekte spielen – wenn man zwischen Kompetenz und Performanz unterscheidet – aber vor allem bei der Performanz eine Rolle. Zieht man wiederum den Vergleich zu latenten Konstrukten und deren Manifestation auf der Verhaltensebene zu Rate, so lässt sich das Problem am Beispiel der Intelligenz verdeutlichen: Eine Person, die in früheren Messungen einen Prozentrang von 97 erreicht hat, bei der die *latente* Eigenschaft „Intelligenz" also hoch ausgeprägt ist, kann dennoch in einem in einer bestimmten Situation durchgeführten Intelligenztest (also auf der *manifesten* Ebene) durch solche motivationalen oder volitionalen Situationseinflüsse, eine schlechte Testleistung zeigen. Dies ändert jedoch nichts an der Tatsache, dass die Person nach wie vor hochintelligent ist.

Ähnlich lässt sich dies z.B. bei mathematischer Kompetenz beobachten: Eine Person kann durchaus eine hoch ausgeprägte Kompetenz zum Lösen verschiedenartigster mathematischer Probleme haben, auch wenn die Performanz eventuell aufgrund motivationaler Einschränkungen suboptimal bleiben kann.

Aufgrund der oben ausgeführten Betrachtungsweisen soll Kompetenz im Folgenden aufgefasst werden als ein latentes Fähigkeitskonstrukt zum Lösen von bestimmten Problemen, welches durch Beobachtung und Messung der Performanz erfassbar gemacht werden kann. Im Gegensatz zur allgemeinen Problemlösefähigkeit ist Kompetenz anforderungsbezogen zu verstehen, es gilt also immer zu fragen: „Kompetenz wozu?"

1.4.2 Mathematische Kompetenz

Nachdem dargelegt wurde, wie der Begriff Kompetenz allgemein zu verstehen ist, soll speziell auf die mathematische Kompetenz eingegangen werden. Diese soll nun zunächst definiert werden, bevor Ansätze zu ihrer Klassifikation dargestellt werden. Zuvor ist darüber hinaus ein kurzer Überblick über ihre Entwicklung zu geben.

1.4.2.1 Definition

Wie kann mathematische Kompetenz aufgefasst werden? Man könnte nun minimalistisch vorgehen und die oben formulierte Definition von Kompetenz anpassen, indem der Ausdruck „bestimmte Probleme" durch „mathematische Probleme" präzisiert wird. Dies ist eine erste Definition von mathematischer Kompetenz. Niss (2004, S. 120) beschreibt mathematische Kompetenz als die "[…] ability to understand, judge, do and use mathematics in a variety of intra- and extramathematical contexts and situations in which mathematics play or could play a role." Auch in dieser Definition findet man die oben angedeutete Situationsbezogenheit von Kompetenzen. Vorsichtigerweise sollte man vielleicht statt von intra- *und* extramathematischen von intra- *und/oder* extramathematischen Kontexten sprechen. Andernfalls würde dies heißen, dass „reine" (im Sinne von nicht angewandter Mathematik) Mathematiker der Definition zufolge keine mathematische Kompetenz besäßen. Durch die Betonung extramathematischer Kontexte wird deutlich, dass zur mathematischen Kompetenz auch gehört, reale Probleme mathematisch zu lösen – also modellieren zu können. Modellierungskompetenz kann daher als eine von mehreren Teilkomponenten mathematischer Kompetenz betrachtet werden. Welches sind aber diese anderen Komponenten?

Wenn wir fragen: „Kompetenz wozu?" und dann bei der mathematischen Kompetenz darauf antworten: „Kompetenz Mathematik zu betreiben", dann bleibt die Frage offen, was es denn eigentlich heißt, Mathematik zu betreiben. Anders ausgedrückt: welche Fähigkeiten machen mathematische Kompetenz eigentlich aus?

1.4.2.2 Klassifikationen mathematischer Teilkompetenzen

Für Blomhoj & Jensen (2007) reicht es nicht aus, *die* mathematische Kompetenz zu definieren. Sie spezifizieren acht mathematische Teilkompetenzen, wobei vier sich damit beschäftigen, Fragen über die Mathematik, mit Mathematik und innerhalb der Mathematik zu stellen und zu beantworten. Vier weitere Teilbereiche beschäftigen sich damit, mit mathematischer Sprache und Werkzeugen umzugehen (siehe Tabelle 2). Zu der ersten Gruppe gehört auch die Modellierungskompetenz.

Zu kritisieren ist allerdings die Tatsache, dass die hier formulierten Teilkompetenzen recht vage formuliert sind und allein ihre Auflistung nicht ausreicht, um sich ein Bild davon zu machen, was mathematische Kompetenz im Einzelnen beinhaltet.

Tabelle 2: Mathematische Kompetenzen nach Blomhoj & Jensen (2007)

To ask and answer in, with and about mathematics:
• reasoning competence • modelling competence • problem tackling competence • mathematical thinking competence
To deal with mathematical language and tools:
• representing competence • symbol and formalism competence • communicating competence • aids and tools competence

Eine andere Klassifikation mathematischer Kompetenzen findet sich in den Bildungsstandards der Kultusministerkonferenz für das deutsche Schulsystem (KMK, 2005; Köller, 2005). Diese formulieren Standards, die alle Schüler am Ende ihrer schulischen Laufbahn erreichen sollen. Die Standards im Fach Mathematik untergliedern mathematische Fähigkeiten auf zwei Dimensionen in fünf bzw. sechs qualitativ distinkte Bereiche (siehe Abbildung 3). Auf der inhaltlichen Seite sind dies die mathematischen Leitideen (Zahl, Messen, Raum und Form, Funktionaler Zusammenhang sowie Daten und Zufall), die verschiedene Themengebiete der Mathematik betreffen.[2] Auf der anderen Seite stehen diesen Leitideen so genannte mathematische Kompetenzen gegenüber, die unterschiedliche Fertigkeiten formulieren, mit denen im Rahmen der Leitideen gearbeitet wird (Argumentieren, Problemlösen, Modellieren, Darstellungen verwenden, mit symbolischen, formalen und technischen Elementen umgehen sowie Kommunizieren). Somit enthält jedes mathematische Problem mindestens zwei Komponenten: Eine inhaltliche Komponente und eine Komponente, die darüber Auskunft gibt, in welcher Weise die Inhalte zu verarbeiten sind. Die Taxonomie der Bildungsstandards lässt sich in Einklang bringen mit derjenigen von Blomhoj & Jensen (2007), allerdings ergänzt um konkrete mathematische Anwendungsbereiche.

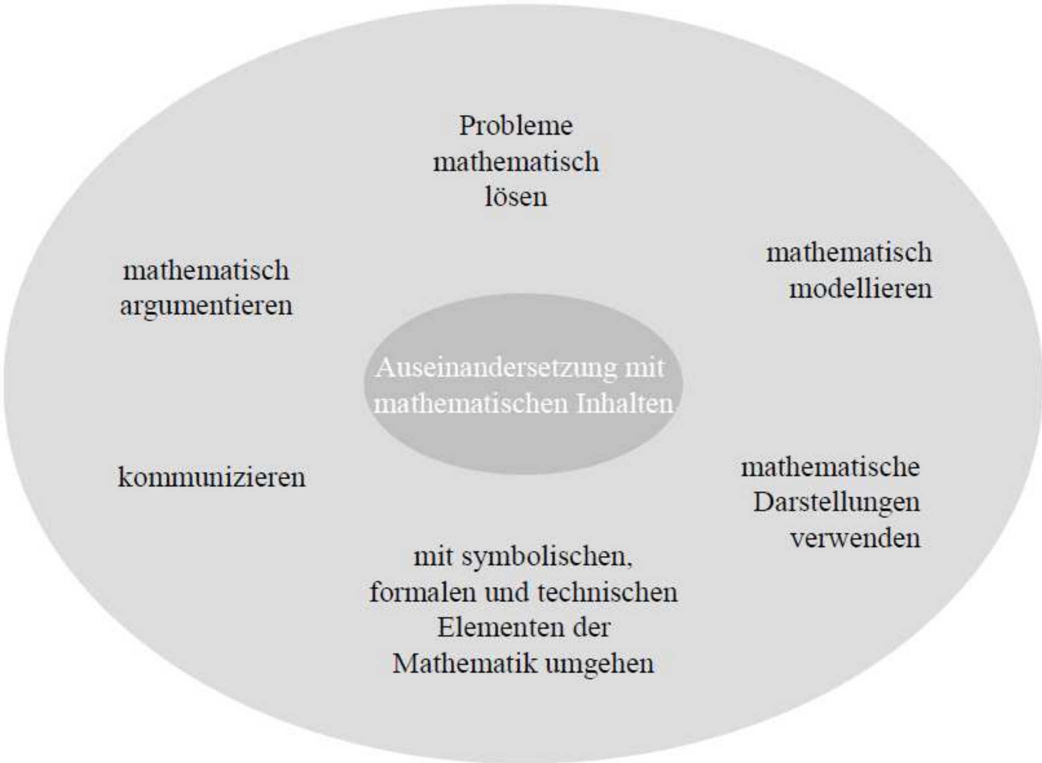


Abbildung 3: Bildungsstandards im Fach Mathematik (aus KMK, 2005)

Die in den Bildungsstandards realisierte Beschreibung von grundlegenden Dimensionen ergibt (in Kombination mit den dort formulierten hierarchisch angeordneten

[2] Dies sind die in den Standards für den mittleren Bildungsabschluss formulierten Anforderungen. Die Anforderungen für den Primarbereich sowie diejenigen für die Sekundarstufe II weichen geringfügig davon ab.

Anforderungsniveaus) ein Kompetenzmodell für das Fach Mathematik (Klieme, 2004).

Wünschenswert wären natürlich Kompetenzmodelle, die theoretisch und empirisch besser begründet sind. Solche Modelle liegen bis dato jedoch leider nicht vor. Bis dahin sind die in der Bundesrepublik vielseitig eingesetzten Bildungsstandards als gängiges Kompetenzmodell zu betrachten.

Um die zu Beginn dieses Teilkapitels gestellte Frage zu beantworten, kann mathematische Kompetenz also betrachtet werden als die Fähigkeit, die im Kompetenzmodell formulierten mathematischen Kompetenzen einzusetzen. Dass dies bei der Auseinandersetzung mit mathematischen Inhalten geschieht, spiegelt den Situationsbezug der Kompetenzen wieder.

Nachdem nun die Fragen beantwortet wurden, was Kompetenz ist und was mathematische Kompetenz ist, sind noch zwei weitere Fragen offen, die beantwortet werden sollen, bevor wir uns der Modellierungskompetenz zuwenden. Wie kann man mathematische Kompetenz messen? Und: Wie entsteht mathematische Kompetenz?

1.4.2.3 Messung mathematischer Kompetenz

Wie in der Definition bereits ausgeführt wurde, ist Kompetenz als ein latentes Konstrukt zu verstehen. Latente Konstrukte haben aber die Eigenschaft, dass sie der Beobachtung nicht direkt zugänglich, sprich: nicht direkt messbar sind.

Die Messung kann allerdings durch den manifesten Gegenpart der Kompetenz, die Performanz realisiert werden. Hasselhorn, Marx & Schneider (2005) sprechen statt von „Performanz" von „Leistung", meinen aber dasselbe, wenn sie schlussfolgern, Kompetenz könne nicht direkt gemessen werden, sehr wohl könne aber Leistung gemessen werden und aus ihr Rückschlüsse auf die Kompetenz gezogen werden.

Bei der Konstruktion von Testverfahren zur Erfassung latenter Dispositionen sind drei Aspekte zu beachten (Rost & Spada, 1978, siehe Abbildung 4).

- Latente, nicht beobachtbare Konstrukte können messbar gemacht werden, indem man sie operationalisiert. Das sog. Messmodell gibt dann Auskunft darüber, wie das latente Konstrukt mit den manifesten Indikatoren zusammenhängt.
- Bei der anschließenden Messung des Merkmals werden den durch Beobachtung gewonnenen Erfahrungen Zahlen zugeordnet, mit anderen Worten: einem empirischen Relativ wird ein numerisches Relativ zugeordnet.
- Letztlich können Verhaltensdispositionen aus ökonomischen Gründen nicht an der Gesamtheit aller derjenigen Personen, die die jeweilige Eigenschaft besitzen, gemessen werden. Aus der Grundgesamtheit ist daher eine Stichprobe zu ziehen, von der unter bestimmten Bedingungen wiederum Rückschlüsse auf die Gesamtpopulation möglich sind.

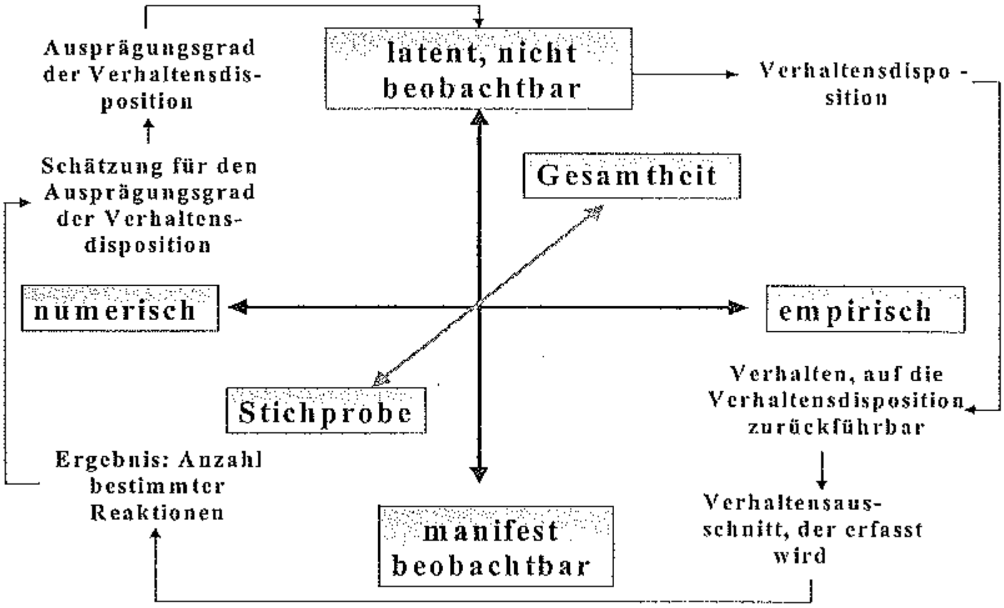


Abbildung 4: Messung einer Verhaltensdispostion mit Hilfe eines Testmodells (nach Rost und Spada 1978, S. 60)

Andere Autoren umgehen allerdings die Unterscheidung in Kompetenz und Performanz, indem sie beide Konzepte gleichsetzen, so z.B. Henning und Keune (2007, S.226): „Competence is here understood as a measurable variable, in the sense that level of competence can be inferred by observing the behavior of students.“

Eine Gleichsetzung beider Konzepte hat allerdings den Nachteil, dass hierbei der bei jeder Schlussweise von Indikatoren auf die zugehörigen Verhaltenspositionen auftretende Messfehler nicht berücksichtigt wird.

1.4.2.4 Entwicklung mathematischer Kompetenz

Wie kann man sich vorstellen, dass mathematische Kompetenz entsteht? Gehen wir von einem fiktiven Schüler A. der zu untersuchenden Zielgruppe am Ende der vierten Klasse aus. Einen großen Teil seiner Kompetenz hat A. wahrscheinlich in den letzten vier Jahren gewonnen, in denen er (beinahe) täglich dem Mathematikunterricht beiwohnte. Der formale Unterricht leistet also zweifelsohne einen bedeutenden, vielleicht sogar den größten Beitrag zur Kompetenzentwicklung in der Mathematik. Allerdings wäre es naiv anzunehmen, dass A. als *tabula rasa* seinen ersten Schultag angetreten hat (Richter & Brügelmann, 1994). Zu vielfältig sind die Erkenntnisse über die Entwicklung von Fähigkeiten und Kenntnissen, die als Vorläuferfähigkeiten für mathematische Kompetenz fungieren und bereits im Kleinkind- und Vorschulalter erworben werden (Weißhaupt, Peucker & Wirtz, 2006). Die wichtigsten Theorien und Befunde hierzu werden im übernächsten Abschnitt dargestellt. Darüber hinaus besteht die Kontroverse darüber, ob - und wenn ja welche - Fähigkeiten sogar von Geburt an vorhanden sind. Der nächste Abschnitt widmet sich daher einigen Befunden zur Anlage-Umwelt-Problematik.

1.4.2.4.1 Angeborene Fähigkeiten

Im Zusammenhang mit mathematischen Fähigkeiten ist nach wie vor die Diskussion um „nature or nurture“ als ausschlaggebende Komponente für die Entwicklung im Gange. Zentral ist hier die Frage, ob mathematische Basisfähigkeiten angeboren sind. Unter solche Basisfähigkeiten fasst man etwa einen angeborenen Zahlensinn (also ein Verständnis für Zahlen) oder darüber hinausgehend grundlegende Additionsfähigkeiten. Zwei Forschungsrichtungen können zu entsprechenden Betrachtungen herangezogen werden, einerseits die Untersuchung von Tieren, andererseits Untersuchungen an Säuglingen und Neugeborenen.

Experimente an Tieren machen sich die Tatsache zu nutze, dass eine genetische Verwandtschaft zwischen Menschen und anderen Säugetieren besteht. Gelingt es nachzuweisen, dass Tieren mathematische Fähigkeiten zueigen sind, so liegt zumindest die Vermutung nahe, dass dieselben auch uns Menschen angeboren sind, da sie unserem phylogenetischen Erbe entsprechen. Tatsächlich schildern z.B. Gallistel & Gelmann (2005) in einem Review Experimente, die auf so etwas wie basale Additions- und Subtraktionsfähigkeiten bei Säugetieren hindeuten. Dahaene (1997) beschreibt ebenfalls eine Reihe solcher Studien, zusammen mit dem Hinweis, Psychologen seien durch den Skandal um den „klugen Hans“ leider allzu schnell verleitet, die Ergebnisse solcher Tierexperimente zu kritisch zu betrachten und grundsätzlich Artefakte zu vermuten. Skepsis über einen „angeborenen Zahlensinn“ bei Tieren ist leider deshalb angebracht, weil auch entsprechende Studien an Säuglingen immer wieder zu widersprüchlichen Ergebnissen führen.

Oft zitiert werden die Arbeiten von Wynn (1992), die bewiesen haben will, dass fünf Monate alte Säuglinge über grundlegende Additions- und Subtraktionsfähigkeiten verfügen. Wynn arbeitete mit Habituierungsexperimenten, in denen sie sich die Tatsache zu Nutze machte, dass Säuglinge ihre Aufmerksamkeit bevorzugt auf neue und daher überraschende Situationen und Ereignisse richten. Sie konnte zeigen, dass ihre Probanden Situationen länger beachteten, in denen „falsche Ergebnisse“ gezeigt wurden (ein Gegenstand plus ein weiterer Gegenstand ergibt drei Gegenstände). Noch heute streitet man sich über die Bedeutung dieser Ergebnisse.

Während auf der einen Seite Neo-Nativisten wie Dahaene (1997) hierin den endgültigen Beleg für einen angeborenen Zahlensinn sehen, stehen andere Autoren den entsprechenden Befunden nach wie vor kritisch gegenüber. Bryant & Nuñes (2002) weisen auf die Möglichkeit hin, dass bei Ergebnissen der Form 1+1=1 Säuglinge Überraschung darüber zeigen, dass der status quo sich nicht ändert, obwohl ein Objekt hinzugefügt wurde und eben nicht über die Tatsache, dass hier fehlerhaft addiert wird. Entsprechende Folgestudien Wynns, die diese Möglichkeit ausschließen sollten, beinhalten jedoch keine Subtraktions- sondern nur Additions“aufgaben“ der Form 1+1=2 versus 1+1=3. Auch hier führen Bryant & Nuñes (2002) Alternativerklärungen an: Vielleicht steigt das Interesse der Säuglinge mit der Anzahl der Objekte, oder sie brauchen einfach länger, um drei Objekte visuell zu verarbeiten als lediglich zwei. Sie warnen außerdem zur Vorsicht, da Wynns Ergebnisse zwar in einigen, nicht jedoch in allen Folgestudien repliziert werden konnten.

Auch Krajewski (2005) äußerst Zweifel an der Validität diverser Habituierungsexperimente und vertritt im Allgemeinen die Ansicht, was hier gemessen werde, sei vielmehr das Erkennen von Quantitäten im Sinne von größeren oder kleineren Flächen als das konkreter Anzahlen. Auch weist sie darauf hin, dass sich die Verwunderung von Säuglingen angesichts der Tatsache „1+1=3“ auch darauf zurückführen lässt, dass solch ein Ergebnis nicht nur mathematisch, sondern eben auch physikalisch unmöglich ist.

Festzuhalten bleibt, dass die Befundlage zum jetzigen Zeitpunkt weder ausreicht, um Menschen und Tieren angeborene mathematische Fähigkeiten zu- noch abzusprechen. Zwei Aussagen lassen sich aber aus dem aktuellen Forschungsstand dennoch mit einiger Sicherheit ableiten:

- Vorgänge wie das Zählen beruhen nicht auf angeborenen mathematischen Prinzipien, wie dies Gelman und Gallistel für ihre fünf Zählprinzipien (s.u) propagieren, sondern sie entwickeln sich langsam und kontinuierlich aufgrund von Erfahrungen mit dem Zählenlernen (Karmiloff-Smith, 1992).

- Angeboren oder zumindest sehr früh in der Entwicklung erlernt zu sein scheint nicht ein Zahlensinn, sondern vielmehr eine eher vage Mengenbewusstheit, die sich weniger auf diskrete Anzahlen, als vielmehr auf die Wahrnehmung und Unterscheidung kontinuierlicher Größen bezieht.

1.4.2.4.2 Vorläuferfähigkeiten

Was aber passiert in den langen Jahren zwischen Geburt und Schuleintritt? Beobachtet man Vorschulkinder, so kann man feststellen, dass diese bereits über erstaunliche mathematische Fähigkeiten verfügen: Grassmann, Klunter, Köhler, Mirwald, Raudies & Thiel (2002) berichten unter anderem, dass

- 91% der von ihnen untersuchten Vorschulkinder Mengen bis 7 abzählen können,
- 64% leichte Additionsaufgaben lösen können,
- 54% dies auch ohne Abzählen gelingt und
- 59% von 10 rückwärts zählen können.

Hier spricht man allerdings noch nicht von mathematischer Kompetenz. In der Literatur findet man für Basisfähigkeiten von Vorschulkindern eher den Begriff „*Vorläuferfähigkeiten*“. Vorläuferfähigkeiten sind noch keine eigentlichen Kompetenzen, sondern fungieren als Prädiktoren für diese. Sie beschreiben damit Fähigkeiten, die bereits vor Schuleintritt gegeben sind, die aber den Erfolg in der Schule in gewissem Maße vorhersagen können, da sie eine Basis für sich später entwickelnde Kompetenzen bilden (Riebel & Jäger, 2008). Hierbei ist grundsätzlich zu unterscheiden zwischen allgemeinen und spezifischen Vorläuferfähigkeiten. Allgemeine Vorläuferfähigkeiten sind von Bedeutung für die Entstehung einer ganzen Reihe von späteren Einzelkompetenzen. Kammermeyer (2004) spricht hier von einem Schrotschussverfahren: Sprache, Wahrnehmung, Verhalten, Motorik und Intelligenz seien zwar alle *irgendwie* wichtig, man wisse aber nicht, wie diese Faktoren kausal zusammenwirken. Besser sei es daher, sich auf spezifische Vorläuferfähigkeiten zu konzentrieren, in diesem Fall also auf mathematische Vorläuferfähigkeiten.

Warum dieses Vorgehen praktikabler ist, macht Abbildung 5 deutlich. Links ist der Zusammenhang zwischen allgemeinen Vorläuferfähigkeiten und der Kompetenzentwicklung (in schulisch und außerschulisch relevanten Bereichen, darunter auch mathematische Kompetenz und Schriftspracherwerb) dargestellt. Nicht nur haben alle diese Vorläuferfähigkeiten einen Einfluss auf die Kompetenz, es muss auch angenommen werden, dass sie sich gegenseitig beeinflussen. Das daraus entstehende Geflecht von gegenseitigen Wirkungen ist daher kaum durchschaubar und nur schwer zu untersuchen. Im rechten Wirkmodell wird der Zusammenhang zwischen spezifischen Vorläuferfähigkeiten und den sich daraus zu entwickelnden Kompetenzen dargestellt. Bei der mathematischen Kompetenz ist dies mengen- und zahlenbezogenes Vorwissen (siehe unten), beim Schriftspracherwerb die phonologische Bewusstheit. Durch eine Orientierung an diesem zweiten Wirkmodell soll die Bedeutung der allgemeinen Vorläuferfähigkeiten keinesfalls abgestritten werden. Es soll hiermit nur deutlich gemacht werden, dass für eine gezielte Förderung von Vorläuferfähigkeiten, mit dem Ziel Beeinträchtigungen bei der Kompetenzentwicklung zu verhindern, eine Konzentration auf Vorläuferfähigkeiten, deren Wirkung leichter spezifiziert und überprüft werden kann, angezeigt ist.

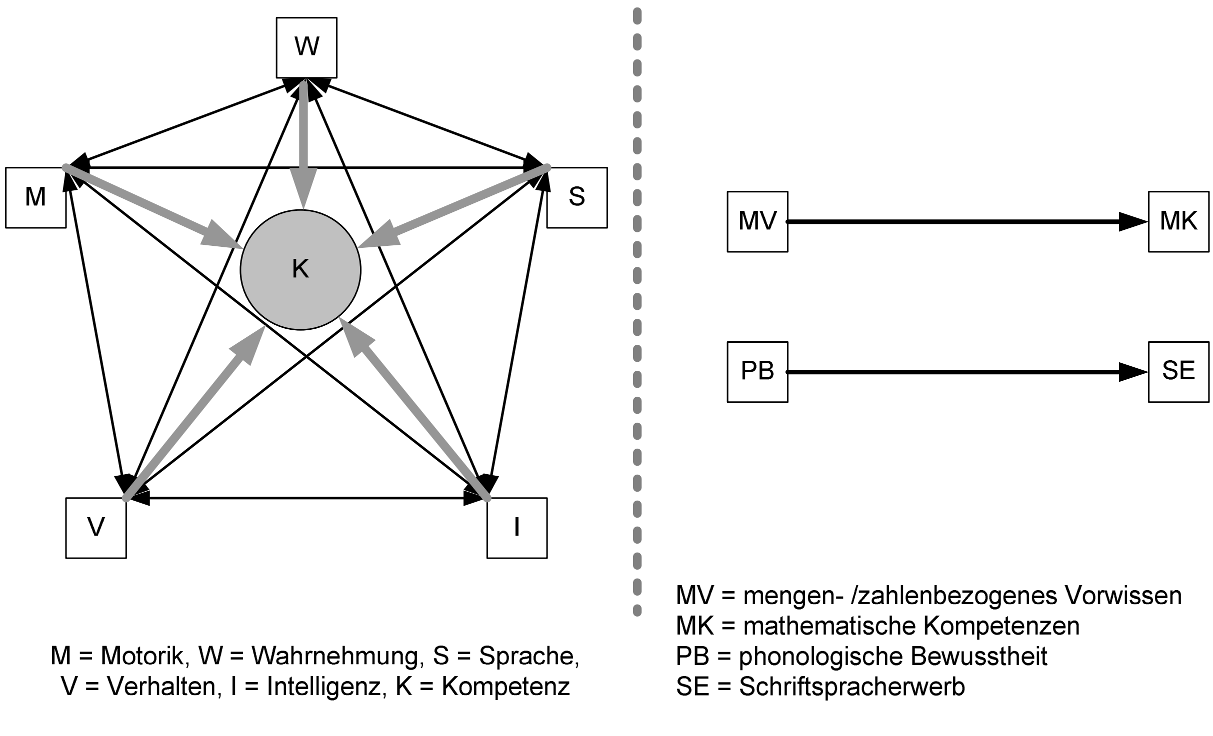


Abbildung 5: Wirkmodelle von allgemeinen (links) und spezifischen (rechts) Vorläuferfähigkeiten

Mathematische Vorläuferfähigkeiten sind besonders im Hinblick auf Dyskalkulie von Bedeutung (Kaufmann & Nuerk, 2006; Mack, 2005). Die Dyskalkulie ist nach ICD-10: F81.2 eine Rechenstörung, die sich in verminderter Schulleistung in den Bereichen grundlegender Rechenfähigkeiten manifestiert und sich nicht auf eine Intelligenzminderung zurückführen lässt (Baumann & Perrez, 2005). Da sich die Störung definitionsgemäß erst *nach* Schuleintritt diagnostizieren lässt, konzentrieren sich gängige diagnostische Verfahren (z.B. ZAREKI-R, DEMAT-Reihe) auf die Diagnose einer eventuell gegebenen Rechenstörung in der Zielgruppe der Grundschulkinder.

Auch wenn sich die Dyskalkulie als Schulleistungsstörung im strengen Sinn erst nach Schuleintritt diagnostizieren lässt, so ist es dennoch nicht unmöglich, bereits im Kindergartenalter Kinder zu identifizieren, die einem erhöhten Risiko für die Entwicklung der Störung ausgesetzt sind. Genau hierfür sind aber wiederum Vorläuferfähigkeiten relevant: Sind diese bei einem Individuum nicht ausgebildet, so ist das Risiko erhöht, dass im späteren Entwicklungsverlauf die Ausbildung der mathematischen Kompetenz leidet. Hierfür müssen aber zunächst folgende Sachverhalte realisiert werden: Die Vorläuferfähigkeiten müssen identifiziert und es muss kausal nachgewiesen werden, dass die Abwesenheit dieser Vorkenntnisse zu Rechenschwierigkeiten führt (Fritz & Ricken, 2005). Dies gelang Krajewski (2003) in einer Langzeitstudie. Sie konnte zeigen, dass sich erste Rechenfähigkeiten aufgrund der Verknüpfung zweier wichtiger Vorläuferfähigkeiten entwickeln, auf welche nun genauer eingegangen werden soll, Mengenwissen und Zahlenwissen.

1.4.2.4.2.1 Mengenwissen

Kleine Mengen können sowohl von Kindern, als auch von Erwachsenen „auf einen Blick“ erkannt werden, eine Fähigkeit, die man als „subitizing“ bezeichnet (Dahaene, 1997; Fritz & Ricken, 2005). Hierbei wird allerdings vermutet, dass dieses Phänomen rein wahrnehmungsbasiert ist und mit Zählen nichts zu tun hat. Simon & Cabrera (1995) konnten nämlich zeigen, dass die Reaktionszeiten beim subitizing unabhängig von der Anzahl der zu erfassenden Elemente sind. Würde hier tatsächlich ein Zählvorgang in Gange gesetzt werden, so müsste die Dauer dieses Vorgangs mit der Anzahl der Elemente ansteigen. Uneinigkeit besteht wiederum darin, ob die Fähigkeit zum subitizing angeboren ist. Fest steht wohl, dass es bereits sehr früh auftritt und sich auch bei ansonsten extrem leistungsschwachen Kindern finden lässt (Fritz & Ricken, 2005). Durch den Spracherwerb werden Kinder später dann dazu befähigt, Mengen und Beziehungen zwischen beiden nicht nur wahrzunehmen, sondern diese Wahrnehmungen auch zu kommunizieren.

1.4.2.4.2.2 Zahlenwissen

Die Fähigkeit zum Zählen entwickelt sich relativ bald nach dem Spracherwerb. Über die Art und Weise wie dies von statten geht, gibt es verschiedene Theorien. Als wichtigste Vertreter können wohl Gelman & Gallistel (1978) einerseits sowie Fuson (1988) andererseits betrachtet werden. Gelman & Gallistel postulieren fünf Zählprinzipien, welche die natürlichen Zahlen beschreiben und deren Kenntnis Voraussetzung für ein sinnvolles Zählen ist. Die Prinzipien sind Stabilität der Zahlwortreihe (man darf beim Zählen also nicht die Reihenfolge der Zahlwörter vertauschen), Eins-zu-eins Zuordnung (jedem zu zählenden Gegenstand wird genau ein Zahlwort zugeordnet), Kardinalität (das beim Zählen als letztes genannte Wort bezeichnet das letzte Objekt und gleichzeitig die Anzahl der Objekte, also die Mächtigkeit der Menge), Anordnungsbeliebigkeit (in welcher Reihenfolge man Objekte zählt ist gleichgültig, solange man jedes Objekt genau einmal berücksichtigt) und Abstraktionsprinzip (Alles ist zählbar) (Lorenz, 2005). Weniger haltbar an der Theorie Gelman & Gallistels (1978) ist die Tatsache, dass sie postulieren, diese Prinzipien seien angeboren und in den neurologischen Strukturen des menschlichen Gehirns verankert. Mithilfe dieser Prinzipien – so die Autoren – steuert das Gehirn die Wahrnehmung von Mengen und Zahlen und es trägt dazu bei, diese sinnvoll interpretieren zu können. Als Beleg hierfür wird die Fähigkeit zum subitizing angeführt. Wie oben bereits erläutert wurde, wird beim subitizing jedoch nicht gezählt, da es sich um eine rein wahrnehmungsbasierte Mengenerfassung handelt. Demnach liegen diesem Vorgang wohl auch keine Zählprinzipien zugrunde. Zwar ist allein hiermit die Theorie der angeborenen Zählprinzipien noch nicht widerlegt, jedoch ist ihr eine wichtige Argumentationsbasis entzogen. Heute geht man daher vielmehr davon aus, dass sich die Zählprinzipien erst mit der Zeit – durch die aktive Auseinandersetzung mit Zahlen, Mengen und Zählvorgängen – entwickeln (Karmiloff-Smith, 1992).

Fuson (1988) vertritt eine andere Richtung: Sie geht von Anfang an nicht von angeborenen Fähigkeiten aus, sondern davon, dass diese sich schrittweise entwickeln. Sie beschreibt eine Ablauffolge von Entwicklungsschritten, wie Kinder sie typischerweise durchlaufen, wenn sie Zählen lernen. Dieser Prozess vollzieht sich im Alter von vier bis sieben oder acht Jahren und umfasst folgende Stufen (Fuson, 1988, S.45-55):

- *String level*: Zahlwörter sind aneinandergereiht und werden wie ein Wort behandelt („einszweidrei“). Kinder sind nicht in der Lage zu erkennen, welche und wie viele Zahlwörter in dieser Liste stecken und was sie bedeuten.

- *Unbreakable list level*: Die Zahlwörter sind nun differenzierbar und können als separat erkannt werden. Die Zahlwortreihe kann aber nur von vorne nach hinten durchlaufen werden und beginnt immer bei eins („eins-zwei-drei“).
- *Breakable chain level*: Ab hier kann die Zahlwortreihe nun von einem beliebigen Punkt aus gestartet werden („fünf-sechs-sieben“).
- *Numerable chain level*: Kinder erfassen in zunehmendem Maße die Bedeutung von Zahlen. In dieser Stufe wird das Kardinalitätsprinzip von Zahlen verstanden. Zahlen können von nun an eingesetzt werden, um kleine Mengen zu addieren und zu subtrahieren (Bermejo, 1996).
- *Bidirectional chain level*: In der letzten Phase können Kinder von jedem beliebigen Punkt der beherrschten Zahlwortreihe aus auch rückwärts zählen.

Kardinalität, die von Gelman & Gallistel (1978) als angeboren betrachtet wird, taucht laut Fuson (1988) also erst relativ spät im Entwicklungsprozess auf. Auch das Eins-zu-eins-Prinzip entwickelt sich erst auf der zweiten Stufe. Erst nachdem einzelne Zahlwörter voneinander unterschieden werden können, können diese auch jeweils einem zu zählenden Gegenstand zugeordnet werden.

1.4.2.4.2.3 Erste Rechenfähigkeiten

Die Verbindung von Mengenbewusstsein und Beherrschen des Zählens bildet die Basis für die Entwicklung erster Rechenfähigkeiten. Einfache Additions- und Subtraktionsaufgaben können gelöst werden, beinhalten in diesem ersten Stadium jedoch immer noch das Zählen, oft mithilfe der Finger. Dies dauert natürlich noch vergleichsweise lange und ist nur mit recht kleinen Zahlen möglich. Hierbei greifen Kinder auf verschiedene Strategien zurück. Für frühe Entwicklungsstufen ist die sog. SUM – Strategie typisch: Bei der Addition zweier Zahlen (2+3) wird bis zur ersten Zahl gezählt (eins – zwei), und dann werden der zweiten Zahlen entsprechend weitere Zählschritte hinzugefügt (drei – vier – fünf). Später wird diese Strategie durch eine effizientere und damit auch weniger fehleranfällige Strategie (MIN) ersetzt. Bei der Addition zweier Zahlen wird dann nicht mehr automatisch beim ersten Summanden angefangen, sondern bei der jeweils größeren Zahl (Shaffer, 2002).

Während der Grundschulzeit vollzieht sich eine immer weiter fortschreitende Verinnerlichung des Zählvorgangs, der dann - wenn überhaupt - nur noch mental vonstatten geht. Aebli (1976) beschreibt diesen Verinnerlichungsprozess anhand der folgenden vier Phasen: Anfangs wird tatsächlich mit konkreten, in der Realität vorhandenen Objekten gerechnet (drei rote Bauklötze und vier weiße Bauklötze). In einem zweiten Schritt werden aus tatsächlichen dreidimensionalen Gegenständen dann in der Vorstellung der Kinder zweidimensionale schematische Abbilder der Gegenstände. Kinder stellen sich dann lediglich noch vor, wirklich zwei Mengen von Bauklötzen vor sich zu haben. Sie sind von nun an auch in der Lage, schematische Zeichnungen (z.B. ein Kreis als Skizze eines Balls) zu erkennen und mit diesen zu arbeiten. Im dritten Schritt treten symbolische Darstellungen durch arabische Ziffern an die Stelle der zuvor beschriebenen Bilder. Es wird also noch mehr abstrahiert. Das Symbol „3“ kann dann stellvertretend für jede Menge stehen, die aus drei wie auch immer gearteten Objekten besteht. Krajewski (2005) sieht an genau dieser Stelle die Verschmelzung zwischen Mengenwissen und Zahlenwissen. In der vierten Stufe entsteht dann aus prozeduralem Wissen über Additionsvorgänge deklaratives Wissen einerseits über konkrete Ergebnisse, die dann direkt aus dem Gedächtnis abgerufen werden können (man denke an das Kleine Einmaleins) und anderseits über Algorithmen zum Lösen komplexerer Aufgaben (wie zum Beispiel das Vorgehen beim schriftlichen Dividieren). Zu Beachten bleibt bei der Verwendung von verschiedenen Strategien, dass nicht immer die schlechtere Strategie komplett von der besseren abgelöst und danach nicht mehr eingesetzt wird. Vielmehr besitzen Kinder und auch Erwachsene ein Repertoire von Strategien. Bei Unsicherheit wird auf ältere Strategien zurückgegriffen, die zwar vielleicht aufwendiger sind, die jedoch der betreffenden Person wesentlich leichter fallen, weil sie darin geübt ist (Bjorklund & Rosenblum, 2001).

Auch wenn über den genauen Entwicklungsverlauf mathematischer Kompetenzen weiterhin nur spekuliert werden kann, so steht doch fest, dass bereits vor Schuleintritt eine ganze Reihe bedeutender Entwicklungsschritte vor sich geht. Weinhold Zulauf und Kollegen bezeichnen das Kindergartenalter als sensitive Zeitspanne für die Entwicklung numerischer Fähigkeiten (Weinhold Zulauf, Schweiter & von Aster, 2003). Hieraus wird abgeleitet, dass eine möglichst frühe Förderung mathematischer Kompetenz angezeigt ist.

Fassen wir zusammen: In der zu untersuchenden Zielgruppe, die sich am Ende ihrer Grundschulzeit befindet, kann man von mindestens drei Quellen ausgehen, aus denen sich jeweils die gegebene mathematische Kompetenz speist:

- Basisfähigkeiten, die entweder angeboren sind, oder sich bereits sehr früh entwickeln (auch wenn diese sich vornehmlich auf Wahrnehmungsphänomene begrenzen),
- Vorläuferfähigkeiten, die sich in der Zeitspanne von der Geburt bis zum Schuleintritt sowohl spontan als auch durch gezielte Frühförderung entwickeln,
- Kompetenzen, die im formalen Mathematikunterricht in den ersten vier Schuljahren vermittelt werden.

Die unterschiedlichen Einflüsse auf die Kompetenz spielen im Verlauf der weiteren Betrachtungen keine Rolle mehr. Es darf jedoch nicht vergessen werden, dass die mathematische Kompetenz am Ende des vierten Schuljahres nicht alleine auf den Grundschulunterricht zurückzuführen ist, sondern dass dies lediglich eine von drei Komponenten darstellt.

1.4.3 Modellierungskompetenz

Oben wurden bereits die in den Bildungsstandards formulierten sechs Grundkomponenten der mathematischen Kompetenz dargestellt. Eine dieser sechs Teilkomponenten ist die Modellierungskompetenz. Fasst man Modellierungskompetenz als die Kompetenz zu Modellieren auf, so schließt sich die Frage an, welche Fähigkeiten Modellieren beinhaltet.

Modellieren wurde bereits definiert als das Abschreiten der im Modellierungszyklus formulierten Stadien und Prozesse. Die meisten Autoren definieren Modellierungskompetenz daher als die Fähigkeit, den Modellierungszyklus erfolgreich zu durchschreiten, so beispielsweise Blomhoj & Jensen (2003, S. 126): „By mathematical modelling competence we mean being able to autonomously and insightfully carry through all aspects of a mathematical modelling process in a certain context.” Blum et al. (2007, S. 12) benennen in ihrer Definition zusätzlich die konkreten Stadien: „So *mathematical modelling competency* means the ability to identify relevant questions, variables, relations or assumptions in a given real world problem situation, to translate these into mathematics and to interpret and validate the solution of the resulting mathematical problem in relation to the given situation, as well as the ability to analyse or compare given models by investigating the assumptions being made, checking properties and scope of a given model etc. in short: modelling competency in our sense denotes the ability to perform the processes that are involved in the construction and investigation of mathematical models.” Diese Definition wird den nachfolgenden Ausführungen zugrunde gelegt.

Maaß (2006, S. 117) weitet die Kompetenzdefintion aus, indem sie auch Aspekte der Performanz miteinbezieht: „Modelling competencies include skills and abilities to perform modelling processes appropriately and goal-oriented as well as the willingness to put these into action.” Ob eine solche Erweiterung sinnvoll ist, bleibt fraglich, da Kompetenz zunächst einmal die Fähigkeit beschreibt, die eine Person in die Lage versetzt, eine bestimmte Handlung auszuführen. Ob die betreffende Person dann aber in einer konkreten Situation auch gewillt oder in der Lage ist, diese Fähigkeit anzuwenden, ist eigentlich eher ein (motivationaler) Aspekt der Performanz. Die theoretisch vorhandene Kompetenz dürfte dadurch jedoch nicht in Frage gestellt werden.

Kaiser & Schwarz (2006) verwenden eine noch breitere Definition von Modellierungskompetenzen, indem sie zwei Bereiche miteinbeziehen: Zum einen sei unter

Modellierungskompetenzen eine Sammlung von Einzelkompetenzen zu verstehen, nämlich diejenigen zur Meisterung der folgenden Teilprozesse:

- Vereinfachen des Problems und Konzentration auf bearbeitbare Fragen,
- Identifikation zentraler Variablen und ihrer Beziehungen,
- Formulierung von Annahmen und Beschreibung des Problems,
- Konstruktion und Auswahl adäquater mathematischer Beschreibungen des Problems und Entwicklung von Lösungen innerhalb des Modells,
- Interpretation der Lösungen innerhalb des realen Problemkontexts und Evaluation der Angemessenheit von Lösungen.

Zum anderen gehören zur Modellierungskompetenz jedoch auch weitere Einzelkompetenzen, die vom konkreten Modellierungszyklus als losgelöst zu betrachten sind. Dazu gehören die Kompetenz, mithilfe der Aktivierung von Metawissen über den Modellierungszyklus zu reflektieren, die Einsicht in die Beziehung zwischen Mathematik und Realität, soziale Kompetenzen für die Bearbeitung von Modellierungsproblemen in Gruppenarbeit sowie kommunikative Kompetenzen zur Darstellung der Ergebnisse.

Auch Maaß (2006, S.113) bezieht solche Randkompetenzen in die Definition mit ein: „While there is a broad consensus that modelling competencies include certain sub-competencies like setting up a real model or mathematizing such a model, it is often said that those sub-competencies relating to the process of modelling are not sufficient to characterize modelling competencies." Sie leitet aus ihren Fallstudien fünf Dimensionen von Modellierungskompetenzen ab:

- Unterkompetenzen, die man braucht, um die einzelnen Schritte des Modellierungsprozesses durchzulaufen (siehe Tabelle 3),
- Metakognitive Kompetenzen,
- Kompetenz, reale Probleme zu strukturieren und zielgerichtet auf das Prozessende hinzuarbeiten,
- Kompetenzen, in Bezug auf den Modellierungsprozess und diese Argumentation verschriftlichen zu können,
- Kompetenzen, die Möglichkeiten zu sehen, welche die Mathematik für die Lösung realer Probleme bietet und diese Möglichkeiten als positiv zu betrachten.

All diese von Maaß (2006) und Kaiser & Schwarz (2006) aufgelisteten Randkompetenzen könnten natürlich durchaus einen Einfluss auf die Qualität der Lösung von Modellierungsaufgaben haben. Trotzdem sollten diese Aspekte nicht als Teilkomponenten der Modellierungskompetenz aufgefasst werden, obwohl zugestanden werden muss, dass einige dieser Kompetenzen mit der Modellierungskompetenz (z.B. metakognitive Kompetenzen) oder mit der dazugehörigen Performanz (z.B. soziale Kompetenz, Motivation) korrelieren dürften.

Für die Operationalisierung eines Konstruktes ist es darüber hinaus von Vorteil, eine möglichst enge und präzise Definition zu verwenden. Beschränkt man sich auf die aus dem Modellierungszyklus abgeleiteten Teilkompetenzen *Strukturieren*, *Mathematisieren*, *Verarbeiten*, *Interpretieren* und *Validieren*, so kann man sich nun wiederum jede diese Einzelkompetenzen vornehmen, noch einen Schritt weiter in die Tiefe gehen, und der Frage nachgehen, was die jeweilige Einzelkompetenz ausmacht. Maaß (2004, 2006, 2007) setzt genau dieses Vorgehen um und listet für jeden Schritt im Modellierungszyklus wiederum Unterkompetenzen auf, die nötig sind um den jeweiligen Schritt durchzuführen. Tabelle 3 gibt die dort spezifizierten Unterkompetenzen wieder.

Tabelle 3: Unterkompetenzen nach Maaß (2006, S.116f.)

1	Competencies to understand the real problem and to set up a model based on reality	**Competency** • to make assumptions for the problem and simplify the situation; • to recognize quantities that influence the situation, to name them and to identify key variables; • to construct relations between the variables; • to look for available information and to differentiate between relevant and irrelevant information;
2	Competencies to set up a mathematical model from the real model	**Competency** • to mathematize relevant quantities and their relations • to simplify relevant quantities and their relations if necessary and to reduce their number and complexity; • to choose appropriate mathematical notations and to represent situations graphically;
3	Competencies to solve mathematical questions within this mathematical model	**Competency** • to use heuristic strategies such as division of the problem into part problems, establishing relations to similar or analog problems, rephrasing the problem, viewing the problem in a different form, varying the quantities or the available data etc.; • to use mathematical knowledge to solve the problem;
4	Competencies to interpret mathematical results in a real situation	**Competency** • to interpret mathematical results in extramathematical contexts; • to generalize solutions that were developed for a special situation; • to view solutions to a problem by using appropriate mathematical language and/or to communicate about the solutions;
5	Competencies to validate the solution	**Competency** • to critically check and reflect on found solutions; to review some parts of the model or again go through the modelling process if solutions do not fit the situation; • to reflect on other ways of solving the problem or if solutions can be developed differently; • to generally question the model.

Zusammenfassend lässt sich folgende Hierarchie der bis jetzt betrachteten Kompetenzen festhalten: Die Fähigkeit, mathematische Probleme in den unterschiedlichsten Situationen und je nach Anforderung auf verschiedene Art und Weise zu lösen, bezeichnet die mathematische Kompetenz. Mathematische Kompetenz gliedert sich laut Bildungsstandards in die sechs Kompetenzen „Argumentieren“, „Problemlösen“, „Modellieren“, „Darstellungen verwenden“, „mit symbolischen, formalen und technischen Elementen umgehen“ sowie „Kommunizieren“. Die Modellierungskompetenz als Teil der mathematischen Kompetenz lässt sich nun wiederum in die im Modellierungszyklus spezifizierten Teilkompetenzen „*Strukturieren*“, „*Mathematisieren*“, „*Verarbeiten*“, „*Interpretieren*“ und „*Validieren*“ aufteilen. Jede dieser Teilkompetenzen umfasst nun ihrerseits eine Reihe von Unterkompetenzen. Abbildung 6 stellt den Zusammenhang zwischen den Kompetenzarten schematisch in einem Kompetenzmodell dar.

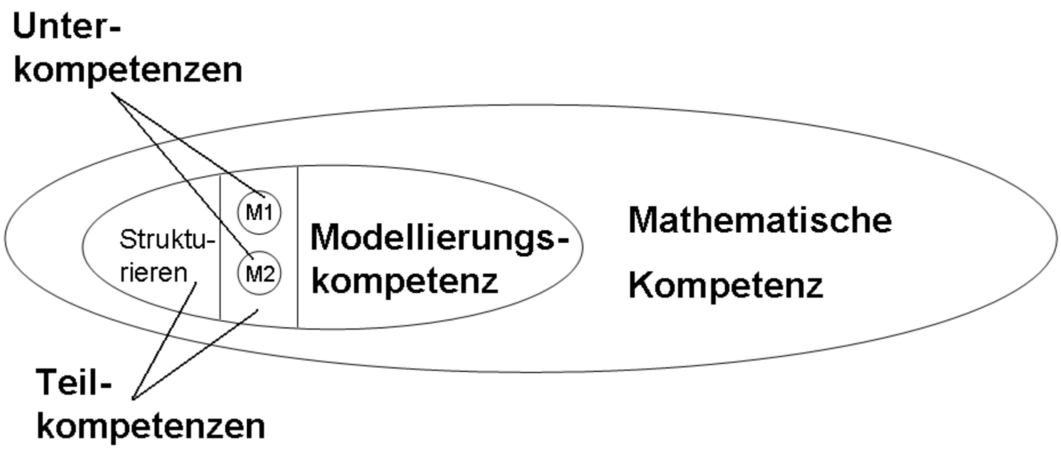


Abbildung 6: Kompetenzmodell

1.5 Bedarf für die Diagnostik von Modellierungskompetenz am Ende der Primarstufe

In den vorhergehenden Abschnitten wurde deutlich, welch bedeutende Rolle der Modellierungskompetenz zufällt. Diese ist Teil der mathematischen Kompetenz. Die Mathematik aber ist – so Baumert und Kollegen (Baumert, Lehmann, Lehrke, Schmitz, Clausen, Hosenfeld, Köller & Neubrand, 1997, S. 58) - „als Werkzeug gewissermaßen Teil der kulturellen Alphabetisierung." Daraus folgt im Umkehrschluss, dass ein Nichtbesitzen grundlegender mathematischer Kompetenzen gleichbedeutend ist mit einer Art von kulturellem Analphabetismus. Diesen kann und darf sich eine moderne Gesellschaft aus Gründen der Chancengleichheit und aus volkswirtschaftlichen Erwägungen nicht leisten.

In der Realität zeigt sich nun aber leider, dass gerade Modellierungsaufgaben deutschen Schülern besonders schwer fallen. Die PISA-Studien haben gezeigt, dass in der Klassenstufe 9 nur 44% der untersuchten Personen ein ausreichendes Niveau der mathematischen Grundbildung erreichen (Klieme et al., 2001). Dies verwundert wenig, denn bereits nach Abschluss der Primarstufe sind die mathematischen Kompetenzen vieler Schüler defizitär. Die IGLU-Studie (Walther et al., 2003) erfasste mathematische Kompetenz anhand von fünf Kompetenzstufen. Dabei bezeichnet die Kompetenzstufe III die Fähigkeiten zur Arbeit mit einfachen Modellen und stellt somit eine Minimalkompetenz für den Umgang mit Modellierungsproblemen dar. 18.6% der in IGLU untersuchten Schüler erreichten diese dritte Kompetenzstufe nicht, weitere 39.8% kamen nicht über diese Stufe hinaus.

Aus diesen beiden Sachverhalten – der Notwendigkeit zur Modellierungsfähigkeit einerseits und den Defiziten in der Bevölkerung andererseits – lässt sich das Erfordernis ableiten, Modellierungskompetenzen zu fördern.

Vor dem Hintergrund der IGLU-Ergebnisse wird besonders deutlich, dass man nicht erst im Sekundarbereich damit beginnen kann, Modellierungsaufgaben im Unterricht einzusetzen. Viel besser ist es, Schülern Modellierungskompetenzen schon zu einem Zeitpunkt zu vermitteln, wo die zu modellierenden Inhalte noch relativ einfach sind. Beginnt man damit nämlich erst, wenn komplexere mathematische Themen behandelt worden sind, so überfordert man die Betroffenen: Einerseits sollen sie schwierige mathematische Inhalte umsetzen, gleichzeitig dies aber auf eine Art und Weise tun, die sie vorher nicht gelernt haben.

Ein Großteil der Literatur zu Modellieren und Modellierungskompetenzen konzentriert sich auf den tertiären Bildungsbereich, ein geringerer Teil beschäftigt sich mit Modellieren in der Sekundarstufe. Nur sehr wenige Arbeiten beschäftigen sich mit Modellieren in der Grundschule. Diese einseitige Orientierung ist darin begründet, dass Modellierungskurse zunächst an Universitäten eingeführt wurden, um die Absolventen besser auf praktische Aufgaben in ihrem späteren Berufsfeld vorzubereiten (Galbraith, 2007).

Dennoch plädieren einige Autoren (English, 2003; Greer et al., 2007; Lamon, 2003; Usiskin, 2007) dafür, Modellieren bereits möglichst frühzeitig zu vermitteln und Schülerinnen und Schülern zunehmend komplexere Probleme vorzusetzen. Möchte man Menschen ermöglichen, mathematisches Modellieren als „A way of life“ zu betrachten, so darf man ihnen den Zugang zu dieser Sichtweise nicht verwehren, so English (2003).

Geht man nun noch einen Schritt weiter und betrachtet die Notwendigkeit der Förderung von Modellierungsfähigkeiten als gegeben, so schließt sich eine weiteres Problem an: Davon ausgehend, dass Modellieren eine Reihe von Teilkompetenzen umfasst (siehe Kapitel 1.4.3), ist es für eine gezielte Förderung angezeigt, für jeden Lerner Stärken und Schwächen zu identifizieren, um dann gezielt die Entwicklung derjenigen Teilkompetenzen zu unterstützen, die noch defizitär sind. Es ist also erforderlich, Wege zu finden, wie man solche Teilkompetenzen erfassen kann.

Das Diagnostizieren und Fördern sollte wie bereits angedeutet möglichst frühzeitig geschehen. Das Ende der vierten Klassenstufe ist für eine solche Bestandsaufnahme besonders günstig. Einerseits, weil Schüler nach den ersten vier Schuljahren über ein ausreichend großes Basiswissen im mathematischen Bereich verfügen, um sinnvolle Modellierungsaufgaben bearbeiten zu können; andererseits weil sich gerade gegen Ende der vierten Klasse im Rahmen von weiteren Schullaufbahnentscheidungen eine Bestandsaufnahme des aktuellen Wissensstandes anbietet.

Zielsetzung der vorliegenden Arbeit ist es daher, erste empirische Grundlagen für die Entwicklung eines diagnostischen Instrumentes zu schaffen, welches Teilkompetenzen beim mathematischen Problemlösen erfasst. Ein solches Instrument ist zunächst für die Zielgruppe der Grundschulabgänger zu realisieren, da gerade vor der für diese Personen anstehenden Transitionsphase Lernstandsanalysen besonders angezeigt sind.

1.6 Diagnostik von Modellierungskompetenzen

Wie kann nun eine Diagnostik von Modellierungskompetenzen erfolgen? Die für die Diagnostik von Modellierungskompetenzen relevante Fragestellung, aus der sich alle folgenden Teilschritte im diagnostischen Prozess ableiten, lautet: „Bei welchen Teilkompetenzen mathematischen Modellierens besteht bei Proband P noch Förderbedarf?"

Betrachtet man die direkt nach der Fragestellung auftretenden Stadien des diagnostischen Prozesses (Jäger, 1983; siehe Abbildung 7) dann wird deutlich, dass zunächst noch eine ganze Reihe von Fragen offen ist, wie sich dieser Prozess im Rahmen der angedeuteten Fragestellung zu vollziehen hat:

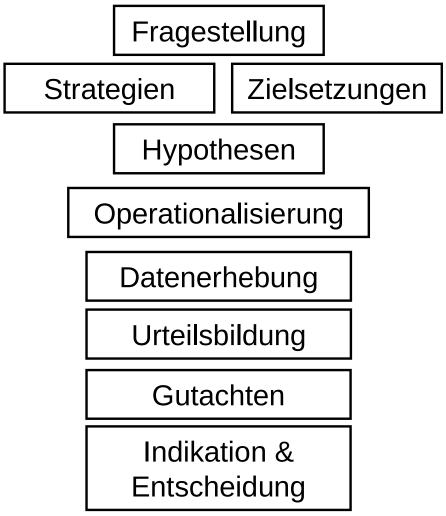


Abbildung 7: Stadien des diagnostischen Prozesses

- Während sich auf der Ebene von Makrostrategien (Jäger & Petermann, 1999) relativ eindeutig eine Zuordnung des Vorgehens zu Selektionsstrategien vornehmen lässt, muss auf der Ebene der Mikrostrategien sorgfältig erwogen werden, wie beispielsweise die Vorgabe von Items zu realisieren ist. Längerfristig ist mit Sicherheit die Verwendung von adaptiven Strategien sinnvoll, da auf diesem Wege schnell und präzise eine Einschätzung des Kompetenzgrades erfolgen kann.
- Hinsichtlich der diagnostischen Zielsetzung des zu entwickelnden Verfahrens muss diskutiert werden, ob eine Einordnung der Ergebnisse norm- oder

kriteriumsorientiert erfolgen sollte (Schott, 1999). Es stellt sich die Frage, inwieweit die Orientierung an einer altersgruppenbezogenen Norm sinnvoll ist, wenn Befunde existieren, dass ein Großteil der Normgruppe Defizite im Bereich der zur Untersuchung stehenden Kompetenz aufweist. Alternativ können Testergebnisse auch kriteriumsorientiert bewertet werden, also unabhängig von der Verteilung des Merkmals in der Bezugsgruppe. Andererseits ist eine Bewertung ohne jeden Bezug zur Altersgruppe, die sich lediglich an Lernzielen orientiert, eine sehr idealistische Vorgehensweise, die sich in der Praxis nur schwer durchsetzen dürfte.

- Wie sind Modellierungskompetenzen zu operationalisieren? Neben der naheliegenden Variante der Operationalisierung mithilfe von Testverfahren sind auch Interviews oder Verhaltensbeobachtungen denkbar. Verschiedene Ansätze zur Umsetzung sowie jeweilige Vor- und Nachteile werden in Kapitel 1.6.1, sowie in Kapitel 2 diskutiert.
- Aspekte der Datenauswertung und Datenintegration werden in Abschnitt 4.1.4 aufgegriffen und erörtert.

Im Folgenden wird nun zunächst auf Möglichkeiten eingegangen, wie Modellierungskompetenzen operationalisiert werden können. Zwei Forschungsrichtungen sind hierbei zu berücksichtigen: Zum einen sind dies allgemeine Testverfahren zur Erfassung von mathematischer Kompetenz, da diese oftmals Unterskalen zu Modellierungsaufgaben besitzen. Zum anderen ist auf die Methoden einzugehen, die eingesetzt wurden, um Modellierungskompetenz zu erforschen.

1.6.1 Operationalisierungsansätze aus dem Bereich Diagnostik von Mathematikleistungen

Im deutschsprachigen Raum gibt es eine Reihe von Verfahren, die sich mit der Messung von mathematischen Kompetenzen in der Zielgruppe der Grundschulabgänger befassen. Auf die wichtigsten dieser Verfahren soll im Folgenden kurz eingegangen werden:

- ZAREKI-R (von Aster, Weinhold Zulauf & Horn, 2006)

  Zielsetzung des ZAREKI-R ist die Diagnostik von Rechenschwächen mit dem Anspruch der Berücksichtigung der diesen Schwächen zugrunde liegenden Neurophysiologie. Obwohl die Unterskala „Textaufgaben" eine Reliabilität von .85 aufweist, eignet sich das Verfahren doch denkbar schlecht zur Diagnostik der Fähigkeit, Text- oder Modellierungsaufgaben lösen zu können. Dies liegt darin begründet, dass aufgrund von nur sechs Aufgaben, denen aufgrund der klassischen Testkonstruktion eine gleiche Schwierigkeit zugeschrieben wird, nur sehr schlecht zwischen hoher und niedriger Fähigkeit differenziert werden kann. So entsprechen sechs gelöste Aufgaben einem Prozentrang von PR=100, fünf gelöste Aufgaben jedoch nur einem Prozentrang von PR=11. Dass hier keine unterschiedlichen Itemschwierigkeiten berücksichtigt werden, ist vor allem deshalb fragwürdig, da im Testmanual Item 11.3 als besonders schwierig bezeichnet wird.

  Des Weiteren muss berücksichtigt werden, dass von den 764 Personen der Normierungsstichprobe allein 343 Personen aus dritten Klassen in der Schweiz stammen, weshalb eine Übertragbarkeit der Normen sowohl auf andere Klassenstufen als auch auf die Population deutscher Schüler fraglich bleibt.

- DEMAT 4 (Gölitz, Roick & Hasselhorn, 2006)

  Auch die Zielsetzung des DEMAT 4 liegt im frühen Erkennen von möglichen Rechenschwächen. Der Test basiert auf der Schnittmenge der Lehrpläne aus 16 Bundesländern und erfasst damit die zentralen mathematischen Kompetenzen, die in der Grundschulzeit vermittelt werden. Neben einer Skala für die vier Grundrechenarten im Zahlenraum bis eine Million enthält das Verfahren auch eine Skala „Sachrechnen", welche neben Größenvergleichen auch klassische Textaufgaben enthält. Diese Skala weist zwar eine durchaus akzeptable innere Konsistenz auf (α=.74), kann aber mit nur vier Items ebenfalls nur schlecht zwischen verschiedenen Fähigkeitsausprägungen unterscheiden.

- HRT 1-4 (Haffner, Baro, Parzer & Resch, 2005)
  In bewusster Abgrenzung zu den Tests der DEMAT-Reihe orientiert sich der HRT 1-4 nicht an Lehrplänen, sondern versucht vielmehr, mathematische Basiskompetenzen unabhängig von Lehrplänen und Bundesländern zu erfassen. Die Autoren gehen sogar soweit zu sagen, das Verfahren sei darüber hinaus auch kulturübergreifend einsetzbar, jedoch sind die entsprechenden Belege hierfür noch zu liefern. Geprüft wird weniger Schulwissen als „[Schul]mathematik-relevante“ Kompetenzen. Die Subtests sind sehr differenziert und testen zuverlässig das Beherrschen der vier Grundrechenarten sowie einiger weiterer numerisch-logischer Fähigkeiten. Zu bedauern ist, dass hierbei auf Textaufgaben komplett verzichtet wird. Auch wenn diese an sich etwas Schultypisches sind, so sind doch die ihnen eng verwandten Modellierungsaufgaben durchaus nicht nur schul- sondern auch alltagsrelevant.

1.6.2 Operationalisierungsansätze aus der Forschung zu Modellierungskompetenzen

Aus den Beschreibungen der Testverfahren wurde deutlich, dass diese zur Diagnostik von Modellierungskompetenzen eher nicht geeignet sind. Die Gründe hierfür lassen sich wie folgt zusammenfassen:

1. Verfahren, die ein derart breites Spektrum an Kompetenzen abdecken, können die Einzelkompetenzen nur schlecht erfassen und eignen sich somit nicht dafür, zwischen guten und schlechten Modellierern zu differenzieren.
2. Die Skalen zu Textaufgaben sind nicht geeignet, die Einzelfacetten von Modellierungskompetenzen getrennt und unabhängig voneinander zu erfassen.

Auf die zweite Begründung soll nun noch etwas näher eingegangen werden. Es klingt zunächst paradox, wenn man behauptet, Modellierungsaufgaben seien ungeeignet, um Modellierungskompetenzen zu erfassen. Wie könnte man die Kompetenz einer Person besser messen, als durch die direkte Beobachtung bei der Lösung einer Aufgabe?

Angenommen man konstruiert einen Test, der aus PISA-Items mit Modellierungscharakter besteht. Man lässt Probanden diese Aufgaben lösen und ihr Vorgehen möglichst detailliert schriftlich erläutern. Konnte eine Person aus einem ausführlichen Text mit zahlreichen redundanten Informationen die wichtigsten gegebenen Größen identifizieren und angeben, welche Größe gesucht ist, so gibt man ihr die volle Punktzahl auf den Schritt „*Strukturieren*." Gelingt es, die Formel zur Berechnung der gesuchten Größe aufzustellen, so vergibt man Punkte für „*Mathematisieren*" usw. All dies erinnert stark daran, wie im Schulalltag das Lösen von Textaufgaben bewertet wird.

Dies ist ein durchaus vernünftiges Vorgehen, wenn man das Ziel hat, Modellierungskompetenz in ihrer Gesamtheit zu erfassen. Man kann hieraus aber keine Aussagen über die Einzelkompetenzen ableiten, also beispielsweise behaupten: Schüler X kann gut mathematische Ergebnisse interpretieren, weil er für diesen Schritt immer die volle Punktzahl erhalten hat. Woran liegt dies? Die Antwort ist durch das Prinzip der lokalen stochastischen Unabhängigkeit begründet, welche hier nicht gegeben ist.

Die probabilistische Testtheorie setzt voraus, dass die Lösungswahrscheinlichkeit eines Items nur von zwei Größen abhängig ist, nämlich der Itemschwierigkeit einerseits und der Fähigkeit der Person andererseits. Keinesfalls darf die Lösungswahrscheinlichkeit eines Items abhängig sein vom Lösen eines anderen Items innerhalb des gleichen Tests. Wäre dies nämlich der Fall, so wäre die Annahme der lokalen stochastischen Unabhängigkeit verletzt (Bühner, 2006).

Betrachtet man nun aber die einzelnen Schritte beim Durchlaufen des Modellierungszyklus als einzelne Items, so sind spätere Schritte im Prozess immer abhängig von der erfolgreichen Bewältigung der vorhergehenden Schritte. Dies ist auch aus einer praktischen Perspektive einleuchtend: Hat man im ersten Schritt nicht erfolgreich die wichtigsten Größen identifizieren können, so kann man diese im zweiten Schritt nicht in einer mathematischen Gleichung ausdrücken usw.

Hierin wird deutlich, dass man die einzelnen Modellierungskompetenzen nur dann zufriedenstellend erfassen kann, wenn man diese unabhängig voneinander untersucht. Houston (2007) unterscheidet dementsprechend zwischen Verfahren, welche die Einzelkompetenzen erfassen und dem sogenannten „Holistic Assessment“, bei welchem der gesamte Modellierungszyklus anhand einer einzigen Aufgabe durchlaufen und bewertet wird. Diese Differenzierung bezüglich der Diagnostik ist vergleichbar mit der von Blomhoj & Jensen (2003) in Bezug auf die pädagogische Praxis. Sie unterscheiden zwischen holistischem und atomistischem Vorgehen, wobei bei letzterem Schülern isoliert Kompetenzen zu ausgewählten Einzelfacetten der Modellierungskompetenz vermittelt werden.

Wie man beim Unterrichten von Modellierungskompetenzen auch immer vorgeht – bei der Diagnostik muss ein atomistischem Vorgehen eingeschlagen werden. Dadurch ist gewährleistet, dass man tatsächlich auch Aussagen über Einzelkompetenzen ableiten kann.

Welche Ansätzen zur Diagnostik bestehen bereits? In diesem Zusammenhang sind drei verschiedene diagnostische Zugänge zu nennen: Lehrerbeurteilungen aufgrund von Projektberichten, Beobachtungsverfahren und Tests, z.B. in Form von Multiple-Choice-Verfahren (Galbraith, 2007). Im Folgenden wird nacheinander auf die einzelnen Verfahren eingegangen, und es werden exemplarische Vertreter der jeweiligen Kategorien vorgestellt.

1.6.2.1 Lehrerbeurteilungen

Eine auch im Schulalltag verbreitete Methode ist die Bewertung der Modellierungskompetenz durch den Lehrer anhand eines bearbeiteten Modellierungsprojektes. Die ersten systematischen Betrachtungen zu dieser Vorgehensweise (Houston, 2007) finden sich bei Hall (1984) und Berry & Le Masurier (1984). Beide beschäftigen sich mit der Frage, wie man Modellierungsprojekte im tertiären Bildungsbereich am besten erfassen kann. Hall (1984) spricht sich dafür aus, Modellierungskompetenzen im Hinblick auf drei Kriterien zu bewerten, die er als „content", „presentation" und „drive" bezeichnet. Jeder dieser drei Aspekte enthält wiederum Einzelleistungen, wobei für jede dieser Einzelleistungen separat eine Note vergeben wird.

Grob zusammengefasst befasst sich „content" mit den technischen Aspekten des Modellierens, „presentation" mit der Qualität des Projektberichtes und „drive" mit der Originalität des Ansatzes und mit dem Management des Modellierungsprojektes. Während Hall hier einerseits einen ersten Denkansatz liefert, Modellierungskompetenzen möglichst differenziert zu erfassen, macht er sich andererseits mit der Berechnungsvorschrift für die Gesamtnote wieder einen Strich durch die Rechnung. Er fordert nämlich, die Einzelnoten nicht mithilfe des arithmetischen, sondern mithilfe des geometrischen Mittels zu einer Gesamtbewertung zu verrechnen. Hierbei werden die Einzelleistungen nicht addiert, sondern miteinander multipliziert, und aus dem Endprodukt wird die Wurzel gezogen. In der Konsequenz wird die Gesamtleistung 0, wenn ein einziger Multiplikator den Wert 0 hat. Hall begründet dieses unorthodoxe Vorgehen damit, dass es realitätsnaher sei: Wer im richtigen Leben bei einem Modellierungsvorgang an einer einzigen Stelle komplett versagt, für den ist das Endergebnis nicht erreichbar.

Diese Methode mag die Alltagsrealität des Modellierens gut simulieren, für diagnostische Zwecke ist sie aber gänzlich ungeeignet, insbesondere dann, wenn man zwischen Stärken und Schwächen des Probanden bei unterschiedlichen Teilkompetenzen differenzieren möchte.

Auch Berry & Le Masurier (1984) gehen der Frage nach, wie man größere Modellierungsprojekte von Studierenden benoten kann. Ihre Bemühungen richteten sich besonders darauf, ein eher formatives denn summatives Bewertungssystem zu erfinden, das wirklich Punkte für den Prozessablauf gibt und nicht das ideale Modell des Experten bevorzugt. Die Vorgehensweise deckt sich mit der von Hall: Es werden jeweils Unterpunkte vergeben für die Kategorien „Abstract", „Formulation", „Initial Model", „Data", „Revisions to the Model" sowie „Conclusions".

Wie bereits angedeutet, muss von beiden Verfahren am Ende Abstand genommen werden, da sie die lokale stochastische Unabhängigkeit verletzen. Aus testtheoretischer Sicht wären weiterhin die mangelnden Informationen bezüglich der Gütekriterien zu kritisieren. Es wird hier nicht deutlich, inwieweit solche Bewertungssysteme objektiv und reliabel sind, ganz zu schweigen davon, ob man anhand einer einzigen Aufgabe die allgemeine Fähigkeit zum Lösen von Modellierungsaufgaben valide erfassen kann. Dennoch sind diese ersten Schritte auf dem Weg zur Diagnostik von Modellierungskompetenzen nicht unbedeutend, denn sie ziehen als erste die Möglichkeit in Betracht, eine stufenweise Punktevergabe entlang des Modellierungszyklus zu realisieren. Man sollte sich hierbei übrigens nicht von den Kategoriennamen verwirren lassen, die auf den ersten Blick nur wenig mit dem oben beschriebenen Prozessmodell zu tun haben scheinen: Houston & Neill (2003a) gelingt es, sowohl die Kategorien nach Hall (1984) als auch die von Berry & Le Masurier (1984) den Stufen im Modellierungszyklus zuzuordnen.

1.6.2.2 Beobachtungsverfahren

Das Verfahren von Berry & Le Masurier (1984) hat den Anspruch, mehr den Prozess des Modellierens zu bewerten als die letztlichen Ergebnisse. Dennoch werden Informationen über diesen Prozess gewonnen, indem man sich verschiedene Stellen im Endprodukt (dem Projektbericht) genauer anschaut.

Ein besseres Verfahren, um den Modellierungszyklus direkt am Geschehen zu betrachten, bieten diverse Beobachtungsverfahren. Hier sind einerseits Videostudien zu nennen, andererseits aber auch Protokolle lauten Denkens sowie Ratingskalen, quasi als Mischfrom zwischen Lehrerbeurteilungs- und Beobachtungsverfahren.

Greefrath (2006) filmte in Fallstudien jeweils zwei Schüler beim Lösen von Aufgaben. Im Anschluss wurde codiert, wie viel Zeit für Planung, Datenbeschaffung, Datenverarbeitung und Kontrolle aufgewendez wurde. Greefraths Ziel ist es hier lediglich, Zeitanteile für die jeweiligen Phasen zu identifizieren. Weiter geht der Ansatz von Borromeo Ferri (2006), die ebenfalls Gruppenaufgaben filmte, mit dem Ziel, genauere Informationen über den Ablaufprozess des Modellierungszyklus zu erhalten.

Das Vorgehen an sich ist vergleichbar mit Protokollen lauten Denkens. Statt jedoch Schüler gezielt zum lauten Denken aufzufordern, sind diese in Zweiergruppen gezwungen, ihre Gedanken laut auszusprechen, um mit ihren Gruppenmitgliedern aktiv und kooperativ zusammenarbeiten zu können.

Auch mithilfe von Ratingskalen können Beobachtungen realisiert werden. Zwar kann man Ratingskalen auch einsetzen, um eine fertig bearbeitete Aufgabe hinsichtlich verschiedener Gesichtspunkte einzuschätzen (Wie beispielsweise „Did the student identify the key mathematical focus of the problem? Were relevant variables identified? …“ aus dem Verfahren von Ideka & Stephen, 1998, S.227), sie können aber auch im Sinne einer formativen Bewertung eingesetzt werden, um während des Modellierungszyklus jeweils Punkte für verschiedene Einzelkompetenzen zu vergeben.

Bei der Frage nach der besten Variante des Modellierungszyklus wurde darauf hingewiesen, dass diese nur beantwortet werden kann, wenn man die jeweilige Zielsetzung (didaktisch, diagnostisch oder aus der Perspektive des Erkenntnisgewinns) in Betracht zieht. Gleiches gilt auch für die Frage nach der besten Methode zur Erfassung von Modellierungskompetenzen. Zu Forschungszwecken sind Videoanalysen mit Sicherheit praktikabel und liefern eine Menge interessanter Informationen über interne und externe Prozessabläufe. Als diagnostisches Instrument sind sie jedoch wenig nützlich.

1.6.2.3 Multiple-Choice-Verfahren

Alle bisher beschriebenen Verfahren haben zwar offenbar den Anspruch an eine atomistische Betrachtungsweise, da sie intensiv die einzelnen Stufen des Modellierungszyklus untersuchen, unterliegen jedoch gleichzeitig immer noch der größten Schwäche holistischen Vorgehens, indem sie den gesamten Modellierungszyklus anhand einer großen Aufgabe untersuchen und damit die lokale stochastische Unabhängigkeit zwischen den Einzelkompetenzen verletzen.

Zwar schrieb einer der Pioniere auf dem Gebiet der Diagnostik von Modellierungskompetenzen: „The only direct method of assessing modelling skills is through a project done over A long period.” (Hall, 1984, S. 143), jedoch ist diese Aussage mittlerweile deutlich in Zweifel zu ziehen.

Eine ebenfalls praktikable Methode ist nämlich die Verwendung von Multiple-Choice-Fragen, die sich auf unterschiedliche Stufen im Modellierungszyklus beziehen und jeweils unterschiedliche Problemstellungen behandeln.

In diese Kategorie fallen die zahlreichen Arbeiten von Haines und Kollegen (Crouch & Haines, 2003, 2004; Haines & Crouch, 2001, 2005; Haines, Crouch & Davis, 2000; Haines, Crouch & Davis, 2001; Haines, Crouch & Fitzharris, 2003; Izard, Haines, Crouch, Houston & Neill, 2003), die mit ihrem Test eine Lücke schließen wollen, da „little attention has so far been paid to recognizing attainment levels of students at defined stages of a modelling process (Haines & Crouch, 2001, S. 129)". Die Autoren haben ein Verfahren entwickelt mit dem Ziel, einen „Snapshot" von Modellierungskompetenzen zu gewinnen, also das Modellieren an verschiedenen Stellen des Kreislaufs punktgenau zu erfassen, ohne dass der ganze Modellierungszyklus durchlaufen werden muss. Dies entspricht im Gegensatz zu den vorher beschriebenen Verfahren zum ersten Mal tatsächlich einem atomistischen Blickwinkel auf Modellierungskompetenzen.

Der Fragebogen existiert in zwei Parallelversionen zu je zwölf Aufgaben, die in ca. 20-30 Minuten zu bearbeiten sind.

Die Einzelitems beziehen sich auf unterschiedliche Stufen im Modellierungszyklus, sind aber in jeweils unterschiedliche Settings eingebunden. Die Autoren begründen dieses Vorgehen damit, dass sie Verwirrungen und gegenseitige Beeinflussungen ausschließen wollten. Dies ist nur eine andere Ausdrucksweise dafür, dass in diesem Testverfahren die lokale stochastische Unabhängigkeit angestrebt ist.

Abbildung 8 enthält zur Veranschaulichung zwei Items aus Version A, wobei sich das erste mit Schritt 1: *Strukturieren* und das zweite mit Schritt 2: *Mathematisieren* beschäftigt.

Item 1:
Consider the real world problem (do **not** try to solve it!):
A bus stop position has to be placed along a road on a new bus route. A covered shelter will be provided. Where should the stop be placed so that the greatest number of people will be encouraged to use the service? The bus company wants people to use the service but cannot lay on buses on demand.
Which **one** of the following assumptions do you consider the least important in formulating a simple mathematical model?

A: Assume that just one bus shelter will be erected.
B: Assume that the road is straight
C: Assume that the weather is twice as likely to be dry as it is to be wet.
D: Assume that the bus runs to a half-hour timetable
E: Assume that customers will not walk great distances to catch a bus

Item 6:
Which **one** of the following options most closely models the height of a sunflower while it is growing (in terms of time t))

A: $1-e^{-t}$
B: $(1-t)^2$
C: t
D: $t-t^2$
E: $\frac{1}{1+e^{-t}}$

Abbildung 8: Beispielitems aus Haines, Crouch & Davis (2001)

Das gesamte Testverfahren wurde in mehreren Studien erprobt, sowohl von den Autoren als auch von Dritten, so beispielsweise von Houston & Neill (2003a; 2003b) in Irland, sowie von Lingefjärd & Holmquist in Dänemark (2005). Kaiser & Schwarz (2006) setzten darüber hinaus eine deutsche Version in der gymnasialen Oberstufe ein. Die Items stellten sich als reliabel (r_{AB}=.88) heraus und sind darüber hinaus Rasch-konform, was für die Güte des Tests spricht. Da die Distraktoren nicht alle als „gleich falsch“ angesehen werden können, sondern einige falsche Antworten durchaus auf ein mathematisches Verständnis hindeuten können, wird ein Partial-Credit-Model zur Auswertung herangezogen. Die Parallelität der Verfahren kann als gewährleistet betrachtet werden, da sich die Schwierigkeiten der jeweiligen Parallelitems nie um mehr als eine Standardabweichungen unterscheiden. Dies liegt bereits in der Konstruktion der Parallelaufgaben begründet, wobei das gleiche Problem jeweils nur in einem anderen Kontext dargestellt wird (Straßenbahnhaltestellen statt Bushaltestellen, Check-In-Schalter am Flughafen statt Supermarktkassen).

Somit kann dieser Ansatz im Hinblick auf eine standardisierte, individuelle Diagnostik als Methode der Wahl betrachtet werden, da hier exklusiv einerseits Modellierungskompetenzen unabhängig voneinander untersucht werden und andererseits gleichzeitig die lokale stochastische Unabhängigkeit gewahrt bleibt. Videostudien oder andere Beobachtungsverfahren sind wiederum besser geeignet für stark individuumszentrierte Fallstudien, die weit ins Detail gehen, sowie für hypothesengenerierende Untersuchungen.

Ein weiterer – durch die Unabhängigkeit der Einzelitems – begründeter Vorteil des Verfahrens liegt darin, dass innerhalb kurzer Zeit eine ganze Reihe von Modellierungssituationen bearbeitet werden können. Gelingt es einem Proband nicht, eine Aufgabe richtig zu verstehen, so entgehen ihm nur Punkte für dieses eine Item. Wird hingegen nur eine einzige Aufgabe gestellt, die der Proband nicht versteht, so dürfte es ihm schwer fallen, im ganzen darauffolgenden Modellierungszyklus noch Punkte zu erlangen. Dies beantwortet wohl deutlich die von Galbraith (2007, S.84) aufgeworfene Frage: „Can modelling ability be estimated, on the basis of performance on a single common problem, or is a pattern of performance across several problems needed to provide a reliable estimate?"

Eine Schwäche des Instrumentes von Haines et al. (z.B. Haines & Crouch, 2001) liegt allerdings darin, dass es sich auf den Übergang von der realen Welt zur Mathematik, also eigentlich nur auf die Phasen 1 und 2 beschränkt. Um den ganzen Modellierungszyklus zu erfassen müsste ein Test entwickelt werden, der alle Einzelphasen abdeckt. Des Weiteren kritisiert Galbraith (2007, S.83): „A challenge remains to relate performance on such sub-skills to overall modelling ability." Testverfahren wie das von Haines et al. müssen also auch den Nachweis der Validität erbringen.

Dies kann zum Beispiel dadurch realisiert werden, dass neben dem Instrument selbst auch noch komplette Modellierungsaufgaben vorgegeben werden, so dass im Anschluss eine Validierung durch die Korrelation beider Leistungsmaße vorgenommen werden kann.

Aus der Sichtung der Literatur wird deutlich, dass sich für die Perspektive der Diagnostik folgende Konsequenzen ergeben:

- Als Grundlage zur individuellen Förderung müssen Wege gefunden werden, möglichst frühzeitig – also schon in der Primarstufe – Modellierungskompetenz zu erfassen.
- Dies muss möglichst differenziert erfolgen, weshalb eine Orientierung an den Teilkompetenzen angezeigt ist. Diese müssen unabhängig voneinander erfasst werden, damit aufgezeigt werden kann, wo jeweilige Stärken und

Schwächen eines Schülers liegen und worauf sich infolgedessen die Förderung vornehmlich konzentrieren sollte.

- Ein Testverfahren, dass dies leisten soll, muss den klassischen psychometrischen Gütekriterien genügen und insbesondere den Nachweis der Validität liefern.

Damit dies sinnvollerweise realisiert werden kann, muss jedoch noch eine weitere Bedingung erfüllt sein. Der Modellierungszyklus lässt sich auffassen als ein Wirkmodell (Gollwitzer & Jäger, 2007), das die Vorgänge beim mathematischen Modellieren sowohl in deskriptiver als auch in präskriptiver Weise verdeutlicht. Dieses Wirkmodell gilt es jedoch zuerst empirisch zu überprüfen, bevor es als Grundlage für die Entwicklung von Diagnoseinstrumenten eingesetzt werden kann.

2. Fragestellung und Operationalisierung

Aus den bisherigen Ausführungen wurde die Relevanz deutlich, die der Diagnostik von Modellierungskompetenzen beizumessen ist.

Da der Modellierungskreislauf als Modell jedoch bis dato in erster Linie auf theoretischen Überlegungen gründet, ist es zunächst angezeigt, der Frage nachzugehen, ob der Modellierungszyklus überhaupt als deskriptives Modell des Vorgehens beim Lösen von mathematischen Textaufgaben geeignet ist. Es sind demnach die folgenden Fragen zu beantworten:

1. Lassen sich die im Modellierungszyklus genannten Teilkompetenzen *Strukturieren*, *Mathematisieren*, *Verarbeiten*, *Interpretieren* und *Validieren* tatsächlich beobachten?

2. Welche der zu den jeweiligen Teilkompetenzen formulierten Unterkompetenzen lassen sich beobachten?

3. Treten die Teilkompetenzen in der im Modellierungszyklus postulierten Reihenfolge auf?

Diese Fragen sollen in einer ersten – eher fallstudienartig angelegten – Untersuchung beantwortet werden, bei der an einer geringen Anzahl von Versuchspersonen der Modellierungsprozess möglichst genau nachvollzogen werden soll.

Des Weiteren ist darauf aufbauend zu untersuchen, inwieweit der Modellierungszyklus als normatives Modell dienen kann. Wenn also tatsächlich der Modellierungszyklus als deskriptives Modell geeignet ist, so muss im nächsten Schritt untersucht werden, ob die Qualität bzw. Erfolgswahrscheinlichkeit beim Lösen einer Aufgabe tatsächlich von allen postulierten Teilkompetenzen abhängig ist.

Wie im theoretischen Teil herausgearbeitet wurde, wäre es nützlich, über ein Instrument zu verfügen, mit dem man die unterschiedlichen Modellierungskompetenzen unabhängig voneinander zuverlässig erfassen kann. Um ein solches Instrument zu entwickeln, müssen auf der Basis der theoretischen Grundlage zunächst Items entwickelt werden, die dann dahingehend zu überprüfen sind, ob sie überhaupt Modellierungskompetenz messen können und wenn ja, wie gut. Ersteres betrifft den Aspekt der Validität, letzteres den der Reliabilität.

Diese letztgenannten Aspekte sollen in einer zweiten, größer angelegten Studie untersucht werden, wofür zunächst Items zu den in der ersten Studie beobachteten Teilkompetenzen zu konstruieren sind, die dann an Schülern am Ende der vierten Klassenstufe zu erproben sind. Dabei stehen folgende Fragen im Vordergrund:

1. Wie ist es um die Qualität der konstruierten Items bestellt? Messen sie Modellierungskompetenz objektiv, reliabel und valide?

2. Trägt jede durch diese Items operationalisierte Teilkompetenz einen wichtigen Teil zur allgemeinen Modellierungskompetenz bei, so dass tatsächlich alle Teilkompetenzen beherrscht werden müssen, um erfolgreich zu modellieren?

2.1 Operationalisierung des Konstrukts „Modellierungskompetenz“

Um all diese Fragestellungen beantworten zu können, bedarf es einer intensiven Beschäftigung mit der Modellierungskompetenz von Personen. Da diese jedoch als latentes Konstrukt nicht direkt beobachtbar ist, muss sie durch Operationalisierung beobachtbar gemacht werden (Jäger & Petermann, 1999).

Welche Art der Operationalisierung jeweils zielführend ist, hängt ab von der Art der Fragestellung. Für Studie 1, in der es um die deskriptive Funktion des Modellierungszyklus geht, sind Beobachtungsverfahren angezeigt.

Um aber – wie in Studie 2 – zu überprüfen, ob die Beherrschung einer Kompetenz letztlich einen Vorteil bringt, so dass man auch von einem normativen Modell sprechen kann, reicht es nicht alleine aus zu beobachten, sondern die Leistung muss direkt gemessen werden. Hierzu sind dann Testverfahren die Methode der Wahl.

Im folgenden Abschnitt wird zunächst darauf eingegangen, wie man Modellierungskompetenz allgemein messbar machen kann und welche Besonderheiten bzw. Schwierigkeiten sich daraus ergeben. Im Anschluss daran werden die Unterkompetenzen, welche bisher nur grob dargestellt wurden, erneut im Hinblick darauf thematisiert, was sich jeweils hinter diesen verbirgt. Auf der Grundlage dieser Erläuterungen wird dann darauf eingegangen, wie die Teilkompetenzen für die beiden Studien jeweils operationalisiert wurden.

2.1.1 Operationalisierung von Modellierungskompetenz

Weiter oben wurden bereits die verschiedenen Facetten mathematischer Kompetenz angesprochen, insbesondere die im KOM-Projekt identifizierten acht Teilkompetenzen (Niss, 2003), von denen eine die Modellierungskompetenz darstellt. Die Autoren des KOM-Projektes wehren sich gegen die „Syllabusitis“ und argumentieren, man könne mathematische Kompetenz nicht gleichsetzen mit „proficiency in mathematical substance matter“ (Blomhoj & Jensen, 2007, S.46) und dann einfach Bereiche der Mathematik auflisten. Vielmehr argumentieren sie für einen zweidimensionalen Ansatz, bei dem auf der einen Seite mathematische Inhalte stehen, auf der anderen Seite die von ihnen identifizierten acht Kompetenzen.

Dies ähnelt stark den in den deutschen Bildungsstandards (KMK, 2005) formulierten mathematischen Kompetenzen einerseits und den mathematischen Leitideen andererseits. Beschäftigt man sich nun mit der Konstruktion bzw. der Klassifikation von Items, so lassen sich Kompetenzen und Leitideen im Sinne einer Tyler-Matrix (Gollwitzer & Jäger, 2007) anordnen (siehe Abbildung 9):

	Leitidee 1	Leitidee 2	Leitidee 3	Leitidee 4	Leitidee 5
Kompetenz 1					
Kompetenz 2					
Kompetenz 3					
Kompetenz 4					
Kompetenz 5					
Kompetenz 6					

Abbildung 9: Kompetenzen und Leitideen

Hierbei ist es durchaus möglich, dass ein Item mehreren Leitideen gleichzeitig zugeordnet wird, wenn z.B. eine Aufgabe zu Strahlensätzen (Leitidee 3) bei ihrer Lösung den Umgang mit Bruchtermen (Leitidee 1) erfordert. Gleichzeitig dürften viele Aufgaben auch mehrere Kompetenzen erfassen.

Beschäftigt man sich ausschließlich mit Modellierungsproblemen, so beschränkt sich die Kompetenzdimension auf die Kompetenz 3: Modellieren. Allerdings sind auch hierbei wieder Ausdifferenzierungen zu betrachten: Jedes Modellierungsitem kann Anforderungen an alle oder nur einen Teil der im Modellierungszyklus ablaufenden Teilkompetenzen stellen. Die dazugehörige Tyler-Matrix entspricht Abbildung 10:

	Leitidee 1	Leitidee 2	Leitidee 3	Leitidee 4	Leitidee 5
Strukturieren					
Mathematisieren					
Verarbeiten					
Interpretieren					
Validieren					

Abbildung 10: Leitideen und Teilkompetenzen

Ein ähnliches Modell schlägt Jensen (2007) vor, dieses enthält jedoch drei Dimensionen: "Degree of Coverage", "Radius of Action" sowie "Technical Level".

“Degree of Coverage” bezeichnet hierbei den oder die Bereiche des Modellierungskreislaufes, die von der Aufgabenstellung abgedeckt werden. „Radius of Action“ bezieht sich auf die jeweils angesprochene Domäne der Mathematik – in unserem Modell also auf die jeweilige mathematische Leitidee. Hinzu kommt die dritte Dimension „Technical Level“, welche den Komplexitätsgrad der Aufgabe hinsichtlich der innermathematischen Anforderungen charakterisiert. Man könnte hier auch von Anforderungsgrad oder Itemschwierigkeit sprechen. Dieser ist höher bei einer Aufgabe, die das Aufstellen und Lösen einer quadratischen Gleichung erfordert, als bei einer Aufgabe, in der lediglich einige Zahlenwerte addiert werden müssen.

Auf Basis dieser drei Komponenten Jensens resultiert ein dreidimensionales Modell der Kompetenz, wie es in Abbildung 11 dargestellt ist.

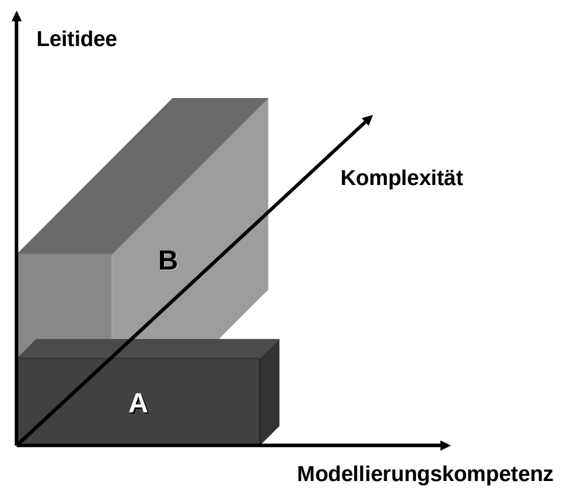


Abbildung 11: Dimensionsmodell der Kompetenz

Aus Abbildung 11 wird deutlich, dass jedes Item und dementsprechend auch die Itemschwierigkeit – was dann bei der Auswertung und Itemanalyse von Bedeutung ist – von (mindestens) drei Größen beeinflusst wird. In Abbildung 11 sind beispielhaft zwei Aufgaben A und B eingezeichnet. Während A ein größeres Spektrum an Modellierungskompetenz erfasst, beinhaltet B mehrere Leitideen und zeichnet sich durch eine größere Komplexität aus.

Enthält nun ein Diagnoseinstrument beide Items A und B, und ergibt sich in einer fiktiven Stichprobe eine höhere Schwierigkeit für Item B als für Item A, so stellen sich folgende Fragen:

- Haben die Probanden nicht die ausreichende mathematische Kompetenz, um Aufgaben mit einem höheren Komplexitätsgrad zu lösen?
- Oder haben die Probanden Schwierigkeiten mit einigen Leitideen, mit anderen aber nicht?
- War Item B vielleicht sogar hinsichtlich seiner Anforderungen an die Modellierungskompetenz schwieriger, obwohl weniger Teilschritte des Modellierungszyklus angewandt werden mussten, diese aber besonders komplex waren?

An diesen Fragen wird deutlich, dass beide Items nicht geeignet sind, um Modellierungskompetenzen (oder jedwede andere der genannten Dimensionen) zu erfassen, da nicht mit Sicherheit festgestellt werden kann, auf welches Konstrukt die Schwierigkeit (und andere Eigenschaften, so z.B. Korrelationen mit der Gesamtskala) zurückzuführen ist.

Die Eindimensionalität des Verfahrens kann also nur künstlich hergestellt werden, indem entsprechende Items so konstruiert werden, dass die jeweils nicht interessierenden Dimensionen konstant gehalten werden.

Da sich bei der vorliegenden Arbeit das Augenmerk auf Modellierungskompetenzen richtet, wurden die Aufgaben für beide Studien so konstruiert, dass gleichzeitig zwei Dinge gewährleistet sind: Erstens wurde die Leitidee konstant gehalten, indem sich die Aufgaben immer auf die „Leitidee 2: Messen“ bezogen. Gleiches gilt für den Komplexitätsgrad – da die Fähigkeit zum Rechnen nur von sekundärem Interesse ist, wurden jeweils Aufgaben konstruiert, bei denen nur einfache Operationen an ebenfalls einfachen Zahlenwerten vorzunehmen sind. Zweitens wurde bezüglich der Modellierungskompetenz sicher gestellt, dass sich Aufgaben nicht in ihrem „Degree of Coverage“ unterscheiden: In der ersten Studie wurden lediglich Aufgaben eingesetzt, die den gesamten Modellierungszyklus abdecken, in der zweiten Studie wurden die verwendeten Multiple-Choice-Items so konstruiert, dass jeweils nur ein Schritt des Modellierungskreislaufs untersucht wird, womit auch wiederum eine Vergleichbarkeit der Aufgaben untereinander gewährleistet ist.

2.1.2 Operationalisierung der Unterkompetenzen

Nach Maaß (2004) sind den Teilkompetenzen mathematischen Modellierens die in Tabelle 4 wiedergegebenen Unterkompetenzen zuzuordnen (Übersetzung durch Autorin).

Tabelle 4: Unterkompetenzen

	Strukturieren
S1	Textverstehen / Problem in eigenen Worten wiedergeben
S2	Größen benennen: Gegeben / Gesucht
(S3	Beziehungen zwischen Variablen rekonstruieren)
S4a	Suche nach problemrelevanten Informationen
S4b	Trennung zwischen relevanter – irrelevanter Information
	Mathematisieren
M1	Mathematisieren von Größen und Beziehungen (Zahlwörter => arabische Ziffern) (Verbalisierungen => Operationen)
(M2	Deren Anzahl und Komplexität verringern)
M3	Angebrachte mathematische Notationen / Skizzen (= formal korrekte Schreibweise)
	Verarbeiten
V1	Mathematisches Wissen zur Problemlösung nutzen (z.B. korrekt schriftlich Rechnen)
V2	Heuristische Strategien („geschickt Rechnen“)
	Interpretieren
I1	Interpretation im außermathematischen Kontext
I2	Verallgemeinerung der Lösung (fak.)
I3	Kommunikation der Ergebnisse in math. Sprache

Zu bedenken ist allerdings, dass dieses Schema, wie so viele Arbeiten im Bereich mathematisches Modellieren, für den tertiären Bildungsbereich entwickelt wurde und schon auf den sekundären Sektor nur mit Abstrichen zu übertragen ist. Will man die von Maaß (2004) genannten Unterkompetenzen auch als dem Modellierungsverhalten von Grundschülern zugrunde liegend betrachten, so ist nachvollziehbar, dass in einem gewissen Rahmen Einschränkungen in Kauf zu nehmen sind. Dies betrifft zum einen die Tatsache, dass nicht alle Unterkompetenzen berücksichtig werden können, da diese sich auf Vorgehensweisen beziehen, die nur bei der Bearbeitung von komplexen Aufgaben zu finden sein dürften. Die Teilkompetenzen S3 und M2 müssen daher aus den an entsprechender Stelle aufgeführten Gründen ausgeklammert werden. Zum anderen gilt für die verbleibenden Unterkompetenzen, dass sich die konkreten Operationalisierungen auf Vereinfachungen oder im Extremfall nur Vorstufen der bei komplexen Problemen benötigten Unterkompetenzen beschränken müssen.

Auf der Grundlage der Aufsplittung in Unterkompetenzen lässt sich der Modellierungszyklus feinkörniger darstellen, als dies bisher geschah (siehe Abbildung 12):

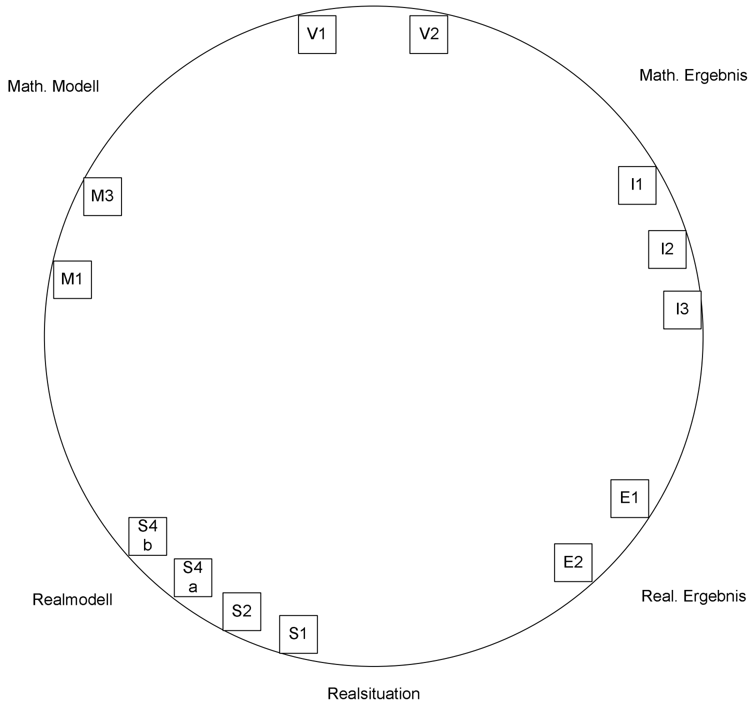


Abbildung 12: Modellierungszyklus mit Unterkompetenzen

Im Folgenden sollen nun die Unterkompetenzen – besonders im Hinblick auf deren Operationalisierung – genauer erläutert werden.

Strukturieren:

Strukturieren ist der erste Schritt im Modellierungszyklus und beinhaltet in erster Linie das Herausfinden des eigentlichen Problems. Bei Textaufgaben geht es zudem darum, sich ein „Bild" von der Situation zu machen – sich also vorzustellen, wie die reale Situation überhaupt aussieht. Davon ausgehend wird dann abstrahiert und sich auf die wesentlichen – problemrelevanten – Merkmale der Situation konzentriert.

S1: *„To make assumptions for the problem and simplify the situation"*. Diese Fähigkeit bezieht sich auf das Verstehen und Vereinfachen des Aufgabentextes. Die Aufgabe muss zunächst gelesen werden. Wird sie verstanden, so entsteht eine mentale Repräsentation der Situation im Gedächtnis des Lesers. Diese ist nicht mehr gebunden an den Wortlaut der Aufgabe, sondern der Proband ist nun in der Lage, die Problemstellung in eigenen Worten wiederzugeben und die zentralen Fragen und Merkmale hervorzuheben. Operationalisiert werden kann diese Kompetenz durch Items zum Textverständnis, z.B. in Form von Verständnisfragen oder Paraphrasierungen. S1 kann als zentraler Vertreter der Kategorie „*Strukturieren*“ betrachtet werden.

Bei S2 „*To recognize quantities that influence the situation, to name them and to identify key variables.*“ geht es um das Erkennen der für die Problemstellung relevanten Größen. Dieser Schritt ist verwandt mit dem im Mathematikunterricht üblichen Vorgehen, Schüler identifizieren zu lassen, was bei einer Aufgabenstellung als „gegeben“ und was auf der anderen Seite als „gesucht“ zu betrachten ist. Hierbei wird im Verständnisprozess noch einen Schritt weiter gegangen, als dies in S1 der Fall ist: Aus der nunmehr mental repräsentierten Problemstellung werden diejenigen Schlüsselkonzepte herausgefiltert, aus denen sich im nächsten Schritt die mathematische Version der Problemstellung formulieren lässt. Auf dieser Basis lässt sich auch eine Operationalisierung der Unterkompetenz realisieren, nämlich durch Vorgabe von Aufgabentexten zu denen dann jeweils die gegebenen und gesuchten Größen zu identifizieren sind.

Die unter S3 genannte Unterkompetenz „*to construct relations between the variables*” wird im Folgenden nicht weiter beachtet und aus dem zu konstruierenden Testverfahren ausgeklammert. Diese Entscheidung liegt zum einen darin begründet, dass das Erkennen von Beziehungen zwischen Variablen in mathematischen Textaufgaben für Grundschüler nicht vorgesehen ist, da keine komplexen Variablengeflechte zu berücksichtigen sind; und zum anderen in der dem ersten Faktor zu Grunde liegenden Tatsache, dass eine solche Anforderung Grundschüler überfordern würde.

Im Zuge der unter S2 angesprochenen Identifikation der problemrelevanten Information wird im Aufgabentext und in eventuellen Abbildungen in dessen Nähe nach den konkreten Informationen (z.B. Zahlenwerten) gesucht, die benötigt werden, um das Problem erfolgreich zu lösen. Hierbei sind zwei Teilaspekte von Bedeutung: Zum einen ist dies die Suche nach verfügbaren Informationen S4a „*To look for available information*“, zum anderen – zumindest bei komplexeren Problemen – das Unterscheiden von relevanten und für die Aufgabenlösung überflüssigen Informationen, S4b „*to differentiate between relevant and irrelevant information*“. Es kann hierbei angenommen werden, dass auch *nach* dem Verstehensprozess der Aufgabentext ein zweites Mal durchgelesen oder zumindest überflogen wird, um eben diese relevanten Größen für die nächsten Schritte verfügbar zu machen. Die Bedeutung der beiden Schritte soll an zwei Extrembeispielen verdeutlicht werden:

Die Suche nach verfügbarer Information kann durchaus groteske Formen annehmen. Bildungswissenschaftler wie Verschaffel und seine Kollegen (Verschaffel, Greer & De Corte, 2000) kritisieren, dass Schüler Textaufgaben lediglich nach Zahlenwerten absuchen würden, um diese dann „irgendwie“ halbwegs sinnvoll miteinander zu verrechnen. Um ein Verstehen des Textes würde sich nur in den allerseltensten Fällen bemüht. In unserer Sprache heißt das: Schüler neigen dazu, den Schritt S1 zu überspringen, sich also keine Gedanken über die Problemsituation zu machen, geschweige denn sich diese tatsächlich vorzustellen oder gar an ihrem Realitätsgehalt zu zweifeln. Vielmehr versuchen sie S2, also das Erkennen der Zielsetzung und der wichtigsten Größen der Aufgabe aus den mithilfe von S4a herausgefilterten Informationen und einigen Schlüsselwörtern zu realisieren. Ein von Dehaene (1997, S.138) berichtetes Beispiel illustriert, wie dies auch Erwachsenen passiert, wenn sie einen Aufgabentext nicht wirklich sorgfältig durchlesen:

“Judy owns five dolls, which is two fewer than Cathy.
How many dolls does Cathy have?”

Liest man den Text nicht aufmerksam, ist man verleitet ohne weiteres Nachdenken die Antwort „drei“ zu geben. Bestimmte Schlüsselwörter wie „weniger als“ aktivieren eher stärker das Schema Subtraktion, obwohl hier eigentlich addiert werden muss. Entsprechende Fehler können nur dann vermieden werden, wenn der Schritt S1 sorgfältig durchgeführt wird. Verschaffel et al. (2000) konnten anhand gezielt unsinnig konstruierter Aufgaben zeigen, dass die meisten Schüler (untersucht wurden hier Personen aus dem Sekundarbereich) dies eben nicht tun. Sie legten den Probanden Aufgaben vor, die eigentlich nicht zu lösen sind, stellten jedoch fest, dass die Mehrzahl der Schüler trotzdem die enthaltenen Zahlen miteinander verrechnete, um ein (inhaltlich völlig sinnfreies) Ergebnis zu präsentieren. Zwei Beispiele (aus Verschaffel et al., 2000, S.19f.):

Bsp. 1: “What will be the temperature of water in a container if you pour 1 liter of water at 80° and 1 liter of 40° into it ?”

Bsp. 2: “John’s best time to run 100 meters is 17 seconds. How long will it take him to run 1 kilometer?”

83% der untersuchten Schüler gaben an, das Wasser aus Aufgabe 1 habe eine Temperatur von 120°C und ganze 97% vertraten ohne Zögern die Ansicht, der Läufer würde einen Kilometer in 170 Sekunden laufen. Diese Ergebnisse machen deutlich, dass bei der Informationssuche in S4a kein wirkliches Textverständnis erfolgt – was umso mehr die Bedeutung von S1 unterstreicht.

Auf der anderen Seite zeigt sich, dass auch S4b – das Trennen von relevanter und irrelevanter Information - durchaus nicht trivial ist. Im Rahmen einer unveröffentlichten Studie des Zentrums für empirische pädagogische Forschung aus dem Jahr 2006 wurde 327 Schülern der Klassen 8-10 an Hauptschulen, Realschulen und Gymnasien folgende Frage vorgelegt:

Ein 43.5 Jahre alter Bauer verkauft 3.45 kg Äpfel für 4.53 €.
Was kosten 45.3 kg Äpfel?
Welche in der Aufgabenstellung genannte Zahl muss man für die Lösung dieser Aufgabe nicht weiter beachten?

a)	43.5	b)	45.3
c)	3.45	d)	4.53

Es zeigte sich, dass 22% aller Befragten nicht in der Lage waren, die Aufgabe korrekt zu lösen. Woran genau dies liegt, ist schwer zu entscheiden und wird noch zu untersuchen sein. Nahe liegt jedoch der Verdacht, dass es Schülern schwer fällt zu erkennen, welche Informationen im Aufgabentext relevant und welche irrelevant sind – die Vermutung liegt also nahe, dass die unter S4b bezeichnete Kompetenz nur mangelhaft ausgeprägt ist.

Die mit der Informationssuche zusammenhängenden Unterkompetenzen S4a und S4b lassen sich operationalisieren, indem man Problemstellungen vorlegt, die sowohl problemrelevante als auch problemirrelevante Informationen enthalten, wobei für die problemrelevanten Informationen gilt, dass sie nicht immer direkt auf den ersten Blick aus dem Aufgabentext erschlossen werden können, sondern zum Beispiel aus Zusatztexten, Karten oder Skizzen abzulesen sind.

Zusammengenommen lässt sich *Strukturieren* auch auffassen als das Verstehen einer Aufgabenstellung im kognitionspsychologischen Sinn, wobei es dem Modellierer gelingen muss, die wesentlichen Inhalte eines Textes zu erfassen, um daraus eine mentale Repräsentation des Problems konstruieren zu können.

Mit der aktiven Suche nach Information endet der erste Schritt im Modellierungszyklus, an den sich der zweite Schritt – das *Mathematisieren* der Problemstellung anschließt.

Mathematisieren

Das *Mathematisieren* von verbalen Aufgabenelementen hat je nach Art der Aufgabenstellung einen größeren bzw. kleineren Stellenwert innerhalb des Lösungsprozesses. Es sind Arten von Aufgaben denkbar, in denen nur sehr wenig mathematisiert werden muss. Man betrachte beispielhaft folgende Problemstellung:

Lisa ist 7 Jahre alt, Anna ist 5 Jahre älter.
Wie alt ist Anna?

Hinter dieser Aufgabenstellung versteckt sich die Rechnung „7+5=☐". Die Mathematisierung – also die Übersetzung des sprachlichen in einen numerisch-logischen Code – ist in diesem Falle recht trivial. Die beiden zu verrechnenden Zahlen sind bereits als arabische Ziffern vorgegeben, übersetzt werden muss nun nur noch die entsprechende Rechenoperation. „Älter" ist neben „mehr als" und anderen Relationsbegriffen eines der typischen Schlüsselwörter, das auf eine Addition hindeutet. Das oben zitierte Beispiel von Dehaene (1997) zeigt allerdings auch, dass solche Schlüsselwörter in die Irre führen können.

Anspruchsvolle Modellierungsaufgaben verlangen im Gegensatz zu diesem Beispiel komplexere Übersetzungsprozesse. Dies kann geschehen, indem Zahlenwerte im analogen Format präsentiert werden, indem es dem Problemlöser erschwert wird, die richtigen Zahlen und Operationen aus dem Text herauszulesen oder indem mehrere Operationen hintereinander geschaltet werden.

Die erste der drei Subkategorien im zweiten Schritt, M1 „to *mathematize relevant quantities and their relations*“, betrifft – wie dies auch schon beim *Strukturieren* der Fall war – den eigentlichen Kern der Teilkompetenz. Es geht hierbei um das Übersetzen von Größen und Relationen in die mathematische Sprache, also um das Erkennen von Zahlen und Größen sowie Operationen, die sich hinter verbalen Formulierungen verbergen.

Die Operationalisierung der Unterkompetenz gelingt dadurch, dass die zu übersetzenden Größen und Relationen in einem Format dargestellt werden, das möglichst wenig Ähnlichkeit mit der mathematischen Sprache hat, um diese dann der korrespondierenden mathematisierten Version zuordnen zu lassen.

M2 entfällt bei der Betrachtung von Grundschulkindern. Die Kompetenz, Anzahl und Komplexität der gefundenen Variablen zu reduzieren („*to simplify relevant quantities and their relations if necessary and to reduce their number and complexity*“) dürfte für diese Zielgruppe noch nicht relevant sein. Wie bereits erwähnt bezieht sich das von Maaß (2004) zitierte Schema auf Modellierungsaufgaben im Rahmen höherer Bildungsbereiche. Die dort zum Einsatz kommenden Aufgaben sind wesentlich komplexerer Natur, selten gibt es nur eine richtige Lösung, manchmal kann diese auch nur annähernd bestimmt werden. Hierfür können dann nicht alle in Frage kommenden Größen in Betracht gezogen werden und das Problem muss entsprechend hinsichtlich seiner Komplexität reduziert werden. Man denke z.B. an das von Borromeo Ferri (2006) berichtete Beispiel, die Oberfläche eines Porsches berechnen zu wollen. Eine von Schülern bevorzugte Lösungsstrategie ist hierbei, das Auto als einen aus verschiedenen Quadern zusammengesetzten Körper zu betrachten und dessen Oberfläche zu berechnen. Abrundungen werden dafür der Einfachheit halber ignoriert. Dass die Oberfläche dann auch nur näherungsweise angegeben werden kann, ist eine logische Konsequenz.

Bei Grundschulkindern sind natürlich auch Schätz- oder Überschlagsaufgaben denkbar, um die Operationalisierung eines der Kompetenz M2 verwandten Teilkonstruktes zu realisieren.

Hingegen ist M3 „*to choose appropriate mathematical notations and to represent situations graphically*“, also die Verwendung einer formal korrekten Schreibweise, von der ersten Klasse an eine wünschenswerte Kompetenz im Repertoire der Schüler und wird dementsprechend auch von Anfang an gelehrt. Hierunter fällt noch nicht das Rechnen an sich, sondern die Art und Weise, wie Ansätze und Rechenwege dargestellt werden. Dazu gehört die Verwendung mathematischer Symbole, sowie das formal korrekte Aufschreiben von Rechnungen und deren Teilschritten (z.B. das saubere Untereinanderschreiben von Zahlen beim schriftlichen Addieren). Die Operationalisierung der Kompetenz ist möglich, indem das Wissen um korrekte Darstellungsweisen in Form von Multiple-Choice-Items abgefragt wird, aber auch durch die Bewertung von Aufzeichnungen der jeweiligen Probanden.

Verarbeiten

Unter „*Verarbeiten*“ fallen all diejenigen Kompetenzen, die sich auf die rein innermathematische Bearbeitung des Problems beziehen. Der Übersetzungsvorgang des Problems in die mathematische Sprache ist abgeschlossen, die Übersetzung der Lösung in die verbale Sprache noch nicht in Gang gesetzt.

Während bei komplexeren Aufgaben an dieser Stelle z.B. Gleichungen auf- und umgestellt werden, beschränkt sich der Verarbeitungsschritt bei Grundschülern noch allein auf das Rechnen. Die Unterkompetenzen V1 und V2 sind in ihrer Operationalisierung daher daran orientiert, wie Probanden beim Ausrechnen der vorgegebenen Problemstellungen vorgehen.

Die Kompetenz V1, „*to use mathematical knowledge to solve the problem*“, kann immer dann als gegeben betrachtet werden, wenn Personen versuchen, Probleme mithilfe mathematischer Strategien zu lösen. Dies dürfte bei Aufgaben, die im Umfeld von schulischem Mathematikunterricht vorgegeben werden, die Regel sein. Allerdings sind durchaus auch andere Strategien denkbar: Viele Probleme lassen sich durch Schätzen, Probieren oder Raten zumindest annähernd lösen. Verfügt ein Schüler nicht über die für ein Problem relevanten mathematischen Kompetenzen (V1), so kann es notwendig werden, auf eine dieser Ersatzstrategien auszuweichen.

V2 „*to use heuristic strategies such as division of the problem into part problems, establishing relations to similar or analog problems, rephrasing the problem, viewing the problem in a different form, varying the quantities or the available data etc.*” bezieht sich wiederum auf Strategien wie sie beim Lösen von komplexen Problemen angewandt werden. An sich kann die Kompetenz daher im Grundschulalter nicht erfasst werden. Allerdings kann man auch hier schon das Verwenden von einfachen Heuristiken beobachten, so zum Beispiel das „geschickte Rechnen“ oder andere sinnvolle Abkürzungen und Erleichterungen im Problemlöseprozess.

Grob lässt sich *Verarbeiten* im Primarbereich zwar mit Rechnen gleichsetzen, es gehören aber auch andere kognitive Prozesse dazu. Man denke z.B. an den Abruf von Wissen aus dem Langzeitgedächtnis, was zum Beispiel bei der Anwendung des Einmaleins oder diverser Rechenregeln zum Tragen kommt (Seitz & Schumann-Hengsteler, 2000; Uesaka, Manolo & Ichikawa, 2007).

Interpretieren

Wurde das Problem innermathematisch erfolgreich verarbeitet, so ist das Resultat dieses Verarbeitungsprozesses zurück in die verbale Sprache zu übersetzen. Gleichzeitig werden diese Ergebnisse auch im Kontext der realen Situation interpretiert.

Die Unterkompetenz I1 „*to interpret mathematical results in extramathematical contexts*" kann wieder als bester Repräsentant der Gesamtkategorie verstanden werden, da sie den Kern des Schrittes „*Interpretieren*" auf den Punkt bringt. In der schulischen Realität lässt sich diese Kompetenz leicht daran erkennen, ob Personen in der Lage sind, einen korrekten Antwortsatz zu formulieren, der die ursprüngliche Frage, wie sie aus der Aufgabenstellung herauszulesen ist, präzise beantwortet. Das kritische Moment ist hierbei, tatsächlich wieder auf sprachlichem Weg zu kommunizieren und nicht mehr mit Zeichen und Symbolen. Die Antwort auf die Frage

Ein Swimmingpool ist drei Meter breit, fünf Meter lang und zwei Meter tief.
Wie viel Wasser passt in den Pool?

muss lauten: „In den Pool passen 30.000 Liter Wasser" und nicht „30.000" oder gar „30 m^3", was zwar das Endergebnis der Rechnung, nicht jedoch eine Antwort auf die Frage darstellt. Die Operationalisierung der Kompetenz erfolgt daher über das Erkennen und Anwenden von korrekten Übersetzungsprozessen.

Die zweite Unterkompetenz I2 „*to generalize solutions that were developed for a special situation*" ist nicht bei allen Aufgaben praktikabel und muss daher als fakultativ betrachtet werden. Es geht hierbei darum, Lösungen zu verallgemeinern, also Antworten zu formulieren, die vom konkreten Problem losgelöst und allgemein ausgedrückt sind. Man betrachte hierzu das in der Vorstudie eingesetzte Item:

Familie Grün sammelt Regenwasser zum Gießen ihrer Blumen.
Die Abbildung rechts zeigt die 20-Liter-Tonne am Montagmorgen.
Wird das Regenwasser bis Sonntagabend ausreichen,
wenn Herr Grün jeden Abend einen Liter braucht?

Die Aufgabe wird zusammen mit einer Abbildung (siehe Anhang) präsentiert, die eine genau zur Hälfte gefüllte 20-Liter-Tonne anzeigt. I1 drückt die Fähigkeit aus, aus dem ausgerechneten Ergebnis 10–7·1l=3l einen Antwortsatz der Art „Das Regenwasser wird bis Ende der Woche ausreichen.“ zu formulieren. I2 beinhaltet die Befähigung zur darüber hinausgehenden, verallgemeinernden Reflektion über den Sachverhalt, zum Beispiel in Form einer Aussage wie: „Wenn man jeden Tag eine bestimmte Anzahl an Litern zum Gießen braucht, dann muss man pro Woche mindestens das siebenfache dieser Anzahl an Regenwasser auffangen.“ Dementsprechend kann auch die Operationalisierung über entsprechende verallgemeinernde Aussagen realisiert werden.

I3, die dritte Unterkompetenz, wird umschrieben als „*to view solutions to a problem by using appropriate mathematical language and/or to communicate about the solutions*“, also als die Fähigkeit, Ergebnisse in mathematischer Sprache zu kommunizieren. Dies scheint auf den ersten Blick etwas verwirrend, da doch *Interpretieren* an sich bedeutet, gerade nicht mehr in mathematischer Sprache zu kommunizieren. Gemeint ist hier aber etwas anderes: Es geht bei der Kommunikation von Lösungen darum, korrekte Begrifflichkeiten zu verwenden. Hiermit ist nicht die symbolische mathematische Sprache in Form von Ziffern, Rechenzeichen etc. gemeint, sondern das Verwenden von Ausdrücken, wie sie in der Mathematik üblich, zugleich aber auch in der Alltagssprache verständlich sind, zum Beispiel die Verwendung von Ausdrücken wie Länge, Höhe, Breite, Fläche, Strecke, etc.

Zu beachten ist allerdings, dass I3 keine Kompetenz auf der gleichen Ebene wie I1 und I2 darstellt. Im Gegenteil steht sie diesen eher diametral gegenüber und kann als die Art und Weise aufgefasst werden, in der die Kompetenzen I1 und I2 umgesetzt werden. Daher ist auch die Operationalisierung mit den beiden anderen Unterkompetenzen verbunden.

Validieren

Betrachtet man den Modellierungszyklus als normatives Modell, so endet er mit einem Validierungsschritt, der die Lösung des Problems auf seine Richtigkeit überprüft. Ist die Lösung nicht zufriedenstellend, wird der Prozess von neuem durchlaufen.

Maaß (2004) nennt für den Validierungsschritt folgende relevante Unterkompetenzen:

• *to critically check and reflect on found solutions; to review some parts of the model or again go through the modelling process if solutions do not fit the situation;*

• *to reflect on other ways of solving the problem or if solutions can be developed differently;*

• *to generally question the model*

Aus der Auflistung wird deutlich, dass die hier formulierten Unterkompetenzen vor allem bei komplexen Modellierungszyklen sinnvoll sind, bei denen es nicht „die" richtige Lösung gibt, sondern eine solche immer nur annäherungsweise bestimmt werden kann und daher viele Wege zielführend sind. Für die Aufgaben im Alltag von Grundschulkindern ist dies wenig realistisch. Daher soll für den letzten Schritt eine andere Systematik vorgeschlagen werden, die nur auf einer relativ groben Ebene zwischen zwei Arten von Ergebnisüberprüfungen unterscheidet.

E1 soll hierbei definiert werden als Indikator für jegliche Art der *innermathematischen* Validierung. Innermathematische Validierung heißt hierbei, dass die Lösung nicht (wie dies bei den drei Einzelkompetenzen nach Maaß der Fall ist) im Hinblick auf die ursprüngliche Fragestellung überprüft wird, sondern lediglich die in Schritt 3 „*Verarbeiten*" angelaufenen Prozesse wiederum mithilfe mathematischen Wissens überprüft werden. Dies kann z.B. geschehen über die Durchführung von Gegenproben, Überschlagsrechnungen oder das Zurückgreifen auf mathematische Regeln und Gesetze (z.B. das Wissen, dass bei der Addition zweier 4-stelliger Zahlen keine 6-stellige Zahl herauskommen kann; dass beim Teilen von geraden Zahlen durch 2 kein Rest stehen bleiben darf etc.).

Hingegen soll E2 verstanden werden als *außermathematische* Validierung, sprich als die Überprüfung der Lösung im Zusammenhang mit der Realsituation. Idealerweise hat sich der Problemlösende im ersten Schritt ein „Bild“ von der Situation gemacht. Man betrachte beispielhaft folgendes Problem:

> Ein Bauer fährt mit dem Traktor einmal um seinen quadratischen Grundbesitz. Der Acker hat eine Fläche von 25 Hektar. Nach 10 min. ist er wieder an seinem Startpunkt angekommen. Wenn der Bauer immer gleich schnell gefahren ist, wie hoch war dann die Geschwindigkeit des Traktors?

Die Lösung lautet 12 km/h. Hat sich ein Proband nun verrechnet – was beim Rechnen mit Einheiten häufig der Fall ist – so könnte er als Lösung 120 km/h erhalten haben. Durchläuft der Proband den Modellierungszyklus nun wie vorgeschrieben, so nimmt er die Lösung nicht einfach hin, sondern integriert diese in sein Vorstellungsbild von der Problemsituation. Die Vorstellung, dass ein Bauer mit dem Traktor auf einem Feldweg so schnell fährt wie ein PKW auf der Autobahn sollte entsprechende Zweifel wecken, ob die Lösung so korrekt sein kann. Idealerweise wird dann eine innermathematische Validierung in Gange gesetzt, um zu Überprüfen, ob ein Rechenfehler vorliegt. Ist dies nicht der Fall, so muss eventuell sogar die Korrektheit der Mathematisierung in Zweifel gezogen und entsprechend überprüft werden.

2.1.3 Operationalisierung als Beobachtungsitems

Zielsetzung der Voruntersuchung ist es zu identifizieren, welche theoretisch postulierten Modellierungskompetenzen tatsächlich beobachtet werden können. Hierfür wurde ein Protokollbogen erstellt, der die einzelnen Teilkompetenzen – operationalisiert durch die Unterkompetenzen – auflistet, zusammen mit einer Spalte zum Eintragen des Zeitpunkts, zu dem die Kompetenz beobachtet werden konnte. Um Modellierungskompetenzen jedoch beobachten zu können, müssen zunächst einmal Situationen geschaffen werden, in welchen diese potenziell überhaupt auftreten können. Hierfür wurden fünf Modellierungsaufgaben konstruiert, bei welchen alle Unterkompetenzen auftreten können, bzw. einen Beitrag zur Lösung der Aufgabe leisten. Alle fünf Aufgaben finden sich im Anhang, das Prinzip soll hier nun anhand von Aufgabe 2 verdeutlich werden: Abbildung 13 zeigt die Aufgabe, Tabelle 5 gibt wieder, an welcher Stelle die jeweiligen Unterkompetenzen umgesetzt werden können.

Spanischer Ackerbau

Im schönen Spanien liegen die Dörfer Desastro und Fortuna. Beide leben vom Getreideanbau. Da Fortuna eine wesentlich größere Fläche zum Bebauen hat, bringt die Ernte immer weit mehr ein, als im kleinen Desastro.
Pro Flächeneinheit Ackerland erntet man in Fortuna 3 Säcke Getreide. Die Bauern aus Desastro setzen dieses Jahr einen neuen Dünger ein und ernten pro Flächeneinheit 5 Säcke Getreide. Nach drei Jahren werden es sogar 6 Säcke sein.

Kann Desastro dieses Jahr nun endlich mehr Getreide ernten als Fortuna?

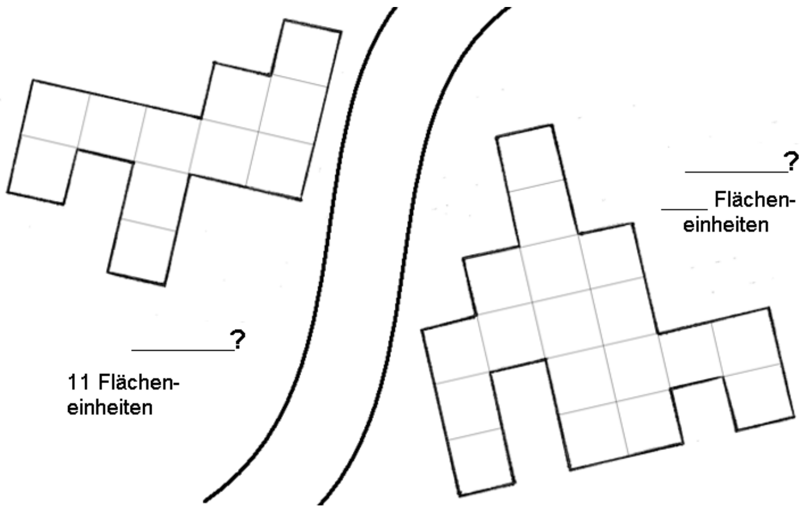


Abbildung 13: Beispielaufgabe

Tabelle 5: Umsetzung der Unterkompetenzen

	Strukturieren	
S1	Textverstehen / Problem in eigenen Worten wiedergeben	Die erfolgreiche Zuordnung des Ortsnamens zu der passenden Skizze zeigt, dass ein zentrales Element des Aufgabentextes verstanden wurde
S2	Größen benennen: Gegeben / Gesucht	Eine Äußerung dahingehend, was auszurechnen ist, kann als Auftreten von S2 aufgefasst werden.
S4a	Suche nach problemrelevanten Informationen	Zur Lösung der Aufgabe müssen die Felder in den Skizzen gezählt werden, es muss also lösungsrelevante Information gesucht werden.
S4b	Trennung zwischen relevanter – irrelevanter Information	Im Aufgabentext sind Informationen enthalten, die nicht relevant sind für die Lösung der Aufgabe. Diese müssen erkannt werden und dürfen nicht in den Lösungsansatz integriert werden.
	Mathematisieren	
M1	Mathematisieren von Größen und Beziehungen (Zahlwörter => arabische Ziffern) (Verbalisierungen => Operationen)	Das Übersetzen der achtzehn Kästchen in die Zahl 18, sowie das Erkennen der geforderten mathematischen Operation (Menge pro Flächeneinheit sollte das Schema „Multiplikation“ aktivieren).
M3	Angebrachte mathematische Notationen / Skizzen (= formal korrekte Schreibweise)	Das saubere Untereinanderschreiben von Zahlen bei der schriftlichen Multiplikation ist ein Indikator dieser Unterkompetenz.
	Verarbeiten	
V1	Mathematisches Wissen zur Problemlösung nutzen (z.B. korrekt schriftlich Rechnen)	Immer dann, wenn ein Proband eine Aufgabe durch Rechnen zu lösen versucht (statt z.B. durch Raten), hat er mathematisches Wissen eingesetzt.
V2	Heuristische Strategien („geschickt Rechnen“)	Heuristische Strategien sind schwerer zu beobachten, da die Kategorie nur vage formuliert ist. Sie kann beobachtet werden, wenn Probanden den Rechenweg verkürzen, bei dieser Aufgabe z.B. wenn 5 * 11 nicht schriftlich berechnet wird, sondern der Proband erkennt, dass es sich hierbei um (10*11)/2 handelt.

	Interpretieren	
I1	Interpretation im außermathematischen Kontext	Interpretationsbemühungen liegen dann vor, wenn das Ergebnis wiederum verstanden wird. Am Ende der Rechnung liegen zwei Zahlen vor: 54 und 55. Der Proband muss erkennen, dass die größere Zahl Desastro zuzuordnen ist.
I2	Verallgemeinerung der Lösung (fak.)	Die Lösung kann nun – muss aber nicht – darüber hinaus verallgemeinert werden, z.B. in der Form: Weniger Ackerland kann dadurch wettgemacht werden, dass man besseren Dünger verwendet. Solche allgemeinen Aussagen aus den Aufgaben abzuleiten, ist allerdings bei Grundschülern wohl eher selten zu erwarten.
I3	Kommunikation der Ergebnisse in math. Sprache	Die Kommunikation in mathematischer Sprache kann als gegeben gelten, wenn in der Interpretation des Probanden die korrekten Ausdrücke Fläche, Menge, Anzahl etc. vorkommen.
	Validieren	
E1	Innermathematische Validerung	Innermathemtisch lassen sich Ergebnisse validieren, wenn ein Proband nachrechnet: Kann 11 mal 5 mehr sein als 3 mal 18?
E2	Außermathematische Validierung	Außermathematische Validierung hingegen bezieht sich auf den inhaltlichen Sinn der Fragestellung, wenn der Proband z.B. überlegt: Kann es sein, dass man mit besserem Dünger die geringere Fläche wettmachen kann?

Auf Grundlage der fünf Aufgaben kann analog hierzu jeweils eingeschätzt werden, welche Unterkompetenzen ein Proband beherrscht oder nicht.

Zur Validierung der Leistungsitems für die Hauptuntersuchung wird ebenfalls auf Beobachtungsdaten zurückgegriffen, nämlich auf eine Einschätzung der Modellierungskompetenz einzelner Schüler durch die Lehrkraft. Die Operationalisierung erfolgt wiederum eng orientiert an der Formulierung der Unterkompetenzen, indem die Lehrkraft jeweils gebeten wird einzuschätzen, wie oft ein Schüler bei der Umsetzung einer Kompetenz Schwierigkeiten hat.

2.1.4 Operationalisierung als Leistungsitems

Möchte man den Modellierungszyklus auf seine Gültigkeit als normatives Modell überprüfen, so reicht es nicht, zu beobachten, ob eine Kompetenz gezeigt wird oder nicht. Vielmehr muss tatsächlich die Leistung gemessen werden. Ein *Bemühen* um Textverständnis allein kann nicht als Prädiktor für die Qualität der Lösung einer Aufgabe fungieren, wohl aber das nachweislich *erfolgte* Textverstehen.

Zu diesem Zweck, sowie schon aus der Fragestellung zur Entwicklung eines Testverfahrens heraus, ist es daher vonnöten, echte Leistungsitems zu entwickeln.

Die Items zur Erfassung der jeweiligen Teilkompetenzen wurden als Multiple-Choice-Items realisiert. Multiple-Choice-Aufgaben sind zur Erfassung von Leistungsdaten weit verbreitet und zeichnen sich im Vergleich zu offenen Antwortformaten (wie sie bei Textaufgaben allgemein üblich sind) insbesondere durch die Ökonomie sowie die Objektivität bei der Auswertung aus (Lissmann & Jäger, 2008).

Für die Konstruktion von Multiple-Choice-Items gibt es eine ganze Reihe von Leitfäden (für einen Überblick vgl. Haladyna, Downing & Rodriguez, 2002), deren praktische Bedeutsamkeit sich jedoch auch nach der jeweils anvisierten Zielsetzung zu bemessen hat. So ist es durchaus denkbar, dass allgemein bekannte „No-Gos“ der Itemformulierung bewusst in Kauf genommen werden, um eine konkrete Fragestellung zu untersuchen.

Dies wurde in der vorliegenden Arbeit z.B. im Bereich der Strukturierungsaufhaben umgesetzt. Es wird in der Regel nämlich davon abgeraten so genannte „tricky items“ zu verwenden (Haladyna, 2004), Items also, die den Anwender bewusst in die Irre führen bzw. ihn zwingen, den Itemtext genau durchzulesen, da eine voreilige Beantwortung zur falschen Antwortalternative führt. Solche Items wurden hier nun aber gezielt konstruiert, um Textverstehen zu erfassen. Man kann davon ausgehen, dass ein Item der Form *„Lars ist 5 Jahre alt, seine Schwester Lisa ist doppelt so alt. Lars kann nur halb so schnell laufen wie Lisa. Lars braucht morgens 5 Minuten um in den Kindergarten zu laufen. Wie lange braucht Lisa, um in die Schule zu laufen?“* nur dann gelöst werden kann, wenn der Proband erkennt, dass die Entfernung zur Schule gar nicht bekannt ist.

Schwieriger als die Formulierung des Itemstammes und der dazugehörigen richtigen Lösung ist die Konstruktion adäquater Distraktoren. Lissmann & Jäger (2008) empfehlen, eine mittlere Attraktivität der Distraktoren anzustreben, da zu attraktive Antworten zu sehr mit der richtigen Lösung konkurrieren, andererseits die ganz offensichtlich falschen Distraktoren das Antwortangebot (und damit die Schwierigkeit der Aufgabe) reduzieren. Anders gesagt: Die Erhöhung der Attraktivität eines Distraktors erhöht die Wahrscheinlichkeit der Wahl auch durch leistungsstarke Personen, umgekehrt führt eine verringerte Attraktivität der Distraktoren dazu, dass auch leistungsschwache Personen recht einfach die richtige Lösung identifizieren können.

Diese Tatsache muss dem Testkonstrukteur jedoch nicht unbedingt zum Nachteil gereichen, im Gegenteil können die Distraktoren gezielt attraktiver bzw. weniger attraktiv gestaltet werden, um so die Schwierigkeit der Aufgabe zu manipulieren. Oder – wenn man einen Schritt weiter gehen möchte – es können Distraktoren eingeführt werden, deren Attraktivität so hoch ist, dass sogar von einer teilweise richtigen Lösung gesprochen werden kann, für die man dann im Rahmen eines sog. Partial-Credit-Models entsprechend Teilpunkte vergeben kann (Rost, 2004).

Wie lässt sich nun aber im konkreten Fall die Attraktivität eines Distraktors erhöhen? Haladyna (Haladyna, 2004) empfiehlt, hierzu typische Schülerfehler einzusetzen. In der Praxis ist es nun aber alles andere als eindeutig, was als „typischer“ Fehler gelten kann und was nicht. Man könnte an dieser Stelle zwar auf Expertenratings zurückgreifen, letztlich kann jedoch nur die Empirie zeigen, ob ein Distraktor sich bewährt oder nicht.

Als Argument gehen Multiple-Choice-Tests lässt sich lediglich die – im Vergleich zu offenen Antworten – erhöhte Ratewahrscheinlichkeit anführen. Lissmann & Jäger (2008) schlagen daher die folgenden Maßnahmen vor:

- Die Testpersonen sind so anzuleiten, dass sie auf jeden Fall eine Lösungsmöglichkeit ankreuzen sollen, auch wenn sie die richtige Antwort nicht wissen. Somit ergibt sich für Personen, die lediglich raten, kein Vorteil gegenüber denen, die ihre Wissenslücken nicht zu verbergen suchen.
- Eine Erhöhung der Distraktorenanzahl führt zu einer verringerten Wahrscheinlichkeit, per Zufall die richtige Antwortalternative zu treffen. Rodriguez (2005) plädiert jedoch für ein Optimum an drei Antwortalternativen, da eine Erhöhung der Alternativen i.d.R. zu immer unplausibleren Distraktoren führt.
- IRT-Modelle sind in der Lage, die Ratewahrscheinlichkeit statistisch zu berücksichtigen. Dies geschieht in sog. 3PL (three-parameter-logistic) Modellen, bei welchen neben der Itemschwierigkeit auch Trennschärfe und Rate-

wahrscheinlichkeit berücksichtigt werden. Diese Modelle werden jedoch aufgrund ihrer Komplexität in der Praxis selten eingesetzt (Rost, 2004).

- Auch über die Testlänge lässt sich das Problem der Ratewahrscheinlichkeit angehen, indem nämlich eine erhöhte Anzahl von Testitems zwangsläufig zu einer erhöhten Reliabilität führt, wodurch trotz einer gewissen Ratewahrscheinlichkeit zuverlässige Aussagen aus den Testdaten abgeleitet werden können.

Die als Grundlage für die Hauptuntersuchung entwickelten Items wurden aufbauend auf den Ergebnissen der Vorstudie erstellt (siehe Abschnitt 3.4), so dass für jede Teilkompetenz neun Items vorlagen.

3. Studie 1: Qualitative Voruntersuchung

Die erste Untersuchung dient, wie in Kapitel 2 ausgeführt, der Überprüfung des Modellierungszyklus als deskriptives Modell. Im folgenden Kapitel wird zunächst die Untersuchungsmethode dargestellt, dann werden die Ergebnisse der Studie präsentiert und im Anschluss diskutiert. Diese Ergebnisse und die sich daraus ergebenden Implikationen bilden gleichsam die Grundlage für die in Kapitel 4 beschriebene Hauptuntersuchung.

Die Überprüfung als deskriptives Modell beinhaltet die Frage, welche der theoretisch postulierten Teilkompetenzen sich tatsächlich beobachten lassen. Man hat es daher mit einer unbestimmten Existenzhypothese zu tun. Eine solche kann bestätigt werden, sobald man einen Fall identifizieren kann, für den die Hypothese zutrifft. Widerlegen lässt sie sich andererseits nicht, da es unmöglich ist, alle denkbaren Fälle zu überprüfen (Hussy & Jain, 2002).

3.1 Methode

Wie im theoretischen Teil bereits ersichtlich wurde, gibt es verschiedene Methoden und Ansätze, um Modellieren zu beobachten und zu bewerten. Diese sollen nun erneut bezüglich Ihrer Eignung für die Zielsetzung von Studie 1 betrachtet werden, woraus sich der Bedarf an einer bisher in dieser Form zur Untersuchung von Modellierungsprozessen noch nicht verwendeten Methode ableiten lässt. Im Gegensatz zu Kapitel 1.6.2, wo es allgemein um die in der Literatur vorzufindenden Operationalisierungswege ging, liegt der Fokus hier nun auf konkreten Umsetzungen, um die für die Vorstudie relevanten Fragestellungen zu beantworten. Die oben aufgeführten Verfahren werden hierfür allerdings erneut kurz aufgegriffen und in Bezug auf ihre Angemessenheit für die vorliegenden Fragestellungen diskutiert.

3.1.1 Überblick über mögliche Untersuchungsmethoden

Der Modellierungszyklus ist ein in mehreren Teilschritten von statten gehender Prozessablauf. Interessiert man sich dafür, welche Teilschritte hierbei auftreten und wann, so ist eine Untersuchungsmethode indiziert, bei der nicht nur anhand des Endprodukts auf Einzelprozesse geschlossen wird, sondern bei der diese Prozesse direkt untersucht werden.

Tabelle 6 gibt nochmals einen Überblick über die verschiedenen Forschungsansätze, die sich zur Untersuchung von Modellierungsabläufen und Modellierungskompetenzen in der Literatur finden.

Tabelle 6: Untersuchungsansätze für Modellierungsabläufe

Methode	**Exemplarische Vertreter**
Projektarbeit – Beurteilung des Modellierungsvorgangs durch Rekonstruktion der Ergebnisse	Blomhoy & Jensen, 2003; Caron & Bélair, 2007
Projektarbeit – Beurteilung des Modellierungsvorgangs durch Lehrerurteile anhand von Kriterienlisten	Berry & Le Masurier, 1984; Dunne & Galbraith, 2003; Hall, 1984
Videographie von in Gruppen bearbeiteten Modellierungsproblem	Borromeo Ferri, 2006; Busse & Kaiser, 2003
Multiple – Choice – Tests	Haines & Crouch, 2005; Houston & Neill, 2003
Concept Maps	Hasemann & Mansfield, 1995; Maaß, 2006

Bei der Verwendung von Projekten, die von Lehrern anhand von Kriterienlisten beurteilt werden oder bei denen aus dem Ergebnis der Ablauf des Modellierungsprozesses rekonstruiert wird, kann kaum von einer reliablen Erfassung des Geschehens beim Modellieren ausgegangen werden. Auch klassische Multiple-Choice-Items sind in diesem Falle wenig zielführend, da auch hier lediglich das Ergebnis beobachtet werden kann, nicht aber die Art und Weise wie ein Modellierer zu diesem gelangt ist.

Mehr Aufschluss gibt das tatsächliche Beobachten von Modellierungsprozessen, wie es z.B. bei Borromeo Ferri (2006) oder bei Busse & Kaiser (2003) umgesetzt wurde. Eine Gruppe von mindestens zwei Schülern bearbeitet zusammen eine Aufgabe. Aus der Interaktion der beiden Personen soll letztlich auf die Vorgänge beim Modellieren geschlossen werden, da man davon ausgeht, dass beide Modellierer miteinander kommunizieren müssen, um gemeinsam zu einer Lösung zu finden und sie auf diese Art und Weise ihre Gedanken verbalisieren. Nachteil ist hierbei jedoch, dass nicht wirklich der *individuelle* Problemlöseprozess untersucht wird und man nicht sicher sein kann, dass bei einer gemeinsamen Lösung bei den einzelnen Personen die gleichen kognitiven Prozesse und in gleicher Reihenfolge ablaufen, wie wenn ein Modellierer allein vorgeht. Es verbleiben bei dieser Vorgehensweise also erhebliche Zweifel an der Validität der Ergebnisse.

Maaß (2006) verwendet einen Mix von Methoden zur Erfassung des Modellierungsvorganges, unter anderem kommen so genannte Concept Maps zum Einsatz. Bei dieser auf Novak (1990) zurückgehenden Methode erhalten Probanden zu einem bestimmten Thema eine Reihe von Karten mit Begriffen oder Abbildungen, zusammen mit der Instruktion, diese so auf einem Blatt Papier zu verteilen, dass verwandte Konzepte nahe beieinander liegen und nicht oder kaum verwandte Konzepte entsprechend räumlich voneinander separiert werden. Gleichzeitig sollen Beziehungen zwischen den Konzepten mithilfe von Pfeilen dargestellt werden. Theoretischer Hintergrund ist die Vermutung, dass Versuchspartner die Konzepte angelehnt an die Art und Weise darstellen, wie diese auch in ihren eigenen Wissensstrukturen repräsentiert sind. Es geht hierbei allerdings weniger um die genaue Repräsentation der Wissensstrukturen, sondern darum, einen Einblick in die Denkprozesse von Probanden zu erhalten und ihnen das Erlernen von neuen Sachverhalten zu erleichtern. Hasemann & Mansfield (1995) machen daher in ihrem Überblicksartikel klar, dass Concept Mapping in erster Linie eine Instruktionsmethode ist und dass es fraglich bleibt, ob es sich hierbei auch um eine valide Forschungsmethode handelt.

Keine der genannten Methoden scheint daher optimal zu sein, um die Zielsetzung von Studie 1 sinnvoll angehen zu können. Am nächsten kommt dem Anspruch wohl noch die Methode der Videoanalysen, die jedoch so abgewandelt werden müsste, dass nur individuelle Lösungsstrategien erfasst werden. Geht man davon aus, den gleichen Versuchsaufbau mit nur einer Person zu realisieren, die dann statt mit einem Mitschüler mit sich selbst oder einem neutralen Versuchsleiter interagiert, so entspricht dies nichts anderem als dem lauten Denken.

3.1.2 Die Methode lauten Denkens

Wie auch bei Videoanalysen liegt der Vorteil der Methode lauten Denkens darin, dass der *Prozess* der Informationsverarbeitung rekonstruiert werden kann – nicht nur deren Ergebnis. Auf diese Art und Weise ist es möglich zu untersuchen, welche der von Maaß (2006) herausgearbeiteten theoretischen Komponenten tatsächlich Teil des in der Realität ablaufenden Modellierungszyklus sind.

Die Methode lauten Denkens ist an sich ein Oberbegriff für verschiedene Untersuchungsmethoden, deren gemeinsamer Nenner in einer Instruktion besteht, welche die jeweiligen Probanden in irgendeiner Form dazu auffordert laut auszusprechen, was ihnen durch den Kopf geht.

Von Bühler im Rahmen der Würzburger Schule bereits 1907 eingesetzt, um kognitive Prozesse beim Problemlösen zu erfassen, wurde die Methode in Zeiten der Vorherrschaft des Behaviorismus wegen ihrer Ähnlichkeit mit der Introspektion heftig kritisiert (Weidle & Wagner, 1982).

Tatsächlich muss durchaus kritisch die Frage gestellt werden, inwieweit eine solche Methode die üblichen Gütekriterien erfüllen kann. Hierbei ist insbesondere die Validität der Daten von Interesse: Ist das, was Versuchspersonen berichten, wirklich ein getreues Abbild des jeweiligen kognitiven Prozesses?

Ericsson & Simon (1980; 1993) stellten zur Beantwortung dieser Frage eine Theorie des lauten Denkens auf, deren Ziel es ist, derartige Verzerrungen vorhersehbar zu machen. Die Autoren unterscheiden zwischen drei verschiedenen Arten lauten Denkens, abhängig von der jeweils zugrunde liegenden Instruktion bzw. Art der zu bearbeitenden (primären) Aufgabe.

Die Theorie basiert auf der Annahme von Kurz- (KZG) und Langzeitgedächtnis (LZG). Denken ist ein Prozess, der in einzelnen Unterschritten („Steps“ S) abläuft, deren Ergebnisse jeweils im KZG abgelegt werden. Auf der ersten Ebene lauten Denkens (Abbildung 14) werden diese Ergebnisse direkt verbalisiert. Die Bearbeitung der primären Aufgabe wird hierdurch nicht beeinflusst.

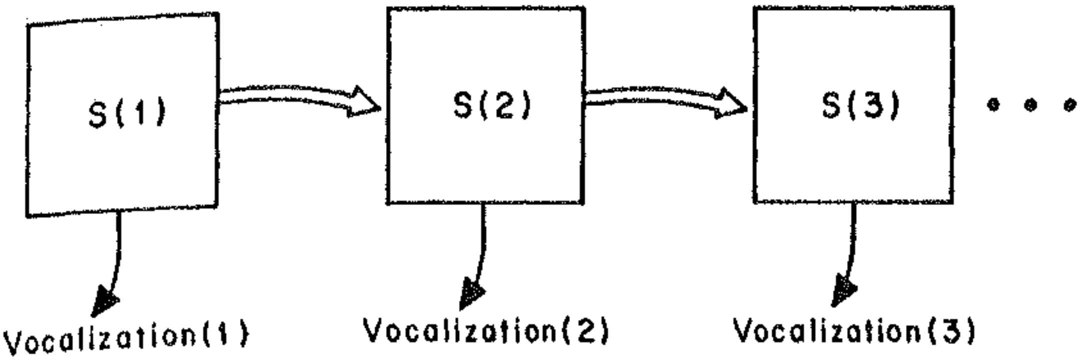


Abbildung 14: Lautes Denken auf Ebene 1 (Ericsson & Simon, 1993, S. 17)

Ericsson und Simon zufolge funktioniert diese (optimale) Version lauten Denkens jedoch nur, wenn rein verbal codierte Gedächtnisinhalte verbalisiert werden. Sobald jedoch in das Denken einer Person z.B. visuelle Informationen miteinbezogen werden – man denke in diesem Zusammenhang an die von Borromeo Ferri (2006) untersuchten mathematischen Denkstile: manche Personen denken mehr „in Bildern" als andere – so müssen diese Informationen vom Probanden zunächst in ein sprachliches Format übersetzt werden, bevor sie verbalisiert werden können.

Dieser Vorgang erfordert Zeit, weshalb auf dieser zweiten Ebene lauten Denkens, die Primäraufgabe dahingehend beeinflusst ist, dass für deren Bearbeitung ein höherer Zeitaufwand vonnöten ist.

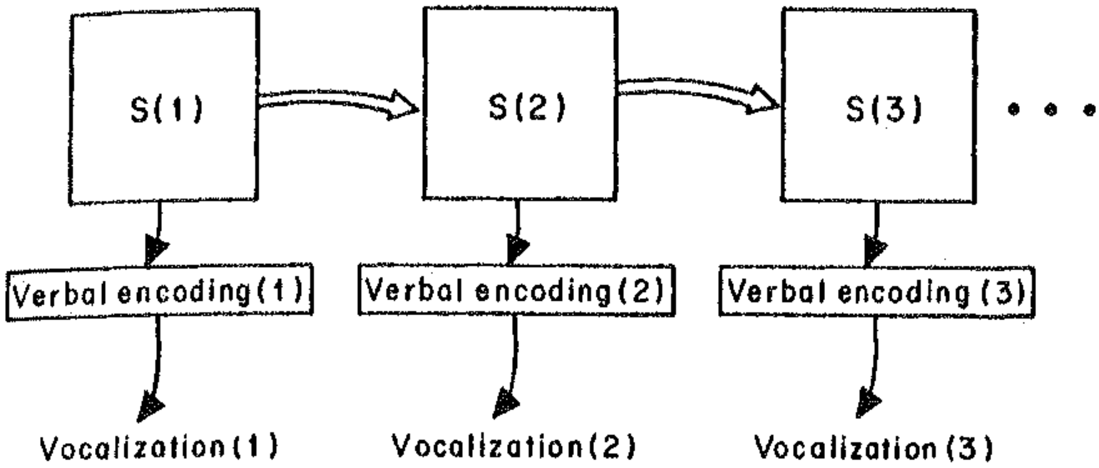


Abbildung 15: Lautes Denken auf Ebene 2 (Ericsson & Simon, 1993, S. 17)

Abbildung 15 verdeutlicht dies: Zwar erfolgt wieder nach jedem Schritt eine Verbalisierung der im KZG zwischengespeicherten Information, zwischengeschaltet ist jedoch jeweils ein Enkodierungsschritt. Aus der Grafik wird auch deutlich, dass dadurch die natürliche Abfolge der Schritte (S(1) bis S(3)) nicht tangiert wird. Kann man davon ausgehen, dass bei einer Untersuchung lautes Denken der Ebene 2 abläuft, so können aus den Aussagen der Versuchspartner Schlussfolgerungen über den Ablauf und die Reihenfolge kognitiver Prozesse gezogen werden, nicht jedoch über deren zeitliche Dauer.

Noch stärkere Auswirkungen auf die Primäraufgabe hat lautes Denken, wenn es auf der dritten Ebene abläuft (siehe Abbildung 16).

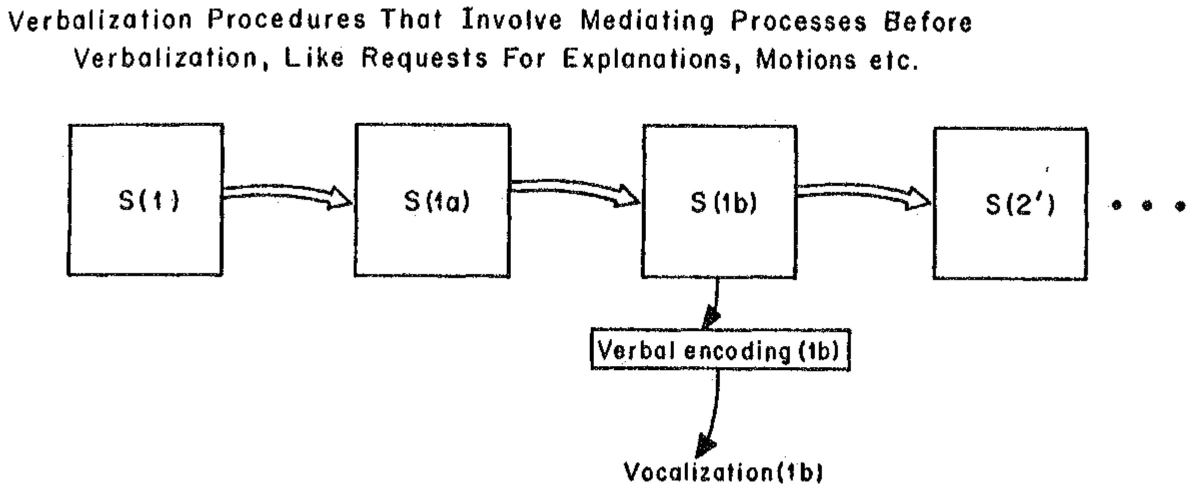


Abbildung 16: Lautes Denken auf Ebene 3 (Ericsson & Simon, 1993, S. 17)

Wenn die Instruktion Tätigkeiten wie die Beschreibung motorischer Aktivitäten beinhaltet oder wenn Versuchspartner gebeten werden, Erklärungen für ihr Verhalten abzugeben, so können zwei zusätzliche Schritte auftreten, die der Verbalisierung – und einer eventuellen Enkodierung – vorangestellt sind.

Zum einen ist dies das Filtern von verfügbaren Informationen dahingehend, ob diese berichtet werden oder nicht. Zum anderen können Interferenzen auftreten, wenn von den Probanden Informationen gefordert werden, die sich eben nicht im KZG befinden, sondern die erst aus dem LZG abgerufen werden müssen. Fragt man beispielsweise nach Erklärungen oder Rechtfertigungen, so kann man auf Informationen zugreifen, die dem Probanden bei Methoden der Ebene 1 gar nicht bewusst gewesen wären. Dieser Mehrwert an Information muss jedoch mit gravierenderen Auswirkungen auf die Primäraufgabe bezahlt werden. Neben einer noch längeren Bearbeitungsdauer kann sich die *Struktur* der Aufgabenbearbeitung ändern, zum Beispiel die Reihenfolge von kognitiven Prozessen. Wenn aber der zu untersuchende kognitive Prozess verändert wird, muss man die Frage stellen, ob die Methode dann überhaupt als geeignet betrachtet werden kann, um diesen Prozess zu untersuchen. Auf den hieraus resultierenden Trade-off zwischen Informationsvielfalt und Validität wird noch näher eingegangen.

Unabhängig von der Ebene der Instruktion sind weitere Bedenken und Probleme gegenüber der Methode lauten Denkens zu benennen, die man sowohl bei der Konzeption von Studien als auch bei der Interpretation der hiermit gewonnenen Daten berücksichtigen muss.

Weidle & Wagner (1982) nennen hierbei folgende Punkte: Als erstes gilt es, das Problem der Kapazität und Auswahl zu berücksichtigen. Obwohl die Methode lautes „Denken“ heißt, umfasst sie eigentlich alle möglichen kognitiven Prozesse, darunter zum Beispiel auch Wahrnehmungen und Empfindungen. Zwar kann man Personen instruieren, nur diejenigen Informationen auszusprechen, die für die Lösung des jeweiligen Primärproblems Relevanz besitzen oder besitzen könnten, allerdings setzt man hier einen Filterprozess in Gang, der sich – wie aus Ebene 3 nach dem Modell von Ericsson & Simon (1980) Modell ersichtlich – eventuell nachteilig auf die Lösung des Problems auswirkt.

Das zweite Problem – und gleichzeitig eine der größten Herausforderungen, der sich die Methode lauten Denkens zu stellen hat – ist das der Bewusstheit. Nicht alle Kognitionen laufen allerdings bewusst ab. Nur diese können aber über die Methode lauten Denkens erfasst werden. Vor allem automatisierte Vorgänge im Rahmen prozeduralen Wissens sind dem Bewusstsein gar nicht zugänglich. Hinzu kommt, so Weidle & Wagner (1982, S.85), dass „es große interindividuelle Unterschiede in der Fähigkeit gibt, sich der eigenen kognitiven Prozesse bewusst zu werden.“

Dirttens gilt es zu bedenken, dass eine Studie, in der die Methode lauten Denkens realisiert wird, nicht allein als eine Repräsentation kognitiver Prozesse betrachtet werden kann, sondern dass sich Versuchsleiter und Proband in einer sozialen Situation gegenüber stehen (vgl. hierzu auch Sasaki, 2003). Weidle & Wagner (1982) empfehlen daher den Versuchsleitern, ihr Verhalten an den Grundsätzen der nicht-direktiven Gesprächstherapie zu orientieren und sich dem Versuchspartner gegenüber möglichst einfühlsam und nichtwertend zu verhalten. Im Zusammenhang mit der Arbeit mit Mathematikaufgaben ist dies besonders dann relevant, wenn Probanden fehlerhaft vorgehen.

Viertens kann (vgl. Ebene 2 im Modell von Ericsson & Simon) es Schwierigkeiten dabei geben, nicht-sprachliche Gedächtnisinhalte verbal auszudrücken. Insbesondere kommen hier auch wieder interindividuelle Unterschiede zum Tragen, da es immer Probanden geben wird, die sich besser, und solche die sich schlechter ausdrücken können (Olson, Duffy & Mack, 1984).

Fünftens ist das Problem der sozialen Erwünschtheit zu nennen. Probanden werden eventuell nicht alle ihre Gedanken aussprechen, oder - noch gravierender – falsche Gedanken berichten, um dem Versuchsleiter zu gefallen.

Weitere Bedenken zur Methode äußern Russo, Johnson & Stephens (1989), die sich intensiv mit der Validität von verbalen Daten – wie sie auch beim lauten Denken anfallen – beschäftigen. Unterschieden werden zwei Gründe für mangelnde Validität: Zum einen die Reaktivität, zum anderen die Nonveridikalität. Während Reaktivität die Veränderung des Problemlöseprozess durch das laute Denken betrifft (man denke wieder an die Ebenen nach Ericsson & Simon), bezieht sich Nonveridikalität auf das Nichtberichten von Gedanken einerseits und auf das Berichten von gar nicht wirklich aufgetauchten Gedanken andererseits.

Für die Reaktiviät nennen Russo et al. (1989) potenzielle Gründe: Das laute Denken kann ein Mehr an kognitiven Ressourcen erfordern, weshalb zwei mögliche Konsequenzen auftreten können, die für den Versuchsleiter zwar beide nicht wünschenswert sind, von denen die Reaktivität in den meisten Fällen jedoch das kleinere Übel darstellen dürfte: „They [die Probanden] can withdraw resources from the primary process and devote them to verbalization, risking reactivity, or they can temporarily suspend verbalization, a violation of veridicality. (S.764)."

Im Gegensatz zu den bisher genannten Effekten, können jedoch auch weitere Faktoren die Lösung der primären Aufgabe positiv beeinflussen, was natürlich dennoch eine Einschränkung der Validität darstellt:

- Die Motivation der Probanden, das Problem möglichst gut und schnell zu lösen könnte im Gegensatz zu alltäglichen Problemlösesituationen gesteigert sein, da die Betreffenden dem Versuchsleiter gegenüber ein besonders gutes Bild von sich selber darbringen wollen.
- Das auditorische Feedback – also das Hören der eigenen Verbalisierungen – kann als hilfreiche Stütze im Problemlöseprozess wahrgenommen werden, kann aber unter Umständen diesen auch beeinträchtigen.
- Außerdem kann das durch die Instruktion angeleitete bewusste Reflektieren über sonst automatisch ablaufende Prozesse zum Erinnern an alte und zum Entdecken neuer Strategien führen, wodurch Probanden ihr Problemlöseverhalten verbessern können.

Ist rein lautes Denken als Methode für die hier vorliegende Fragestellung geeignet? Hierbei sind zwei Gesichtspunkte zu beachten: Einerseits betrifft dies die Fragestellung an sich (Kann lautes Denken die beim mathematischen Modellieren ablaufenden Problemlöseprozesse abbilden?), andererseits aber auch die anvisierte Zielgruppe, sprich: Kann lautes Denken erfolgreich bei 10-Jährigen Kindern eingesetzt werden?

Lautes Denken wurde schon von Beginn an bevorzugt eingesetzt, um Problemlöseprozesse abzubilden. Auch um den Lösungsprozessen bei mathematischen Textaufgaben auf den Grund zu gehen, wurde lautes Denken bereits eingesetzt. Flaherty (1975) konnte zeigen, dass lautes Denken keinen Einfluss auf den grundsätzlichen Problemlöseprozess hat, dass jedoch signifikant häufiger Rechenfehler gemacht werden, als wenn Probleme ohne gleichzeitiges lautes Denken gelöst werden.

Allerdings (vgl. das oben aufgeführte Problem der Bewusstheit) ergeben sich gravierende Probleme, wenn Probanden über Schritte im Problemlöseprozess berichten sollen, die unbewusst ablaufen, da sie aufgrund von Übung im Sinne prozeduralen Wissens automatisiert worden sind. Weidle & Wagner (1982) vermerken hierzu: „Zur Methode des lauten Denkens meint Duncker, dass die besten Protokolle nur eine unvollständige Darstellung dessen sein können, was sich tatsächlich in den Köpfen der Vpn abspielt. Manchmal sind den Vpn bestimmte Lösungsphasen so selbstverständlich, dass sie gar nicht explizit geäußert werden, oder die Vp bemerkt nicht, dass sie bereits einen Schritt weiter im Lösungsprozess fortgeschritten ist (S. 90)."

Gerade beim Lösen von Modellierungsaufgaben ist jedoch damit zu rechnen, dass viele Prozesse (z.B. S4a Suche nach problemrelevanten Informationen) sich dem Bewusstsein des Modellierers entziehen und daraus resultierend auch beim lauten Denken nicht verbalisiert werden.

Auch die Frage nach der Zielgruppe stimmt bedenklich: Zwar finden sich in der Literatur einige Studien, in denen lautes Denken bei Kindern eingesetzt wurde (z.B. Camp, Blom, Hebert & Doorninck, 1977), doch findet man selten eine Diskussion über die Angemessenheit der Methode. Eine Ausnahme bildet Grüßing (2002), die die Methode bei Grundschulkindern einsetzte, um Strategien bei der Bewältigung von räumlich-geometrischen Anforderungen zu untersuchen. Offenbar sind gerade bei Kindern die interindividuellen Unterschiede bezüglich der Fähigkeit, „sich der eigenen kognitiven Prozesse bewusst zu werden und sie in Sprache umzusetzen [S.39]" ziemlich groß.

Neben der kognitiven Psychologie wird lautes Denken auch im Zuge des sog. Usability Testing von Softwareentwicklern eingesetzt um festzustellen, wie gut Personen mit bestimmten Programmen zurechtkommen. Ein Vergleich verschiedener

Methoden (Fragebogen, lautes Denken und Interview) zeigte, dass mit lautem Denken in der Zielgruppe der 8-14 Jährigen die meisten Informationen erfasst werden konnten (Donker & Markopoulos, 2002). Dennoch ist – zumindest bei einigen Kindern – aufgrund der oben genannten zu erwartenden interindividuellen Unterschiede in der Fähigkeit zum lauten Denken - mit erheblichen Schwierigkeiten beim Einsatz der Methode zu rechnen.

3.1.3 Adaption der Methode lauten Denkens

Wie sind nun die oben genannten zu erwartenden Schwierigkeiten, nämlich das zu erwartende Kommunikationsproblem sowie das Problem der Bewusstheit, am Besten in den Griff zu bekommen? Russo et al. (1989) vertreten die Meinung, Verzerrungen bei der Bearbeitung der Primäraufgabe seien absolut inakzeptabel, wohingegen es weniger problematisch sei, wenn Probanden nicht alle der von ihnen erlebten Kognitionen berichteten. Bedenkt man jedoch die Zielsetzung der Studie, die darin liegt, möglichst viele der eventuell ablaufenden Phasen – seien diese nun bewusst oder unbewusst – aufzudecken, so wäre eher eine umgekehrte Betrachtungsweise indiziert.

Es ist von zentraler Bedeutung, möglichst alle erlebten Kognitionen zu identifizieren, selbst wenn dadurch das Vorgehen bei der Lösung der Primäraufgabe beeinflusst wird. Eine solche Verzerrung ist ohnehin zu erwarten, da eine Untersuchung der Modellierungskompetenzen nicht im Rahmen der Ebene-1-Modelle lauten Denkens zu realisieren ist. Schließlich ist bei mathematischem Wissen damit zu rechnen (siehe Erläuterungen zum Kompetenzbegriff), dass ein wesentlicher Anteil dieses Wissens prozeduraler Natur ist und daher automatisch – und teilweise auch unbewusst – abläuft (Markman, 1999). Um dieses Wissen trotzdem erfassen zu können, muss es also zunächst bewusst gemacht werden – dies erfordert aber wiederum Reflexion und Abrufen von Informationen aus dem LZG. Trotz der oben skizzierten Beeinträchtigungen und Nachteile ist also ein Modell der Ebene 3 nach Ericsson & Simon (1980) für die vorliegende Fragestellung unumgänglich.

Um eine höhere Authentizität herzustellen, als dies gewöhnlich bei lautem Denken üblich ist, wurden die Probanden in Studie 1 instruiert, sich vorzustellen, die Versuchsleiter seien völlig unbewandert im Lösen von mathematischen Problemen und wurden dazu aufgefordert, ihnen daher Schritt für Schritt zu erklären, wie sie bei der Problemlösung vorzugehen hätten. Dies erleichtert den Grundschulkindern den Umgang mit der zunächst seltsam anmutenden Methode und reduziert die Wahrscheinlichkeit von Kommunikationsproblemen. Um zu vermeiden, dass unbewusste Zwischenschritte beim Problemlösen nicht verbalisiert würden, wurden zwei Maßnahmen realisiert: Erstens verblieb – zumindest scheinbar – die formale Rolle des Problemlösers beim Versuchsleiter. Um im Problem einen Schritt weiter zu kommen, war der Proband daher gezwungen, diesen Schritt zu verbalisieren, damit der Versuchsleiter ihn ausführen konnte. Zweitens wurde das Überspringen von Zwischenschritten verhindert, indem der Versuchsleiter bei nicht ausreichend konkreten Anweisungen Nachfragen stellte. (Z.B. „Jetzt rechnest Du das aus." – „Und wie soll ich das machen?").

Trotz aller Praktikabilität der Methode steht außer Frage, dass diese nicht frei von Problemen ist. Daher gilt es, die mit der Studie gewonnen Daten mit Vorsicht zu interpretieren und besonders die beiden vorherrschenden Einschränkungen im Rahmen der Ebene-3-Modelle lauten Denkens zu berücksichtigen: Zum einen dürfte die Bearbeitungszeit der Aufgabe wesentlich länger sein, als dies ohne lautes Denken der Fall wäre. Zum anderen kann nicht mit Sicherheit davon ausgegangen werde, dass die Struktur des Modellierungskreislaufs, sprich: die Reihenfolge der auftauchenden einzelnen Modellierungskompetenzen getreue Abbildungen des im Stillen ablaufenden Modellierungszyklus sind. Diese Gefahr resultiert aus der Anwendung des Ebene-3-Modells und gleichzeitig aus der Tatsache, dass Probanden bei der beschriebenen Versuchsanordnung gebeten werden, ihr Verhalten zu erklären. Ericsson & Simon (1993, S.86) warnen hiervor ausdrücklich: „Another kind of complementary instruction is a request for *explanation*. In order to get as full an understanding as possible of the subjects' processes, they are asked to explain their thinking. […] As follows from our earlier discussion, inducing the subject to explain his solution very likely changes the structure of his thought processes."

Dieses Problem kann aber nicht als spezifisch für die hier verwendete Versuchsanordnung betrachtet werden, sondern vielmehr als eine allgemeine Schwierigkeit bei Protokollen lauten Denkens: „TOL [thinking out loud] data should not be taken as *direct* reflections of thought processes but rather as data which are correlated with underlying thought processes.", so Olson et al. (1984, S. 254).

Trotz all dieser Bedenken, bleibt die anvisierte Methode immer noch die beste Alternative, da sie als einzige wirklichen Zugang zu kognitiven Prozessen gewährt. Dass diese eventuell keine deckungsgleichen Abbilder zu denjenigen Prozessen sein dürften, die ablaufen, wenn die jeweiligen Personen alleine und im Stillen arbeiten, ist hierbei in Kauf zu nehmen.

3.1.4 Material und Versuchsanordnung

Jedem Probanden wurden die fünf Modellierungsaufgaben nacheinander vorgelegt. Die Probanden wurden dabei instruiert, den Versuchsleiter schrittweise durch den Lösungsprozess zu führen.

Der Versuchleiter selbst war für die Fragestellung blind, und nur im Hinblick auf das von ihm zu zeigende Verhalten gegenüber dem Probanden geschult worden. Auf diese Weise sollte verhindert werden, dass der Versuchsleiter durch eventuelle Suggestivfragen gezielt bei denjenigen Teil- und Unterkompetenzen nachfragte, die der Proband ausgelassen hatte.

Mit im Untersuchungsraum saß außerdem eine weitere Person, die über die Äußerungen der Probanden Protokoll führte und sich Notizen darüber machte, welche Unterkompetenzen bei den Aussagen der Probanden identifiziert werden konnten und in welcher Reihenfolge diese auftraten. Die Protokollandin war vorher mit dem theoretischen Hintergrund vertraut gemacht worden und in mehreren Probedurchläufen für den Umgang mit den Protokollbögen geschult worden.

Zu Beginn jeder Aufgabe wurde eine Stoppuhr in Gang gesetzt. Immer wenn die Protokollandin eine Unterkompetenz entdeckte, notierte sie die bis dahin verstrichene Zeit im Protokollbogen. Der Protokollbogen (siehe Abbildung 17) listet alle von Maaß (2004) genannten Unterkompetenzen zu den fünf Teilschritten im Modellierungszyklus und bietet bei jeder Unterkompetenz Platz für die Notierung von Auftretenszeiten.

3.1.5 Beschreibung der Stichprobe

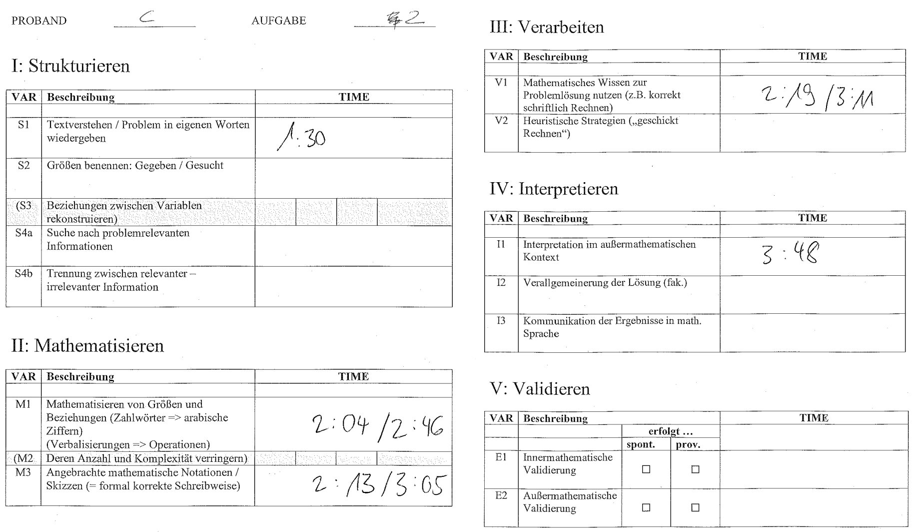
Da Studie 1 in erster Linie den Charakter einer Voruntersuchung hat, wurde die Anzahl der Probanden auf N=8 beschränkt, welche dann in jeweils ca. 60-minütigen Sitzungen untersucht wurden. Da jeder der Probanden alle fünf Aufgaben bearbeitete, ließen sich so in der Studie N=40 Modellierungsverläufe beobachten

Die Versuchspersonen wurden in Kooperation mit zwei Schulen und einer Hausaufgabenbetreuungsstelle akquiriert, Teilnahmebedingung war, dass die Probanden keine Schwierigkeiten in Mathematik haben durften (die letzte Zeugnisnote der Teilnehmer in Mathematik lag im Durchschnitt bei 1.75, die Note bei der letzten Klassenarbeit in Mathematik bei 1.62).

6 Versuchspersonen waren männlich, zwei weiblich. Mit Ausnahme eines Probanden, der erst 9 Jahre alt war, waren alle Teilnehmer 10 Jahre alt und besuchten die vierte oder fünfte Klasse.

3.1.6 Auswertung der Daten

Grundlage für die Auswertung der Sitzungen bildeten die Untersuchungsprotokolle der einzelnen Aufgaben, aus denen der Verlauf des jeweiligen Modellierungszyklus rekonstruiert wurde. Abbildung 17 zeigt einen exemplarischen Protokollbogen, der wie folgt zu lesen ist:



PROBAND C AUFGABE 2

I: Strukturieren

VAR	Beschreibung	TIME
S1	Textverstehen / Problem in eigenen Worten wiedergeben	1:30
S2	Größen benennen: Gegeben / Gesucht	
(S3	Beziehungen zwischen Variablen rekonstruieren)	
S4a	Suche nach problemrelevanten Informationen	
S4b	Trennung zwischen relevanter – irrelevanter Information	

II: Mathematisieren

VAR	Beschreibung	TIME
M1	Mathematisieren von Größen und Beziehungen (Zahlwörter => arabische Ziffern) (Verbalisierungen => Operationen)	2:04 / 2:46
(M2	Deren Anzahl und Komplexität verringern)	
M3	Angebrachte mathematische Notationen / Skizzen (= formal korrekte Schreibweise)	2:13 / 3:05

III: Verarbeiten

VAR	Beschreibung	TIME
V1	Mathematisches Wissen zur Problemlösung nutzen (z.B. korrekt schriftlich Rechnen)	2:19 / 3:11
V2	Heuristische Strategien („geschickt Rechnen“)	

IV: Interpretieren

VAR	Beschreibung	TIME
I1	Interpretation im außermathematischen Kontext	3:48
I2	Verallgemeinerung der Lösung (fak.)	
I3	Kommunikation der Ergebnisse in math. Sprache	

V: Validieren

VAR	Beschreibung	erfolgt … spont.	erfolgt … prov.	TIME
E1	Innermathematische Validierung	☐	☐	
E2	Außermathematische Validierung	☐	☐	

Abbildung 17: Protokollbogen

Nach 1:30 Minuten berichtete der Proband von Textverstehen bzw. gab das Problem in eigenen Worten wieder. Nach 2:04 Minuten erfolgte die Anweisung an den Versuchsleiter, wie einige Informationen aus dem Aufgabentext in die mathematische Sprache zu übersetzen seien (dementsprechend erfolgte 9 Sekunden später die Anweisung, diese Informationen in formal korrekter Schreibweise zu Papier zu bringen). Nach 2:19 Minuten wurden diese nunmehr mathematischen Informationen verarbeitet – sprich: der Proband gab die Anweisung etwas auszurechnen.

Die Aufgabe war nun aber noch nicht gelöst, denn der Proband war erst bei einem Zwischenergebnis angelangt. Nach 2:46 Minuten wurden die zur endgültigen Lösung noch fehlenden Informationen ihrerseits auch mathematisiert, nach 3:05 Minuten formal korrekt niedergeschrieben und nach 3:11 Minuten mithilfe mathematischen Wissens in ein mathematisches Ergebnis umgewandelt, sowie letztlich nach 3:48 Minuten interpretiert.

Der so aus den Protokollen herauszulesende Modellierungsverlauf wurde jeweils graphisch veranschaulicht, indem der Weg des Probanden in chronologischer Abfolge nachgezeichnet wurde.

Abbildung 18 zeigt beispielhaft einen solchen Modellierungsverlauf, abgeleitet aus dem Protokollbogen der Abbildung 17.

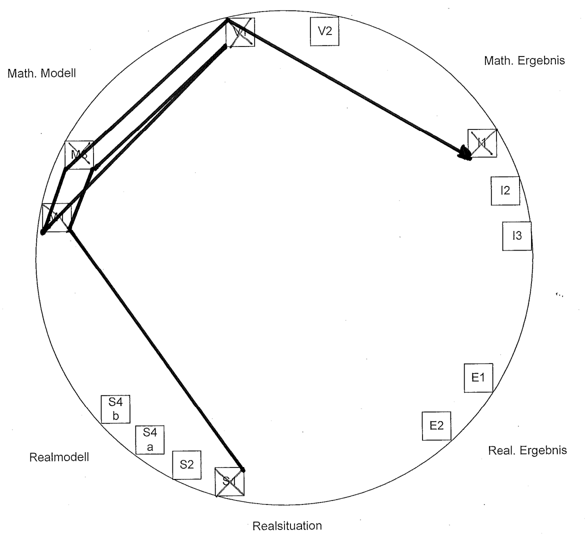


Abbildung 18: Exemplarischer Modellierungsverlauf

Neben der Aufzeichnung des Modellierungsverlaufs wurden aus den Protokollbögen die folgenden Informationen extrahiert:

- Erstens wurde neben der Reihenfolge der auftretenden Unterkompetenzen eine Übersicht über die Anzahl der aufgetretenen Unterkompetenzen erstellt, inklusive einer Information darüber, zu welchen Teilkompetenzen überhaupt keine Umsetzung von jeweiligen Unterkompetenzen beobachtet werden konnten. (So wird z.B. aus dem in Abbildung 17 dargestellten Protokoll ersichtlich, dass in diesem Fall die Teilkompetenz „*Validieren*“ nicht realisiert und dass lediglich eine von vier Unterkompetenzen der Teilkompetenz „*Strukturieren*“ umgesetzt wurde).
- Zweitens wurde jeweils erfasst, inwiefern die Probanden eventuell vom theoretisch vorgegebenen Modellierungsverlauf abgewichen waren. Hintergrund zu dieser Auswertung ist die Fragestellung, inwieweit die Modellierungsverläufe wirklich zyklisch ablaufen, oder ob nicht vielmehr Rückkoppelungen von jeder Teilkompetenz zu jeder anderen Teilkompetenz möglich sind. Wie können solche Abweichungen nun aber (über die rein graphische Darstellung hinaus) numerisch abgebildet werden? Hierzu wurde eine Logik zur Vergabe von Abweichungspunkten entwickelt, wobei die Zahl der vergebenen Abweichungspunkte umso größer wird, je mehr der Modellierungsverlauf eines Individuums vom theoretisch postulierten Modell abweicht. Wie solche Abweichungen inhaltlich zu interpretieren sind, wird in Kapitel 3.3 diskutiert. Tabelle 7 zeigt wie viele Abweichungspunkte jeweils beim Übergang von einer Teilkompetenz zur anderen vergeben wurden.

Tabelle 7: Quantifizierung der Abweichung vom Modellierungszyklus

Zu ⇨ Von ⇩	Strukturieren	Mathematisieren	Verarbeiten	Interpretieren	Validieren
Strukturieren	0	0	1	2	0
Mathematisieren	1	0	0	1	0
Verarbeiten	2	1	0	0	0
Interpretieren	3	2	1	0	0
Validieren	0	0	0	0	0

- Verbindungen innerhalb einer Teilkompetenz (also z.B. von I1 nach I3) werden nicht mit Abweichungspunkten bewertet, da die verschiedenen Unterkompetenzen lediglich zur besseren Übersichtlichkeit nummeriert sind. Dahinter verbirgt sich jedoch keine zeitliche Reihung der Unterkompetenzen.
- Da der Modellierungszyklus beim Durchschreiten der Teilkompetenzen die Reihenfolge *Strukturieren – Mathematisieren – Verarbeiten – Interpretieren – Validieren* postuliert, werden die Verbindungen einer Teilkompetenz zur jeweils darauf folgenden Teilkompetenz nicht mit Abweichungspunkten versehen.
- Während also eine Verbindung von *Strukturieren* nach *Mathematisieren* theoriekonform ist und nicht mit Abweichungspunkten versehen wird, stellt eine rückwärts gerichtete Verbindung von *Mathematisieren* zu *Strukturieren* eine Verletzung des theoretisch postulierten Ablaufes dar. Verbindungen zur jeweils vorhergehenden Teilkompetenz werden daher mit einem Abweichungspunkt bewertet.
- Auch das Überspringen eines Stadiums ist als nicht theoriekonform zu betrachten und daher mit einem Abweichungspunkt zu bewerten (wenn also nach dem *Strukturieren* direkt zum *Verarbeiten* übergegangen wird, ohne dass ein erkennbarer Mathematisierungsschritt vollzogen wird). Das Überspringen von zwei Stadien wird dementsprechend mit zwei Abweichungspunkten bewertet.
- Zwei Abweichungspunkte werden auch dann vergeben, wenn der Modellierer zwei Schritte zurück geht, wenn der Verlauf also sowohl rückwärts gerichtet ist, als auch ein Stadium übersprungen wird (beispielsweise von *Interpretieren* zurück zu *Mathematisieren*). Beim rückwärtigen Überspringen von zwei Teilkompetenzen werden entsprechend drei Abweichungspunkte vergeben.
- Gleichzeitig wird die Teilkompetenz „*Validieren*“ aus dem soeben beschriebenen Schema ausgeschlossen, das heisst Verbindungen zum Validierungsschritt und von dort aus zurück zu anderen Stadien sind zu jeder Zeit theoriekonform und werden nicht mit Abweichungspunkten versehen. Grund hierfür ist die Annahme, dass das Überwachen und Überprüfen eigener Vorgehensweisen im Sinne metakognitiver Prozesse während des gesamten Geschehens auftreten kann und nicht auf das Ende des Problemlöseprozesses beschränkt sein muss.

3.2. Ergebnisse

Die Ergebnisse der qualitativen Voruntersuchung werden im Folgenden dargestellt und anschließend diskutiert. Letztlich werden aus den Ergebnissen einige Konsequenzen für die Hauptuntersuchung abgeleitet.

3.2.1 Gütekriterien

Bevor die Ergebnisse der ersten Studie dargestellt werden, soll kurz auf die psychometrischen Gütekriterien des verwendeten Protokollbogens eingegangen werden. Da die Studie nur einen ersten groben Überblick über die Vorgehensweise von Modellierern ermöglichen sollte und daher eher explorativen Charakter hat, müssen die an die Gütekriterien anzulegenden Maßstäbe nicht ganz so streng sein wie bei hypothesenprüfenden Untersuchungen.

3.2.1.1 Objektivität

Im Zuge der Schulung der Protokollandin wurde auch die Objektivität des Instrumentes geprüft.

Hierfür wurden nach zwei Probedurchläufen drei Modellierungsverläufe einer 10 jährigen Probandin sowohl von der Protokollandin als auch von der Autorin beobachtet und anhand des Protokollbogens bewertet.

Zur Berechnung der Interraterreliabilität wurde im Anschluss verglichen, wie oft die beiden Rater jeweils gleiche, und wie oft sie jeweils unterschiedliche Urteile abgaben. Aus den so entstandenen Kreuztabellen wurde für jeden der Verläufe Cohen’s Kappa errechnet.

Tabelle 8: Interraterreliabilität

	Rater B: Prozess aufgetreten			Rater B: Prozess nicht aufgetreten		
Verlauf	1	2	3	1	2	3
Rater A: Prozess aufgetreten	10	9	8	0	1	1
Rater A: Prozess nicht aufgetreten	0	0	0	3	3	4

In Tabelle 8 sind die Ergebnisse der beiden Ratings gegenübergestellt. Im ersten Verlauf traten nach beiden Einschätzungen dieselben zehn 10 Unterkompetenzen auf, woraus sich ein Kappa-Wert von $\kappa=1.00$ ergibt. Im zweiten und dritten Verlauf wichen Rater A und Rater B bei jeweils einer Kategorie in ihrem Urteil voneinander ab, dies entspricht den Kappa-Werten $\kappa=.81$, $p=.03$ bei Verlauf 2 bzw. $\kappa=.83$; $p=.02$ bei Verlauf 1.

Diese Werte sprechen für eine sehr hohe Übereinstimmung zwischen beiden Ratern (Wirtz & Caspar, 2002) und aufgrund dessen auch für eine mehr als ausreichende Objektivität des Verfahrens.

3.2.1.2 Reliabilität

Zur Bestimmung der Reliabilität des Ratingverfahrens wurde die innere Konsistenz berechnet. Da innere Konsistenz und Eindimensionalität nichts miteinander zu tun haben (Schermelleh-Engel & Werner, 2007), können zu deren Berechnung alle 13 Items der Ratingskala hinzugezogen werden, auch wenn angenommen wird, dass den Items verschiedene latente Dimensionen zugrunde liegen.

Analog zu Cronbach's Alpha kann die innere Konsistenz bei dichotomen Daten mithilfe der Kuder-Richardson-Formel (KR-20) bestimmt werden, wobei der Reliabilitätskoeffizient abhängig ist von Itemschwierigkeit, Varianz der Testrohwerte und Anzahl der Items.

Unter Berücksichtigung aller 13 Items bei 40 Fällen ergibt sich eine Reliabilität von $r_{tt}=.64$.

Entfernt man die Items I2, E1 und E2, die aufgrund ihrer Schwierigkeit von 1,00 eine Nullvarianz aufweisen, so erhöht sich die Reliabilität auf $r_{tt}=.66$. Auch dieser Wert liegt allerdings noch etwas unter dem häufig geforderten Mindestwert von .70 (Jäger & Petermann, 1999).

3.2.1.3 Validität

Auf mögliche Einschränkungen der Validität verbaler Daten durch Prozesse wie Reaktivität und Nonveridikalität wurde bereits hingewiesen (siehe Kapitel 3.1.2). Auch bei der vorliegenden Studie können derartige Effekte nicht ausgeschlossen werden.

Insbesondere muss davon ausgegangen werden, dass trotz entsprechender Anleitung Probanden nicht alle ablaufenden Prozesse verbalisieren, schon allein deshalb, weil nicht angenommen werden kann, dass sämtliche Prozesse dem Bewusstsein zugänglich sind.

Nur weil Unterkompetenzen nicht beobachtet werden konnten, ist daher nicht der Schluss zulässig, dass diese bei Modellierern nicht auftreten.

Umgekehrt kann jedoch argumentiert werden, dass von der Beobachtung eines Prozesses auf dessen tatsächliche Existenz geschlossen werden kann. Das ein Prozess fälschlicherweise identifiziert wird, in Wahrheit jedoch gar nicht aufgetreten ist, kann natürlich in Einzelfällen durchaus möglich sein.

Betrachtet man jedoch eine große Anzahl von Modellierungsverläufen, so kann davon ausgegangen werden, dass Unterkompetenzen, die bei einer hohen Prozentzahl der Verläufe identifiziert werden können, auch tatsächlich auftreten.

Ob durch den Modellierungsprozess die jeweilige Aufgabe richtig gelöst werden konnte oder nicht, kann im Übrigen nicht als Indikator für die Validität herangezogen werden. Die Korrektheit der Lösung einer Aufgabe ist zunächst vollkommen unabhängig vom alleinigen Auftreten einer Teil- oder Unterkompetenz zu betrachten. Bei einer Klassenarbeit rechnet jeder Schüler (er verarbeitet also), trotzdem kann das Ergebnis dieses Verarbeitungsprozesses falsch sein. Da in der ersten Studie jedoch nur erfasst wurde, *ob* ein Modellierer bestimmte Kompetenzen realisiert, nicht aber *wie gut* er diese umzusetzen versteht, ist eine Betrachtung von Zusammenhängen mit der Lösung zur Bestimmung der Validität nicht sinnvoll.

3.2.2 Auftretende Teilkompetenzen

Welche der theoretisch postulierten Teilkompetenzen des Modellierungszyklus treten nun auch tatsächlich in Modellierungsverläufen von Grundschulabgängern auf? Tabelle 9 listet Anzahl und Prozentsatz der Modellierungsverläufe auf, in denen die jeweiligen Teilkompetenzen von den Probanden realisiert wurden.

Tabelle 9: Auftretende Teilkompetenzen in Studie 1

Teilkompetenz	Aufgetreten in Anzahl der Fälle	 Prozent der Fälle
Strukturieren	29	72.5
Mathematisieren	36	90
Verarbeiten	32	80
Interpretieren	34	85
Validieren	0	0

Die ersten vier Teilkompetenzen konnten wenn auch nicht bei allen, so doch bei den meisten Probanden identifiziert werden. Dass es auch jeweils Fälle gibt, in denen Teilkompetenzen nicht realisiert werden, ist durchaus denkbar. Besonders auffällig ist hingegen die Tatsache, dass die Teilkompetenz *Validieren* bei keinem Probanden und bei keiner Aufgabe beobachtet werden konnte.

3.2.3 Auftretende Unterkompetenzen

Abgesehen von „*Validieren*" konnten also alle Teilkompetenzen bei einem Großteil der Modellierungsverläufe beobachtet werden. Wie sieht es aber nun eine Ebene tiefer, bei den Unterkompetenzen aus? Abbildung 19 zeigt die absoluten Häufigkeiten des Auftretens, Tabelle 10 gibt zudem die prozentualen Häufigkeiten und den Wortlaut der Unterkompetenz auf dem Beobachtungsbogen wieder.

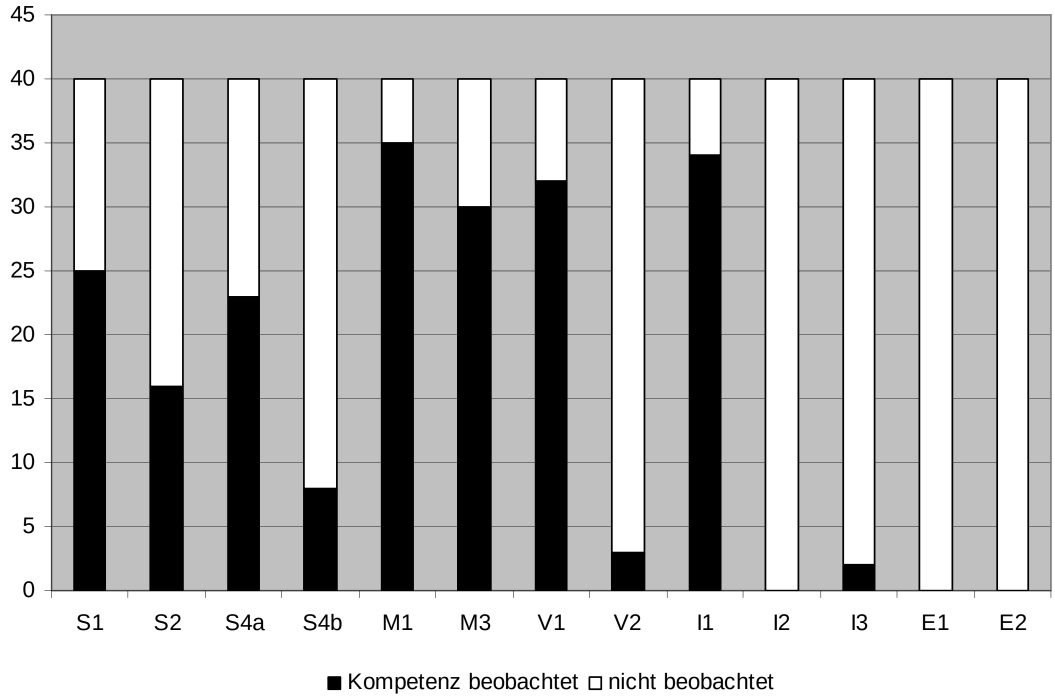


Abbildung 19: Auftretende Unterkompetenzen

Festzuhalten sind bei der Analyse der Ergebnisse folgende Punkte:

1. Die im Abschnitt 2.3. theoretisch als Hauptindikatoren der Teilkompetenzen definierten Unterkompetenzen (S1 für *Strukturieren*, M1 für *Mathematisieren*, V1 für *Verarbeiten* und I1 für *Interpretieren*) treten innerhalb ihrer jeweiligen Teilkompetenzen im Vergleich zu anderen Unterkompetenzen am häufigsten auf.
2. Die Unterkompetenzen V2, I2, und I3 sowie E1 und E2 treten nicht oder doch so selten auf, dass sie als empirisch irrelevant für den Problemlöseprozess betrachtet werden können.

Tabelle 10: Auftretende Unterkompetenzen

Variable	Item	Anzahl	Prozent
S1	Textverstehen / Problem in eigenen Worten wiedergeben	25	63
S2	Größen benennen: Gegeben / Gesucht	16	40
S4a	Suche nach problemrelevanten Informationen	23	58
S4b	Trennung zwischen relevanter – irrelevanter Information	8	20
M1	Mathematisieren von Größen und Beziehungen (Zahlwörter => arabische Ziffern) (Verbalisierungen => Operationen)	35	88
M3	Angebrachte mathematische Notationen / Skizzen (= formal korrekte Schreibweise)	30	75
V1	Mathematisches Wissen zur Problemlösung nutzen (z.B. korrekt schriftlich Rechnen)	32	80
V2	Heuristische Strategien („geschickt Rechnen“)	3	8
I1	Interpretation im außermathematischen Kontext	34	85
I2	Verallgemeinerung der Lösung	0	0
I3	Kommunikation der Ergebnisse in math. Sprache	2	5
E1	Textverstehen / Problem in eigenen Worten wiedergeben	0	0
E2	Größen benennen: Gegeben / Gesucht	0	0

3.2.4 Abweichungen vom theoretisch postulierten Modellierungszyklus

Inwieweit beschreibt der Modellierungszyklus tatsächlich die Art und Weise, wie Personen beim Lösen von Modellierungsaufgaben vorgehen? Eine Antwort auf diese Frage gibt die nähere Betrachtung zweier Arten von Abweichungen vom vollständigen Modellierungszyklus. Einerseits ist dies der *Wegfall von Einzelprozessen*, sprich: von Teilkompetenzen, andererseits sind dies *Abweichungen in der Reihenfolge des Abschreitens der Teilkompetenzen*.

Zunächst zum Wegfall der Teilkompetenzen: In allen untersuchten Modellierungsverläufen fehlte die Teilkompetenz *Validieren* komplett. Wie oft darüber hinaus die anderen Teilkompetenzen entfielen, ist in Tabelle 9 dargestellt. Auf der Ebene der einzelnen Verläufe gilt es darüber hinaus festzuhalten, ob und wie oft sogar mehr als zwei Teilkompetenzen komplett entfielen.

Bei 21 der 40 untersuchten Modellierungsverläufe und somit bei etwas mehr als der Hälfte entfiel nur ein Prozess, das *Validieren*. Bei 12 Fällen wurde auf einen weiteren Prozess verzichtet, bei 4 Fällen sogar auf insgesamt drei Teilkompetenzen. Bei drei Extremfällen konnte sogar nur eine Teilkompetenz beobachtet werden.

Auch bei den Unterkompetenzen konnten Abweichungen dergestalt festgestellt werden, dass bei weitem nicht alle der theoretisch postulierten Unterkompetenzen beobachtet werden konnten. In den meisten Fällen wurde zwar aus jeder Teilkompetenz eine Unterkompetenz (meistens der o.g. Hauptindikator) gezeigt, ansonsten ging die Tendenz aber eher zu einer mittleren Anzahl an auftretenden Unterkompetenzen. Theoretisch wäre ein Maximum von 13 Unterkompetenzen möglich, das empirische Maximum liegt jedoch bei 9 beobachteten Unterkompetenzen. Das – nur in einem Fall aufgetretene – Minimum liegt bei nur einer einzigen Unterkompetenz. Im Mittel kamen die Modellierer mit 5.2 Unterkompetenzen aus, die Standardabweichung lag bei SD=1.99.

Auch die Reihenfolge der auftretenden Teilkompetenzen wich in vielen Fällen deutlich vom theoretisch postulierten Modell ab. Tabelle 11 zeigt, wie oft welche Art der Abweichung (Auslassungen, Rückschritte, sowie Rückschritte und Auslassungen) beobachtet werden konnte und mit welcher Häufigkeit dies geschah. Gleichzeitig bleibt allerdings festzuhalten, dass in den meisten Fällen keine (57.5% der Fälle) oder nur eine (17.5% der Fälle) Abweichung beobachtet werden konnte, sprich: Drei Viertel aller Modellierungsverläufe entsprechen fast oder ganz dem theoretisch postulierten Schema.

Tabelle 11: Verteilung der Abweichungspunkte

Zu ⇨ / Von ⇩	Strukturieren	Mathematisieren	Verarbeiten	Interpretieren	Validieren
Strukturieren			3	3	
Mathematisieren	4			4	
Verarbeiten	2	5			
Interpretieren	-	3	-		
Validieren					

Im Folgenden sollen die beobachteten Abweichungen an einigen Beispielen graphisch veranschaulicht werden. In Abbildung 20 sind zunächst vier beispielhafte Verläufe dargestellt, in denen keine Abweichungen beobachtet werden konnten. Verlauf 04 macht deutlich, dass die Abweichungspunkte lediglich Rückschritte und das Überspringen von Phasen erfassen, nicht aber das vorzeitige Beenden des Modellierungszyklus. Eine Situation, wie in Verlauf 04 dargestellt, ergibt sich beispielsweise dann, wenn ein Proband nach dem Berechnen eines Ergebnisses dieses im Raum stehen lässt, ohne ihm noch durch eine Interpretation Sinn zu verleihen und es auf die Ausgangsfrage zu beziehen.

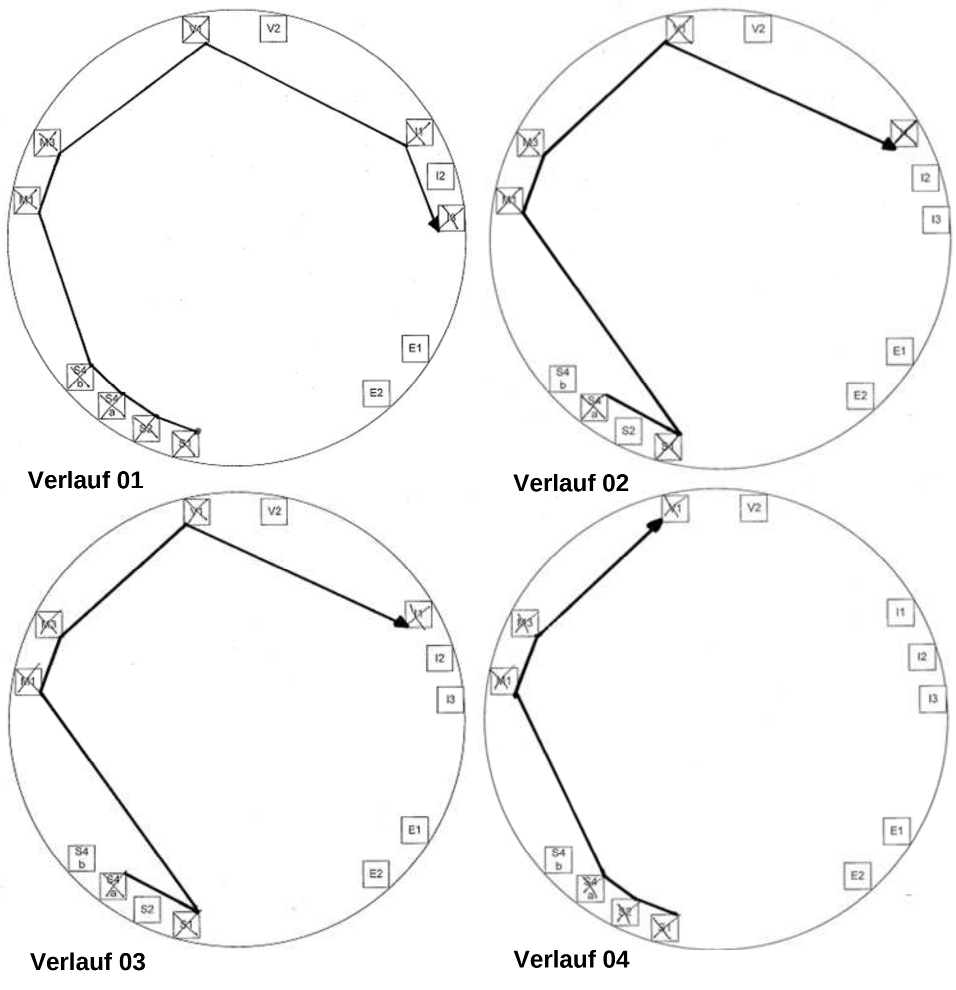


Abbildung 20: Verläufe ohne Abweichungspunkte

In Abbildung 21 sind hingegen Verläufe abgebildet, die geringfügig vom theoretisch postulierten Ablauf abweichen. Dennoch ist das Grundmuster insofern noch wenig verletzt, als dass sich die Verläufe noch recht stark „am Rand entlang“ vollziehen, der Kreis wird also nicht von einer Seite auf die andere durchkreuzt. Das liegt daran, dass bei den hier dargestellten Fällen, bei denen ein (Verlauf 03 und 04) oder zwei (Verlauf 01 und 02) Abweichungspunkte vergeben werden, nur geringfügige Abweichungen in Form von Rückkoppelungen auftreten. Modellierer gehen hier anscheinend noch mal „einen Schritt zurück“, beispielsweise um ihre Informationsbasis zu erweitern.

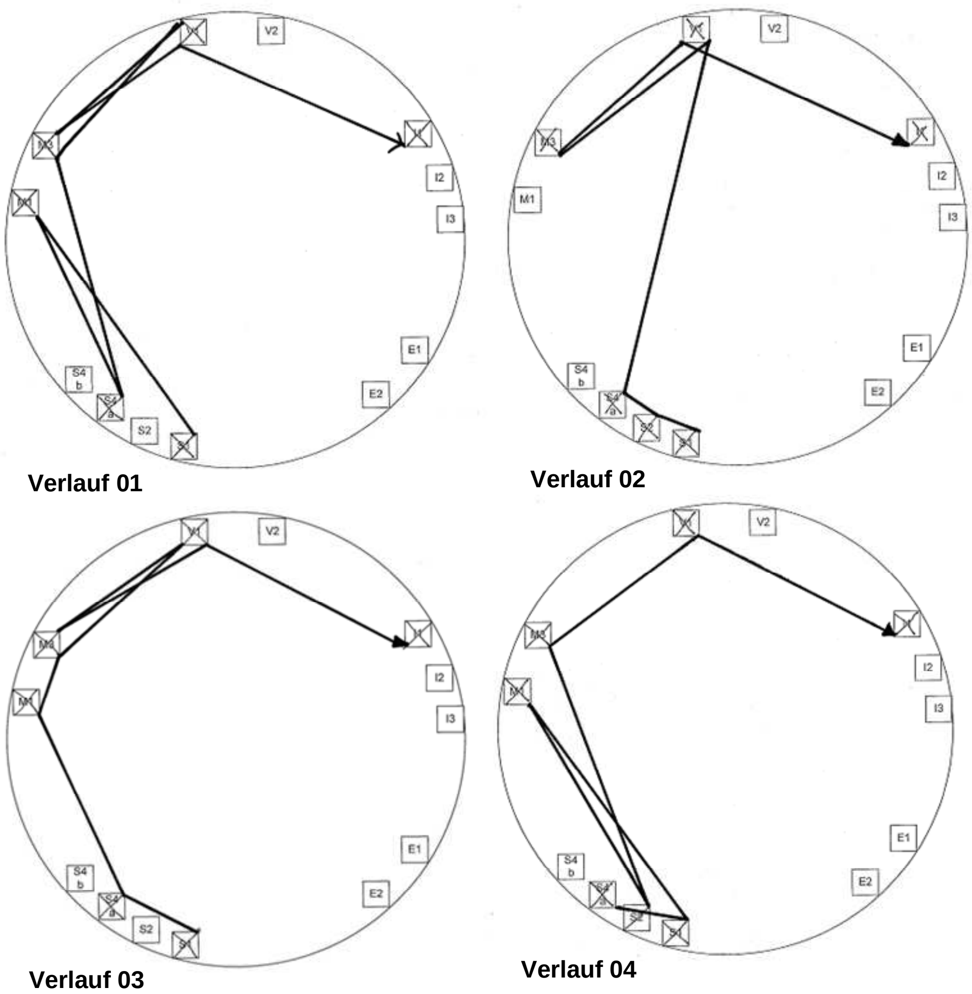


Abbildung 21: Verläufe mit einem oder zwei Abweichungspunkten

Die Verläufe in Abbildung 22 weichen hingegen deutlich vom Grundmuster ab. Die Kreise, die den Modellierungszyklus darstellen, werden stark durchkreuzt. Trotzdem bleibt (mit Ausnahme von Verlauf 03) erkennbar, wo sich Anfang und Ziel des Prozesses befinden. Auch hier geht es letztlich um das Abschreiten der Stadien im Uhrzeigersinn. Im Vergleich zu anderen, einfacheren Verläufen beobachtet man jedoch noch mehr Rückkoppelungen.

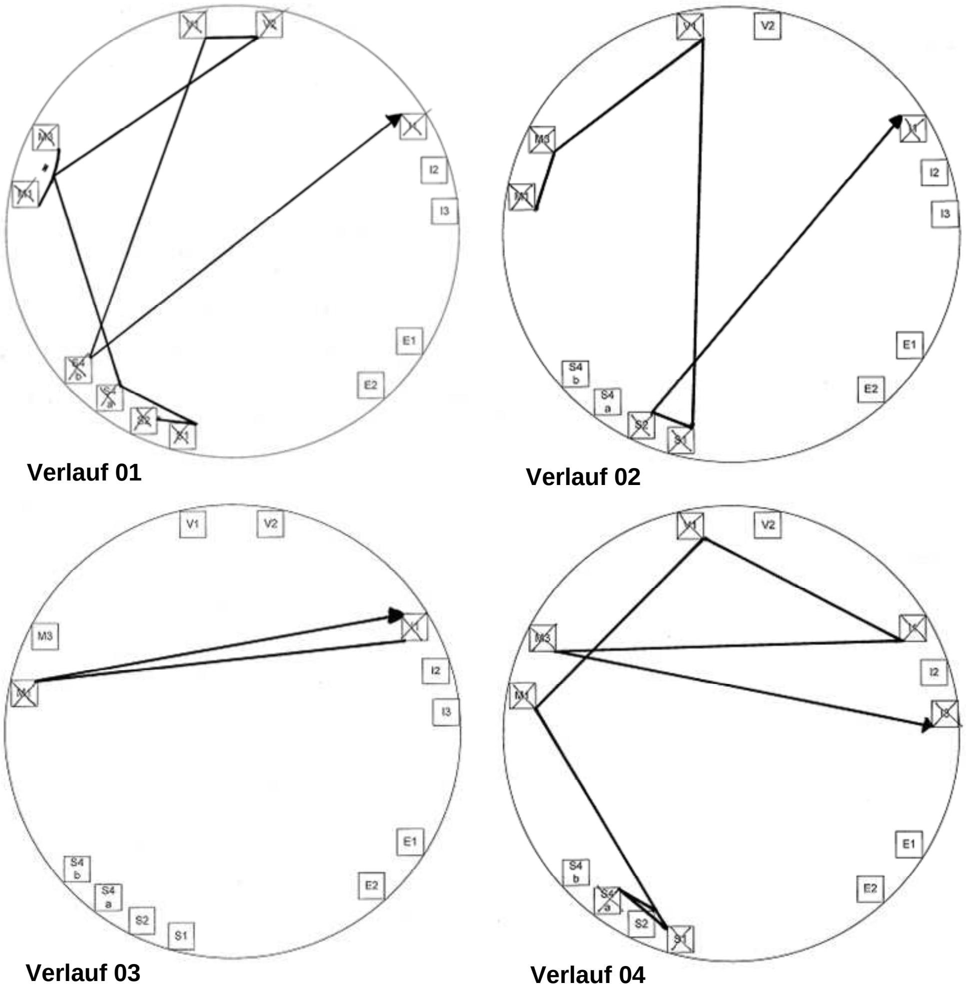


Abbildung 22: Verläufe mit drei und vier Abweichungspunkten

Anhand dieser graphischen Darstellungen wurden nun einige der empirischen Beobachtungen von Abweichungen zum Modellierungszyklus aufgezeigt. Wie solche Abweichungen gedeutet werden können, ist nun zu diskutieren.

3.3 Diskussion

Die im vorhergehenden Abschnitt dargelegten Ergebnisse basieren auf einem Forschungsansatz, der auf die vorliegende Fragestellung bisher noch nicht angewandt wurde. Umso wichtiger ist daher die Gewährleistung der klassischen Gütekriterien, deren Erfüllung als Grundvoraussetzung für die Interpretierbarkeit der Ergebnisse gelten muss.

Die berichtete Objektivität, gemessen an der Beurteilerübereinstimmung in Bezug auf die Protokolle, kann als gut bis sehr gut eingestuft werden. Die Objektivität als Grundlage für weitere Gütekriterien ist daher als gegeben zu betrachten.

Etwas kritischer sieht es allerdings bei der Reliabilität aus. Die innere Konsistenz verpasst – wenn auch nur knapp – das Kriterium für einen zufriedenstellenden Reliabilitätskoeffizienten, liegt aber noch im akzeptablen Vertrauensintervall. Für Forschungszwecke ist dies ausreichend, für die Entwicklung eines Testverfahrens zur Erfassung von Modellierungskompetenzen sind jedoch deutlich höhere Werte zu fordern.

Über die Validität liegen keine empirischen Daten vor, weshalb sie auch am kritischsten zu bewerten ist. Hier hat man es allerdings mit einer generellen Schwäche der Methode lauten Denkens auf Ebene 3 nach Ericsson & Simon (1980) zu tun. Grundsätzlich bleibt bei einem solchen Untersuchungsansatz immer die Frage offen, ob und inwieweit der Prozess genauso verlaufen wäre, wenn kein lautes Denken stattgefunden hätte.

Annähernd Sicherheit verschafft daher die Anzahl und Varianz der Beobachtungen. Die 40 Modellierungsverläufe basieren auf verschiedenen Modellierern und verschiedenen Aufgaben. Werden Prozesse oder Muster also in mehreren Fällen beobachtet, so kann daraus eine Überzufälligkeit abgeleitet werden. Einzelbeobachtungen sind hingegen grundsätzlich kritisch zu bewerten, insbesondere dann, wenn sie sich vor dem theoretischen Hintergrund nur schwer interpretieren lassen.

3.3.1 Auftretende Teilkompetenzen

Die Ergebnisse der ersten Studie zeigen, dass die Teilkompetenzen *Strukturieren*, *Mathematisieren*, *Verarbeiten* und *Interpretieren* bei den allermeisten Modellierungsverläufen identifiziert werden können. Man muss sich nun die Frage stellen, wie es überhaupt möglich sein kann, Aufgaben zu lösen, wenn einer dieser Teilprozesse wegfällt. Hierfür sind durchaus plausible Szenarien denkbar:

Strukturieren gilt als grundlegender Schritt, um zunächst einmal die Aufgabenstellung zu identifizieren. Entällt dieser Schritt, so wird gerechnet, ohne dass überhaupt der Versuch unternommen wird zu verstehen, worin das eigentliche Problem liegt. Probanden „scannen" den Aufgabentext nach Zahlen und bestimmten Schlüsselwörtern, wie „weniger", „mehr als" etc. und mathematisieren diese Angaben quasi „blind", um sie im nächsten Schritt weiterzuverarbeiten. Zu der Frage, wie man an Textaufgaben herangehen soll, gab ein 10jähriges Mädchen zu Protokoll: „Du schaust einfach, wo in dem Text Zahlen stehen und schreibst sie raus. Und dann kuckst Du, wie man da irgendwie eine Aufgabe draus machen kann."

Dass ein Verzicht auf den Strukturierungsprozess gerade bei komplexen Aufgaben der Qualität der Lösung nicht zuträglich sein kann, liegt auf der Hand. Bei einem Viertel der hier untersuchten Modellierungsverläufe konnten solche Prozesse aber nicht beobachtet werden.

Es wäre nun sicher verfrüht, in diesem Mangel an Strukturierungsprozessen die Erklärung dafür zu suchen, dass viele Schüler beim Lösen von Modellierungsaufgaben scheitern. Es lohnt sich jedoch durchaus, diese Hypothese weiter zu verfolgen.

Modellierungsprozesse ohne erkennbare Mathematisierung wurden auch beobachtet, wenn auch weitaus seltener. Wie lässt sich dies interpretieren?

Betrachten wir zunächst den Ausfall der nächsten Teilkompetenz: In 20% der Modellierungsverläufe konnte kein *Verarbeiten* beobachtet werden, es fand also kein beobachtbares Rechnen oder irgendeine Anwendung mathematischen Wissens statt. Dies ist durchaus plausibel, da Probanden Aufgaben schließlich auch nicht-mathematisch lösen können, indem sie die Antwort raten oder schätzen.

Weitaus schwieriger ist es nun aber, Erklärungen dafür zu finden, dass bei einer Aufgabe keine Informationen aus dem Text mathematisiert werden. An dieser Stelle ist eine nähere Betrachtung der Überschneidung zwischen Ausfällen beim *Mathematisieren* und Ausfällen beim *Verarbeiten* angezeigt.

Bei einer Kreuztabellierung beider Ausfälle stellt man folgendes fest: In acht von 40 Fällen wurde nicht verarbeitet, in 4 von 40 Fällen wurde nicht mathematisiert. In allen vier Fällen, in denen nicht mathematisiert wurde, wurde in der Folge auch nicht verarbeitet. Man hat es hier anscheinend in allen Fällen mit Verläufen zu tun, in denen die Probanden auf eine mathematische Herangehensweise verzichten und versuchen, die Aufgaben auf anderem Weg zu lösen. Der Lösungsprozess verzichtet in diesen 4 Fällen anscheinend also komplett auf Übersetzungsvorgänge und vollzieht sich rein in der außermathematischen Welt.

Auch ein Wegfall des Interpretationsschrittes, wie er in 6 Fällen beobachtet werden konnte, muss dem Modellierer nicht unbedingt auffallen, ist er doch bereits im Schritt „*Verarbeiten*“ zu einer Lösung des Problems gelangt. Verzichtet wird dann lediglich darauf, die so gewonnene Lösung wieder zurück auf das ursprüngliche Problem zu übertragen.

Es ist kaum verwunderlich, dass Modellierungsverläufe auch ohne anschließende Validierung des Ergebnisses erfolgreich sein können. Dennoch erstaunt es, dass in keinem einzigen Fall eine wie auch immer geartete Überprüfung des Endergebnisses beobachtet werden konnte.

Betrachtet man den Modellierungszyklus als deskriptives Modell, so ist die Kompetenz *Validieren* also anscheinend fehl am Platz.

Dies führt uns zu einer ersten Revision des Modellierungszyklus, dargestellt in Abbildung 23. Im Vergleich zur bisherigen Betrachtungsweise wollen wir den Modellierungszyklus nicht mehr gleichzeitig als Deskription und Präskription auffassen, sondern den Zyklus in seiner deskriptiven Funktion an die aktuellen Befunde anpassen.

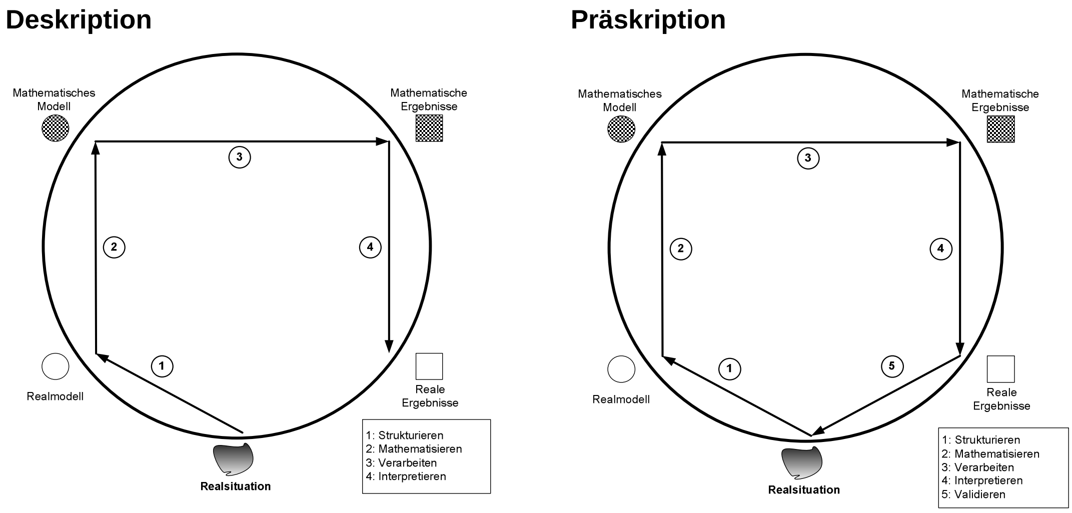


Abbildung 23: Erste Revision des Modellierungszyklus

3.3.2 Auftretende Unterkompetenzen

Auf der Ebene der auftretenden Unterkompetenzen konnte beobachtet werden, dass die als Hauptindikatoren für die jeweilige Teilkompetenz verstandenen Unterkompetenzen jeweils am häufigsten auftraten. Dies bestätigt die von der Theorie vorgegebene Schlüsselrolle der jeweiligen Indikatoren. Für weitere Untersuchungen und insbesondere im Hinblick auf die Generierung von Items zur Erfassung der Kompetenzen sollen daher diese jeweils am häufigsten aufgetretenen Kompetenzen operationalisiert werden. Im Einzelnen sind dies:

- Beim *Strukturieren* die Unterkompetenz S1: Verstehen des Textes bzw. Erfassen der Problemstellung
- Beim *Mathematisieren* die Unterkompetenz M1: Mathematisieren von Größen und Beziehungen
- Beim *Verarbeiten* die Unterkompetenz V1: Mathematisches Wissen zur Problemlösung nutzen
- Beim *Interpretieren* die Unterkompetenz I1: Interpretation der Lösung im außermathematischen Kontext

Bei der Betrachtung von Abbildung 19 bzw. Tabelle 10 wird allerdings deutlich, dass auch die Unterkompetenzen S4a (Suche nach problemrelevanter Information) und M3 (Angebrachte mathematische Notationen / formal korrekte Schreibweise) bei mehr als der Hälfte der Modellierungsverläufe zu beobachten sind und ihnen daher auch eine entsprechende Wichtigkeit beizumessen ist.

Grundsätzlich gilt es festzuhalten, dass – betrachtet man den Modellierungszyklus als deskriptives Modell – nicht alle Unterkompetenzen von Maaß (2004) sinnvollerweise in einem solchen deskriptiven Modell angesiedelt werden können.

Dies gilt einerseits für sämtliche Unterkompetenzen der Teilkompetenz *Validieren*, andererseits aber auch für die Unterkompetenz I2 (Verallgemeinerung der Lösung), welche in keinem einzigen Modellierungsverlauf beobachtet werden konnte. Auch bei denjenigen Unterkompetenzen, die in weniger als der Hälfte aller Verläufe beobachtet werden konnten, ist eine Platzierung im deskriptiven Modell anzuzweifeln. In Bezug auf die Unterkompetenzen innerhalb des Modellierungszyklus lässt sich demnach Folgendes schlussfolgern:

Aufgrund der Tatsache, dass viele Unterkompetenzen nur bei einem Teil der Verläufe oder gar nicht beobachtet werden konnten, ist es für nachfolgende Betrachtungen und Untersuchungen des Phänomens Modellierungskompetenz angezeigt, diese insbesondere auf der Makroebene zu betrachten, nämlich lediglich auf der Ebene der Teilkompetenzen. Die Unterkompetenzen, die jeweils als zentral für die Teilkompetenzen zu betrachten sind, können dann als Grundlage für die Operationalisierung der Teilkompetenz dienen.

3.3.3. Abweichungen

Die im vorhergehenden Abschnitt beschriebenen Arten der Abweichung vom postulierten Modellierungszyklus äußerten sich durch den Wegfall eines oder mehrerer Einzelstadien. Auch wenn solche Abweichungen beobachtet werden konnten, so lässt sich doch festhalten, dass in den meisten Fällen die im Modellierungszyklus formulierten Stadien beobachtet werden können und dieser sich daher – wenn auch in leicht revidierter Form – als deskriptives Modell eignet. Dies gilt allerdings nur für die ersten vier Prozesse und nicht für die Teilkompetenz *Validieren*.

Diese Aussage betrifft jedoch zunächst lediglich die *Existenz* der Teilkompetenzen, nicht aber die *Reihenfolge* des Auftretens. Um dem Modellierungszyklus tatsächlich seine Angemessenheit zur Deskription von Modellierungsvorgängen zusprechen zu können, muss jedoch auch der Frage nachgegangen werden, inwieweit Modellierer die vorgegebene Reihenfolge tatsächlich einhalten.

Die Datenlage (Tabelle 11) zeigt, dass Abweichungen durch Überspringen von Stadien einerseits und durch rückwärtiges Beschreiten des Modellierungszyklus durchaus nicht selten auftreten, da sie in 42.5% aller Verläufe beobachtet werden konnten: Aus den in den Abbildungen 20 und 21 dargestellten Abläufen wird deutlich, dass solche Abweichungen in der äußeren Form an Rückkoppelungsschleifen erinnern, was gleichzeitig auch eine plausible Erklärung für das Auftreten solcher Abweichungen darstellt:

Modellierungsaufgaben lassen sich oftmals in Teilprobleme zerlegen, weshalb es durchaus sinnvoll sein kann, diese Teilprobleme nacheinander abzuarbeiten, also den Text zu *Strukturieren*, die relevanten Informationen für Teilproblem 1 zu suchen (S4a), sie zu *Mathematisieren* (M1), zu *Verarbeiten* (V1), dann die entsprechenden Informationen für Teilproblem 2 zu suchen und mit der Lösung von Teilproblem 1 in Verbindung zu bringen (S4a) , diese zu *Mathematisieren*, *Verarbeiten*, usw. Während ein Teil der Modellierer so vorzugehen scheint, durchlaufen andere den Modellierungszyklus tatsächlich strikt nach dem theoretischen Modell.

Festhalten lässt sich daher: Die Abfolge beim Modellieren kann dem Modellierungszyklus in seiner Ursprungsversion zwar folgen, allerdings wird dies nicht von allen Modellierern so umgesetzt. An dieser Stelle kommen anscheinend interindividuelle Unterschiede zum Vorschein, denen auch im deskriptiven Modell Rechnung getragen werden muss.

Es wird daher vorgeschlagen, das in Abbildung 23 revidierte deskriptive Modell noch um einen weiteren Aspekt an die aus der Empirie gewonnen Informationen anzupassen und wie folgt zu erweitern: Neben dem direkten Ablauf sind Verknüpfungen zwischen sämtlichen vier noch verbleibenden Stadien jederzeit möglich. Aufgrund dieser Befunde gilt es also, den Modellierungszyklus nunmehr als rekursives Modell aufzufassen (siehe Abbildung 24).

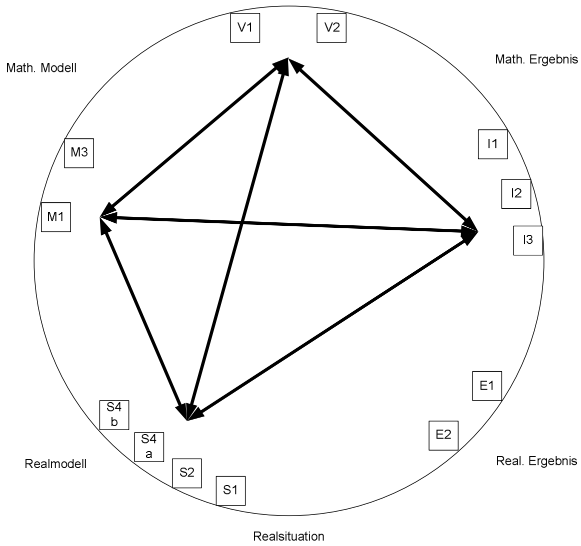


Abbildung 24: Deskriptives Modell des Modellierungszyklus

Dieses Modell enthält im Gegensatz zu dem in der Theorie postulierten keine Validierungsphase und erlaubt Rückkoppelungen zwischen allen Phasen. Sofern der Modellierungszyklus als deskriptives Modell eingesetzt werden soll, muss die bisherige Version durch die in Abbildung 24 dargestellte ersetzt werden.

3.4 Abschließende Bemerkungen

Zielsetzung der ersten Studie war es, eine Antwort auf die Frage zu finden, ob und inwieweit sich der in der Literatur berichtete Modellierungszyklus überhaupt als deskriptives Modell für die Stadien und Prozesse eignet, welche beim mathematischen Modellieren ablaufen. Hierbei wurde durch die Deskription der relevanten Prozesse beim mathematischen Modellieren eine Inventarisierung der Stadien des Modellierungszyklus realisiert (Jäger, 1983; Schmidtke, 1999).

Die Frage nach dieser Eignung kann mithilfe der in Kapitel 2 aufgeworfenen Fragen beantwortet werden:

1. Lassen sich die im Modellierungszyklus genannten Teilkompetenzen *Strukturieren*, *Mathematisieren*, *Verarbeiten*, *Interpretieren* und *Validieren* tatsächlich beobachten?

2. Welche der zu den jeweiligen Teilkompetenzen formulierten Unterkompetenzen lassen sich beobachten?

3. Treten die Teilkompetenzen in der im Modellierungszyklus postulierten Reihenfolge auf?

Zusammenfassend lässt sich festhalten:

1. Die ersten vier Teilkompetenzen lassen sich bei den allermeisten Verläufen beobachten, wodurch gerechtfertigt ist, dass diese in ein deskriptives Modell des Vorgehens beim Modellieren aufgenommen werden. Dies gilt allerdings nicht für die Teilkompetenz *Validieren*, welche in keinem einzigen Verlauf beobachtet werden konnte.

2. Auf der Ebene der Unterkompetenzen lassen sich vor allem diejenigen Kompetenzen identifizieren, welchen eine Schlüsselrolle für die Beherrschung der Teilkompetenz zugeschrieben werden kann. Es bietet sich daher an, sich dieser als Grundlage zur Operationalisierung der Teilkompetenz in Testitems zu bedienen. Wenig praktikabel erscheint hingegen die Berück-

sichtigung sämtlicher theoretisch postulierter Unterkompetenzen, da nicht alle häufig genug auftreten um eine Aufnahme dieser in das deskriptive Modell zu berücksichtigen.

3. In vielen beobachteten Verläufen ist die Reihenfolge der Einzelstadien als deutlich vom Idealmodell abweichend zu klassifizieren.

Aufgrund dieser Befunde bleibt daher festzuhalten, dass sich der postulierte Modellierungszyklus an sich nicht als deskriptives Modell eignet, da er zu viele Restriktionen enthält, von denen die empirisch gewonnenen Beobachtungen allzu deutlich abweichen. Für zukünftige Betrachtungen wird daher die in Abbildung 24 dargestellte Revision des deskriptiven Modells als Grundlage vorgeschlagen.

Nachdem nun die Frage nach der Angemessenheit des Modellierungszyklus als deskriptives Modell beantwortet ist, gilt es noch die Frage nach dem normativen Modell zu beantworten. Dies soll in Studie 2 geschehen, worin auch Möglichkeiten zur unabhängigen Erfassung der Teilkompetenzen erprobt werden sollen.

4. Studie 2: Hauptuntersuchung

Nachdem in der Voruntersuchung die im deskriptiven Model zu verortenden Teilkompetenzen identifiziert wurden, ist als nächstes eine Strategie zu ermitteln, wie diese erfasst werden können. Als Operationalisierung dienen Multiple-Choice-Items, deren Entwicklung in Abschnitt 2.1.4 beschrieben wurde. Neben der Erprobung von Items zur Erfassung der verschiedenen Teilkompetenzen ist eine zweite Zielsetzung der Hauptuntersuchung die Überprüfung der Annahme, dass der Modellierungszyklus als normatives Modell geeignet ist. Es ist also grundsätzlich die Hypothese zu prüfen, ob das Beherrschen der einzelnen Teilkompetenzen tatsächlich auch einen Einfluss auf die Modellierungskompetenz in ihrer Gesamtheit hat, also auf die Fähigkeit zum Lösen von kompletten Modellierungsaufgaben.

Im Folgenden werden zunächst Aufbau und Durchführung der Studie beschrieben und im Anschluss deren Ergebnisse dargestellt und diskutiert.

Die vorliegende Untersuchung soll letztlich als Basis dafür dienen, ein Instrument zur Diagnostik von Modellierungskompetenzen zu entwickeln. Dieses dürfte in einigen Teilen demjenigen Instrument ähneln, welches während der Hauptuntersuchung eingesetzt wurde, z.B. dadurch, dass es eine Teilmenge der Items enthalten dürfte und ähnlich einzusetzen ist. Um unterscheiden zu können, von welchem Verfahren die Rede ist, soll das in der Studie zum Einsatz kommende Material als Testverfahren zur Erfassung von Modellierungskompetenzen (TEMOD) 1 bezeichnet werden, das daraus resultierende endgültige Verfahren als TEMOD 2.

4.1. Methoden

Im Folgenden werden kurz die der Untersuchung zugrunde liegenden Materialen und die Art der Versuchsdurchführung beschrieben.

4.1.1. Material und Versuchsanordnung

Zu den vier in der Voruntersuchung beobachteten Teilkompetenzen wurden jeweils neun Multiple-Choice-Items entwickelt, bei denen jeweils vier Antwortalternativen vorgegeben waren. Die Aufgaben können als prototypische Aufgaben für Modellierungsaufgaben in der Primarstufe betrachtet werden. Da die Items nicht intuitiv entwickelt wurden, um dann allein empirisch zu ihren Referenzkonstrukt zugeordnet zu werden, sondern als direkte Operationalisierung der Definition der Teilkompetenzen betrachtet werden können, handelt es sich hierbei um eine rationale Itemkonstruktion (Lösel, 1999). Sämtliche Items sind zusammen mit ihren wichtigsten Kennwerten und Informationen über die Verteilung der Antworten über die Distraktoren in Abschnitt 4.2.1 dargestellt.

Da die Bearbeitung aller 36 Testitems für Schüler der vierten Klassen eine große Belastung dargestellt hätte, wurden die Items auf zwei Testhefte verteilt, die an zwei unterschiedlichen Tagen zu bearbeiten waren. Jedes Testheft enthielt dabei jeweils die Hälfte aller Items zu einer Teilkompetenz. Um zu vermeiden, dass einerseits Schüler zusammenarbeiten konnten, andererseits die Itemschwierigkeiten mit Ermüdungseffekten konfundiert würden, wurden von jedem Testheft zwei Versionen A und B erstellt, wobei bei Form B die Items in Vergleich zu Form A in umgekehrter Reihenfolge abgedruckt waren. Dadurch wurde eine Pseudoparallelform realisiert.

Um einen Zusammenhang zwischen der Beherrschung der einzelnen Teilkompetenzen und dem erfolgreichen Lösen von Modellierungsaufgaben messen zu können, wurden zudem in jedem Testheft drei Textaufgaben mit Modellierungscharakter vorgegeben. Vier der insgesamt sechs Aufgaben stammten aus dem DEMAT 4+ (Gölitz et al., 2006), zwei weitere waren Abwandlungen von Textaufgaben aus einem Unterrichtswerk für die Klassenstufe 4 (Rinkens & Hönisch, 2005).

Im Testheft für den ersten Testtag wurden zudem die Variablen Alter, Geschlecht, Muttersprache und die Mathematiknote in der letzten Klassenarbeit sowie im letzten Zeugnis erfragt.

Für die Lehrkräfte wurde analog zu den Unterkompetenzen ein Ratingverfahren konstruiert, bei dem diese für jeden Schüler einzuschätzen hatten, ob dieser „immer“, „oft“, „manchmal“, „selten“ oder „nie“ Schwierigkeiten beim Umsetzen der jeweiligen Unterkompetenz hat. Das Verfahren (Ratingverfahren zur Erfassung von Modellierungskompetenzen, kurz: REMOD) ist im Anhang wiedergegeben.

4.1.2 Durchführung der Untersuchung

Zur Akquise der benötigten Versuchspersonen (angestrebt war eine minimale Stichprobe von N=200 Personen) wurden 185 Grundschulen in Rheinland-Pfalz angeschrieben und um ihre Teilnahme bei der Studie gebeten.

Von allen angeschriebenen Schulen meldeten sich 16 Klassen für die Testung an. Mit jeder Klasse wurden zwei Termine für die jeweiligen Testtage vereinbart. Die Testdurchführung erfolgte vor Ort an den Schulen durch einen von drei geschulten Testleitern. Um die Bedingungskonstanz möglichst hoch zu halten, mussten sich die Testleiter an eine detaillierte Instruktion halten (diese ist im Anhang wiedergegeben).

Nach der Testung wurden die Daten in eine Datenbank eingepflegt und im Anschluss daran wurde aufgrund von Vergleichen von individuellen Ergebnissen mit Gesamtmittelwerten für jeden Schüler eine individuelle Rückmeldung über sein Abschneiden bezüglich der Teilkompetenzen erstellt und an die Klassen verschickt.

Die Zuordnung der beiden Testtage, sowie die Zuordnung der Rückmeldungen zu den Schülern erfolgte anonym über die Klassenbuchnummer.

Als eine Schwierigkeit bei der Stichprobenziehung ist sicherlich der Aspekt der Selbstselektion in Betracht zu ziehen. Es muss davon ausgegangen werden, dass sich gerade besonders engagierte Schulen eher zur freiwilligen Teilnahme bereit erklärten als andere. Für die Verteilung der Kennwerte ist daher wohl mit leichten Deckeneffekten zu rechnen. Für die erste Erprobung von Items ist eine Verschiebung des Mittelwertes als unproblematisch zu betrachten, solange eine ausreichende Varianz der Daten gegeben ist. Für eine eventuelle spätere Normierung ist es allerdings von zentraler Bedeutung, echte Zufallsstichproben zu gewinnen.

4.1.3 Beschreibung der Stichprobe

Insgesamt nahmen N=325 Personen an der Untersuchung teil, wodurch eine hinreichend große Stichprobe für die nachfolgend beschriebenen statistischen Analysen gegeben ist.

Entsprechend dem für das Ende der vierten Klassenstufe üblichen Alter waren die untersuchten Schüler im Mittel zehn Jahre alt, mit einer Standardabweichung von etwa einem halben Jahr (M=9.90, SD=.49). Die Altersverteilung (siehe Abbildung 25) reichte von einem Minimum (N=1) von acht Jahren bis zu einem Maximum (ebenfalls N=1) von zwölf Jahren. Mehr als drei Viertel der Schüler (76.9%) waren zum Zeitpunkt der Testung zehn Jahre alt. Von zwei Personen lagen keine Angaben zum Alter vor.

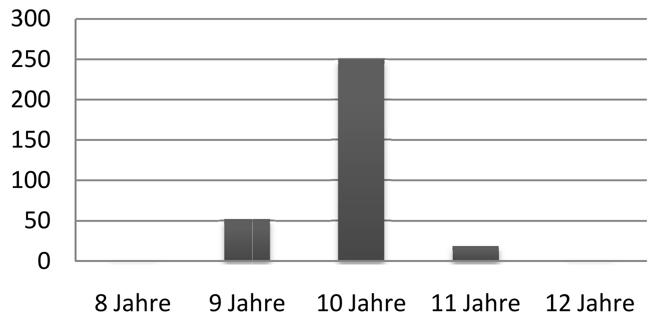


Abbildung 25: Altersverteilung in Studie 2

Von den befragten Personen waren 155 weiblich (47.7%) und 167 männlich (51.4%), drei Schüler machten keine Angaben zu Ihrem Geschlecht. Bezüglich des Geschlechts kann also von einer annähernden Gleichverteilung ausgegangen werden.

Als weitere soziodemographische Variable wurde die Muttersprache erhoben. Nur zehn von 325 Kindern gaben an, eine andere Muttersprache als Deutsch zu haben, dies entspricht einem Prozentsatz von lediglich 3.1% (2 Personen hatten keine Angaben zu ihrer Muttersprache gemacht). Vergleicht man diese Zahl mit der vom statistischen Bundesamt veröffentlichten Zahl von 10.6% Ausländern an deutschen Grundschulen (destatis, 2008), so wird deutlich, dass Ausländerkinder in der vorliegenden Stichprobe deutlich unterrepräsentiert sind.

Dies lässt sich zum einen dadurch begründen, dass sämtliche teilnehmenden Schulen aus ländlichen Gegenden stammten und zum anderen durch die bereits erwähnte Selbstselektion der Schulen. Es könnte z.B. vermutet werden, dass Schulen mit höherem Ausländeranteil oder Schulen, die in sozialen Brennpunkten verortet sind, mit geringerer Wahrscheinlichkeit freiwillig an Untersuchungen wie der vorliegenden teilzunehmen wünschen.

Die Muttersprache der Teilnehmer wurde unter anderem deshalb erhoben, weil ausgeschlossen werden sollte, dass eventuell auftretendes mangelndes Textverständnis auf ein mangelndes Beherrschen der deutschen Sprache zurückzuführen ist. Allerdings zeigte sich, dass die Subgruppe der Ausländerkinder sich in ihrer Testleistung – gemessen am Gesamtscore - nicht signifikant von der restlichen Stichprobe unterschied ($t=.182$, $df=321$, $p=.862$), weshalb diese Teilgruppe auch nicht von weiteren Analysen ausgeschlossen werden muss.

Bezüglich der Größe der jeweiligen Schulklassen kann die Anzahl der Tests pro untersuchter Klasse aber nur schlecht als Indikator für die tatsächliche Klassengröße herangezogen werden, da hierbei lediglich die Zahl der teilnehmenden Schüler berücksichtigt würde. Sie sagt nichts über die Gesamtzahl der Schüler in einer Klasse aus. Vor der Untersuchung wurde die Einverständniserklärung der Eltern eingeholt. Lag diese nicht vor, so nahm der betreffende Schüler nicht an der Testung teil. Daher wurde ein anderer Zugang zur näherungsweisen Schätzung der Klassengröße herangezogen: Von jedem untersuchten Schüler lag für die Anonymisierung der Daten lediglich die Klassenbuchnummer vor. Für jede der teilnehmenden Klassen wurde die jeweils größte Klassenbuchnummer identifiziert; diese Zahl repräsentiert im Folgenden die geschätzte Klassengröße. Die Variable verteilt sich wie folgt über die untersuchten 16 Klassen: Im Mittel umfassten die Klassen 23 Schüler, bei einer Standardabweichung von 4 Schülern ($M=22.54$, $SD=3.79$). In der kleinsten Klasse waren 16, in der größten 32 Schüler.

4.1.4 Auswertung der Daten

Die Analyse der Items in Bezug auf Schwierigkeiten, Verteilungen, Reliabilität und Kriteriumsvalidität wurden mithilfe von SPSS 16 durchgeführt.

Zur Untersuchung der Konstruktvalidität des Verfahrens wurde die Methode der konfirmatorischen Faktorenanalyse mit MPLUS (Version 4.1) verwendet. Faktorenanalysen werden hierfür sinnvollerweise eingesetzt, da auf diesem Wege überprüft werden kann, wie hoch die einzelnen Indikatoritems mit ihren dazugehörigen latenten Konstrukten korrelieren (Brown, 2006). Da im vorliegenden Fall eine genaue Theorie darüber vorliegt, welche Items zu welchen Faktoren gehören, ist die konfirmatorische Faktorenanalyse (CFA) angezeigt (Moosbrugger & Schermelleh-Engel, 2007).

Im Gegensatz zur explorativen Faktorenanalyse, bei der keine Vorstellung über die Zusammenhänge zwischen den Variablen vonnöten ist, muss bei der CFA a priori spezifiziert werden:

- Wie viele Faktoren man erwartet;
- Welches Item auf welchem Faktor laden soll;
- Welcher Zusammenhang zwischen Items und Faktoren zu erwarten ist und
- Welche Beziehungen zwischen den Faktoren zu erwarten sind.

Auf der Grundlage der theoretischen Ausführungen (siehe Kapitel 1.3 sowie Kapitel 2.1.2 können all diese Fragen beantwortet werden, womit der Einsatz der CFA gerechtfertigt ist (Heck, 1998):

- Entsprechend der Konstruktion des Testverfahrens sind vier Faktoren zu erwarten: *Strukturieren*, *Mathematisieren*, *Verarbeiten* und *Interpretieren*.
- Jedes Item lädt auf demjenigen Faktor, zu dessen Messung es konstruiert wurde.
- Zwischen Item und Faktor ist daher jeweils eine Korrelation von mindestens λ=.30 zu verlangen, da man ab diesem Punkt von einem mittelgroßen Effekt spricht. Ideal wären darüber hinaus große Effekte von λ>.50. Da die Items jeweils nur eine Teilkompetenz messen sollen und dürfen, werden keine Nebenladungen auf anderen Konstrukten neben dem Referenzfaktor zugelassen. Items die solche Nebenladungen aufweisen, sind zu selektieren.

- Da alle vier Faktoren Teilaspekte der Modellierungskompetenz sind, sind hohe Korrelationen zwischen diesen zu erwarten und werden dementsprechend im Modell zugelassen.

Zur Beantwortung der Frage nach dem Modellierungszyklus als normatives Modell wird auf die Regressionsanalyse zurückgegriffen, bei der die Modellierungskompetenz (gemessen durch die Skala Textaufgaben) als abhängige Variable zu erklären ist. Sie wird durch die vier Prädiktoren *Strukturieren*, *Mathematisieren*, *Verarbeiten* und *Interpretieren* vorhergesagt.

Allerdings ist die einfache Regressionsanalyse bei Klumpenstichproben nicht ohne weiteres bedenkenlos einsetzbar. Vielmehr gibt es mehrere Gründe, die dafür sprechen, in Clustern organisierte Daten mehrebenenanalytisch auszuwerten (Schwetz & Subramanian, 2005):

Erstens kann so der Tatsache Rechnung getragen werden, dass keine echte Zufallsstichprobe vorliegt, sondern eine Klumpenstichprobe.

Zweitens finden Unterricht und Lernen – auch in Bezug auf die Modellierungskompetenz – in Schulklassen statt. Schüler einer Klasse lernen unter ähnlichen Vorrausetzungen und Bedingungen. Diesen innerhalb einer Klasse gleichen, zwischen Klassen hingegen variierenden Kontextbedingungen muss Rechnung getragen werden.

Sofern eine Stichprobe de facto eine Clusterstruktur aufweist, dieser aber bei der statistischen Auswertung der Daten nicht Rechnung getragen wird, kommt es zu einer fehlerhaften Ermittlung von Standardfehlern (Snijders & Bosker, 1999). Übliche Verfahren wie Regressions- oder Varianzanalyse sind bei solchen Daten nicht angemessen, da bei diesen von der Unabhängigkeit der Beobachtungen (zu sehen in unkorrelierten Fehlervarianzen) als Vorausetzung ausgegangen wird. Da sich die Individuen in den Klumpen jedoch ähnlicher sind als zwischen den Klumpen, kann eben diese Unabhängigkeit nicht angenommen werden, sondern es muss von Autokorrelationen ausgegangen werden.

Bei jeder Regressionsanalyse werden zwei wichtige Parameter geschätzt, mit denen sich die Regressionsgerade eindeutig definieren lässt. Dies ist zum einen der y-Achsenabschnitt, also derjenige y-Wert bei dem die Regressionsgerade die y-Achse (bei x=0) schneidet. Dieser auch als Intercept bezeichnete Wert stellt den Mittelwert der abhängigen Variablen dar. Der zweite Parameter ist die Steigung der Regressionsgeraden (oder Slope), die verdeutlicht, wie stark der Einfluss der jeweiligen Prädiktorvariablen auf das Kriterium ist.

Während die einfache Regression jeweils nur den Mittelwert über alle Klassen für Intercept und Slope schätzt, geht die Mehrebenenanalyse davon aus, dass diese Parameter zwischen Klassen variieren können (Ditton, 1998).

Im sog. Random-Intercept-Modell wird für jede Klasse ein eigener y-Achsen-Abschnitt geschätzt, im Random-Slope-Modell darüber hinaus sogar für jede Klasse eine eigene Steigung, so dass letztlich für jede Klasse eine eigene Regressionsgerade geschätzt wird (Goldstein, 1999).

Damit die Regressionskonstante sinnvoll interpretiert werden kann, müssen die Prädiktorvariablen am Gesamtmittelwert zentriert werden. Hierfür wird lediglich von jedem beobachteten Wert der Mittelwert der jeweiligen Prädiktorvariable subtrahiert, so dass sich ein neuer Mittelwert von Null ergibt. Da es sich hierbei um eine Lineartransformation handelt, hat dies keine weiteren Auswirkungen auf die zu schätzenden Parameter.

Die Regressionsanalyse zur Überprüfung des Modellierungszyklus als normatives Modell wurde aus oben genannten Gründen mehrebenanalytisch mithilfe von MLwiN Version 2.02 realsiert.

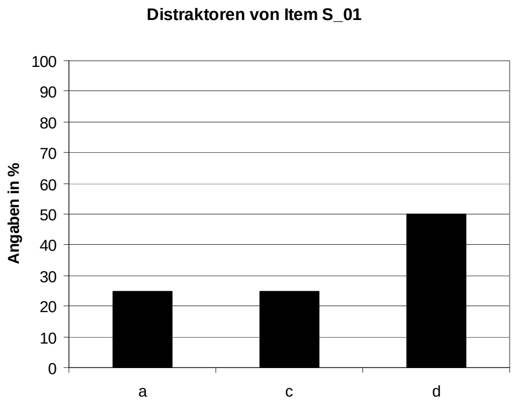
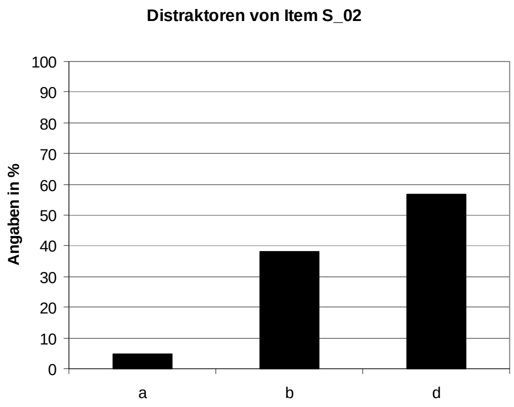
4.2. Ergebnisse

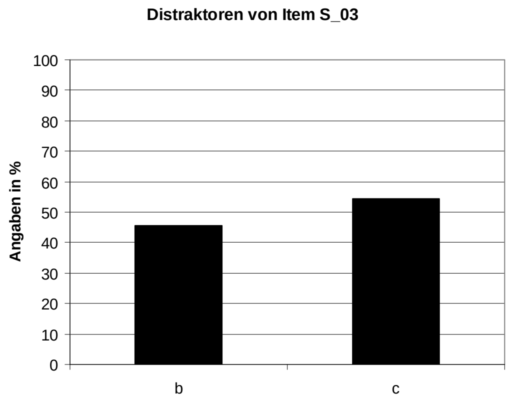
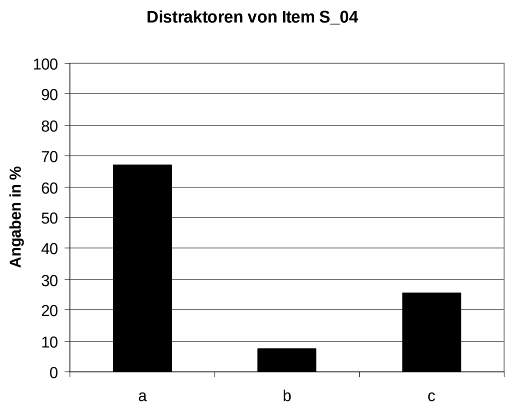
Im Folgenden werden die Ergebnisse der Hauptuntersuchung dargestellt, wobei zunächst auf die Analyse der verwendeten Items eingegangen wird und auf die Frage, welche dieser Items sich als Grundlage für ein Testverfahren eignen. Mit dem gleichen Hintergrund wird im Anschluss auf die Gütekriterien des TEMOD 1 eingegangen, welche sowohl die Rechtfertigung von Schlüssen aus den Daten der Hauptuntersuchung liefern als auch eine Basis für die Gütekriterien des TEMOD 2 darstellen. Letztlich wird die Hypothese überprüft, ob der Modellierungszyklus sich als normatives Modell eignet.

4.2.1 Analyse der Items

Im Folgenden sollen nun die in der Hauptuntersuchung verwendeten Items zusammen mit ihren wichtigsten Eigenschaften dargestellt werden. Für jede Teilkompetenz werden die neun Items präsentiert, sowie jeweils ihre Schwierigkeit und die Faktorladung (siehe Abschnitt 4.2.2.3.2) als Indikator für die Trennschärfe. Um die Güte eines Multiple-Choice-Items zu bewerten sind darüber hinaus die Distraktoren zu betrachten – im Idealfall verteilen sich die falsch Antwortenden gleichmäßig über die Distraktoren, denn dann sind diese alle gleich attraktiv. Daher wird die Attraktivität der Distraktoren jeweils in einem Diagramm dargestellt.

4.2.1.1 Strukturieren

Item Nr. S_01 Lena und ihre Oma Irene gehen in den Zoo. Da Lena gerne in den Streichelzoo geht, kaufen sie an der Kasse gleich zwei Packungen Tierfutter. Im Zoorestaurant isst Lena ein Eis, während ihre Oma eine Bratwurst bestellt. Welchen Satz hat die Oma beim Zoobesuch wohl tatsächlich gesagt? a) „Guten Tag, zwei Eintrittskarten für Kinder bitte.“ **b) „Guten Tag, zwei Packungen Tierfutter bitte.“** c) „Guten Tag, zwei Eis bitte.“ d) „Guten Tag, zwei Portionen Bratwurst mit Sauerkraut bitte.“	Distraktoren:  Itemschwierigkeit: .889 Faktorladung: .659
Item Nr. S_02 Frau Weresch spart für ein neues Auto. 6512 € hat sie bereits gespart. Doch nun braucht sie einen neuen Fernseher, der 390 € kostet, die Waschmaschine muss für 210 € repariert werden und ihr Mann schenkt ihr einen Diamantring für 800 €. Für neue Kleider gibt sie 411 € aus. Für welchen Gegenstand hat Frau Weresch das meiste Geld ausgegeben? a) Fernseher b) Diamantring **c) Kleidung** d) Auto	Distraktoren  Itemschwierigkeit: .332 Faktorladung: .421

Item Nr. S_03 Ein kleines Schweinchen ist bei seiner Mutter ausgezogen und möchte sich nun ein eigenes Haus aus Stroh bauen. Dafür bestellt es beim Bauern 100 Strohballen. Der Bauer liefert wie versprochen am nächsten Tag das Stroh. Allerdings bringt er statt 100 nur 95 Strohballen. Das kleine Schweinchen merkt aber nicht, dass es betrogen wurde, denn es ist zu faul um nachzuzählen. Was ist passiert? **a) Der Bauer hat zuwenig Stroh geliefert.** b) Der Bauer hat zuviel Stroh geliefert. c) Das Schweinchen hat sich verzählt. d) Der Bauer hat kein Stroh gebracht.	Distraktoren:  Itemschwierigkeit: .932 Faktorladung: .995
Item Nr. S_04 Lars ist 5 Jahre alt, seine Schwester Lisa ist doppelt so alt. Lars kann nur halb so schnell laufen wie Lisa. Lars braucht morgens 5 Minuten um in den Kindergarten zu laufen. Wie lange braucht Lisa, um in die Schule zu laufen? a) Halb so lang. b) Gleich lang. c) Doppelt so lang. **d) Das kann man nicht sagen.**	Distraktoren:  Itemschwierigkeit: .354 Faktorladung:.480

Item Nr. S_05

Maren wiegt die Zutaten für einen Kuchen ab:

500 g Mehl
190 g Zucker
170 g Butter
160 g Schokolade

Sie verrührt die Zutaten und backt den Teig dann 60 Minuten lang bei 180° C.

Welche Zahl hat Marens Küchenwaage heute nicht angezeigt?

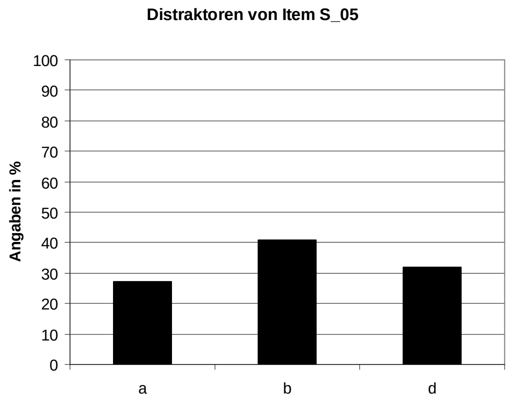
a) 160

b) 170

c) 180

d) 190

Distraktoren:



Itemschwierigkeit: .895

Faktorladung: .742

Item Nr. S_06

Das Herz ist der Motor unseres Körpers. Es pumpt das Blut regelmäßig hindurch. Das Herz eines Kindes schlägt etwa hundertmal pro Minute (also sechstausendmal in jeder Stunde), bei Erwachsenen schlägt es 3600 mal pro Stunde.

Bei wem schlagen die Herzen schneller?

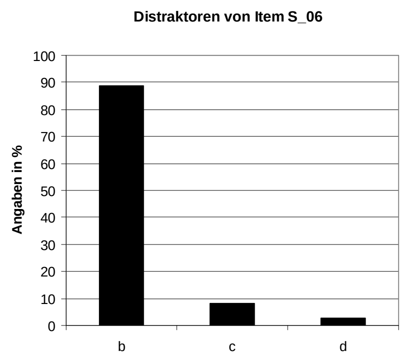
a) Kinder

b) Erwachsene

c) Beide gleich

d) Kann man nicht sagen

Distraktoren:



Itemschwierigkeit: .889

Faktorladung: -

<table>
<tr>
<td>

Item Nr. S_07

Charly hat zwei Meerschweinchen, Hummel und Pünktchen. Hummel frisst in zwei Tagen einen Beutel Futter, Pünktchen in drei Tagen.

Wer frisst das meiste Futter?

a) Charly

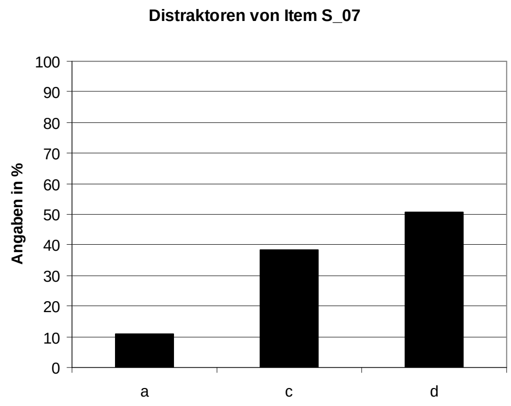
b) Hummel

c) Pünktchen

d) Alle fressen gleich viel.

</td>
<td>

Distraktoren



Itemschwierigkeit: .797

Faktorladung: .229

</td>
</tr>
<tr>
<td>

Item Nr. S_08

Anna, Lisa, Paula und Tina laufen um die Wette. Die Lehrerin notiert die Platzierungen. Lisa braucht die meiste Zeit, Anna war die Schnellste. Tina war schneller als Paula. Wie sieht der Notizzettel der Lehrerin aus?

a) **1. Anna**
2. Tina
3. Paula
4. Lisa

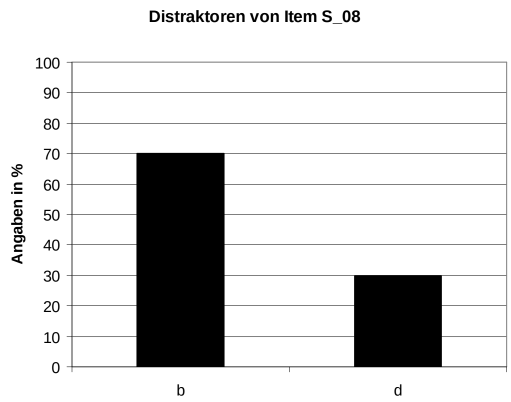
b) 1. Lisa
2. Anna
3. Tina
4. Paula

c) 1. Lisa
2. Anna
3. Paula
4. Tina

d) 1. Tina
2. Paula
3. Lisa
4. Anna

</td>
<td>

Distraktoren:



Itemschwierigkeit: .960

Faktorladung: .634

</td>
</tr>
</table>

Item Nr. S_09	Distraktoren:
Delia und Manuela gehen jeden Morgen zusammen in den Kinder- garten. Delia läuft zuerst zu Manuela, dann gehen sie zusammen in den Kindergarten. Wer muss morgens zuerst sein Haus verlassen? a) Manuela **b) Delia** c) Beide müssen gleich früh aus dem Haus. d) Das kann man nicht sagen.	 Itemschwierigkeit: .914 Faktorladung: .273

Mit Blick auf eine endgültige Zusammenstellung von Items für ein künftiges Testverfahren sind mehrere Kriterien zur Beurteilung heranzuziehen. Zunächst sollte die Itemschwierigkeit nicht zu hoch oder zu niedrig sein. In beiden Fällen ist davon auszugehen, dass beinahe alle Probanden ein Item lösen bzw. nicht lösen, es kann daher nicht gut zwischen Probanden unterschiedlicher Fähigkeitsniveaus differenzieren und ist aus statistischer Sicht wertlos. Allerdings können sehr leichte oder sehr schwere Items aus anderen Gründen in den Test integriert werden, z.B. (bei leichten Items) um auch sehr schwache Probanden zu motivieren.

Für das TEMOD 2 Verfahren sollen allerdings nur solche Items verwendet werden, deren Schwierigkeit p zwischen 10% und 90% liegt.

Als weiteres Kriterium ist die Ladung des Items auf den jeweiligen Faktor (in diesem Fall *Strukturieren*) zu betrachten (Eine ausführliche Darstellung der Faktorenanalysen über die vier Teilkompetenzen erfolgt in Kapitel 4.2.2.3.2). Diese kann ab einem Wert von .30 als akzeptabel betrachtet werden, da man standardisierte Ladungskoeffizienten als Korrelationen interpretieren und bei einer Ladung ab .30 von einem mittelstarken Effekt ausgegangen werden kann.

Als drittes Kriterium sind die Distraktoren zu Rate zu ziehen. Die Verteilung der falsch Antwortenden über die Distraktoren sollte ungefähr gleich sein, als Kriterium wird daher definiert, dass idealerweise kein Distraktor von weniger als 10% der falsch Antwortenden gewählt wurde.

Tabelle 12 gibt nun für die Teilkompetenz *Strukturieren* und die dazugehörigen Items wieder, inwieweit diese Kriterien erfüllt sind oder nicht.

Tabelle 12: Items für Strukturieren

Item	$0.1 < p < 0.9$	$\Lambda > .30$	Dis.
S1	✓	✓	✓
S2	✓	✓	
S3		✓	
S4	✓	✓	
S5	✓	✓	✓
S6	✓		
S7	✓		✓
S8		✓	
S9	✓		✓

Leider ist Tabelle 12 zu entnehmen, dass lediglich die Items S1 und S5 alle drei Kriterien erfüllen. Item S1 lädt sehr hoch auf dem Referenzfaktor, ist allerdings ein vergleichsweise einfaches Item, wie der Schwierigkeitsindex deutlich macht. Ähnliches gilt für Item S5.

Vernachlässigt man das am wenigsten bedeutsame Kriterium, nämlich dasjenige bezüglich der Distraktoren, so eignen sich darüber hinaus auch die Items S2 und S4 zur Aufnahme in das TEMOD 2 Verfahren. Bei Item S2 ist Antwortalternative a) problematisch, die nur selten gewählt wurde. Dies verwundert wenig, da es darum geht, den höchsten Geldbetrag zu identifizieren und a) dem Fernseher der geringste Geldbetrag von allen vier Antwortalternativen zuzuordnen ist. Die Itemschwierigkeit zeigt, dass lediglich ein Drittel aller Probanden die richtige Lösung finden konnte. Dies lässt sich recht eindeutig auf das mangelnde Verständnis des Aufgabentextes zurückführen, der sich bei genauem Lesen deutlich erschließen sollte. Auch Item S4 konnte lediglich von einem guten Drittel der Probanden gelöst werden. Mangelnde Sorgfalt beim Lesen des Aufgabentextes lassen die Probanden übersehen, dass Lars und Lisa nicht den gleichen Weg gehen, man daher also keine Aussagen über die jeweilige Dauer machen kann. Da unter der Voraussetzung, dass beide den gleichen Weg gehen, allerdings Antwortalternative a) richtig wäre, wird diese auch von einem weitaus größeren Anteil von Probanden gewählt als die anderen Distraktoren. Dies ist insofern naheliegend, als dass diese noch weiter von der richtigen Lösung entfernt sind als Antwort a).

Alle weiteren Items sind aufgrund der beiden ersten Kriterien für den Einsatz als Testitems ungeeignet: S3 ist zum einen zu leicht, zum anderen ist der Distraktor d) anscheinend so offensichtlich falsch, dass er überhaupt keine Attraktivität besitzt. Item S6 fällt komplett weg, da dieses bei der Faktorenanalyse (siehe 4.2.2.3.2) überhaupt nicht mit dem Konstrukt in Verbindung gebracht werden konnte. Bei S7 ist die Ladung mit .229 zu gering und bei S8 ist zum einen die Schwierigkeit zu hoch, zum anderen fällt auch hier ein Distraktor komplett weg. S9 hingegen scheitert sowohl daran, dass es zu leicht ist, als auch an der zu geringen Ladung auf dem Referenzfaktor.

4.2.1.2 Mathematisieren

<table>
<tr>
<td>
Item Nr. M_01

Wie kann man die Antwort zu dieser Aufgabe ausrechnen?

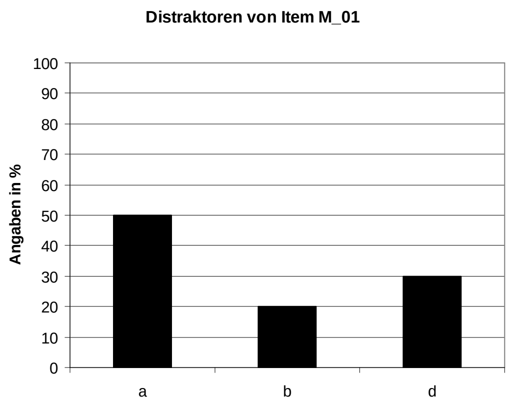
„Bei der Geburt ist ein Blauwalbaby ungefähr drei Tonnen schwer. Nach zwei Jahren ist das Blauwalbaby neunmal so schwer wie bei seiner Geburt. Wie viel wiegt es?“

a) 3 + 9 =

b) 3 + 2 =

c) 3 · 9 =

d) 3 · 2 =
</td>
<td>
Distraktoren:



Itemschwierigkeit: .859

Faktorladung: .535
</td>
</tr>
<tr>
<td>
Item Nr. M_02

Familie Klöbner möchte ein Haus kaufen. Das Haus kostet dreihunderttausendzehn Euro. Herr Klöbner sagt: „Von meiner Mutter bekommen wir dreiundvierzigtausendachthundert Euro geschenkt!“

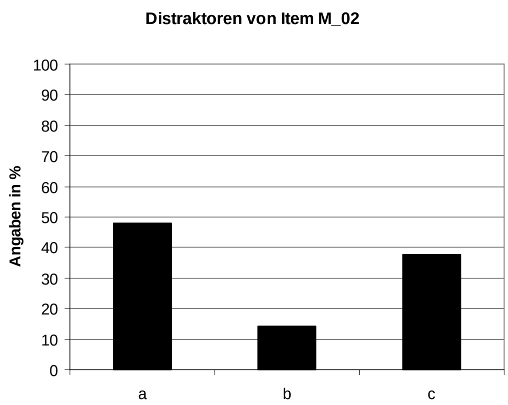
Wie können die Klöbners ausrechnen, wieviel sie selbst noch bezahlen müssen?

a) 310000 – 43800

b) 31000 : 4380

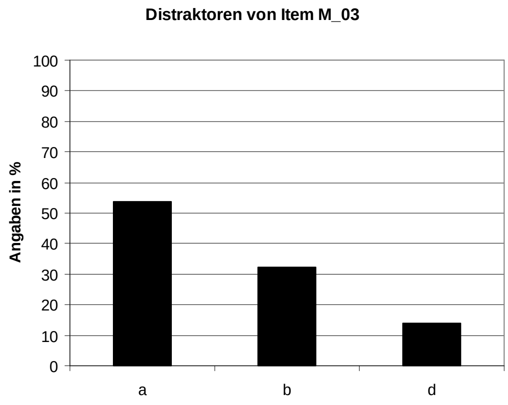
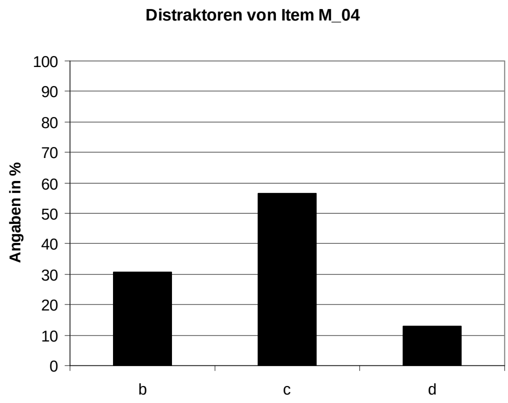
c) 300010 : 4380

d) 300010 – 43800
</td>
<td>
Distraktoren:



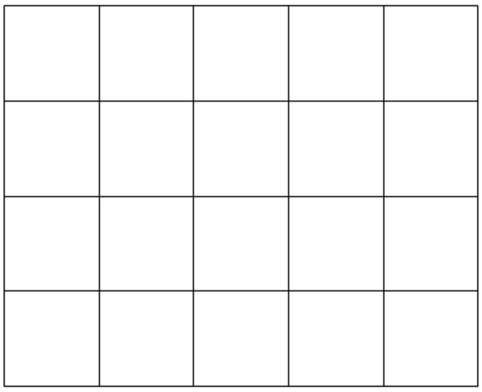
Itemschwierigkeit: .739

Faktorladung: .559
</td>
</tr>
</table>

Item Nr. M_03 Der Kilometerzähler von Frau Gronewald zeigt 86 386 Kilometer. Die nächste Inspektion ist bei hunderttausend Kilometer fällig. Frau Gronewald rechnet aus, wie weit sie noch fahren kann. a) 100000 + 86386 = b) 10000 – 86386 = **c) 100000 – 86386 =** d) 1000000 – 86386	Distraktoren:  Itemschwierigkeit: .640 Faktorladung: . 369
Item Nr. M_04 Vor einer Baustelle auf der Autobahn A 65 ist ein 3 km langer Stau auf der rechten Spur. Jedes Auto braucht ca. 4m Platz. Frau Meier vom Verkehrsamt möchte abschätzen, wie viele Autos ungefähr in diesem Stau stehen. Wie muss sie rechnen? **a) 3000 : 4 =** b) 3 : 4 = c) 3000 - 4 = d) 300 – 4 =	Distraktoren:  Itemschwierigkeit: .868 Faktorladung: .460

Item Nr. M_05

Herr Ackermann besitzt ein Waldstück (20 Flächeneinheiten groß). Er möchte den Wald in vier gleich große Teile aufteilen und an seine Söhne verschenken.



Wie kann er ausrechnen, wie groß jeder Teil ist?

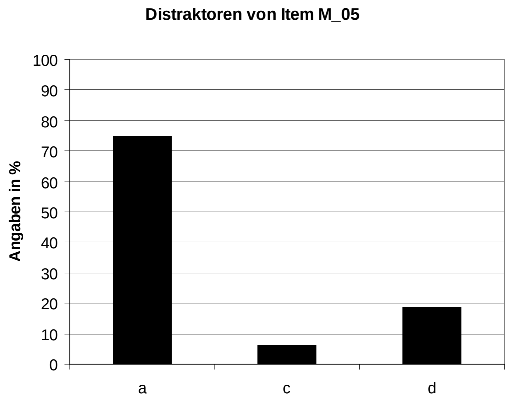
a) 20 – 5 =

b) 20 : 4 =

c) 20 : 3 =

d) 20 – 5 =

Distraktoren:



Itemschwierigkeit: .911

Faktorladung: .875

Item Nr. M_06

Lisa hat zu ihrem Kindergeburtstag drei Kinder eingeladen, sie selbst und ihre beiden Geschwister wollen natürlich auch an der Feier teilnehmen. Lisas Mama hat 30 Muffins gebacken. Lisa möchte nun ausrechnen, wie viele Muffins jeder essen kann. Wie soll sie rechnen?

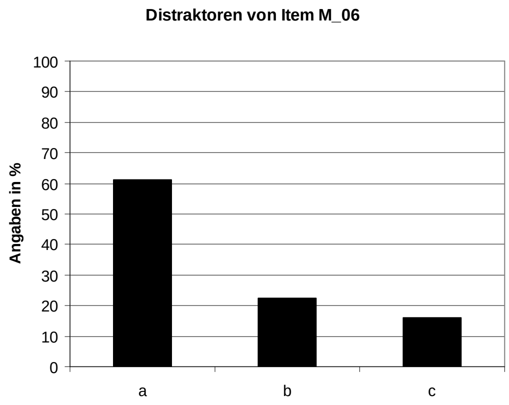
a) 30 - 3 =

b) 30 - 6 =

c) 3 + 2 + 1 =

d) 30 : 6 =

Distraktoren:



Itemschwierigkeit: .865

Faktorladung: .787

Item Nr. M_07

Frau Müller kauft Weihnachtsgeschenke für ihre fünf Enkel. Für jeden gibt sie 20 € aus. Wie muss Frau Müller rechnen, um herauszufinden, wie viel Geld sie insgesamt ausgegeben hat?

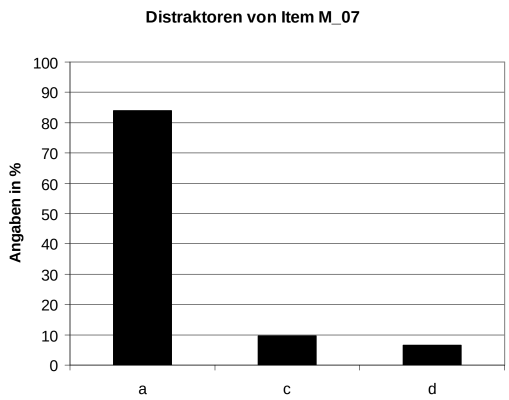
a) 20 : 5 =

b) 20 · 5 =

c) 20 + 5 =

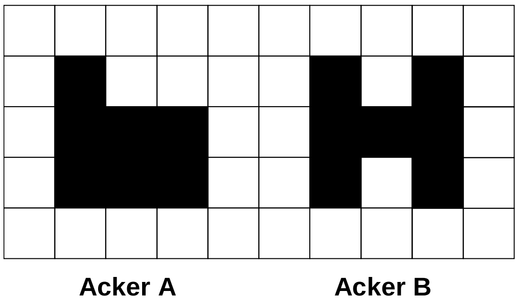
d) 20 – 5 =

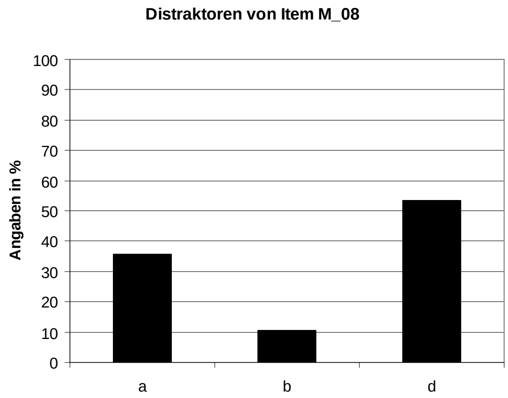
Distraktoren:

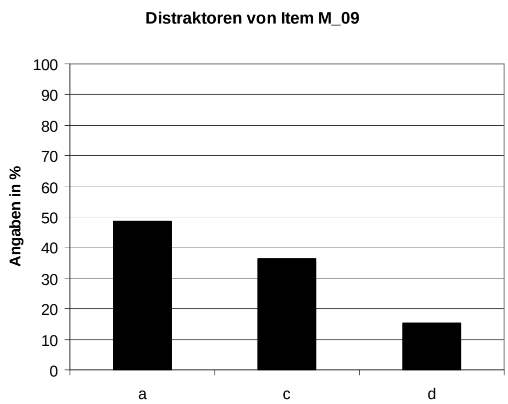


Itemschwierigkeit: .868

Faktorladung: .754

<table>
<tr>
<td>
Item Nr. M_08


Bauer Schmitt besitzt einen Acker (Acker A) und kauft nun einen zweiten (Acker B) hinzu. Wie kann er ausrechnen, wie viele Flächen Ackerland er insgesamt besitzt?
a) 5 + 9 =
b) 6 + 8 =
c) 7 + 7 =
d) 8 + 7 =
</td>
<td>
Distraktoren:


Itemschwierigkeit: .871
Faktorladung: .748
</td>
</tr>
<tr>
<td>
Item Nr. M_09
Herr Tietze kauft einen Strauß Rosen für seine Frau. Er hat 21 Euro dabei, jede Rose kostet 3 Euro. Herr Tietze fragt sich nun, ob sein Geld für mehr als 4 Rosen ausreicht. Hilf ihm auszurechnen, wie viele Rosen er kaufen kann.
a) 21 · 4 =
b) 21 : 3 =
c) 21 – 4 =
d) 21 + 3
</td>
<td>
Distraktoren:


Itemschwierigkeit: .868
Faktorladung: .803
</td>
</tr>
</table>

Die Items zu *Mathematisieren* werden analog zu den Strukturierungsitems beurteilt. Tabelle 13 zeigt, dass fast alle Items (mit Ausnahme von M5 und M7) die geforderten Kriterien erfüllen.

Tabelle 13: Items zu Mathematisieren

Item	0.1 < p < 0.9	Λ > .30	Dis.
M1	✓	✓	✓
M2	✓	✓	✓
M3	✓	✓	✓
M4	✓	✓	✓
M5		✓	
M6	✓	✓	✓
M7	✓	✓	
M8	✓	✓	✓
M9	✓	✓	✓

Bei Item M5 ist zunächst einmal die Schwierigkeit ein Ausschlusskriterium, darüber hinaus ist Distraktor a) wesentlich attraktiver als die beiden anderen. Es wäre zu untersuchen, ob sich die Schwierigkeit des Items erhöhen ließe, wenn es ohne die beigefügte Skizze dargeboten würde. M7 scheitert am Kriterium der Distraktoren: über 80% der falsch Antwortenden wählen Antwort a). Dies könnte daran liegen, dass die Division der Multiplikation näher ist als der Addition oder der Subtraktion. Wenn der Aufgabentext beim Modellierer dann zunächst das Schema Multiplikation/Division aktiviert, und im nächsten Schritt erst genauer zwischen beiden Operationen differenziert wird, so liegt nahe, dass die Wahrscheinlichkeit Distraktor a) zu wählen höher ist als die für die Distraktoren c) oder d).

Alle anderen Items können grundsätzlich weiterhin eingesetzt werden, wobei allerdings auffällt, dass alle Items tendenziell eher leicht sind.

4.2.1.3 Verarbeiten

<table>
<tr>
<td>Item Nr. V_01

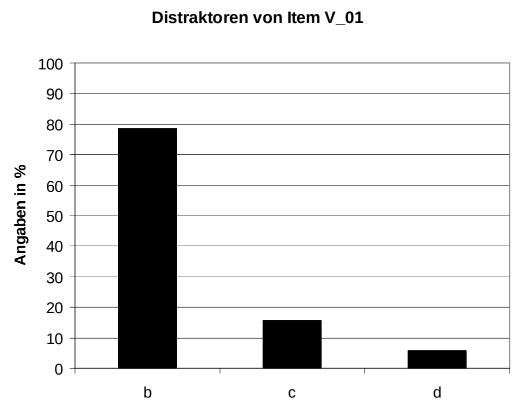
21400 : 50 =

a) 428

b) 4280

c) 4028

d) 402</td>
<td>Distraktoren:



Itemschwierigkeit: .612

Faktorladung: .436</td>
</tr>
<tr>
<td>Item Nr. V_02

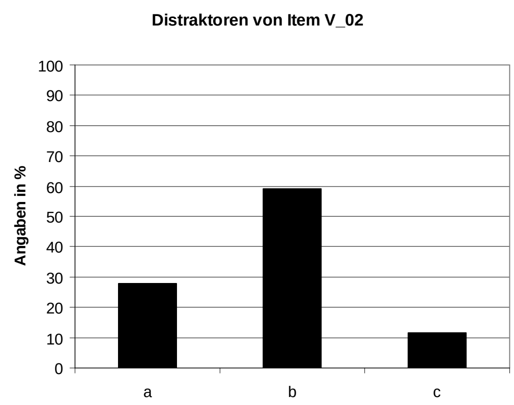
798 · 320 =

a) 3990

b) 25536

c) 39900

d) 255360</td>
<td>Distraktoren:



Itemschwierigkeit: .714

Faktorladung: .460</td>
</tr>
</table>

Item Nr. V_03

```
  10001
 -
  5555
```

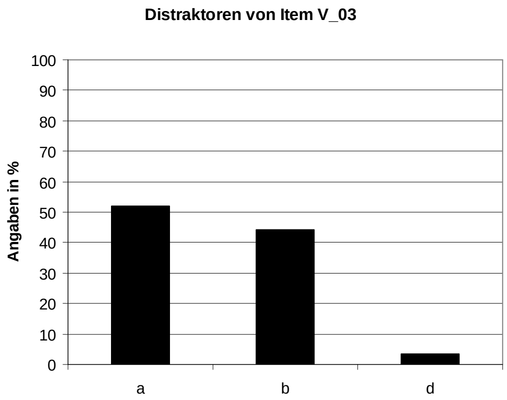
a) 5556

b) 15556

c) 4446

d) 14446

Distraktoren:



Itemschwierigkeit: .643

Faktorladung: .453

Item Nr. V_04

```
  4957
 +
 7266
```

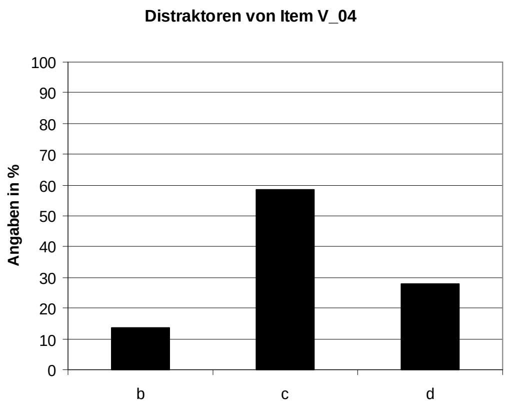
a) 2309

b) 3711

c) 12223

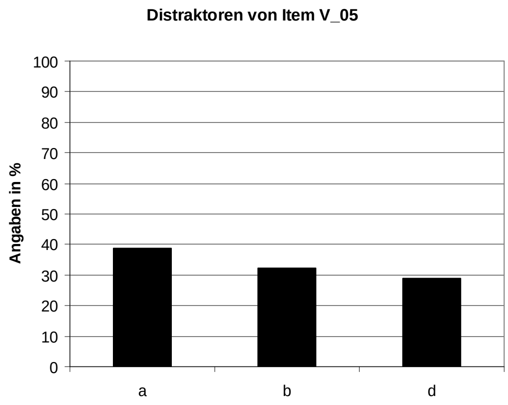
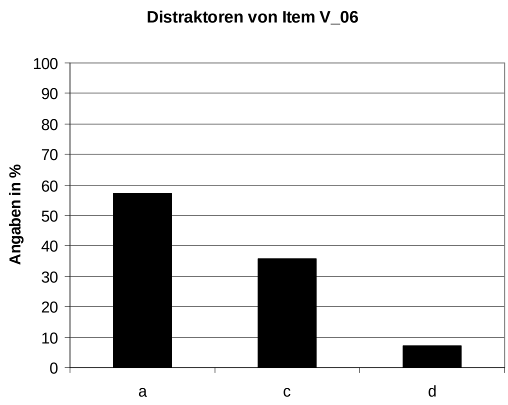
d) 2319

Distraktoren:



Itemschwierigkeit: .662

Faktorladung: .220

Item Nr. V_05 931 : 7 = a) 131 b) 132 **c) 133** d) 134	Distraktoren:  Itemschwierigkeit: .889 Faktorladung: .653
Item Nr. V_06 2555 : 5 = a) 1214 **b) 511** c) 411 d) 51	Distraktoren:  Itemschwierigkeit: .914 Faktorladung: .915

<table>
<tr>
<td>Item Nr. V_07

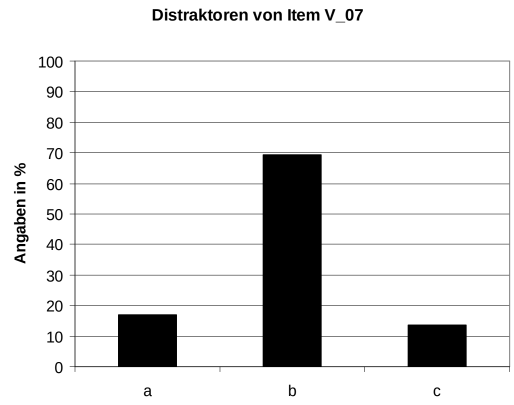
30205 · 25 =

a) 759705

b) 211435

c) 754005

d) 755125</td>
<td>Distraktoren:



Itemschwierigkeit: .782

Faktorladung: .718</td>
</tr>
<tr>
<td>Item Nr. V_08

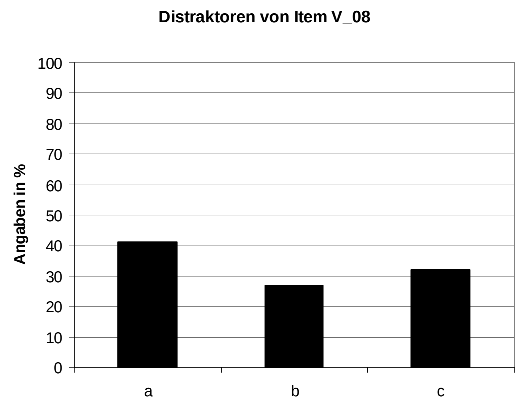
7634
-3565

a) 4079

b) 11199

c) 4131

d) 4069</td>
<td>Distraktoren:



Itemschwierigkeit: .794

Faktorladung: .618</td>
</tr>
</table>

Item Nr. V_09

```
  1066
 +5574
```

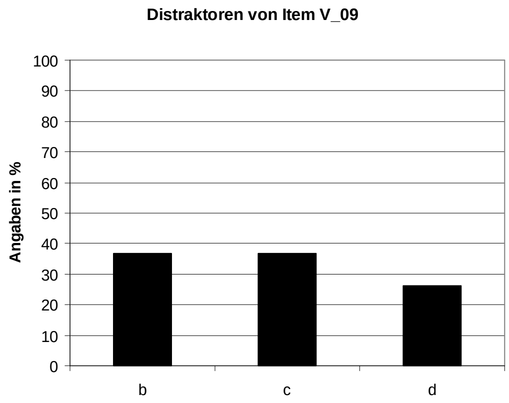
a) 6640

b) 6610

c) 6630

d) 6530

Distraktoren:



Itemschwierigkeit: .910

Faktorladung: .846

Bei der Teilkompetenz *Verarbeiten* müssen aufgrund der drei Kriterien die vier Items V1, V3, V4 und V6 wegfallen (siehe Tabelle 14). Da es sich bei den Itemstämmen nur um rein innermathematische Rechenaufgaben handelt und bei den Antwortalternativen nur um unterschiedliche Zahlenwerte, ist es bei dieser Teilkompetenz weitaus schwieriger, Erklärungen für die Nichterfüllung der Kriterien zu finden. Sämtliche Distraktoren wurden konstruiert, indem der Problemlösealgorithmus durchgeführt und an verschiedenen Stellen Fehler eingebaut wurden (z.B. Weglassen der „1" in der nächsten Spalte bei Addition über den Zehner, verrutschen in der Zeile, etc.). Die daraus resultierenden unterschiedlichen Verteilungen über die Distraktoren (die bei V1, V3 und V6 so gravierend waren, dass die Items selektiert werden müssen) zeigen dann allerdings auf, dass einige Fehler bei den Probanden anscheinend häufiger gemacht werden als andere.

Tabelle 14: Items zu Verarbeiten

Item	$0.1 < p < 0.9$	$\Lambda > .30$	Dis.
V1	✓	✓	
V2	✓	✓	✓
V3	✓	✓	
V4	✓		✓
V5	✓	✓	✓
V6		✓	
V7	✓	✓	✓
V8	✓	✓	✓
V9	✓	✓	✓

4.3.1.4 Interpretieren

<table>
<tr>
<td>Item Nr. I_01

Miriam soll die Geschwindigkeit eines Buses ausrechnen. Sie weiß wie lange der Bus braucht um eine Strecke von 120 Kilometern zu fahren, nämlich genau 2 Stunden. Sie hat die Aufgabe schon gerechnet:

120 km : 2 h = 60

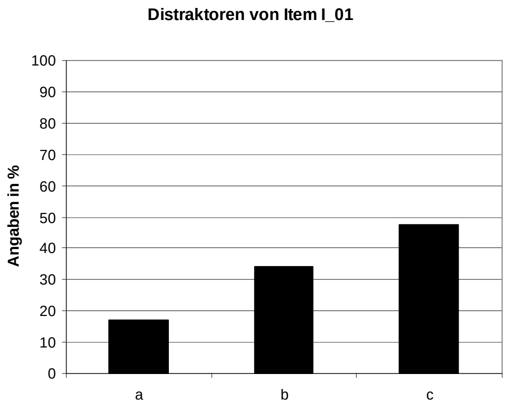
Was soll Miriam nun als Antwortsatz schreiben?

a) Der Bus braucht 60 Stunden.

b) Der Bus fährt 60 Kilometer.

c) Der Bus braucht 60 Minuten.

d) Der Bus fährt 60 Kilometer pro Stunde.</td>
<td>Distraktoren:



Itemschwierigkeit: .723

Faktorladung: .653</td>
</tr>
<tr>
<td>Item Nr. I_02

Die Klasse 4b hat diese Aufgabe als Hausaufgabe auf:

Familie Meier (Herr Meier, Frau Meier und die beiden Kinder) essen eine Pizza mit 18 Stücken. Wie viele Stücke kriegt jeder?

Laura hat gerechnet:

18 : 4 = 4 R 2

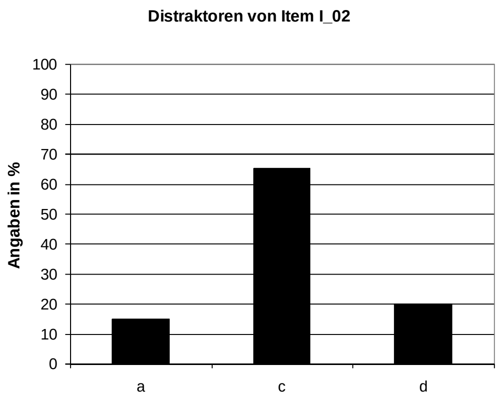
Sie überlegt nun, was sie als Antwortsatz schreiben soll. Vier Möglichkeiten fallen ihr ein. Eine ist allerdings falsch. Findest Du sie?

a) Jeder kriegt vier Pizzastücke.

b) Jeder kriegt fünf Pizzastücke.

c) Jeder kriegt vier Pizzastücke und zwei bleiben übrig.

d) Jeder kriegt vier Pizzastücke, aber es bleibt etwas übrig.</td>
<td>Distraktoren:



Itemschwierigkeit: .640

Faktorladung: .608</td>
</tr>
</table>

Item Nr. I_03

Frau Kleinmüller misst jeden Montag ihre kleine Tochter und schreibt auf, wie viel sie gewachsen ist.

Diese Woche hat Frau Kleinmüller in den Kalender gekritzelt:

1,02 + 3 = 1,05

Die Tochter ist im letzten Monat also

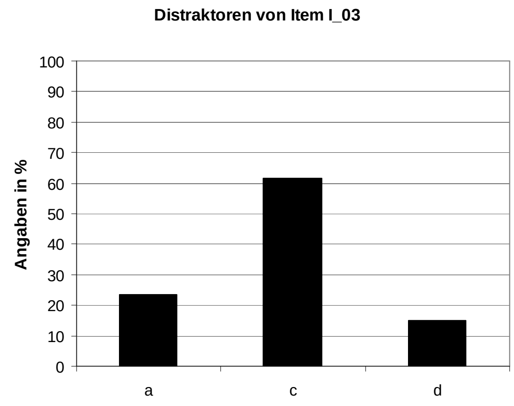
a) 3 m gewachsen

b) 3 cm gewachsen

c) 1,05 cm gewachsen

d) 1,05 m gewachsen

Distraktoren:



Itemschwierigkeit: .806

Faktorladung: 660

Item Nr. I_04

Monika zählt ihr Geld. Sie hat 21 € gespart. Nun rechnet sie:

21€ - 3€ + 5€ = 19€

Was könnte sie sich beim Rechnen gedacht haben?

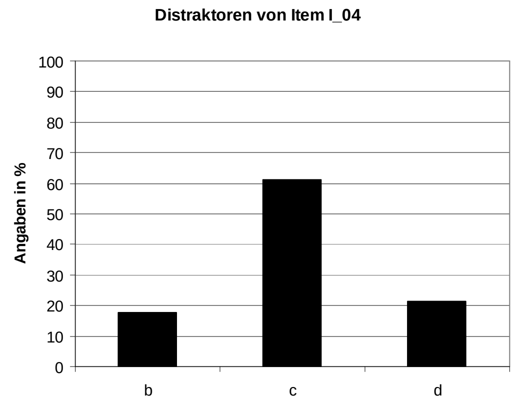
a) Von Oma Lina bekomme ich noch 5€ Taschengeld, am Montag kaufe ich für 3€ Bonbons, dann bleiben mir noch 19€.

b) Wenn meine Schwester mir noch 3€ leiht und mein Bruder mir 5€ leiht, dann habe ich 19€.

c) Nächste Woche kaufe ich für 5€ Schokolade, 3€ schuldet mir mein Freund Stefan noch, dann habe ich 19€.

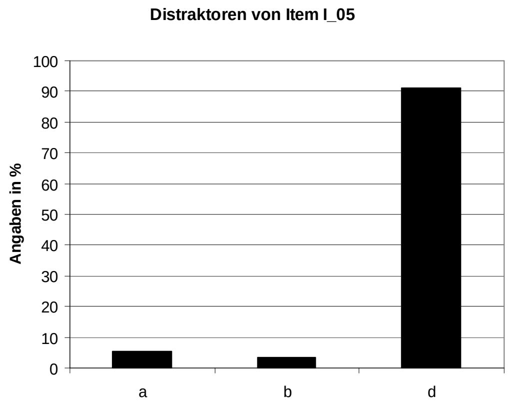
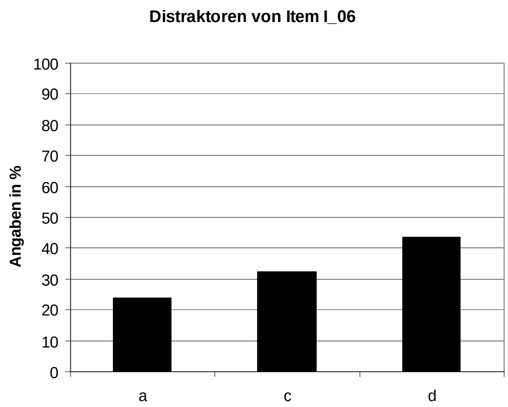
d) Wenn ich noch irgendwo 3€ auftreiben kann, dann habe ich zusammen mit meinen 5€ Taschengeld insgesamt 19€ gespart.

Distraktoren:



Itemschwierigkeit: .622

Faktorladung: .422

Item Nr. I_05 Paulas Katze hat 5 Junge bekommen. Paulas Freundinnen Anne und Liese wollen die Kätzchen untereinander aufteilen. Paula rechnet 5:2 = 2 R 1 Nun schreibt sie Ihren Freundinnen eine SMS: a) Jeder kriegt etwas mehr als 2 Kätzchen. b) Jeder kriegt 3 Kätzchen. **c) Jeder kriegt 2 Kätzchen.** d) Jeder kriegt 2 Rest 1 Kätzchen.	Distraktoren:  Itemschwierigkeit: .440 Faktorladung: .220
Item Nr. I_06 Mutter strickt einen Wollschal. Nach 5 Tagen ist der Schal bereits einen halben Meter lang. Sie rechnet ½ m : 5 = 10 cm Was hat Mutter ausgerechnet? a) Wie lange sie für einen 10 cm langen Schal braucht. **b) Wie viele cm sie durchschnittlich jeden Tag gestrickt hat.** c) Wie lange sie noch stricken muss, bis der Schal fertig ist. d) Wie viele cm sie schon gestrickt hat.	Distraktoren:  Itemschwierigkeit: .622 Faktorladung: .539

Item Nr. I_07

Tim kauft 4 Tierbücher zu je 20€ und ein Kinderlexikon. Insgesamt bezahlt er 119€.

Er rechnet nun den Preis des Lexikons aus:

$4 \cdot 20 = 80$

$119 - 80 = 39$

Was sagt Tim zu seiner Mutter:

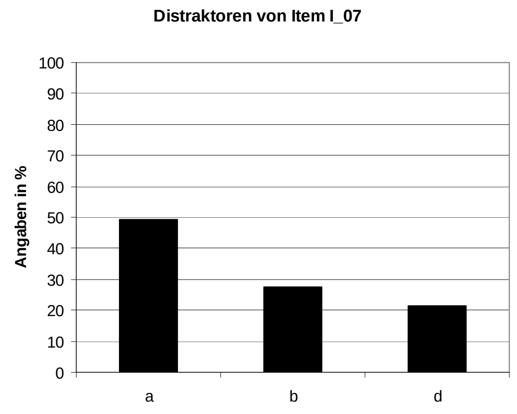
a) Insgesamt habe ich 80€ bezahlt.

b) Jedes Buch kostet 39€.

c) Das Kinderlexikon kostet 39€.

d) Alle Bücher haben gleich viel gekostet.

Distraktoren:



Itemschwierigkeit: .791

Faktorladung: 582

Item I_08

Eine Aufgabe in einer Klassenarbeit lautet:

Für einen Kuchen braucht man zwei Päckchen Butter zu je 250g, 570g Mehl und 300g Zucker. Was wiegen alle Zutaten?

Antje rechnet:

$2 \cdot 250 = \underline{500}$

$500 + 570 + 300 = \underline{\underline{1370}}$

Was schreibt sie als Antwortsatz?

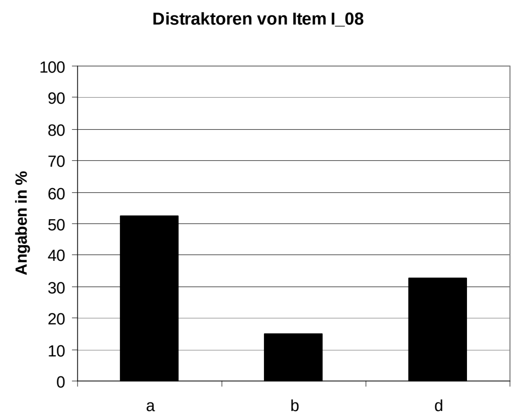
a) Alle Zutaten wiegen 13kg und 70g.

b) Die Zutaten wiegen 500g.

c) Alle Zutaten wiegen 1kg und 370g.

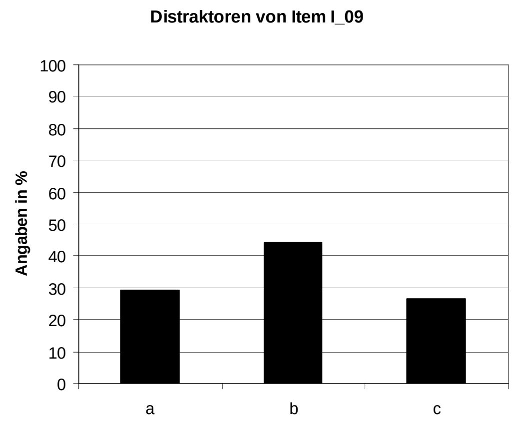
d) Alles zusammen wiegt 1570 Kilo.

Distraktoren:



Itemschwierigkeit: .671

Faktorladung: .500

Item Nr. I_09	Distraktoren:
Christoph hat ausgerechnet, was sein Vater wöchentlich an Taschengeld für ihn und seine Schwester ausgibt: 2 · 3€ = 6€ Er schreibt auf: a) Wenn jedes Kind 2 Euro kriegt, sind es insgesamt 6. b) 3 Kinder brauchen zusammen 6€. c) Jedes der beiden Kinder bekommt 6€. **d) Wenn jedes der beiden Kinder 3€ bekommt, braucht Papa jede Woche 6€.**	 Itemschwierigkeit: .883 Faktorladung: .521

Wie Tabelle 15 zu entnehmen ist, eignen sich die Interpretationsitems mit Ausnahme von I5 alle zur Aufnahme in ein Testverfahren. Einerseits fiel es einem Großteil der Probanden schwer, die richtige Antwort zu I5 zu identifizieren, andererseits entschieden sich über 90% der falsch Antwortenden für Distraktor d). Außerdem ist die Faktorladung mit .220 zu gering, um das Item als Indikator für die Teilkompetenz *Interpretieren* aufzufassen.

Tabelle 15: Items zu Interpretieren

Item	$0.1 < p < 0.9$	$\Lambda > .30$	Dis.
I1	✓	✓	✓
I2	✓	✓	✓
I3	✓	✓	✓
I4	✓	✓	✓
I5	✓		
I6	✓	✓	✓
I7	✓	✓	✓
I8	✓	✓	✓
I9	✓	✓	✓

Bildet man aus den jeweils verbleibenden Items Skalen, so zeigt sich aufgrund der Tatsache, dass der Schwierigkeitsindex bei allen verwendeten Items vergleichsweise hoch ist, bei drei von vier Skalen eine linksschiefe Verteilung der Summenscores über die einzelnen Unterskalen (siehe Abbildungen 26 bis 29).

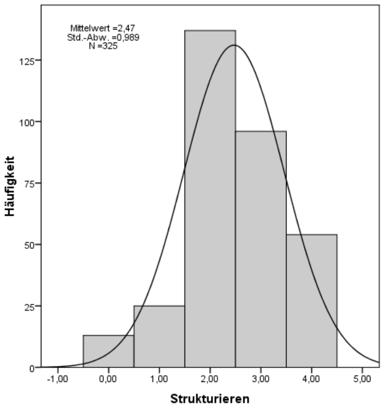


Abbildung 26: Skala Strukturieren

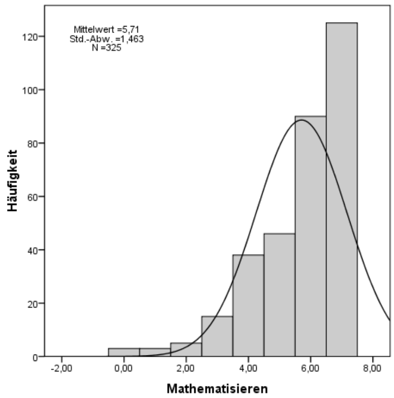


Abbildung 27: Skala Mathematisieren

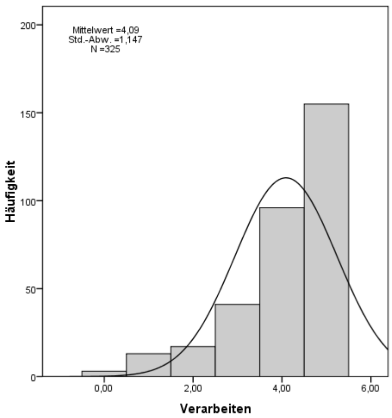


Abbildung 28: Skala Verarbeiten

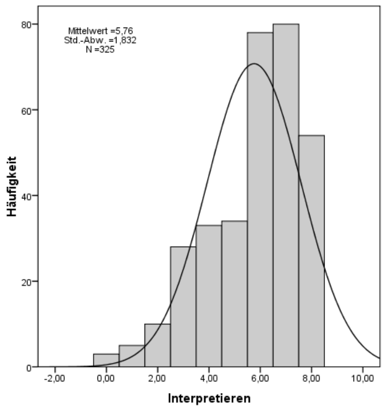


Abbildung 29: Skala Interpretieren

4.2.2 Gütekriterien des Testverfahrens

Sollen die in der vorliegenden Untersuchung erprobten Items langfristig mit der Zielsetzung eingesetzt werden, Modellierungskompetenzen zu diagnostizieren, so müssen diese Items und das aus ihnen konstruierte Testverfahren bestimmten Gütekriterien genügen.

Eine weitere mit Studie 2 zu beantwortenden Fragen ist die, ob die Untersuchungsmethode ausreichend objektiv, reliabel und valide ist, um die Grundlage für eine förderungsorientierte Diagnostik zu bilden.

Im Folgenden soll daher die Testform, welche in der Hauptuntersuchung eingesetzt wurde, auf die psychometrischen Hauptgütekriterien überprüft werden. Zu beachten gilt, dass ein endgültiges Testverfahren letztlich nicht alle Items enthalten wird, sondern lediglich eine Selektion, wie dies oben bei der Diskussion der Itemcharakteristika bereits angedeutet wurde. Insofern kann sich die Prüfung der Gütekriterien auch nicht auf den endgültigen Test beziehen, vielmehr wird dies – im Rahmen seiner Normierung – eine erneute Überprüfung erforderlich machen. Dennoch können die hier berichteten Kennwerte durchaus als Anhaltspunkte betrachtet werden. Oder anders gesagt: Sofern sich die Gütekriterien als ausreichend gut herausstellen, können sie als Basis für die Güte des späteren Tests betrachtet werden. Sofern sie sich aber nicht als hinreichend herausstellen, ist gar nicht erst eine Berechtigung dafür gegeben, mit den Items weiter zu arbeiten.

4.2.2.1 Objektivität

Vergleicht man den hier gewählten Ansatz mit anderen Untersuchungsmethoden zu mathematischer Modellierungskompetenz (siehe Abschnitt 1.6.2), so stellt sich die Wahl eines standardisierten Multiple-Choice-Tests zur Individualdiagnostik ganz offensichtlich als objektivste aller bisher diskutierten Methoden dar. Auswertungs- und Interpretationsobjektivität können aufgrund der klar definierten Lösungen und Skalenzugehörigkeiten ohne weitere Prüfung als gegeben betrachtet werden.

Der einzige Punkt, an dem sich dem Testleiter überhaupt die Möglichkeit bietet, sich nicht objektiv zu verhalten, liegt bei der Instruktion der Testpersonen. Wenn Testleiter sich z.B. nicht manualkonform verhalten, dann betrifft dies den Aspekt der Durchführungsobjektivität. Diese kann hier relativ leicht empirisch überprüft werden. Sofern sich alle Testleiter gleich eng an den Vorgaben orientieren, sollten sich für die Klassen, welche von Testleiter A instruiert und beaufsichtigt wurden, keine signifikanten Leistungsunterschiede zu Klassen ergeben, die von Testleiter B oder C instruiert und beaufsichtigt wurden. Verletzt hingegen ein Testleiter beispielsweise das Gebot, keine Hilfestellungen zu geben, so sollten „seine“ Klassen signifikant besser abschneiden, als andere Klassen, die keine Hilfestellungen erhalten haben.

Zur Überprüfung der Durchführungsobjektivität wurde eine einfaktorielle Varianzanalyse (ANOVA) gerechnet, mit „Testleiter“ als unabhängiger und dem Gesamtscore des Tests als abhängiger Variable. Da bei zwei Schuklassen mehr als ein Testleiter anwesend war, wurden diese beiden Klassen aus der Analyse ausgeschlossen, da sie nicht eindeutig einer der beiden Personen zugeordnet werden konnten. Somit verblieb für die ANOVA eine Stichprobe von N=293 Personen.

Tabelle 16: ANOVA zur Überprüfung der Objektivität

	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen	18.179	2	9.090	.286	.752
Innerhalb der Gruppen	9227.241	290	31.818		
Gesamt	9245.420	292			

Wie erwartet ergaben sich mit F(292)=.286, p=.752 keinerlei signifikante Unterschiede zwischen den verschiedenen Testleitern (siehe Tabelle 16). Die Objektivität des Verfahrens kann daher als gegeben betrachtet werden. Dies ist besonders insofern von Bedeutung, als dass die Objektivität eine notwendige Bedingung für Reliabilität und diese wiederum eine notwendige Bedingung für die Validität darstellt.

4.2.2.2 Reliabilität

Wie zuverlässig sind die in Studie 2 zum Einsatz gekommen Items? Analog zum Vorgehen bei Studie 1 wurde auch hier die Reliabilität über die innere Konsistenz des gesamten Verfahrens bestimmt.

Da es sich ebenfalls um dichotome Daten handelt, kommt die Kuder-Richardson-Formel (KR-20) erneut zum Tragen. Auf der Basis aller 36 Items ergibt sich ein Wert von α=.845.

4.2.2.3 Validität

Gerade bei der Entwicklung neuer Verfahren zur Untersuchung eines Konstrukts spielt die Überprüfung der Validität eine tragende Rolle.

Aus diesem Grunde sollen mehrere Methoden zur Beurteilung des Validitätsaspektes herangezogen werden.

4.2.2.3.1 Kriteriumsvalidität

Für einen neu zu entwickelnden Test liegt es in der Natur der Sache, dass die Messung der Kriteriumsvalidität problembehaftet ist. Gerade aus der Tatsache, dass es keine objektiven, reliablen und validen Methoden gibt, um Modellierungskompetenzen unabhängig voneinander zu erfassen, ergibt sich schließlich überhaupt die Rechtfertigung für die Konstruktion des Testverfahrens. An welchen Außenkriterien lässt sich der Test dann validieren? Hierzu wurde auf drei Kriterien zurückgegriffen: Schulnoten, Kompetenz zum Lösen von Textaufgaben sowie Beurteilung der Modellierungskompetenz durch den Mathematiklehrer.

Zunächst zu den Schulnoten. Ein Zusammenhang zwischen Schulnote und Testleistung ist aus folgendem Grund zu erwarten: Die Schulnote im Fach Mathematik soll – so zumindest ihr Anspruch – eine Aussage über den Ausprägungsgrad der mathematischen Kompetenz eines Schülers machen. Modellierungskompetenz ist aber ein bedeutender Teilaspekt der mathematischen Kompetenz, weshalb der Gesamtwert im Test in der Lage sein sollte, einen Teil der Varianz in den Noten zu erklären.

Die erfragten Schulnoten streuten bei der Frage nach der „Note in der letzten Klassenarbeit“ von *sehr gut* bis *ungenügend* (siehe Abbildung 30); bei der Frage nach der „Note im letzten Zeugnis“ lagen die Werte zwischen *sehr gut* und *mangelhaft* (siehe Abbildung 30). In der letzten Klassenarbeit lag die Note im Mittel bei M=2.44 (SD=.988), im letzten Zeugnis lag der Mittelwert bei M=2.29 (SD=.885). Beide Variablen verteilen sich leicht rechtsschief.

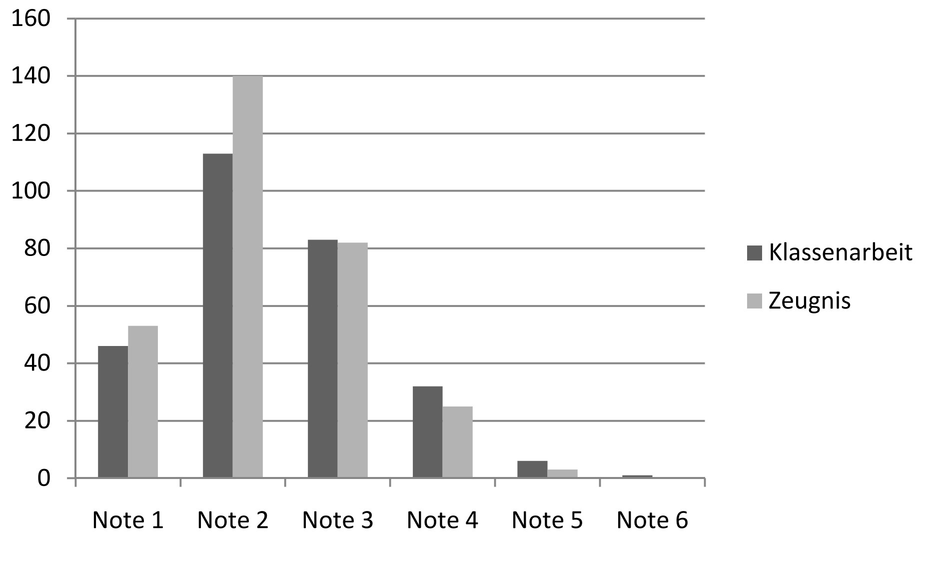


Abbildung 30: Notenverteilung über die Vp in Studie 2

Die Note in der letzten Klassenarbeit korrelierte mit der letzten Zeugnisnote zu .628 (p<.000), worin sich einerseits die relative Stabilität der mathematischen Kompetenz, andererseits aber auch die Möglichkeit zu Veränderungen spiegelt. Die Höhe der Korrelation entspricht einem großen Effekt.

Korreliert man die Note in der letzten Klassenarbeit mit dem Gesamtscore, so ergibt sich ein mittelstarker Effekt bei einer Korrelation -.470 (p<.000). Zu der Note im letzten Zeugnis ergibt sich ein Zusammenhang von -.499 (p<.000). Die höhere Korrelation bei der Zeugnisnote könnte eventuell dadurch erklärt werden, dass diese ein breiteres Spektrum an Kompetenzen erfasst als eine einzige Klassenarbeit, in der in aller Regel nur ein oder zwei Themenkomplexe abgefragt werden. Wenn also angenommen werden kann, dass das Konstrukt Modellierungskompetenz auf die Zeugnisnote mehr Einfluss hat als auf die Klassenarbeitsnote, so ist schlüssig, dass in der Zeugnisnote auch mehr Varianz durch die Modellierungskompetenz aufgeklärt werden kann.

Schulnoten beinhalten – betrachtet man sie als Indikator für mathematische Kompetenz – aber neben der Modellierungskompetenz noch viele andere Facetten. Zur Validierung des Verfahrens ist es daher angezeigt, auf Leistungsindikatoren zurückzugreifen, welche näher am zu untersuchenden Konstrukt liegen. Dies ist bei Textaufgaben der Fall, bei denen alle vier Teilkompetenzen des Modellierens realisiert werden müssen, um zu einer erfolgreichen Lösung zu gelangen. Wie oben bereits diskutiert liegt die Schwäche solcher Aufgaben darin, dass die Teilkompetenzen nicht unabhängig voneinander betrachtet werden können. Dennoch sollte sich ein substanzieller Zusammenhang zeigen zwischen der Modellierungskompetenz, gemessen durch den Gesamtscore, und der Fähigkeit zum Lösen von Textaufgaben, gemessen durch die Anzahl der richtig gelösten Textaufgaben. Von den sechs Textaufgaben wurden im Mittel 4 gelöst (M=4.035, SD=1.691), die Verteilung ist damit leicht linksschief (s. Abbildung 31).

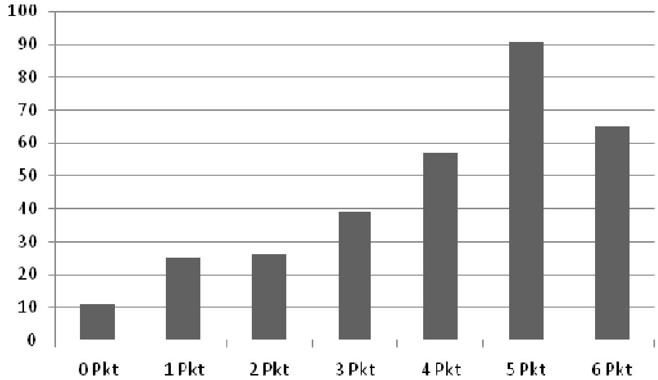


Abbildung 31: Verteilung über die Lösungspunkte in den Textaufgaben

Die Korrelation zwischen Textaufgaben- und dem Gesamtscore beim Modellieren weist mit $r = .546$ ($p<.000$) einen starken Effekt auf.

Als dritte Variable zur Kriteriumsvalidierung wurde die Einschätzung der Modellierungskompetenz durch den Mathematiklehrer mit dem REMOD-Verfahren (siehe Anhang) hinzugezogen. Bildet man aus sämtlichen elf Items zur Erfassung der Modellierungskompetenz eine Skala, so erweist sich diese mit einer interne Konsistenz von $\alpha=.972$ als höchst reliabel.

Bei elf Variablen liegt das theoretische Minimum der Skala bei elf Punkten (für den Fall, dass für einen Schüler bei jeder Unterkompetenz angekreuzt wurde „dem Schüler fällt es *nie* schwer, …"), das theoretische Maximum bei 55 Punkten (für den Fall, dass bei allen Variablen angekreuzt wurde „dem Schüler fällt es *immer* schwer, …"). Im Mittel erreichten die untersuchten Schüler $M=25.24$ Punkte, bei einer Standardabweichung von $SD=9.43$ Punkten.

Der Zusammenhang zwischen Lehrerurteil und Gesamtscore der Modellierungskompetenz liegt bei $r=-.590$, $p<.000$, was einem großen Effekt entspricht. Allerdings ist bei der Einschätzung durch Lehrer mit einer starken Konfundierung durch die Schulnote zu rechnen. Betrachtet man daher die Partialkorrelation zwischen Testleistung und Lehrerurteil, kontrolliert für die Zeugnisnote, so verbleibt immerhin (obwohl die Schulnote ja auch Modellierungskompetenz beinhaltet) ein Zusammenhang von $r=-.298$, $p<.000$ und damit ein annähernd mittelgroßer Effekt.

4.2.2.3.2 Konstruktvalidität

Die bisherigen Betrachtungen zur Validität des Verfahrens beschränkten sich auf die Ebene des Gesamtscores. Es ist jedoch darüber hinaus nachzuweisen, dass die postulierten vier Teilkompetenzen auch tatsächlich vier hinreichend unterschiedliche Konstrukte beschreiben. Andernfalls wäre der Bedarf nach einer Methode zur unabhängigen Erfassung der Teilkompetenzen obsolet. Die Konstruktvalidität kann also dadurch nachgewiesen werden, dass die theoretisch postulierte Faktorenstruktur mithilfe konfirmatorischer Faktorenanalysen bestätigt wird.

Das auf Passung zu den Daten zu prüfende Faktorenmodell ergibt sich aus der Theorie, auf deren Basis die Items konstruiert wurden. Zu testen ist demnach ein Vier-Faktoren-Modell mit den vier latenten Dimensionen *Strukturieren*, *Mathemati-*

sieren, *Verarbeiten* und *Interpretieren*, denen als manifeste Indikatoren jeweils die neun Items aus dem TEMOD 1 – Verfahren zuzuordnen sind.

Da die Items jeweils nur als Indikatoren für eine Teilkompetenz fungieren sollen, wurden keine Nebenladungen zugelassen. Alle vier Teilkompetenzen sind letztlich nur unterschiedliche Facetten eines gemeinsamen Konstruktes, das als Modellierungskompetenz bezeichnet wird. Es muss daher angenommen werden, dass die Interkorrelationen der vier Faktoren untereinander hoch sind. Korrelationen werden daher erlaubt und frei geschätzt.

Aus den neun Items pro Faktor und den sechs Interkorrelationen ergeben sich in der Summe 42 zu schätzende Parameter. Die Korrelationsmatrix, die als Ausgangspunkt für die Modellschätzung dient, enthält 666 Elemente, womit das Modell deutlich überidentifiziert ist.

Da die Indikatoren dichotom sind, wurde als Schätzverfahren der WLSMV-Algorithmus eingesetzt (Bentler & Chou, 1987; Schermelleh-Engel, Moosbrugger & Müller, 2003).

Im ersten Durchlauf zeigte sich ein Problem in Bezug auf Item S06, da nur .3% der Varianz in diesem Item durch den Faktor *Strukturieren* erklärt werden können und das Item daher offensichtlich nicht zu dem theoretisch postulierten latenten Konstrukt passt. Die von MPLUS produzierten Modification Indices empfehlen zur Verbesserung des Modellfits, das Item sowohl auf dem Faktor *Mathematisieren*, als auch auf dem Faktor *Interpretieren* laden zu lassen. Item S06 wurde daher aus den weiteren Analysen entfernt.

Im zweiten Durchlauf gab es seitens MPLUS keine weiteren Modification Indices. Abbildung 32 zeigt die Zuordnung der Items zu ihren jeweiligen Faktoren und die dazugehörigen Ladungen. Da jedes Item nur auf einem Faktor lädt, lässt sich die Residualvarianz jeweils berechnen durch $\delta=1-\lambda^2$.

Alle Items laden signifikant ($p < .0001$) auf ihrem Referenzfaktor, wobei jedoch vier Items (S07, S09, V4 und I5) Ladungen von <. 30 aufweisen, was einer Varianzaufklärung von weniger als 10% entspricht, weshalb diese Items selektiert und nicht im TEMOD 2 – Verfahren zum Einsatz kommen sollten. Auch unter dem Blickwinkel der Effektstärke ist dieses Vorgehen gerechtfertigt.

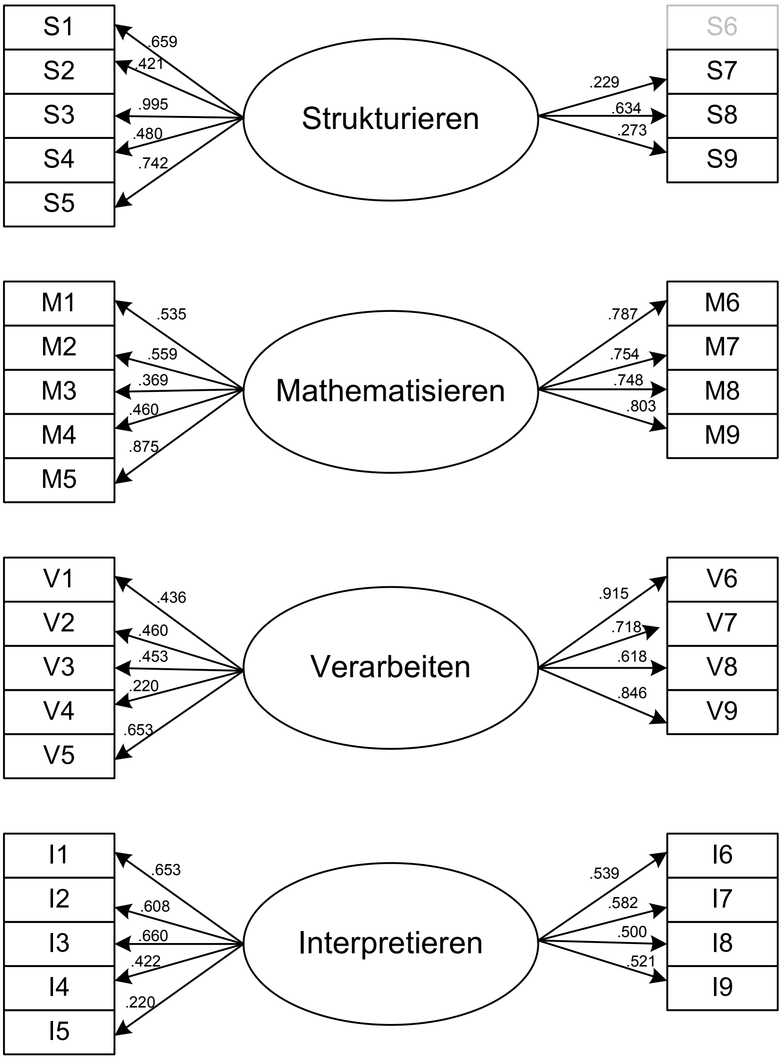


Abbildung 32: CFA mit vier Faktoren

Die Passung des Modells auf die Daten kann an mehreren Kriterien abgelesen werden (Bollen, 1989; Heck, 1998; Schermelleh-Engel et al., 2003):

Sowohl der RMSEA (Root Mean Square Error of Approximation) als auch der χ^2-Test messen die Passung zwischen theoretischer und empirischer Kovarianzmatrix. Der RMSEA weist mit .045 auf eine gute Passung zwischen theoretischer und empirischer Kovarianzmatrix hin (ab RMSEA von <.05 spricht man von gutem Fit). Der χ^2-Test kann bei Verwendung des WLSMV-Schätzverfahrens nicht interpretiert werden.

Zum Vergleich des Modells mit einem Baseline- oder Nullmodell, bei dem alle Parameter auf Null fixiert werden, lassen sich CFI (Comparative Fit Index) und TLI (Tucker Lewis Index oder Nonnormed Fit Index) heranziehen. Während der CFI mit .898 für einen schlechten Modellfit spricht, liegt der TLI mit .942 im akzeptablen Bereich und sogar nahe an der Grenze zum guten Bereich (der CFI sollte mindestens einen Wert von .95, der TLI einen Wert von mindestens .90 aufweisen).

Aus Tabelle 17 ist die Interkorrelation der Faktoren zu entnehmen. Am deutlichsten sticht der Zusammenhang zwischen den Teilkompetenzen *Mathematisieren* und *Interpretieren* hervor, der mit .995 darauf hinweist, dass es sich – aus rein statistischer Sicht – um das gleiche Konstrukt handelt.

Tabelle 17: Interkorrelationen bei 4 Faktoren

	S	M	V	I
Strukturieren		.952	.792	.882
Mathematisieren			.827	.995
Verarbeiten				.831
Interpretieren				

Auch aus inhaltlicher Sicht ist der Zusammenhang zwischen beiden Variablen nicht von der Hand zu weisen, bezeichnen sie doch lediglich spiegelbildlich die Übersetzungprozesse zwischen realer Welt und mathematischem Modell.

Um herauszufinden, wie sich die Auffassung beider Teilkompetenzen als gemeinsame latente Dimension auf die Modellgüte auswirkt, wurde ein zweites Modell spezifiziert, bei dem die Items M01-M09 sowie die Items I01-I09 auf einem gemeinsamen Faktor laden. Abbildung 33 stellt dies graphisch zusammen mit den aus der erneuten Modellschätzung resultierenden Faktorladungen dar.

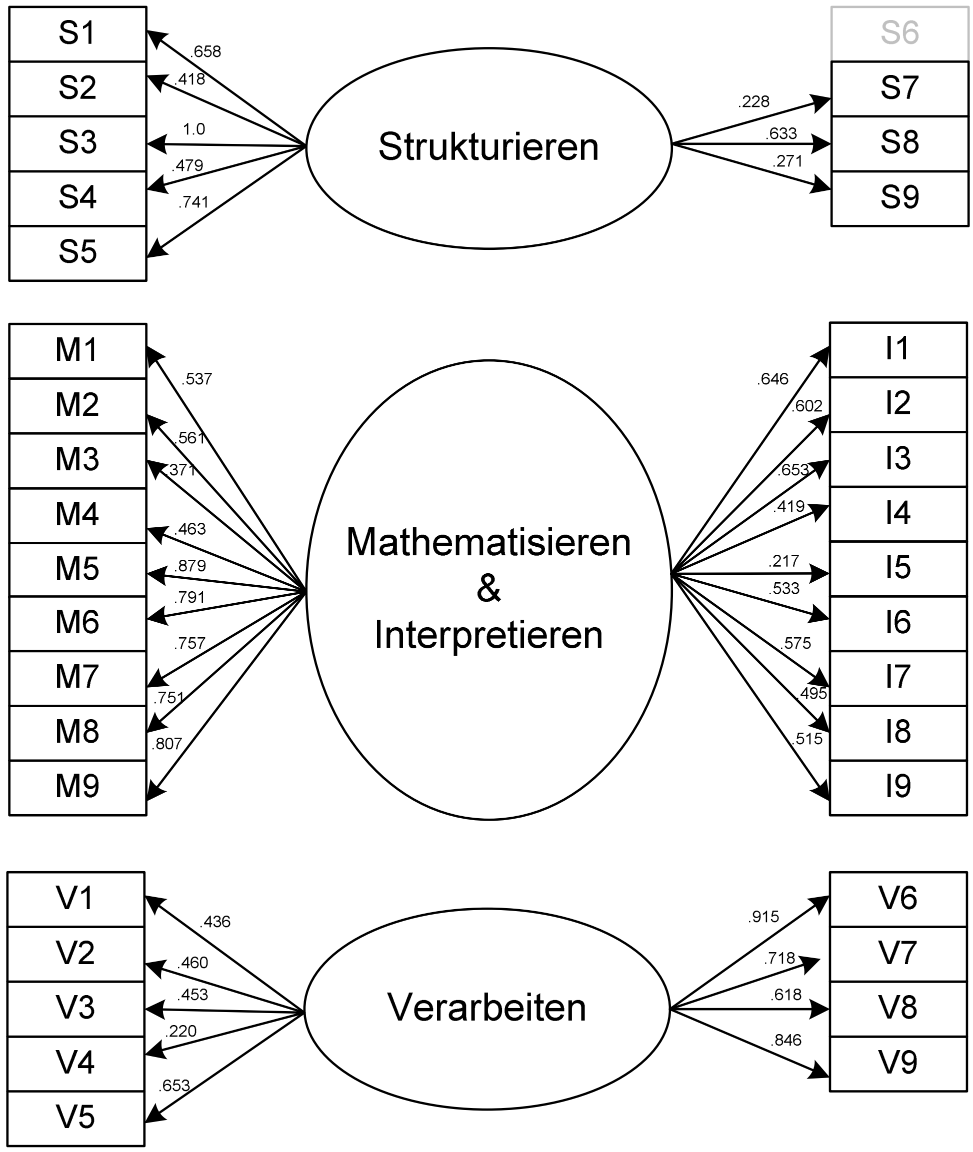


Abbildung 33: CFA mit drei Faktoren

Die Revision des Modells führt zu keinen Veränderungen im Modellfit, weshalb beide Varianten aus rein technischer Sicht als gleichwertig betrachtet werden müssen. Auch nach der gemeinsamen Auffassung von *Mathematisieren* und *Interpretieren* verbleiben (Tabelle 18) äußerst hohe Interkorrelationen zwischen den einzelnen Faktoren, die jeweils großen Effekten entsprechen.

Tabelle 18: Interkorrelationen bei 3 Faktoren

	S	MI	V
Strukturieren		.929	.792
Mathematisieren / Interpretieren			.830
Verarbeiten			

Diese Ergebnisse zeigen, dass die Teilkompetenzen untereinander teilweise schwer unterscheidbar sind und zum großen Teil etwas Ähnliches messen. Dies ist durchaus theoriekonform, denn dieses gemeinsame Konstrukt ist Modellierungskompetenz. Dennoch gibt es Unterschiede zwischen den Teilkompetenzen, denn wie die Modellpassung zeigt, konnte die theoretisch postulierte Faktorenstruktur bestätigt werden.

4.2.3 Der Modellierungszyklus als normatives Modell

Neben der Beurteilung der Testgütekriterien war eine zweite zentrale Fragestellung von Studie 2, inwieweit sich der Modellierungszyklus als normatives Modell eignet. Inhaltlich bedeutet ein normatives oder präskriptives Modell, dass eine bestimmte Anzahl (evtl. auch eine bestimmte Reihenfolge) von Stadien durchlaufen werden muss, damit das Ergebnis am Ende optimal ist. Wenn der Modellierungszyklus ein präskriptives Modell ist, dann bedeutet dies im vorliegenden Fall also, dass die Qualität der Lösung davon abhängt, dass die Teilkompetenzen entsprechend beherrscht und umgesetzt werden.

Die Nullhypothese lautet dementsprechend: Mindestens eine der theoretisch postulierten Teilkompetenzen liefert keinen signifikanten Erklärungswert für die Varianz in der Fähigkeit zum Modellieren (hier gemessen durch den Score im Bereich Textaufgaben) über den Einfluss der anderen Variablen hinaus. Dementsprechend muss mindestens eins der β-Gewichte in einer multiplen Regression nicht signifikant von 0 abweichen, um die H_0 beizubehalten.

Für die Prüfung dieser Hypothese wurde aufgrund der oben genannten Gründe bezüglich der Unzulänglichkeit gewöhnlicher Regressionsanalysen die Methode der Mehrebenenanlyse favorisiert.

Bei der Mehrebenenanalyse beginnt man zunächst mit dem Aufstellen eines Nullmodells, bei welchem noch keine unabhängige Variable aufgenommen wird. Die Modellgleichung enthält nur die abhängige Variable (in diesem Fall „text" für den Gesamtscore der Textaufgaben) und einen konstanten Term β_{0ij}. Dies stellt den y-Achsenabschnitt (Intecept) der gemeinsamen Regressionsgerade für alle Klassen dar und setzt sich zusammen aus dem Gesamtmittelwert β_0 und den Residuen. Die Residuen stellen die Abweichungen der einzelnen Messwerte von der Regressionsgeraden dar.

Bei der Mehrebenenanalyse geht man davon aus, dass sich diese Residualvarianz durch zwei Faktoren erklären lässt, nämlich die Varianz innerhalb von Klassen (also diejenige auf Individualebene) e_{0ij} und die Varianz auf Klassenebene u_{0j}. Der Index j bedeutet, dass ein Parameter auf der Klassenebene variiert, der Index i, dass er auf Individualebene variiert.

Im Nullmodell kann nun untersucht werden, wie viel Varianz auf welcher Ebene erklärt werden kann. Würde die Zugehörigkeit zu einer Klasse für die Modellierungskompetenz keine Rolle spielen, so dürfte auf Klassenebene keine Varianz erklärt werden. Abbildung 34 verdeutlicht die soeben erläuterten Zusammenhänge anhand der entsprechenden Formeln.

Equations

$$\mathrm{text}_{ij} \sim \mathrm{N}(XB,\ \Omega)$$

$$\mathrm{text}_{ij} = \beta_{0ij}\mathrm{cons}$$

$$\beta_{0ij} = \beta_0 + u_{0j} + e_{0ij}$$

$$\begin{bmatrix} u_{0j} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_u)\ :\ \Omega_u = \begin{bmatrix} \sigma^2_{u0} \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_e)\ :\ \Omega_e = \begin{bmatrix} \sigma^2_{e0} \end{bmatrix}$$

*-2*loglikelihood(IGLS Deviance)* = **1207,404(314 of 325 cases in use)**

Abbildung 34: Nullmodell vor Parameterschätzung

Nach der Parameterschätzung durch MLwiN zeigt sich folgendes Ergebnis (Abbildung 35). Die Werte vor den Klammern stellen die jeweils geschätzten Parameter dar, die Werte in den Klammern die dazugehörigen Standardfehler. Teilt man die Regressionsparameter durch ihre Standardfehler, so lässt sich an diesem t-verteilten Quotienten die Signifikanz des Parameters ablesen. Ab einem Wert von 1.96 bei zweiseitigen bzw. ab einem Wert von 1.65 bei einseitigen Hypothesen kann ein Parameter als signifikant auf dem Niveau von α=.05 betrachtet werden. Da die Zusammenhänge zwischen Prädiktoren und Kriterium jeweils als positiv erwartet werden, kann der Grenzwert für einseitige Hypothesentests zugrunde gelegt werden.

Equations

$$\text{text}_{ij} \sim \mathrm{N}(XB,\ \Omega)$$

$$\text{text}_{ij} = \beta_{0ij}\text{cons}$$

$$\beta_{0ij} = 4{,}027(0{,}156) + u_{0j} + e_{0ij}$$

$$\begin{bmatrix} u_{0j} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_u)\ :\ \Omega_u = \begin{bmatrix} 0{,}254(0{,}138) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_e)\ :\ \Omega_e = \begin{bmatrix} 2{,}594(0{,}212) \end{bmatrix}$$

*-2*loglikelihood(IGLS Deviance)* = **1207,404(314 of 325 cases in use)**

Abbildung 35: Nullmodell nach Parameterschätzung

Der Gesamtmittelwert liegt bei 4.027 richtig gelösten Aufgaben. Aus der u-Matrix und der e-Matrix lässt sich ableiten, dass auf Klassenebene zwar weitaus weniger Varianz erklärt wird als auf Individualebene, dass jedoch auch die Klassenebene (.254/.138=1.84) einen signifikanten Beitrag zur Varianzerklärung in der Modellierungskompetenz leistet.

Dies zeigt sich auch, wenn man für jede Klasse eine eigene Regressionsgerade schätzen lässt, wobei diese im Fall des Nullmodells noch keine Vorhersagen aufgrund der Prädiktoren erlaubt, sondern lediglich die verschiedenen Intercepts aufgrund der unterschiedlichen Klassenmittelwerte zum Ausdruck bringt (Abbildung 36).

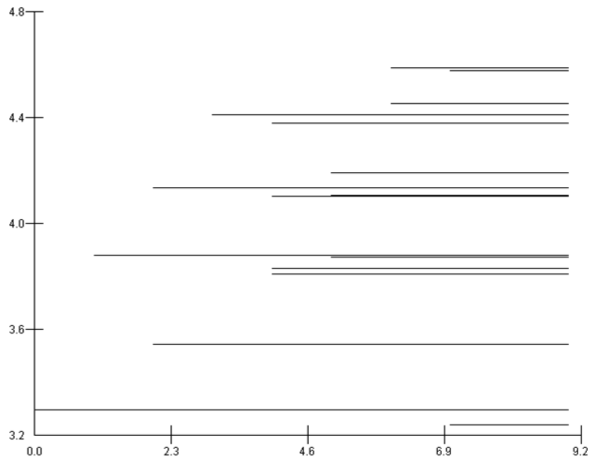


Abbildung 36: Regressionsgeraden im Nullmodell

Diese hier zu beobachtende Streuung der Intercepts rechtfertigt eine weitere Betrachtung der Fragestellung auf der Basis von zwei Ebenen.

Im nächsten Schritt gilt es, den Einfluss der jeweiligen Teilkompetenzen als Prädiktoren für die Lösungswahrscheinlichkeit zu untersuchen. Dabei soll zunächst für jede Teilkompetenz ein eigenes Modell spezifiziert und im Anschluss alle Teilkompetenzen in eine einzige multiple Regressionsgleichung integriert werden.

Abbildung 37 zeigt die Regressionsgleichung für *Strukturieren* vor der Parameterschätzung.

Equations

$$\text{text}_{ij} \sim \mathrm{N}(XB, \Omega)$$

$$\text{text}_{ij} = \beta_{0ij}\text{cons} + \beta_1 \text{strukt\_c}_{ij}$$

$$\beta_{0ij} = \beta_0 + u_{0j} + e_{0ij}$$

$$\begin{bmatrix} u_{0j} \end{bmatrix} \sim \mathrm{N}(0, \Omega_u) : \Omega_u = \begin{bmatrix} \sigma^2_{u0} \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim \mathrm{N}(0, \Omega_e) : \Omega_e = \begin{bmatrix} \sigma^2_{e0} \end{bmatrix}$$

*-2*loglikelihood(IGLS Deviance)* = 1168,363(314 of 325 cases in use)

Abbildung 37: Random-Intercept-Modell für Strukturieren vor der Parameterschätzung

Hier werden nun vier Parameter geschätzt: Zunächst sind dies das Intercept β_{0ij} und das Regressionsgewicht von *Strukturieren* β_1. Die Indizes bei β_{0ij} symbolisieren, dass dies kein fixer Parameter ist, sondern dass der Mittelwert variiert. Er setzt sich zusammen aus dem Gesamtmittelwert β_0 und zwei Fehleranteilen u und e. Der dritte zu schätzende Parameter σ^2_{u0} beschreibt denjenigen Anteil dieser Varianz, der auf Klassenebene zu finden ist, wohingehend der vierte Parameter σ^2_{e0} denjenigen Anteil beschreibt, der lediglich auf der Individualebene besteht. Es wird bei diesem Modell also davon ausgegangen, dass sich Klassen in ihren Leistungen in der AV unterscheiden, dass die Variable *Strukturieren* einen Einfluss auf die Variable Textaufgaben hat, dass die Größe dieses Zusammenhangs (ausgedrückt durch die nicht über Klassen variierende Steigung der Regressionsgeraden β_1) aber über alle Klassen hinweg gleich ist.

Abbildung 38 gibt nun die Ergebnisse der Parameterschätzung wieder.

Equations

$$\text{text}_{ij} \sim \mathrm{N}(XB, \Omega)$$

$$\text{text}_{ij} = \beta_{0ij}\text{cons} + 0{,}465(0{,}072)\text{strukt\_c}_{ij}$$

$$\beta_{0ij} = 3{,}970(0{,}161) + u_{0j} + e_{0ij}$$

$$\left[u_{0j}\right] \sim \mathrm{N}(0,\ \Omega_u) \ : \ \Omega_u = \left[0{,}294(0{,}146)\right]$$

$$\left[e_{0ij}\right] \sim \mathrm{N}(0,\ \Omega_e) \ : \ \Omega_e = \left[2{,}269(0{,}186)\right]$$

*-2*loglikelihood(IGLS Deviance)* = 1168,363(314 of 325 cases in use)

Abbildung 38: Random-Intercept-Modell für Strukturieren nach der Parameterschätzung

Strukturieren hat einen signifikanten Einfluss (.465/.072=6.46) auf die Modellierungsleistung, und die Klasse, in der sich ein Schüler befindet, spielt bei der Vorhersage seiner Leistung ebenfalls eine signifikante Rolle (.294/.146=2.01). Hierbei sei angemerkt, dass der Einfluss der Klasse lediglich auf die abhängige Variable zu sehen ist und mit der Teilkompetenz *Strukturieren* nichts zu tun hat. Der Effekt zeigt sich so auch im Nullmodell, welches keine Prädiktoren enthält. Daher wird dieser in den folgenden Modellen, die die anderen Teilkompetenzen als Prädiktoren enthalten, logischerweise repliziert und muss nicht mehr bei jedem Modell im Hinblick auf seine Signifikanz erwähnt werden.

Das Random-Intercept-Modell kann dann als gültig betrachtet werden, wenn die Intercepts signifikant zwischen Klassen variieren. Dies ist hier der Fall, womit der Einsatz der Mehrebenenanalyse gegenüber der einfachen Regression gerechtfertigt ist. Dies gilt, wie eben ausgeführt, auch für die anderen Teilkompetenzen, da der y-Achsenabschnitt schon vor der Aufnahme von Prädiktoren ins Modell feststeht. Den Abbildungen 39 bis 41 sind die Random-Intercept-Modelle für *Mathematisieren*, *Verarbeiten* und *Interpretieren* zu entnehmen.

Equations

$$\text{text}_{ij} \sim \mathrm{N}(XB, \Omega)$$
$$\text{text}_{ij} = \beta_{0ij}\text{cons} + 0{,}426(0{,}053)\text{math\_c}_{ij}$$
$$\beta_{0ij} = 3{,}964(0{,}153) + u_{0j} + e_{0ij}$$
$$\left[u_{0j}\right] \sim \mathrm{N}(0,\ \Omega_u)\ :\ \Omega_u = \left[0{,}262(0{,}132)\right]$$
$$\left[e_{0ij}\right] \sim \mathrm{N}(0,\ \Omega_e)\ :\ \Omega_e = \left[2{,}137(0{,}175)\right]$$

*-2*loglikelihood(IGLS Deviance)* = 1148,915(314 of 325 cases in use)

Abbildung 39: Random-Intercept-Modell für Mathematisieren

Equations

$$\text{text}_{ij} \sim \mathrm{N}(XB, \Omega)$$
$$\text{text}_{ij} = \beta_{0ij}\text{cons} + 0{,}402(0{,}050)\text{vera\_c}_{ij}$$
$$\beta_{0ij} = 3{,}982(0{,}134) + u_{0j} + e_{0ij}$$
$$\left[u_{0j}\right] \sim \mathrm{N}(0,\ \Omega_u)\ :\ \Omega_u = \left[0{,}175(0{,}102)\right]$$
$$\left[e_{0ij}\right] \sim \mathrm{N}(0,\ \Omega_e)\ :\ \Omega_e = \left[2{,}171(0{,}178)\right]$$

*-2*loglikelihood(IGLS Deviance)* = 1149,530(314 of 325 cases in use)

Abbildung 40: Random-Intercept-Modell für Verarbeiten

Equations

$$\text{text}_{ij} \sim \mathrm{N}(XB, \Omega)$$
$$\text{text}_{ij} = \beta_{0ij}\text{cons} + 0{,}424(0{,}045)\text{inte\_c}_{ij}$$
$$\beta_{0ij} = 3{,}981(0{,}136) + u_{0j} + e_{0ij}$$
$$\left[u_{0j}\right] \sim \mathrm{N}(0,\ \Omega_u)\ :\ \Omega_u = \left[0{,}190(0{,}104)\right]$$
$$\left[e_{0ij}\right] \sim \mathrm{N}(0,\ \Omega_e)\ :\ \Omega_e = \left[2{,}022(0{,}166)\right]$$

*-2*loglikelihood(IGLS Deviance)* = 1128,749(314 of 325 cases in use)

Abbildung 41: Random-Intercept-Modell für Interpretieren

Die Regressionsgewichte liegen alle in einem ähnlichen Bereich und sind jeweils signifikant, ebenso wie die Residuen auf Klassenebene, womit das Random-Intercept-Modell für alle vier Teilkompetenzen als angemessen betrachtet werden kann.

Am Beispiel *Strukturieren* wird nun graphisch verdeutlicht, was dies bedeutet: Im Random-Intercept-Modell wird für jede Klasse eine eigene Regressionsgleichung geschätzt. Da jedoch die Steigung auf β_1 fixiert wurde, variieren diese Gleichungen, und die aus ihnen resultierenden Regressionsgeraden wie sie in Abbildung 42 wer-den, lediglich bezüglich ihrer y-Achsenabschnitte. Für die Regressionsgeraden er-geben sich wie beim Nullmodell parallele Linien, die nun aber mit einer Steigung versehen sind. Dies bedeutet, dass es in Bezug auf die Modellierungskompetenz gute und schlechte Klassen gibt.

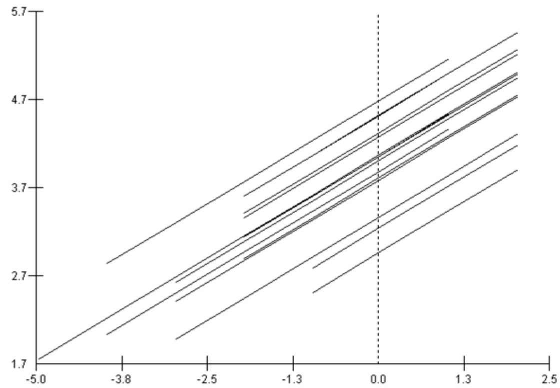


Abbildung 42: Regressionsgeraden im Random-Intercept-Modell

Nachdem es offensichtlich angezeigt ist, die vorliegenden Daten als Random-Intercept-Modelle zu spezifizieren, muss in einem zweiten Schritt der Frage nachgegangen werden, ob sogar ein Random-Slope-Modell auf die Daten passt.

Das Random-Slope Modell für *Strukturieren* vor der Parameterschätzung ist in Abbildung 43 dargestellt.

Equations

$$\text{text}_{ij} \sim \mathrm{N}(XB, \Omega)$$

$$\text{text}_{ij} = \beta_{0ij}\text{cons} + \beta_{1j}\text{strukt\_c}_{ij}$$

$$\beta_{0ij} = \beta_0 + u_{0j} + e_{0ij}$$

$$\beta_{1j} = \beta_1 + u_{1j}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_u)\ :\ \Omega_u = \begin{bmatrix} \sigma^2_{u0} & \\ \sigma_{u01} & \sigma^2_{u1} \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_e)\ :\ \Omega_e = \begin{bmatrix} \sigma^2_{e0} \end{bmatrix}$$

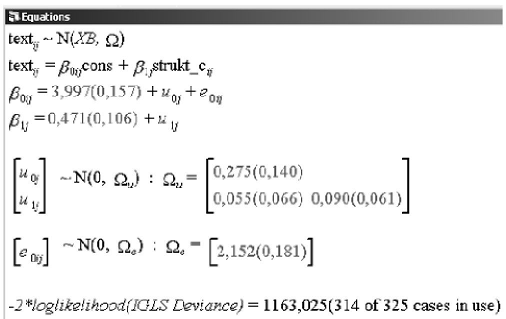
Abbildung 43: Random-Slope-Modell für Strukturieren vor der Parameterschätzung

Im Unterschied zum Random-Intercept-Modell werden zusätzlich zu den variierenden y-Achsenabschnitten auch variierende Steigungen der Regressionsgeraden zugelassen. Das Regressionsgewicht wird von β_1 zu β_{1j}, da es nun zwischen den Klassen variiert. Das in die allgemeine Regressionsgleichung einzufügende β-Gewicht stellt dann lediglich den Mittelwert über alle geschätzten β-Gewichte dar. Dieses β_{1j} setzt sich zusammen aus diesem Mittelwert und einem auf Klassenebene variierenden Fehlerterm u_{1j}, für den nun auch eine Varianz σ^2_{u1} geschätzt wird. Diese Varianz der Steigungen ist wie folgt zu interpretieren: Für den Fall, dass es eine signifikant von Null verschiedene Varianz der Steigungen zwischen Klassen gibt, bedeutet dies, dass in einigen Klassen die Prädiktorvariable einen größeren Einfluss auf das Kriterium hat als in anderen Klassen.

Auf den vorliegenden Fall bezogen würde das bedeuten, dass es Klassen gibt, in denen *Strukturieren* einen größeren Einfluss auf die Modellierungsfähigkeit hat als in anderen Klassen oder anders ausgedrückt: Es gibt Klassen, in denen das Verfügen über die Teilkompetenz *Strukturieren* die Lösungswahrscheinlichkeit einer Aufgabe mehr erhöht als in anderen Klassen. Ein solcher Effekt lässt sich nicht in Einklang mit dem theoretischen Hintergrund zu Modellierungskompetenzen bringen, weshalb auch keine signifikant variierenden Steigungen und somit keine Angemessenheit des Random-Slope-Modells erwartet werden, weder für *Strukturieren* noch für die weiteren Teilkompetenzen.

Gibt es allerdings einen signifikant variierenden Slope, so kann man darüber hinaus untersuchen, ob dieser mit dem variierenden Intercept in Zusammenhang steht. In MLwiN wird dies über die Kovarianz zwischen y-Achsenabschnitt und Steigung gemessen, σ^2_{uo1}. Inhaltlich würde eine signifikant von Null verschiedene Kovarianz bedeuten, dass mit steigendem Leistungsniveau der Klasse (hoher Intercept) der Einfluss der Teilkompetenz (hoher Slope) auf die Modellierungsfähigkeit zunimmt. Da auch dies theoretisch nicht sinnvoll ist, wird keine solche Kovarianz erwartet.

In Abbildung 44 ist zunächst das Random-Slope-Modell für *Strukturieren* nach erfolgter Parameterschätzung dargestellt.


Equations

$$\text{text}_{ij} \sim N(XB, \Omega)$$
$$\text{text}_{ij} = \beta_{0ij}\text{cons} + \beta_{1j}\text{strukt\_c}_{ij}$$
$$\beta_{0ij} = 3{,}997(0{,}157) + u_{0j} + e_{0ij}$$
$$\beta_{1j} = 0{,}471(0{,}106) + u_{1j}$$
$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0{,}275(0{,}140) & \\ 0{,}055(0{,}066) & 0{,}090(0{,}061) \end{bmatrix}$$
$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 2{,}152(0{,}181) \end{bmatrix}$$

-2*loglikelihood(IGLS Deviance) = 1163,025(314 of 325 cases in use)

Abbildung 44: Random-Slope-Modell für Strukturieren nach der Parameterschätzung

Hieraus ist ersichtlich, dass weder der Slope signifikant variiert (.090/.061=1.48), noch dass der Slope mit dem Intercept korreliert (.055/.066=.83), was wenig verwunderlich ist in Anbetracht der Tatsache, dass es im Slope de facto keine Varianz gibt. Abbildung 45 veranschaulicht dies graphisch: Die auf Basis des Random-Slope-Modells geschätzten Regressionsgeraden variieren zwar leicht in ihrer Steigung, die meisten Linien sind aber nach wie vor annähernd parallel.

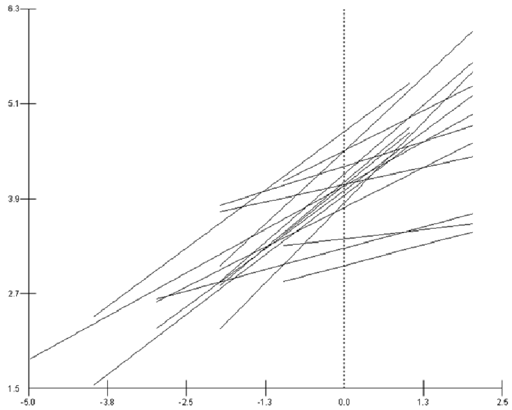


Abbildung 45: Regressionsgeraden im Random-Slope-Modell

Die nachfolgenden Abbildungen 46 bis 48 zeigen die Parameterschätzungen auf Grundlage des Random-Slope-Modells für *Mathematisieren*, *Verarbeiten* und *Interpretieren*.

Equations

$$\text{text}_{ij} \sim \mathrm{N}(XB,\ \Omega)$$

$$\text{text}_{ij} = \beta_{0ij}\text{cons} + \beta_{1j}\text{math\_c}_{ij}$$

$$\beta_{0ij} = 3{,}971(0{,}145) + u_{0j} + e_{0ij}$$

$$\beta_{1j} = 0{,}428(0{,}061) + u_{1j}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_u) \ :\ \Omega_u = \begin{bmatrix} 0{,}224(0{,}119) & \\ 0{,}033(0{,}035) & 0{,}013(0{,}019) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_e) \ :\ \Omega_e = \begin{bmatrix} 2{,}115(0{,}177) \end{bmatrix}$$

*-2*loglikelihood(IGLS Deviance)* = 1147,598(314 of 325 cases in use)

Abbildung 46: Random-Slope-Modell für Mathematisieren

Equations

$$\text{text}_{ij} \sim \mathrm{N}(XB,\ \Omega)$$

$$\text{text}_{ij} = \beta_{0ij}\text{cons} + \beta_{1j}\text{vera\_c}_{ij}$$

$$\beta_{0ij} = 3{,}977(0{,}132) + u_{0j} + e_{0ij}$$

$$\beta_{1j} = 0{,}399(0{,}056) + u_{1j}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_u) \ :\ \Omega_u = \begin{bmatrix} 0{,}164(0{,}099) & \\ 0{,}020(0{,}029) & 0{,}009(0{,}017) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_e) \ :\ \Omega_e = \begin{bmatrix} 2{,}150(0{,}180) \end{bmatrix}$$

*-2*loglikelihood(IGLS Deviance)* = 1148,989(314 of 325 cases in use)

Abbildung 47: Random-Slope-Modell für Verarbeiten

Equations

$$\text{text}_{ij} \sim \mathrm{N}(XB, \Omega)$$
$$\text{text}_{ij} = \beta_{0ij}\text{cons} + \beta_{1j}\text{inte\_c}_{ij}$$
$$\beta_{0ij} = 3{,}981(0{,}136) + u_{0j} + e_{0ij}$$
$$\beta_{1j} = 0{,}424(0{,}045) + u_{1j}$$
$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_u)\ :\ \Omega_u = \begin{bmatrix} 0{,}190(0{,}104) & \\ 0{,}000(0{,}000) & 0{,}000(0{,}000) \end{bmatrix}$$
$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_e)\ :\ \Omega_e = \begin{bmatrix} 2{,}022(0{,}166) \end{bmatrix}$$

*-2*loglikelihood(IGLS Deviance)* = **1128,749(314 of 325 cases in use)**

Abbildung 48: Random-Slope-Modell für Interpretieren

Aus den Schätzergebnissen wird deutlich, dass – wie zu erwarten – weder die Varianz des Slope noch die Kovarianz zwischen Intercept und Slope signifikant von Null verschieden sind.

Das Random-Slope-Modell kann daher zugunsten des Random-Intercept-Modells für alle vier Teilkompetenzen verworfen werden.

Auf der Basis dieser Ergebnisse lässt sich nun die ursprüngliche Fragestellung nach dem Modellierungszyklus als normativem Modell mithilfe eines mehrebenenanalytischen Gesamtmodells beantworten, wobei alle Teilkompetenzen in eine einzige multiple Regressionsgleichung auf der Basis eines Random-Intercept-Modells aufgenommen werden. Wie bereits erwähnt, ist für den Nachweis der Eignung als normatives Modell zu erwarten, dass jede einzelne Prädiktorvariable für sich genommen einen signifikanten Beitrag zur Erklärung der Varianz in der Modellierungsleistung erbringen kann. Abbildung 49 zeigt die spezifizierte Modellgleichung zusammen mit den geschätzten β-Gewichten und ihren Standardfehlern.

Equations

$$\text{text}_{ij} \sim \mathrm{N}(XB, \Omega)$$
$$\text{text}_{ij} = \beta_{0ij}\text{cons} + 0{,}207(0{,}070)\text{strukt\_c}_{ij} + 0{,}117(0{,}063)\text{math\_c}_{ij} + 0{,}212(0{,}052)\text{vera\_c}_{ij} + 0{,}233(0{,}055)\text{inte\_c}_{ij}$$
$$\beta_{0ij} = 3{,}936(0{,}141) + u_{0j} + e_{0ij}$$
$$\begin{bmatrix} u_{0j} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_u)\ :\ \Omega_u = \begin{bmatrix} 0{,}222(0{,}111) \end{bmatrix}$$
$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_e)\ :\ \Omega_e = \begin{bmatrix} 1{,}778(0{,}146) \end{bmatrix}$$

*-2*loglikelihood(IGLS Deviance)* = **1091,390(314 of 325 cases in use)**

Abbildung 49: Gesamtmodell

Die β-Gewichte erscheinen auf den ersten Blick gering, es gilt jedoch zu bedenken, dass es sich hier jeweils um den von allen anderen Prädiktorvariablen bereinigten Einfluss einer Teilkompetenz handelt, der Wert ist also jeweils im Sinne einer Partialkorrelation zu interpretieren. Alle ß-Gewichte sind signifikant, wodurch untermauert wird, dass jede Teilkompetenz einen wichtigen Beitrag zur Erklärung der Modellierungskompetenz liefert.

Wie lässt sich aber nun die Güte dieses Gesamtmodells überprüfen? Es ist der Frage nachzugehen, wie viel Varianz im Kriterium durch die Prädiktoren gemeinsam erklärt werden kann. Hierfür wurde der multiple Determinationskoeffzient R^2 berechnet. Da MLwiN keine solchen Kennwerte ausgibt, geschah dies manuell in SPSS. R^2 berechnet sich aus der quadrierten Korrelation zwischen den beobachteten Werten y und den durch die Regressionsgleichung prädizierten Werten $\hat{y}$ (Wirtz & Nachtigall, 1998). Letztere setzen sich laut Regressionsgleichung zusammen aus:

$$\hat{y} = \beta_{0ij} + \beta_1\,(strukt) + \beta_2\,(math) + \beta_3\,(vera) + \beta_4(inte)$$

In diesem Fall also:

$$\hat{y} = 3.936 + .207\,(strukt) + .117(math) + .212\,(vera) + .233(inte)$$

Aufgrund dieser Gleichung wurde für jede Versuchsperson der erwartete Wert berechnet. R^2 berechnet sich dann aus

$$R^2 = [r\,(y_i, \hat{y}_i]^2$$

und bezeichnet den Anteil der durch das Modell erklärten Varianz in der Variable Textaufgaben (y). Für R^2 ergibt sich ein Wert von .3025 ($p<.001$). Wie ist ein erklärter Varianzanteil von ca. 30% zu bewerten?

Aus R^2 als Maß der multiplen Korrelation lässt sich ein Effekstärkemaß f^2 berechnen, wobei ab .15 von einem mittleren, ab .35 von einem großen Effekt gesprochen wird (Bortz & Döring, 2002). Hierfür wird der Anteil der erklärten Varianz durch den Anteil der nicht-erklärten Varianz dividiert, wodurch sich hier ein f^2 von .43 ergibt, was einem großen Effekt entspricht.

4.3. Diskussion

Die im vorhergehenden Abschnitt dargestellten Ergebnisse zu Items, Gütekriterien und normativem Modell werden nun nacheinander diskutiert. Eine allgemeine Diskussion und Einordnung der Ergebnisse von Vor- und Hauptuntersuchung erfolgt in Kapitel 5.

4.3.1 Analyse der Items

Die Analyse der Items des Tests zur Erfassung der Modellierungskompetenz zeigt, dass die konstruierten Aufgaben im Großen und Ganzen geeignet sind, in einem TEMOD 2 Verfahren eingesetzt zu werden. Allerdings gibt es einzelne Items, die sich als ungeeignet erwiesen haben und daher ausgeschlossen werden sollten.

Dies ist ein bei der Konstruktion von Tests durchaus zu erwartendes Ergebnis, da sich die Qualität von Items letztlich immer nur empirisch bestimmen lässt. Lediglich bei der Skala *Strukturieren* ist es als kritisch zu betrachten, dass nur vier geeignete Items verbleiben. Hier wären mehr Indikatoren dieses Teilkonstrukts durchaus wünschenswert. Für die Konstruktion eines in der Praxis einzusetzenden Testverfahrens wäre daher zur Erhöhung der Reliabilität die Konstruktion und Erprobung weiterer Strukturierungsitems in Erwägung zu ziehen.

Bei der Stichprobenbeschreibung wurde bereits darauf hingewiesen, dass durch die Selbstselektion der teilnehmenden Schulen Deckeneffekte bei der Items zu erwarten sind. Wie erwartet traten diese Effekte ein. In der Population wird allerdings eine annähernde Normalverteilung für die jeweiligen Summenscores erwartet, was an einer repräsentativen Stichprobe zu überprüfen wäre.

Wie können individuelle Ergebnisse des TEMOD 2 – Tests aussehen, auf deren Grundlage dann eine Förderung eingeleitet werden kann? Hierfür kann für jeden teilnehmenden Schüler ein Profil erstellt werden. Abbildung 50 verdeutlicht dies anhand von fünf zufällig ausgewählten Schülern aus der untersuchten Stichprobe.

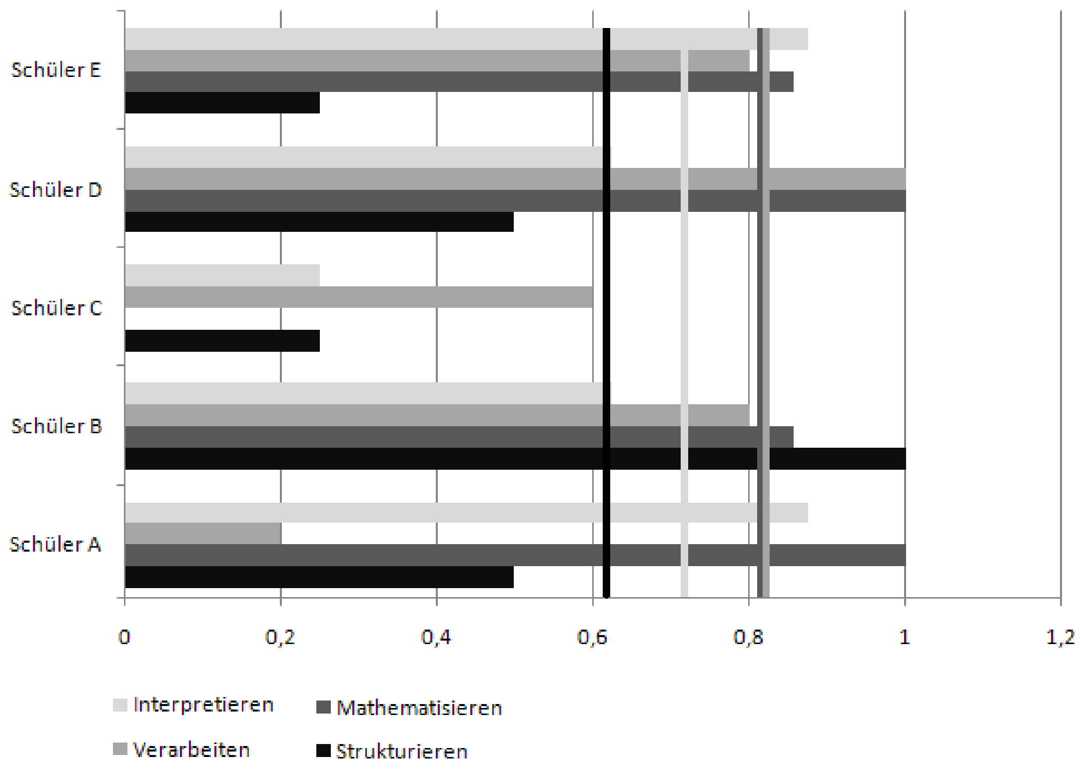


Abbildung 50: Ergebnisprofile

Da die Skalen zu den einzelnen Teilkompetenzen nach der Itemselektion unterschiedlich viele Aufgaben enthalten, wurde für die Berechnung der Skalenmittelwerte und der Testergebnisse der Schüler A bis E der jeweilige Wert der Skalen durch die Anzahl der Items dividiert, so dass für jede Skala ein theoretisches Minimum von 0 Punkten und ein theoretisches Maximum von 1.0 Punkten resultiert.

Die senkrechten Balken zeigen die Gesamtmittelwerte für die jeweilige Teilkompetenz aus der Untersuchungsstichprobe an, die waagerechten Balken das Abschneiden der Schüler in Bezug auf die jeweilige Teilkompetenz. Obwohl die Mittelwerte alle recht nahe beieinander liegen, zeigt Abbildung 50, dass es doch deutliche *intraindividuelle* Unterschiede bei Schülern hinsichtlich der verschiedenen Teilkompetenzen gibt.

Schüler A hat deutliche Schwächen beim *Strukturieren* und *Verarbeiten*, in denen er entsprechend gefördert werden müsste, wohingegen ihm die Übersetzungskompetenzen *Mathematisieren* und *Interpretieren* keine Schwierigkeiten bereiten. Schüler C hat in allen Bereichen Defizite, wohingegen Schüler E nur Schwierigkeiten beim *Strukturieren* hat.

Anhand dieser Beispiele wird deutlich, wie eine individuelle Diagnostik und darauf aufbauende Förderung aussehen könnte. Notwendig werden dann natürlich eine Reihe von Übungsmaterialen, die ähnlich aussehen, wie die im Test verwendeten Items, mit denen dann die unterschiedlichen Teilkompetenzen tatsächlich auch unabhängig voneinander trainiert werden können.

4.3.2 Gütekriterien des Testverfahrens

Die im folgenden Abschnitt diskutierten Gütekriterien beziehen sich zunächst auf das Instrument in der Form, wie es in der Hauptuntersuchung eingesetzt wurde. Die Güte des aus der vorliegenden Arbeit resultierenden Verfahrens ist zweifelsohne gesondert zu überprüfen. Insgesamt kann davon ausgegangen werden, dass die Güte der hier verwendeten Items, welche als Grundlage für ein späteres Verfahren dienen sollen, als notwendige – wenn auch nicht als hinreichende – Bedingung für die Qualität des Endproduktes betrachtet werden können.

4.3.2.1 Objektivität

Wie für ein standardisiertes Multiple-Choice-Verfahren wenig verwunderlich, stellt die Objektivität in der Hauptuntersuchung kein Problem dar. Bedenkt man nun jedoch die angedachte Anwendung des Testverfahrens, so müssen mögliche Einschränkungen diskutiert werden. Setzt man das TEMOD-Testverfahren als Gruppendiagnostikum ein, so ist zumindest innerhalb einer Klasse die Vergleichbarkeit der Ergebnisse recht leicht zu gewährleisten, da das Testverfahren dann jedem Schüler unter gleichen Bedingungen vorgelegt werden kann. Dennoch ist selbst in diesem Fall nicht auszuschließen, dass Lehrkräfte – sofern sie als Testleiter fungieren – einigen Schülern, aber eben nicht allen, inhaltliche Hilfestellungen beim Bearbeiten der Testaufgaben geben.

Da das Verfahren aber insbesondere als Grundlage zur individuellen Förderung eingesetzt werden soll, sind Individualtestungen – z.B. im Rahmen von Nachhilfeunterricht – die wohl wahrscheinlichere Variante. Das individuelle Testergebnis muss dann mit einer noch zu gewinnenden Norm verglichen werden und kann inhaltlich nur dann sinnvoll interpretiert werden, wenn jede Testperson unter vergleichbaren Bedingungen getestet wird.

Bedeutsam ist daher die angemessene Instruktion der Testleiter, die als Grundlage für eine objektive Testdurchführung zu sehen ist. Nur dadurch ist Bedingungskonstanz zu realisieren.

4.3.2.2 Reliabilität

Bei der Diskussion der Ergebnisse aus Studie 1 wurde darauf hingewiesen, dass für den Test in Studie 2 eine höhere Reliabilität anzustreben ist, als sie in der Voruntersuchung erreicht werden konnte. Das Testverfahren kann diesem Anspruch gerecht werden und weist eine zufriedenstellende bis gute Reliabilität auf. Warum diese in Studie 2 höher war als in Studie 1 liegt nahe: Bei der Voruntersuchung wurde das Auftreten versus Nicht-Auftreten der verschiedenen Teilkompetenzen lediglich durch Beobachtungen gemessen, in der Hauptstudie hingegen durch Leistungsdaten erfasst. Dass Beobachter – auch bei lautem Denken – Verhalten nicht genauso zuverlässig erfassen können, wie dies bei Testdaten der Fall ist, ist trivial. Man kann sich an dieser Stelle nun fragen, wie dann die hohe Reliabilität bei den Ratings der Mathematiklehrer zustande kam, wenn es sich hierbei doch auch um Fremdeinschätzungen handelt. Hier gilt es zu bedenken, dass Lehrkräfte zu ihrem Urteil nicht allein aufgrund von Beobachtungen kommen, sondern auch durch das Hinzuziehen von „harten“ Daten, wie Testergebnisse in Klassenarbeiten. Auch dürften die Beobachtungen von Lehrkräften reliabler sein, da sich ihre Beobachtungen über längere Zeiträume und verschiedene Situationen erstrecken, wohingegen die Testleiter in der Voruntersuchung jeweils nur wenige Stunden mit ihren Testpersonen zu tun hatten.

4.3.2.3 Validität

Der Validität als Gütekriterium kommt deshalb eine so große Rolle zu, weil sie die Rechtfertigungsgrundlage für den Einsatz eines Testverfahrens bildet. Es gilt den Nachweis zu bringen, dass der TEMOD-Test auch tatsächlich Modellierungskompetenz misst und nicht etwa ein völlig anderes Konstrukt.

Um dies sicherzustellen, wurden zwei Arten der Validität eingehend untersucht: Kriteriums- und Konstruktvalidität.

4.3.2.3.1 Kriteriumsvalidität

Alle drei zur Validierung hinzugezogenen Außenkriterien weisen eine Nähe zur Modellierungskompetenz auf. Dabei deckt keines die Modellierungskompetenz hinreichend gut ab, worin auch die Notwendigkeit zur Konstruktion des vorliegenden Testverfahrens begründet liegt. Abbildung 51 verdeutlicht einige Aspekte der Gemeinsamkeiten und Unterschiede zwischen den verschiedenen Konstrukten.

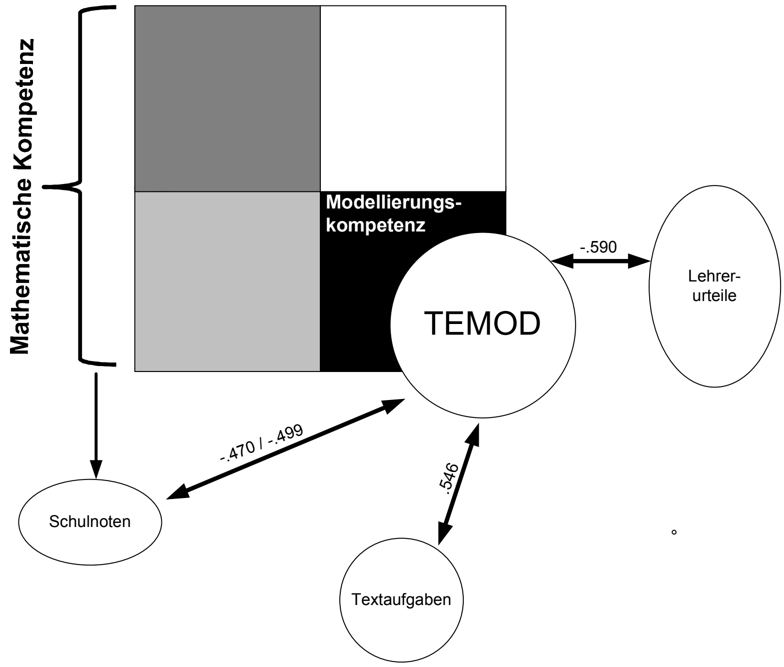


Abbildung 51: Außenkriterien zur Validierung
Anmerkung: Beim TEMOD-Score und bei den Textaufgaben gehen hohe Werte mit einer hohen Anzahl von gelösten Aufgaben einher. Lehrerurteile beziehen sich darauf, wie oft Schüler bei verschiedenen Aspekten des Modellierens Schwierigkeiten haben – hohe Werte bedeuten daher eine geringe Modellierungskompetenz. Auch bei Schulnoten bedeuten niedrige Werte hohe Leistungsfähigkeit und umgekehrt. Die Vorzeichen bei Noten und Lehrerurteilen sind daher negativ.

Der in Abbildung 51 durch ein Quadrat dargestellten mathematischen Kompetenz kommt im Mathematikunterricht eine tragende Rolle zu. Wie in den Bildungsstandards formuliert, ist sie das Ziel, das alle Schüler am Ende ihrer Schulzeit erreicht haben sollen. Erfasst wird sie innerhalb des Bildungssystems unter anderem durch Noten, jedoch in den letzten Jahren zunehmend auch auf anderen Wegen – in der Grundschule z.B. in der IGLU-Studie (Walther et al., 2003) oder mithilfe von Vergleichsarbeiten (Hosenfeld, 2005).

Es kann davon ausgegangen werden, dass diese Art der Leistungserhebung weit weniger messfehlerbehaftet ist, als dies bei Schulnoten der Fall ist, schon allein aufgrund der bei Schulnoten unklaren Zuordnung zu norm- oder kriteriumsorientierten Verfahren (Schott, 1999). Dennoch sind Schulnoten Indikatoren für mathematische Kompetenz und daher auch für die Modellierungskompetenz, welche – wie in Abbildung 51 entsprechend angedeutet – eine der Facetten mathematischer Kompetenz darstellt. Das TEMOD-Verfahren hat die Zielsetzung, die Modellierungskompetenz möglichst genau zu erfassen. Dennoch wird eingeräumt, dass aufgrund von Messfehlern, Schwierigkeiten bei der Operationalisierung, eventuell noch unbekannten Störvariablen etc. keine perfekte Abdeckung des Konstrukts möglich ist, wohl aber eine bessere Erfassung als dies mithilfe anderer Verfahren realisiert werden kann.

Die hohe Korrelation zwischen Testverfahren und Schulnoten kann als erster Indikator für die Validität des TEMOD-I Verfahrens betrachtet werden, insofern als es anscheinend Aspekte der mathematischen Kompetenz misst.

Welche Aspekte der mathematischen Kompetenz sind dies aber im Einzelnen? Ist es tatsächlich Modellierungskompetenz, die hier gemessen wird? Die Betrachtungen zu den anderen beiden Außenkriterien legen dies nahe. Modellierungskompetenz wird benötigt, um Textaufgaben zu lösen. Auch bei Textaufgaben handelt es sich um Leistungsdaten, daher sind diese ebenso wie das Testverfahren in Abbildung 51 durch Kreise dargestellt. Schulnoten und Lehrerbeurteilungen hingegen durch Ellipsen, da sie zum einen auf verschiedenen Datengrundlagen beruhen, zum anderen nicht als voneinander unabhängig betrachtet werden können. Die Fähigkeit zum Lösen von Textaufgaben kann als abhängige Variable betrachtet werden, die von den verschiedenen Teilkompetenzen des Modellierens beeinflusst ist. Je besser die einzelnen Teilkompetenzen beherrscht werden, desto höher ist die Erfolgswahrscheinlichkeit (siehe auch Abschnitt 4.2.3 zur Überprüfung dieser Theorie). Textaufgaben sind daher das wichtigste Außenkriterium zur Validierung des Verfahrens. Der hohe Zusammenhang zwischen der Fähigkeit zum Lösen von Textaufgaben einerseits und dem Gesamtscore andererseits kann daher als weiterer Indikator für die Validität des Verfahrens betrachtet werden. Als letztes Kriterium wurde die Beurteilung der Modellierungskompetenz durch die Lehrkräfte hinzugezogen. Aufgrund der Tatsache, dass die hierfür verwendeten Items eine 1:1 Umsetzung der Unterkompetenzen darstellen, kann der Zusammenhang zwischen diesen Urteilen und den TEMOD-I Items gleichzeitig als ein Maß für die Güte der Operationalisierung des Konstrukts Modellierungskompetenz durch eben diese Items betrachtet werden. Der deutliche Effekt dieses Zusammenhangs bestätigt diese Güte einerseits, andererseits fungiert er als dritter Indikator für die Kriteriumsvalidität des Verfahrens.

Dass die Korrelation mit Schulnoten kleiner ist als die mit Textaufgaben und diese wiederum geringer als die mit den Urteilen, ist darüber hinaus ebenfalls theoriekonform: Der Zusammenhang wird höher mit zunehmender Nähe des Außenkriteriums zum zu untersuchenden Konstrukt.

4.3.2.3.2 Konstruktvalidität

Die konfirmatorischen Faktorenanlysen legen nahe, dass auch die Konstruktvalidtät des Verfahrens als gegeben betrachtet werden kann. Dies lässt sich zum einen aus der Tatsache ableiten, dass alle Items signifikant auf ihren Referenzfaktoren laden und die Zuordnung insofern eindeutig ist, dass bei allen Items (außer S06, welches selektiert wurde) keine Nebenladungen auf andere Faktoren in den Modifkationsindizes angezeigt werden. Die verbleibenden 35 Items messen also 4 unterschiedliche Teilkompetenzen, denen sie sich auch eindeutig zuordnen lassen.

Auch sprechen die Kriterien aus den Fit-Indizes zum größeren Teil für die Güte des Modells, auch wenn der CFI in dieser Hinsicht eine Ausnahme darstellt.

Die hohe Korrelation zwischen *Mathematisieren* einerseits und *Interpretieren* andererseits ließ es interessant scheinen, der Frage nachzugehen, ob ein gemeinsamer Faktor für beide Teilkompetenzen einen besseren Modellfit zur Folge hat, sprich: die Daten besser beschreiben kann. Dieser verändert sich jedoch im revidierten Modell nicht – was wenig verwunderlich ist, bedenkt man, dass aus statistischer Sicht eine Gleichsetzung zweier Faktoren, die ohnehin fast perfekt miteinander korrelieren, eine kaum merkliche Veränderung im Modell darstellt.

Trotzdem liegt es nahe, sich mit der Frage zu beschäftigen, ob *Mathematisieren* und *Interpretieren* tatsächlich als zwei unterschiedliche Teilkonstrukte aufgefasst werden sollten, wenn die Empirie beide für identisch befindet. Grundsätzlich lassen sich für beide Perspektiven – und somit auch für beide Messmodelle – stützende Argumente finden.

Mathematisieren beschreibt den Übersetzungsvorgang von Informationen aus dem realen Leben in die mathematische Sprache der Ziffern, Zeichen, Symbole und Operatoren, oder anders ausgedrückt: die Übertragung von einem Symbolsystem in ein anderes. *Interpretieren* beschreibt nichts anderes als den genau umgekehrten Prozess. In beiden Fällen geht es also um Übersetzungsprozesse, die sich nicht inhaltlich voneinander unterscheiden, sondern sich lediglich spiegelbildlich zueinander verhalten. Dies erklärt die hohe Korrelation zwischen beiden Faktoren und spricht dafür, beide Teilkompetenzen als eine zu behandeln.

Andererseits darf nicht vergessen werden, dass sich die aus den empirischen Ergebnissen abgeleiteten Informationen nicht tatsächlich auf die kognitiven Prozesse beziehen, sondern auf die Operationalisierung dieser kognitiven Prozesse, in diesem Fall in Form der verwendeten Items. Die Items zu *Mathematisieren* und *Interpretieren* unterscheiden sich in ihrer Form lediglich dadurch voneinander, dass beim *Mathematisieren* die mathematischen Sprachelemente in den Antwortalternativen befinden, beim *Interpretieren* hingegen im Itemstamm. Ob die Testpersonen jedoch die Items in der vorgegebenen Reihenfolge lösen, sprich: Die im Itemstamm in einer Sprache formulierte Information erfassen und in eine andere Sprache übersetzen und diese dann mit den Antwortalternativen vergleichen oder ob sie eher nach dem Ausschlussverfahren vorgehen und die einzelnen Antwortalternativen übersetzen, um sie mit dem Inhalt des Itemstammes zu vergleichen, kann aus den Ergebnisdaten nicht geschlossen werden.

Beim Einsatz von Multiple-Choice-Items kann sich der Modellierer also aussuchen, welchen Übersetzungsweg er gehen möchte, während beim Modellieren „in vivo" keine Alternativen vorgegeben sind und man daher immer von vorne nach hinten arbeiten muss. Es kann deshalb nicht ausgeschlossen werden, dass sich hinter der scheinbaren Identität von *Mathematisieren* und *Interpretieren* ein Artefakt verbirgt.

Aufschluss geben könnten hierzu Studien, bei denen man Testpersonen beim Bearbeiten solcher Multiple-Choice-Aufgaben laut denken lässt um dann zu vergleichen, welche Unterschiede sich in den kognitiven Prozessen beim *Mathematisieren* im Vergleich zum *Interpretieren* ergeben.

Da sich aus der Trennung beider Teilkompetenzen keine Nachteile ergeben (selbst die Testökonomie lässt sich kaum merklich verbessern, wenn man nur drei statt vier Faktoren untersucht), liegt es nahe, das Vier-Faktoren-Modell beizubehalten, solange unklar ist, was sich hinter dem hohen Zusammenhang zwischen beiden Teilkompetenzen verbirgt.

Die Tatsache, dass das Vier-Faktoren-Modell in der Literatur zu Modellieren als etabliert gilt, sollte jedoch nicht als Grundlage für eine Rechtfertigung dieses Modells dienen dürfen. Im Gegenteil lässt es zu wünschen übrig, dass sich die gesamte theoretische Diskussion und die sich aus ihr ergebenden praktischen Anwendungen auf ein Modell stützen, das nie empirisch auf seine Gültigkeit überprüft wurde.

Die Sinnhaftigkeit des Modellierungszyklus ließe sich noch weiter anzweifeln, betrachtet man die großen Interkorrelationen zwischen einzelnen Teilkompetenzen.

Strukturieren und *Mathematisieren* / *Interpretieren* haben immerhin auch 86,3% gemeinsame Varianz. Macht eine unabhängige Betrachtung der verschiedenen Teilkompetenzen überhaupt noch Sinn, wenn die Wahrscheinlichkeit extrem hoch ist, dass jemand der in einem der vier Bereiche Schwierigkeiten hat, in den anderen ebenfalls gefördert werden muss?

An dieser Stelle gilt es zu bedenken, dass die Zusammenhänge hoch, aber eben nicht perfekt sind. Abgesehen von der Übereinstimmung zwischen *Mathematisieren* und *Interpretieren* gibt es zwischen den Konstrukten eben auch Varianzanteile, die sich nicht gemeinsam erklären lassen. Eben diese Anteile bilden die Grundlage für intraindividuelle Unterschiede, die zu unterschiedlichen Leistungsprofilen führen können, bei denen eine Teilkompetenz stärker ausgeprägt ist als die andere. Der hohe Zusammenhang – sogar zwischen *Mathematisieren* und *Interpretieren* – bleibt darüber hinaus ein statistischer und kein deterministischer, sprich: es ist anzunehmen, dass es sehr wohl Personen gibt, bei denen eine Kompetenz sehr gut, die andere überhaupt nicht beherrscht wird. Bei eben diesen Personen lohnt sich eine gezielte Förderung. Und sie aus der Gruppe von denen herauszufinden, die ein homogenes Leistungsprofil aufweisen, ist mit dem vorliegenden Verfahren relativ einfach. Abbildung 50 verdeutlichte diesen Sachverhalten anhand einiger zufällig ausgewählter Probanden. Denn dass Modellierungskompetenz im Allgemeinen diagnostiziert und bei entsprechenden Mängeln gefördert werden muss, steht außer Frage und ändert sich auch nicht durch die offensichtliche Vermengung der theoretisch postulierten Teilkonstrukte. Als redundant kann ein Faktor nämlich nur dann angesehen werden, wenn man aufgrund der Werte der anderen Faktoren das Abschneiden des Probanden sicher prädizieren kann (Moosbrugger & Hartig, 2002) und dies ist bei der Modellierungskompetenz nicht der Fall.

4.3.2.4 Nebengütekriterien

Neben den Hauptgütekriterien Objektivität, Reliabilität und Validität gibt es bei der Beurteilung eines Testverfahrens auch noch sieben Nebengütekriterien zu beachten, auf welche im Folgenden kurz eingegangen werden soll (Kubinger, 2003).

Das Gütekriterium der *Skalierung* bezieht sich darauf, dass hinter der beim Messen vorgenommen Zuordnung von einem empirischen zu einem numerischen Relativ ein mathematisch-statistisches Messmodell zu stehen hat. Dies bezieht sich z.B. auf die im Rahmen der klassischen Testtheorie nicht überprüfte Annahme, dass ein Summenscore überhaupt als adäquater Indikator für die Fähigkeitsausprägung einer Person angesehen werden kann. Dahinter steckt die Frage nach der

Dimensionalität eines Verfahrens: Das Bilden von Summenscores ist letztlich nur dann gerechtfertigt, wenn alle Items das gleiche latente Konstrukt messen. Beim TEMOD Verfahren haben die konfirmatorischen Faktorenanalysen gezeigt, dass das Unterscheiden von *Mathematisieren* und *Interpretieren* statistisch gesehen eventuell nicht sinnvoll ist. Aus inhaltlichen und vor allem auch aus didaktischen Gründen ist diese Trennung aber dennoch angezeigt.

Bevor das vorliegende Testverfahren tatsächlich zuverlässig zur Diagnostik von Modellierungskompetenz eingesetzt werden kann, ist eine *Normierung* erforderlich – möglichst an einer repräsentativen, zumindest jedoch an einer spezifisch repräsentativen Stichprobe. Spezifisch repräsentativ meint hierbei (im Vergleich zu global repräsentativ), dass zwar nicht alle Merkmale sich so verteilen wie in der Population aber zumindest die für die Fragestellung (bekanntlich) relevanten Merkmale (Bortz & Döring, 2002). Dies betrifft in diesem Fall z.B. den – wie oben bereits erwähnt – durch die Selbstselektion der Schulen eventuell verzerrten Anteil nichtdeutscher Schüler in der Untersuchungsstichprobe.

Was die *Testökonomie* angeht, so wurde an anderer Stelle bereits ausführlich darauf eingegangen, dass durch den hier verwendeten diagnostischen Zugang in kürzester Zeit eine Fülle an präzisen und förderrelevanten Informationen gewonnen werden kann, wie es bisher im deutschsprachigen Raum und in Bezug auf die anvisierte Zielgruppe nicht möglich ist.

Auch die *Nützlichkeit* des Verfahrens ist nicht in Zweifel zu ziehen, betrachtet man die im theoretischen Teil diskutierten Aspekte der Wichtigkeit mathematischen Modellierens. Es darf an dieser Stelle nicht vergessen werden, dass sich gerade aus der Diagnostik als solche wiederum neue Ansatzpunkte für die Didaktik der Modellierungskompetenz ergeben. Dies kann beispielsweise darin begründet liegen, dass aus einem differenzierten Leistungsprofil einer Klasse auch differenzierter Unterricht effektiv realisiert werden kann, quasi im Sinne einer individuellen Förderung von verschiedenen Subgruppen, deren Stärken und Schwächen sich jeweils voneinander unterscheiden.

Kubinger (2003, S. 202) definiert *Zumutbarkeit* als die Gestaltung eines Testes dahingehend, dass „die Tp physisch sowieso und psychisch in Bezug auf ihre energetisch-motivationale und emotionale Beanspruchung geschont wird.“ Vergleicht man den Test zur Erfassung von Modellierungskompetenzen mit einer für die Altersgruppe üblichen Klassenarbeit in Mathematik, so kann davon ausgegangen werden, dass das Testverfahren von vergleichbarer Zumutbarkeit ist. Dies resultiert aus dem ähnlichen Aufbau der Items einerseits und aus der Länge des Tests andererseits. Für Grundschulkinder dürfte zwar das Multiple-Choice-Format zur Leistungserfassung eine relativ neue Methode darstellen, mit der Testpersonen jedoch leicht vertraut gemacht werden können.

Was die *Unverfälschbarkeit* des Verfahrens angeht, so ist diese schon allein deshalb als tendenziell eher unproblematisch zu betrachten, als dass es sich hier um einen Leistungstest handelt, der bekanntlich nur nach unten verfälscht werden kann, beispielsweise bei mangelnder Motivation der Testperson. Es gilt daher, zum einen möglichst abwechslungsreiche Items einzusetzen und zum anderen den Schülern im Sinne eines „informed consent“ (Bortz & Döring, 2002) zu verdeutlichen, warum eine zuverlässige Diagnostik für sie von so großer Relevanz ist.

Lediglich die *Fairness* ist kritisch zu betrachten: Die Items sind (abgesehen vom Bereich „*Verarbeiten*“) deutlich textlastig und benachteiligen daher eventuell Probanden nichtdeutscher Herkunft, denen es weniger am Textverständnis als vielmehr am Sprachverständnis mangelt. Da eine nichtsprachliche Form des Verfahrens schwer denkbar ist, kann zum jetzigen Zeitpunkt nur die Empfehlung gegeben werden, entsprechende Variablen bei der Interpretation der Ergebnisse einschränkend zu berücksichtigen.

Alles in allem fällt die Beurteilung des Verfahrens in Bezug auf die Gütekriterien positiv aus, womit für den Einsatz des TEMOD in der Praxis die entsprechende Rechtfertigungsgrundlage geschaffen wäre.

4.3.3 Der Modellierungszyklus als normatives Modell

Seit einigen Jahren wird in der Bildungsforschung der Bedarf nach Betrachtung der Klassen- und Schulebene zur Erklärung von Schülerleistungen akzeptiert. Bedenkt man, dass für die Entwicklung der Lesekompetenz die Schule, in der ein Schüler sich befindet, weitaus mehr Varianz erklären kann als die individuellen Hintergrundvariablen Familienhintergrund, Alter und Geschlecht zusammengenommen (Mortimore, Sammons, Stoll, Lewis & Ecob, 1988), so wird deutlich, dass es nicht mehr zeitgemäß sein kann, Leistungsdaten im Bereich der Schul- und Bildungsforschung nur auf Individualebene zu betrachten (Helmke & Weinert, 1997). Auch im Bereich der Mathematikleistung konnten Effekte der Klassen- bzw. Schulebene nachgewiesen werden (Young, 1998; Webster & Fisher, 2000; Schreiber, 2002).

Diesem Umstand wurde hier Rechnung getragen, eine Vorgehensweise, die sich a posteriori dadurch rechtfertigen lässt, dass es in der Tat auf Klassenebene Varianz in der Modellierungskompetenz gibt. Die Erklärung dafür liegt auf der Hand: Da Schüler in einer Schulklasse gemeinsam (also unter gleichen zeitlichen, räumlichen und situativen Bedingungen sowie durch die gleiche Lehrkraft) ihre Modellierungskompetenz erworben haben ist es naheliegend, dass es bezüglich des Leistungsniveaus innerhalb einer Klasse weniger Unterschiede gibt, als zwischen den Klassen (Schwetz, 2003).

Die Ergebnisse auf der Basis der einfachen Regressionen zeigen, dass ein Random-Intercept-Modell die Daten passend beschreibt, da hierbei dem Umstand Rechnung getragen wird, dass die Klasse bei der Erklärung der Modellierungsleistung eine Rolle spielt. Die Random-Slope-Modelle mussten hingegen erwartungsgemäß verworfen werden. Die Bedeutung der einzelnen Teilkompetenzen ist über alle Klassen hinweg annähernd gleich.

Gleichzeitig konnte durch die multiple Regression gezeigt werden, dass jede einzelne Teilkompetenz einen wichtigen Beitrag für die Modellierungsleistung liefert. Da die β-Gewichte der einzelnen Teilkompetenzen alle in einem ähnlichen Bereich liegen, kann davon ausgegangen werden, dass letztlich auch auf die Förderung jeder einzelnen Kompetenz gleich viel Wert zu legen ist.

Darüber hinaus kann der durch das Modell erklärte Varianzanteil als großer Effekt bezeichnet werden. Daraus kann gefolgert werden, dass in der Tat jede dieser Kompetenzen beherrscht werden muss, um erfolgreich modellieren zu können. Dies bestätigt nicht nur den Modellierungszyklus in seiner theoretisch postulierten Form an sich, sondern auch seine Funktion als normatives Modell. Andererseits zeigt jedoch auch der große Anteil an nicht erklärter Varianz, dass es neben den Teilkompetenzen noch eine Reihe von anderen Variablen geben muss, die für das erfolgreiche Modellieren relevant sind. Über diese können lediglich Vermutungen angestellt werden. Arbinger (1997) nennt drei Arten von Wissen, die zum Lösen von mathematischen Textaufgaben von Bedeutung sind. Neben bereichsbezogenem Wissen (was analog zur Verarbeitungskompetenz angesehen werden kann) und strategischem Wissen (was sich z.B. in den Strukturierungskompetenzen widerspiegelt) ist auch metakognitives Wissen vonnöten. Dieses kann aufgefasst werden als Wissen um den Modellierungszyklus, welches zentral für die adäquate Steuerung des eigenen Vorgehens beim Modellieren zu betrachten ist. Vielleicht – aber hierbei begibt man sich deutlich in den Bereich der Spekulation – gehen gute Modellierer eher den direkten Weg entlang des Modellierungszyklus und weisen weniger Abweichungen auf, wie sie in Studie 1 untersucht wurden.

5. Zusammenfassende Diskussion und Ausblick

Der in Kapitel 1 dargelegte theoretische und empirische Hintergrund zu Modellieren lässt einerseits deutlich werden, welch tragende Rolle Modellierungskompetenz spielt und wie wichtig es ist, diese bereits frühzeitig zu diagnostizieren und entsprechend zu fördern. Gleichzeitig wird aus großen Forschungsprojekten wie IGLU (Walther et al., 2003) und PISA (PISA – Konsortium Deutschland, 2001) deutlich, dass die Modellierungskompetenz bei einem großen Teil der deutschen Schüler nicht so entwickelt ist, wie dies wünschenswert wäre. Andererseits zeigt sich beim Studium der Literatur zum Themenkreis um Modellieren und Modellierungskompetenzen aber auch ein erheblicher Mangel an empirischen Studien zur Untermauerung der die Bildungsforschung beherrschenden Theorien.

Die meisten Arbeiten stammen aus dem pädagogischen oder mathematisch-didaktischen Bereich, weshalb man die in der psychologischen Forschung üblichen Verfahren und Vorgehensweisen weithin vermisst. Zum einen hapert es daher oft bereits an der theoretischen Fundierung von Konstrukten. Man denke z.B. an die Bildungsstandards, wo Problemlösen und Modellieren zwei unterschiedliche Dimensionen von mathematischen Kompetenzen beschreiben sollen. Man kann jedoch Modellieren als Problemlösen auffassen, auch wenn nicht umgekehrt jede Art des mathematischen Problemlösens auch als Modellieren bezeichnet werden kann, da es auch mathematische Probleme gibt, die nicht im Bezug zur Alltagswelt stehen, z.B. das Beweisen eines mathematischen Satzes. Es wäre durchaus möglich, Bildungsstandards vor dem Hintergrund kognitionspsychologischer Theorien zu formulieren, dies ist jedoch nicht der Fall. Neben dieser Freiheit von Theorie kommt erschwerend noch die Freiheit an Empirie hinzu: So sind die Bildungsstandards eben leider auch nicht empirisch belegt – z.B. durch Faktorenanalysen in ihrer Dimensionalität bestätigt. Gleiches gilt auch wiederum für Modellieren:

Der Modellierungszyklus wird in der Literatur als gängiges Modell gehandelt, um Modellierungsprozesse beim Lösen von mathematischen Problemen zu beschreiben, womit ihm eine deskriptive Funktion zuteil wird. Die Studien zur Überprüfung dieser deskriptiven Funktion sind dünn gesät und stützen sich auf Untersuchungsmethoden, die nur als eingeschränkt valide betrachtet werden können (z.B. Borromeo Ferri, 2006).

Andererseits dient der Modellierungszyklus aber auch als didaktische Grundlage, wenn Lernenden Modellierungskompetenzen vermittelt werden sollen. In den zahlreichen (v.a. auf Universitätsniveau, z.B. Geng, 2003) existierenden Modellierungskursen wird vermittelt, dass die im Modellierungszyklus formulierten Stadien und Prozesse nacheinander abgearbeitet werden müssen, um zu einer optimalen Lösung zu gelangen. Dies aber macht den Modellierungszyklus gleichzeitig auch zu einem normativen Modell. Hierfür aber fehlt nun tatsächlich jedwede empirische Grundlage.

Letztlich besteht darüber hinaus eine Lücke bei der diagnostischen Betrachtung von Modellierungskompetenzen. Zwar wurde von Haines und Kollegen (z.B. Haines & Crouch, 2001) herausgearbeitet, dass das unabhängige Erfassen der Teilkompetenzen mithilfe von Multiple-Choice-Items die Methode der Wahl bei der Diagnostik von Modellierungskompetenzen darstellt – was sicherlich als Durchbruch zu werten ist – gleichzeitig erbringt das von den Autoren entwickelte und von vielen anderen eingesetzte und für eigene Zwecke adaptierte Instrument nicht den Nachweis der Validität. Zu zeigen gilt, dass mit solchen Items tatsächlich auch Modellierungskompetenz gemessen wird.

Gleichzeitig beschränken sich die mit dem Haines-Verfahren unternommenen diagnostischen Bemühungen allein auf den sekundären und tertiären Bildungsbereich. Dabei muss klar sein, dass gerade eine besonders frühe Förderung wichtig und zielführend ist.

Aus diesen drei offen bleibenden Fragestellungen leiteten sich die Zielsetzungen der vorliegenden Arbeit ab, nämlich die Prüfung des Modellierungszyklus auf Eignung als deskriptives sowie als normatives Modell und die Entwicklung und Validierung von Items zur Erfassung von Modellierungskompetenz am Ende der vierten Jahrgangsstufe.

Die Ergebnisse der ersten Studie deuten darauf hin, dass der Modellierungszyklus in seiner bisherigen Form nicht als angemessenes deskriptives Modell fungieren kann, da einerseits die Teilkompetenz *Validieren* scheinbar realiter nicht zum Einsatz kommt und anderseits die Reihenfolge der Stadien vielmehr ein rekursives Modell nahelegt. Ein neues – an die Datenlage angepasstes – Modell wurde daher vorgeschlagen. Fraglich bleibt, worin die beobachteten interindividuellen Unterschiede in der Reihenfolge der Bearbeitung von Teilkompetenzen zu begründen sind. Dies zu untersuchen ist zunächst besonders deshalb schwierig, da noch keine

Hypothesen darüber vorliegen, welche Variablen es sein könnten, die das Vorgehen beim Modellieren beeinflussen. Weitere und intensivere Untersuchungen mithilfe Protokollen lauten Denkens könnten eingesetzt werden um diesbezüglich Hypothesen zu generieren und anschließend zu überprüfen.

Die Eignung als normatives Modell hingegen ließ sich durch Analysen aufgrund der in Studie 2 erhobenen Daten bestätigen. Insofern rechtfertigt die Empirie die Existenz des Modellierungszyklus in seiner theoretisch postulierten Form. Wie die mehrebenenanalytische multiple Regression zeigt, sind die vier Teilkompetenzen in der Form, wie sie durch das entwickelte Testverfahren erfasst werden, jeweils von signifikanter Bedeutung für die Erklärung der Lösungswahrscheinlichkeit.

Kritisch anmerken lässt sich sicherlich, dass hierbei per se keine Aussagen über die Reihenfolge der Teilprozesse getroffen werden. Selbst wenn man zeigen kann, dass die schrittweise Aufnahme von Teilkompetenzen als Prädiktoren in eine Regressionsgleichung jeweils einen deutlichen Zuwachs in R^2 zur Folge haben (eine entsprechende Analyse findet sich im Anhang, sie muss jedoch aufgrund der Vernachlässigung der Mehrebenenstruktur mit Vorsicht interpretiert werden), so zeigt dies zwar, dass jede zusätzlich aufgenommene Teilkompetenz einen Erklärungswert über denjenigen der anderen Teilkompetenzen hinaus aufweist und bestätigt so gewissermaßen die inkrementelle Validität des Modells in sich, sagt jedoch noch nichts darüber aus, ob auch die Modellierungsprozesse an sich in dieser Reihenfolge ablaufen müssen, um in einer optimalen Lösung zu resultieren.

Hier sind weitere Studien notwendig, die sich gezielt der Frage annehmen, ob die im deskriptiven Modell des Modellierungszyklus formulierten rekursiven Prozesse einen Einfluss auf die Qualität der Aufgabenlösung haben und wenn ja, ob dieser positiver oder negativer Natur ist.

Der nicht durch die Modellierungskompetenzen zu erklärende Anteil der Varianz in den Textaufgaben weist jedoch auch darauf hin, dass es anscheinend weitere – bisher noch nicht bekannte – Variablen gibt, die das Ergebnis des Problemlöseprozesses beeinflussen. Denkbar ist zum Beispiel der Einfluss von Gedächtnisleistung. Typische Strategien, wie das Vorgehen beim schriftlichen Dividieren, sind im Langzeitgedächtnis abgespeichert und müssen bei der Bearbeitung von konkreten Aufgaben abgerufen und eingesetzt werden. Auch das Einmaleins, welches als Grundlage für viele Aufgaben in der vierten Jahrgangsstufe gilt, wurde von den Schülern zuvor wie ein Gedicht auswendig gelernt. Schwierigkeiten beim Abruf dieses verbalen Codes können zwar theoretisch wettgemacht werden, indem man sich das Ergebnis herleitet, dürften jedoch den Problemlöseprozess trotzdem beeinträchtigen (Imbo & Vandierendonck, 2007).

Des Weiteren gibt es Untersuchungen, die sich dem Phänomen „Math Anxiety“, also Angst vor Mathematik widmen. Auch diese könnte beim Modellieren insofern eine Rolle spielen, dass selbst beim grundsätzlichen Beherrschen aller Teilkompetenzen der tatsächliche Lösungsprozess durch Math Anxiety beeinträchtigt wird. Denn diese wirkt bekanntermaßen vermittelt über das Arbeitsgedächtnis: Angst braucht kognitive Kapazität, die dann entsprechend in geringerem Maße zum Lösen der Primäraufgabe zur Verfügung steht (Ashcraft & Kirk, 2001). Wenn dies schon bei einfachen Rechenaufgaben gilt, so kann vermutet werden, dass dies bei komplexen Modellierungsaufgaben erst recht der Fall ist, da gerade das Abschreiten der diversen Stadien und Prozesse entlang des Modellierungszyklus metakognitive Prozesse erfordern. Neben Angst ist auch die von der Person selbst wahrgenommene Schwierigkeit beim Lernen von Mathematik als Einflussfaktor auf Mathematikleistungen (im Bereich der Arithmetik) identifiziert worden (Karzmark, 2009).

Inwieweit auch Persönlichkeitseigenschaften für die interindividuellen Unterschiede in mathematischen Leistungstest eine Rolle spielen können, ist schwer zu beantworten, da man bisher vergeblich nach entsprechenden Studien sucht. Lediglich eine Untersuchung von Hogan & Parlapiano (2008) beschäftigt sich mit dem Zusammenhang von „Quantitative Estimation Skill“ (was jedoch wiederum nur wenig Überschneidung mit Mathematik im engeren Sinn hat) und den Faktoren des 16 PF, wobei sich lediglich ein Zusammenhang mit dem Faktor Independence (in der deutschen Version entspricht dies den Faktor E: Soziale Anpassung vs. Selbstbehauptung), der sich jedoch theoretisch nur schwer erklären lässt. Auch im Bereich Persönlichkeit sind also weitere Studien vonnöten um sich Klarheit über eventuelle weitere Einflussgrößen zu verschaffen.

Zunächst aber verschaffen die für die Hauptuntersuchung entwickelten Items eine Möglichkeit, Modellierungskompetenzen von Grundschulabgängern objektiv, reliabel und valide zu messen. Weitere Untersuchungen sind allerdings nötig, um einen aus den Items resultierenden Test zu normieren, damit Einzelergebnisse besser interpretierbar werden.

Damit ist zunächst eine diagnostische Grundlage geschaffen, die dann als Ausgangsbasis für eine individuelle Förderung von Schülern betrachtet werden kann.

Hier sind dann die Lehrkräfte gefragt, z.B. durch differenzierten Unterricht gezielt auf die im Ergebnisprofil deutlich gewordenen Stärken und Schwächen der Schüler einzugehen.
Die Bedeutung von Lehrkräften darf aber keinesfalls dergestalt unterschätzt werden, dass angenommen wird, diese kämen erst zu diesem Zeitpunkt ins Spiel. Im Gegenteil wird bei den in Kapitel 4.2.3 berichteten Mehrebenenanalysen an dem auf Klassenebene erklärbaren Varianzanteil deutlich, dass die Zugehörigkeit zu einer bestimmten Schulklasse einen signifikanten Einfluss auf die Modellierungskompetenz eines Schülers hat. Eine auf Klassenebene variierende Variable ist die der jeweiligen Lehrkraft. Es kann daher vermutet werden, dass es eine Rolle spielt, wie gut oder schlecht Lehrer in der Lage sind, Modellierungskompetenzen entsprechend zu vermitteln.
Eine Befragung des Bildungsbarometers (Jäger-Flor & Jäger, 2008) zeigt allerdings, dass 31,5% aller Schüler ihren Mathematiklehrern die Note 4 oder schlechter auf ihre Fähigkeit, mathematische Sachverhalte verständlich zu erklären, geben würden. Um Schülern Modellierungskompetenzen zu vermitteln, muss also bereits bei der Professionalisierung von Lehrkräften begonnen werden. Diese müssen gezielt ausgebildet sein dahingehend, was unter Modellierungskompetenzen zu verstehen ist und wie man diese erkennen und entsprechend fördern kann.

Literatur

Aebli, H. (1976). *Grundformen des Lehrens: Eine allgemeine Didaktik auf kognitionpsychologischer Grundlage.* Stuttgart: Klett.

Arbinger, R. (1997). *Psychologie des Problemlösens: Eine abwendungsorientiere Einführung.* Darmstadt: Primus.

Arendasy, M., Sommer, M. & Glück, J. (2004). Dimensionalitaet und differenzielle Validitaet von Textaufgaben. Zum Einfluss von Bearbeitungsstrategien. *Zeitschrift für pädagogische Psychologie, 18*, 231-243.

Ashcraft, M. H. & Kirk, E. P. (2001). The Relationships Among Working Memory, Math Anxiety, and Performance. *Journal of Experimental Psychology: General, 130*(2), 224-237.

Barbosa, J. C. (2003). What is mathematical modelling? In S. Lamon, W. Parker & K. Houston (Hrsg.), *Mathematical modelling: A way of life. ICTMA 11* (S. 227-234). Chichester: Horwood Publishing.

Baumann, U. & Perrez, M. (2005). *Lehrbuch Klinische Psychologie, Psychotherapie: Klassifikation, Diagnostik, Ätiologie, Intervention.* Bern: Huber.

Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Tillmann, K.-J. & Weiß, M. (2000). *Internationales und nationales Rahmenkonzept für die Erfassung von mathematischer Grundbildung in PISA.* Berlin: Max-Planck-Institut für Bildungsforschung.

Baumert, J., Lehmann, R., Lehrke, M., Schmitz, B., Clausen, M., Hosenfeld, I., Köller, O. & Neubrand, J. (1997). *TIMSS - Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde.* Opladen: Leske & Budrich.

Beck, B. & Klieme, E. (2007). Einleitung. In B. Beck & E. Klieme (Hrsg.), *DESI: Sprachliche Kompetenzen. Konzepte und Messung* (S. 1-8). Weinheim: Beltz.

Bentler, P. M. & Chou, C.-P. (1987). Practical issues in structural equation modeling. *Sociological Methods and Research, 16*(1), 78-117.

Bermejo, V. (1996). Cardinality development and counting. *Developmental Psychology, 22*, 263-268.

Berry, J. S. & Le Masurier, D. (1984). O.U. students do it by themselves. In J. S. Berry, D. N. Burghes, I. D. Huntley, D. J. G. James & A. O. Moscardini (Hrsg.), *Teaching and applying mathematical modelling* (S. 48-85). Chichester: Ellis Horwood.

Bjorklund, D. E. & Rosenblum, K. E. (2001). Children's use of multiple and variable counting strategies in a game context. *Developmental Science, 4*, 183-193.

Blomhoj, M. & Jensen, T. H. (2003). Developing mathematical modelling competence: conceptual clarification and educational planning. *Teaching Mathematics and its Applications, 22*(3), 123-139.

Blomhoj, M. & Jensen, T. H. (2007). What's all the fuss about competencies? In W. Blum, P. L. Galbraith, H. Henn & M. Niss (Hrsg.), *Modelling and Applications in Mathematics Education: The 14th ICMI Study* (S. 45-56). Berlin: Springer.

Blum, W. (2003). ICMI Study 14: Applications and Modelling in Mathematics Education. *Educational Studies in Mathematics, 51*, 149-171.

Blum, W., Galbraith, P. L., Henn, H. & Niss, M. (2007). *Modelling and Applications in Mathematics Education: The 14th ICMI Study.* Berlin: Springer.

Blum, W., Neubrand, M., Ehmke, T., Senkbeil, M., Jordan, A., Ulfig, F. & Carstensen, C. H. (2004). Mathematische Kompetenz. In P.-K. Deutschland (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland - Ergebnisse des zweiten internationalen Vergleichs* (S.47-92). Münster: Waxmann.

Blum W. & Leiss, D. (2007). How do students and teachers deal with mathematical modeling problems. The example “Filling up”. In C. Haines & W. Blum (Hrsg.), *Mathematical Modelling (ICTMA 12): Education, Engineering and Economics* (S. 1622-1633). Chichester: Horwood.

Bollen, K. A. (1989). *Structural equation modeling with latent variables*. New York: Wiley.

Borromeo Ferri, R. (2006). Theoretical and empirical differentiations of phases in the modelling process. *Zentralblatt für Didaktik der Mathematik, 38*(2), 86-95.

Borromeo Ferri, R. & Kaiser, G. (2006). Perspektiven zur Modellierung im Mathematikunterricht - Analysen aktueller Ansätze, *Beiträge zum Mathematikunterricht 2006 in Osnabrück* (S. 50-52). Hildesheim: Franzbecker.

Borromeo Ferri, R., Leiss, D. & Blum, W. (2006). Der Modellierungskreislauf unter kognitionspsychologischer Perspektive, *Beiträge zum Mathematikunterricht 2006 in Osnabrück* (S. 53-55). Hildesheim: Franzbecker.

Bortz, J. & Döring, N. (2002). *Forschungsmethoden und Evaluation*. Berlin: Springer

Brase, G. L. (2002). "Bugs" built into the system: How privileged representations influence mathematical reasoning across the lifespan. *Learning and Individual Differences, 12*(2002), 391-409.

Brown, T.A. (2006). *Confirmatory factor analysis for applied research*. New York: The Guilford Press

Bühler, K. (1907). Tatsachen und Probleme einer Psychologie der Denkvorgänge. *Archiv für Psychologie, 9*, 297-305

Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion. 2. Auflage.* München: Pearson Studium.

Busse, A. & Kaiser, G. (2003). Context in application and modelling - an empirical approach. In Q. Ye, W. Blum, K. Houston & Q. Jiang (Hrsg.), *Mathematical Modelling in Education and Culture: ICTMA 10 :* . Chichester: Horwood Publishing.

Bryant, P. & Nuñes, T. (2002). Children's Understanding of Mathematics. In U. Goswhani (Hrsg.), *Blackwell handbook of childhood cognitive development* (S. 412-439). Malden: Blackwell.

Camp, B. W., Gaston E. Blom, Hebert, F. & Doorninck, W. J. v. (1977). "Think Aloud": A Program for Developing Self-Control in Young Aggressive Boys. *Journal of Abnormal Child Psychology, 5*(2), 157-169.

Caron, F. & Bélair, J. (2007). Exploring university stundents' competencies in modelling. In C. R. Haines, P. L. Galbraith, W. Blum & S. Khan (Hrsg.), *Mathematical modelling. ICTMA 12. Education, Engineering and Economics:* (S.120-129). Chichester: Horwood Publishing.

Carraher, T. N., Carraher, D. & Schliemann, A. D. (1985). Mathematics in the streets and in the schools. *British Journal of Developmental Psychology, 3*, 21-29.

Chomsky, N. (1968). *Language and Mind*. New York: Harcourt, Brace & World

Crouch, R. M. & Haines, C. R. (2003). *Do you know which students are good mathematical modellers? Some research developments. Technical Report No. 83.* University of Hertfordshire: Department of Physics, Astronomy and Mathematics.

Crouch, R. M. & Haines, C. R. (2004). Mathematical modelling: transitions between the real world and the mathematical model. *International Journal of Mathematical Education in Science and Technology, 35*(2), 197-206.

Dahaene, S. (1997). *Number Sense.* Oxforfd: University Press.

destatis. (2008). *Ausländeranteil sinkt mit steigendem Bildungslevel*. Wiesbaden: Statistisches Bundesamt.

Ditton, H. (1998). *Mehrebenenanalyse*. Weinheim: Juventa

Domahs, F., Delazer, M. & Nuerk, H. (2006). What makes multiplication facts difficult? Problem size or neighborhood consistency? *Experimental Psychology, 53*(4), 275-182

Donker, A. & Markopoulos, P. (2002). A comparison of think-aloud, questionnaires and interviews for testing usability with children. In X. Faulkner, J. Finlay & F. Detienne (Hrsg.), *People and computers XVI: Memorable yet invisible. Proceedings of HCI 2000* (S. 305-316). London: Springer.

Dossey, J. A. (1992). The nature of Mathematics: Its role and its influence. In D.A. Grouws (Hrsg.), *Handbook of research on mathematics teaching and learning* (S. 39-48). New York: Macmillan

Dunne, T. & Galbraith, P. L. (2003). Mathematical modelling as pedagogy - impact of an immersion program. In Q. Ye, W. Blum, K. Houston & Q. Jiang (Hrsg.), *Ma-*

thematical Modelling in Education and Culture: ICTMA 10 (S. 16-30). Chichester: Horwood Publishing.

English, L. (2003). Mathematical modeling with young learners. In S. Lamon, W. Parker & K. Houston (Hrsg.), *Mathematical modelling: A way of life. ICTMA 11* (S. 3-18). Chichester: Horwood Publishing.

Ericsson, K. A. & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*(3), 215-250.

Ericsson, K. A. & Simon, H. A. (1993). *Protocol analysis: verbal reports as data.* Cambridge: MIT Press.

Flaherty, E. G. (1975). The thinking aloud technique and problem soving ability. *Journal of Educational Psychology, 68*(6), 223-225.

Fritz, A. & Ricken, G. (2005). Früherkennung von Kindern mit Schwierigkeiten im Erwerb von Rechenfähigkeiten. In M. Hasselhorn, H. Marx & W. Schneider (Hrsg.), *Diagnostik von Mathematikleistungen. Tests und Trends - Neue Folge Band 4* (S. 5-28). Göttingen: Hogrefe.

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., Schatschneider, C. & Fletcher, J. M. (2006). The Cognitive Correlates of Third-Grade Skill in Arithmetic, Algorithmic Computation, and Arithmetic Word Problems. *Journal of Educational Psychology, 98*(1), 29-43.

Fuson, K. C. (1988). *Children's counting and concept of numers.* New York: Springer.

Galbraith, P. L. (2007). Beyond the low hanging fruit. In W. Blum, P. L. Galbraith, H. Henn & M. Niss (Hrsg.), *Modelling and Applications in Mathematics Education: The 14th ICMI Study.* (S. 79-88). Berlin: Springer.

Gallistel, C. R. & Gelmann, R. (2005). Mathematical Cognition. In K. Holyoak & R. Morrison (Hrsg.), *The Cambridge handbook of thinking and reasoning.* (S. 559-588). Cambridge: Cambridge University Press.

Gelman, R. (2000). The epigenesis of mathematical thinking. *Journal of Applied Developmental Psychology*(21), 27-37.

Gelmann, R. & Gallistel, C. R. (1978). *The child's understanding of number.* Cambridge, MA: Harvard Unoversity Press.

Geng, W. (2003). "How to model mathematically" table and its applications. In Q. Ye, W. Blum, K. Houston & Q. Jiang (Hrsg.), *Mathematical modelling in education and culture: ICTMA 10* (S. 67-80).

Gölitz, D., Roick, T. & Hasselhorn, M. (2006). Deutscher Mathematiktest für vierte Klassen (DEMAT 4). Göttingen: Beltz Test.

Goldstein, H. (1999). Multilevel Statistical Models. London: Institute of Education.

Gollwitzer, M. & Jäger, R. S. (2007). *Evaluation.* Weinheim: BeltzPVU.

Grassmann, M., Klunter, M., Köhler, E., Mirwald, E., Raudies, M. & Thiel, O. (2002). Mathematische Kompetenzen von Schulanfängern. *Potsdamer Studien zur Grundschulforschung, 30*.

Greefrath, G. (2006). Prozessanalysen von Modellierungsaufgaben, *Beiträge zum Mathematikunterricht 2006 in Osnabrück* (S. 223-226). Hildesheim: Franzbecker.

Greeno, J. G., Moore, J. L. & Smith, D.R. (1993). Transfer of situated learning. In D. Detterman & R. J. Sternberg (Hrsg.), *Transfer on trial: Intelligence, Cognition and Instruction* (S. 99-167). Norwood: Aplex.

Greer, B. (1992). Multiplication and division as models of situations. In D. A. Grouws (Hrsg.), *Handbook of research on mathematics teaching and learning* (S. 276-295). New York: Macmillan.

Greer, B., Verschaffel, L. & Mukhopadhyay, S. (2007). Modelling for life: Mathematics and children's experience. In W. Blum, P. L. Galbraith, H. Henn & M. Niss (Hrsg.), *Modelling and Applications in Mathematics Education: The 14th ICMI Study* (S. 89-98). Berlin: Springer.

Grüßing, M. (2002). Wieviel Raumvorstellung braucht man für Raumvorstellungsaufgaben? Strategien von Grundschulkindern bei der Bewältigung räumlich-geometrischer Anforderungen. *Zentralblatt für Didaktik der Mathematik, 34*(2), 37-45.

Haffner, J., Baro, K., Parzer, P. & Resch, F. (2005). *Heidelberger Rechentest. Erfassung mathematischer Basiskompetenzen im Grundschulalter.* Göttingen: Hogrefe.

Haines, C. R. & Crouch, R. M. (2001). Recognizing constructs within mathematical modelling. *Teaching Mathematics and it's Applications, 20*(3), 129-138.

Haines, C. R. & Crouch, R. M. (2005). Applying mathematics: making multiple-choice questions work. *Teaching Mathematics and its Applications, 24*(2-3), 107-133.

Haines, C. R., Crouch, R. M. & Davis, J. (2000). *Mathematical Modelling Skills: A Research Instrument. Technical Report No. 55.* University of Hertfordshire: Department of Mathematics.

Haines, C. R., Crouch, R. M. & Davis, J. (2001). Understanding Student's Modellings Skills. In J. F. Matos, W. Blum, K. Houston & S. P. Carreira (Hrsg.), *Modelling and Mathematics Education* (S. 366-380). Chichester: Horwood Publishing.

Haines, C. R., Crouch, R. M. & Fitzharris, A. (2003). Deconstructing mathematical modelling: Approaches to problem solving. In Q. Ye, W. Blum, K. Houston & Q. Jiang (Hrsg.), *Mathematical Modelling in Education and Culture: ICTMA 10* (S. 41-53). Chichester: Horwood Publishing.

Haladyna, T. M. (2004). *Developing and validating multiple-choice test items.* Mahwah: Lawrence Erlbaum.

Haladyna, T. M., Downing, S. M. & Rodriguez, M. C. (2002). A review of multiple-choice-item-writing guidelines for classroom assessment. *Applied measurement in education, 15*(3), 309-333.

Hall, G. G. (1984). The assessment of modelling projects. In J. S. Berry, D. N. Burghes, I. D. Huntley, D. J. G. James & A. O. Moscardini (Hrsg.), *Teaching and applying mathematical modelling* (S. 173-148). Chichester: Ellis Horwood.

Hasemann, K. & Mansfield, H. (1995). Concept mapping in research on mathematical knowledge development: background, methods, findings and conclusions. *Educational Studies in Mathematics, 29*, 45-72.

Hasemann, K. & Stern, E. (2002). Die Förderung des mathematischen Verständnisses anhand von Textaufgaben – Ergebnisse einer Interventionsstudie in Klassen des 2. Schuljahres. *Journal für Mathematik-Didaktik, 2002*(3/4), 222-242

Hasselhorn, M., Marx, H. & Schneider, W. (2005). Diagnostik von Mathematikleistungen, -kompetenzen und - schwächen: Eine Einführung. In M. Hasselhorn, H. Marx & W. Schneider (Hrsg.), *Diagnostik von Mathematikleistungen. Tests und Trends - Neue Folge Band 4* (S. 1-4). Göttingen: Hogrefe.

Heck, R. H. (1998). Factor analysis: Exploratory and confirmatory approaches. In G. A. Marcoulides (Hrsg.), *Modern methods for business research* (S. 177-215). Mawah, London: Lawrence Erlbaum Accociates.

Hegarty, M., Mayer, R. E. & Monk, C.A. (1995). Comprehension of Arithmetic word problems: a comparison of successful and unsuccessful problem solvers. *Journal of Educational Psychology, 87*(1), 18-32

Helmke, A. & Weinert, F.E. (1997). Bedingungsfaktoren schulischer Leistung. In Weinert, F.E. (Hrsg.), *Enzyklopädie der Psychologie. Psychologie des Unterrichts und der Schule. Serie I. Band 3* (S. 71-176). Göttingen: Hogrefe

Henning, H. & Keune, M. (2007). Levels of modelling competencies. In W. Blum, P. L. Galbraith, H. Henn & M. Niss (Hrsg.), *Modelling and Applications in Mathematics Education. The 14th ICMI Study.* (S. 225-232). Berlin: Springer.

Hogan, T. P. & Parlapiano, C. A. (2008) Personality factors related to quantitative estimation skill: confirmation and extension. *Psychological Research, 103,* 189-198

Hosenfeld, I. (2005). Bildungsstandards auf dem Prüfstand – Der Nutzen von Vergleicharbeiten. *Schulmanagement*, *5*, 8-10

Houston, K. (2007). Assessing the "phases" of mathematical modelling. In W. Blum, P. L. Galbraith, H. Henn & M. Niss (Hrsg.), *Modelling and Applications in Mathematics Education: The 14th ICMI Study*(S. 249-256) *.* Berlin: Springer.

Houston, K. & Neill, N. (2003a). Assessing modelling skills. In S. Lamon, W. Parker & K. Houston (Hrsg.), *Mathematical modelling: A way of life. ICTMA 11* (S. 155-164). Chichester: Horwood Publishing.

Houston, K. & Neill, N. (2003b). Investigating students' modelling skills. In Q. Ye, W. Blum, K. Houston & Q. Jiang (Hrsg.), *Mathematical Modelling in Education and Culture: ICTMA 10* (S. 54-66). Chichester: Horwood Publishing.

Hussy, W., Jain, A. (2002). Experimentelle Hypothesenprüfung in der Psychologie. Göttingen: Hogrefe

Ideka, T. & Stephens, M. (1998). The influence of problem format on student’s approaches to mathematical modeling. In P. Galbraith, W. Blum, G. Booker, I. Huntley (Hrsg.), Mathematical Modelling, Teaching and Assessment in a Technology-Rich World. Chichester: Horwood Publishing (S. 223-232).

Imbo, I. & Vandierendonck, A. (2007). Do multiplication and division strategies rely on executive and phonological working memory resources? *Memory & Cognition, 35*(7), 1759-1771.

Izard, J., Haines, C. R., Crouch, R. M., Houston, K. & Neill, N. (2003). Assessing the impact of teaching mathematical modelling: Some implications. In S. Lamon, W. Parker & K. Houston (Hrsg.), *Mathematical modelling: A way of life. ICTMA 11* (S. 165-178). Chichester: Horwood Publishing.

Jäger, R. S. (1983). *Der diagnostische Prozess. Eine Diskussion psychologischer und methodischer Randbedingungen.* Göttingen: Hogrefe.

Jäger, R. S. & Petermann, F. (Hrsg.). (1999). *Psychologische Diagnostik. 4. Auflage*. Weinheim: BeltzPVU.

Jäger-Flor, D. & Jäger, R.S. (2008). Bildungsbarometer zum Thema „Mathematik“. Landau: Verlag Empirische Pädagogik

Jensen, T. H. (2007). Assessing mathematical modelling competency. In C. R. Haines, P. L. Galbraith, W. Blum & S. Khan (Hrsg.), *Mathematical Modelling. ICTMA 12. Education, Engineering and Economics* (S. 141-147). Chichester: Horwood Publishing.

Kaiser, G. & Schwarz, B. (2006). Modellierungskompetenzten - Entwicklung im Unterricht und ihre Messung, *Beiträge zum Mathematikunterricht 2006 in Osnabrück* (S. 56-58). Hildesheim: Franzbecker.

Kammermeyer, G. (2004). *Schulfähigkeit als Brücke zwischen Kindertagesstätte und Grundschule.*: Vortrag aus der Fachmesse 'kita bildet', 2003. Verfügbar unter: http://www.kita-bildet.de/downloads/Referat_Kammermeyer-1.pdf [29.04.08].

Karmiloff-Smith, A. (1992). The child as a Mathematician. In A. Karmiloff-Smith (Hrsg.), *Beyond modularity: A developmental perspective on cognitive science, Chapter 4*. Cambridge: MIT Press.

Karzmark, P. (2009). The effect of cognitive, personality, and background factors on the WAIS-III Arithmetic Subtest. *Applied Neuropsychology, 16*(1), 49-53

Kaufmann, L. & Nuerk, H.-C. (2006). Die Entwicklung des Rechnens und dessen Störrungen: Genese, Modelle, Diagnostik und Intervention. *Zeitschrift für Legasthenie und Dyskalkulie, 2006*(2), 11-16.

Kilpatrick, J. (1992). A history of research in mathematics education. In D. A. Grouws (Hrsg.), *Handbook of research on mathematics teaching and learning* (S. 3-38). New York: Macmillan.

Kim, K. H. (2007). A structural equation model of the factors impacting problem success of Algebra Word Problem: Reading, math, working memory and conceptual knowledge. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 68*(6-A), 2321?

Klauer, K. J. (1987). *Kriteriumsorientierte Tests. Lehrbuch der Theorie und Praxis lehrzielorientierten Messens.* Göttingen: Hogrefe.

Klieme, E. (2004). Was sind Kompetenzen und wie lassen sie sich messen? *Auszug aus Pädagogik, 6*, 10-13.

Klieme, E., Hartig, J. & Rauch, D. (2008). The concept of competence in educational contexts. In J. Hartig, E. Klieme & D. Leutner (Hrsg.), *Assessment of competencies in educational contexts* (S. 3- 22). Göttingen: Hogrefe

Klieme, E., Neubrand, M. & Lüdtke, O. (2001). Mathematische Grundbildung: Testkonzeption und Ergebnisse. In P.-K. Deutschland (Hrsg.), *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.

KMK. (2005). *Bildungsstandards im Fach Mathematik für den Hauptschulabschluss.* Neuwied: Luchterhand.

Köller, O. (2005). Neue Besen kehren gut: Das Institut zur Qualitaetsentwicklung im Bildungswesen soll die nationalen Bildungsstandards in Deutschland ueberpruefen. *Psychologie in Erziehung und Unterricht, 52*(4), 281-286.

Krajewski, K. (2003). *Vorhersagen von Rechenschwächen in der Grundschule.* Hamburg: Dr Kovac.

Krajewski, K. (2005). Vorschulische Mengenbewusstheit von Zahlen und ihre Bedeutung für die Früherkennung von Rechenschwäche. In M. Hasselhorn, H. Marx & W. Schneider (Hrsg.), *Diagnostik von Mathematikleistungen. Tests und Trends - Neue Folge Band 4* (S. 49-70). Göttingen: Hogrefe.

Kubinger, K. D. (2003). Gütekriterien. In K. D. Kubinger & R. S. Jäger (Hrsg.), Schlüsselbegriffe der psycholigischen Diagnostik. Weinheim: BeltzPVU (S.195-204).

Lamon, S. (2003). Modelling in elementary school: Helping young students to see the world mathematically. In S. Lamon, W. Parker & K. Houston (Hrsg.), *Mathematical modelling: A way of life. ICTMA 11* (S. 19-36). Chichester: Horwood.

Lingefjärd, T. & Holmquist, M. (2005). To assess students' attitudes, skills and competencies in mathematical modeling. *Teaching Mathematics and its Applications, 24*(2-3), 123-133.

Lissmann, U. & Jäger, R. S. (2008). Multiple-Choice-Aufgaben. *Journal für Lehrerinnen- und Lehrerbildung, 8*(1), 45-50.

Lorenz, J. H. (2005). Diagnostik mathematischer Basiskompetenzen im Vorschulalter. In M. Hasselhorn, H. Marx & W. Schneider (Hrsg.), *Diagnostik von Mathematikleistungen. Tests und Trends - Neue Folge Band 4* (S. 29-48). Göttingen: Hogrefe.

Lösel, F. (1999). Persönlichkeitsdaten (Tests). In R. S. Jäger & F. Petermann (Hrsg.). *Psychologische Diagnostik. 4. Auflage.* (S. 362-380). Weinheim: BeltzPVU

Maaß, K. (2004). *Mathematisches Modellieren im Unterricht - Ergebnisse einer empirischen Studie.* Hildesheim: Franzbecker.

Maaß, K. (2006). What are modelling competencies? *Zentralblatt für Didaktik der Mathematik, 38*(2), 113-142.

Maaß, K. (2007). Modelling in class: What do we want students to learn? In C. R. Haines, P. L. Galbraith, W. Blum & S. Khan (Hrsg.), *Mathematical Modelling. ICTMA 12. Education, Engineering and Economics* (S. 63-79). Chichester: Horwood Publishing.

Mack, W. (2005). Entwicklung von Zählkompetenz und das Problem Dyskalkulie. *Forschung Frankfurt, 1*(2005), 32-35.

Markman, Arthur B. (1999). *Knowledge Representations*. Mahwah: Lawrence Erlbaum

McClelland, D. C. (1973). Testing for competence rather than intelligence. *American Psychologist, 28*, 1-14

Moosbrugger, H. & Hartig, J. (2002). Factor analysis in personality research: Some artefacts and their consequences for psychological assessment*. Psychologische Beiträge, 44*, S. 136-158

Moosbrugger, H. & Schermelleh-Engel, K. (2007). Explorative und konformatorische Faktorenanalyse. In. H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 307-324). Berlin: Springer.

Mortimore, P., Sammons, P., Stoll, L., Lewis, D. & Ecob, R. (1988). *School Matters*. London: Open Books.

Neubrand, M. (2003). “Mathematical literacy”/„Mathematische Grundbildung“. *Zeitschrift für Erziehungswissenschaft, 6*(3), 338-356.

Niss, M. (2004). Mathematical competencies and the learning of mathematics: The danish KOM project. In K. Gagatsis & S. Papastavridis (Hrsg.), *3rd Mediterranean Conference on Mathematical Education* (S. 115-124). Athens: Hellenic Mathematical Society and Cyprus Mathematical Society.

Novak, J. D. (1990). Concept mapping: A useful tool for science education. *Journal of Research in Science Teaching, 27*(10), 937-949.

Olson, G. M., Duffy, S. A. & Mack, R. L. (1984). Thinking-out-loud as a method for studying real-time comprehension process. In D. E. Kieras & M. A. Just (Hrsg.), *New methods in reading comprehension research* (S. 253-286). Hillsdale, NJ: Erlbaum.

Ottesen, J. T. (2001). Do not ask what mathematics can do for modelling. In D. Holton (Hrsg.), *The teaching and learning of mathematics at university level: an ICMI study* (S. 335-346). Dordrecht: Kluwer.

Penrose, O. (1978). How can we teach Mathematical Modelling? *Journal of Mathematical Modelling for Teachers, 1*(2), 31-42

PISA Konsortium Deutschland. (2001). *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich.* Opladen: Leske + Budrich.

PISA Konsortium Deutschland. (2004). *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland - Ergebnisse des zweiten internationalen Vergleichs.* Münster: Waxmann.

Pollak, H. (1979). The interaction between Mathematics and other school subjects. *New Trends in Mathematics teaching, Volume IV*: Paris: UNESCO

Pollak, H. (2003). A history of the teaching of modelling. In G. Stanic & J. Kilpatrick (Hrsg.), *A History of School Mathematics* (S. 647-671). Reston: NCTM.

Pollak, H. (2007). Mathematical modelling: A conversation with Henry Pollak. In W. Blum, P. L. Galbraith, H. Henn & M. Niss (Hrsg.), *Modelling and Applications in Mathematics Education: The 14th ICMI Study.* (S. 109-120). Berlin: Springer.

Rhymer, K. N. & Cates, G. L. (2006). Student Performance on and Preferences for Mathematics Word Problems: An Investigation of the Effects of Explicit Timing and Interspersing Procedures. *School Psychology Quarterly, 21*(2), 34-45.

Richter, S. & Brügelmann, H. (1994). Der Schulanfang ist keine Stunde Null. In H. Brügelmann & S. Richter (Hrsg.), *Wie wir schreiben lernen* (S. 62-78). Opladen: Leske & Budrich.

Riebel, J. & Jäger, R. S. (2008). Kompetenzen von Schulanfängern. *Psychologie in Erziehung und Unterricht, 55*(2), 132-143.

Rinkens, H. & Hönisch, K. (2005). Welt der Zahl 4. Braunschweig: Bildungshaus Schulbuchverlage

Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: a meta-analysis of 80 years of research. *Educational Measurement, 24*(2), 3-13.

Rost, J. (2004). *Lehrbuch Testtheorie - Testkonstruktion.* Bern: Verlag Hans Huber.

Rost, J. & Spada, H. (1978). Probabilistische Testtheorie. In K.-J. Klauer (Hrsg.), *Handbuch der pädagogischen Diagnostik* (S. 59-97). Düsseldorf: Schwann.

Russo, E. J., Johnson, E. J. & Stephens, D. L. (1989). The validity of verbal protocols. *Memory and Cognition, 17*(6), 759-769.

Sasaki, T. (2003). Recipient orientation in verbal report protocols: methodological issues in concurrent think-aloud. *Second Language Studies, 22*(1), 1-54.

Schermelleh-Engel, K., Moosbrugger, H. & Müller, H. (2003). Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research Online, 8*(2), 23-74.

Schermelleh-Engel, K. & Werner, C. (2007). Methoden der Reliabilitätsbestimmung. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 113-134). Berlin: Springer.

Schmid, I. (2008). *Personalisierung in Textaufgaben: Lässt sich das Arbeitsgedächtnis von Grundschulkindern durch die Personalisierung mathematischer Textaufgaben entlasten?* ?: Verlag Dr. Müller.

Schmidtke, A. (1999). Diagnostik als Inventarisieren. In R. S. Jäger & F. Petermann (Hrsg.), *Psychologische Diagnostik. 4. Auflage* (S. 263-268). Weinheim: BeltzPVU

Schott, F. (1999). Kriteriumsorientierte Diagnostik. In R. S. Jäger & F. Petermann (Hrsg.), *Psychologische Diagnostik. 4. Auflage* (S.226-231). Weinheim: BeltzPVU.

Schreiber, J. B. (2002). Institutional and student factors and their influence on advanced Mathematics achievement. *The Journal of Educational Research, 95* (5), 274-286

Schwetz, H. (2003). Die Klasse macht den Unterschied. Mehrebenenanalytische Untersuchungen der Effekte von Unterricht. Landau: Verlag Empirische Pädagogik

Schwetz, H. & Subramanian, S. V. (2005). Einführung in die Mehrebenenanalyse mit MLwin. Von der Regressionsanalyse zum Random-Slope-Modell. Landau: Verlag Empirische Pädagogik.

Seitz, K. & Schumann-Hengsteler, R. (2000). Mental multiplication and working memory. *European Journal of Cognitive Psychology, 12*(4), 552-570.

Shaffer, D. (2002). *Developmental Psychology.* Belmont: Wadsworth/Thompson Learning.

Simon, T. J. & Cabrera, A. (1995). Evidence for subitizing as a stimulus-limited processing phenomenon. Proceedings of the seventeenth annual conference of the Cognitive Science Society. Hillsdale: Erlbaum

Snijders, T.A.B. & Bosker, R.J. (1999). *Multilevel Analysis*. London: Sage

Sommer, M., Arendasy, M. & Glück, J. (2004). Mathematisches Denken bei Zweitklaesslern: Werden unterschiedlich schwierige Textaufgaben unterschiedlich verarbeitet? *Psychologie in Erziehung und Unterricht, 51*(2), 99-112.

Spada, H. (2005). *Lehrbuch Allgemeine Psychologie*. Bern: Huber

Uesaka, Y., Manolo, E. & Ichikawa, S. I. (2007). What kinds of perceptions and daily learning behaviors promote students` use of diagrams in problem solving. *Learning and Instruction, 17*, 322-335.

Usiskin, Z. (2007). The arithmetical operations as mathematical models. In W. Blum, P. L. Galbraith, H. Henn & M. Niss (Hrsg.), *Modelling and Applications in Mathematics Education. The 14th ICMi Study.* (S. 257-264). Berlin: Springer.

Verschaffel, L., Greer, B. & De Corte, E. (2000). *Making sense of word problems.* Lisse, NL: Swets and Zeitlinger.

von Aster, M., Weinhold Zulauf, M. & Horn, R. (2006). *ZAREKI-R - Neuropsychologische Testbatterie für Zahlenverarbeitung und Rechnen bei Kindern. Manual.* Frankfurt: Harcourt Test Services.

Walther, G., Geiser, H., Langeheine, R. & Lobemeier, K. (2003). Mathematische Kompetenzen am Ende der vierten Jahrgangsstufe. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, G. Walther & R. Valtin (Hrsg.), *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Klassenstufe im internationalen Vergleich*(S.189-226)*.* Münster: Waxmann.

Webb, N. (1992). Assessment of students' knowledge of Mathematics: Steps towards a theory. In D. A. Grouws (Hrsg.), *Handbook of research on mathematics teaching and learning* (S. 661-683). New York: Macmillan.

Webster, B. J. & Fisher, D.L. (2000). Accounting for variation in Science and Mathematics achievement: a multilevel analysis of Australian Data. *School Effectiveness and School Improvement, 11*(3), 339-360

Weidle, R. & Wagner, A. C. (1982). Die Methode des Lauten Denkens. In G. L. Huber & H. Mandl (Hrsg.), *Verbale Daten. Eine Einführung in die Grundlagen und Methoden der Erhebung und Auswertung* (S. 81- 103). Weinheim: Beltz.

Weinert, F.E. (2001). Vergleichende Leistungsmessung in Schulen - eine umstrittene Selbstverständlichkeit. In F.E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S.17-31). Weinheim: Beltz.

Weinhold Zulauf, M., Schweiter, M. & von Aster, G. (2003). Das Kindergartenalter: sensitive Periode für die Entwicklung numerischer Fähigkeiten. *Kindheit und Entwicklung, 12*, 222-230.

Weißhaupt, S., Peucker, S. & Wirtz, M. (2006). Diagnose mathematischen Vorwissens im Vorschulalter und Vorhersage von Rechenleistungen und Rechenschwierigkeiten in der Grundschule. *Psychologie in Erziehung und Unterricht, 53*, 236-245.

Wirtz, C. & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen.* Göttingen: Hogrefe.

Wirtz, N. & Nachtigall, C. (1998). *Deskriptive Statistik*. Weinheim: Juventa

Wynn, K. (1992). Addition and Subtraction by human infants. *Nature, 358*, 749-750.

Young, D. J. (1998). Rural and urban differences in student level achievement in Science and Mathematics: A multilevel analysis. *School Effectiveness and School Improvement, 9*(4), 386-418

Anhang

Anhang A: Verzeichnis der Abbildungen

Anhang B: Verzeichnis der Tabellen

Anhang C: Modellierungsaufgaben aus Studie 1 mit Operationalisierungsvorschriften

Anhang D: Instruktionen für Testleiter zur TEMOD – Testung (Studie 2)

Anhang E: REMOD – Ratingverfahren für Lehrkräfte

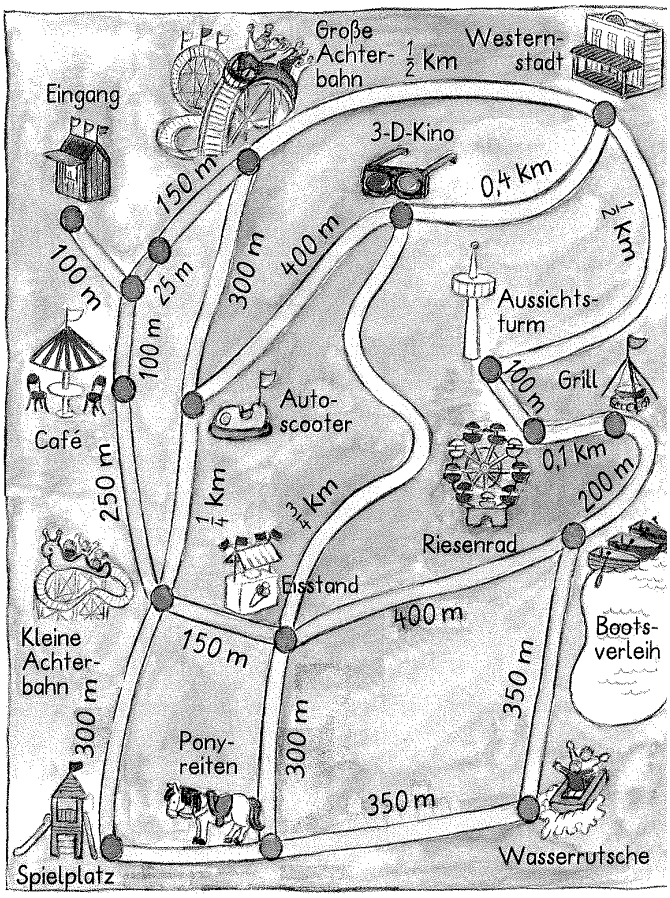
Anhang F: Sequentielle multiple Regression

ANHANG A:
Verzeichnis der Abbildungen

ANHANNG B:
Verzeichnis der Tabellen

ANHANG C:
Modellierungsaufgaben aus Studie 1 mit Operationalisierungsvorschriften

Freizeitpark



Uli fährt gerade auf der Wasserrutsche, als seine Mutter auf dem Handy anruft.
„Hallo Uli! Bitte komm auf dem kürzesten Weg zur kleinen Achter-bahn!“

Welchen Weg sollte Uli einschlagen? An welchen Stationen kommt er vorbei?

I: Strukturieren

VAR	Beschreibung allgemein	Aufgabenspezifisch
S1	Textverstehen / Problem in eigenen Worten wiedergeben	• Hinweise darauf, dass das Problem verstanden wurde = Uli befindet sich an Punkt X, seine Mutter an Punkt Y • Vp gibt den Text der Aufgabenstellung in eigenen Worten wieder
S2	Größen benennen: Gegeben / Gesucht	• Äußerungen darüber, welche Größen gegeben und gesucht sind (geht einen Schritt weiter als S1): Gesucht ist die Länge der Strecke X….
(S3	Beziehungen zwischen Variablen rekonstruieren)	
S4a	Suche nach problemrelevanten Informationen	• Suche der Ausgangs- und Zielposition auf der Karte • Suche nach der Strecke: Ponyreiten – Spielplatz (Sprechblase) • Zusammentragen von einzelnen Streckenlängen
S4b	Trennung zwischen relevanter – irrelevanter Information	• Unterscheidung von Zahlenwerten, die zur Problemlösung benötigt werden und solchen, die nicht benötigt werden – weiterer Verlauf der Ernte in drei Jahren wird als unnötig erkannt

II: Mathematisieren

VAR	Beschreibung allgemein	Aufgabenspezifisch
M1	Mathematisieren von Größen und Beziehungen (Zahlwörter => arabische Ziffern) (Verbalisierungen => Operationen)	• „Addieren“ von Streckenabschnitten wird ausdrücklich genannt • Identifikation der kürzesten Strecke: A < B. • Sprechblase => Umwandlung in arabische Ziffer (Nachfrage VL: Wie schreibt man das hin?) • Größenvergleich darüber, welche Zahl am größten ist
(M2	Deren Anzahl und Komplexität verringern)	
M3	Angebrachte mathematische Notationen / Skizzen (= formal korrekte Schreibweise)	• **Formal** korrektes Vorgehen bei der Anweisung zum schriftlichen Addieren der Einzelstrecken • Z.B. Zahlen richtig untereinander schreiben (noch nicht rechnen)

III: Verarbeiten

VAR	Beschreibung allgemein	Aufgabenspezifisch
V1	Mathematisches Wissen zur Problemlösung nutzen (z.B. korrekt schriftlich Rechnen)	• Einsatz einer **mathematischen** Lösungsstrategie, unabhängig von deren Qualität (also z.B. Rechenfehler) – nicht codieren, wenn z.B. Streckenlängen per Augenmaß verglichen werden, oder nur verbale Beschreibungen der Strecke erfolgen, ohne Längenvergleiche
V2	Heuristische Strategien („geschickt Rechnen“)	• Einsatz von Strategien, um den Lösungsprozess zu vereinfachen oder zu beschleunigen • Hier z.B. die 350 m des jeweils ersten Streckenabschnittes nicht mitzuverrechnen, da diese Information für alle drei Stecken gleich und somit für den Vergleich irrelevant ist

IV: Interpretieren

VAR	Beschreibung allgemein	Aufgabenspezifisch
I1	Interpretation im außermathematischen Kontext	• Formulierung eines Antwortsatzes, der sich auf die Ausgangsfrage bezieht • Formulierung dieser Antwort sprachlichen Format. Also: „Uli sollte zum Ponyreiten, dann zum….“ statt „die Kürzeste Stecke hat 950 m“
I2	Verallgemeinerung der Lösung (fak.)	• Formulierung einer Antwort, die vom konkreten Problem losgelöst und **allgemein** ausgedrückt ist • „Wenn man von X zu Y gelangen will, dann ist Strecke z, die beste Lösung, weil ….“
I3	Kommunikation der Ergebnisse in math. Sprache	• (Allgemeine) Reflektion über die **Lösung** in mathematischer Sprache (womit nicht die Nennung eines reinen Zahlenwerts als Lösung gemeint ist!!!!) = Bei I2 oder I2 werden mathematische Ausdrücke verwendet! • Sondern die Verwendung von korrekten mathematischen Ausdrücken in der **sprachlichen** Kommunikation der Ergebnisse • Ausdrücke wie „länger als“ (nicht: schneller als), „Strecke“ (nicht: Weg). Etc.

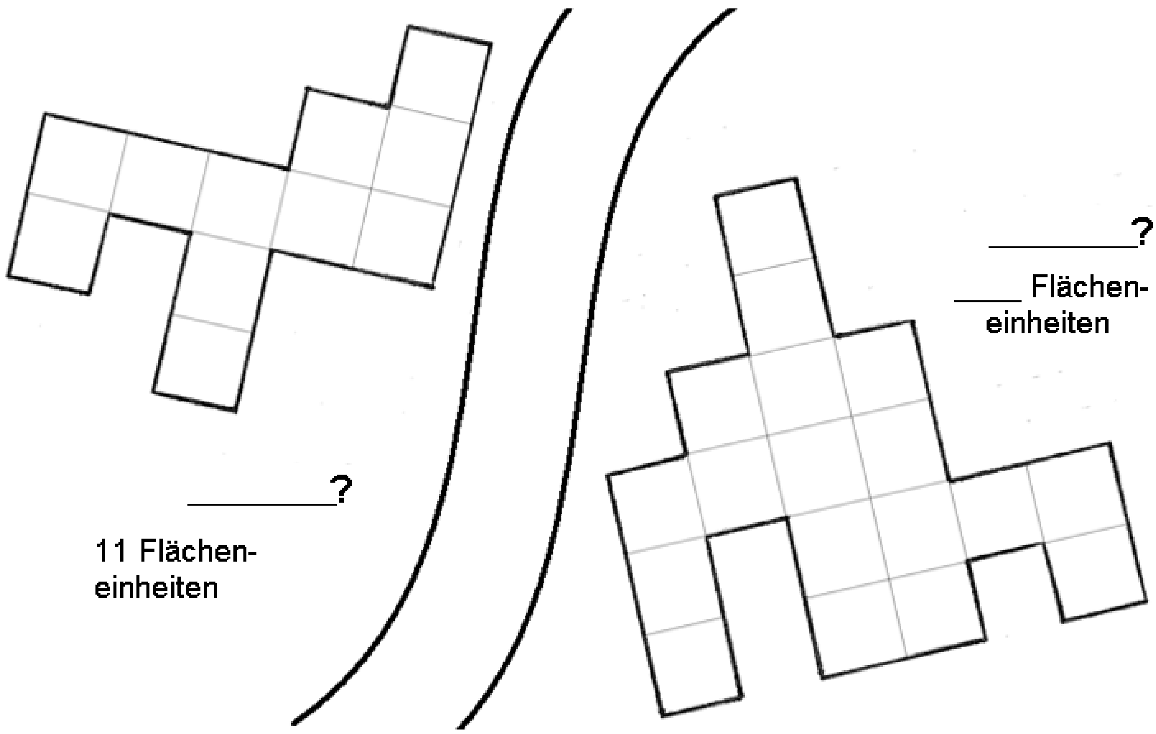
V: Validieren

VAR	Beschreibung	
	Beschreibung	**Aufgabenspezifisch**
E1	Innermathematische Validierung	• Jedwedes kritisches Hinterfragen der Richtigkeit eines Schrittes im Lösungsprozess – bezogen auf innermathematische Aspekte • Z.B. Gegenprobe, Regeln und Gesetze („Bei der Addition zweier dreistelliger Zahlen kann keine zweistellige Zahl herauskommen, ich muss mich also verrechnet haben).
E2	Außermathematische Validierung	• Jedwedes kritisches Hinterfragen der Richtigkeit eines Schrittes im Lösungsprozess – bezogen auf außermathematische Aspekte • Z.B. Überprüfung an einem entsprechenden Vorstellungsbild aus der Realität: „Kann es sein, dass innerhalb eines Freizeitparks zwei Punkte um 40km auseinander liegen?

Spanischer Ackerbau

Im schönen Spanien liegen die Dörfer Desastro und Fortuna. Beide leben vom Getreideanbau. Da Fortuna eine wesentlich größere Fläche zum Bebauen hat, bringt die Ernte immer weit mehr ein, als im kleinen Desastro.
Pro Flächeneinheit Ackerland erntet man in Fortuna 3 Säcke Getreide. Die Bauern aus Desastro setzen dieses Jahr einen neuen Dünger ein und ernten pro Flächeneinheit 5 Säcke Getreide. Nach drei Jahren werden es sogar 6 Säcke sein.

Kann Desastro dieses Jahr nun endlich mehr Getreide ernten als Fortuna?



I: Strukturieren

VAR	Beschreibung allgemein	Aufgabenspezifisch
S1	Textverstehen / Problem in eigenen Worten wiedergeben	• Hinweise darauf, dass das Problem verstanden wurde = Fortuna hat eine größere Anbaufläche, im Gegenzug hat Desastro eine bessere Ernte • Vp ordnet Namen korrekt der Grafik zu • Vp gibt den Text der Aufgabenstellung in eigenen Worten wieder
S2	Größen benennen: Gegeben / Gesucht	• Äußerungen darüber, welche Größen gegeben und gesucht sind (geht einen Schritt weiter als S1): Gesucht ist die Menge des jeweils geernteten Getreides
(S3	Beziehungen zwischen Variablen rekonstruieren)	
S4a	Suche nach problemrelevanten Informationen	• Abzählen der Flächeneinheiten in der Grafik • Suche nach Anzahl der Säcke pro m² aus dem Aufgabentext
S4b	Trennung zwischen relevanter – irrelevanter Information	• Unterscheidung von Streckeninformationen, die zur Problemlösung benötigt werden und solchen, die nicht benötigt werden

II: Mathematisieren

VAR	Beschreibung allgemein	Aufgabenspezifisch
M1	Mathematisieren von Größen und Beziehungen (Zahlwörter => arabische Ziffern) (Verbalisierungen => Operationen)	• Flächeneinheiten werden als arabische Ziffern ausgedrückt und hingeschrieben • Größenvergleich darüber, welche Zahl am größten ist
(M2	Deren Anzahl und Komplexität verringern)	
M3	Angebrachte mathematische Notationen / Skizzen (= formal korrekte Schreibweise)	• **Formal** korrektes Vorgehen bei der Anweisung zum schriftlichen Multiplizieren der Getreidenmenge • Z.B. Zahlen richtig untereinander schreiben (noch nicht rechnen)

III: Verarbeiten

VAR	Beschreibung allgemein	Aufgabenspezifisch
V1	Mathematisches Wissen zur Problemlösung nutzen (z.B. korrekt schriftlich Rechnen)	• Einsatz einer **mathematischen** Lösungsstrategie, unabhängig von deren Qualität (also z.B. Rechenfehler) – nicht codieren, wenn z.B. geschätzt oder geraten wird
V2	Heuristische Strategien („geschickt Rechnen“)	• Einsatz von Strategien, um den Lösungsprozess zu vereinfachen oder zu beschleunigen

IV: Interpretieren

VAR	Beschreibung allgemein	Aufgabenspezifisch
I1	Interpretation im außermathematischen Kontext	• Formulierung eines Antwortsatzes, der sich auf die Ausgangsfrage bezieht • Formulierung dieser Antwort im sprachlichen Format. Also: „Desastro kann endlich aufholen.“ statt „54 Säcke und 55 Säcke“
I2	Verallgemeinerung der Lösung (fak.)	• Formulierung einer Antwort, die vom konkreten Problem losgelöst und **allgemein** ausgedrückt ist • „Mehr Säcke pro Flächeneinheit anzubauen ist eine gute Methode…“ ODER „Wenn Fortuna diese neue Methode einsetzt, werden sie bald wieder aufgeholt haben“
I3	Kommunikation der Ergebnisse in math. Sprache	• (Allgemeine) Reflektion über die **Lösung** in mathematischer Sprache (womit nicht die Nennung eines reinen Zahlenwerts als Lösung gemeint ist!!!!) = Bei I2 oder I2 werden mathematische Ausdrücke verwendet! • Sondern die Verwendung von korrekten mathematischen Ausdrücken in der **sprachlichen** Kommunikation der Ergebnisse • Ausdrücke wie Menge, Fläche, etc.

V: Validieren

VAR	Beschreibung	
	Beschreibung	**Aufgabenspezifisch**
E1	Innermathematische Validierung	• Jedwedes kritisches Hinterfragen der Richtigkeit eines Schrittes im Lösungsprozess – bezogen auf innermathematische Aspekte • Z.B. Gegenprobe, Regeln und Gesetze („Bei der Multiplikation zweier ungerader Zahlen kann nicht wieder eine ungerade Zahl herauskommen, ich muss mich also verrechnet haben“).
E2	Außermathematische Validierung	• Jedwedes kritisches Hinterfragen der Richtigkeit eines Schrittes im Lösungsprozess – bezogen auf außermathematische Aspekte • Z.B. Überprüfung an einem entsprechenden Vorstellungsbild aus der Realität: „Kann es sein, dass beim Einsatz einer effektiveren Methode ein schlechteres Ergebnis herauskommt?“

Geburtstagparty

Die Zwillinge Rudi und Willi möchten für Ihre Geburtstagsfeier Spaghetti Bolognese kochen. Dazu fehlen ihnen noch folgende Zutaten:

- Nudeln
- Hackfleisch
- Tomaten
- Parmesankäse

Das Fleisch haben sie bereits beim Metzger für vier Euro gekauft. Nun kaufen sie im Supermarkt noch die restlichen Zutaten ein. Die gesamten Kosten wollen sie teilen. Was muss jeder bezahlen?

I: Strukturieren

VAR	Beschreibung allgemein	Aufgabenspezifisch
S1	Textverstehen / Problem in eigenen Worten wiedergeben	• Hinweise darauf, dass das Problem verstanden wurde = Zusammenstellen der Einzelzutaten • Hinweise, dass das Ziel der Aufgabe identifiziert wurde: Herauszufinden, was die Zutaten insgesamt kosten • Vp gibt den Text der Aufgabenstellung in eigenen Worten wieder
S2	Größen benennen: Gegeben / Gesucht	• Äußerungen darüber, welche Größen gegeben und gesucht sind (geht einen Schritt weiter als S1): Gesucht ist der zu zahlende Preis pro Person
(S3	Beziehungen zwischen Variablen rekonstruieren)	
S4a	Suche nach problemrelevanten Informationen	• Herauslesen der Preise aus dem Angebotsprospekt • Herauslesen der bereits gegebenen Preisinformation für Hackfleisch aus dem Aufgabentext
S4b	Trennung zwischen relevanter – irrelevanter Information	• Unterscheidung von benötigten Zutaten und weiteren Lebensmitteln im Angebotsprospekt, die nicht gebraucht werden

II: Mathematisieren

VAR	Beschreibung allgemein	Aufgabenspezifisch
M1	Mathematisieren von Größen und Beziehungen (Zahlwörter => arabische Ziffern) (Verbalisierungen => Operationen)	• Vier Euro aus Aufgabentext werden in 4 € übersetzt • Sammlung der Zutaten wird als Addition erkannt, Aufteilung des Preises als Division : 2
(M2	Deren Anzahl und Komplexität verringern)	
M3	Angebrachte mathematische Notationen / Skizzen (= formal korrekte Schreibweise)	• **Formal** korrektes Vorgehen bei der Anweisung zum schriftlichen Addition / Division • Z.B. Zahlen richtig untereinander schreiben (noch nicht rechnen)

III: Verarbeiten

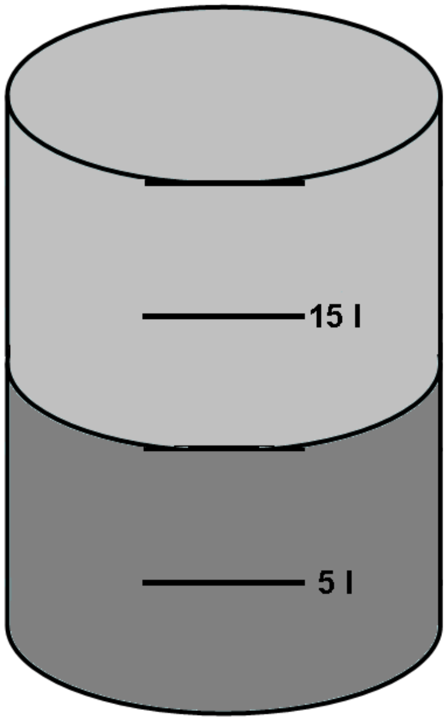
VAR	Beschreibung allgemein	Aufgabenspezifisch
V1	Mathematisches Wissen zur Problemlösung nutzen (z.B. korrekt schriftlich Rechnen)	• Einsatz einer **mathematischen** Lösungsstrategie, unabhängig von deren Qualität (also z.B. Rechenfehler) – nicht codieren, wenn z.B. geschätzt oder geraten wird
V2	Heuristische Strategien („geschickt Rechnen“)	• Einsatz von Strategien, um den Lösungsprozess zu vereinfachen oder zu beschleunigen

IV: Interpretieren

VAR	Beschreibung allgemein	Aufgabenspezifisch
I1	Interpretation im außermathematischen Kontext	• Formulierung eines Antwortsatzes, der sich auf die Ausgangsfrage bezieht • Formulierung dieser Antwort im sprachlichen Format. Also: „Jeder der Zwillinge bezahlt….“ statt „ X Euro“
I2	Verallgemeinerung der Lösung (fak.)	• Formulierung einer Antwort, die vom konkreten Problem losgelöst und **allgemein** ausgedrückt ist • „Zum Kochen des Gerichtes benötigt man Zutaten im Wert von X €“
I3	Kommunikation der Ergebnisse in math. Sprache	• (Allgemeine) Reflektion über die **Lösung** in mathematischer Sprache (womit nicht die Nennung eines reinen Zahlenwerts als Lösung gemeint ist!!!!) = Bei I2 oder I2 werden mathematische Ausdrücke verwendet! • Sondern die Verwendung von korrekten mathematischen Ausdrücken in der **sprachlichen** Kommunikation der Ergebnisse • Ausdrücke wie Menge, Fläche, etc.

Regenwasser

Familie Grün sammelt Regenwasser zum Gießen ihrer 13 Blumen. Die Abbildung rechts zeigt die 20-Liter-Tonne am Montagmorgen. Wird das Regenwasser bis Sonntagabend ausreichen, wenn Herr Grün jeden Abend einen Liter braucht?



I: Strukturieren

VAR	Beschreibung allgemein	Aufgabenspezifisch
S1	Textverstehen / Problem in eigenen Worten wiedergeben	• Hinweise darauf, dass das Problem verstanden wurde = Ist das Regenwasser ausreichend • Vp gibt den Text der Aufgabenstellung in eigenen Worten wieder
S2	Größen benennen: Gegeben / Gesucht	• Äußerungen darüber, welche Größen gegeben und gesucht sind (geht einen Schritt weiter als S1):Gesucht ist die benötigte Menge an Regenwasser, diese ist dann mit den 10l zu vergleichen
(S3	Beziehungen zwischen Variablen rekonstruieren)	
S4a	Suche nach problemrelevanten Informationen	• Ablesen des Wasserstandes aus der Skizze • Herauslesen der Informationen aus dem Text: 7 x 1l
S4b	Trennung zwischen relevanter – irrelevanter Information	• Erkennen, dass die Anzahl der Blumen nicht lösungsrelevant ist

II: Mathematisieren

VAR	Beschreibung allgemein	Aufgabenspezifisch
M1	Mathematisieren von Größen und Beziehungen (Zahlwörter => arabische Ziffern) (Verbalisierungen => Operationen)	• Jeden Abend einen Liter wird in 7 x 1l = 7l übersetzt • Verbrauch des Wassers wird als Subtraktion erfasst. Oder die Anzahl der gebrauchten Liter wird über Größenvergleich zu den 10 l Ausgangsmenge in Beziehung gesetzt
(M2	Deren Anzahl und Komplexität verringern)	
M3	Angebrachte mathematische Notationen / Skizzen (= formal korrekte Schreibweise)	• **Formal** korrektes Vorgehen bei der Anweisung zum schriftlichen Addieren / Subtrahieren • Z.B. Zahlen richtig untereinander schreiben (noch nicht rechnen)

III: Verarbeiten

VAR	Beschreibung allgemein	Aufgabenspezifisch
V1	Mathematisches Wissen zur Problemlösung nutzen (z.B. korrekt schriftlich Rechnen)	• Einsatz einer **mathematischen** Lösungsstrategie, unabhängig von deren Qualität (also z.B. Rechenfehler) – nicht codieren, wenn z.B. geschätzt oder geraten wird
V2	Heuristische Strategien („geschickt Rechnen“)	• Einsatz von Strategien, um den Lösungsprozess zu vereinfachen oder zu beschleunigen

IV: Interpretieren

VAR	Beschreibung allgemein	Aufgabenspezifisch
I1	Interpretation im außermathematischen Kontext	• Formulierung eines Antwortsatzes, der sich auf die Ausgangsfrage bezieht • Formulierung dieser Antwort im sprachlichen Format. Also: „Das Wasser reicht aus“ statt „3l“
I2	Verallgemeinerung der Lösung (fak.)	• Formulierung einer Antwort, die vom konkreten Problem losgelöst und **allgemein** ausgedrückt ist • „Wenn man jeden Tag 1l Wasser zum Giesen braucht, dann muss es in der Woche mehr als 7 Liter regnen“
I3	Kommunikation der Ergebnisse in math. Sprache	• (Allgemeine) Reflektion über die **Lösung** in mathematischer Sprache (womit nicht die Nennung eines reinen Zahlenwerts als Lösung gemeint ist!!!!) = Bei I2 oder I2 werden mathematische Ausdrücke verwendet! • Sondern die Verwendung von korrekten mathematischen Ausdrücken in der **sprachlichen** Kommunikation der Ergebnisse • Ausdrücke wie Menge, Fläche, etc.

V: Validieren

VAR	Beschreibung	
	Beschreibung	**Aufgabenspezifisch**
E1	Innermathematische Validierung	• Jedwedes kritisches Hinterfragen der Richtigkeit eines Schrittes im Lösungsprozess – bezogen auf innermathematische Aspekte • Z.B. Gegenprobe, Regeln und Gesetze („10 – 3 kann nicht 8 sein, da 8 + 3 nicht 10 ist“).
E2	Außermathematische Validierung	• Jedwedes kritisches Hinterfragen der Richtigkeit eines Schrittes im Lösungsprozess – bezogen auf außermathematische Aspekte • Z.B. Überprüfung an einem entsprechenden Vorstellungsbild aus der Realität: „Kann es sein, dass in einer Regentonne, die nur 20 Liter umfasst 25 Liter drinbleiben?“

Schokoladen - Wette

Eine Aktion des Pfälzer-Wald-Vereins:
Für jeden Beutel Nüsse, den ein Kind gesammelt hat, erhält es fünfzig Gramm Schokolade.

Tina und ihre Schwester Laura machen bei der Aktion mit. Nach 3 Stunden hat
Tina 7 Beutel mit Nüssen gesammelt, als Laura sagt:

Kann Tina schon ihre Nüsse eintauschen gehen, oder sollte sie noch weiter sammeln?

I: Strukturieren

VAR	Beschreibung allgemein	Aufgabenspezifisch
S1	Textverstehen / Problem in eigenen Worten wiedergeben	• Hinweise darauf, dass das Problem verstanden wurde = Hat Tina schon genug gesammelt? • Vp gibt den Text der Aufgabenstellung in eigenen Worten wieder
S2	Größen benennen: Gegeben / Gesucht	• Äußerungen darüber, welche Größen gegeben und gesucht sind (geht einen Schritt weiter als S1):Gesucht ist die Menge an Schokolade, die Tinas Sammlung entspricht
(S3	Beziehungen zwischen Variablen rekonstruieren)	
S4a	Suche nach problemrelevanten Informationen	• Zusammentragen der Information aus Aufgabentext und Sprechblase
S4b	Trennung zwischen relevanter – irrelevanter Information	• Erkennen, dass die vergangenen Stunden keine Rolle spielen.

II: Mathematisieren

VAR	Beschreibung allgemein	Aufgabenspezifisch
M1	Mathematisieren von Größen und Beziehungen (Zahlwörter => arabische Ziffern) (Verbalisierungen => Operationen)	• Fünfzig Gramm wird in 50 g übersetzt. • 50 g je Beutel wird als wiederholte Addition = Multiplikation erkannt • Aufgabenstellung wird als Größenvergleich erkannt
(M2	Deren Anzahl und Komplexität verringern)	
M3	Angebrachte mathematische Notationen / Skizzen (= formal korrekte Schreibweise)	• **Formal** korrektes Vorgehen bei der Anweisung zum schriftlichen Multiplizieren • Z.B. Zahlen richtig untereinander schreiben (noch nicht rechnen)

III: Verarbeiten

VAR	Beschreibung allgemein	Aufgabenspezifisch
V1	Mathematisches Wissen zur Problemlösung nutzen (z.B. korrekt schriftlich Rechnen)	• Einsatz einer **mathematischen** Lösungsstrategie, unabhängig von deren Qualität (also z.B. Rechenfehler) – nicht codieren, wenn z.B. geschätzt oder geraten wird
V2	Heuristische Strategien („geschickt Rechnen“)	• Einsatz von Strategien, um den Lösungsprozess zu vereinfachen oder zu beschleunigen

IV: Interpretieren

VAR	Beschreibung allgemein	Aufgabenspezifisch
I1	Interpretation im außermathematischen Kontext	• Formulierung eines Antwortsatzes, der sich auf die Ausgangsfrage bezieht • Formulierung dieser Antwort im sprachlichen Format. Also: „Tina kann aufhören“ statt „350 g“
I2	Verallgemeinerung der Lösung (fak.)	• Formulierung einer Antwort, die vom konkreten Problem losgelöst und **allgemein** ausgedrückt ist •
I3	Kommunikation der Ergebnisse in math. Sprache	• (Allgemeine) Reflektion über die **Lösung** in mathematischer Sprache (womit nicht die Nennung eines reinen Zahlenwerts als Lösung gemeint ist!!!!) = Bei I2 oder I2 werden mathematische Ausdrücke verwendet! • Sondern die Verwendung von korrekten mathematischen Ausdrücken in der **sprachlichen** Kommunikation der Ergebnisse • Ausdrücke wie Menge, Fläche, etc.

V: Validieren

VAR	Beschreibung	
	Beschreibung	Aufgabenspezifisch
E1	Innermathematische Validierung	• Jedwedes kritisches Hinterfragen der Richtigkeit eines Schrittes im Lösungsprozess – bezogen auf innermathematische Aspekte • Z.B. Gegenprobe, Regeln und Gesetze („36 ist kein Element der 7er-Reihe, ich muss mich verrechnet haben“).
E2	Außermathematische Validierung	• Jedwedes kritisches Hinterfragen der Richtigkeit eines Schrittes im Lösungsprozess – bezogen auf außermathematische Aspekte • Z.B. Überprüfung an einem entsprechenden Vorstellungsbild aus der Realität: „Kann es sein, dass Tina 20 kg Schokolade bekommt?“

Anhang D:
Instruktionen für Testleiter zur TEMOD – Testung (Studie 2)

TEMOD-Testung
Testverfahren zur **E**rfassung von **Mod**ellierungskompetenzen

Instruktionen für Testleiter/innen

Liebe/r Testleiter/in,

bei der Testung mit dem TEMOD-Verfahren ist es wichtig, dass alle getesteten Klassen unter den gleichen (also standardisierten) Bedingungen erfasst werden.

Halten Sie sich daher bei Ihren Testungen möglichst genau an die auf den folgenden Seiten dargestellten Anweisungen zur Testdurchführung.

1. Testtag

Die nachfolgende Tabelle gibt Ihnen einen Überblick über den Ablauf des ersten Testtages.

Minuten Testungszeit	Zu bearbeitende Aufgaben
0-10	Vorstellung / Instruktion
10-20	Block G
20-40	Block V
40-55	Block S
55-65	Pause
65-85	Block I
85-100	Block M

Stellen Sie sich zu Beginn der Testung kurz bei den Schüler/innen vor.

Teilen Sie den Schüler/innen zunächst die Testaufgaben aus und zwar so, dass die Versionen A und B immer abwechselnd ausgeteilt werden, damit die Schüler/innen nicht voneinander abschreiben. (Zu Ihrer Information: Die Versionen A und B erhalten exakt die gleichen Aufgaben, allerdings innerhalb der Blöcke jeweils in umgedrehter Reihenfolge.)

Erklären Sie den Schüler/innen dann mithilfe der folgenden Instruktion den Ablauf der Testung. Halten Sie sich dabei unbedingt an den Wortlaut der Instruktion. Nur so ist sichergestellt, dass Sie bei jeder Testung die Schüler/innen gleich instruieren.
„Bei unserer Studie geht es um Textaufgaben. Vielleicht habt ihr selbst manchmal Probleme Textaufgaben zu lösen, oder kennt Mitschüler, die Schwierigkeiten mit Textaufgaben haben. Wir werden heute eine ganze Menge von Aufgaben beantworten. Nur zwei davon sind normale Textaufgaben und ein paar sind ganz normale Rechenaufgaben, wie ihr sie aus dem Unterricht kennt.
Alle anderen Aufgaben sind etwas ungewöhnlich. Ihr müsst bei diesen Aufgaben nämlich absolut nichts rechnen!!! Ich sage dann dazu, wann diese Aufgaben anfangen.
Bei diesen Aufgaben müsst ihr immer nur die Texte lesen, dann sind vier mögliche Antworten gegeben. Nur eine davon ist jeweils richtig, die kreuzt ihr bitte an.
Es ist immer nur eine Antwort richtig. Kreuzt also bei jeder Aufgabe auch immer nur ein Kästchen an.

Ich zeige Euch das kurz mal (Letzte Seite hochhalten). Ihr lest hier die Aufgabe durch, da ist immer eine Frage enthalten. Dann schaut ihr welche vier Antworten richtig ist und kreuzt das Kästchen (zeigen) vor der richtigen Antwort an.

Wenn ihr eine Aufgabe nicht lösen könnt, weil sie zu schwer ist, oder ihr sie nicht versteht, dann geht bitte zur nächsten Aufgabe über. Wenn am Ende noch Zeit ist, könnt ihr wieder zu den schwereren Aufgaben zurückgehen.

Wenn ihr alle Aufgaben auf einer Seite bearbeitet habt, dann könnt ihr auf die nächste Seite blättern. Wenn aber am Ende der Seite so ein Stopp-Zeichen ist (hochhalten, zeigen), dann blättert bitte nicht um, sondern wartet, bis ich das Zeichen dazu gebe.

Wenn ihr eine Aufgabe nicht beantworten könnt oder nicht sicher seid, welches die richtige Antwort ist, dann kreuzt bitte diejenige an, von der ihr denkt, dass sie noch am ehesten richtig ist.

Wenn ihr Fragen habt, meldet Euch kurz, dann komme ich zu euch. Das hört sich vielleicht schwierig an, aber wir beginnen jetzt erstmal mit etwas, was ihr schon kennt.“

!!! Achtung bei Rückfragen während der Testung: Den Schüler/innen dürfen auf gar keinen Fall inhaltliche Tipps zum Lösen der Aufgaben gegeben werden. Antworten Sie bei inhaltlichen Fragen immer nur *„Ich weiß die Antwort auch nicht. Lies Dir genau den Text durch und versuche, es herauszufinden.“*
Sorgen Sie auch unbedingt dafür, dass die Lehrer/innen nicht helfen – wenn Sie ihnen erklären, dass das die Ergebnisse verfälscht, sind diese in der Regel kooperativ.

Instruieren Sie die Schüler zum ersten Aufgabenblock:
„Bitte blättert in Euren Tests alle auf die Seite 2 (Seite 2 hochhalten und zeigen!). Auf dieser Seite seht ihr ganz normale Textaufgaben. Versucht die Aufgaben zu bearbeiten.
Für Rechnungen könnt ihr das leere Blatt benutzen, das ich euch ausgeteilt habe. Schreibt in Eure Tests dann bitte nur die Antworten, nachdem ihr die Aufgabe ausgerechnet habt.
Wenn die erste Aufgabe Euch zu schwer ist, geht bitte zur nächsten Aufgabe über und beendet die erste, wenn danach noch Zeit ist.
Wenn ihr fertig seid, wartet ab, bis ich sage, dass die nächsten Aufgaben drankommen. Bitte blättert vorher nicht um.
Ihr habt 10 Minuten Zeit. Los!"
(Starten Sie die Stoppuhr und warten Sie 10 Minuten ab!).
Gehen Sie während dieser 10 Minuten durch die Reihen und kontrollieren Sie, ob die Schüler sich so verhalten, wie sie instruiert wurden und beantworten Sie eventuelle Fragen.

Nach Ablauf der 10 Minuten sagen Sie:
„Stopp! Legt bitte alle die Stifte einen Moment weg. Wir gehen gleich zum nächsten Aufgabenblock über!"

Instruieren Sie die Schüler nun für den zweiten Aufgabenblock (Block I).
„Jetzt dürft ihr umblättern auf die nächste Seite. Die Aufgaben, die jetzt kommen sind ganz normale Rechenaufgaben, wie ihr sie sicher aus dem normalen Unterricht kennt. Der einzige Unterschied ist, dass die richtige Lösung unten schon mit dabei steht. Wenn ihr die Aufgaben ausgerechnet habt, dann sucht einfach Eure Lösung in der Liste unten und kreuzt die richtige Antwort an.
Wenn ihr eine Aufgabe nicht lösen könnt, oder sie zu schwer ist, haltet euch nicht so lange an der Aufgabe auf, sondern versucht erst, die anderen Aufgaben zu lösen.
Wenn ihr am Ende des Blattes angekommen seid, dürft Ihr auf die nächste Seite umblättern, wo ihr weitere Aufgaben findet.
Danach blättert bitte nicht weiter als das Stoppzeichen (zeigen!). Ihr habt dafür nun 20 Minuten Zeit. Los!"
(Starten Sie die Stoppuhr und warten Sie 20 Minuten ab!).

!!!WICHTIG!!!
Während der 20 Minuten Testphase gehen Sie bitte mit der Lehrerin und dem Klassenbuch zu jedem einzelnen Schüler und tragen Sie die Klassenbuchnummer auf die erste Seite des Testheftes ein!
Lassen Sie dabei auf jeden Fall **kein Kind aus**! Die Klassenbuchnummer muss unbedingt vermerkt werden, damit die Ergebnisse der beiden Testtage einander zugeordnet werden und damit die Ergebnisse an die Kinder rückgemeldet werden können!

Nach Ablauf der 20 Minuten sagen Sie:
„Stopp! Legt bitte alle die Stifte einen Moment weg. Wir gehen gleich zum nächsten Aufgabenblock über!"

Instruieren Sie nun die Schüler für den dritten Aufgabenblock:
„Blättert bitte Um zur nächsten Seite! Ab jetzt müsst ihr für den Rest der Testung absolut nichts mehr rechnen!" Bei den nächsten drei Aufgabenblöcken lest ihr euch bitte nur den Text und die Fragen durch und kreuzt dann unten die richtige Antwort an. Ihr müsst nichts rechnen.
Es ist immer nur eine Antwort richtig. Wenn ihr eine Aufgabe nicht beantworten könnt oder nicht sicher seid, welches die richtige Antwort ist, dann kreuzt bitte diejenige an, von der ihr denkt, dass sie noch am ehesten richtig ist.
Wenn ihr eine Aufgabe nicht lösen könnt, oder sie zu schwer ist, haltet euch nicht so lange an der Aufgabe auf, sondern versucht erst, die anderen Aufgaben zu lösen.
Wenn ihr am Ende des Blattes angekommen seid, dürft Ihr auf die nächste Seite umblättern, wo ihr weitere Aufgaben findet.
Danach blättert bitte nicht weiter als das Stoppzeichen.
Ihr habt dafür nun 15 Minuten Zeit. Los!"
(Starten Sie die Stoppuhr und warten Sie 15 Minuten ab!).

Geben Sie dem Lehrer/der Lehrerin nun einen Stapel REMOD-Bögen und eine REMOD-Instruktion. Bitten Sie den Lehrer/die Lehrerin für jedes Kind einen Bogen auszufüllen.

Nach Ablauf der 15 Minuten sagen Sie:
„Stopp! Legt bitte alle die Stifte einen Moment weg. Wir machen nun 10 Minuten lang Pause! Nach der Pause geht es noch ein bisschen weiter, aber das meiste haben wir schon geschafft!"

Warten Sie nach der Pause, bis alle Kinder wieder an ihrem Platz sitzen. Beginnen Sie dann mit der Instruktion für den vierten Aufgabenblock (Block I).

„Ihr dürft nun Bei den Aufgaben die nun kommen, müsst Ihr auch nicht rechnen. Bitte lest Euch einfach die Fragen durch, und kreuzt die richtige Antwort an.
Wenn ihr eine Aufgabe nicht lösen könnt, oder sie zu schwer ist, haltet euch nicht so lange an der Aufgabe auf, sondern versucht erst, die anderen Aufgaben zu lösen.
Wenn ihr am Ende des Blattes angekommen seid, dürft Ihr auf die nächste Seite umblättern, wo ihr weitere Aufgaben findet.
Danach blättert bitte nicht weiter als das Stoppzeichen.
Ihr habt dafür nun 20 Minuten Zeit. Los!"
(Starten Sie die Stoppuhr und warten Sie 20 Minuten ab!).

Nach Ablauf der 20 Minuten sagen Sie:
„Stopp! Legt bitte alle die Stifte einen Moment weg. Jetzt kommt der letzte Aufgabenblock. Ihr dürft umblättern. Kreuzt bitte wieder die richtige Antwort zu jeder Aufgabe an. Wenn ihr eine Aufgabe nicht lösen könnt, oder sie zu schwer ist, haltet euch nicht so lange an der Aufgabe auf, sondern versucht erst, die anderen Aufgaben zu lösen.
Wenn ihr am Ende des Blattes angekommen seid, dürft Ihr auf die nächste Seite umblättern, wo ihr weitere Aufgaben findet."
Ihr habt dafür nun 15 Minuten Zeit. Los!"
(Starten Sie die Stoppuhr und warten Sie 15 Minuten ab!).

Nach Ablauf der 15 Minuten sagen Sie:
„Stopp. Bitte legt alle die Stifte weg. Nehmt Euch nun bitte noch die erste Seite und füllt diese aus."

Sammeln Sie dann die Testhefte ein. Achten Sie darauf, dass die Schüler die erste Seite des Testheftes ausgefüllt haben. Sammeln Sie auch die REMOD-Bögen von der Lehrkraft ein und packen Sie alles zusammen in ein Paket.

Versehen Sie das gesamte Testmaterial auf jeden Fall mit der Beschriftung welche Klasse (a, b, c, d) an welcher Schule getestet wurde, damit wir die Ergebnisse auch zuordnen können!

Während der Testung:
- Gehen Sie während der Testung durch die Reihen und überprüfen Sie, ob die Schüler die Instruktionen befolgen.
- Beim Beginn jedes Aufgabenblockes gehen Sie bitte als allererstes einmal durch die Reihen und überprüfen Sie, ob jedes Kind auf der richtigen Seite ist!

2. Testtag

Die nachfolgende Tabelle gibt Ihnen einen Überblick über den Ablauf des zweiten Testtages.

Minuten Testungszeit	Zu bearbeitende Aufgaben
0-10	Vorstellung / Instruktion
10-20	Block G
20-40	Block V
40-55	Block S
55-65	Pause
65-85	Block M
85-100	Block I

Begrüßen Sie die Klasse und stellen Sie sich nochmals kurz vor.

Teilen Sie den Schüler/innen zunächst die Testaufgaben aus und zwar so, dass die Versionen A und B immer abwechselnd ausgeteilt werden, damit die Schüler/innen nicht voneinander abschreiben. (Zu Ihrer Information: Die Versionen A und B erhalten exakt die gleichen Aufgaben, allerdings innerhalb der Blöcke jeweils in umgedrehter Reihenfolge.)

Erinnern Sie die Schüler/innen nochmals an das Vorgehen:

„Wir gehen heute genauso vor wie beim letzten Mal. Ich erkläre Euch aber zur Erinnerung noch mal die wichtigsten Punkte:

Wenn ihr eine Aufgabe nicht lösen könnt, weil sie zu schwer ist, oder ihr sie nicht versteht, dann geht bitte zur nächsten Aufgabe über. Wenn am Ende noch Zeit ist, könnt ihr wieder zu den schwereren Aufgaben zurückgehen.

Wenn ihr alle Aufgaben auf einer Seite bearbeitet habt, dann könnt ihr auf die nächste Seite blättern. Wenn aber am Ende der Seite so ein Stopp-Zeichen ist (hochhalten, zeigen), dann blättert bitte nicht um, sondern wartet, bis ich das Zeichen dazu gebe.

Wenn ihr eine Aufgabe nicht beantworten könnt oder nicht sicher seid, welches die richtige Antwort ist, dann kreuzt bitte diejenige an, von der ihr denkt, dass sie noch am ehesten richtig ist.

Wenn ihr Fragen habt, meldet Euch kurz, dann komme ich zu euch. Das hört sich vielleicht schwierig an, aber wir beginnen jetzt erstmal mit etwas, was ihr schon kennt.“

!!! Achtung bei Rückfragen während der Testung: Den Schüler/innen dürfen auf gar keinen Fall inhaltliche Tipps zum Lösen der Aufgaben gegeben werden. Antworten Sie bei inhaltlichen Fragen immer nur *„Ich weiß die Antwort auch nicht. Lies Dir genau den Text durch und versuche, es herauszufinden.“*
Sorgen Sie auch unbedingt dafür, dass die Lehrer/innen nicht helfen – wenn Sie ihnen erklären, dass das die Ergebnisse verfälscht, sind diese in der Regel kooperativ.

Instruieren Sie die Schüler zum ersten Aufgabenblock:
„Bitte blättert in Euren Tests alle auf die Seite 2 (Seite 2 hochhalten und zeigen!). Auf dieser Seite seht ihr ganz normale Textaufgaben. Versucht die Aufgaben zu bearbeiten.
Für Rechnungen könnt ihr das leere Blatt benutzen, das ich euch ausgeteilt habe. Schreibt in Eure Tests dann bitte nur die Antworten, nachdem ihr die Aufgabe ausgerechnet habt.
Wenn die erste Aufgabe Euch zu schwer ist, geht bitte zur nächsten Aufgabe über und beendet die erste, wenn danach noch Zeit ist.
Wenn ihr fertig seid, wartet ab, bis ich sage, dass die nächsten Aufgaben drankommen. Bitte blättert vorher nicht um.
Ihr habt 10 Minuten Zeit. Los!“
(Starten Sie die Stoppuhr und warten Sie 10 Minuten ab!).
Gehen Sie während dieser 10 Minuten durch die Reihen und kontrollieren Sie, ob die Schüler sich so verhalten, wie sie instruiert wurden und beantworten Sie eventuelle Fragen.

Nach Ablauf der 10 Minuten sagen Sie:
„Stopp! Legt bitte alle die Stifte einen Moment weg. Wir gehen gleich zum nächsten Aufgabenblock über!"

Instruieren Sie die Schüler nun für den zweiten Aufgabenblock (Block I).
„Jetzt dürft ihr umblättern auf die nächste Seite. Die Aufgaben, die jetzt kommen sind ganz normale Rechenaufgaben, wie ihr sie sicher aus dem normalen Unterricht kennt. Der einzige Unterschied ist, dass die richtige Lösung unten schon mit dabei steht. Wenn ihr die Aufgaben ausgerechnet habt, dann sucht einfach Eure Lösung aus der Liste unten und kreuzt die richtige Antwort an.
Wenn ihr eine Aufgabe nicht lösen könnt, oder sie zu schwer ist, haltet euch nicht so lange an der Aufgabe auf, sondern versucht erst, die anderen Aufgaben zu lösen.
Wenn ihr am Ende des Blattes angekommen seid, dürft Ihr auf die nächste Seite umblättern, wo ihr weitere Aufgaben findet.
Danach blättert bitte nicht weiter als das Stoppzeichen (zeigen!). Ihr habt dafür nun 15 Minuten Zeit. Los!"
(Starten Sie die Stoppuhr und warten Sie 15 Minuten ab!).

!!!WICHTIG!!!
Während der 15 Minuten Testphase gehen Sie bitte mit der Lehrerin und dem Klassenbuch zu jedem einzelnen Schüler und tragen Sie die Klassenbuchnummer auf die erste Seite des Testheftes ein!
Lassen Sie dabei auf jeden Fall **kein Kind aus**! Die Klassenbuchnummer muss unbedingt vermerkt werden, damit die Ergebnisse der beiden Testtage einander zugeordnet werden und damit die Ergebnisse an die Kinder rückgemeldet werden können!

Nach Ablauf der 20 Minuten sagen Sie:
„Stopp! Legt bitte alle die Stifte einen Moment weg. Wir gehen gleich zum nächsten Aufgabenblock über!"

Instruieren Sie nun die Schüler für den dritten Aufgabenblock:
„Blättert bitte Um zur nächsten Seite! Ab jetzt müsst ihr für den Rest der Testung absolut nichts mehr rechnen!" Bei den nächsten drei Aufgabenblöcken lest ihr euch bitte nur den Text und die Fragen durch und kreuzt dann unten die richtige Antwort an. Ihr müsst nichts rechnen.
Es ist immer nur eine Antwort richtig. Wenn ihr eine Aufgabe nicht beantworten könnt oder nicht sicher seid, welches die richtige Antwort ist, dann kreuzt bitte diejenige an, von der ihr denkt, dass sie noch am ehesten richtig ist.
Wenn ihr eine Aufgabe nicht lösen könnt, oder sie zu schwer ist, haltet euch nicht so lange an der Aufgabe auf, sondern versucht erst, die anderen Aufgaben zu lösen.
Wenn ihr am Ende des Blattes angekommen seid, dürft Ihr auf die nächste Seite umblättern, wo ihr weitere Aufgaben findet.
Danach blättert bitte nicht weiter als das Stoppzeichen.
Ihr habt dafür nun 20 Minuten Zeit. Los!"
(Starten Sie die Stoppuhr und warten Sie 20 Minuten ab!).

Nach Ablauf der 15 Minuten sagen Sie:
„Stopp! Legt bitte alle die Stifte einen Moment weg. Wir machen nun 10 Minuten lang Pause! Nach der Pause geht es noch ein bisschen weiter, aber das meiste haben wir schon geschafft!"

Warten Sie nach der Pause, bis alle Kinder wieder an ihrem Platz sitzen. Beginnen Sie dann mit der Instruktion für den vierten Aufgabenblock (Block I).

„Ihr dürft nun Bei den Aufgaben die nun kommen, müsst Ihr auch nicht rechnen. Bitte lest Euch einfach die Fragen durch, und kreuzt die richtige Antwort an.

Wenn ihr eine Aufgabe nicht lösen könnt, oder sie zu schwer ist, haltet euch nicht so lange an der Aufgabe auf, sondern versucht erst, die anderen Aufgaben zu lösen.
Wenn ihr am Ende des Blattes angekommen seid, dürft Ihr auf die nächste Seite umblättern, wo ihr weitere Aufgaben findet.
Danach blättert bitte nicht weiter als das Stoppzeichen.
Ihr habt dafür nun 20 Minuten Zeit. Los!"
(Starten Sie die Stoppuhr und warten Sie 20 Minuten ab!).

Nach Ablauf der 20 Minuten sagen Sie:
„Stopp! Legt bitte alle die Stifte einen Moment weg. Jetzt kommt der letzte Aufgabenblock. Ihr dürft umblättern. Kreuzt bitte wieder die richtige Antwort zu jeder Aufgabe an. Wenn ihr eine Aufgabe nicht lösen könnt, oder sie zu schwer ist, haltet euch nicht so lange an der Aufgabe auf, sondern versucht erst, die anderen Aufgaben zu lösen.
Wenn ihr am Ende des Blattes angekommen seid, dürft Ihr auf die nächste Seite umblättern, wo ihr weitere Aufgaben findet."
Ihr habt dafür nun 15 Minuten Zeit. Los!"
(Starten Sie die Stoppuhr und warten Sie 15 Minuten ab!).

Nach Ablauf der 15 Minuten sagen Sie:
„Stopp. Bitte legt alle die Stifte weg. Nehmt Euch nun bitte noch die erste Seite und füllt diese aus."

Sammeln Sie dann die Testhefte ein. Achten Sie darauf, dass die Schüler die erste Seite des Testheftes ausgefüllt haben. Sammeln Sie auch die REMOD-Bögen von der Lehrkraft ein und packen Sie alles zusammen in ein Paket.

Versehen Sie das gesamte Testmaterial auf jeden Fall mit der Beschriftung welche Klasse (a, b, c, d) an welcher Schule getestet wurde, damit wir die Ergebnisse auch zuordnen können!

Während der Testung:

- Gehen Sie während der Testung durch die Reihen und überprüfen Sie, ob die Schüler die Instruktionen befolgen.
- Beim Beginn jedes Aufgabenblockes gehen Sie bitte **als allererstes** einmal durch die Reihen und überprüfen Sie, ob jedes Kind auf der richtigen Seite ist!

Anhang E:
REMOD – Ratingverfahren für Lehrkräfte

REMOD

Liebe Lehrerin, lieber Lehrer,

In unserem Forschungsprojekt „Modellierungskompetenzen beim mathematischen Problemlösen“ gehen wir der Frage nach, **welche unterschiedlichen Teilprozesse beim Lösen von mathematischen Textaufgaben relevant sind** und wie man diese möglichst effizient und unabhängig voneinander erfassen kann.

Der zu entwickelnde Test hat zum **Ziel**, aufzuzeigen, bei welchen dieser lösungsrelevanten Teilprozesse der jeweilige Schüler/die jeweilige Schülerin noch **Schwächen** hat bzw. wo seine/ihre **individuellen Stärken** liegen. Auf der Grundlage dieses Testergebnisses ist es dann möglich, gezielt auf den Schüler/die Schülerin **zugeschnittene Förderstrategien** einzusetzen.

Um die Güte unserer bisher entwickelten Testaufgaben zu überprüfen, würden wir gerne auch **Ihre Einschätzung der mathematischen Kompetenz** Ihrer Schüler/innen mit erheben.
Wenn die in unseren Tests erfassten Leistungen mit Ihrem Expertenurteil konform gehen, ist dies ein weiterer Hinweis auf die Qualität unseres Verfahrens.

Wir bitten Sie daher, für jeden Schüler/jede Schülerin ihrer Klasse unseren **REMOD-Fragebogen** auszufüllen.

- Tragen Sie bitte zuallererst die **Klassenbuchnummer** des jeweiligen Schülers/der jeweiligen Schülerin in das dafür vorgesehene Feld ein.
- Auf dem Fragebogen sind eine Reihe von **Kompetenzen aufgelistet**, mit denen einige Ihrer Schüler/innen eventuell Schwierigkeiten haben könnten. Versuchen Sie bitte jeweils ein Urteil abzugeben ob der Schüler/die Schülerin hinsichtlich der jeweiligen Kompetenz ***nie*****,** ***selten*****,** ***manchmal*****,** ***oft*** **oder** ***immer*** **Schwierigkeiten hat.**
- Bei einigen der aufgelisteten Kompetenzen fällt Ihnen eine solche Einschätzung durch ihre Erfahrung mit dem Schüler/der Schülerin im Unterricht und bei der Korrektur von Hausaufgaben oder Klassenarbeiten sicherlich leicht. Es wird aber vielleicht auch Kompetenzen geben, bei denen es Ihnen **schwer fällt, eine zuverlässige Einschätzung abzugeben**, kreuzen Sie in diesem Fall bitte die letzte Kategorie: „schwer feststellbar“ an.
- Vergeben Sie bitte abschließend eine **Gesamtnote** von 1-6 dafür, wie hoch Sie **im Allgemeinen die mathematischen Fähigkeiten** des Schülers/der Schülerin einschätzen.

Wenden Sie sich bei Verständnisfragen gerne an unsere Testleiter.

Herzlichen Dank für Ihre Mitarbeit!

REMOD

Klassenbuchnummer ___________________	nie	selten	manchmal	oft	immer	schwer fest- stellbar
Der Schülerin/dem Schüler fällt es schwer, den **Sinn** eines gelesenen Textes zu **verstehen**.	☐	☐	☐	☐	☐	☐
Der Schülerin/dem Schüler fällt es schwer, festzustellen welche **Größen** bei einer Textaufgabe **gegeben** und welche **gesucht** sind.	☐	☐	☐	☐	☐	☐
Der Schülerin/dem Schüler fällt es schwer, aus einem **Aufgabentext** die **relevanten Informationen** herauszu-lesen.	☐	☐	☐	☐	☐	☐
Der Schülerin/dem Schüler fällt es schwer, **zwischen relevanten und irrelevanten Informationen** im Auf-gabentext zu **unterscheiden**.	☐	☐	☐	☐	☐	☐
Der Schülerin/dem Schüler fällt es schwer, zu erkennen, ob bei einer Textaufgabe **addiert, subtrahiert, multipliziert oder dividiert** werden muss.	☐	☐	☐	☐	☐	☐
Der Schülerin/dem Schüler fällt es schwer, **Zahlwörter in Ziffern zu übersetzen**.	☐	☐	☐	☐	☐	☐
Der Schülerin/dem Schüler fällt es schwer, eine **mathematisch korrekte Schreibweise** zu benutzen.	☐	☐	☐	☐	☐	☐
Der Schülerin/dem Schüler fällt es schwer, **mathematisches Wissen** zur Problemlösung nutzen (d.h. Rechen-aufgaben so zu lösen wie er/sie es gelernt hat).	☐	☐	☐	☐	☐	☐
Der Schülerin/dem Schüler fällt es schwer, „**geschickt zu rechnen**“.	☐	☐	☐	☐	☐	☐
Der Schülerin/dem Schüler fällt es schwer, nach dem Lösen einer Text-aufgabe einen **Antwortsatz zu formu-lieren**, der die ursprünglich gestellte Frage beantwortet.	☐	☐	☐	☐	☐	☐
Der Schülerin/dem Schüler fällt es schwer, Lösungen von Textaufgaben auch **auf andere Situationen zu ver-allgemeinern**.	☐	☐	☐	☐	☐	☐

Allgemein würde ich die **mathematische Kompetenz** des Schülers / der Schülerin mit der **Note** _____ bewerten.

Anhang F:
Sequentielle multiple Regression

```
REGRESSION
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT score_textaufgaben
 /METHOD=ENTER score_strukturieren
 /METHOD=ENTER score_mathematisieren
 /METHOD=ENTER score_verarbeiten
 /METHOD=ENTER score_interpretieren.
```

Aufgenommene/Entfernte Variablen[b]

Modell	Aufgenommene Variablen	Entfernte Variablen	Methode
1	score_strukturieren[a]	.	Eingeben
2	score_mathematisieren[a]	.	Eingeben
3	score_verarbeiten[a]	.	Eingeben
4	score_interpretieren[a]	.	Eingeben

a. Alle gewünschten Variablen wurden aufgenommen.

b. Abhängige Variable: score_textaufgaben

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,319[a]	,102	,099	1,60509
2	,440[b]	,193	,188	1,52333
3	,511[c]	,261	,254	1,46000
4	,551[d]	,304	,295	1,41964

a. Einflußvariablen : (Konstante), score_strukturieren

b. Einflußvariablen : (Konstante), score_strukturieren, score_mathematisieren

c. Einflußvariablen : (Konstante), score_strukturieren, score_mathematisieren, score_verarbeiten

d. Einflußvariablen : (Konstante), score_strukturieren, score_mathematisieren, score_verarbeiten, score_interpretieren

Koeffizienten[a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
		B	Standardfehler	Beta		
1	(Konstante)	,950	,527		1,801	,073
	score_strukturieren	,436	,073	,319	5,937	,000
2	(Konstante)	-,491	,556		-,883	,378
	score_strukturieren	,267	,075	,195	3,541	,000
	score_mathematisieren	,346	,058	,327	5,949	,000
3	(Konstante)	-1,294	,554		-2,336	,020
	score_strukturieren	,216	,073	,158	2,963	,003
	score_mathematisieren	,236	,059	,223	3,963	,000
	score_verarbeiten	,285	,053	,289	5,345	,000
4	(Konstante)	-1,006	,543		-1,854	,065
	score_strukturieren	,155	,072	,113	2,153	,032
	score_mathematisieren	,100	,066	,095	1,524	,129
	score_verarbeiten	,233	,053	,236	4,379	,000
	score_interpretieren	,244	,056	,273	4,344	,000

a. Abhängige Variable: score_textaufgaben

Lebenslauf

	Persönliche Daten
Name	Julia Sybille Riebel
Geburtsdatum	28.05.1983
Geburtsort	Landau
Nationalität	Deutsch
	Schule
1989 - 1993	Grundschule Ilbesheim
1993 - 2002	Otto – Hahn – Gymnasium Landau Abschluss: Abitur
	Studium und Beruf
2002 - 2008	Studium der Psychologie an der Universität Koblenz · Landau Abschluss: Diplom
2004 - 2006	Wissenschaftliche Hilfskraft im Fachbereich 8 der Universität Koblenz · Landau – Abteilung Entwicklungspsychologie
2006 - 2008	Wissenschaftliche Hilfskraft am Zentrum für empirische pädagogische Forschung der Universität Koblenz · Landau
2008 -	Wissenschaftliche Mitarbeiterin am Zentrum für empirische pädagogische Forschung der Universität Koblenz · Landau

Eidesstattliche Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbstständig und ohne unerlaubte Hilfe angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit (einschließlich Abbildungen), die anderen Ursprungs sind in jedem Einzelfall mit Angabe des Urhebers als solche kenntlich gemacht habe.

Des weiteren versichere ich, dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie noch nicht veröffentlich ist sowie dass ich mich noch nicht anderweitig um einen Doktorgrad beworben habe bzw. einen solchen bereits besitze.

Die dem Verfahren zugrunde liegende Promotionsordnung des Fachbereichs Psychologie der Universität Koblenz-Landau vom 25. Oktober 2007 ist mir bekannt.

Landau, 3.11.2009 ____________________________