BEZIRKSREGIERUNG ARNSBERG

IN ZUSAMMENARBEIT MIT DER RUHR-UNIVERSITÄT BOCHUM

fermat

Förderung des Erfassens mathematischer Texte

S. Alff, Dr. Ch. Fuhrmann, Dr. M. Funcke, H. Gündogdu,

B. Heise, E. Klute, S. Vlam, R. Vockenroth

12. Februar 2015

Ein Arbeitsbericht der Projektgruppe zu Entwicklung und Validierung von Diagnosetests zum allgemeinen Text-Lese-Verständnis, fachsprachlichen Text-Lese-Verständnis, Symbol- und Zahlen-Verständnis sowie Mathematisieren.

Inhaltsverzeichnis

Kapitel 1 Einleitung

In vielen öffentlichen Diskussionen klagen Ausbildungsbetriebe über die Schwierigkeit, geeignete Auszubildende für ihre Unternehmen zu finden. Dies wird oft auf eine unzureichende Lese- und Rechenkompetenz von Schülerinnen und Schülern zurückgeführt.

Auch im schulischen Umfeld ist wahrnehmbar, dass Schülerinnen und Schüler komplexe, textbasierte Problemstellungen in der Mathematik häufig nicht zufriedenstellend lösen können. Das Lösen solcher Problemstellungen repräsentiert die Fähigkeit, berufliche Anforderungen zu meistern.

Im Juni 2010 sind einige Mathematiklehrerinnen und –lehrer an Berufskollegs der Bezirksregierung Arnsberg gerne der Einladung des Fachleiters Herrn Poelke zu einem Erfahrungsaustausch an der Ruhr-Universität Bochum gefolgt. Die Veranstaltung war mit dem Titel „Individuelle Förderung Mathematik" versehen und wurde vom federführenden Professor des Lehrstuhls für Berufs- und Wirtschaftspädagogik, Dr. Harney, unter dem Arbeitstitel „Textverständnis Mathematik" geführt.

Wie passen Textverständnis und individuelle Förderung in der Mathematik zusammen? Im Verlauf des Treffens wurde schnell deutlich, dass alle anwesenden Lehrerinnen und Lehrer vielfach ähnliche Beobachtungen und Erfahrungen in ihren Klassen und Kursen gesammelt hatten.

Als ein großes „Problemfeld" kristallisierte sich der Bereich der „Textaufgaben" heraus, nämlich dass Schülerinnen und Schüler sich mathematischen Textaufgaben (zum Beispiel in Klausuren und Abschlussprüfungen) nicht logisch und gezielt nähern, sondern häufig wild das vorhandene Zahlenmaterial auswählen und in mehr oder weniger zuvor gelernte Formeln pressen, um dann irgendwie sinnsuchend ein Ergebnis in einem Antwortsatz unterzubringen.

Aus diesem Erfahrungsaustausch entwickelte sich die Projektgruppe ***fermat***[1], die es sich zur Aufgabe gemacht hat, die Lesekompetenz im Mathematikunterricht zu diagnostizieren und zu fördern. ***fermat*** ist eine Wortschöpfung aus **F**örderung des **Er**fassens **ma**thematischer **T**exte.

Nach Einschätzung der Arbeitsgruppe ist die mathematische Text-Lese-Kompetenz ein bedeutender Schlüssel, um sich in der Arbeitswelt erfolgreich zu behaupten.

Im Rahmen einer Fortbildungsreihe der Bezirksregierung Arnsberg und der Ruhr-Universität Bochum hat ***fermat*** dazu einen Diagnosetest entwickelt und validiert. Dieses Instrument ermöglicht es den individuellen Förderbedarf von Schülerinnen und Schülern im Bereich des mathematischen Text-Lese-Verständnisses zu ermitteln. Im Anschluss können die Schülerinnen und Schüler ihre diagnostizierten Defizite gezielt aufarbeiten.

Der Diagnosetest soll möglichst vielen Berufskollegs landesweit kostenlos zugänglich gemacht werden. Dazu gibt es eine webbasierte Onlineversion mit automatischer Auswertung der Testfragen.

Im Folgenden findet sich ein Arbeitsbericht der Projektgruppe.

[1] Die Gruppe ***fermat*** setzt sich aus Mathematiklehrerinnen und -lehrern der Berufskollegs Eugen-Schmalenbach-Berufskolleg (Dr. Michael Funcke, Rainer Vockenroth), Kaufmannsschule II Hagen (Birgit Heise), Berufskolleg des Kreises Olpe (Sybille Vlam, Stephanie Alff) sowie Cuno Berufskolleg I in Hagen (Eva Klute) zusammen. Ergänzt wird die Gruppe durch Christoph Fuhrmann von der Ruhr-Universität Bochum.

Kapitel 2 Projektmotivation

2.1 Textaufgaben im Schulalltag des Berufskollegs

Das Lösen von mathematischen Textaufgaben stellt für die Schülerinnen und Schüler des Berufskollegs, mit ihrer stark differierenden kulturellen, sozialen und ethnischen Vielfalt eine besondere Herausforderung dar.

Der Versuch, Textaufgaben an alltägliche oder berufliche Situationen anzubinden und entsprechend zu formulieren, ist durch die unterschiedlichen Lebenswelten der Schüler schwierig. Daher bleibt die oft postulierte motivierende Wirkung von Textaufgaben aus.

Das Aufzeigen der Möglichkeit, alltägliche Probleme mit Hilfe von Mathematik zu lösen, zeigt eher eine demotivierende Wirkung, denn Textaufgaben werden im Kontext des sowieso „ungeliebten" Unterrichtsfach Mathematik meist als Höchstschwierigkeit angesehen: Neben den ohnehin schwierig zu erlernenden instrumentellen, mathematischen Fähigkeiten fordern Textaufgaben zusätzliche Kompetenzen aus anderen Domänen bei der Lösung ein.

2.2 Textaufgaben in der Wahrnehmung von Schülerinnen und Schülern

In der Bearbeitung von Textaufgaben könnten Schülerinnen und Schüler erkennen, dass Mathematik nicht nur ein in sich abgeschlossenen formales Wissens- und Wissenschaftsgebiet ist, sondern ein „mächtiges Werkzeug bei der Lösung unterschiedlichster Probleme"[2] sein kann.

Tatsächlich bestehen bei der Beschäftigung mit mathematischen Textaufgaben im Unterricht Schwierigkeiten. Hinzu kommt die grundsätzlich ablehnende Haltung von Schülerinnen und Schülern gegen diese Aufgabensorte.

[2] Siehe dazu: Blum, Neubrand u.a. 2003, S.47.

2.3 Textaufgaben in der Wahrnehmung von Lehrerinnen und Lehrern

Textaufgaben tragen zusätzliche, problembesetzte Aspekte in mathematische Aufgabenkontexte ein, die im Rahmen des üblichen Mathematikunterrichts von Lehrerinnen und Lehrern erkannt und beachtet werden müssen.

Erst wenn Lehrerinnen und Lehrern bewusst ist, welche Kompetenzen zur Lösung von Textaufgaben benötigt werden und welche „Minimalausstattung" an Ausprägung die Schülerinnen und Schüler in diesen Kompetenzen besitzen müssen, kann Unterricht entsprechend gestaltet werden. Ziel muss sein, die individuellen Problemlagen zu erkennen und Schülerinnen und Schülern dementsprechend dann auch individuell zu fördern.

Auf Seiten der Lehrerinnen und Lehrer fehlt vielfach sowohl ein diagnostisches Werkzeug zur Analyse der für Textaufgaben benötigten Kompetenzen als auch das entsprechende Fördermaterial.

2.4 Lesefähigkeiten beim Lösen von Textaufgaben

Es stellt sich somit die Frage nach den Bedingungen des Erfolgs beim Lösen von rein textlich präsentierten Aufgaben, die mittels mathematischer und textanalytischer Kompetenzen gelöst werden müssen.

Zur Bewältigung von mathematischen Textaufgaben müssen die Schülerinnen und Schüler über basale Lesefähigkeiten im Sinne der von PISA 2000 definierten *reading literacy* verfügen.[3] Dies bedeutet, dass sie das in der Textaufgabe dargestellte Phänomen in seiner Sinnhaftigkeit erfassen und entsprechend deuten können.[4] Sie müssen in der Lage sein, die für den Problemkontext der Textaufgabe relevanten Informationen aus dem Text zu entnehmen.

---

[3] Siehe dazu: OECD 2003, 2006.

[4] Die hier an dieser Stelle beschriebene Entdeckung des Phänomens der Aufgabe meint hier noch nicht den von Blum u.a. (Blum 2003) dargestellten Beginn des Prozess des mathematischen Modellierens.
Ausgangspunkt des kybernetischen Prozesses des mathematischen Modellierens nach Blum ist die Strukturierung einer textlich präsentierten Realsituation. Die Strukturierung setzt voraus, dass das textlich dargestellte Phänomen inhaltlich verstanden wurde.

Eine weitere semantische Schwierigkeit für die richtige Inhaltserfassung der Textaufgabe durch die Schülerinnen und Schüler ergibt sich durch die Übersetzung der Begriffe der Alltagssprache in die Formensprache der Mathematik. Begriffe wie *überragt*, *mindestens*, *weiter*, *teurer* etc. müssen durch ihre mathematische Form substituiert werden. Dies wird unter dem Begriff mathematische Lesekompetenz verstanden.

2.5 Hierarchie von Kompetenzen

Dies bedeutet: Bevor Schülerinnen und Schüler die Kompetenzen anwenden können, die allgemein unter dem Stichwort Mathematik verstanden werden (instrumentelles Rechnen), müssen sie über Kompetenzen verfügen, von denen zumindest die *reading literacy* nicht Bestandteil der Bildungsstandards im Fach Mathematik ist.

Erst wenn bei den Schülerinnen und Schülern eine der Textaufgabe angemessen hohe Ausprägung der Lesekompetenz[5] und der mathematischen Lesekompetenz vorhanden ist, kann es im Rahmen der Bearbeitung einer Textaufgabe zur Anwendung der originär im Mathematikunterricht vermittelten Kompetenzen kommen. Erst ab diesen „textlichen“ Kompetenzausprägungen ergibt sich die Sinnhaftigkeit der Anwendung der Kompetenzen *Mathematisieren* und *Modellieren*. Denn ist der Text der Aufgabe in seiner Phänomenologie nicht verstanden, erfolgt ein Modellieren oder Mathematisieren ohne Bezug zum Problem der Aufgabe: ein willkürliches Problem wird bearbeitet.

Die Bearbeitung des Algorithmus des Rechnens im Sinne einer instrumentellen, technischen Kompetenz erfolgt als letzter Schritt.

[5] Lesekompetenz meint hier eine übergeordnete Kompetenz, die auch das Lesen diskontinuierlicher Texte z.B. Grafiken beinhaltet. Das "Lesen" von Grafiken, Abbildungen etc. (Sumfleth, Schüttler 1995, Winn 1987) stellt jedoch grundsätzlich eine eigene Kompetenz dar, die zu einem späteren Zeitpunkt im Rahmen des erfolgreichen Lösens von mathematischen Textaufgaben getestet werden soll.

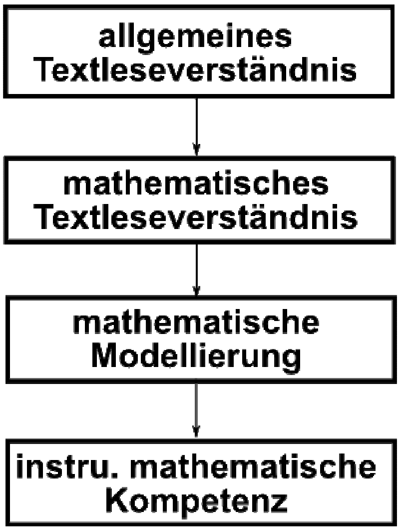


Abbildung 1: Lösungsalgorithmus einer mathematischen Textaufgabe

2.6 Auswirkung auf Unterricht und Diagnose

An der Lösung einer Textaufgabe sind so zumindest die Kompetenzen *reading literacy*, mathematische Lesekompetenz, Mathematisieren/Modellidentifikation[6] und instrumentelles Rechnen beteiligt.

Unterrichtsplanung und individuelle Förderung müssen auf diese Kompetenzen, die an der Lösung von Textaufgaben beteiligt sind, abgestimmt werden. Entsprechend müssen die Ausprägungen der Schüler in den einzelnen Kompetenzen diagnostiziert werden. Ein dazu zu entwickelnder Test muss Aufgaben beinhalten, die eindeutig einer der Kompetenzen zugewiesen werden können, oder aber von einer Kompetenz dominiert werden. Eine große Bedeutung kommt daher der Item- bzw. Fragen-Konstruktion und der Unidimensionalität der Aufgaben zu.[7]

[6] Zur Unterscheidung von Mathematisieren und Modellidentifikation siehe weiter unten.

[7] In der Item-Response-Theorie (IRT) ist der Begriff der Unidimensionalität in der Testaufgabenentwicklung verwurzelt und Voraussetzung für das Prinzip der lokalen stochastischen Unabhängigkeit (Lord 1980, Lord, Novick 1968, Wilson 2005, Fischer 1974, Rost 2004, Baker, Kim 2004). Die Testaufgabenentwicklung und die Pilotierung der Aufgaben stellen in der IRT die Unidimensionalität als einen modellimmanenten Prozess sicher. Dies stellt einen Gegensatz zur klassischen Testtheorie dar, in der die Berechnung der Reliabilität eine der Testentwicklung und Testdurchführung nachgelagerte Überprüfung der Passung ist, eine post-Analyse.

Andernfalls ist Diagnose nicht möglich. Die Forderung der Unidimensionalität der zu entwickelnden Testaufgaben unterscheidet diese von „üblichen“ Klausuraufgaben. Die Aufgabenentwicklung stellt daher unter der diagnostischen Prämisse und bei Anwendung des hier präferierten Rasch Modells [8] eine Herausforderung dar, die mehrere Aufgabenpilotierungszyklen in Anspruch nehmen kann.[9]

2.7 Allgemeine Anforderung an einen Diagnosetest

Da für das erfolgreiche Lösen von Textaufgaben neben mathematischen Fähigkeiten auch Text-Lese-Fähigkeiten Voraussetzung sind, muss der im Rahmen von *fermat* entwickelte Test sowohl klassisch mathematische Fähigkeiten als auch basales, fachsprachliches und symbolhaftes Text-Lese-Verstehen der Schülerinnen und Schüler überprüfen.

Nur in der individuellen Analyse der Ausprägung der Kompetenzen, die an der Lösung von Textaufgaben beteiligt sind, ist diagnostizierbar, warum Schülerinnen und Schüler an Textaufgaben scheitern. Danach kann individuelle Förderung nachhaltig greifen.

[8] Das Rasch Modell ist ein Modell der Item-Response-Theorie und wird unter anderem bei der Analyse der PISA Erhebungen verwendet (u.a. PISA 2003).

[9] Siehe dazu: Wilson 2003, Liu 2007, Arnold 2002, Wuttke 2008.

KAPITEL 3 WISSENSCHAFTLICHE EINORDNUNG

3.1 Anbindung an PISA

Die im Rahmen des Projekts *fermat* in Kapitel 2 aufgeführten Kompetenzen zur Lösung einer Textaufgabe[10] - *reading literacy*, mathematische Lesekompetenz, Mathematisieren bzw. Modellidentifikation und instrumentelles Rechnen - stehen in einem ähnlichen Verhältnis zueinander wie die Kompetenzen, die PISA 2003 für das Konzept des mathematischen Modellierens, die *mathematical literacy*, angibt.[11]

Nach PISA setzt sich das mathematische Modellieren aus mehreren Teilkompetenzen zusammen: Strukturieren, Mathematisieren, Verarbeiten, Interpretieren, Validieren. Durch eine zyklische Verknüpfung bilden diese Kompetenzen den Begriff der *mathematical literacy* ab.

Diese zyklische Verknüpfung bindet den Ausgangspunkt der Modellierung, die Realsituation, an die Ergebnisse des Modellierungsprozesses an. Die Resultate der miteinander verknüpften Prozesse müssen interpretierend auf die Realsituation rückbezogen werden.

Die Kompetenzen selbst befinden sich dabei, wie der in *fermat* beschriebene Algorithmus zum Lösen der Aufgabe, in einer hierarchischen Beziehung: Ohne eine der Realsituation angepasste Strukturierung ist kein Mathematisieren möglich, ohne dieses keine Verarbeitung usw.[12]

Der nationale Ergänzungstest in PISA 2003 legt zudem einen Schwerpunkt auf die Kompetenz der instrumentellen Rechenfertigkeiten – technisches Rechnen –, die dem Kompetenzbereich des Verarbeitens im mathematischen Modellierungsprozess zuzuordnen ist.

---

[10] Siehe zum verwendeten Kompetenzmodell auch Kapitel 5.
[11] Siehe dazu: Blum u.a. 2002, S. 47.
[12] Siehe dazu Kapitel 5.

3.2 *fermat*: reading literacy und mathematical literacy

Die Kompetenzen zum Lösen von Textaufgaben weisen inhaltliche wie relationale Ähnlichkeiten zum mathematischen Modellierungszyklus auf. Beispielsweise werden in beiden Kompetenzmodellen instrumentelle und mathematisierende Kompetenzen ausgewiesen, die in der gleichen hierarchischen Beziehung zueinander stehen. Unterschiede gibt es bei der Analyse der Strukturierungskompetenz. Diese wird im Algorithmus zum Lösen von Textaufgaben detaillierter betrachtet als im mathematischen Modellieren.

Das Strukturieren einer Textaufgabe, d.h. einer textlich repräsentierten Realsituation, wird im Rahmen von *fermat* nicht nur auf die Schwierigkeiten im Sinne des mathematischen Modellierens hin analysiert, sondern zusätzlich auch auf die textlichen Hindernisse: Die *reading literacy* wird in den Kontext der *mathematical literacy* gestellt.

3.3 Modellversuch Vocational Literacy

Die Bedeutung, gerade im Kontext von Textaufgaben, das Text-Lese-Verstehen zu untersuchen, wird durch das Projekt „Vocational Literacy – Methodische und sprachliche Kompetenzen in der beruflichen Bildung (VOLI)“[13] deutlich.

Der Begriff *vocational literacy* wird dabei als die Summe der sprachlichen Fertigkeiten definiert, die in beruflichen Zusammenhängen relevant sind. Dies bezieht sich einerseits auf kommunikative Fähigkeiten, anderseits auf das Strukturieren von Fachtexten bzw. technischen Texten. Aus diesen müssen Informationen entnommen werden, um sie auf berufliche Anforderungen anwenden zu können.

Als Prototyp einer solchen Anforderung kann in Übertragung auf *fermat* eine Textaufgabe gelten, in der ein textlich präsentiertes Anwendungsproblem mathematisch gelöst werden soll. Denkbar ist die Anwendung einer neuen Gesetzeslage auf die Berechnung der Sozialversicherungen oder die Berücksichtigung neuer DIN-Normen für die Materialprüfung.

VOLI legt seinen Schwerpunkt jedoch alleine auf das Text-Lese-Verstehen.

---

[13] Siehe dazu: Biedebach 2006, S. 24-26.

3.4 Zusammenfassung

Das im Rahmen von *fermat* erarbeitete Kompetenzmodell zur Lösung von Textaufgaben verbindet sowohl die in PISA definierte Kompetenz der *mathematical literacy* mit ihrem mathematischen Schwerpunkt sowie die Text-Lese-Kompetenzen, die innerhalb von VOLI für die berufliche Bildung als relevant angesehen werden.

KAPITEL 4 METHODISCHE EINORDNUNG

Geeignete methodische, statistische Instrumente zur Kompetenzdiagnostik stellt die Item-Response-Theorie zur Verfügung und hier im Besonderen das Rasch Modell, mit dessen Hilfe u.a. die Tests für die PISA Untersuchungen entwickelt wurden.[14]

4.1 Testentwicklung mit Hilfe des Rasch Modells - PISA-Logik

Für die Item-Response-Theorie ist die Beziehung zwischen der Kompetenzausprägung einer Person und der Schwierigkeit der kompetenzkorrespondierenden Aufgaben von zentraler Bedeutung. Aus der Wechselbeziehung ergibt sich die Wahrscheinlichkeit einer Person mit gegebener Ausprägung auf einer Kompetenz/Dimension Aufgaben aus einer solchen Domäne zu lösen. Das bedeutet, je höher die Ausprägung der Kompetenz einer Person, desto größer ist die Wahrscheinlichkeit, Aufgaben, die diese Kompetenz repräsentieren, zu lösen.

Modelliert wird dieser Zusammenhang in der Item-Response-Theorie durch eine geeignete Item-Response-Funktion. Das Rasch Modell verwendet dazu den folgenden funktionalen Zusammenhang:

$$
\begin{aligned}
p(x_{ik}=1|\theta_i, \sigma_k) &= \frac{\exp(\theta_i - \sigma_k)}{(1+\exp(\theta_i - \sigma_k))} \\
p(x_{ik}=0|\theta_i, \sigma_k) &= \frac{1}{(1+\exp(\theta_i - \sigma_k))}
\end{aligned}
\qquad (1)
$$

mit $x_{ik}=1$ als der richtigen Lösung und $x_{ik}=0$ als der nicht richtigen Lösung der Aufgabe k von Person i, θ_i als der Ausprägung der Person i auf der betrachteten Kompetenz/Dimension, σ_k als der Aufgabenschwierigkeit der Aufgabe k und p als der entsprechenden Lösungs- oder Nicht-Lösungswahrscheinlichkeit; bzw.

---

[14] Siehe dazu: Adam, Wu 2000 und Baumert et.al. 2002.

$$p(x_{ik}|\theta_i, \sigma_k) = \frac{\exp(x_{ik}(\theta_i - \sigma_k))}{(1 + \exp(\theta_i - \sigma_k))} \qquad (2)$$ [15]

4.2 Item-Response-Funktion

Die Item-Response-Funktion des Rasch Modells gehört zur Familie der logistischen Funktionen und hat folgenden Verlauf für eine Aufgabe mit Aufgabenschwierigkeit Null:

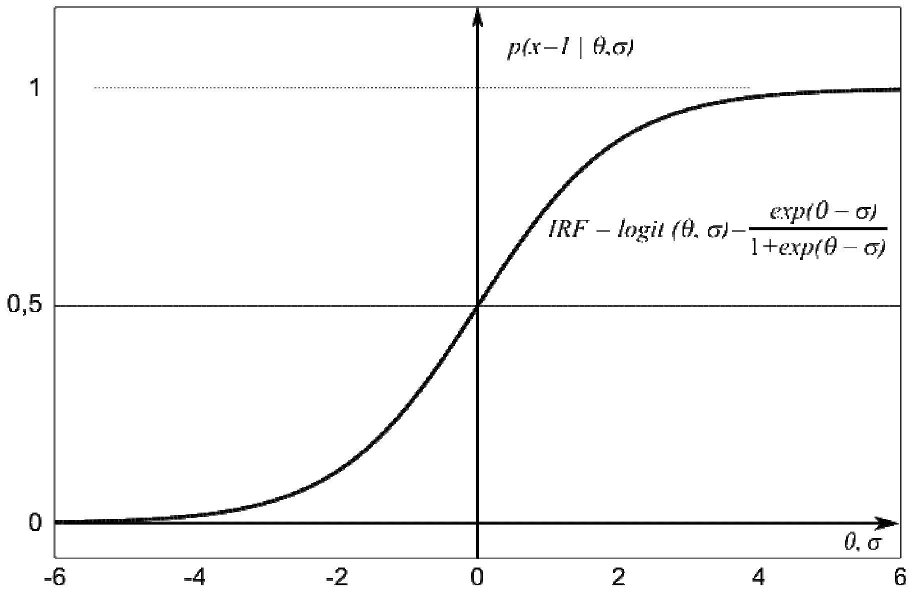


Abbildung 2: Verlauf einer Item-Response-Funktion mit Itemschwierigkeit/Itemlocation = 0 für Rasch Modelle.

Auf der x-Achse werden sowohl die Aufgabenschwierigkeit wie auch die Personenfähigkeit gleichermaßen abgetragen. Die Wahrscheinlichkeit der Lösung einer Aufgabe ergibt sich dann durch die Differenz der Ausprägung der Person auf einer Kompetenz und der Aufgabenschwierigkeit. Die Lage der Aufgabenschwierigkeit ist dabei grundsätzlich als der Wendepunkt der logistischen Funktion definiert.

15 Siehe dazu: Rost 2004, Rasch 1961; Goldstein 1980; Swaminathan 1983, Andersen 1973, Birnbaum in Lord, Novick 1968.

Der Verlauf der logistischen Funktion ist dem Verlauf der Verteilungsfunktion der Standardnormalverteilung – der kumulativen Dichtefunktion – sehr ähnlich. Jedoch wird in der Item-Response-Theorie die Funktion nicht aufgrund ihrer und mit ihren statistischen Eigenschaften verwendet, sondern als psychologisch plausibler Regressionszusammenhang, der Lösungswahrscheinlichkeit über der Kompetenz betrachtet.[16]

4.3 Merkmale des Rasch Modells

Die Besonderheit des Rasch Modells ist, dass es eine in der logistischen Funktionsfamilie einzigartige messtechnische Eigenschaft aufweist. Der Vergleich zweier Personen ist unabhängig von den verwendeten Aufgaben. Dies wird mit dem Begriff spezifische Objektivität beschrieben und bedeutet: Liegt für eine Population ein Aufgaben- bzw. Itemuniversum modellkonformer Aufgaben/Items vor, so ist es unerheblich, welche Items aus diesem Universum für einen Test ausgewählt werden. Der Vergleich der Messergebnisse der getesteten Personen liefert immer das gleiche Resultat.

Diese Eigenschaft ergibt sich unter anderem dadurch, dass für alle Aufgaben ein gleich steiler Funktionsverlauf angenommen wird.

Aus den mathematischen Eigenschaften des Modells ergibt sich, dass, wenn das Modell gültig ist, alleine die Anzahl der durch eine Person gelösten Aufgaben (Score) alle Informationen über deren Kompetenzausprägung enthält. Der Personenscore ist somit eine suffiziente Statistik. Das bedeutet, dass das Rasch Modell nur den Score der Personen als Information für das Modell benötigt. Die Hinzunahme weiterer Informationen, wie z.B. die der Antwortmuster der Personen, erzeugt für das Modell keine zusätzliche Information, die für die Kompetenzausprägungsberechnung verwertet werden könnten.

Voraussetzung hierfür ist, dass die Aufgaben nur mit einer Kompetenz/Dimension korrelieren, denn andernfalls wären keine Aussagen bezüglich einer bestimmten Kompetenzausprägung möglich.

[16] Siehe dazu: Baker, Kim 2004, Lord, Novick 1968.

4.4 Abgrenzung zur klassischen Testtheorie

In der modellhaften Verbindung von Personenfähigkeit und Itemschwierigkeit ist der wesentliche Unterschied der Item-Response-Theorie zur klassischen Testtheorie begründet. In der klassischen Testtheorie erfolgt die Analyse der Items immer erst ex post, denn die Berechnungen zur Reliabilität der Items erfolgen immer erst nach absolviertem Test. Die Items sind nur indirekt über den Gesamttest Bestandteil der Theorie, die nur den „wahren Gesamttestwert", der sich aus empirischem Gesamttestwert (Score) und einem Gesamtfehler zusammensetzt, kennt. Die Items gehören sozusagen nicht zum Modell.

Der Prozess der Aufgabenentwicklung für ein Rasch Modell ist ein iterativer und langwieriger Prozess in dem die entworfenen Aufgaben immer wieder auf ihre Eindimensionalität und ihre Konformität mit dem Rasch Modell überprüft werden, geändert oder verworfen werden. Die Aufgabenentwicklung selbst stellt somit sicher, dass die Aufgaben reliabel und valide sind, also zuverlässig das messen, was zu messen ist.

4.5 Faktorenanalyse zur Überprüfung der Unidimensionalität der entwickelten Items

Die Aufgaben aller Tests sollen jeweils nur eine Dimension bzw. innerhalb der Dimensionen vorher definierte Unterdimensionen abbilden. Dazu werden entsprechende Aufgaben entwickelt und auf die Hypothese hin überprüft, dass die generierten Aufgaben nur eine Kompetenz oder Unterkompetenz abbilden. Diese Überprüfung erfolgt mit einem Verfahren auf Basis der Regressionsrechnung, der Faktorenanalyse.

Eine Idee der Faktorenanalyse ist es, zu einem gegebenem Datenbestand korrelative Zusammenhänge von Items zu identifizieren, sogenannte Faktoren.[17]

Zu jeder der definierten Dimensionen muss ein eigenständiger Test mit entsprechenden Items entwickelt werden. Jeder Test für sich diagnostiziert die Ausprägung der Personen auf den Kompetenzen mit ihren eventuellen Unterkompetenzen. In einer „Längsschnittbetrachtung" jeder einzelnen Person über alle Tests hinweg kann ermittelt werden, in welchen Kompetenzen Förderbedarf existiert. Erfolgreich eine Textaufgabe lösen sollten nur die

[17] Siehe dazu: Bortz 1999; Wirtz, Nachtigall 2004; Hair 2010.

Personen, die auf allen vier Kompetenzen ein noch zu definierendes Mindestmaß an Fähigkeit aufweisen.

Die Durchführung einer beispielhaften Bestimmung der Parameter eines Rasch Modells mittels der Software R findet sich im Anhang 1.

Kapitel 5 Kompetenzmodell

5.1 Festlegung des Begriffs Kompetenz

Franz Weinert versteht unter *Kompetenz* 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.“.[18]

Nach Jost Reischmann wird *Kompetenz* als Handlungsmöglichkeit in einem gegebenen Kontext, verstanden, der gleichzeitig 3 Elemente umfasst:

„a) Die handelnde Person muss über die einschlägigen Fähigkeiten verfügen.
b) Es muss erlaubt und möglich sein, dass diese Person dieses Handeln ausführt.
c) Es müssen die notwendigen materiellen Ressourcen zur Verfügung stehen.
Handlungsmöglichkeit meint dabei nicht einfache Routinefertigkeiten, sondern schließt Elemente wie Komplexität, Selbständigkeit, an Standards orientierte Qualität, Verstehen und Wissen ein und ist auf einen beschreibbar umrissenen Handlungskontext bezogen.“[19]

Nach den Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss[20] trägt der Mathematikunterricht mit seiner Sprache, seinen Symbolen, Bildern und Formeln in der Bedeutung für die Beschreibung und Bearbeitung von Aufgaben und Problemen inner- und außerhalb der Mathematik zur Bildung der Schülerinnen und Schüler bei.

Nach diesem Bildungsstandard im Fach Mathematik für den Mittleren Schulabschluss sollen folgende 5 Kompetenzen erworben werden:

---

[18] Aus Weinert 2001, S. 27.

[19] Reischmann 2004.

[20] Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss, S. 6 ff.

- **(K1) Mathematisch argumentieren**

  mathematische Argumentationen entwickeln, Beschreiben und Begründen von Lösungswegen

- **(K2) Probleme mathematisch lösen**

  geeignete heuristische Hilfsmittel, Strategien und Prinzipien zum Problemlösen auswählen und anwenden; Plausibilität der Ergebnisse überprüfen

- **(K3) Mathematisch modellieren**

  den Bereich oder die Situation, die modelliert werden soll, in mathematische Begriffe, Strukturen und Relationen übersetzen; in dem jeweiligen mathematischen Modell arbeiten

- **(K4) Mathematische Darstellungen verwenden**

  verschiedene Formen der Darstellung von mathematischen Objekten und Situationen anwenden, interpretieren und unterscheiden; Beziehungen zwischen Darstellungsformen je nach Situation und Zweck auswählen und zwischen ihnen wechseln

- **(K5) Mit symbolischen, formalen und technischen Elementen der Mathematik umgehen**

  mit Variablen, Termen, Gleichungen, Funktionen, Diagrammen und Tabellen arbeiten; symbolische und formale Sprache in natürliche Sprache übersetzen und umgekehrt

- **(K6) Kommunizieren**

  Überlegungen, Lösungswege bzw. Ergebnisse dokumentieren, verständlich darstellen und präsentieren; die Fachsprache adressatengerecht verwenden, Äußerungen von anderen und Texte zu mathematischen Inhalten verstehen und überprüfen

„Diese Kompetenzen werden immer im Verbund erworben und angewendet.“ [21]

[21] Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss, Beschluss der Kultusministerkonferenz vom 4.12.2003, S. 7.

5.2 Beteiligte Kompetenzen für das Lösen von mathematischen Textaufgaben

Welche Kompetenzen fehlen oder sind bei Schülerinnen und Schülern nicht in ausreichender Weise ausgeprägt, wenn sie Mathematikaufgaben, die ihnen in Form von Textaufgaben vorgelegt werden, nicht bewältigen können?

Mathematische Fachtexte haben ihre eigenen Besonderheiten. Die Leseschwierigkeiten und die mathematischen Verständnisschwierigkeiten hängen sehr eng zusammen und sind häufig kaum zu trennen.

Dass ein eventuelles Scheitern an sogenannten Textaufgaben nicht nur an einer mangelnden Ausprägung instrumenteller mathematischer Kompetenz liegen kann, ist offensichtlich, denn das Lösen von Textaufgaben setzt basal die Fähigkeit des Lesens voraus.

5.2.1 Allgemeinsprachliches Text-Lese-Verständnis

Die Schülerinnen und Schüler müssen über die Kompetenz verfügen, die für die Lösung der Aufgabe relevanten Informationen im Text identifizieren zu können. Beschreibt das mathematische Problem ein Phänomen der Lebenswelt der Schülerinnen und Schüler, so werden diese zunächst auf ein allgemeines Text-Lese-Verständnis zurückgreifen, um die Informationen im Text herauszufiltern.

5.2.2 Fachsprachliches Text-Lese-Verständnis

In einem weiteren Schritt müssen die als relevant identifizierten Informationen aus ihrem umgangssprachlichen Darbietungskontext gelöst werden. Umgangssprachlich dargebotene Beziehungen zwischen jeweils relevanten Objekten müssen als Fachsprache erkannt werden, um sie in einem weiteren Schritt in mathematische Bezüge und entsprechende Symbolik übersetzen zu können. Die Schülerinnen und Schüler müssen über mathematische/fachsprachliche Lesekompetenz verfügen.

5.2.3 Modellierung / Mathematisierung eines Problemkontextes

Ist sichergestellt, dass der Text in seinem Sinngehalt verstanden wurde, die benötigten Informationen extrahiert werden konnten und die Beziehung zwischen den Objekten erkannt wurden, kann in einem nächsten Schritt überprüft werden, ob die Schülerinnen und Schüler in der Lage sind, die gewonnenen Erkenntnisse mathematisch zu modellieren und in einen mathematischen Ansatz zu übersetzen.

5.2.4 Instrumentelles Rechnen

Erst im letzten Schritt innerhalb des Lösungswegs einer mathematischen Textaufgabe kommt es zur Anwendung instrumenteller mathematischer Fähigkeit, indem die aus dem Text heraus generierte Mathematisierung des Problems unter Verwendung der entsprechenden Werte, mit Hilfe instrumenteller Fähigkeiten „gelöst wird”.

5.3. Kompetenzmodell nach *fermat*

Innerhalb des Projekts *fermat* werden verschiedene Kompetenzen betrachtet, die sich jeweils aus unterschiedlichen Aspekten zusammensetzen und teilweise aufeinander aufbauen.

I:	**Allgemeinsprachliches Text-Lese-Verständnis**
II:	**Fachsprachliches Text-Lese-Verständnis**
III:	**Symbol- und Zahlen-Verständnis**
IV:	**Kompetenz Mathematisieren**
V:	**Kompetenz Modellierung**

Wie oben und in Kapitel 3 angedeutet, bildet das allgemeinsprachliche Text-Lese-Verständnis (Kompetenz I) eine Voraussetzung für ein fachsprachliches Text-Lese-Verständnis (Kompetenz II). Ein ausreichendes Verständnis der Symbol- und Zahldarstellung (Kompetenz III) kann in Teilen auch ohne die vorigen Kompetenzen existieren. Die Fähigkeit des Mathematisierens (Kompetenz IV) erscheint ohne grundsätzliche Beherrschung der ersten drei Kompetenzen schwierig.

5.3.1 Kompetenz I: Allgemeinsprachliches Text-Lese-Verständnis

Unter dem allgemeinsprachlichen Text-Lese-Verständnis wird innerhalb des Projekts das sinnentsprechende Erfassen eines Textes verstanden. Dies bezieht sich auf das Entnehmen von

Einzelinformationen sowie das Erfassen von Zusammenhängen. In beiden Bereichen lassen sich verschiedene inhaltliche Ebenen und unterschiedliche Schwierigkeiten feststellen.

Die kognitive Leistung des Lesenden sollte hierbei nicht nach den Inhalten des Textes, sondern nach dem Schwierigkeitsgrad des Erkenntnisgewinns beurteilt werden. Auf dieser Basis können mathematische Fragestellungen sinnvoll bearbeitet werden.

Die folgenden Ausführungen beziehen sich auf den Text „Knapp 6000 neue Auszubildende in Hagen im Herbst 2009“, der im Rahmen der Diagnosetests entwickelt wurde.[22]

Erfassen von Einzelinformationen

Einzelinformationen können direkt aus dem Text abgelesen oder aus mehreren Textpasssagen zusammengestellt werden. Im Zusammenhang damit ist u.a. denkbar:

- Daten / Zahlen entnehmen, z.B. 5959 neue Auszubildende;
- Personen oder Personengruppen, Orte, Jahreszahlen identifizieren, z.B. die Gruppe *Auszubildende zum Koch in Hagen* oder die Zuordnung *im Vorjahr* zu *2008*;
- Einzelne Wörter verstehen, z.B. *unversorgt.*

Erfassen von Zusammenhängen

Um Zusammenhänge aus dem Text zu entnehmen, ist es erforderlich Folgerungen aus Textaussagen zu schließen, Argumentationen nachzuvollziehen und logische oder sprachliche Verbindungen herzustellen. Im Hinblick darauf sind folgende Aktivitäten denkbar:

- Bezüge innerhalb eines Textes identifizieren;

  Um z.B. festzustellen, dass nicht alle Abschlüsse in den Zeilen 19 und 20 genannt werden, kann man die Zahlen mit der Zahl neuer Auszubildender vergleichen.

- Chronologie eines Textes herstellen, z.B. Vergleichen von Angaben aus dem Vorjahr mit aktuellen Angaben aus 2009;

[22] Siehe dazu Kapitel 6 und Anhang 2.

- Zentrale Aussagen eines Textes nachvollziehen, z.B. *Es gibt mehr neue Auszubildende 2009*.

Eine Textanalyse wie im Deutschunterricht ist für das grundlegende, mathematische Verständnis eines Textes nicht zwingend erforderlich. So ist eine Unterteilung des Textes in sinnzusammenhängende Abschnitte hilfreich, meist ist dies aber bereits durch das Druckbild vorgegeben. Zusammenhänge und Informationen im Text können auch ohne eine Sinnunterteilung des Textes erkennbar sein. Wichtig ist ein verständnisorientiertes Lesen zur Entnahme der für spätere Aufgabenstellungen relevanten Angaben.

Anforderungsniveau

Durch den entwickelten Diagnosetest werden in allen Kompetenzen mindestens zu erfüllende Anforderungsniveaus festgestellt. Diese Niveaus orientieren sich an den Anforderungen, denen Schülerinnen und Schüler beim Verstehen von mathematikhaltigen Texten begegnen.[23]

Die Anforderung, die im Rahmen des Diagnosetests untersucht wird, ist ein quasi fehlerfreies Verständnis von Informationen und Zusammenhängen, das beim Eintritt ins Berufskolleg vorhanden sein sollte. Schwächen in dieser Kompetenz sollten im Bereich einer grundlegenden Deutschförderung aufgearbeitet werden. Kompetenz I wird als eine Grundlage für das fachsprachliche Text-Lese-Verständnis überprüft.

5.3.2 Kompetenz II: Fachsprachliches Text-Lese-Verständnis

Unter fachsprachlichem Text-Lese-Verständnis wird im Rahmen von *fermat* das Verstehen von mathematischem Fachvokabular und nicht-mathematischen Vokabular in mathematischen Zusammenhängen zusammengefasst.

Neben dem Erfassen und Umsetzen mathematischen Fachvokabulars, z.B. *Subtraktion*, *Faktor*, *Produkt*, *Summe*, *addieren*, *dividieren* wird das dem allgemeinsprachlichen Text-Lese-Verständnis entliehene Erfassen mathematischer Begriffe untersucht. Dazu zählen auch

---

[23] Siehe dazu Kapitel 8.

Wörter und Textaussagen wie z. B. *die Mehrzahl, meist*, *fast*, *verringert sich* und auch begrenzende bzw. vergleichende Ausdrücke wie. *genau so viel*, *höchstens* oder *mehr als.*

In dem vom *fermat* ausgewählten Text beziehen sich die Testfragen im Wesentlichen auf vergleichende, umschreibende Wörter und Verben, die eine Veränderung angeben. So sind auch die gewählten Textformulierungen zur Feststellung der Kompetenz II in einem allgemeinsprachlichen Sinnzusammenhang zu verstehen und liegen in unterschiedlichen Schwierigkeitsstufen (leicht, mittelschwer, schwer) vor. Die Aufforderung „Addieren Sie …“ oder „Multiplizieren Sie …“ ist von sehr vielen Schülerinnen und Schülern leicht zu verstehen. Ist diese Rechnung jedoch als Ergebnis eines in einem Text umschriebenen Sachverhalts als Teil der Lösungsstrategie zu verwenden, so muss von einem höheren Schwierigkeitsgrad ausgegangen werden.

An der nachfolgenden Beispieltabelle soll gezeigt werden, dass es verschiedene allgemeinsprachliche Ausdrücke für einen mathematischen Begriff geben kann.

Tabelle 1: Beispiele für Zusammenhang zwischen allgemeinsprachlichen und mathematischen Begriffen

allgemeinsprachlich verwendete Begriffe	**mathematische Begriffe**
vergrößern – verkleinern hinzufügen – vermindern zunehmen – abnehmen sinken - steigen	addieren – subtrahieren Summe – Differenz
verdoppeln – halbieren das Doppelte – die Hälfte	Multiplizieren – dividieren Produkt – Quotient

5.3.3 Kompetenz III: Symbol- und Zahlen-Verständnis

Kompetenz III betrachtet das Interpretieren von logischen Rechenzeichen wie >, >, =, >=, von Einheiten wie *m^2, Sekunden, €, Kilogramm* und von Verhältniszahlen wie beispielsweise Angaben in Prozent oder Bruchzahlen. Gängige Abkürzungen wie *Mio* für *Millionen* müssen in Vollwörter übersetzt werden können und umgekehrt.

Ob die Schülerinnen und Schüler Bruchteile von Wortdarstellungen in mathematische Sprache umsetzen und große Zahlen von Wortform in Ziffern (und umgekehrt) übersetzen können, ist zu betrachten. Als Beispiele sind hier zu nennen *ein Drittel* in *1/3* oder *22,5 Millionen* in *22500000*.

Gerade bei dieser Kompetenz kann man vermuten, dass das fachsprachliche Text-Lese-Verständnis teilweise Voraussetzung für das Verstehen von mathematischen Symbolen ist, z.B. *mehr als* mit dem Symbol >. Insofern korrespondieren Fragen zur Kompetenz II mit denen zur Kompetenz III.

5.3.4 Kompetenz IV: Mathematisieren

Das Übersetzen des Modells eines realen Problems oder einer Fragestellung in eine mathematische Formel oder einen Rechenweg ist mit dem Begriff Mathematisieren gemeint. Das Mathematisieren wird vorwiegend durch die folgenden Fähigkeiten beschrieben:

- Text in Formel übersetzen (Schlüsselwörter);
- Formel in den Zusammenhang übersetzen;
- Rechenweg auswählen und zuordnen;
- relevante Zahlen in den Rechenweg einordnen;
- Auswahl eines Rechenweges hinterfragen.

Das Modellbilden gehört nicht zur Kompetenz Mathematisieren, sondern wird im Modellierungskreislauf als Schritt vor dem Mathematisieren verstanden. So beschreibt ein

Modell näherungsweise die Realität durch die Auswahl von Fakten sowie das Bilden von Ideen unter Berücksichtigung der zur Verfügung stehenden mathematischen Methoden. „Die Modellbildung abstrahiert mit dem Erstellen eines Modells von der Realität, weil diese meist zu komplex ist, um diese genau abzubilden.“[24]

Vorausgesetzte Rechenverfahren

Um zu Mathematisieren müssen Rechenverfahren vorausgesetzt werden. Beim Eintritt ins Berufskolleg ist davon auszugehen, dass eine Beherrschung der Bruchrechnung, des Dreisatzes, der Prozentrechnung, des Lösens linearer Gleichungen und der Zinsrechnung grundlegend gegeben ist. Daher werden in den Aufgaben zur Mathematisierung Modelle verwendet, bei denen diese Rechenverfahren als mathematische Methoden zu Grunde liegen.

Anforderungsniveau

Die Anforderung, die in dem Diagnosetest untersucht wird, ist eine weitgehend fehlerfreie Beherrschung des Mathematisierens auf der beschriebenen, grundlegenden Niveaustufe. Eine Förderung kann sich an den oben aufgeführten Aspekten innerhalb der einzelnen Kompetenzen orientieren. Die im Diagnosetest fehlerhaft beantworteten Aufgaben können Hinweise auf den zu fördernden Kompetenzbereich liefern und die Auswahl des Fördermaterials unterstützen.

Feststellung des Förderbedarfs

Einige Kompetenzen bauen wie beschrieben teilweise aufeinander auf. Demnach beeinflussen sich die Kompetenzen auch in Bezug auf den Förderbedarf. Es kann kein reines Stufenmodell bei der Beurteilung des Förderbedarfs geben. Förderbedarf sollte immer individuell abgestimmt werden und die Ergebnisse der unterschiedlichen Kompetenztests berücksichtigen. Hinweise zur individuellen Förderung von Schülerinnen und Schülern finden sich im Kapitel 8.

[24] Aus: Ritter 2005.

Kapitel 6 Textauswahl

6.1 Texte in der Mathematik

Die Auseinandersetzung mit mathematischem Text-Lese-Verständnis beinhaltet eine Beschäftigung mit Texten und deren Beschaffenheit.

Zwei Textsorten sind hier zu unterscheiden. Zum einen mathematische Texte in Form von Aufgabenstellungen, zum anderen mathematikhaltige Texte wie zum Beispiel Artikel aus Zeitungen oder Zeitschriften.

6.1.1 Mathematische Texte

Zur erst genannten Sorte zählen Textaufgaben, die im Mathematikunterricht und in Klausuren allgegenwärtig sind. Texte aus Textaufgaben sind in der Regel kurz und prägnant. Sie sind sehr exakt formuliert und beinhalten eine Fülle von Informationen, die oft mit Hilfe von Abkürzungen, Fachbegriffen oder Abstraktionen ausgedrückt werden. Die formale Sprache, der sich meist bedient wird, ist für Schülerinnen und Schüler zum Teil gewöhnungsbedürftig. Diese formale Sprache hat für manche Schülerin oder manchen Schüler eher eine abschreckende als eine motivierende Wirkung. Im Schulalltag müssen sich die Schülerinnen und Schüler jedoch mit diesen Texten auseinandersetzen und einen verständnisvollen Umgang mit dieser Art von Texten erlernen.

6.1.2 Mathematikhaltige Texte

Mathematikhaltige Texte sind nicht immer mathematisch eindeutig formuliert. Die Informationsdichte ist in der Regel nicht so hoch wie bei mathematischen Texten und sie beinhalten weniger mathematische Fachbegriffe. Dadurch können Redundanzen auftreten. Der Text kann für die Bearbeitung einer Aufgabe nicht relevante, vielleicht sogar überflüssige Informationen beinhalten. Hier müssen die Schülerinnen und Schüler strukturieren können und wichtige Informationen von Überflüssigem unterscheiden lernen.

6.2 Anforderungen an den Diagnosetext

Um als Basis einer Testkonstruktion dienen zu können, ist es unerlässlich, dass der Text genau formuliert und somit eindeutig zu verstehen ist. Da dies vorausgesetzt werden muss, scheiden viele Zeitungsartikel sehr schnell aus. Bei anderen Texten, wie zum Beispiel dem Artikel: „Das zweite Internet“[25] wurde allerdings erst nach genauerer Arbeit mit dem Text klar, dass er sich nicht für den Test eignet.

Im Rahmen von *fermat* wurde immer wieder auf der Suche nach geeigneten, authentischen Texten die Frage diskutiert: Darf man Änderungen, Erweiterungen, Aktualisierungen und Kürzungen an den Texten vornehmen, um diese dann für eigene Zwecke nutzen zu können? Aus urheberrechtlichen Gründen wurden diese Fragen schließlich verneint.

Weitere Kriterien an einen Text sind der Schwierigkeitsgrad, die Textlänge und die Thematik.

Schwierigkeitsgrad

Der Schwierigkeitsgrad darf nicht zu komplex sein, so dass die Schülerinnen und Schüler nicht nach wenigen Zeilen schon demotiviert und überfordert aufhören zu lesen. Er darf aber auch nicht zu einfach sein, damit genügend Fragen und Aufgabenstellungen zum Text formuliert werden können.

Textlänge

Eine Textlänge zwischen 200 und 300 Wörter erscheint sinnvoll, denn der Test sollte insgesamt zeitlich nicht zu lang werden, so dass er auf jeden Fall in einer Einzelstunde (45 Minuten) durchführbar ist.

Thematik

Die Thematik sollte einen Bezug zum Berufskolleg und zur Lebenswelt der Schülerinnen und Schüler haben.

[25] Siehe dazu: Mai 2010. Erst bei genauer Analyse kamen inhaltliche Ungereimtheiten und zweideutige Formulierungen zum Vorschein. Zum Beispiel geht aus dem Text nicht eindeutig hervor, ob die verwendeten Begriffen: „Besucher“, „Mitglieder“ und „Nutzer“ synonym verwendet werden können oder ob unterschiedliche Personengruppen damit gemeint sind.

Darüber hinaus sollte der Text ohne Graphiken und Diagramme auskommen, um die Eindimensionalität gewährleisten zu können. Das Lesen von Graphiken und Diagramme setzt eine weitere Kompetenz voraus, die auch sehr interessant ist und ggf. in einem weiterführenden Test mit aufgenommen werden kann.

6.3 Textauswahl zum Diagnosetest

Als Grundlage für den Test dient ein mathematikhaltiger Text, in dem über die Ausbildungssituation in Hagen berichtet wird.[26]

Um Urheberrechtsverletzungen auszuschließen und die Eindeutigkeit der Formulierungen gewährleisten zu können, wurde der ausgewählte Text in Anlehnung an einen Bericht über Studentenzahlen frei formuliert. Dies hat den Vorteil, dass der Text immer wieder überarbeitet und angepasst werden kann. Zum Beispiel wurden die Zeile 20-25 später hinzugefügt um weitere Fragemöglichkeiten zu erhalten. Nichtsdestotrotz können nicht alle Kriterien und Unterkompetenzen mit einem Text abgeprüft werden, z.B. die Chronologie ist in diesem Text nicht adäquat abprüfbar. Desweiteren sind eine Vielzahl von mathematischen Begriffen und Symbolen, die überprüft werden sollen, nicht Bestandteile dieses Textes.

Es gibt daher im zweiten Teil des Tests Aufgaben und Fragestellungen die unabhängig von dem Text bearbeitet werden können. Der Textumfang dieser Fragen ist wie bei gewöhnlichen Textaufgaben, es wurde jedoch auf Eindimensionalität geachtet.[27]

[26] Siehe dazu Anhang 2.

[27] Siehe dazu Kapitel 7.

Kapitel 7 Fragenentwicklung

7.1 Anforderungen an den Diagnosetest

Mit der festgelegten Zielsetzung, die im bisherigen Schulleben erworbene Lesekompetenz von Berufskollegschülerinnen und -schülern im Fach Mathematik zu diagnostizieren und daran anknüpfend notwendige Fertigkeiten adäquat auszubauen, ergibt sich nun die Notwendigkeit der Auswahl eines Testverfahrens. Informationen über den Kompetenzstand der Schülerinnen und Schüler können auf vielfältige Art gewonnen werden. Zur Testkonstruktion und zur Typologie von Testaufgaben gibt eine große Bandbreite an Möglichkeiten. Testtheoretische Grundlagen werden beispielsweise von Jordan zusammengefasst.[28]

Im Folgenden wird nun dargestellt, welche Rahmenbedingungen eine Diagnose im Sinne von *fermat* in der Schule vorfindet und welche konkreten Anforderungen sich daraus für einen Test ergeben.

7.1.1 Objektivität

Der Diagnosetest soll Objektivität besitzen. Im Schulalltag wird der Test von Lehrerinnen und Lehrern durchgeführt. Das bedeutet, dass zum Teil auch testdiagnostisch ungeschulte Personen das Testverfahren problemlos einsetzen können sollen.

7.1.2 Unabhängigkeit/Standardisierung

Zudem muss sichergestellt sein, dass der Test von unbegrenzt vielen Schülerinnen und Schülern unter den gleichen Bedingungen bearbeitet werden kann. Die Auswertung muss standardisiert und unabhängig von den Lehrkräften sein, die den Test durchführen, d.h. das

[28] Siehe dazu: Jordan 2011, Kapitel 7.

Testergebnis einer Schülerin oder eines Schülers bzw. die Vergabe von Testpunkten für bestimmte Antworten ist völlig unbeeinflusst vom der auswertenden Lehrkraft.

7.1.3 Praktikabilität

Die Testung der Lesekompetenz im Fach Mathematik ist eine zusätzliche Aufgabe für Mathematiklehrerinnen und -lehrer. Um eine hohe Akzeptanz des Tests zu erzielen, ist es förderlich, den Arbeits- bzw. Korrekturaufwand für die einzelne Lehrkraft minimal zu halten. Um die Lesekompetenz von Schülerinnen und Schülern umfassend und gesichert zu diagnostizieren, ist ein punktuell eingesetzter Test sicherlich nicht ideal. Wünschenswert wären kontinuierliche Beobachtungen in unterschiedlichen Lernsituationen, die durch Interviews oder Befragungen unterstützt werden. Dies ist jedoch im regulären Schulalltag von den Fachlehrerinnen und Fachlehrern kaum zu leisten.

Eine computerunterstützte Auswertung ist sinnvoll, da sie die Aspekte der Objektivität und des möglichst geringen Arbeitsaufwandes berücksichtigt. Darüber hinaus kann diese maschinelle Auswertung leicht übersichtlich und schnell verfügbar gestaltet werden.

7.1.4 Wissenschaftliche Auswertung

Der Diagnosestest im Rahmen von *fermat* soll nach der Item-Response-Theorie, die auf Georg Rasch zurückzuführen ist, ausgewertet werden. Auf dieser Testtheorie basiert auch die Auswertung der PISA-Erhebung aus den Jahren 2000 und 2002. Damit ist Forderung nach der Unidimensionalität der Aufgaben verbunden.

7.1.5 Fazit: Multiple –Choice-Aufgabenformat

Die oben beschriebenen Rahmenbedingungen finden in hohem Maße bei einer Testgestaltung mit Multiple-Choice-Aufgaben Berücksichtigung.

Geeignete Aufgaben bestehen aus einer geschlossenen Frage, für die mehrere vorformulierte Antworten zur Verfügung stehen. Die Schülerin bzw. der Schüler wählt eine oder mehrere richtige Antworten aus. Die Anzahl der Distraktoren, d.h. der nicht zutreffenden Antwortmöglichkeiten, variiert dabei.

In der Pre-Testphase des *fermat*-Projekts wurde allerdings festgestellt, dass erhebliche Schwierigkeiten auftreten, zu einzelnen Aufgaben eine Vielzahl von sinnvollen Distraktoren bzw. von unabhängigen Antwortmöglichkeiten zu entwickeln. Aus diesem Grund ist die Mehrzahl der Aufgaben bzw. Items dieses-Testverfahrens als Entscheidungsfrage konzipiert.

Im Rahmen des Modellversuchs VOLI weist Efing darauf hin, dass „man durch Multiple-Choice-Aufgaben ein hohes Maß an Validität erzielen“ kann, „sofern man die Lesekompetenz von Schülerinnen und Schülern diagnostizieren möchte“. Es ist sichergestellt, dass „tatsächlich das Leseverstehen und nicht etwa die Schreibkompetenz […] getestet wird.“[29]

7.2 Gestalterische Aufbereitung der Aufgaben

Die Gestaltung der Aufgaben ist in der Pilotierungsphase so gehalten, dass einer getroffenen Aussage die Auswahlmöglichkeiten *richtig* oder *falsch* folgen. Bei der Umsetzung der Diagnosetests auf *moodle* steht der Auswahl *falsch* die Alternative *wahr* gegenüber.[30]

Beispiel 1:

1. Es gab 14,6 Prozent mehr neue Hagener Auszubildende im Herbst 2009, dann waren es im Vorjahr weniger neue Auszubildende.

○ richtig ○ falsch

Gibt es für eine Frage mehrere Antwortmöglichkeiten, müssen diese Möglichkeiten einzeln mit *richtig* oder *falsch* bejaht bzw. negiert werden.

Beispiel 2:

[29] Aus: Efing, Institut für Qualitätsentwicklung 2006, S. 13.

[30] Siehe dazu auch Kapitel 9.

"Fast 1 % aller Hagener Schulabgänger" bedeutet...		
	richtig	falsch
8. ... etwas mehr als 1 % aller Hagener Schulabgänger.	○	○
9. ... etwas weniger als 1 % aller Hagener Schulabgänger.	○	○
10. ... exakt 1 % aller Hagener Schulabgänger.	○	○
11. ... deutlich weniger als 1 % aller Hagener Schulabgänger.	○	○

Diese Gestaltung hat den Vorteil, dass die Schülerinnen und Schüler nicht per Ausschlussverfahren oder durch Wahl der Mehrfachauswahl Rückschlüsse auf die Lösung ziehen können, denn die Schülerinnen und Schüler müssen zu jeder Antwortmöglichkeit die Entscheidung treffen, ob diese zutrifft oder nicht.

7.3 Wissenschaftliche Einordnung des Aufgabenformats

Im Kontext des Berufskollegs befassten sich zwei Projekte mit der Testung der Lesefähigkeit von Schülerinnen und Schülern. Diese werden im Folgenden kurz vorgestellt und dabei deren unterschiedliche verwendeten Aufgabenformate diskutiert.

7.3.1 Modellversuch *Vocational Literacy*

Der hessische Modellversuch „Vocational Literacy – Methodische und sprachliche Kompetenzen in der beruflichen Bildung“ (VOLI) aus den Jahren 2003 bis 2006 umfasst mehrere Testkomponenten.[31] Der entwickelte „Baukasten Lesediagnose“ besteht aus fünf Bausteinen, die jeweils aus einem oder mehreren Texten und Aufgaben unterschiedlichen Schwierigkeitsgrades zusammengesetzt sind. Der Lehrer wählt, abgestimmt auf die Lerngruppe, einen oder mehrere Tests aus dem „Baukasten Lesediagnose“ aus, führt diese durch und wertet sie aus. So soll die Lesekompetenz einzelner Schüler individuell und

[31] Siehe dazu auch Kapitel 3.

differenziert diagnostiziert werden.[32] Eine erste, eher intuitiv geprägte Einschätzung der Lehrkraft über die Lesekompetenz der Klasse bzw. einzelner Schülerinnen und Schüler ist notwendiger Bestandteil der Text- und Testauswahl. Dies wird bei der Testung im Rahmen von *fermat* vermieden; alle Schülerinnen und Schüler bearbeiten einen identischen Test.

7.3.2 Test nach Jordan

Jordan hat für den vom ihm entwickelten und durchgeführten „Test zum mathematischen Textverständnis“ als Antwortformat die individuelle Lösungsangabe in speziell vorgesehenen, fett markierten Feldern gewählt. Für Nebenrechnungen steht die Rückseite der Arbeitsblätter zur Verfügung.[33] Bei der Auswertung kann es allerdings zu Schwierigkeiten kommen, wenn eine Schülerin oder ein Schüler eine Aufgabe nicht vollständig oder fehlerhaft gelöst hat. Hier ist die Objektivität des Auswertens möglicherweise nicht immer gegeben, d.h. die Wahl des Auswerters beeinflusst möglicherweise die Vergabe von Punkten für bestimmte Antworten. Als ein Aufgabenbeispiel für die Testkonzeption von Jordan sei hier die Bungalow-Aufgabe (Item 21, Bestimmen der Länge einer Leiter) aufgeführt. Von den Schülerinnen und Schülern wird bei der Bearbeitung der Bungalow-Aufgabe erwartet, die für das Bestimmen der Leiterlänge notwendige Formel im Zusammenhang mit dem Satz von Pythagoras auszuwählen, die relevanten Informationen aus dem Text zu entnehmen, diese adäquat in die Formel einzusetzen sowie den zur Lösungsbestimmung notwendigen Algorithmus (Auflösen einer Gleichung durch Radizieren einer Summe von Quadratzahlen) korrekt anzuwenden.[34]

Die Auswertung der Schülerlösungen kann bei komplexen Aufgaben- und Antwortformat gegebenenfalls nicht mehr als eindeutig und objektiv angesehen werden. Die Verwendung von Multiple Choice vermeidet das skizzierte Problem der Auswerterobjektivität.

[32] Siehe dazu: Biedebach 2006, S. 24-26.
[33] Siehe dazu: Jordan 2011, Anhang D.
[34] Ebd., S. 206.

7.4 Weitere Aspekte zum Multiple-Choice-Testverfahren

7.4.1 Fehlervermeidung beim Ankreuzen der Lösung

Probleme können sich bei einem Multiple-Choice-Testverfahren darin zeigen, dass dieses Format in deutschen Schulen nur wenig gebräuchlich ist und Schülerinnen und Schüler bisher relativ ungeübt damit sind. So haben beispielsweise in den ersten PISA-Runden über 10% der Schülerinnen und Schüler, trotz Hinweis auf eine Einzelauswahl, bei einzelnen Aufgaben mehr als eine Antwort angekreuzt.[35] Dem wird bei *fermat* durch die vorrangige binäre Lösungsmöglichkeit entgegengewirkt.

7.4.2 Gleichgewichtung der Items

Der Schwierigkeitsgrad der einzelnen Aufgaben kann durch verschiedene Faktoren, z.B. den sprachlichen Anspruch oder der Anzahl der Wahlmöglichkeiten variieren. Eine Gewichtung von einzelnen Aufgaben, d.h. das Zuweisen von mehr Punkten bei schweren oder wichtigeren Fragen, findet beim Diagnosetest im Rahmen von *fermat* nicht statt.

Bei der Aufgabenentwicklung haben wir zunächst eine Einteilung der Aufgaben in leicht, mittel und schwer vorgenommen. Nach ersten Auswertungen in der Pre-Testphase wurde jedoch deutlich, dass eine solche Einstufung nicht zielführend ist. Mit der in der Sekundarstufe I erworbenen Vorbildung und den sprachlichen Kompetenzen sollte es allen Schülerinnen und Schülern bei Eintritt in das Berufskolleg möglich sein, die Fragen des Diagnosetests, zumindest im Hinblick auf die definierten Kompetenzen I und II, korrekt zu beantworten.

7.4.3 Zeitlicher Rahmen

Die Motivation und die Konzentrationsfähigkeit der Schülerinnen und Schüler findet beim Testverfahren Berücksichtigung. Der Diagnosetest darf nicht zu lang sein, da sonst die Möglichkeit besteht, dass Schülerinnen und Schüler die einzelnen Fragen und Antworten nicht mehr detailliert lesen und Aufgaben daher nach dem Zufallsprinzip lösen.

[35] Siehe dazu: Wuttke 2007, S. 171 ff.

Im Rahmen des VOLI-Modellversuchs konstatiert Efing, dass „sowohl die Test- als auch Interviewergebnisse der Lehrer zeigen, dass sich Berufsschüler nicht länger als 45-60 Minuten am Stück konzentrieren können“[36]. Konzentration, Motivation sowie Leistungs- und Anstrengungsbereitschaft beeinflussen laut Efing das Testergebnis im „Baukasten Lesediagnose“.

Die Durchführungsdauer des Tests im Rahmen von *fermat* soll in der Regel 45 Minuten nicht übersteigen. Damit wird auch der Möglichkeit des Erratens der richtigen Lösung, die das Antwortformat Multiple-Choice bietet und die die Validität des Diagnosetests beeinträchtigen könnte, entgegengewirkt.[37] Ausgeschlossen werden kann ein zufälliges Auswählen der Antwort beispielsweise durch Testverweigerer allerdings nicht.

[36] Aus: Efing 2006, S. 51.
[37] Siehe dazu: Kubinger 2005, S. 158–165.

KAPITEL 8 PILOTIERUNG: Ermittlung von individuellen Förderempfehlungen auf der Basis von Scoreklassen

Um die individuellen Förderbedarfe der Schülerinnen und Schüler angeben zu können, werden die Antworten eines individuellen Tests nach den einzelnen Kompetenzbereichen aufgeschlüsselt und der Score einer bestimmten Kompetenz ermittelt. Dabei ist der Score die Anzahl der richtig beantworteten Items, wobei ein Item eine in Bezug auf die Kompetenz unidimensionale Aufgabe oder Teilaufgabe darstellt.

Die Art der Kompetenzausprägung wird dabei einer Farbe zugeordnet. Dabei steht die Farbe Rot für *geforderte Kompetenzausprägung nicht erreicht*, Gelb für *Kompetenzausprägung teilweise erreicht* und die Farbe Grün für *Kompetenzausprägung erreicht*. Der Übergangsbereich (Gelb) wurde geschaffen, um situative Kontexte wie zum Beispiel Krankheit, mangelhafte Konzentration oder Störungen in der Beurteilung des Förderbedarfs berücksichtigen zu können.

8.1 Das allgemeinsprachliche Text-Lese-Verständnis

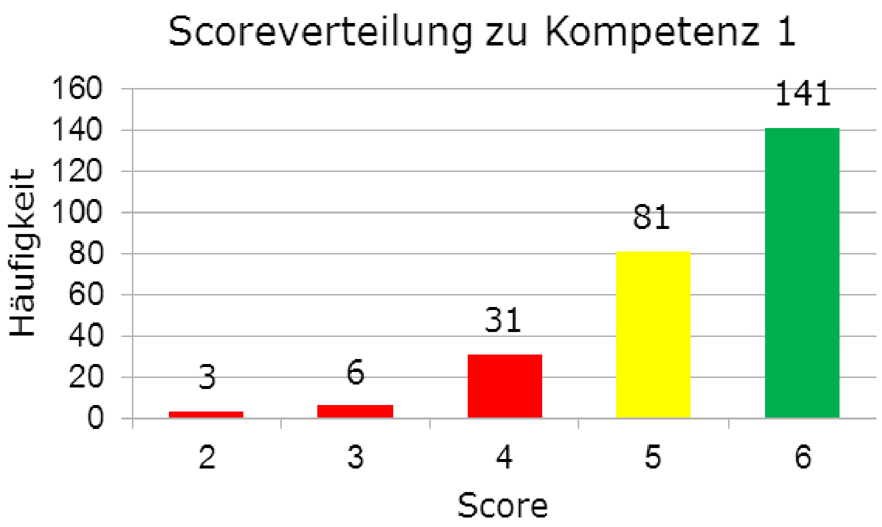


Abbildung 3. Scoreverteilung zum allgemeinsprachlichen Text-Lese-Verständnis.

Abbildung 3 zeigt eine Scoreverteilung, wie sie sich aus 262 beantworteten Tests[38] ergibt. Insgesamt wurden 6 Items zum allgemeinsprachlichen Text-Lese-Verständnis abgefragt. Es ist zu erkennen, dass die niedrigen Scores (0 bis 4), sehr selten vertreten sind. Dies zeigt, dass der Schwierigkeitsgrad des Tests und der Items für die Zielgruppe angemessen ist.

Dem Score 6 wurde die Klasse „Grün“ zugeordnet, die Farbe bedeutet, dass die Schülerinnen und Schüler alle Items korrekt beantwortet haben und daher bezüglich dieser Kompetenz keinen Förderbedarf haben.

Die Klasse „Gelb“ wurde auf Score 5 beschränkt, da hier ein Scorebereich vorliegt, für den nicht eindeutig diagnostiziert werden kann, ob der Score 6 evtl. aufgrund einer Flüchtigkeit oder ähnlichem nicht erreicht wurde.

Die zum Score 4 und kleiner nochmals rapide abfallenden Häufigkeiten zeigen, dass hier größere Probleme im allgemeinsprachlichen Text-Lese-Verständnis vorliegen müssen, daher wurde ihnen die Klasse „Rot“ für Förderbedarf zugeordnet.

Insgesamt ergibt sich hier für 40 von 262 Schülerinnen und Schüler ein klarer Förderbedarf im Bereich allgemeinsprachliches Text-Lese-Verständnis.

[38] Der Test (Test 1) wurde im Mai 2012 mit einem Stichprobenumfang von n=262 durchgeführt.

8.2 Fachsprachliches Text-Lese-Verständnis

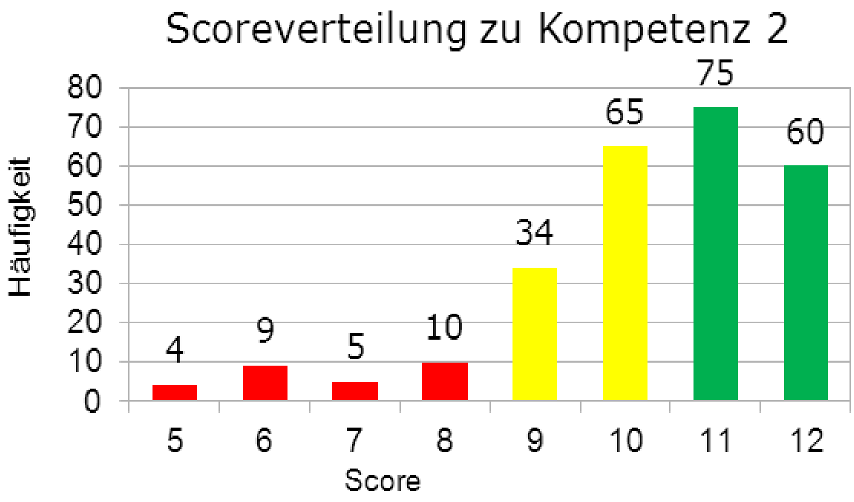


Abbildung 4. Scoreverteilung zum fachsprachlichen Text-Lese-Verständnis.

Die Vorgehensweise bei der Klasseneinteilung für das fachsprachliche Text-Lese-Verständnis basiert auf demselben Datensatz und erfolgte analog der Vorgehensweise beim allgemeinsprachlichen Text-Lese-Verständnis.

Dass in der Scoreverteilung in Abbildung 4 von Score 11 zu Score 12 eine Abnahme der Häufigkeit zu erkennen ist bedeutet, dass es sich hier um lösbare, aber anspruchsvolle Items handelt. Es wurde festgelegt, dass hier die beiden Scores 11 und 12 der Klasse „Grün" zugeordnet werden.

Auch der Übergangsbereich zu den besonders niedrigen Häufigkeiten stellt sich hier breiter dar, weshalb die beiden Scores 9 und 10 der Klasse gelb zugeordnet wurden.

Die Scores kleiner als 9 wurden insgesamt nur 28-mal erreicht, was zeigt, dass diese Schülerinnen und Schüler unbedingten Förderbedarf im fachsprachlichen Text-Lese-Verständnis haben. Diese Scores wurden der „Roten"-Klasse zugeordnet.

8.3 Symbol- und Zahlen-Verständnis

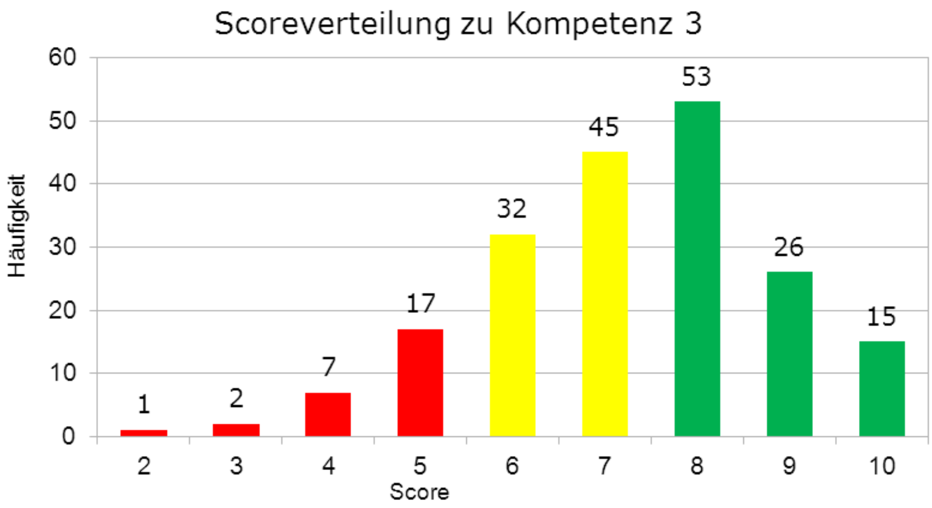


Abbildung 5. Scoreverteilung zum Symbol- und Zahlen-Verständnis.

Die Aufteilung der Klassen für die Kompetenz Symbol- und Zahlen-Verständnis basiert auf einem Datensatz von 198 beantworteten Tests[39].

An den abnehmenden Häufigkeiten zu den Scores 9 und 10 ist zu erkennen, dass dieser Testteil deutlicher differenziert als die bisherigen Testteile. Dies liegt wahrscheinlich auch daran, dass der Schwierigkeitsgrad der Items höher ist.

Die Klasse „Grün" wurde von Score 8 bis 10 festgelegt. Der Übergangsbereich liegt bei den Scores 6 und 7. Die Scores darunter wurden der Klasse „Rot" zugeordnet.

8.4 Beschreibung des Musters bei der Einteilung in Scoreklassen

Im Folgenden wird die Klassenfestlegung der ersten drei Kompetenzen erläutert. Die Einteilung der Klassen erfolgt aus dem Gesamtergebnis von durchgeführten Tests. Danach wurde nach subjektiver Einschätzung der Mitglieder der Projektgruppe *fermat* abgewogen, ob

[39] Dieser Test (Test 2) wurde im Dezember 2012 mit einem Stichprobenumfang von n=198 durchgeführt.

die Anforderung des Tests dem Kompetenzstand der Schülerinnen und Schüler entspricht. Analysiert man diese Festlegung ergibt sich folgendes Bild aus den drei oben gezeigten Auswertungsdiagrammen (Abbildung 3 bis Abbildung 5).

Die Festlegung für die Klasse Grün erfolgte ab dem Score, bei dem 50% bis 55% der Schülerinnen und Schüler den maximalen Score oder weniger erreichen. Die Klasse Gelb wurde so gelegt, dass ca. 30% bis 35% der Schülerinnen und Schüler einen Score unterhalb der Klasse Grün erzielen. Und der Klasse Rot werden die Scoreergebnisse zugeordnet, die 15% bis 20% der Schülerinnen und Schüler erreicht haben und unterhalb der Ergebnisse der Klasse Gelb liegen. Bezüglich aller drei Kompetenzen liegt der am häufigsten erzielte Score im Bereich der Klasse Grün.

Zusätzlich wurden die Unterschiede zwischen den Scoreergebnissen betrachtet. Zwischen den Klassen Gelb und Rot ist jeweils eine deutliche Abnahme der Fallzahlen festzustellen, die nach unserer Auffassung eine Grenze zwischen Kompetenzausprägungen darstellt. Für den Übergang von Grün zu Gelb ist diese sprunghafte Abnahme nicht zu erkennen. Die Abgrenzung erfolgte hier derart, dass der am häufigsten erzielte Score der Klasse Grün zugeordnet wird.

Zusammenfassend ergibt sich die in der folgenden Tabelle dargestellte Klassenaufteilung für die verschiedenen Kompetenzen:

Tabelle 2: Klassen von erreichten Scores

Kompetenz	grün	gelb	rot
Allgemeinsprachliches Text-Lese-Verständnis (6 Items)	6	5	0-4
Fachsprachliches Text-Lese-Verständnis (12 Items)	11-12	9-10	0-8
Symbol- und Zahlen-Verständnis (10 Items)	8-10	6-7	0-5

8.5 Zuordnung der Scoreklassen zum Förderbedarf

Grundsätzlich wird nach der Einteilung in die drei Klassen der Förderbedarf für jede Kompetenz einzeln ermittelt, da wir davon ausgehen, dass die Kompetenzen unabhängig voneinander sind. Das Auswerteverfahren basiert ebenfalls auf dieser Annahme. Für die Klasse Grün besteht kein Förderbedarf in der jeweiligen Kompetenz, für die Klasse Rot besteht Förderbedarf in der jeweiligen Kompetenz. Bei Schülerinnen und Schülern, die der Klasse Gelb in einer Kompetenz zugeordnet werden, muss die jeweilige Lehrkraft entscheiden, ob sie die Schülerin oder den Schüler fördern möchte.

Während einer Datenanalyse wurde versucht, eine Abhängigkeit der Kompetenzen untereinander herauszufinden. So könnte man vermuten, dass jemand, der in Kompetenz II in der Klasse Grün landet, auch in der Kompetenz I ein besseres Ergebnis erzielt. (Wer mathematisches Text-Lese-Verständnis hat, der hat auch allgemeines Text-Lese-Verständnis.) Diese Analyse führt allerdings in ein Dilemma. Wenn wir ein unidimensionales Testverfahren wählen, dann können wir damit nicht zeigen, dass es Abhängigkeiten von mehr als einer Kompetenz gibt. Denn dann wäre das Testverfahren nicht mehr unidimensional. Ferner war die Gesamtanzahl der durchgeführten Tests für eine solche Aussage zu klein. Während sich in den Tests 1 und 2 jeweils diese Abhängigkeit hätte zeigen können, spiegelte sich diese in der Summe der beiden Stichproben nicht wieder. Die Testreliabilität wäre damit verletzt gewesen.

Nach mehreren Testdurchläufen stellte sich die Frage, ob eine zuverlässige und statistisch reliable Aussage über die Kompetenzausprägung möglich ist. Bei der Untersuchung des allgemeinsprachlichen Text-Lese-Verständnisses werden 6 Items untersucht. Die Wahrscheinlichkeit, alle Fragen durch Raten richtig zu beantworten, beträgt damit 1/64. Für die Kompetenzausprägung „teilweise erreicht“ ist die Ratewahrscheinlichkeit 6/64. Insgesamt besteht bei der geringen Anzahl von Items die Möglichkeit, dass Schülerinnen und Schüler mit Förderbedarf durch den Test nicht entdeckt werden.

Es sollte also ausgeschlossen werden, dass Scoreklassen, die keine Förderung nach sich ziehen, durch Raten erzielt werden können. Dies ist durch Erhöhen der Anzahl der Items möglich. Für die Modellierung nach dem Rasch-Modell wird dadurch zwar keine zusätzliche Information gewonnen, jedoch erhöht sich die Reliabilität der Aussagen.

Kapitel 9 Technische Umsetzung

Die technische Umsetzung stellte die Gruppe von Anfang an vor große Herausforderungen, da die Auswertung der Daten hinsichtlich einer Förderempfehlung, wie auch hinsichtlich der weiteren Pilotierung automatisiert erfolgen sollte.

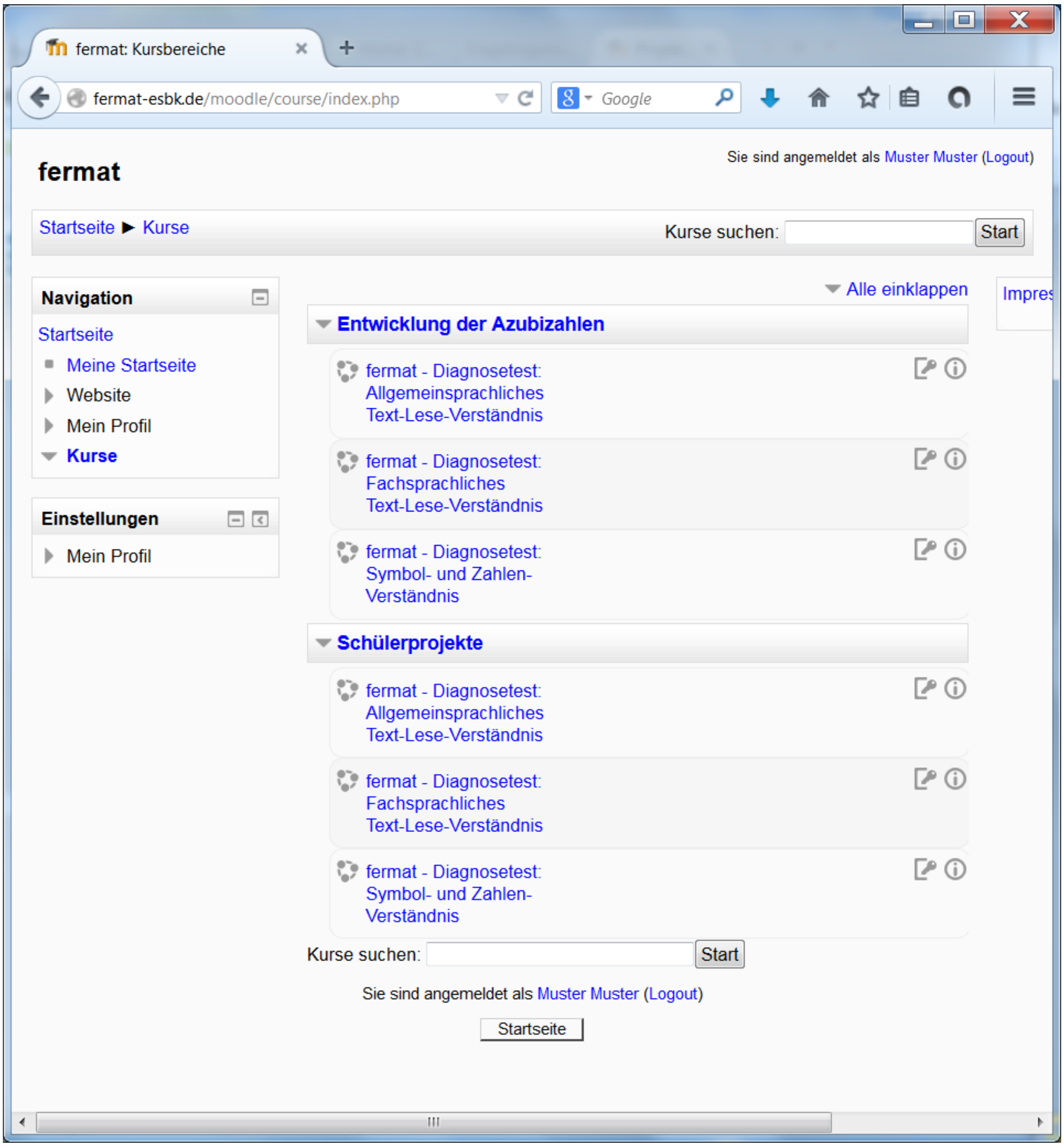
Die Analyse für eine Förderempfehlung beschränkt sich auf das Auszählen richtig und falsch beantworteter Items. Für die Pilotierung hingegen sind alle Einzelantworten relevant, da neben der Faktorenanalyse in diesem Schritt ggf. entschieden werden muss, ob Items zusammengefasst werden können oder müssen.

9.1 Erste Versuche mit GrafStat

Erste Versuche den Test online verfügbar zu machen erfolgten mit dem Programm GrafStat von Uwe W. Diener (http://www.grafstat.de/), da dies auf einfache Weise eine Erstellung von Fragebögen ermöglicht und gleichzeitig eine gute Datenbankinfrastruktur zur Verfügung stellt. Leider konnten jedoch einige unserer Anforderungen an die Darstellung der Webseite, an die Archivierung von veränderten Fragen usw. nicht ohne viel Handarbeit am html-Code umgesetzt werden. Zusätzlich sollte den Teilnehmenden Schülerinnen und Schüler direkt ihre Förderempfehlung mitgeteilt werden, da GrafStat ein Werkzeug für Umfragen mit anonymen Umfrageergebnissen ist, wurde nach Alternativen gesucht.

9.2 Moodle

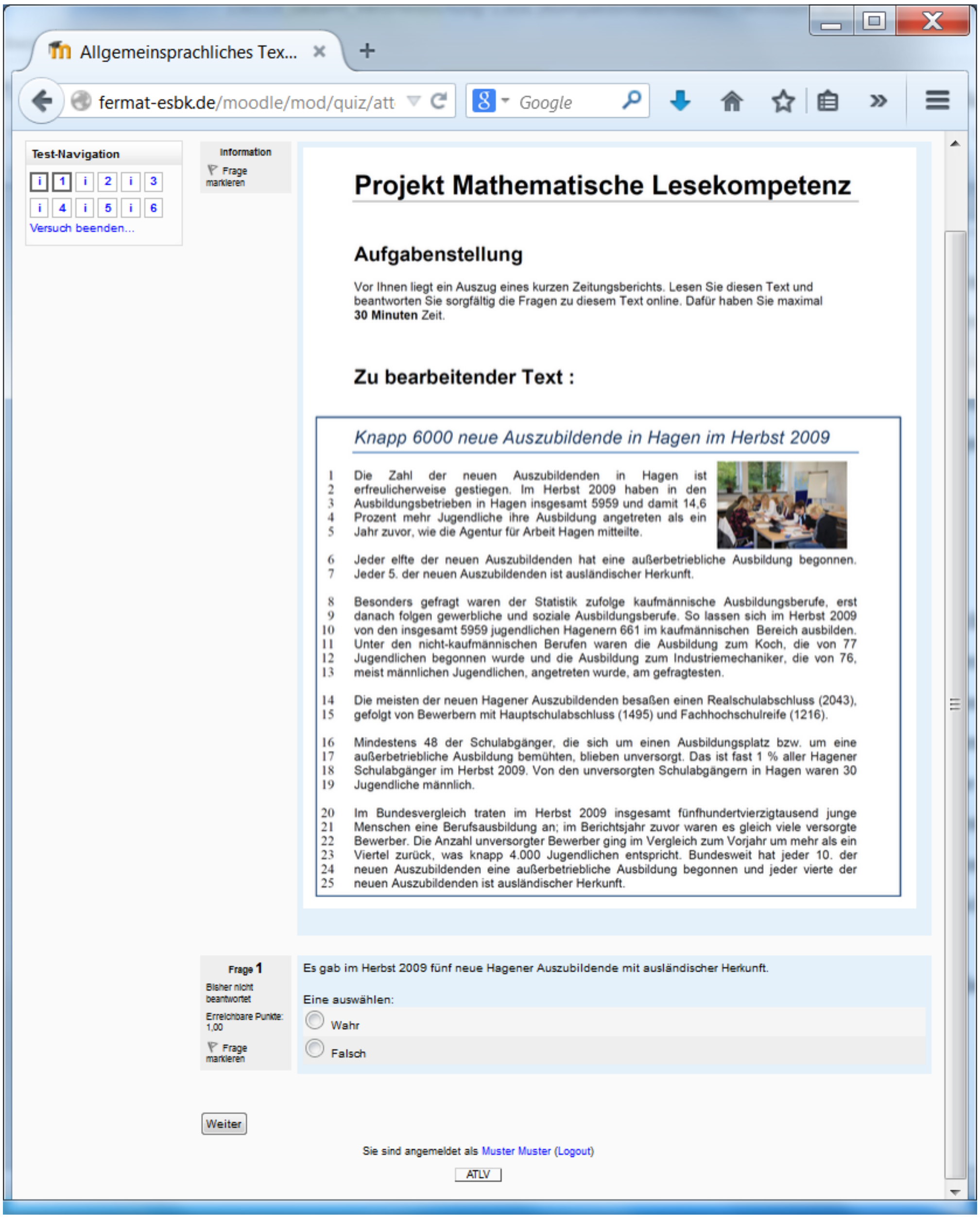
Als Lernplattform mit Onlinetestfunktion an Schulen hat sich inzwischen Moodle etabliert. Moodle bietet im Wesentlichen die erforderlichen Funktionen bzw. lässt sich durch Erweiterungspakete um benötigte Funktionen erweitern. Inzwischen wurden von *fermat* zwei Onlinetests in moodle eingepflegt und können unter http://fermat-esbk.de/ durchgeführt werden.



9.2 Technische Herausforderungen

Als Hinsichtlich der Nutzung der Tests im Schulalltag bestehen jedoch noch einige spezielle Probleme. Zum Beispiel verläuft der Selbstregistrierungsprozess von Schülern im Moodlesystem normalerweise so ab, dass sich die Schülerinnen und Schüler unter Angabe einer Emailadresse anmelden, eine Mail mit einem Bestätigungslink erhalten und sich damit vollständig für die Benutzung freischalten können. Erst dann kann einem Kurs (bei fermat

Test) beigetreten werden und der Test durchgeführt werden. Leider sind die Mails der Schüler in den Schulen oft nicht abrufbar und daher ist diese Methode für *fermat* nicht praktikabel. Alternativ wurde das aktuell verwendete Moodlesystem so eingestellt, dass eine Mailüberprüfung nicht stattfindet und sich so jeder auf der Seite registrieren kann, dabei werden bewusst Sicherheitsrisiken eingegangen. Wir hoffen auf professionelle Hilfe an dieser Stelle, entweder um das System abzusichern, oder um den Registrierungsprozess sicherer zu gestalten.

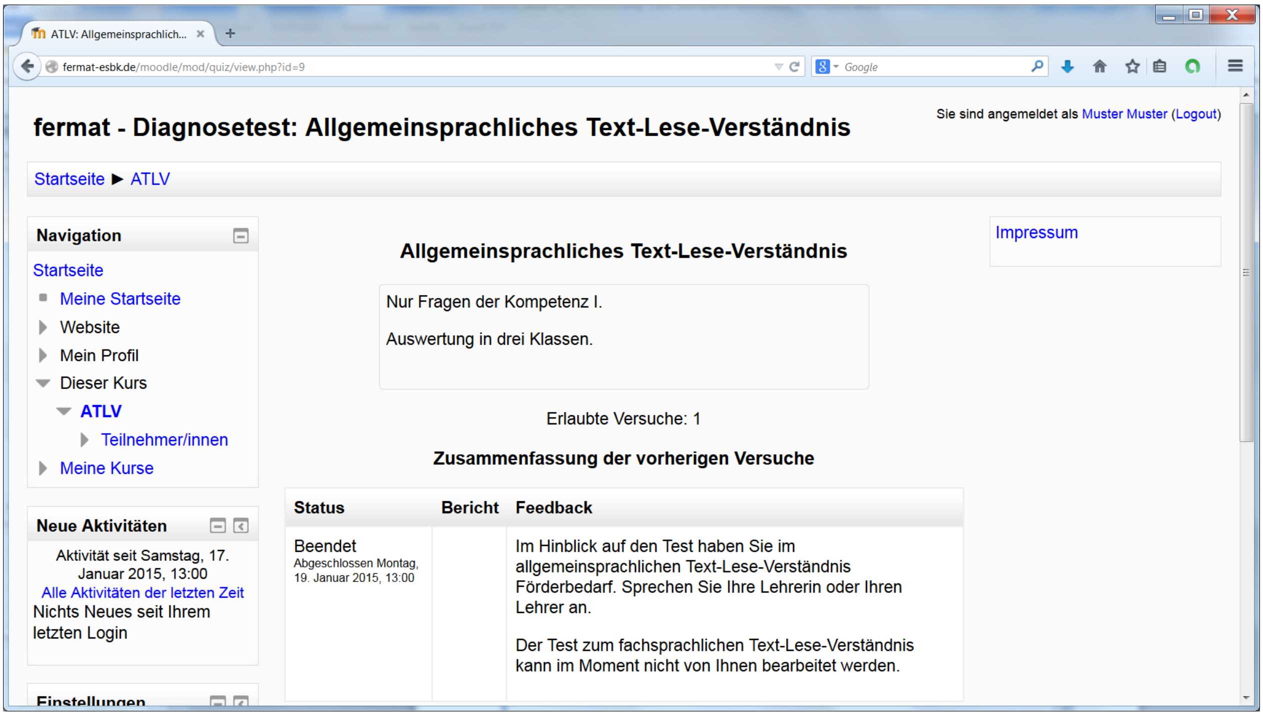


9.3 Testdurchführung

Nach einer erfolgreichen Selbstregistrierung, wie im vorhergehenden Abschnitt beschrieben, können die Schülerinnen und Schüler, nach Vorgabe der betreuenden Lehrerinnen und Lehrer einen der Tests aussuchen. Zur Einschreibung ist ein Kennwort erforderlich, welches verhindern soll, dass beliebige Tests ausgewählt werden können. Dieses Kennwort ist zurzeit nur durch persönlichen Kontakt mit den Mitgliedern der ***fermat***-Gruppe erhältlich. Nach erfolgter Einschreibung, beginnt der Test mit Fragen zur ersten Kompetenz Allgemeinsprachliches Text-Lese-Verständnis.

Bei erfolgreicher Beantwortung der Fragen dieser Kompetenz wird der betreffenden Schülerin bzw. dem betreffenden Schüler das nächste Kennwort für den Test der zweiten Kompetenz bekannt gegeben usw.

Nach jeder Test Stufe bekommen die Schülerinnen und Schüler eine Förderempfehlung auf der Basis ihrer Antworten.



9.4 Datenauswertung

Die für die weitere Testentwicklung erforderlichen Daten werden von Moodle in einer Datenbank gespeichert. Das Moddlefrontend erlaubt es diese Daten zu exportieren und mit den beschriebenen Methoden zu analysieren.

Kapitel 10 Ausblick

10.1 Erfahrungen mit *fermat*

Auf lexikalischer Ebene zeigt es sich, dass Schülerinnen und Schüler häufig nicht über ausreichende Wortbildungskenntnisse verfügen, um Komposita, Verben mit Präfix oder abgeleitete Nomen bzw. Adjektive zu verstehen.[40] Zum Beispiel ist *flächenmäßig* ein abgeleitetes Adjektiv, das sich aus dem Nomen *Fläche* und dem Suffix *mäßig* (geht auf Maß zurück) zusammensetzt und von den Schülerinnen und Schülern adäquat dekodiert werden muss.

Auf syntaktischer Ebene entstehen vermehrt Schwierigkeiten beim Verstehen von Passivkonstruktionen, Attributen (mit und ohne Präposition) sowie Haupt- und Nebensatzkonstruktionen. Passivformulierungen sind beispielsweise besonders für leseungeübte Schüler schwierig, da die handelnde Person nicht explizit genannt wird.

Deshalb sind bei der Erstellung von Text- bzw. Fördermaterial auf sprachlich exakte und eindeutige Formulierungen sowie die Vermeidung von sprachlichen Stolpersteinen zu achten. Die Grundprinzipien der Textoptimierung zur Prüfungserstellung von Schlenker-Schulte/Wagner sind als Leitfaden sehr hilfreich. Prüfungstexte, die auf der Wort-, Satz- und Textebene aufbereitet wurden, ermöglichen Schülerinnen und Schülern einen verbesserten Zugang zu den Aufgaben.

10.2 Möglichkeiten zum Erstellen von Fördermaterial

Das Landesinstitut für Lehrerbildung und Schulentwicklung in Hamburg hat im Jahre 2008 einen Aufsatz zur Förderung der Lesekompetenz in der Mathematik im Bereich der

[40] Siehe dazu: Günther 2013, S. 11/12.

Sekundarstufe I herausgegeben, in dem vielfältige Anregungen zum Erstellen von Fördermaterialien benannt werden.[41]

Folgende, exemplarische Aufgaben können als Grundlage für Fördermaterial dienen.

Aufgabe 1: Begrenzende und vergleichende Ausdrücke verwenden

Beispiele: *genau so viel*, *höchstens, mindestens*, *mehr als,* ...

Kompetenz: Allgemeinsprachliches Text-Lese-Verständnis

Ergänzen Sie den Lückentext sinnvoll mit den Wörtern aus dem untenstehenden Kasten.

um etwa 490	exakt 24	mehr als 500	unter 2:34
knapp 40	genau 32,18	zwischen 20	weniger als 2
genau so viel wie 48,27			

Der Ursprung des modernen Marathonlaufs

Der Geschichtsschreiber Herodot berichtet über den griechischen Boten Pheidippides, der _______________ v. Chr. von Athen in _______________ Tagen nach Sparta gelaufen war, um Hilfe im Krieg gegen die Perser zu suchen. Daraus formten _______________ Jahre danach Plutarch und Lukian von Samosata eine Legende, der zufolge ein Läufer sich nach dem Sieg der Athener in der Schlacht von Marathon auf den _______________ Kilometer langen Weg nach Athen gemacht habe und dort nach der Verkündung seiner Botschaft „Wir haben gesiegt" tot zusammengebrochen sei.

Bei den ersten Langstreckenläufen der Neuzeit dachte niemand an diese Historie. Als Vergnügen für die Zuschauer wurden im ausgehenden 18. Jahrhundert die ersten Langstreckenläufe in der Art eines sportlichen Wettkampfes ausgetragen.

[41] Siehe dazu: Bergunde 2008, S. 9 und 13.

In England und den Vereinigten Staaten absolvierten die „Fußläufer“ Strecken _______________ Meilen (das entspricht _______________ km) und 30 Meilen (das ist _______________ km). Im Jahr 1808 lief ein Mann mit dem Namen Blewet über _______________ Meilen (38,62 km) eine Zeit von _______________ Stunden.

Erst 1921 hat der internationale Verband für Leichtathletik (IAAF) die Distanz von ***genau 42,195*** Kilometer als offizielle Streckenlänge für einen Marathonlauf festgelegt.

Quelle: http://de.wikipedia.org/wiki/Marathonlauf#Ursprung_des_modernen_Marathonlaufs

Lösung

Der Ursprung des modernen Marathonlaufs

Der Geschichtsschreiber Herodot berichtet über den griechischen Boten Pheidippides, der ***um etwa 490*** v. Chr. von Athen in ***weniger als 2*** Tagen nach Sparta gelaufen war, um Hilfe im Krieg gegen die Perser zu suchen. Daraus formten ***mehr als 500*** Jahre danach Plutarch und Lukian von Samosata eine Legende, der zufolge ein Läufer sich nach dem Sieg der Athener in der Schlacht von Marathon auf den ***knapp 40*** Kilometer langen Weg nach Athen gemacht habe und dort nach der Verkündung seiner Botschaft „Wir haben gesiegt“ tot zusammengebrochen sei.

Bei den ersten Langstreckenläufen der Neuzeit dachte niemand an diese Historie. Als Vergnügen für die Zuschauer wurden im ausgehenden 18. Jahrhundert die ersten Langstreckenläufe in der Art eines sportlichen Wettkampfes ausgetragen.

In England und den Vereinigten Staaten absolvierten die „Fußläufer“ Strecken ***zwischen 20*** Meilen (das entspricht ***genau 32,18*** km) und 30 Meilen (das ist ***genau so viel wie 48,27*** km). Im Jahr 1808 lief ein Mann mit dem Namen Blewet über ***exakt 24*** Meilen (38,62 km) eine Zeit von ***unter 2:34*** Stunden.

Erst 1921 hat der internationale Verband für Leichtathletik (IAAF) die Distanz von ***genau 42,195*** Kilometer als offizielle Streckenlänge für einen Marathonlauf festgelegt.

Aufgabe 2: Begrenzende / vergleichende Ausdrücke in Symbole umsetzen

Beispiele: *genau so viel*, *höchstens, mindestens*, *mehr als, … in <, =. >*

Kompetenz: Symbol- und Zahlen-Verständnis

Ordnen Sie die Ausdrücke aus dem untenstehenden Kasten den mathematischen Symbolen < (kleiner als), = (gleich) und > (größer als) zu.

um etwa 490	exakt 24	mehr als 500	unter 2:34
knapp 40	genau 32,18	zwischen 20	weniger als 2
höchstens 34	genau so viel wie 48,27		

< **kleiner als**	= **gleich**	> **größer als**

Nicht zuzuordnen sind:

____

Lösung

< kleiner als	= gleich	> größer als
höchstens 34	genau 32,18	mehr als 500
unter 2:34	exakt 24	
knapp 40	genau so viel wie 48,27	
weniger als 2		

Nicht zuzuordnen sind:

um etwa 490, zwischen 20 (und)

Aufgabe: Mathematische Begriffe kennen

Beispiele: *multiplizieren, dividieren, …*

Kompetenz: Mathematisches Text-Lese-Verständnis

Ergänzen Sie den Lückentext sinnvoll mit den Wörtern aus dem untenstehenden Kasten.

durch	**multipliziert**	**addiert**
subtrahiert	**dividiert**	

400 – 40 → Die Zahl 40 wird von der Zahl 400 ______________ .

400 : 40 → Die Zahl 400 wird ______________ die Zahl 40 ______________ .

400 + 40 → Zur Zahl 400 wird die Zahl 40 ______________ .

400 · 40 → Die Zahl 400 wird mit der Zahl 40 ______________ .

Lösung

400 – 40 → Die Zahl 40 wird von der Zahl 400 ***subtrahiert***.

400 : 40 → Die Zahl 400 wird ***durch*** die Zahl 40 ***dividiert*** .

400 + 40 → Zur Zahl 400 wird die Zahl 40 ***addiert***.

400 · 40 → Die Zahl 400 wird mit der Zahl 40 ***multipliziert*** .

10.3 Ausblick Fördermaterialien

Der Bereich der Fördermaterialien wird im kommenden Halbjahr bis Juli 2015 im Mittelpunkt der Arbeit von ***fermat*** stehen.

Kapitel 11 Anhang

11.1 Anhang 1

Beispielhafte Durchführung der Bestimmung der Parameter eines Rasch Modells mittels der Software R

Da das Softwarepaket R mit den jeweils benötigten Zusatzpaketen unter freier Lizenz zur Verfügung steht, werden die Parameterberechnungen für ein Rasch Modell mit Hilfe dieser Software realisiert.

Die Software steht unter folgendem Link zum Download zur Verfügung:

http://cran.r-project.org/

R ist Kommandozeilen-basiert. Um den Umgang etwas zu vereinfachen sollte der R-Commander installiert werden.

Dazu muss im Befehlsfenster von R (Terminalfenster) der Befehl

```
install.packages("Rcmdr")
```

eingegeben werden.

Der Start des R-Commanders erfolgt durch das Laden der Bibliothek:

```
library(Rcmdr)
```

Eine Anleitung und Download-Links findet man unter:

http://www.uni-ulm.de/fileadmin/website_uni_ulm/mawi.inst.110/lehre/ss09/WiStat/R.pdf

http://socserv.mcmaster.ca/jfox/Misc/Rcmdr/installation-notes.html

Der R-Commander

Falls Befehle manuell eingegeben werden müssen, so erfolgt dies im oberen Fenster des R-Commanders (Skriptfenster). Durch den Button „Befehl ausführen” wird der Befehl ausgeführt, in dem sich der Cursor befindet, bzw. es werden die markierten Befehle ausgeführt.

Import

Stata Dateien und CSV-Dateien können über den Menüpunkt „Datenmanagement” importiert werden. Speichern und öffnen von Skriptdateien erfolgt unter „Datei”. Unter „Datenmanagement” können auch Manipulationen an den Daten vorgenommen werden wie z.B. das Erstellen einer neuen Datenmatrix aus einer geladenen Datenmatrix. Dies kann nötig sein, da R-Befehle immer nur auf ganzen Matrizen arbeiten und daher die Datenmatrizen nur die für die Berechnung notwendigen Daten enthalten dürfen.

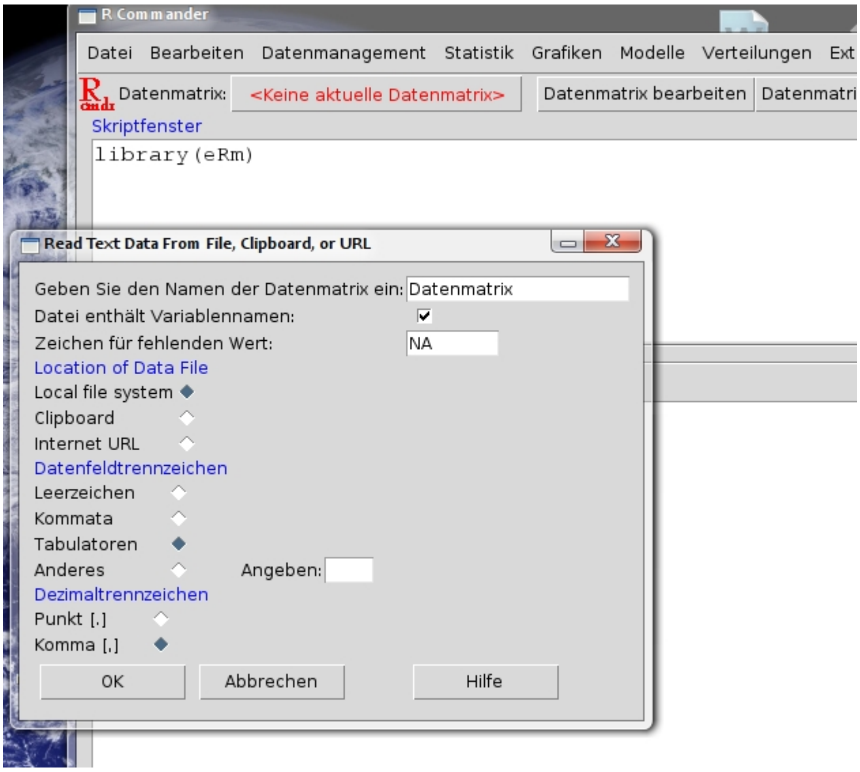


Abbildung 3: Menü zum Datenimport einer CSV-Datei.

Abbildung 3 zeigt das Menü zum Datenimport einer CSV-Datei. Angegeben werden kann das Trennzeichen der Daten, das Dezimaltrennzeichen und welcher Wert ein Missing darstellt. Ein Missing ist dabei ein vom Anwender definierter Wert für einen systematisch fehlenden Wert wie z.B. eine nicht beantwortete Aufgabe oder die Nichtangabe des Geschlechts.

Die dazugehörige Befehlsyntax sieht wie folgt aus:

```
> Datenmatrix <- read.table("/home/christoph/Desktop/CT.FU.csv", header=TRUE, sep="\t",
na.strings="NA", dec=",", strip.white=TRUE)
```

<u>Faktorenanalyse zur Überprüfung der Unidimensionalität der entwickelten Items</u>

Wie ausgeführt sollen die Aufgaben aller vier Tests jeweils für sich nur eine Dimension abbilden bzw. innerhalb der Dimensionen vorher definierte Unterdimensionen. Dazu werden entsprechende Aufgaben entwickelt und auf die Hypothese überprüft, dass die generierten Aufgaben nur eine Kompetenz oder Unterkompetenz abbilden. Diese Überprüfung erfolgt mit einem Verfahren auf Basis der Regressionsrechnung, der Faktorenanalyse.

Eine Idee der Faktorenanalyse ist es, zu einem gegebenen Datenbestand korrelative Zusammenhänge von Items zu identifizieren, sogenannte Faktoren.[42] Diesen latenten Strukturen, die das Gemeinsame der Aufgaben darstellen, wird die Eigenschaft zugesprochen, dass genau sie die Kompetenz/Dimension sind, die die Aufgabenlösung verursachen. Idealerweise würde sich für die Aufgabenentwicklung eines Tests daher nur ein Faktor ermitteln lassen. Dies ist aber in der empirischen Realität nicht zu finden. Unidimensionalität wird immer nur approximativ vorfindbar, konstruierbar sein.[43] Konstruierbar sind jedoch Aufgaben, die von einem Faktor dominiert werden.[44] Kennzeichen einer solchen Dominanz ist die Existenz eines Faktors mit einem großen Eigenwert im Gegensatz zu den anderen von der Faktorenanalyse entdeckten Faktoren.

---

[42] Siehe dazu: Bortz 1999; Wirtz, Nachtigall 2004; Hair 2010.

[43] Siehe dazu: Christopherson, Grape 2007; Planinic 2010.

[44] Sihe dazu: Sijtsma, Verweij 1999.

Den Befehl zur Faktorenanalyse stellt R-Commander unter „Statistik – Dimensionsreduktion und Klassifizieren“ zur Verfügung. Die Voreinstellungen aus Abbildung 5 können übernommen werden.

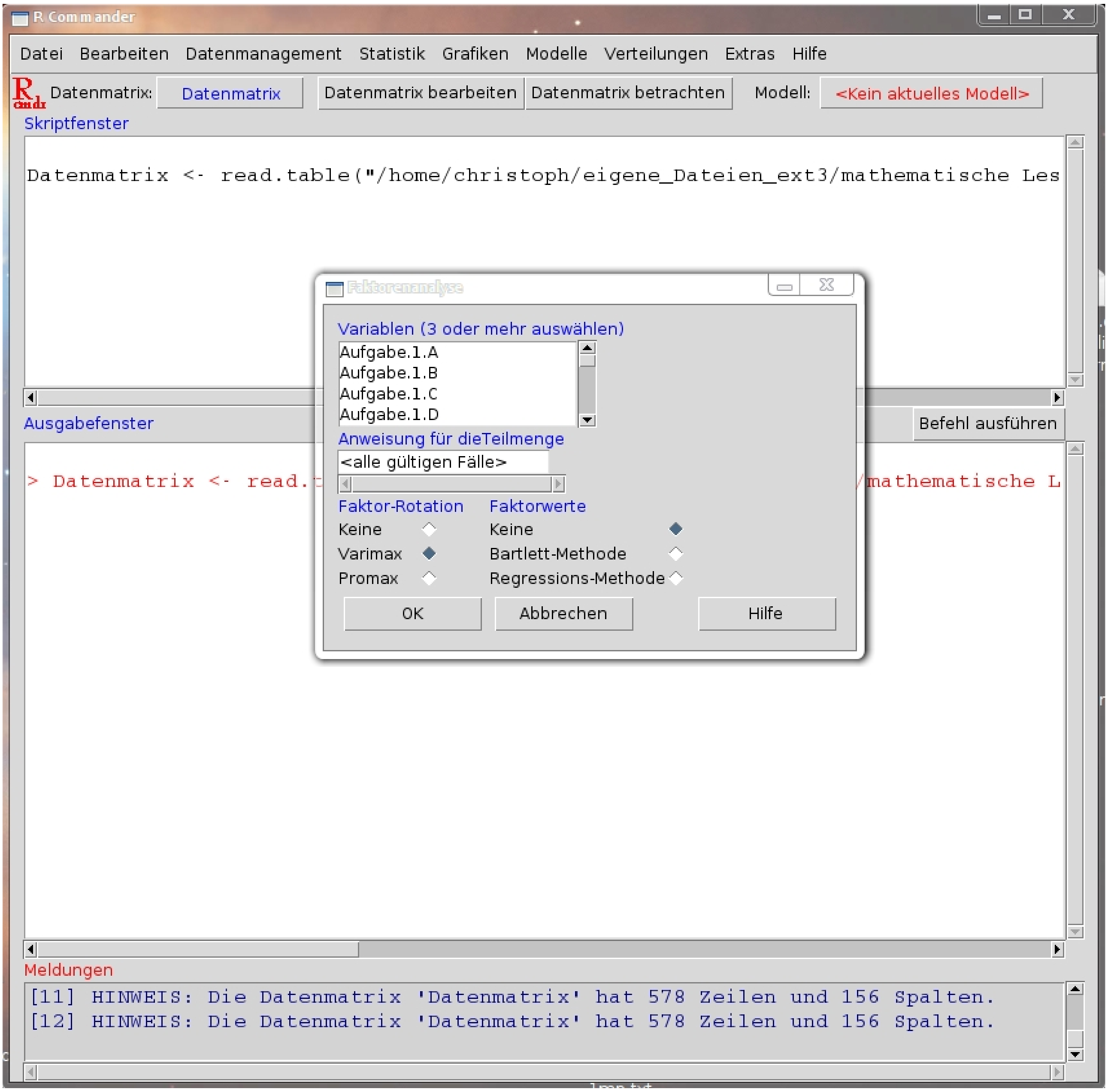


Abbildung 3: Faktorenanalyse mit Hilfe des Maximum-Likelihood-Verfahrens im R-Coamder

Nach der Auswahl der Variablen kann noch die Anzahl der zu extrahierenden Faktoren bestimmt werden: In diesem Fall ein Faktor.

Der erzeugte Output zeigt nach einer langen Liste der Faktorladungen (Korrelation der Aufgaben zum Faktor) den Eigenwert unter „*SS-Loadings*" und etwas tiefer die Varianz der Residuen an:

```
The root mean square of the residuals (RMSR) is  0.06

The df corrected root mean square of the residuals is  0.08
```

Ebenso wie die Reliabilität:

```
Tucker Lewis Index of factoring reliability =  0.312
```

und die Korrelation der Scores mit den Faktoren:

```
Correlation of scores with factors             0.94

Multiple R square of scores with factors       0.88

Minimum correlation of possible factor scores  0.76
```

Für die angegebenen fiktiven Beispielwerte ergibt sich eine geringe Restvarianz (Residuen) und eine hohe Korrelation des Faktors mit dem Score, aber nur eine geringe Reliabilität.

Ein relevanter Eigenwert muss grundsätzlich größer Eins sein, da ansonsten der dazugehörige Faktor gerade einmal so viel Varianzaufklärung besitzt wie eine einzige Variable. Dies bedeutet, dass je höher der Eigenwert ist, desto mehr der gesamten Varianz des Datensatzes klärt der dazugehörige Faktor auf. Maximal kann der Eigenwert einen Wert, der der Anzahl der in der Analyse verwendeten Variablen entspricht, annehmen.

Zudem laden/korrelieren nicht alle Aufgaben gleich gut mit dem Faktor. Als Grenzwert hat sich ein Wert von 0.3 etabliert. Aufgaben, die geringer als dieser Wert mit dem Faktor korrelieren, werden daher von der weiteren Betrachtung ausgeschlossen.

Die weitere Analyse, die Rasch Analyse, betrachtet daher nur noch die Aufgaben, die eine Korrelation größer 0.3 mit dem Faktor aufweisen.

Da die vom R-Commander zur Verfügung gestellte Faktorenanalyse mittels dem Verfahren des Maximum-Likelihood berechnet wird, können die grundsätzlichen Probleme der Korrelationsberechnung von Variablen, die nicht intervallskaliert sind, vernachlässigt werden.

Falls dennoch Varianten der Faktorenanalyse berechnet werden sollen und um beispielsweise die Berechnungsgrundlage der Faktoren auch auf der Tetrachorische-Matrix an Stelle der Korrelationsmatrix basieren zu lassen, so muss das Paket „psych" installiert werden.[45]

Dazu muss im Befehlsfenster der Befehl

```
install.packages(„psych)
```

eingegeben und ausgeführt werden. Über „Extras - Lade Pakete" muss das Paket aber noch geladen werden. Eine Anleitung zu dem Paket findet sich unter:

```
cran.r-project.org/web/packages/psych/psych.pdf
```

<u>Raschanalyse – Package eRm</u>

Mit

```
install.packages("eRm", dependencies = TRUE)
```

wird das Paket zur Raschanalyse mittels des „Conditional Maximum Likelihood" installiert und mit

```
> library(eRm)
```

geladen.[46]

---

[45] Siehe dazu: Werner 2012.

[46] Siehe dazu: Mair et.al. 2007, 20011; Poistingl et.al. 2011.

Das Paket kann neben einem Rasch-Modell noch weitere Modelle berechnen wie z.B. Partial Credit Modell (PCM) oder Rating Scale Modell (RSM), die aber alle nicht die erwähnte messtechnische Besonderheit des Rasch Modells aufweisen, dass Personen unabhängig von der Wahl der Items miteinander verglichen werden können.

RM-Analyse (Rasch Modellanalyse)

```
> res.rm<-RM(Datenmatrix)
```

Der obige Befehl startet die Analyse der *Datenmatrix* mit der Methode **RM** (Rasch Modellanalyse) und übergibt die Ergebnisse der Berechnung an das Objekt *res.rm* (Bezeichnung dieses Objekts frei wählbar). Der Befehl zur Berechnung eines Raschmodells lautet **RM**, der für ein Partial Credit Modell **PCM** und der für ein Rating Scale Modell **RSM**. Zur Berechnung eines anderen Modelltypus muss daher in der obigen Programmzeile nur RM durch PCM oder RSM ersetzt werden und eventuell der Objektname geändert werden.

Ausgabe der Itemlokationen / Aufgabenschwierigkeit als Liste

Dem Befehl zur Ausgabe wird das berechnete Objekt *res.rm* übergeben:

```
> summary(res.rm)

Results of RM estimation:

Call:  RM(X = Datenmatrix)

Conditional log-likelihood: -12789.43
Number of iterations: 123
Number of parameters: 41

Item (Category) Difficulty Parameters (eta) with 0.95 CI:
            Estimate Std. Error lower CI upper CI
CT.FU1.c2     -2.311      0.120   -2.546   -2.075
CT.FU2.c1     -1.564      0.122   -1.803   -1.325
```

CT.FU2.c2	-3.472	0.108	-3.684	-3.260
CT.FU3.c1	-0.005	0.117	-0.235	0.224
CT.FU3.c2	-2.496	0.081	-2.654	-2.338
CT.FU4.c1	0.237	0.089	0.063	0.411
CT.FU4.c2	-0.935	0.080	-1.092	-0.778
CT.FU5.c1	-0.713	0.069	-0.848	-0.578
CT.FU5.c2	0.538	0.137	0.270	0.807
CT.FU6.c1	1.977	0.196	1.594	2.361
CT.FU6.c2	-1.590	0.072	-1.731	-1.449
CT.FU7.c1	3.200	0.276	2.659	3.741
CT.FU7.c2	-0.341	0.081	-0.500	-0.181
CT.FU8.c1	-0.785	0.071	-0.924	-0.645
CT.FU8.c2	-0.286	0.105	-0.492	-0.079
CT.FU9.c1	1.132	0.113	0.910	1.355
CT.FU9.c2	-0.582	0.081	-0.740	-0.424
CT.FU10.c1	0.388	0.136	0.123	0.654
CT.FU10.c2	-2.564	0.080	-2.721	-2.407
CT.FU11.c1	1.514	0.140	1.239	1.788
CT.FU11.c2	-0.960	0.075	-1.107	-0.812
CT.FU12.c1	2.504	0.201	2.110	2.898
CT.FU12.c2	-0.447	0.080	-0.604	-0.289
CT.FU13.c1	2.672	0.267	2.147	3.196
CT.FU13.c2	-1.544	0.072	-1.685	-1.404
CT.FU14.c1	1.280	0.129	1.026	1.533
CT.FU14.c2	-1.060	0.075	-1.207	-0.914
CT.FU15a.c1	0.768	0.102	0.569	0.968
CT.FU15a.c2	-0.750	0.080	-0.907	-0.594
CT.FU15b.c1	1.005	0.110	0.790	1.220
CT.FU15b.c2	-0.682	0.080	-0.838	-0.525
CT.FU15c.c1	0.568	0.094	0.384	0.752
CT.FU15c.c2	-0.641	0.082	-0.802	-0.480
CT.FU15d.c1	1.739	0.139	1.467	2.012
CT.FU15d.c2	-0.219	0.085	-0.386	-0.051
CT.FU15e.c1	2.846	0.213	2.428	3.264

CT.FU15e.c2	0.637	0.107	0.427	0.847
CT.FU15f.c1	3.327	0.287	2.765	3.889
CT.FU15f.c2	-0.150	0.085	-0.316	0.017
CT.FU16.c1	1.379	0.133	1.118	1.639
CT.FU16.c2	-0.979	0.075	-1.127	-0.831

Item Easiness Parameters (beta) with 0.95 CI:

	Estimate	Std. Error	lower CI	upper CI
beta CT.FU1.c1	2.637	0.107	2.427	2.846
beta CT.FU1.c2	2.311	0.120	2.075	2.546
beta CT.FU2.c1	1.564	0.122	1.325	1.803
beta CT.FU2.c2	3.472	0.108	3.260	3.684
beta CT.FU3.c1	0.005	0.117	-0.224	0.235
beta CT.FU3.c2	2.496	0.081	2.338	2.654
beta CT.FU4.c1	-0.237	0.089	-0.411	-0.063
beta CT.FU4.c2	0.935	0.080	0.778	1.092
beta CT.FU5.c1	0.713	0.069	0.578	0.848
beta CT.FU5.c2	-0.538	0.137	-0.807	-0.270
beta CT.FU6.c1	-1.977	0.196	-2.361	-1.594
beta CT.FU6.c2	1.590	0.072	1.449	1.731
beta CT.FU7.c1	-3.200	0.276	-3.741	-2.659
beta CT.FU7.c2	0.341	0.081	0.181	0.500
beta CT.FU8.c1	0.785	0.071	0.645	0.924
beta CT.FU8.c2	0.286	0.105	0.079	0.492
beta CT.FU9.c1	-1.132	0.113	-1.355	-0.910
beta CT.FU9.c2	0.582	0.081	0.424	0.740
beta CT.FU10.c1	-0.388	0.136	-0.654	-0.123
beta CT.FU10.c2	2.564	0.080	2.407	2.721
beta CT.FU11.c1	-1.514	0.140	-1.788	-1.239
beta CT.FU11.c2	0.960	0.075	0.812	1.107
beta CT.FU12.c1	-2.504	0.201	-2.898	-2.110
beta CT.FU12.c2	0.447	0.080	0.289	0.604
beta CT.FU13.c1	-2.672	0.267	-3.196	-2.147
beta CT.FU13.c2	1.544	0.072	1.404	1.685

```
beta CT.FU14.c1    -1.280      0.129   -1.533   -1.026
beta CT.FU14.c2     1.060      0.075    0.914    1.207
beta CT.FU15a.c1   -0.768      0.102   -0.968   -0.569
beta CT.FU15a.c2    0.750      0.080    0.594    0.907
beta CT.FU15b.c1   -1.005      0.110   -1.220   -0.790
beta CT.FU15b.c2    0.682      0.080    0.525    0.838
beta CT.FU15c.c1   -0.568      0.094   -0.752   -0.384
beta CT.FU15c.c2    0.641      0.082    0.480    0.802
beta CT.FU15d.c1   -1.739      0.139   -2.012   -1.467
beta CT.FU15d.c2    0.219      0.085    0.051    0.386
beta CT.FU15e.c1   -2.846      0.213   -3.264   -2.428
beta CT.FU15e.c2   -0.637      0.107   -0.847   -0.427
beta CT.FU15f.c1   -3.327      0.287   -3.889   -2.765
beta CT.FU15f.c2    0.150      0.085   -0.017    0.316
beta CT.FU16.c1    -1.379      0.133   -1.639   -1.118
beta CT.FU16.c2     0.979      0.075    0.831    1.127
```

Die erste Liste gibt die Aufgabenschwierigkeit, die zweite die Aufgabenleichtigkeit an.

Mit

```
> plotICC(res.rm)
```

erhält man die grafische Ausgabe aller Item-Response-Funktionen. Da diese sich im Rasch Modell bis auf die Lage der Itemlokation nicht unterscheiden, also alle den Verlauf der Funktion aus Abbildung 2 aufweisen, ist diese Darstellung informationslos.

Personenparameter-Berechnung des Rasch Modells

Nun zur Bestimmung der Personenparameter. Dazu wird das berechnete Modellobjekt *res.rm* der Methode *person.parameter* übergeben:

```
> personest.rm<-person.parameter(res.rm)
```

Ausgabe jedes einzelnen Personenparameters:

```
> summary(personest.rm)
```

```
Estimation of Ability Parameters

Collapsed log-likelihood: -474.3405
Number of iterations: 19
Number of parameters: 35

ML estimated ability parameters (without spline interpolated values):
              Estimate Std. Err.       2.5 %      97.5 %
theta P1   -0.600197884 0.2564658 -1.10286167 -0.09753409
theta P2   -0.340848567 0.2544781 -0.83961648  0.15791934
theta P3   -0.734039764 0.2613597 -1.24629540 -0.22178413
theta P4   -1.106509651 0.2884724 -1.67190525 -0.54111405
theta P5   -0.666488682 0.2585665 -1.17326970 -0.15970767
theta P6   -0.470015749 0.2542114 -0.96826085  0.02822935
theta P7   -0.534816684 0.2550212 -1.03464903 -0.03498433
theta P8   -0.470015749 0.2542114 -0.96826085  0.02822935
theta P9   -1.383160013 0.3216612 -2.01360435 -0.75271567
theta P10  -1.025520453 0.2808893 -1.57605328 -0.47498763
theta P11  -0.734039764 0.2613597 -1.24629540 -0.22178413
.
theta P909 -0.948437728 0.2745462 -1.48653839 -0.41033707
theta P910 -1.025520453 0.2808893 -1.57605328 -0.47498763
theta P911 -1.192296530 0.2975440 -1.77547197 -0.60912109
theta P912 -1.284016259 0.3084455 -1.88855830 -0.67947422
theta P913 -0.803244056 0.2648999 -1.32243841 -0.28404970
theta P914 -0.874542848 0.2692618 -1.40228620 -0.34679950
theta P915 -0.948437728 0.2745462 -1.48653839 -0.41033707
theta P916 -0.666488682 0.2585665 -1.17326970 -0.15970767
theta P917 -1.192296530 0.2975440 -1.77547197 -0.60912109
theta P918 -1.383160013 0.3216612 -2.01360435 -0.75271567
theta P919 -1.106509651 0.2884724 -1.67190525 -0.54111405
theta P920 -1.750171100 0.3843796 -2.50354134 -0.99680086
theta P921 -1.106509651 0.2884724 -1.67190525 -0.54111405
```

```
theta P922 -1.284016259 0.3084455 -1.88855830 -0.67947422
theta P923 -0.734039764 0.2613597 -1.24629540 -0.22178413
theta P924 -1.383160013 0.3216612 -2.01360435 -0.75271567
theta P925 -1.911048262 0.4192648 -2.73279222 -1.08930431
theta P926 -0.734039764 0.2613597 -1.24629540 -0.22178413
```

Ausgabe der Personenparameter für die Scores:

```
> print(personest.rm)
```

erzeugt eine Ausgabe der Parameter nach Score. Da nur so viele Personenparameter wie Scores existieren, ergeben sich nur so viele Personenparameter wie Scores möglich sind, also von keiner gelösten Aufgabe bis zu allen gelösten Aufgaben.

Grafische Darstellung des Verhältnisses von Personenparameter und Score

Übergabe des Objekts der Personenparameterberechnung *personest.rm* an den *plot*-Befehl:

```
> plot(personest.rm).
```

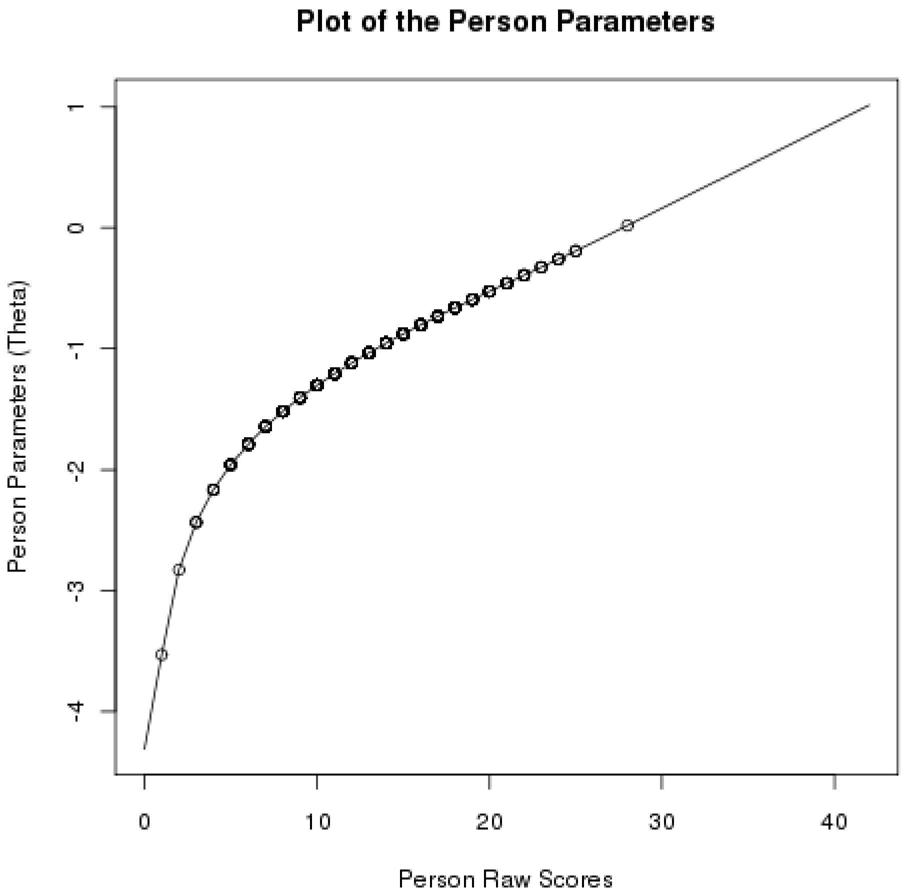


Abbildung 4: Grafische Ausgabe der Personenparameter im Verhältnis zum Score

Grundsätzlich sollte der Verlauf dieser Funktion linear sein. Da der Score einer linearen Hierarchie folgt, sollten auch die dazugehörigen Parameterwerte linear sein. Doch weisen die Parameterschätzer für die Randscores modellbedingt größere Fehler auf als für „mittlere" Scores. Daher ergibt sich die starke Krümmung für niedrige Scores.

Fit-Werte

Infit und Outfit Itemanalyse

Berechnung der Infit und Outfit Werte, Übergabe des Objekts zur Personenparameterbestimmung an die *itemfit*-Methode:

```
> itemfit(personest.rm)

Itemfit Statistics:
             Chisq  df p-value Outfit MSQ Infit MSQ Outfit t Infit t
CT.FU1     872.977 922   0.874      0.946     0.946    -1.80   -1.80
CT.FU2     820.156 922   0.993      0.889     0.912    -3.06   -2.81
CT.FU3     909.147 922   0.612      0.985     0.950    -0.52   -2.31
CT.FU4     922.864 922   0.486      1.000     1.003     0.01    0.18
CT.FU5     931.863 922   0.403      1.010     1.009     0.37    0.36
CT.FU6     969.514 922   0.135      1.050     1.017     2.30    1.14
CT.FU7    1107.488 922   0.000      1.200     1.104     4.52    3.94
CT.FU8     910.275 922   0.602      0.986     0.998    -0.53   -0.07
CT.FU9     967.316 922   0.146      1.048     1.034     1.55    1.55
CT.FU10    780.527 922   1.000      0.846     0.901    -5.20   -4.43
CT.FU11    904.055 922   0.657      0.979     0.979    -0.77   -1.18
CT.FU12    934.962 922   0.376      1.013     1.034     0.36    1.41
CT.FU13    998.070 922   0.041      1.081     1.071     3.59    4.84
CT.FU14    842.995 922   0.970      0.913     0.931    -3.61   -4.14
CT.FU15a   770.567 922   1.000      0.835     0.859    -6.39   -7.35
CT.FU15b   763.033 922   1.000      0.827     0.864    -6.29   -6.79
CT.FU15c   806.985 922   0.997      0.874     0.878    -4.69   -5.96
CT.FU15d  1055.408 922   0.001      1.143     1.069     3.45    2.51
CT.FU15e   959.439 922   0.191      1.039     0.966     0.59   -0.72
CT.FU15f  1042.830 922   0.003      1.130     0.979     2.66   -0.70
CT.FU16    986.021 922   0.070      1.068     1.070     2.62    3.90
```

Die Infit und Outfit Maße geben an, ob ein Item irreguläre Antwortmuster verursacht bzw. ob Personen Antwortmuster besitzen, die weit davon entfernt sind mit dem Rasch Modell konform zu sein. Dies bezeichnet man als lineare Hierarchie der Scores .Der Infit bestimmt, wie wahrscheinlich das Vorkommen eines Antwortmusters ist, der Outfit die mittlere Abweichung eines Antwortmusters.[47]

Die standardisierten Outfit und Infit Maße *Outfit t Infit t* sollen den absoluten Wert von 2 nicht übersteigen. Items, die solche Werte erzeugen, sind mit den Restriktionen des Rasch Modells nicht vereinbar und müssen aus der Testentwicklung entfernt werden – das Modell zwingt die Aufgabenentwicklung über die Modell-Definition in einen Validierungsprozess, der mühsam ist und zudem viele Aufgaben als ungeeignet für das Modell identifiziert.[48]

Andersen LR-Test

Da im Rasch Modell der Vergleich zweier Personen unabhängig von den verwendeten Aufgaben ist und vice versa, so müssten sich in unterschiedlichen Personengruppen, bis auf eine Verschiebung, die gleichen Parameterwerte einstellen. Dies versteht man unter spezifischer Objektivität. Diese Eigenschaft des Rasch Modells dient als Grundlage für den Andersen Likelihood Ratio Test: Es werden zwei Subgruppen gebildet und deren Parameterberechnungen miteinander verglichen. Die Art und Weise der Bildung der Subgruppen kann beeinflusst werden. Als Splittpunkt der Stichprobe zur Berechnung der Parameter in den so entstehenden Subgruppen wurde das arithmetische Mittel der Scores gewählt. Möglich ist aber auch der Median *median* oder ein externes Kriterium wie Alter oder Geschlecht.

Die Nullhypothese geht von der Gültigkeit des Rasch Modells aus, d.h. die Parameterwerte sind in den Subgruppen identisch.

Erzeugung eines LR-Test-Objekts über

```
> lrest.rm<-LRtest(res.rm, splitcr="mean", se=TRUE)
> lrest.rm

Andersen LR-test:
```

[47] Siehe dazu: Glas, Verhelst 1994; Glas 1988; Hardouin 2007, Linacre 2002; Wilson 2005; Planinic 2010.

[48] Siehe dazu: Wilson 2003, 2005; Wuttke 2008, Liu et.al. 2007.

```
LR-value: 172.39
Chi-square df: 41
p-value:  0
```

Für die dargestellten Werte und den Splittpunkt des Mittelwerts müsste die Nullhypothese, dass die Parameter in den Subgruppen identisch sind, verworfen werden, denn der p-Wert ist Null und damit kleiner als irgendein gewähltes Signifikanzniveau - *p-value: 0*. Dies bedeutet, dass mit diesen Items und diesen Aufgaben die Gültigkeit eines Rasch Modells verworfen werden müsste.

Grafische Darstellung des Andersen LR-Tests

Übergabe des LR-Test-Objekts an den plot-Befehl *plotGOF*:

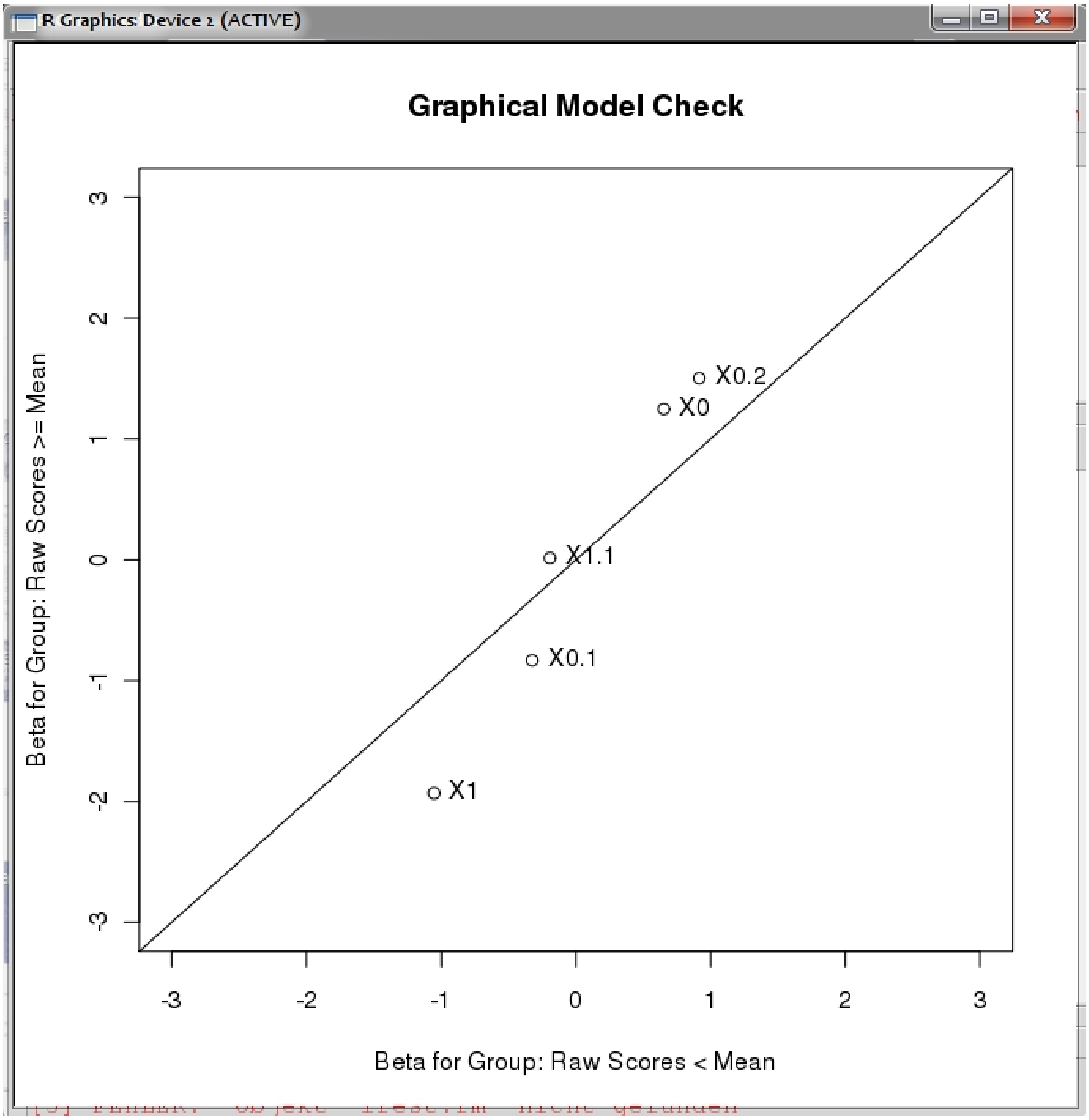


Abbildung 5: Grafische Aufbereitung des GOF-Tests für ein Rasch Modell mit fünf Items.

```
> plotGOF(lrest.rm)
```

<u>Informationsindices</u>

Erzeugung der Informationsindices AIC und BIC: Übergabe des Personenparameterobjekts an die Methode *IC*:

```
> IC<-IC(personest.pcm)
> IC

Information Criteria:
                          value npar      AIC      BIC     cAIC
joint log-lik         -14840.19   76 29832.39 30199.29 30275.29
marginal log-lik      -15665.33   41 31412.66 31610.73 31651.73
conditional log-lik   -12789.43   41 25660.86 25858.92 25899.92
```

Diese Werte stellen eine Entscheidungshilfe bei der Wahl des sehr wahrscheinlich für die Daten gültigen Modells dar: Rasch Modell, PCM oder RSM. Dazu müssten alternativ Berechnungen mit Variationen des Rasch Modells durchgeführt (Hybride Modelle, Mixed Modelle)[49] und die dort erhaltenen Informationsindices mit den hier berechneten Daten verglichen werden.

Es gilt, je kleiner desto besser. Der BIC trägt der vorhandenen Parameterzahl Rechnung, ähnlich dem *adjustierten* R^2 in der multiplen Regressionsrechnung.

[49] Siehe dazu: Rost 2004.

11.2 Anhang 2

Diagnosetext

Knapp 6000 neue Auszubildende in Hagen im Herbst 2009

Die Zahl der neuen Auszubildenden in Hagen ist erfreulicherweise gestiegen. Im Herbst 2009 haben in den Ausbildungsbetrieben in Hagen insgesamt 5959 und damit 14,6 Prozent mehr Jugendliche ihre Ausbildung angetreten als ein Jahr zuvor, wie die Agentur für Arbeit Hagen mitteilte.

Jeder elfte der neuen Auszubildenden hat eine außerbetriebliche Ausbildung begonnen. Jeder 5. der neuen Auszubildenden ist ausländischer Herkunft.

Besonders gefragt waren der Statistik zufolge kaufmännische Ausbildungsberufe, erst danach folgen gewerbliche und soziale Ausbildungsberufe. So lassen sich im Herbst 2009 von den insgesamt 5959 jugendlichen Hagenern 661 im kaufmännischen Bereich ausbilden. Unter den nicht-kaufmännischen Berufen waren die Ausbildung zum Koch, die von 77 Jugendlichen begonnen wurde und die Ausbildung zum Industriemechaniker, die von 76, meist männlichen Jugendlichen, angetreten wurde, am gefragtesten.

Die meisten der neuen Hagener Auszubildenden besaßen einen Realschulabschluss (2043), gefolgt von Bewerbern mit Hauptschulabschluss (1495) und Fachhochschulreife (1216).

Mindestens 48 der Schulabgänger, die sich um einen Ausbildungsplatz bzw. um eine außerbetriebliche Ausbildung bemühten, blieben unversorgt. Das ist fast 1 % aller Hagener Schulabgänger im Herbst 2009. Von den unversorgten Schulabgängern in Hagen waren 30 Jugendliche männlich.

Im Bundesvergleich traten im Herbst 2009 insgesamt fünfhundertvierzigtausend junge Menschen eine Berufsausbildung an; im Berichtsjahr zuvor waren es gleich viele versorgte Bewerber. Die Anzahl unversorgter Bewerber ging im Vergleich zum Vorjahr um mehr als ein Viertel zurück, was knapp 4.000 Jugendlichen entspricht. Bundesweit hat jeder 10. der neuen Auszubildenden eine außerbetriebliche Ausbildung begonnen und jeder vierte der neuen Auszubildenden ist ausländischer Herkunft.

Kapitel 12 Literatur

Abedi, Jamal: „Dimensionality of the NAEP Subscals Scores in Mathematics". ‹http://www.cse.ucla.edu/products/Reports/TECH428.pdf›. In: ‹http://www.cse.ucla.edu/›. 1997, Download am 09.11.2011.

Adams, Ray; Wu, Margaret: PISA 2000 Technical Report. ‹http://www.oecd.org/document/7/0,3343,en_32252351_32236159_33688711_1_1_1_1,00.html›. In: ‹http://www.oecd.org›. 2002, Download am 03.09.2008.

Adams, Ray; Wu, Margaret (Hrsg.): PISA 2000 Technical Report. OECD, Paris 2002.

Andersen, Erling: „Conditional Inference For Multiple-Choice Questionaries". The British Journal of Mathematical & Statistical Psychology. Bd. 26. 1973, S. 31-44.

Andersen, Erling: „A Godness Of Fit Test for The Rasch Model". In: Psychometrika. Bd. 38. 1973, Nr. 1, S. 123-140.

Arnold, Karl-Heinz: „Qualitätskriterien für die standardisierte Messung von Schulleistungen". In: Weinert, Franz E. (Hrsg.): Leistungsmessung in Schulen. Weinheim 2002, S. 117-130.

Baker, Frank B.; Seock-Ho. Kim: Item response theory. Parameter estimation techniques, 2nd edition, New York 2004.

Baumert, Jürgen; Artelt, Cordula; Klieme, Eckhard; Stanat, Petra: „PISA. Programme for International Student Assessment. Zielsetzung, theoretische Konzeption und Entwicklung von Messverfahren“. In: Weinert, Franz E. (Hrsg.): Leistungsmessung in Schulen. Weinheim 2002, S. 285-310.

Bentler, P.M.: „Comparative Fit Indices in Structural Models“. In: Psychological Bulletin. Bd. 107. 1990, S. 238-246.

Bergunde, Manfred: „Förderung der Lesekompetenz in der Mathematik“. In: http://li.hamburg.de/contentblob/3845816/data/download-pdf-auszug-aus-der-li-broschuere-foerderung-der-lesekompetenz-in-der-mathematik.pdf. 2008, Download am 13.01.2015.

Biedebach, W.: „Der Modellversuch Vocational Literacy (VOLI)- Methodische und sprachliche Kompetenzen in der beruflichen Bildung. Konzeption – Erfahrungen – bisherige Ergebnisse.“ In: Efing, CH., Janich, N. (Hrsg.): Förderung der berufsbezogenen Sprachkompetenz. Befunde und Perspektiven. Paderborn 2006, S. 24-26.

Blum, Werner; Neubrand, Michael; Ehmke, Timo et.al.: „Mathematische Kompetenz“. In: PISA-Konsortium Deutschland (Hrsg.): PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland - Ergebnisse des zweiten internationalen Vergleichs. Münster 2003, S. 47-92.

Bortz, Jürgen: Statistik für Sozialwissenschaftler. 5. Auflage. Heidelberg, New York 1999.

Christophersen; Grape: „Die Erfassung latenter Konstrukte mit Hilfe formativer und reflektiver Messmodelle“. In: Albers, Sönke Daniel; Klapper, Udo; Konradt, Achim; Wolf, Walter; Wolf, Joachim (Hrsg.): Methodik der empirischen Forschung. 2. Auflage. Wiesbaden 2007, S. 103-118.

De Boeck, Paul; Leuven, K.U.: „Random Item IRT Models“ In: Psychometrika. Bd. 73. 2008, Nr. 4, S. 533-559.

Efing, Ch., Janich, N. (Hrsg.): Förderung der berufsbezogenen Sprachkompetenz. Befunde und Perspektiven. Paderborn 2006.

Efing, Ch. (Autor); Institut für Qualitätsentwicklung Hessen (Hrsg.): Baukasten Lesediagnose. Wiesbaden 2006.

Efing, Ch.: „Viele sind nicht in der Lage, diese schwarzen Symbole da lebendig zu machen. – Befunde empirischer Erhebungen zur Sprachkompetenz hessischer Berufsschüler“. In: Efing, Ch., Janich, N. (Hrsg.): Förderung der berufsbezogenen Sprachkompetenz. Befunde und Perspektiven. Paderborn 2006.

Fischer, Gerhard H.: „The Linear Logistic Test. Model as an Instrument in Educational Research“ In: Acta Psychologica. Bd. 37. 1973, S. 359-374.

Fischer, Gerhard: Einführung in die Theorie psychologischer Tests. Bern 1974.

Glas, Cees, Verhelst, Norman D.: „Testing the Rasch Model“. In: Fischer, Gerhard H.; Molenaar, Ivo W. (Hrsg.): Rasch Models. Heidelberg, New York 1995, S. 69-96.

Glas, Cees A. W.: „The Derivation of some Tests for the Rasch Model from the Multinomial Distribution“. In: Psychometrika. Bd. 53 1988, Nr. 4, S. 525-546.

Grothmann, René: Euler Math Toolbox. ‹http://eumat.sourceforge.net/›. In: ‹http://eumat.sourceforge.net/german/index.html›. 2010, Download am 24.1.2010.

Günther, K.; Laxczkowiak, J.; Niederhaus, C.; Wittwer, F.: Sprachförderung im Fachunterricht an beruflichen Schulen. Berlin 2013.

Hair, Joseph F.: Multivariate data analysis. 7. Auflage. Pearson, Upper Saddle River 2010.

Hardouin, Jean-Benoit: „Rasch Analysis: Estimation and tests with raschtest“. In: The Stata Journal, Bd. 7 2007, Nr. 1, S. 22-44.

Hattie, John: „Methodology Review: Assessing Unidimensionality of Tests and Items“. In: Applied Psychological Measurement. Bd. 9 1985, S. 139-164.

Jordan, R.: Entwicklung und Validierung eines Testverfahrens zur Ermittlung der Lesekompetenz und des mathematischen Textverständnisses mit empirischer Untersuchung an allgemeinbildenden und berufsbildenden Schulen. Münster 2011.

Kubinger, K.D.: „Objektive psychologisch-diagnostische Verfahren“. In: Rammsayer, T., Weber, H. (Hrsg.): Handbuch der Persönlichkeitspsychologie und Differentiellen Psychologie. Göttingen 2005, S. 158–165.

Kultusministerkonferenz: Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss. Beschluss der Kultusministerkonferenz vom 4.12.2003.

Linacre, John M.: „What do Infit and Outfit, Mean-Square and Standardized mean?“, Rasch Measurement, Bd. 16 2002, Nr. 2, S. 878.

Liu, Jinghua; Harris, Deborah J.; Schmidt, Amy et.al.: „Statistical Procedures Used in College Admissions Testing“. In: Rao, Calyampudi Radhakrishna; Krishnaiah, Paruchuri Rama (Hrsg.): Handbook of Statistics. Bd 26. Elsevier, Amsterdam et.al. 2007, S. 1057-1091.

Lord, Frederic M.; Novick, Melvin R.: Statistical theories of mental test scores. Addison-Wesley, Reading et.al. 1968.

Lord, Frederic M.: Applications of item response theory to practical testing problems. Erlbaum, Hillsdale 1980.

Mai, Jochen. „Das zweite Internet“. In: WirtschaftsWoche Nr. 46. 15.11.2010, S. 84-86.

Mair, Patrick; Hatzinger, Reinhold: „CML based estimation of extended Rasch models with the eRm package in R“. In: Psychology Science. Bd. 49. 2007, Nr. 1, S. 26-43.

Mair, Patrick; Hatzinger; Reinhold; Maier, Marco: Extended Rasch Modeling. Package ‘eRm’. ‹http://r-forge.r-project.org/projects/erm/›.

In: ‹http://cran.r-project.org/web/packages/eRm/index.html›. 2011, Download am 04.12.2011.

OECD: The PISA 2003 Assessment Framework - Mathematics, Reading, Science and Problem Solving Knowledge and Skills.

‹http://www.oecd.org/edu/preschoolandschool/programmeforinternationalstudentassessmentpisa/33694881.pdf›. In: ‹http://www.oecd.org/pisa/pisaproducts/pisa2003/›. 2003, Download am 30.01.2013.

OECD: Assessing Scientific, Reading and Mathematical Literacy. ‹http://www.oecd.org/pisa/pisaproducts/pisa2006/37464175.pdf›.

In: ‹http://www.oecd.org/pisa/pisaproducts/›. 2006, Download am 30.01.2013.

OECD: PISA 2003 Data Analysis Manual. SPSS User. 2nd edition. ‹http://browse.oecdbookshop.org/oecd/pdfs/free/9809031e.pdf›.

In: ‹http://www.oecdbookshop.org/›. 2009, Download am 03.09.2008.

OECD: PISA 2006 - Beispielaufgaben aus dem Naturwissenschaftstest. ‹http://pisa.ipn.uni-kiel.de/PISA06_Science_Beispielaufgaben_Loesungen.pdf›. In: IPN Kiel. ‹http://pisa.ipn.uni-kiel.de› 2006, Download am 18.05.2012.

OECD: PISA 2003 Technical Report,

http://www.oecd.org/edu/school/programmeforinternationalstudentassessmentpisa/35188570.pdf›. In: ‹http://www.oecd.org/pisa›, ‹http://www.oecd.org/pisa/pisaproducts/pisa2003›. 2003, Download am 19.06.2013.

Planinic, Maja; Ivanjek, Lana; Susac, Ana: „Rasch model based analysis of the Force Concept Inventory“. In: Physical Review Special Topics - Physics Education Research. Bd. 6. 2010, S. 010103-1-11.

Poinstingl, Herbert; Mair, Patrick; Hatzinger, Reinhold: Manual zum Softwarepackage eRm (extended Rasch modeling). ‹http://erm.r-forge.r-project.org/eRm_manual.pdf›. In: http://erm.r-forge.r-project.org, ‹http://erm.r-forge.r-project.org/›. 2007, 6.8.2011.

Rasch, Georg: Probabilistic models for some intelligence and attainment tests with a forword and afterword by Benjamin D. Wright. Copenhagen 1980.

Rasch, Georg: „On general Laws and the Meaning of Measurement in Psychology“. In: Neyman, Jerzy (Hrsg.): Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. Berkeley 1961, Bd. IV, S. 321-333.

Rasch, Georg: „An Item Analysis which takes individual differences into account“. In: The British Journal of Mathematical & Statistical Psychology. Bd. 19. 1966, Nr. 1, S. 49-57.

Reischmann, Jost: Kompetenz lehren? Kompetenz- und Performanz-Orientierung in der Andragogik zwischen Didaktik und Organisationsentwicklung. http://www.uni-bamberg.de/andragogik/interne_links/publikationen.htm. 2004, Download am 29.06.2014.

Ritter, Joachim: Historisches Wörterbuch der Philosophie. Basel 2005.

Rizopoulos, Dimitris: „ltm: An R Package for Latent Variable Modeling and Item Response Theory Analyses“. In: Journal of Statistical Software. Bd. 17. 2006, Nr. 5, S. 1-26.

Rost, Jürgen: Lehrbuch Testtheorie – Testkonstruktion. 2. Ausgabe. Bern 2004.

Sijtsma, Klaas; Verweij, Anton C.: „Knowledge of Solution Strategies and IRT Modeling of Items for Transitive Reasoning“. In: Applied Psychological Measurement. Bd. 23. 1999, Nr. 1, S. 55-68.

Skrondal, Anders; Rabe-Hesketh, Sophia: „Multilevel logistic Regression for polytomous Data and Rankings“. In: Psychometrika, Bd. 68. 2003, Nr. 2, S. 267-287.

Stevens, S. S.: „On the Theory of Scales of Measurement“. In: Science. Bd. 7. 1946, Nr. 103, S. 677-680.

Sumfleth, Elke; Schüttler, Susanne: „Linguistische Textverständlichkeitskriterien - Helfen sie bei der Darstellung chemischer Inhalte?“. In: Zeitschrift für Didaktik der Naturwissenschaften. Bd. 1, 1995, S. 55-72.

Swaminathan, Hariharan: „Parameter Estimation in Item Response Models“. In: Hambleton, Ronald (Hrsg.): Applications of Item Response Theory. Educational Research Institut of British Columbia, Vancouver 1983, S. 24-56.

Weinert, Franz E.: „Vergleichende Leistungsmessung in Schulen – Eine umstrittene Selbstverständlichkeit“. In: Weinert, Franz E. (Hrsg.): Leistungsmessungen in Schulen. Weinheim, Basel 2001, S. 27 ff.

Wilson, Mark: On Choosing a Model for Measuring. ‹http://www.dgps.de/fachgruppen/methoden/mpr-online/issue21/mpr122_8.pdf›. In: Methods of Psychological Research Online 2003, Vol. 8, No. 3, pp 1-22. ‹http://www.dgps.de/fachgruppen/methoden/mpr-online/›. 2003, Download am 09.08.2008.

Wilson, Mark: Constructing Measures: An Item Response Modelling Approach. New Jersey, 2005.

Winn, Bill: „Charts, Graphs, and Diagrams in Educational Materials“. In: Willows, Dale M.; Harvey, Houghton A. (Hrsg.): The Psychology of Illustration. Berlin, Heidelberg 1987, S. 152-198.

Wirtz, Markus; Nachtigall, Christof: Wahrscheinlichkeitsrechnung und Inferenzstatistik. Bd. 2. 3. Auflage. Weinheim 2004.

Wu, M.L.; Adams, R..J.; Wilson, M.R.: ConQuest: Generalized item response modelling software. Australian Council for Educational Research, Melbourne 1998.

Wuttke, Joachim: „Uncertainties and Bias in PISA“. In: Hopmann, Stefan; Brinek, Gertrude; Retzl, Martin (Hrsg.): PISA zufolge PISA - PISA According to PISA. Wien 2008, S. 241-264.

Wuttke, Joachim: „Die Insignifikanz signifikanter Unterschiede“. In: Jahnke, T.; Meyerhöfer, W.: PISA & Co – Kritik eines Programms. Hildesheim 2007.