universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Zur Zulässigkeit der Dichotomisierung mehrkategorieller und kontinuierlicher Daten als Grundlage einer Modellierung nach der Item-Response-Theorie“

verfasst von / submitted by

Isabell Romana Baldauf, BSc BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2015 / Vienna 2015

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 840
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Psychologie
Betreut von / Supervisor:	Mag. Dr. Christine Hohensinn

Danksagung

An dieser Stelle möchte ich all jenen danken, die durch die fachliche und persönliche Unterstützung zum Gelingen dieser Arbeit beigetragen haben.

Ganz besonders bedanke ich mich bei Frau Mag. Dr. Christine Hohensinn für die Betreuung der Masterarbeit, für die wertvollen Ratschläge und für die Idee und Inspiration zu diesem Thema.

Mein Dank gilt auch Univ.-Prof. Mag. Dr. Klaus Kubinger für seine konstruktive Kritik und seine Anregungen im Zuge dieser Arbeit. Ich möchte mich an dieser Stelle auch bei ihm für die großartige Lehre, die er am Fachbereich psychologische Diagnostik an der Universität Wien aufgebaut hat, bedanken.

Bedanken möchte ich mich auch bei Mag. Jan Steinfeld für die Unterstützung bei computer- und programmiertechnischen Fragen.

Ein herzliches Dankeschön gilt meiner Familie für ihre Unterstützung, im Besonderen meinen Eltern, die mir immer das Gefühl geben, dass ich alles, was ich wirklich will, auch erreichen kann. Schließlich möchte ich meinem Partner Roland für seine Geduld, für die emotionale Unterstützung und für anregende Diskussionen danken.

Abstract - deutsche Version

Es gibt Arbeiten zur Zulässigkeit der Dichotomisierung mehrkategorieller und kontinuierlicher Daten als Grundlage einer Modellierung nach der Item-Response-Theorie aus dem Jahre 1980 von Jansen und Roskam (Jansen & Roskam, 1986; Roskam & Jansen, 1989; Roskam, 1995), die mathematisch zeigen, dass es unzulässig ist, mehrkategorielle Daten, die zum Beispiel für ein Partial-Credit-Modell passen, zu dichotomisieren und dann ein Rasch-Modell zu berechnen. Trotzdem wird dies in der Testkonstruktionspraxis häufig gemacht oder es werden Kategorien, die selten genutzt werden, zusammengelegt. Die vorliegende Arbeit beschäftigt sich mit den empirischen Auswirkungen einer nachträglichen Dichotomisierung auf die Parameterschätzung und die Modellgeltung. Im Zuge dieser Simulationsstudie wurden Daten gemäß drei Modellen der Item-Response-Theorie generiert, an verschiedenen Stellen dichotomisiert oder mehrkategoriell unterteilt und anschließend verrechnet. Für die Modellierung mehrkategorieller und kontinuierlicher Daten kamen das Partial-Credit-Modell, das Graded-Response-Modell und das Modell für kontinuierliche Antwortskalen nach Müller zum Einsatz. Die Modellgeltung der Modelle der Rasch-Familie wurde mit dem Likelihood-Ratio-Test und den Fit-Indizes geprüft. Die Überprüfung der Modellkonformität des Graded-Response-Modells erfolgte mittels Pearsons χ^2-Test, der Likelihood-Ratio-Statistik G^2 und der M_2^* −Statistik. Es zeigte sich, dass ein nachträgliches Dichotomisieren mehrkategorieller Daten der Rasch-Familie Auswirkungen auf die Parameterschätzung und die Modellgeltung hat. Diese scheinen geringer bei dem Modell für kontinuierliche Antwortskalen nach Müller, sind aber nicht vertretbar, betrachtet man die Verzerrung der Personenparameterschätzungen. Das Graded-Response-Modell ist wie erwartet weniger „empfindlich“ gegenüber einer nachträglichen Dichotomisierung, dennoch sollte auch hier auf ein nachträgliches Zusammenlegen von Antwortmöglichkeiten verzichtet werden, da es zu Verzerrungen bei der Personenparameterschätzung kommen kann.

Schlüsselwörter: Rasch-Modell, Partial-Credit-Modell, Graded-Response-Modell, Modell für kontinuierliche Antwortskalen nach Müller, Dichotomisierbarkeit, Zusammenlegen von Antwortkategorien, Likelihood-Ratio-Test nach Andersen, Fit-Indizes, Pearsons χ^2-Test, Likelihood-Ratio-Statistik G^2, M_2^* −Statistik von Cai & Hansen

Abstract - English version

This paper discusses the compatibility of polytomous and continuous Item Response Theory models with dichotomization of the response continuum. Jansen and Roskam (Jansen & Roskam, 1986; Roskam & Jansen, 1989; Roskam, 1995) have proven mathematically that polytomous Rasch models, for example the Partial Credit Model, are not compatible with dichotomization after graded response data has been collected. Nevertheless, it is common in psychological measurement to collect data in the form of graded responses and then to combine adjacent categories which are rarely used by participants. The present paper deals with the empirical effects on parameter estimation and model validity when adjacent categories are combined. In the course of a simulation study, artificial data according to the Partial Credit Model, the Graded Response Model and the Continuous Rating Scale Model were generated and dichotomized or polytomized at different positions. To test the goodness-of-fit of the Rasch models, Andersen's likelihood ratio test and fit indices were used. To check the assumptions of the Graded Response Model, Pearson's χ^2 statistic, the likelihood ratio statistic G^2 and the M_2^* statistic were calculated. The dichotomization of data according to polytomous Rasch models has an impact on parameter estimation and model validity. The effects appear to be less pronounced for the Continuous Rating Scale Model, but are not unreasonable considering the distortion of the person parameter estimation. The Graded Response Model is less sensitive to a dichotomization of the response continuum. But even here, the person parameter estimations are not invariant to dichotomization.

Keywords: Rasch model, Partial Credit Model, Graded Response Model, Continuous Rating Scale Model, dichotomization, joining categories, Andersen's likelihood ratio test, fit indices, Pearson's χ^2 statistic, likelihood ratio statistic G^2, M_2^* statistic

Inhalt

1 Einleitung

Eine zentrale Aufgabe der Angewandten Psychologie ist die Entwicklung psychologisch-diagnostischer Verfahren. Diese dienen als Messinstrumente latenter Merkmale, wie zum Beispiel des räumlichen Vorstellungsvermögens von Schülern oder Persönlichkeitseigenschaften eines Probanden.

Zur Beschreibung des Zusammenhangs des tatsächlichen Ausprägungsgrades des Merkmals und des gemessenen Ausprägungsgrades werden statistische Modelle verwendet. Eines dieser statistischen Modelle, welches zur Familie der Item Response Theory gehört, ist das dichotome logistische Modell von Rasch, welches im Rest der Arbeit als Rasch-Modell bezeichnet wird. Dieses kann herangezogen werden, wenn die Aufgaben des Messinstrumentes nicht aufbauend sind und dieselbe zugrundeliegende Dimension messen. Für die Beantwortung jeder Aufgabe dürfen allerdings nur zwei Kategorien (z. B. richtig/falsch oder stimme zu/stimme nicht zu) zur Verfügung stehen.

Für die Modellierung mehrkategorieller (z. B. kann eine Person bei jeder Aufgabe zwischen 0 und 3 Punkte erzielen) und kontinuierlicher (z. B. Analogskala) Daten wurden Erweiterungen des Rasch-Modells entwickelt. Diese implizieren allerdings weitere Voraussetzungen, welche in der probabilistischen Testtheorie überprüfbar sind. Eine dieser Voraussetzungen, um die resultierenden Parameter valide schätzen zu können, ist eine gewisse Antwortdichte. Das heißt, es kommt zu Problemen bei der Parameterschätzung (ungeordnete Schwellenparameter), wenn einzelne Antwortmöglichkeiten eines Testmodells in den Daten nur selten auftreten. Grund dafür können unklar formulierte Antwortmöglichkeiten oder ein ungeeignetes Bewertungsschema sein (Strobl, 2012).

Van Wyke und Andrich (2006) beschreiben diesbezüglich folgendes Beispiel: Schulkindern wurde die Aufgabe gestellt, mehrere auf einem Blatt abgebildete kleine Blumen mithilfe eines Lineals zu einem Rechteck zu verbinden. Es konnten 0, 1 oder 2 Punkte erzielt werden. 2 Punkte wurden vergeben, wenn das Rechteck richtig eingezeichnet wurde, 1 Punkt, sobald die richtigen Blumen gewählt wurden, das Rechteck aber „wackelig“ oder schief gezeichnet wurde, und 0 Punkte, wenn die Aufgabe nicht erfüllt wurde. Es zeigte sich, dass das Konzept des Rechteckes entweder

verstanden wurde – dann machten die Kinder auch keine Fehler mehr – oder nicht. Die mittlere Kategorie kam dadurch kaum vor.

Befindet man sich noch in der Testentwicklung[1], kann diese Problematik umgangen werden, indem man die Aufgaben vorab im Zuge eines Pretests vorgibt. Wurden die Daten allerdings schon erhoben, werden die betreffenden Antwortkategorien in der Praxis häufig zusammengelegt, um die Problematik der ungeordneten Schwellenparameter zu umgehen. Bei dem oben angeführten Beispiel könnte man statt der drei Kategorien (0, 1 oder 2 Punkte) die oberen beiden zusammenlegen, dann hätte man nur noch zwei Kategorien (das Kind konnte die Aufgabe erfüllen oder nicht).

Für einen Datensatz, für den Modellgültigkeit bezüglich eines Modells für mehrkategorielle Daten (bspw. Partial-Credit-Modell) angenommen werden kann, gilt nach einer Dichotomisierung allerdings nicht automatisch das Rasch-Modell. Jansen und Roskam (Jansen & Roskam, 1986; Roskam & Jansen, 1989; Roskam, 1995) haben mathematisch gezeigt, dass Modelle der Rasch-Familie nicht invariant gegenüber einer nachträglichen Dichotomisierung sind. Die vorliegende Arbeit beschäftigt sich mit den empirischen Auswirkungen einer nachträglichen Dichotomisierung auf die Parameterschätzung und die Modellgeltung.

Im theoretischen Teil werden das Rasch-Modell, das Partial-Credit-Modell, das Graded-Response-Modell und das kontinuierliche Ratingskalenmodell von Müller sowie deren Modelltests beschrieben. In diesem Zusammenhang wird auch auf die Frage der Zulässigkeit der Dichotomisierung eingegangen. Der zweite Teil der Arbeit beschreibt die durchgeführte Simulationsstudie. Im Anhang ist der vollständige Programmcode zur Simulation angeführt. Abschließend werden die Ergebnisse dargestellt und diskutiert.

[1] Da im Zusammenhang mit Modellen der Rasch-Familie oft von Leistungstests gesprochen wird, wird im Rest der Arbeit der Begriff „Test“ statt „psychologisch-diagnostisches Verfahren“ verwendet.

2 Theoretischer Teil

In diesem Teil der Arbeit werden die wichtigsten Grundlagen zusammengefasst, die für das Verständnis der durchgeführten Simulationsstudie relevant sind. Es werden das Rasch-Modell, das Partial-Credit-Modell, das Graded-Response-Modell und das kontinuierliche Ratingskalenmodell von Müller beschrieben. Des Weiteren wird auf die Dichotomisierbarkeit der einzelnen Modelle eingegangen. Schließlich werden noch die Modelltests beschrieben, die in der Simulationsstudie zur Anwendung kamen.

2.1 Item-Response-Theorie

Die Item-Response-Theorie (IRT) ist ein Teilgebiet der psychologischen Testtheorie (Rost, 2004). Sie umfasst eine Familie probabilistischer Modelle, welche das Antwortverhalten (Response) von Probanden modellieren (Rost, 2004). Das Besondere an den probabilistischen Modellen ist die Trennung der latenten Dimension und der beobachtbaren Variable (z. B. das Testverhalten einer Person), welche als Indikator oder Symptom der latenten Dimension aufgefasst wird (Fischer, 1974). Das heißt, dass das Ergebnis, welches eine Person bei einem Test erzielt, nicht deterministisch von ihren Fähigkeiten abhängt, sondern auch von einer Zufallskomponente mitbeeinflusst wird (Fischer, 1974). Daher wird die Wahrscheinlichkeit für das Auftreten einer bestimmten Reaktion in Abhängigkeit der Fähigkeit einer Person und der Schwierigkeit einer Aufgabe beschrieben.

2.2 Das Rasch-Modell

Das bekannteste Modell der IRT ist das dichotome logistische Modell, welches 1960 von dem dänischen Statistiker Georg Rasch entwickelt wurde. Bei diesem wird die Wahrscheinlichkeit modelliert, dass eine Person v mit bestimmter Fähigkeit θ_v eine Aufgabe i mit Schwierigkeit β_i löst. Die Wahrscheinlichkeit, eine Aufgabe zu lösen oder nicht zu lösen, wird als logistische Funktion modelliert, welche von der Differenz der Fähigkeit der Person und der Schwierigkeit der Aufgabe abhängt.

$$P(X_{vi} = x_{vi}|\theta_v, \beta_i) = \frac{e^{x_{vi}(\theta_v - \beta_i)}}{1 + e^{\theta_v - \beta_i}} \quad (1)$$

X_{vi} stellt hierbei das tatsächliche Antwortverhalten einer Person *v* bei der Aufgabe *i* dar. Die Variable kann nur die Ausprägungen x_{vi}, 0 für nicht gelöst und 1 für gelöst annehmen. Die Wahrscheinlichkeit, eine Aufgabe zu lösen beziehungsweise nicht zu lösen, kann folgendermaßen geschrieben werden:

$$P(X_{vi} = 1|\theta_v, \beta_i) = \frac{e^{(\theta_v - \beta_i)}}{1 + e^{\theta_v - \beta_i}} \quad ; \quad P(X_{vi} = 0|\theta_v, \beta_i) = \frac{1}{1 + e^{\theta_v - \beta_i}} \quad (2)$$

Wie oben erwähnt, beschreibt die logistische Funktion die Wahrscheinlichkeit, eine Aufgabe zu lösen, in Abhängigkeit der Personenfähigkeit und der Aufgabenschwierigkeit (Rost, 2004). Die graphische Darstellung dieser Funktion nennt man aufgabencharakteristische Kurve (ICC, englisch „Item Characteristic Curve“). In der folgenden Graphik sind die ICCs für vier Aufgaben dargestellt.

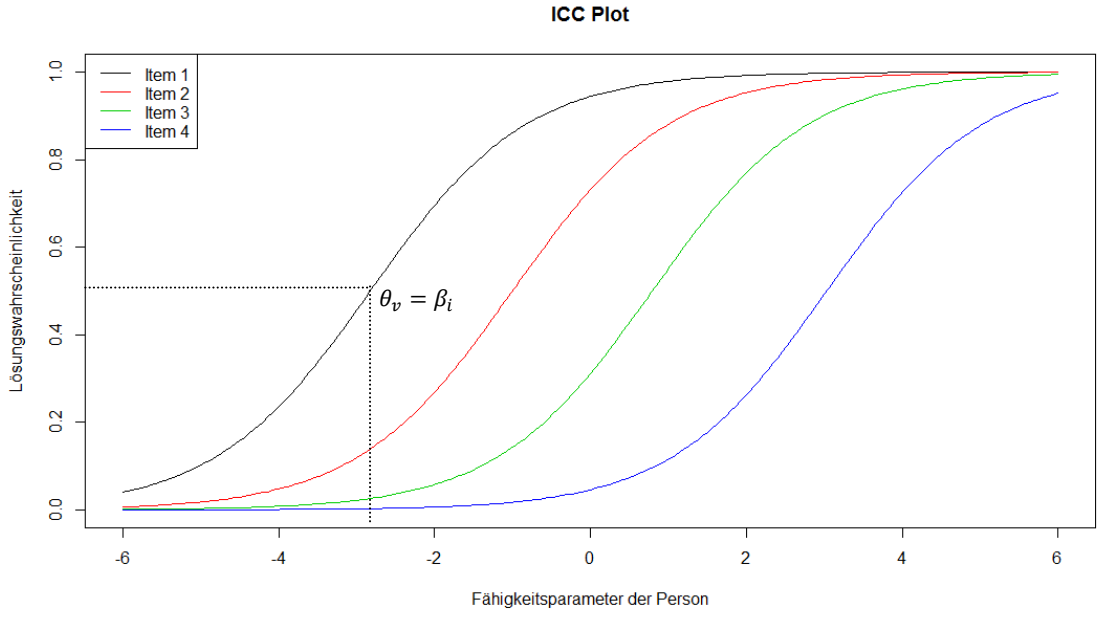


Abbildung 1: Aufgabencharakteristische Kurve (ICC) für vier Aufgaben

Der eingezeichnete Punkt wird als Wendepunkt bezeichnet. An dieser Stelle beträgt die Wahrscheinlichkeit, eine Aufgabe zu lösen, genau 50 %. Das heißt, eine Person ist genauso fähig, wie die Aufgabe schwer ist ($\theta_v = \beta_i$).

Es gibt allerdings Annahmen für das Rasch-Modell, welche erfüllt sein müssen, damit das Modell gilt. Möchte man das Rasch-Modell auf Daten eines Tests anwenden, muss man diese überprüfen.

Trennschärfe

Eine dieser Annahmen ist, dass alle Aufgaben die gleiche Trennschärfe haben müssen. Je höher die Trennschärfe einer Aufgabe ist, desto genauer kann diese zwischen Personen mit unterschiedlichen Fähigkeiten differenzieren. Möchte man das Rasch-Modell verwenden, müssen alle Aufgaben die gleiche Trennschärfe haben (Rost, 2004).

Lokale stochastische Unabhängigkeit

Lokale stochastische Unabhängigkeit bedeutet, dass die Wahrscheinlichkeit, dass eine Person eine Aufgabe löst, unabhängig davon ist, ob sie eine andere Aufgabe gelöst hat. Verletzt wäre diese Annahme, wenn zwei Aufgaben aufeinander aufbauen. „Lokal“ in diesem Zusammenhang bedeutet, dass diese Unabhängigkeit gelten muss, solange man Personen mit derselben Fähigkeit betrachtet (Fischer, 1974).

Spezifische Objektivität

Die spezifische Objektivität im Rasch-Modell bezieht sich auf die Vergleichbarkeit der Fähigkeit von Personen und der Vergleichbarkeit der Schwierigkeit von Aufgaben (Strobl, 2012).

Zwischen zwei Personen ist ein eindeutiger Vergleich möglich, unabhängig davon, welche und wie viele Aufgaben für diesen Vergleich verwendet wurden. Ebenso ist ein eindeutiger Vergleich zweier Aufgaben möglich, unabhängig davon, welche Personen man testet. Des Weiteren muss die Auswahl der Personen und Aufgaben aus der Population nicht mehr zufällig sein, um vergleichbar zu sein. Diese Eigenschaft wird als „Stichprobenunabhängigkeit“ bezeichnet (Fischer, 1974).

Eindimensionalität

Dem Rasch-Modell liegt außerdem die zentrale Eigenschaft zugrunde, dass es eindimensional messen muss. Das bedeutet einerseits, dass alle Aufgaben dieselbe Dimension (Fähigkeit) messen, und andererseits, dass die Beantwortung dieser Aufgaben durch eine Person nur von

ebendieser zugrundeliegenden Dimension abhängen darf. Ein Test zur Mathematik-Kompetenz sollte nur diese messen und nicht zusätzlich sprachliche Fähigkeiten (Strobl, 2012).

Eine weitere Eigenschaft des Rasch-Modells, welche aus der Modellgleichung folgt, sind die suffizienten Statistiken. Als suffiziente oder erschöpfende Statistiken werden Statistiken (Kenngröße, die aus den Daten berechnet wird) bezeichnet, welche die gesamte relevante Information über den unbekannten Parameter enthalten (Fischer, 1974). Im Fall des Rasch-Modells sind die suffizienten Statistiken die Zeilen- und Spalten-Randsummen. Für jede Person enthält die Zeilen-Randsumme die gesamte Information über die Fähigkeit der Person, und die Spalten-Randsumme enthält für jede Aufgabe die gesamte Information über die Schwierigkeit dieser Aufgabe (Rost, 2004).

Parameterschätzung

Es gibt verschiedene Ansätze, um die Fähigkeiten der Personen (Personenparameter) und die Aufgabenschwierigkeiten (Aufgabenparameter) zu schätzen. Im Wesentlichen unterscheiden sie sich darin, ob die Aufgaben- und Personenparameter gleichzeitig (Joint-Maximum-Likelihood-Methode) oder nacheinander (Conditional-Maximum-Likelihood-Methode und Marginal-Maximum-Likelihood-Methode) geschätzt werden. Beiden Methoden liegt die Likelihoodfunktion des Rasch-Modells zugrunde. In der folgenden Arbeit wurde die Conditional Maximum Likelihood (CML) verwendet.

Die Grundidee ist, anhand der beobachteten Daten die Likelihood für bestimmte Personen- und Aufgabenparameter zu maximieren. Die Likelihoodfunktion des Rasch-Modells kann wie folgt geschrieben werden:

$$\boldsymbol{L(\theta_1, \dots, \theta_n; \beta_1, \dots \beta_k | x_{11}, \dots, x_{nk}) = \prod_{v=1}^{n} \prod_{i=1}^{k} \frac{exp(x_{vi}(\theta_v - \beta_i))}{1 + exp(\theta_v - \beta_i)) \rightarrow max}} \quad \textbf{(3)}$$

Mit dieser Formel **(3)** kann man die Likelihood für beliebige Parameter $(\theta_1, \dots, \theta_n;\ \beta_1, \dots, \beta_k)$ gegeben der Daten $x_{11}, \dots, x_{nk}$ berechnen. Ein anschauliches Beispiel zeigen Koller, Alexandrowicz und Hatzinger (2012, S. 30-35).

Die CML-Methode oder auch bedingte Maximum-Likelihood-Methode basiert auf einem zweistufigen Vorgehen. Im ersten Schritt werden die Personenparameter

herauskonditioniert und nur die Aufgabenparameter geschätzt (Strobl, 2012). Im zweiten Schritt werden die Personenparameter geschätzt. Es wird die Wahrscheinlichkeit (Likelihood) der Daten unter der Bedingung der Randsummen betrachtet. Die suffiziente Statistik (Randsummen) kann genutzt werden, um die Likelihoodfunktion so umzuschreiben, dass sie nicht mehr von den Personenparametern abhängt (Strobl, 2012). Die resultierende Likelihoodfunktion kann man allerdings nicht einfach nach den Aufgabenparametern auflösen, daher werden die Maximum-Likelihood-Schätzer mithilfe iterativer Verfahren (durch Computerprogramme) berechnet (Strobl, 2012).

2.3 Modelle für mehrstufige und kontinuierliche Antwortkategorien

Im folgenden Kapitel wird auf die Modelle für mehrstufige und kontinuierliche Antwortkategorien eingegangen, die in der Simulationsstudie betrachtet wurden. Neben den zwei Modellen für mehrstufige Antwortkategorien, dem Partial-Credit-Modell und dem Graded-Response-Modell, wird auch ein Modell für kontinuierliche Antwortkategorien betrachtet. Ein zentraler Unterschied zwischen den Modellen liegt in der Modellierung der Kategoriewahrscheinlichkeit. Die Art der Modellierung beeinflusst auch die Zulässigkeit einer nachträglichen Dichotomisierung, welche im Zuge der Beschreibung der Modelle herausgearbeitet wird.

Das Partial-Credit-Modell (PCM) gehört zur Rasch-Familie, mit all seinen Eigenschaften (Eindimensionalität, suffiziente Statistik, lokale stochastische Unabhängigkeit und spezifische Objektivität) (Strobl, 2012).

Das Graded-Response-Modell (GRM) hingegen gehört nicht zur Rasch-Familie. Aufgrund der Herleitung der Modellgleichung erfüllt das GRM einige in Kapitel 2.1 beschriebene Eigenschaften, wie die lokale stochastische Unabhängigkeit und die Eindimensionalität. Allerdings sind auch einige nicht gegeben, wie die spezifische Objektivität und die Suffizienz. Es wird auch die Forderung aufgegeben, dass alle Aufgaben dieselbe Trennschärfe besitzen müssen (Strobl, 2012).

Das kontinuierliche Ratingskalenmodell von Müller (CRSM, englisch „Continuous Rating Scale Model“) gehört ebenfalls zur Rasch-Familie. Es ist ein Spezialfall des

Modells mit mehrstufigen Antwortkategorien, bei dem die Anzahl der Antwortmöglichkeiten gegen unendlich geht (Müller, 1999).

Zusammenfassend lässt sich sagen, dass Modelle der Rasch-Familie eindeutigere Messergebnisse liefern und statistisch wertvolle Eigenschaften besitzen, aber auch einschränkender in ihren Annahmen sind.

Das Partial-Credit-Modell

Das Partial-Credit-Modell wurde von Masters (1982) entwickelt. Es wird zur Modellierung mehrstufiger Antwortformate herangezogen. Im Rasch-Modell wird nur eine Modellgleichung **(1)** betrachtet. Da es im Rasch-Modell nur zwei Antwortalternativen gibt, wird die Wahrscheinlichkeit einer positiven Antwort modelliert, wobei die Wahrscheinlichkeit einer negativen Antwort ihr Komplement und damit vernachlässigbar ist (Rost, 2004).

Im Partial-Credit-Modell, beziehungsweise in allen Modellen für mehrstufige Antwortformate, wird für jede Antwortmöglichkeit eine Modellgleichung formuliert (Strobl, 2012). Man spricht in diesem Zusammenhang von Kategoriewahrscheinlichkeit und Schwellenwahrscheinlichkeit. Im Rasch-Modell gibt es zwei Kategorie-wahrscheinlichkeiten: die Wahrscheinlichkeit, die Aufgabe zu lösen, und die Wahrscheinlichkeit, die Aufgabe nicht zu lösen. In Abb. 1 wurde die Wahrscheinlichkeit, die Aufgabe zu lösen, betrachtet; die Wahrscheinlichkeit, die Aufgabe nicht zu lösen, ist die Gegenwahrscheinlichkeit und damit redundant (Rost, 2004). Im folgenden Plot sind beide Kategoriewahrscheinlichkeiten aufgetragen.

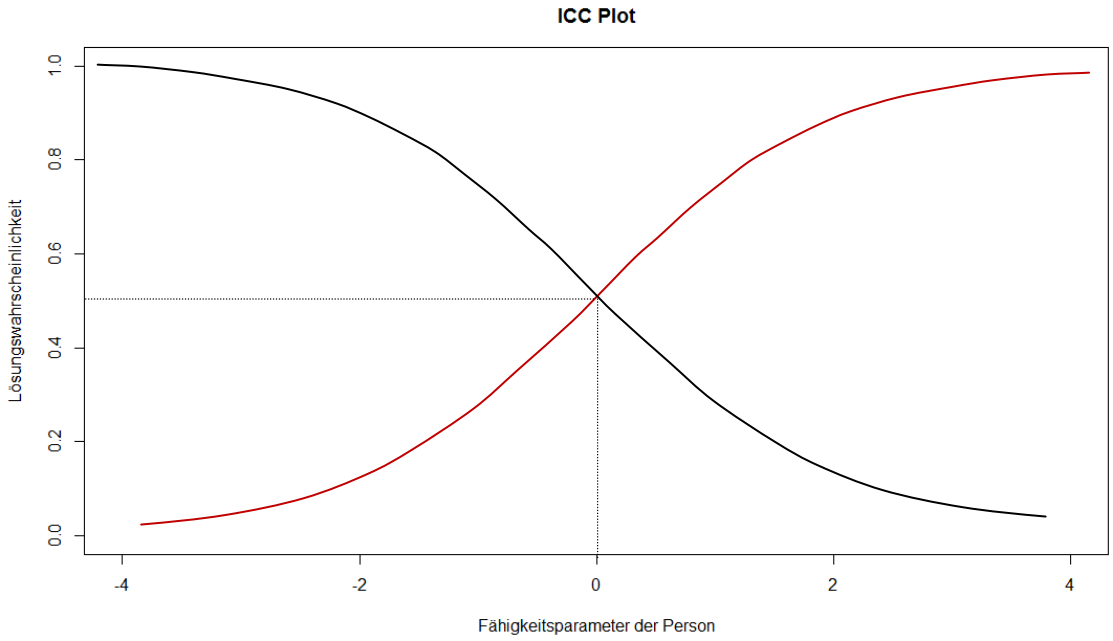


Abbildung 2: ICCs für Lösungswahrscheinlichkeit und Komplement

Die damit verbundene Schwelle liegt bei einer Lösungswahrscheinlichkeit von 50 %, die Person ist gleich fähig wie die Aufgabe schwierig. Ist die Aufgabe schwieriger als die Person fähig, liegt die Lösungswahrscheinlichkeit unter dieser Schwelle (englisch „Threshold"); ist die Person fähiger als die Aufgabe schwierig, liegt die Lösungswahrscheinlichkeit über 50 % (Rost, 2004).

Wie in Kapitel 2.1 angeführt, berechnet man die Lösungswahrscheinlichkeit im Rasch-Modell mit Formel **(2)** (1. Gleichung). Man kann stattdessen aber auch folgende Formel verwenden:

$$\boldsymbol{P(X_{vi} = x_{vi}|\theta_v, \beta_i) = \frac{P(X_{vi} = 1|\theta_v, \beta_i)}{P(X_{vi} = 1|\theta_v, \beta_i) + P(X_{vi} = 0|\theta_v, \beta_i)}} \quad \textbf{(4)}$$

Diese gibt den Anteil der Wahrscheinlichkeit, die Aufgabe zu lösen (in Kategorie 1 zu antworten), an der Wahrscheinlichkeit, in einer der beiden Kategorien zu antworten, an (Rost, 2004). Da sich die beiden Komponenten im Nenner bei nur zwei Antwortmöglichkeiten auf eins (*Wahrscheinlichkeit* + *Gegenwahrscheinlichkeit*) addieren, fällt dieser weg.

In den folgenden Modellen gibt es mehr als zwei Kategoriewahrscheinlichkeiten, da sich die Antwortmöglichkeiten nicht mehr auf „gelöst“ und „nicht gelöst“ beschränken. Daher wird der Begriff der Schwellenwahrscheinlichkeit eingeführt. Bei Modellen mit mehrstufigen Antwortformaten wird für jeweils zwei benachbarte Kategorien die Wahrscheinlichkeit berechnet, in der jeweils höheren zu antworten. Bei einem Modell mit drei Antwortmöglichkeiten wird beispielsweise die Wahrscheinlichkeit betrachtet, in der ersten Kategorie zu antworten, unter der Bedingung, in der ersten oder zweiten Kategorie zu antworten. Diese Wahrscheinlichkeit wird als Schwellenwahrscheinlichkeit bezeichnet (Strobl, 2012).

Betrachtet man nicht mehr zwei benachbarte Kategorien, sondern alle Kategorien gleichzeitig, spricht man von der Kategoriewahrscheinlichkeit (Rost, 2004).

In der folgenden Abbildung sind die Kategoriewahrscheinlichkeiten für eine Aufgabe *i* mit vier Kategorien dargestellt. Die τ_{ij} geben die Schwellenparameter der jeweiligen Kategorie an.

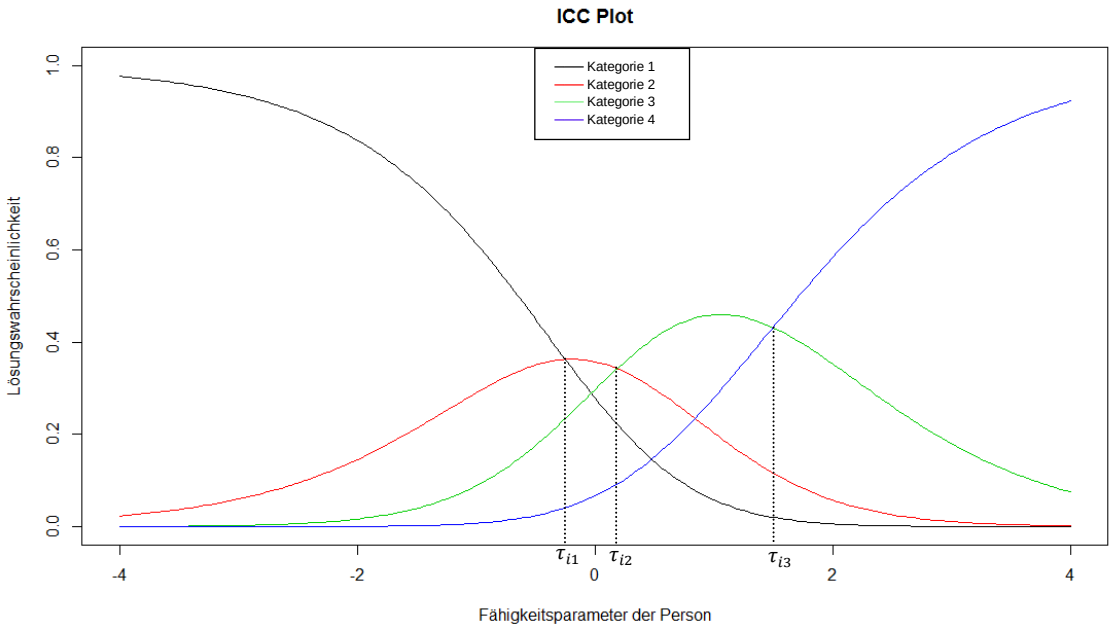


Abbildung 3: CCC Plot für eine Aufgabe des PCM

Man nennt diese Art der Darstellung auch *Category Characteristic Curves* oder *CCCs*.

Die Wahrscheinlichkeit, dass die Antwort einer Person *v* für eine Aufgabe *i* mit m_i Antwortmöglichkeiten in die Kategorie *c* fällt, unter der Bedingung, dass sie in die Kategorie *c* oder *c-1* fällt, ist:

$$\boldsymbol{P(X_{vi} = c | X_{vi} \in \{c-1, c\}, \theta_v, \tau_{ic}) = \frac{e^{\theta_v - \tau_{ic}}}{1 + e^{\theta_v - \tau_{ic}}}} \quad \textbf{(5)}$$

Mit Ausnahme des Parameters τ_{ic} (Schwellenparameter) entspricht diese Formel jener des Rasch-Modells (vergleiche dazu Formel **(1)**) (Strobl, 2012).

Aus dieser bedingten Wahrscheinlichkeit (Schwellenwahrscheinlichkeit) leitete Masters (1982) die unbedingte Wahrscheinlichkeit (Kategoriewahrscheinlichkeit)

$$\boldsymbol{P\left(X_{vi} = c \middle| \theta_v; \tau_{i1}, \dots, \tau_{im_i}\right) = \frac{e^{\sum_{k=0}^{c}(\theta_v - \tau_{ik})}}{\sum_{l=0}^{m_i} e^{\sum_{k=0}^{l}(\theta_v - \tau_k)}}} \quad \textbf{(6)}$$

her, wobei zur Berechnung festgelegt wird, dass $\sum_{k=0}^{0}(\theta_v - \tau_{ik}) = 0$ ist (Strobl, 2012).

Die Auswirkungen eines nachträglichen Dichotomisierens sollen anhand eines Beispiels gezeigt werden. Das Partial-Credit-Modell lässt sich auch in folgender Form schreiben (Ableitung siehe Rost (2004, S. 208-209)).

$$\boldsymbol{P\left(X_{vi} = c \middle| \theta_v; \tau_{i1}, \dots, \tau_{im_i}\right) = \frac{e^{\sum_{k=0}^{c}(\theta_v - \tau_{ik})}}{1 + \sum_{l=1}^{m_i} e^{\sum_{k=0}^{l}(\theta_v - \tau_k)}}} \quad \textbf{(7)}$$

Im Folgenden wird die Kategoriewahrscheinlichkeit $\boldsymbol{P\left(X_{vi} = 2 \middle| \theta_v; \tau_{i1}, \dots, \tau_{im_i}\right)}$ für eine 4-kategorielle Aufgabe anhand der Modellgleichung berechnet.

$$P\left(X_{vi} = 2 \middle| \theta_v; \tau_{i1}, \ldots, \tau_{im_i}\right) = \frac{e^{\sum_{k=0}^{2}(\theta_v - \tau_{ik})}}{1 + \sum_{l=1}^{3} e^{\sum_{k=1}^{l}(\theta_v - \tau_k)}} = \quad (8)$$

$$= \frac{e^{\sum_{k=0}^{2}(\theta_v - \tau_{ik})}}{1 + e^{\sum_{k=1}^{1}(\theta_v - \tau_{ik})} + e^{\sum_{k=1}^{2}(\theta_v - \tau_{ik})} + e^{\sum_{k=1}^{3}(\theta_v - \tau_{ik})}} =$$

$$= \frac{e^{(\theta_v - \tau_{i1}) + (\theta_v - \tau_{i2})]}}{1 + e^{[(\theta_v - \tau_{i1})]} + e^{[(\theta_v - \tau_{i1}) + (\theta_v - \tau_{i2})]} + e^{[(\theta_v - \tau_{i1}) + (\theta_v - \tau_{i2}) + (\theta_v - \tau_{i3})]}}$$

$$| \sum_{k=0}^{0}(\theta_v - \tau_{ik}) = 0$$

Betrachtet man diese Modellgleichung, fällt auf, dass sie im Nenner die Summe der Zähler aller Kategoriewahrscheinlichkeiten aufweist (Rost, 2004). Daraus folgt, dass jede Veränderung der Kategorieanzahl, wie ein Zusammenlegen von Antwortmöglichkeiten, den Nenner und folglich auch die Wahrscheinlichkeit jeder anderen Kategorie verändert. Stimmen die Daten mit den ursprünglichen Kategorien und dem entsprechenden Modell überein, verletzt ein nachträgliches Zusammenlegen der Antwortmöglichkeiten die Gültigkeit des Modells.

Wie auch beim Rasch-Modell, gibt es verschiedene Ansätze, die Personen- und Aufgabenparameter im Partial-Credit-Modell zu schätzen. In der vorliegenden Arbeit wurde die Parameterschätzung mittels Conditional Maximum Likelihood durchgeführt.

Das Graded-Response-Modell

Wie auch das Partial-Credit-Modell (PCM), dient das Graded-Response-Modell (GRM) von Samejima (1969) zur Modellierung mehrstufiger Antwortkategorien. Beim Graded-Response-Modell handelt es sich um ein sogenanntes kumulatives Modell (Strobl, 2012). Im Unterschied zum PCM sind die Kategoriewahrscheinlichkeiten im GRM nicht direkt berechenbar, sondern müssen indirekt über ein sogenanntes Differenzenverfahren berechnet werden.

Im GRM wird die kumulierte Wahrscheinlichkeit betrachtet, dass die Antwort einer Person mindestens in die Kategorie *c* fällt (Strobl, 2012). Es werden nicht, wie beim PCM, immer zwei benachbarte Kategorien betrachtet, sondern es werden alle Antworten auf höhere Kategorien der Aufgabe miteingeschlossen.

Die folgende Formel gibt die Wahrscheinlichkeit an, dass eine Person *v* bei der Aufgabe *i* mindestens in Kategorie *c* antwortet.

$$\boldsymbol{P(X_{vi} \geq c|\theta_v, \beta_{ic}, \alpha_i) = \frac{e^{((\theta_v - \beta_{ic})\alpha_i)}}{1 + e^{((\theta_v - \beta_{ic})\alpha_i)}}} \quad \textbf{(9)}$$

Der Parameter β_{ic} steht für die Schwellenparameter. Der Parameter α_i wird als Diskriminationsparameter bezeichnet. Wie oben schon erwähnt, können die Aufgaben beim GRM unterschiedliche Trennschärfen aufweisen, welche durch den Diskriminationsparameter modelliert werden (Nering & Ostini, 2010). Aus beispielsweise vier Antwortkategorien resultieren drei kumulierte Wahrscheinlichkeiten:

$$\boldsymbol{\pi_{vi,j\geq 1} = P(X_{vi} \geq 1|\theta_v, \beta_{ic}, \alpha_i)} \text{ ... P für Antwort in Kategorie 1 oder höher} \quad \textbf{(10)}$$
$$\boldsymbol{\pi_{vi,j\geq 2} = P(X_{vi} \geq 2|\theta_v, \beta_{ic}, \alpha_i)} \text{ ... P für Antwort in Kategorie 2 oder höher}$$
$$\boldsymbol{\pi_{vi,j\geq 3} = P(X_{vi} \geq 3|\theta_v, \beta_{ic}, \alpha_i)} \text{ ... P für Antwort in Kategorie 3 oder höher}$$

Die Wahrscheinlichkeit einer bestimmten Kategorie *j* wird über die Differenz zweier benachbarter kumulierter Wahrscheinlichkeiten berechnet, beginnend mit der höchsten (Nering & Ostini, 2010).

$$\begin{aligned} \boldsymbol{\pi_{vi,3}} &= \boldsymbol{\pi_{vi,j\geq 3}} \\ \boldsymbol{\pi_{vi,2}} &= \boldsymbol{\pi_{vi,j\geq 2} - \pi_{vi,j\geq 3}} \\ \boldsymbol{\pi_{vi,1}} &= \boldsymbol{\pi_{vi,j\geq 1} - \pi_{vi,j\geq 2}} \\ \boldsymbol{\pi_{vi,0}} &= \boldsymbol{1 - \pi_{vi,j\geq 1}} \end{aligned} \quad \textbf{(11)}$$

Aufgrund dieser Berechnung über die Differenzen spricht man von einem indirekten Modell (Strobl, 2014). Nur die höchste Kategorie kann direkt berechnet werden; die

restlichen Berechnungen erfolgen über die Differenz. Betrachtet man die Formel (11), wird auch bewusst, warum ein nachträgliches Zusammenlegen von Antwortalternativen keinen Einfluss auf die restlichen Kategoriewahrscheinlichkeiten hat. Würde man die oberen beiden Kategorien zusammenlegen, würden sich dadurch die unteren beiden nicht verändern.

Wie dem Rasch-Modell und dem Partial-Credit-Modell liegen auch dem Graded-Response-Modell einige Annahmen zugrunde, wenn auch deutlich weniger als bei den anderen beiden. Die Annahmen der Eindimensionalität und der lokalen stochastischen Unabhängigkeit müssen erfüllt sein. Im Vergleich zum Rasch-Modell liegt allerdings keine spezifische Objektivität und keine „Stichprobenunabhängigkeit“ vor. Die Zulässigkeit unterschiedlicher Trennschärfen wirkt sich auf die Suffizienz aus (Strobl, 2012). Die Randsummen in diesem Modell stellen keine suffizienten Statistiken dar. Dementsprechend können die Aufgaben- und Personenparameter nicht mittels Conditional-Maximum-Likelihood-Methode geschätzt werden.

Ein weiterer Ansatz, die Aufgaben- und Personenparameter zu schätzen, der diese Problematik umgeht, ist die Marginal-Maximum-Likelihood-Methode (Strobl, 2012). Bei dieser Methode werden im ersten Schritt ebenfalls die Aufgabenparameter unabhängig von den Personenparametern geschätzt. Allerdings werden diese nicht, wie bei der Conditional-Maximum-Likelihood-Methode, durch die Randsummen ersetzt, sondern es wird eine Randverteilung (marginale Verteilung) für die Personenparameter angenommen (Strobl, 2012). In den meisten Fällen wird dafür die Normalverteilung herangezogen (Strobl, 2012). Die Likelihood wird mit der marginalen Verteilung multipliziert und anschließend wird über die Personenparameter integriert. Dadurch entsteht die marginale Likelihood, welche nur mehr die Aufgabenparameter als unbekannte Größe enthält (Strobl, 2012). Ist die Verteilungsannahme allerdings unpassend, kann es zu Verzerrungen bei der Schätzung kommen.

Das Modell für kontinuierliche Antwortskalen nach Müller

Müller (1987) entwickelte ein Modell zur Verrechnung kontinuierlicher (analoger) Ratingskalen (CRSM) gemäß der probabilistischen Testtheorie. Ausgehend vom Rating-Scale-Modell (RSM) von Andrich (1978) hat Müller ein Modell für den Grenzfall beschrieben, bei dem die Anzahl der Antwortmöglichkeiten gegen unendlich geht und damit auch die Schwellen unendlich dicht beieinanderliegen.

Das Rating-Scale-Modell ist ein Modell zur Verrechnung mehrstufiger Antwortskalen. Es ist ein Spezialfall des Partial-Credit-Modells, bei dem alle Aufgaben die gleiche Anzahl an Antwortmöglichkeiten aufweisen und die Abstände zwischen den Schwellenparametern gleich sind (Strobl, 2012). Für die Schwellenabstände hat Andrich eine Art Maßeinheit eingeführt und diese als Dispersionsparameter definiert. Müller (1987) erweiterte dieses Modell, indem er eine Antwortskala der Länge *d* mit dem Mittelpunkt *c* betrachtet und diese in *m+1* Intervalle unterteilt. Die Verrechnung erfolgt jeweils über die Mittelpunkte der Intervalle x_i (Müller, 1999).

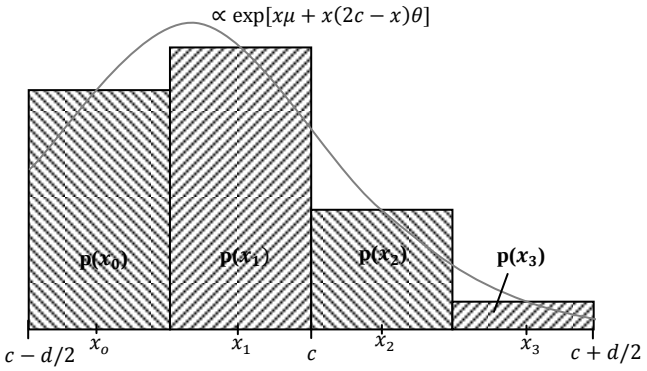


Abbildung 4: Histogramm der Antwortwahrscheinlichkeiten von vier künstlichen Kategorien mit geglätteter Kurve

Mit steigender Anzahl an Intervallen nähern sich die Balken der geglätteten Kurve an. Lässt man $x \rightarrow \infty$ gehen, wird aus der diskreten Zufallsvariable *X* eine kontinuierliche, welche jeden Wert $x \in \left\{c - \frac{d}{2}; c + \frac{d}{2}\right\}$ annehmen kann (Müller, 1999). Die

Wahrscheinlichkeit einer Antwort in einem gegebenen Intervall [a,b] kann durch folgendes Integral bestimmt werden:

$$\boldsymbol{p(a \leq X \leq b|\mu,\lambda) = \frac{\int_a^b exp[x\mu + x(2c - x)\lambda]dx}{\int_a^b exp[t\mu + t(2c - x)\lambda]dt}} \quad \textbf{(12)}$$

wobei μ als Lokationsparameter bezeichnet wird und die Differenz zwischen dem Personen- und Aufgabenparameter ausdrückt (Müller, 1999). Die Variable λ steht für den Dispersionsparameter. Andrich (1978) hat diesen als gleiche Differenz der Schwellenparameter definiert. Lässt man aber die Anzahl der Kategorien gegen unendlich gehen, dann wird die Differenz zweier benachbarter Schwellenparameter null. Daher formulierte Müller (1987) den Dispersionsparameter als gleiche Dichte der Schwellenparameter am Intervall [a,b].

Das Modell für kontinuierliche Antwortskalen gehört aufgrund seiner Herleitung und seiner Eigenschaften ebenfalls zur Familie der Rasch-Modelle. Daraus folgt auch, dass ein nachträgliches Dichotomisieren von Antwortalternativen nicht zulässig ist.

Die Parameter des Modells für kontinuierliche Ratingskalen werden mittels der paarweisen Conditional-Maximum-Likelihood-Methode geschätzt. Aufgrund der Komplexität der Methode wird hier nicht näher darauf eingegangen, sondern an dieser Stelle auf die Arbeit von Müller (1999, S. 123-149) verwiesen.

2.4 Dichotomisierbarkeit mehrkategorieller Rasch-Modelle nach Jansen & Roskam

Jansen und Roskam (Jansen & Roskam, 1986; Roskam & Jansen, 1989; Roskam, 1995) haben mathematisch gezeigt, dass mehrkategorielle eindimensionale Rasch-Modelle nicht invariant gegenüber einer nachträglichen Dichotomisierung sind. Sie argumentieren, dass ein Modell als „dichotomisierbar" angesehen werden kann, wenn es die Bedingung der *ξ-Invarianz* unter der *Joining Assumption* erfüllt. Die Bedingung der *ξ-Invarianz* ist erfüllt, wenn die Einteilung des Antwortkontinuums (dichotom oder mehrkategoriell) keinen Einfluss auf die Personenparameterschätzung hat, ausgenommen einer zulässigen Transformation. Die *Joining Assumption* besagt, dass sich beim Zusammenfassen von Antwortkategorien die beteiligten Wahrscheinlichkeiten addieren. Jansen und Roskam haben gezeigt, dass das eindimensionale mehrkategorielle Rasch-Modell inklusive seiner Verallgemeinerungen und Spezialfälle (Rating-Scale-Modell, Partial-Credit-Modell etc.) unter der Bedingung der *Joining Assumption* fast immer gegen das Prinzip der *ξ-Invarianz* verstößt.

Es ist zu erwähnen, dass es damals zu einer Kontroverse zwischen Jansen & Roskam und Andrich (1995) und in weiterer Folge auch Müller (1995 & 1999) kam. Wie schon erwähnt, gehen Jansen & Roskam von der Modellannahme aus, dass sich die Wahrscheinlichkeiten zweier benachbarter Kategorien bei einem Zusammenlegen addieren. Des Weiteren gehen sie von der Arbeitshypothese aus, dass die Anzahl der Antwortmöglichkeiten keinen Einfluss auf die Präzision der Messung hat und die Messergebnisse nicht beeinflussen sollte. Müller (1999, S.160) schreibt dazu: „*Wir akzeptieren die Idee der θ-Invarianz, nicht jedoch die von Jansen und Roskam verwendete Argumentationsrichtung. Beim Übergang vom diskreten zum kontinuierlichen RS-Modell haben wir ausdrücklich vorausgesetzt, daß das latente Persönlichkeitsmerkmal θ nicht von der Anzahl der Kategoriegrenzen einer Ratingskala, bis hin zu einem Kontinuum, abhängt. Mit unterschiedlich vielen Kategoriegrenzen sind aber unterschiedliche Stimulusbedingungen verbunden, die sich nach unserer Auffassung ähnlich wie unterschiedliche Itemschwierigkeiten auf die Verteilung der manifesten Antworten auswirken. Wir argumentieren daher umgekehrt, daß bei Rasch-Modellen die Vereinigungsannahme verletzt ist. Eine solche Verletzung würden wir nur dann als bedenklich ansehen, wenn die manifesten Kategoriegrenzen*

durch objektive Kriterien definiert wären – was bei den üblichen subjektiven Ratingskalen nicht zutrifft.“ Andrich (1995) hält der Argumentation von Jansen und Roskam entgegen, dass man mehrkategorielle Modelle genau aus dem Grund dichotomen Modellen vorzieht, um eine präzisere Messung zu erhalten. Er zeigt anhand eines anschaulichen Beispiels (Andrich, 1995), dass sich die Kategoriewahrscheinlichkeiten nicht einfach addieren und dass die *Joining Assumption* nicht gelten darf.

Zur Prüfung der Modellgeltung gibt es in der Forschung im Allgemeinen zwei Ansätze: die mathematische Herleitung, wie beispielsweise bei Jansen und Roskam (Jansen & Roskam, 1986; Roskam & Jansen, 1989; Roskam, 1995) und die empirische Überprüfung mittels Simulationsstudien. Die vorliegende Arbeit soll nun als Ergänzung zu den Arbeiten von Jansen und Roskam die vorgestellte Thematik aus empirischer Sicht überprüfen.

2.5 Modellgeltungstests in der Item-Response-Theorie

In diesem Kapitel werden die statistischen Tests zur Überprüfung der Geltung des Rasch-Modells sowie dessen Erweiterungen und des Graded-Response-Modells vorgestellt, welche bei dieser Simulation angewandt wurden. Zur Prüfung der Modellgeltung des Rasch-Modells und des Partial-Credit-Modells wurden der Likelihood-Ratio-Test und die Fit-Indizes berechnet. Die Überprüfung der Modellkonformität des Graded-Response-Modells erfolgte mithilfe von Pearsons χ^2-Test, der Likelihood-Ratio-Statistik G^2 und der M_2^*-Statistik von Cai & Hansen. Auf eine Prüfung der Modellgeltung des Modells für kontinuierliche Antwortskalen nach Müller muss verzichtet werden. Diese ist zwar grundsätzlich durchführbar, wurde aber noch nicht in ein Computerprogramm implementiert.

Likelihood-Ratio-Test nach Andersen

Die Grundlage des Likelihood-Ratio-Tests nach Andersen bildet die Likelihood der Daten. In Kapitel 2.1 zur Parameterschätzung wurde schon erwähnt, dass die Parameter gewählt werden, welche die Likelihoodfunktion maximieren. Je höher die Likelihood ist, desto besser erklärt das Modell die Daten. Mit sogenannten Likelihood-Ratio-Tests lassen sich die Likelihoods zweier Modelle miteinander vergleichen. Dafür wird der Likelihoodquotient, ein Quotient aus den Likelihoods zweier Modelle, berechnet. Grundlage dieser Modelle bilden dieselben Daten. Es können keine Modelle mit unterschiedlicher Datenbasis verwendet werden. So kann man auch nicht prüfen, ob ein Modell besser passt, wenn man Antwortkategorien zusammenlegt (Rost, 2004).

Ausgehend von diesem Grundkonzept hat Andersen (1973) einen Likelihood-Ratio-Test (LRT) zur Prüfung der Personenhomogenität (Teilannahme der spezifischen Objektivität) im Rasch-Modell entwickelt. Dem LRT liegt die Annahme zugrunde, der Test messe in jeder Teilstichprobe dieselbe Fähigkeit oder Eigenschaft, mit anderen Worten, der Test sei für alle Personen homogen. Wenn nun das Rasch-Modell gilt, dürfen sich die Schätzungen der Aufgabenparameter in verschiedenen Teilstichproben nicht unterscheiden (Rost, 2004). Zur Prüfung dieser Hypothese vergleicht der LRT die Schätzungen der Aufgabenparameter von zumindest zwei Teilstichproben mit den Schätzungen der Gesamtstichprobe. Die Parameterschätzung der Teilstichproben und der Gesamtstichprobe erfolgt mittels Conditional-Maximum-Likelihood-Methode. Die

geschätzten Aufgabenparameter für die Gesamtstichprobe $\hat{\beta}$ und die Teilstichproben $\hat{\beta}_u$ für *u=1,...,K* werden in die Likelihood eingesetzt: $cL(r,\hat{\beta})$ und $cL(r,\hat{\beta}_u)$. Zum Vergleich dieser Likelihoods wird ein Quotient (*LQ* für Likelihoodquotient) gebildet:

$$\boldsymbol{LQ} = \frac{\boldsymbol{cL}(\boldsymbol{r},\widehat{\boldsymbol{\beta}})}{\prod_{\boldsymbol{u}=\boldsymbol{1}}^{\boldsymbol{K}-\boldsymbol{1}} \boldsymbol{cL_u}(\boldsymbol{r_u},\widehat{\boldsymbol{\beta}_u})} \quad \textbf{(13)}$$

Gilt das Rasch-Modell, sollten die Aufgabenparameter der Teilstichproben genauso gut passen, wie die der Gesamtstichprobe, das heißt, die Likelihoods sollten annähernd gleich groß sein und der Likelihoodquotient sollte damit nahe bei eins liegen. Um zu prüfen, ob der Likelihoodquotient signifikant von eins abweicht, wurde folgende Teststatistik formuliert (Rost, 2004):

$$T_{LQ} = -2\, ln\, LQ \quad \sim \quad \chi^2_{(K-1)(K-2)} \quad \textbf{(14)}$$

Der Likelihoodquotient wird kleiner 1, wenn der Nenner größer als der Zähler ist, sprich, das Modell mit den Teilstichproben besser passt und die Annahme der Personenhomogenität verletzt ist. Da der Logarithmus einer Zahl kleiner 1 eine negative Zahl ergibt, wird diese mit *-2* multipliziert. Damit werden die Größenverhältnisse umgedreht, sodass bei einer Modellverletzung die Teststatistik groß und der χ^2-Test signifikant wird. Ein signifikantes Ergebnis bedeutet demnach, dass das Rasch-Modell nicht gilt (Strobl, 2012).

Residuen-basierte Personen- und Itemfit-Indizes

Der oben beschriebene Likelihood-Ratio-Test zählt zu den globalen Modelltests, während die im folgenden Teil beschriebenen Fit-Indizes zu den lokalen Modelltests zählen (Rost, 2004). Bei lokalen Modellgeltungstests werden die Antwortmuster einzelner Aufgaben beziehungsweise einzelner Personen bezüglich ihrer Modellkonformität betrachtet. Modellverletzungen auf Personenebene kann man anhand von Personenfit-Indizes und Modellverletzungen auf Aufgabenebene anhand von Itemfit-Indizes identifizieren (Rost, 2004). Man unterscheidet zwischen Likelihood- und Residuen-basierten Fit-Maßen. In der vorliegenden Simulationsstudie wurden Residuen-basierte Fit-Maße berechnet. Als Residuen werden die Abweichungen

zwischen den beobachteten und theoretisch erwarteten Werten bezeichnet (Rost, 2004). Im vorliegenden Fall sind damit die beobachteten und erwarteten Antworten auf einzelne Aufgaben gemeint, x_{vi} (tatsächliche Antwort der Person *v* auf Item *i*) und $E(x_{vi})$. Die Variable p_{vix} gibt die Wahrscheinlichkeit für die beobachtete Antwort an.

Im ersten Schritt wird die Differenz zwischen den tatsächlichen und erwarteten Antworten der Personen berechnet. Diese Differenzen werden anschließend durch die Standardabweichung der erwarteten Antwort dividiert.

$$\boldsymbol{z_{vi} = \frac{(x_{vi} - E(x_{vi}))}{\sqrt{\sum_{x=0}^{m}(x - E(x))^2 p_{vix}}}} \quad \textbf{(15)}$$

$$\boldsymbol{| \; E(x_{vi}) = \sum_{x=0}^{m} x p_{vix}}$$

Damit erhält man für jede Antwort einen standardisierten Wert z_{vi}, einen *Z*-Wert mit Erwartungswert *0* und Varianz *1*. Quadriert man diese *Z*-Werte, damit sich negative und positive Abweichungen nicht aufheben, und addiert sie, erhält man eine χ^2-verteilte Prüfgröße mit Erwartungswert *N* und Varianz *2N* (*N* ist die Anzahl der Summanden, im Fall der Itemfit-Indizes die Personenanzahl und im Fall der Personenfit-Indizes die Anzahl der Aufgaben) (Rost, 2004). Addiert man sie spaltenweise, das heißt, addiert man alle *Z*-Werte einer Aufgabe, spricht man von Itemfit-Indizes; addiert man sie zeilenweise, das heißt, addiert man alle *Z*-Werte einer Person, spricht man von Personenfit-Indizes. Diese Prüfgrößen können nun mit einer χ^2-Verteilung verglichen werden, wobei die äußeren 5 % für eine Modellverletzung sprechen (Rost, 2004). Dies ist nur eines mehrerer Residuen-basierter Fit-Maße. Populäre Fit-Maße, welche auf dem vorgestellten Konzept aufbauen, sind beispielsweise die Infit- und Outfit-Statistiken von Wright und Masters (1990).

Pearsons χ^2-Test

Pearsons χ^2-Test gehört wieder zu den globalen Modelltests, folgt allerdings einem ähnlichen Prinzip wie die Fit-Statistiken. Er geht der Frage nach, ob das zu prüfende Modell zur Vorhersage der empirischen Daten geeignet ist (Rost, 2004). Ein χ^2-Test findet auch in anderen Bereichen der Statistik Anwendung und vergleicht empirisch beobachtete und erwartete Häufigkeiten.

Im Falle der IRT vergleicht er beobachtete und erwartete Patternhäufigkeiten. Das Wort *Antwortpattern* oder *Antwortmuster* bezeichnet den Vektor aller Antworten einer Person oder eines Items. Die Häufigkeiten *p* ergeben sich durch ein Aufrechnen der Personen oder Items, die exakt dasselbe Antwortmuster produziert haben (Rost, 2004). Die Wahrscheinlichkeiten dieser Antwortmuster $\hat{\pi}$ ergeben sich durch die Multiplikation der Einzelwahrscheinlichkeiten (Kategoriewahrscheinlichkeiten) des jeweiligen Modells. Die Teststatistik kann wie folgt berechnet werden, wobei *n* die Stichprobengröße und *j=1,...k* die Antwortpattern angibt (Maydeu-Olivares & Cai, 2006).

$$\chi^2 = \sum_{j=1}^{k} \frac{(p_j - n\hat{\pi}_j)^2}{n\hat{\pi}_j} \quad (16)$$

Die resultierende Teststatistik folgt einer χ^2-Verteilung, wobei große Werte wieder für die Ablehnung des Modells sprechen.

Likelihood-Ratio-Statistik G^2

Die G^2-Statistik (Likelihood-Ratio-Statistik G^2 oder auch G-Test) basiert ebenfalls auf den erwarteten und empirisch beobachteten Patternhäufigkeiten. Wie auch der χ^2-Test, prüft sie die Nullhypothese, dass sich die erwarteten und empirisch beobachteten Häufigkeiten nicht unterscheiden. Die G^2-Statistik bedient sich des natürlichen Logarithmus zweier Likelihoodfunktionen und wird daher auch oft als Likelihood-Ratio-Statistik bezeichnet. Durch die Division zweier Likelihoods kürzt sich einiges weg, und man kann die resultierende Formel wie folgt schreiben (Maydeu-Olivares & Cai, 2006):

$$G^2 = 2\sum_{j=1}^{k} p_j log\left(\frac{p_j}{n\hat{\pi}_j}\right) \quad (17)$$

Wie auch schon die χ^2-Statistik, folgt sie einer χ^2-Verteilung, bei der große Werte für eine Modellverletzung sprechen.

Für beide Statistiken gilt die χ^2-Verteilung allerdings nur, solange die Patternhäufigkeiten nicht zu spärlich besetzt sind. Als Grundregel gilt, dass jedes

Antwortmuster zumindest einmal vorkommen sollte (Rost, 2004). Bei einem dichotomen Antwortformat und zehn Aufgaben wären das $2^{10} = 1024$ Möglichkeiten – eine Anzahl an Möglichkeiten, die schwer zu realisieren ist, insbesondere, sobald die Anzahl an Antwortmöglichkeiten oder Aufgaben steigt. Thissen und Steinberg (1997) postulieren, dass bei einer Anzahl an Antwortmöglichkeiten größer fünf und mehr als sechs Aufgaben eine Approximation durch die χ^2-Verteilung unzulässig ist.

M_2^*-Statistik nach Cai & Hansen

Aufgrund dieser Problematik haben Maydeu-Olivares und Joe (2005) die M_2-Statistik für Modelle mit dichotomem Antwortformat entwickelt, welche von Cai und Hansen (2013) für mehrkategorielle Antwortformate weiterentwickelt wurde. Die M_2- und die M_2^*-Statistik gehören zu den sogenannten *Limited Information Statistics (LIS)*. Im Vergleich zu *Overall Goodness of Fit Statistics*, welche mithilfe der Patternwahrscheinlichkeiten berechnet werden, verwenden *LIS* univariate und bivariate Wahrscheinlichkeiten (univariate und bivariate Momente) (Maydeu-Olivares, 2015). Auf eine Herleitung wird an dieser Stelle verzichtet, es wird aber auf die Arbeiten von Maydeu-Olivares & Joe (2006) und Cai & Hansen (2013) für eine detailliertere Darstellung verwiesen.

Beide Statistiken sind asymptotisch χ^2-verteilt und sensitiver bei Modellverletzungen als Pearsons χ^2-Statistik oder die G^2-Statistik (Cai & Hansen, 2013).

3 Simulationsstudie

3.1 Ziel der Simulationsstudie

Ziel dieser Studie ist es, die Auswirkungen einer nachträglichen Dichotomisierung mehrkategorieller Antwortskalen von der empirischen Seite zu betrachten. Jansen und Roskam (Jansen & Roskam, 1986; Roskam & Jansen, 1989; Roskam, 1995) haben mathematisch gezeigt, dass das Rasch-Modell nicht mehr auf Daten passt, die nachträglich dichotomisiert wurden, wenn das Modell für mehrstufige Antworten (z. B. PCM) für die ursprünglichen Daten gilt. Im Zuge dieser Arbeit soll der Frage nachgegangen werden, wie sich ein nachträgliches Zusammenlegen von Antwortmöglichkeiten auf die Parameterschätzung und die empirische Geltung des Modells auswirkt. Rost (1999, S. 147) betont: „*Diese Modelleigenschaft ordinaler Rasch-Modelle, nicht invariant gegenüber einer nachträglichen Dichotomisierung zu sein, ist von einiger praktischer Relevanz, denn die Dichotomisierung mehrkategorieller Daten ist in vielen empirischen Untersuchungen an der Tagesordnung“*.

An dieser Stelle sollte noch erwähnt werden, dass die verwendeten Modellgeltungstests nicht dafür entwickelt wurden, eine nachträgliche Dichotomisierung zu erkennen. Diese Studie hat einen rein explorativen Charakter und geht der Frage nach, welche Auswirkungen sich im Falle einer nachträglichen Dichotomisierung im Mittel ergeben.

3.2 Methoden

Die Simulation wurde mithilfe der R-Packages, ***eRm*** (Mair, Hatzinger & Maier, 2014), ***mirt*** (Chalmers, 2012), ***pcIRT*** (Hohensinn, 2013) und ***PP*** (Reif, 2014), und des Statistikprogramms ***R*** (R Development Core Team, 2012) durchgeführt. Zur Beschleunigung der Rechenzeit wurde zusätzlich das R-Package ***snow*** (Tierney, Rossini, Li & Sercikova, 2012) verwendet.

Simulationsaufbau

Es wurden pro Modell für mehrstufige Antwortformate (PCM & GRM) 18 Simulationsszenarien und für das Modell für kontinuierliche Antwortskalen (CRSM) drei Simulationsszenarien durchgeführt. Die Anzahl der Personen, die Stelle, an der dichotomisiert wurde, und die Anzahl der Antwortmöglichkeiten wurden variiert,

während die Anzahl von 20 Aufgaben und das Signifikanzniveau von 5 % gleich blieben.

In der folgenden Tabelle (Tabelle 1) ist angeführt, welche Szenarien simuliert wurden. Das Teilungskriterium (*TK*) gibt die Stelle an, an der dichotomisiert wurde. Ein Teilungskriterium *TK < 2* bei vier Antwortmöglichkeiten bedeutet, dass die ersten beiden Kategorien, sprich 0 und 1, zusammengelegt wurden und die „neue" Kategorie 0 bilden und die oberen beiden Kategorien (2 und 3) die „neue" Kategorie 1 bilden.

Tabelle 1: Simulationsszenarien

MODELL	KATEGORIEN	TEILUNGSKRITERIUM (TK)
PCM & GRM		
	4 Antwortmöglichkeiten {0,1,2,3}	TK < 1 {0=0 & 1,2,3=1} TK < 2 {0,1=0 & 2,3=1} TK < 3 {0,1,2=0 & 3=1}
	5 Antwortmöglichkeiten {0,1,2,3,4}	TK < 1 {0=0 & 1,2,3,4=1} TK < 2 {0,1=0 & 2,3,4=1} TK < 4 {0,1,2,3=0 & 4=1}
	6 Antwortmöglichkeiten {0,1,2,3,4,5}	TK < 2 {0,1=0 & 2,3,4,5=1} TK < 3 {0,1,2=0 & 3,4,5=1} TK < 4 {0,1,2,3=0 & 4,5=1}
CRSM		
	Kontinuierlich {0_____1}	Mittelwert Gemäß den empirischen Quartilen Gemäß den theoretischen Quartilen

Die kontinuierlichen Daten wurden einmal dichotom gemäß ihrem Mittelwert und zweimal mehrkategoriell gemäß den empirischen und theoretischen Quartilen unterteilt. Die Stichprobengröße für die in Tabelle 1 durchgeführten Szenarien betrug 1000 bzw. 2000 für die Modelle für mehrstufige Antwortformate und 1000 für das Modell für kontinuierliche Antwortskalen. Jedes Simulationsszenario wurde 10000 Mal wiederholt.

Die Struktur der Szenarien war stets dieselbe:

1. Generierung mehrkategorieller oder kontinuierlicher Daten gemäß den Modellen.
2. Dichotomisierung und/oder mehrkategorielle Unterteilung der Daten.
3. Modellierung der dichotomen, mehrkategoriellen oder kontinuierlichen Daten, einmal mit dem Modell, nach dem die Daten generiert wurden, und einmal mit dem Modell für die zusammengelegten Daten.
4. Berechnung der Korrelationen der Personenparameterschätzungen des ursprünglichen Modells (Modell für mehrstufige oder kontinuierliche Antwortskalen) und des Modells für die nachträglich zusammengelegten Antwortkategorien.
5. Berechnung der Modellgeltungstests

Simulation zum Partial-Credit-Modell

Die mehrkategoriellen Daten gemäß dem Partial-Credit-Modell wurden mithilfe des R-Packages ***PP*** (Reif, 2014) generiert. Es wurden 20 Aufgaben gewählt; dies entspricht einer typischen Anzahl an Aufgaben eines Leistungstests. Die Aufgabenparameter wurden beliebig gewählt (sie befinden sich großteils in einem Bereich zwischen -3 und 3), sind geordnet und wurden Summe-Null normiert. Die genauen Aufgabenparameter können im Simulationsskript nachgelesen werden. Alle verwendeten Befehle sowie der gesamte R-Code sind im Anhang angeführt. Die Personenparameter wurden zufällig aus einer Standardnormalverteilung gezogen. Die mehrkategoriellen Daten gemäß dem Partial-Credit-Modell wurden auf drei verschiedene Arten dichotomisiert (*TK < 1*, *TK < 2 & TK < 3*, für eine Beschreibung siehe Tabelle 1). Daraus resultierten drei Matrizen mit dichotomem Antwortformat (0 und 1 Einträgen). Die Modellierung der mehrkategoriellen Daten erfolgte mit dem Partial-Credit-Modell, die Modellierung der dichotomen Datenmatrizen mit dem Rasch-Modell. Anschließend wurden die Korrelationen der geschätzten Personenparameter des PCM und der drei Rasch-Modelle berechnet. Zur Überprüfung der Modellgeltung wurden sowohl für das PCM als auch für die Rasch-Modelle der Likelihood-Ratio-Test und die Fit-Indizes berechnet. Das Signifikanzniveau wurde für beide Modelltests auf 5 % festgesetzt. Für alle Berechnungen des Likelihood-Ratio-Tests nach Andersen wurde als Teilungskriterium

der *Median des Testscores* verwendet. Mit Ausnahme der Datengenerierung wurden alle Funktionen mithilfe des R-Packages ***eRm*** (Mair, Hatzinger & Maier, 2014) durchgeführt.

Simulation zum Graded-Response-Modell

Die Generierung der mehrkategoriellen Daten gemäß dem GRM sowie alle weiteren Berechnungen dieser Simulation wurden mithilfe des R-Packages ***mirt*** (Chalmers, 2012) durchgeführt. Die Personenparameter wurden, wie auch schon beim PCM, zufällig aus einer Standardnormalverteilung gezogen. Wie in Kapitel 2.2 schon erwähnt, unterscheidet sich das GRM vom PCM dahingehend, dass die Aufgaben unterschiedlich trennscharf sein können. Die simulierten Trennschärfen folgen einer logarithmischen Normalverteilung mit Mittelwert 0,2 und Standardabweichung 0,3. Aufbauend auf den Trennschärfeparametern wurden die Aufgabenparameter festgelegt. Bei diesen wurde darauf geachtet, dass sie nicht zu nah beieinanderliegen, da es sonst zu Problemen bei der Schätzung kam, weil einige Kategoriewahrscheinlichkeiten extrem niedrig waren. Die Verteilung der Trennschärfen und die Berechnung der Aufgabenparameter wurden auf Anraten des Package-Autors so gewählt. Es ist eine Art gängige Praxis, die Trennschärfen gemäß einer logarithmischen Normalverteilung zu modellieren, allerdings könnte auch genauso gut eine Gleichverteilung verwendet werden. Die Überprüfung der Modellgeltung für das mehrkategorielle und dichotome Modell erfolgte mittels Pearsons χ^2-Statistik, G^2-Statistik und M_2^*-Statistik. Auch für diese wurde ein Signifikanzniveau von 5 % festgelegt.

Simulation zum Modell für kontinuierliche Antwortskalen

Die kontinuierlichen Daten gemäß dem Modell von Müller und die Berechnung des Modells für kontinuierliche Ratingskalen wurden mit dem R-Package ***pcIRT*** (Hohensinn, 2013) realisiert. Auch hier wurden die Personenparameter zufällig aus einer Standardnormalverteilung gezogen. Die Aufgabenparameter wurden im Intervall -1,5 und 1,5 gewählt. Der Dispersionsparameter wurde auf 5 festgesetzt. Die kontinuierlichen Daten, die Werte zwischen 0 und 1 annehmen können, wurden einmal dichotomisiert (Teilungspunkt = Mittelwert) und zweimal mehrkategoriell unterteilt (gemäß den empirischen und theoretischen Quartilen). Die Berechnung der Modelle für

dichotome und mehrkategorielle Antwortformate (PCM und Rasch-Modell) sowie die Schätzung der Modellparameter und die Berechnung der Modellgeltungstests (Likelihood-Ratio-Test und Fit-Indizes) erfolgten mit dem R-Package ***eRm*** (Mair, Hatzinger & Maier, 2014).

4 Ergebnisse

Im folgenden Kapitel werden die Ergebnisse der Simulationsstudie in Tabellen dargestellt. Beginnend mit dem Partial-Credit-Modell werden für jedes Modell deskriptive Statistiken zu den Korrelationen der Personenparameterschätzungen sowie die Anzahl vollständig berechneter Signifikanztests und die Anzahl davon signifikanter Ergebnisse angeführt.

4.1 Simulation zum Partial-Credit-Modell

Wie in Kapitel 3.2 schon erwähnt, wurden Daten gemäß dem Partial-Credit-Modell simuliert. Diese wurden anschließend an drei verschiedenen Stellen dichotomisiert. Die resultierenden Matrizen wurden mit dem Rasch-Modell modelliert. Die ursprünglichen Daten wurden mit dem Partial-Credit-Modell verrechnet. Die Personenparameterschätzungen des Partial-Credit-Modells und der Rasch-Modelle wurden korreliert. Dieses Vorgehen wurde 10000 Mal durchgeführt. Von den resultierenden Korrelationskoeffizienten wurden deskriptive Statistiken (Minimum, 25%-Quantil etc.) berechnet. Diese sind in Tabelle 2 für eine Stichprobengröße von 1000 und in Tabelle 3 für eine Stichprobengröße von 2000 angeführt.

Die in Tabelle 2 markierte Zelle lässt sich wie folgt lesen:

Bei einer Stichprobengröße von 1000 und 20 Aufgaben, wobei jede Aufgabe vier Antwortmöglichkeiten besitzt, lag der kleinste Korrelationskoeffizient zwischen dem ursprünglichen (PCM mit vier Antwortmöglichkeiten) und dem dichotomisierten Modell mit TK < 1 bei 0,927. Das Teilungskriterium *TK < 1* bedeutet, dass die Kategorie 0 als 0 kodiert und die Kategorien 1, 2 und 3 zu 1 umkodiert wurden; die resultierenden Daten ergeben die „neue“, dichotome Matrix.

Tabelle 2: Simulation zum PCM - Deskriptive Statistiken zu den Korrelationen der Personenparameter (N=1000)

STICHPROBENGRÖSSE 1000 & 20 AUFGABEN

4 Antwortkategorien		**Minimum**	**25% - Quantil**	**Median**	**Mittelwert**	**75% - Quantil**	**Maximum**	**Standard-abweichung**
	TK < 1	0,9272	0,9380	0,9403	0,9402	0,9424	0,9509	0,00320972
	TK < 2	0,9637	0,9695	0,9705	0,9705	0,9715	0,9760	0,0015043
	TK < 3	0,9260	0,9380	0,9401	0,9401	0,9423	0,9500	0,00316377
5 Antwortkategorien		**Minimum**	**25% - Quantil**	**Median**	**Mittelwert**	**75% - Quantil**	**Maximum**	**Standard-abweichung**
	TK < 1	0,8986	0,9145	0,9173	0,9172	0,9199	0,9317	0,00404234
	TK < 2	0,9492	0,9566	0,9580	0,9580	0,9594	0,9661	0,00210912
	TK < 4	0,9162	0,9346	0,9368	0,9367	0,9388	0,9476	0,00316783
6 Antwortkategorien		**Minimum**	**25% - Quantil**	**Median**	**Mittelwert**	**75% - Quantil**	**Maximum**	**Standard-abweichung**
	TK < 2	0,9403	0,9480	0,9496	0,9496	0,9512	0,9589	0,00244621
	TK < 3	0,9520	0,9590	0,9603	0,9603	0,9616	0,9663	0,00193082
	TK < 4	0,9489	0,9559	0,9573	0,9573	0,9588	0,9640	0,00210378

Tabelle 3: Simulation zum PCM - Deskriptive Statistiken zu den Korrelationen der Personenparameter (N=2000)

STICHPROBENGRÖSSE 2000 & 20 AUFGABEN

4 Antwortkategorien		**Minimum**	**25% - Quantil**	**Median**	**Mittelwert**	**75% - Quantil**	**Maximum**	**Standard-abweichung**
	TK < 1	0,9361	0,9434	0,9449	0,9448	0,9463	0,9526	0,00213055
	TK < 2	0,9674	0,9714	0,9721	0,9721	0,9727	0,9755	0,00103119
	TK < 3	0,9324	0,9409	0,9424	0,9424	0,9439	0,9503	0,00225198
5 Antwortkategorien		**Minimum**	**25% - Quantil**	**Median**	**Mittelwert**	**75% - Quantil**	**Maximum**	**Standard-abweichung**
	TK < 1	0,9022	0,9123	0,9144	0,9143	0,9163	0,9257	0,0030040
	TK < 2	0,9494	0,9541	0,9552	0,9552	0,9563	0,9613	0,00157884
	TK < 4	0,9215	0,9297	0,9314	0,9314	0,9330	0,9396	0,00243722
6 Antwortkategorien		**Minimum**	**25% - Quantil**	**Median**	**Mittelwert**	**75% - Quantil**	**Maximum**	**Standard-abweichung**
	TK < 2	0,9371	0,9445	0,9458	0,9458	0,9470	0,9532	0,00185641
	TK < 3	0,9511	0,9561	0,9571	0,9571	0,9581	0,9625	0,00151492
	TK < 4	0,9472	0,9529	0,9540	0,9540	0,9551	0,9599	0,00163234

Wie man der Tabelle entnehmen kann, sind die Korrelationskoeffizienten durchgehend sehr hoch. Vor allem bei den mittleren Teilungskriterien liegen die Korrelationskoeffizienten im Mittel zwischen 0,95 und 0,97. Die niedrigsten Korrelationskoeffizienten und die größte Streuung traten bei fünf Antwortmöglichkeiten und dem Teilungskriterium $TK < 1$ auf. Ein Korrelationskoeffizient von eins konnte nicht erreicht werden.

An dieser Stelle könnte man argumentieren, dass die Gleichheit der Personenparameterschätzungen nicht relevant ist, sollten die Ränge übereinstimmen. Vergleicht man zwei Personen anhand ihrer kognitiven Fähigkeiten mithilfe eines Intelligenz-Diagnostikums, sollte die fähigere Person auch nach einer Dichotomisierung in der Rangreihung vorne sein. Ausgehend von folgender Überlegung wurden die Rangkorrelationen der Personenparameterschätzungen zur Simulation des Partial-Credit-Modells (N=1000) berechnet. Diese sind in der folgenden Tabelle (Tabelle 4) angeführt. Zum Vergleich wurden auch nochmals die Produkt-Moment-Korrelationen aus Tabelle 2 beigefügt.

Die deskriptiven Statistiken der Produkt-Moment-Korrelation sind in Schwarz angeführt, die Rangkorrelationen in Rot. Es zeigt sich, dass sich die Korrelationskoeffizienten nur minimal unterscheiden, das heißt, dass auch bei der Rangkorrelation ein Korrelationskoeffizient von eins nicht erreicht wird.

Tabelle 4: Simulation zum PCM - Vergleich Produkt-Moment- und Rangkorrelation

STICHPROBENGRÖSSE 1000 & 20 AUFGABEN

4 Antwortkategorien		**Minimum**	**25% - Quantil**	**Median**	**Mittelwert**	**75% - Quantil**	**Maximum**	**Standard-abweichung**
	TK < 1	0,9272	0,9380	0,9403	0,9402	0,9424	0,9509	0,0032097
		0,9311	**0,9427**	**0,9449**	**0,9448**	**0,9470**	**0,9561**	**0,0031256**
	TK < 2	0,9637	0,9695	0,9705	0,9705	0,9715	0,9760	0,0015043
		0,9651	**0,9711**	**0,9721**	**0,9721**	**0,9731**	**0,9773**	**0,0015392**
	TK < 3	0,9260	0,9380	0,9401	0,9401	0,9423	0,9500	0,0031637
		0,9327	**0,9437**	**0,9457**	**0,9457**	**0,9477**	**0,9559**	**0,0029855**
5 Antwortkategorien		**Minimum**	**25% - Quantil**	**Median**	**Mittelwert**	**75% - Quantil**	**Maximum**	**Standard-abweichung**
	TK < 1	0,8986	0,9145	0,9173	0,9172	0,9199	0,9317	0,0040423
		0,8940	**0,9107**	**0,9140**	**0,9139**	**0,9172**	**0,9298**	**0,0048953**
	TK < 2	0,9492	0,9566	0,9580	0,9580	0,9594	0,9661	0,0021091
		0,9432	**0,9526**	**0,9544**	**0,9543**	**0,9562**	**0,9632**	**0,0026943**
	TK < 4	0,9162	0,9346	0,9368	0,9367	0,9388	0,9476	0,0031678
		0,9319	**0,9422**	**0,9442**	**0,9442**	**0,9462**	**0,9537**	**0,0029255**

Tabelle wird fortgesetzt...

6 Antwortkategorien		Minimum	25% - Quantil	Median	Mittelwert	75% - Quantil	Maximum	Standard-abweichung
	TK < 2	0,9403	0,9480	0,9496	0,9496	0,9512	0,9589	0,00244621
		0,9281	**0,9405**	**0,9429**	**0,9428**	**0,9451**	**0,9546**	**0,00341813**
	TK < 3	0,9520	0,9590	0,9603	0,9603	0,9616	0,9663	0,00193082
		0,9465	**0,9545**	**0,9563**	**0,9562**	**0,9579**	**0,9649**	**0,00249262**
	TK < 4	0,9489	0,9559	0,9573	0,9573	0,9588	0,9640	0,00210378
		0,9505	**0,9582**	**0,9596**	**0,9596**	**0,9610**	**0,9661**	**0,00213449**

*** In Schwarz sind die deskriptiven Statistiken zu den Produkt-Moment-Korrelationen angegeben**
*** In Rot sind die deskriptiven Statistiken zu den Rangkorrelationen angegeben**

In Tabelle 5 (Stichprobengröße 1000) und Tabelle 6 (Stichprobengröße 2000) werden die Ergebnisse des Likelihood-Ratio-Tests für die verschiedenen Stichprobengrößen, Antwortmöglichkeiten und Teilungskriterien angeführt.

Tabelle 5: Simulation zum PCM - Ergebnisse des Likelihood-Ratio-Tests (N=1000)

STICHPROBENGRÖSSE 1000 & 20 AUFGABEN

4 Antwortkategorien	**Signifikante Ergebnisse**	**Abbrüche**
Mehrkategorielle Daten	433 (4,33 %)	0
Dichotome Daten TK < 1	9861 (99 %)	39 (0,39 %)
Dichotome Daten TK < 2	2711 (27,11 %)	0
Dichotome Daten TK < 3	9799 (97,99 %)	0
5 Antwortkategorien	**Signifikante Ergebnisse**	**Abbrüche**
Mehrkategorielle Daten	535 (5,35 %)	0
Dichotome Daten TK < 1	9786 (97,9 %)	4 (0,04 %)
Dichotome Daten TK < 2	4571 (46,2 %)	106 (1,06 %)
Dichotome Daten TK < 4	9692 (97,1 %)	19 (0,19 %)
6 Antwortkategorien	**Signifikante Ergebnisse**	**Abbrüche**
Mehrkategorielle Daten	430 (4,35 %)	113 (1,13 %)
Dichotome Daten TK < 2	7200 (72 %)	0
Dichotome Daten TK < 3	1214 (12,14 %)	0
Dichotome Daten TK < 4	9597 (96,07 %)	11 (0,11 %)

Es wird aufgelistet, wie viele Signifikanztests nicht vollständig berechnet werden konnten (Abbrüche) und wie viele davon signifikant waren. Die Prozentzahl in Klammern gibt den relativen Anteil signifikanter Ergebnisse bei der Anzahl der gültigen Durchgänge an. Die Anzahl gültiger Durchgänge weicht dann von 10000 ab, wenn das Rasch-Modell oder der Likelihood-Ratio-Test mit den simulierten Datensätzen nicht berechnet werden konnte.

Beim Likelihood-Ratio-Test kann es beispielsweise zu einem Abbruch kommen, wenn durch das Zusammenlegen der Antwortmöglichkeiten sehr viele Personen einen sehr hohen Testscore erzielt haben. Durch das Teilungskriterium *Median des Testscores* könnte es passieren, dass in der höheren Fähigkeitsgruppe nur mehr Personen sind, die alle Aufgaben gelöst haben. Somit ist in dieser Gruppe kein Item schätzbar und es kommt zu einem Abbruch. Dasselbe gilt natürlich auch, wenn viele Personen einen sehr niedrigen Testscore erzielt haben.

Tabelle 6: Simulation zum PCM - Ergebnisse des Likelihood-Ratio-Tests (N=2000)

STICHPROBENGRÖSSE 2000 & 20 AUFGABEN

4 Antwortkategorien	**Signifikante Ergebnisse**	**Abbrüche**
Mehrkategorielle Daten	409 (4,09 %)	0
Dichotome Daten TK < 1	9996 (100 %)	4 (0,04 %)
Dichotome Daten TK < 2	6247 (62,47 %)	0
Dichotome Daten TK < 3	10000 (100 %)	0
5 Antwortkategorien	**Signifikante Ergebnisse**	**Abbrüche**
Mehrkategorielle Daten	490 (4,9 %)	0
Dichotome Daten TK < 1	9994 (100 %)	6 (0,06 %)
Dichotome Daten TK < 2	8281 (83,4 %)	71 (0,71 %)
Dichotome Daten TK < 4	9996 (99,97 %)	2 (0,02 %)
6 Antwortkategorien	**Signifikante Ergebnisse**	**Abbrüche**
Mehrkategorielle Daten	545 (5,45 %)	6 (0,06 %)
Dichotome Daten TK < 2	9801 (98,01 %)	0
Dichotome Daten TK < 3	3228 (32,28 %)	0
Dichotome Daten TK < 4	9996 (99,96 %)	1 (0,01 %)

Die Anzahl signifikanter Ergebnisse bei der mehrkategoriellen Modellierung bewegt sich in einem Bereich zwischen 409 und 545; das entspricht einem aktualisiertem Risiko 1. Art von 4,09 % bis 5,45 %. Damit liegen die Ergebnisse innerhalb der von Rasch und Guiard (2004) postulierten 20%-Robustheit ($0{,}04 \leq \alpha_{akt} \leq 0{,}06$).

Auffallend ist der Anteil der signifikanten Ergebnisse bei der dichotomen Modellierung. Bei den Teilungskriterien, die mittig teilen (*TK < 2* bei 4 und 5 Antwortkategorien und *TK < 3* bei 6 Antwortkategorien), liegt die Anzahl signifikanter Ergebnisse bei einer Stichprobengröße von 1000 zwischen 12,14 % und 46,2 %.

Diese steigt jedoch mit zunehmender Stichprobengröße (N=2000) auf 32,28 % – 83,4 % an. Bei den oberen und unteren Teilungskriterien zeigen sich noch extremere Ergebnisse. Mit Ausnahme eines Ergebnisses (72 % signifikante Ergebnisse bei 6 Antwortkategorien, *TK < 2* und *N=1000*) kam es bei den restlichen Simulationsszenarien zu 98 % – 100 % signifikanten Ergebnissen. Das heißt, in 98 % – 100 % der Durchläufe zeigte der Likelihood-Ratio-Test eine Modellverletzung auf.

In Tabelle 7 und 8 (Itemfit-Indizes) und Tabelle 9 und 10 (Personenfit-Indizes) sind die Ergebnisse der Berechnung der Fit-Indizes angeführt. In Tabelle 7 (*N*=1000) und 8 (*N*=2000) ist der relative Anteil an Tests angegeben, die eine bestimmte Anzahl an nicht modellkonformen Aufgaben aufweisen. In der Praxis müssten die Aufgaben aus dem Test genommen und das Rasch-Modell neu berechnet werden.

Die erste Zeile von Tabelle 7 kann wie folgt gelesen werden:
Bei 77,19 % der Berechnungen der Itemfit-Indizes kam es zu keinen signifikanten Ergebnissen, bei 20,7 % war ein Item, bei 2,07 % waren zwei Items und bei 0,04 % waren drei Items signifikant. Die gesamte Zeile addiert sich auf 100 %.

Bei dieser Berechnung kam es zu keinen Abbrüchen, damit entsprechen die relativen den absoluten Häufigkeiten. Aufgrund der Übersichtlichkeit wurde darauf verzichtet, die absoluten Häufigkeiten anzugeben.

Tabelle 7: Simulation zum PCM - Ergebnisse der Item-Fit-Indizes (N=1000)

STICHPROBENGRÖSSE 1000 & 20 AUFGABEN

	Anzahl nicht modellkonformer Aufgaben								
4 Antwortkategorien	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
mehrkategorielle Daten	77,19 %	20,7 %	2,07 %	0,04 %	0	0	0	0	0
dichotome Daten TK < 1	1,25 %	8,99 %	29,9 %	36,19 %	18,92 %	4,21 %	0,51 %	0,03 %	0
dichotome Daten TK < 2	8,64 %	26,06 %	32,3 %	20,44 %	9,26 %	2,74 %	0,47 %	0,08 %	0,01 %
dichotome Daten TK < 3	1,19 %	14,62 %	35,28 %	32,69 %	13,41 %	2,6 %	0,2 %	0,01 %	0
5 Antwortkategorien	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
mehrkategorielle Daten	61,91 %	31,58 %	5,99 %	0,51 %	0,01 %	0	0	0	0
dichotome Daten TK < 1	1,99 %	13,42 %	31,84 %	33,24 %	15,98 %	3,21 %	0,31 %	0,01 %	0
dichotome Daten TK < 2	12,04 %	29,01 %	33,34 %	17,69 %	6,33 %	1,39 %	0,19 %	0,01 %	0
dichotome Daten TK < 4	2,36 %	16,9 %	37 %	30,33 %	11,34 %	1,86 %	0,19 %	0,02 %	0
6 Antwortkategorien	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
mehrkategorielle Daten	82,21 %	16,59 %	1,16 %	0,04 %	0	0	0	0	0
dichotome Daten TK < 2	5,49 %	24,78 %	34,36 %	23,8 %	9,09 %	2,12 %	0,34 %	0,02 %	0
dichotome Daten TK < 3	16,57 %	34,34 %	29,63 %	14,12 %	4,4 %	0,79 %	0,15 %	0	0
dichotome Daten TK < 4	1,71 %	10,91 %	28,75 %	32,94 %	19 %	5,66 %	0,97 %	0,06 %	0

Tabelle 8: Simulation zum PCM - Ergebnisse der Item-Fit-Indizes (N=2000)

STICHPROBENGRÖSSE 2000 & 20 AUFGABEN

	Anzahl nicht modellkonformer Aufgaben								
4 Antwortkategorien	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
mehrkategorielle Daten	86,09 %	13,14 %	0,76 %	0,01 %	0	0	0	0	0
dichotome Daten TK < 1	0,07 %	1,08 %	11,5 %	34,84 %	35,6 %	14,27 %	2,5 %	0,14 %	0
dichotome Daten TK < 2	9,06 %	27,44 %	32,7 %	21,01 %	7,46 %	2,05 %	0,25 %	0,02 %	0,01 %
dichotome Daten TK < 3	0,04 %	3,65 %	22,76 %	38,51 %	26,38 %	7,71 %	0,94 %	0,01 %	0
5 Antwortkategorien	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
mehrkategorielle Daten	78,44 %	19,79 %	1,73 %	0,04 %	0	0	0	0	0
dichotome Daten TK < 1	0,33 %	4,91 %	21,51 %	35,84 %	27,69 %	8,74 %	0,97 %	0,01 %	0
dichotome Daten TK < 2	12,87 %	34,08 %	32,45 %	15,1 %	4,8 %	0,63 %	0,06 %	0,01 %	0
dichotome Daten TK < 4	0,69 %	11,72 %	43,67 %	33,51 %	9,44 %	0,89 %	0,07 %	0,01 %	0
6 Antwortkategorien	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
mehrkategorielle Daten	92,61 %	7,24 %	0,14 %	0,01 %	0	0	0	0	
dichotome Daten TK < 2	4,13 %	25,35 %	36,82 %	23,7 %	7,89 %	1,93 %	0,17 %	0,01 %	0
dichotome Daten TK < 3	22,04 %	38,21 %	26,12 %	10,13 %	2,89 %	0,53 %	0,07 %	0,01 %	0
dichotome Daten TK < 4	0,43 %	5,61 %	23,59 %	38,7 %	23,32 %	6,86 %	1,39 %	0,1 %	0

Tabelle 9: Simulation zum PCM - Ergebnisse der Person-Fit-Indizes (N=1000)

STICHPROBENGRÖSSE 1000 & 20 AUFGABEN

	Anzahl nicht modellkonformer Personen							
4 Antwortkategorien	**0**	**{1-20}**	**{21-40}**	**{41-60}**	**{61-80}**	**{81-100}**	**{101-120}**	**{121-140}**
kategorial	0	0	0,02 %	34,61 %	64,94 %	0,43 %	0	0
TK < 1	0,37 %	0	0	1,04 %	74,57 %	24 %	0,02 %	0
TK < 2	0	0	0	0	0,01 %	16,47 %	78,29 %	5,23 %
TK < 3	0	0	0	0,66 %	71,28 %	28,05 %	0,01 %	0
5 Antwortkategorien	**0**	**{1-20}**	**{21-40}**	**{41-60}**	**{61-80}**	**{81-100}**	**{101-120}**	**{121-140}**
kategorial	0,04 %	0	0	0,05 %	42,14 %	57,14 %	0,63 %	0
TK < 1	0,04 %	0	0	0,45 %	62,46 %	36,9 %	0,15 %	0
TK < 2	1,04 %	0	0	0,23 %	57,71 %	40,93 %	0,09 %	0
TK < 4	0,05 %	0	2,66 %	90,9 %	6,39 %	0	0	0
6 Antwortkategorien	**0**	**{1-20}**	**{21-40}**	**{41-60}**	**{61-80}**	**{81-100}**	**{101-120}**	**{121-140}**
kategorial	0	0	0	21,34 %	77,08 %	1,58 %	0	0
TK < 2	0	0	0	7,21%	87,6 %	5,19 %	0	0
TK < 3	0	0	0	8,38 %	86,8 %	4,82 %	0	0
TK < 4	0,02 %	0	0,46 %	76,13 %	23,38 %	0,01 %	0	0

Tabelle 10: Simulation zum PCM - Ergebnisse der Person-Fit-Indizes (N=2000)

STICHPROBENGRÖSSE 2000 & 20 AUFGABEN

	Anzahl nicht modellkonformer Personen							
4 Antwortkategorien	**0**	**{1-36}**	**{37-72}**	**{73-108}**	**{109-144}**	**{145-180}**	**{181-216}**	**{217-253}**
kategorial	0	0	0	1,21 %	94,28 %	4,51 %	0	0
TK < 1	0,04 %	0	0	0	24,97 %	74,89 %	0,1 %	0
TK < 2	0	0	0	0	0	0,11 %	59,44 %	40,45 %
TK < 3	0	0	0	0	22,88 %	77,05 %	0,07 %	0
5 Antwortkategorien	**0**	**{1-36}**	**{37-72}**	**{73-108}**	**{109-144}**	**{145-180}**	**{181-216}**	**{217-253}**
kategorial	0	0	0	0	3,97 %	92,73 %	3,3 %	0
TK < 1	0,06 %	0	0	0	14,25 %	84,67 %	1,02 %	0
TK < 2	0,66 %	0	0	0	4,43 %	92,7 %	2,21 %	0
TK < 4	0	0	0,03 %	89,26 %	10,71 %	0	0	0
6 Antwortkategorien	**0**	**{1-36}**	**{37-72}**	**{73-108}**	**{109-144}**	**{145-180}**	**{181-216}**	**{217-253}**
kategorial	0	0	0	0,96 %	93,81 %	5,23 %	0	0
TK < 2	0	0	0	0,02 %	61,93 %	38,05 %	0	0
TK < 3	0	0	0	0,02 %	54,97 %	44,99 %	0,02 %	0
TK < 4	0	0	0	43,37 %	56,59 %	0,04 %	0	0

4.2 Simulation zum Graded-Response-Modell

Der Aufbau dieser Simulation ist sehr ähnlich wie der des Partial-Credit-Modells. Es wurden künstliche Daten simuliert, die auf das Graded-Response-Modell passen. Variiert wurden die Stichprobengröße, die Anzahl der Antwortmöglichkeiten und die Art der Dichotomisierung. Der Vergleichbarkeit halber entsprechen die Simulationsszenarien denen des Partial-Credit-Modells. Wie auch bei der oben angeführten Simulation, wurden deskriptive Statistiken zu den Korrelationen der Personenparameterschätzungen der mehrkategoriellen und dichotomen Modellierung berechnet. Anschließend wurde die Geltung der Modelle überprüft. Für die Überprüfung der Modellgeltung wurden Pearsons χ^2-Statistik, die G^2-Statistik und die M_2^*-Statistik herangezogen.

In Tabelle 11 sind die deskriptiven Statistiken zu den Korrelationen der Personenparameterschätzungen angeführt. Wie auch schon beim PCM, wurden für eine Stichprobengröße von N=1000 sowohl die Produkt-Moment-Korrelation als auch die Rangkorrelation berechnet. Die deskriptiven Statistiken der Produkt-Moment-Korrelation sind in Schwarz angeführt, die Rangkorrelationen in Rot. Es zeigt sich, dass sich die Korrelationskoeffizienten nur minimal unterscheiden, das heißt, dass auch bei der Rangkorrelation ein Korrelationskoeffizient von eins nicht erreicht wird.

Tabelle 12 beinhaltet die deskriptiven Statistiken zu den Produkt-Moment-Korrelationen der Personenparameterschätzungen bei einer Stichprobengröße von N=2000.

Tabelle 11: Simulation zum GRM - Deskriptive Statistiken zu den Korrelationen der Personenparameter (N=1000)

STICHPROBENGRÖSSE 1000 & 20 AUFGABEN

4 Antwortkategorien	**Minimum**	**25% - Quantil**	**Median**	**Mittelwert**	**75% - Quantil**	**Maximum**	**Standard-abweichung**
TK < 1	0,9647	0,9711	0,9724	0,9723	0,9736	0,9793	0,00189195
	0,9648	**0,9728**	**0,9743**	**0,9742**	**0,9758**	**0,9817**	**0,00226424**
TK < 2	0,9743	0,9792	0,98	0,98	0,9807	0,9839	0,00115584
	0,9766	**0,9814**	**0,9823**	**0,9822**	**0,9831**	**0,9858**	**0,00119504**
TK < 3	0,9765	0,9803	0,9812	0,9812	0,9821	0,9859	0,00130362
	0,9779	**0,9832**	**0,9841**	**0,9841**	**0,9850**	**0,9889**	**0,00138677**
5 Antwortkategorien	**Minimum**	**25% - Quantil**	**Median**	**Mittelwert**	**75% - Quantil**	**Maximum**	**Standard-abweichung**
TK < 1	0,9504	0,9587	0,9605	0,9604	0,9623	0,97	0,0026411
	0,95	**0,9619**	**0,9640**	**0,9638**	**0,966**	**0,9741**	**0,00301234**
TK < 2	0,9641	0,9696	0,9707	0,9706	0,9717	0,9758	0,00158506
	0,967	**0,9734**	**0,9746**	**0,9745**	**0,9757**	**0,9796**	**0,00166402**
TK < 4	0,947	0,9561	0,9581	0,958	0,9599	0,968	0,00278414
	0,9474	**0,9596**	**0,9618**	**0,9616**	**0,9638**	**0,972**	**0,00314748**

Tabelle wird fortgesetzt...

6 Antwortkategorien	Minimum	25% - Quantil	Median	Mittelwert	75% - Quantil	Maximum	Standard-abweichung
TK < 2	0,9615	0,9683	0,9695	0,9695	0,9706	0,9753	0,00171407
	0,9653	**0,9736**	**0,9747**	**0,9747**	**0,9758**	**0,9799**	**0,00165984**
TK < 3	0,9635	0,9683	0,9694	0,9694	0,9705	0,9749	0,00156723
	0,9659	**0,972**	**0,9733**	**0,9732**	**0,9744**	**0,9789**	**0,00177882**
TK < 4	0,9511	0,9591	0,9608	0,9607	0,9623	0,9693	0,00235866
	0,9512	**0,9616**	**0,9635**	**0,9634**	**0,9653**	**0,9724**	**0,00277718**

*** In Schwarz sind die deskriptiven Statistiken zu den Produkt-Moment-Korrelationen angegeben**
*** In Rot sind die deskriptiven Statistiken zu den Rangkorrelationen angegeben**

Tabelle 12: Simulation zum GRM - Deskriptive Statistiken zu den Korrelationen der Personenparameter (N=2000)

STICHPROBENGRÖSSE 2000 & 20 AUFGABEN

4 Antwortkategorien	**Minimum**	**25% - Quantil**	**Median**	**Mittelwert**	**75% - Quantil**	**Maximum**	**Standard-abweichung**
TK < 1	0,9736	0,9776	0,9783	0,9783	0,979	0,9819	0,00106702
TK < 2	0,9748	0,9774	0,978	0,978	0,9785	0,9809	0,00084891
TK < 3	0,9615	0,9663	0,9674	0,9674	0,9684	0,9728	0,00156259
5 Antwortkategorien	**Minimum**	**25% - Quantil**	**Median**	**Mittelwert**	**75% - Quantil**	**Maximum**	**Standard-abweichung**
TK < 1	0,9606	0,9666	0,9677	0,9677	0,9687	0,9734	0,00159414
TK < 2	0,968	0,9722	0,9729	0,9728	0,9735	0,9765	0,00103189
TK < 4	0,9433	0,9502	0,9517	0,9517	0,9532	0,9604	0,00225507
6 Antwortkategorien	**Minimum**	**25% - Quantil**	**Median**	**Mittelwert**	**75% - Quantil**	**Maximum**	**Standard-abweichung**
TK < 2	0,9562	0,9622	0,9632	0,9632	0,9643	0,9684	0,00151553
TK < 3	0,9649	0,9683	0,969	0,969	0,9698	0,9727	0,00107619
TK < 4	0,9594	0,9644	0,9654	0,9653	0,9663	0,9702	0,00138914

Im Vergleich zu den Korrelationskoeffizienten der Simulation des Partial-Credit-Modells fallen die des Graded-Response-Modells deutlich höher aus. Des Weiteren ist zu beachten, dass die Schwankungsbreite der Korrelationskoeffizienten des GRM deutlich enger ist (Min = 0,943 und Max = 0,986) als die des PCM (Min = 0,899 und Max = 0,976). Das heißt, die Korrelationskoeffizienten der unterschiedlichen Dichotomisierungsarten unterscheiden sich nicht so stark. Zur Erinnerung, beim PCM erzielten die mittleren Teilungskriterien durchwegs die höchsten Korrelationskoeffizienten.

In den folgenden Tabellen, Tabelle 13 und Tabelle 14, sind die Ergebnisse zu den Modelltests des Graded-Response-Modells angeführt. Für jeden Modelltest, nämlich Pearsons χ^2-Statistik, die G^2-Statistik und die M_2^*-Statistik, sind die Anzahl der Abbrüche pro Simulationsszenario und die Anzahl der signifikanten Ergebnisse angegeben. Da es bei keinem der Modelltests zu Abbrüchen kam, entsprechen die Prozentzahlen den absoluten Häufigkeiten. Daher wird auf eine Darstellung dieser der Übersicht halber verzichtet.

Die erste Zeile der Tabelle 13 lässt sich wie folgt lesen:
Bei einer Stichprobengröße von 1000 und 20 Aufgaben mit jeweils vier Antwortmöglichkeiten kam es bei der M_2^*-Statistik in 512 von 10000 Durchläufen (5,12 %) zu signifikanten Ergebnissen. Die χ^2-Statistik zeigte in 11,9 % und die G^2-Statistik in 12,94 % der Durchläufe eine Modellverletzung auf.

Tabelle 13: Simulation zum GRM - Modellgeltungstests M_2^*-Statistik, Pearsons χ^2-Statistik, G^2-Statistik (N=1000)

STICHPROBENGRÖSSE 1000 & 20 AUFGABEN

	M_2^*-Statistik		χ^2-Statistik		G^2-Statistik	
4 Antwortkategorien	**sign. Ergebnisse**	**Abbrüche**	**sign. Ergebnisse**	**Abbrüche**	**sign. Ergebnisse**	**Abbrüche**
mehrkategorielle Daten	5,12 %	0	11,9 %	0	12,94 %	0
dichotome Daten TK < 1	5,06 %	0	52,03 %	0	51,62 %	0
dichotome Daten TK < 2	5,17 %	0	7,74 %	0	7,67 %	0
dichotome Daten TK < 3	5,09 %	0	2,36 %	0	2,41 %	0
5 Antwortkategorien	**sign. Ergebnisse**	**Abbrüche**	**sign. Ergebnisse**	**Abbrüche**	**sign. Ergebnisse**	**Abbrüche**
mehrkategorielle Daten	5,22 %	0	7,61 %	0	8,24 %	0
dichotome Daten TK < 1	5,3 %	0	70,04 %	0	69,32 %	0
dichotome Daten TK < 2	5,09 %	0	10,02 %	0	10,04 %	0
dichotome Daten TK < 4	5,12 %	0	2,85 %	0	2,56 %	0
6 Antwortkategorien	**sign. Ergebnisse**	**Abbrüche**	**sign. Ergebnisse**	**Abbrüche**	**sign. Ergebnisse**	**Abbrüche**
mehrkategorielle Daten	5,02 %	0	6,47 %	0	6,75 %	0
dichotome Daten TK < 2	4,86 %	0	27,4 %	0	27,22 %	0
dichotome Daten TK < 3	5,06 %	0	2,68 %	0	2,84 %	0
dichotome Daten TK < 4	5,17 %	0	2,46 %	0	2,69 %	0

Tabelle 14: Simulation zum GRM - Modellgeltungstests M_2^*-Statistik, Persons χ^2-Statistik, G^2-Statistik (N=2000)

STICHPROBENGRÖSSE 2000 & 20 AUFGABEN

	M_2^*-Statistik		χ^2-Statistik		G^2-Statistik	
4 Antwortkategorien	**sign. Ergebnisse**	**Abbrüche**	**sign. Ergebnisse**	**Abbrüche**	**sign. Ergebnisse**	**Abbrüche**
mehrkategorielle Daten	5,03 %	0	4,77 %	0	5,52 %	0
dichotome Daten TK < 1	4,79 %	0	5,69 %	0	5,7 %	0
dichotome Daten TK < 2	4,91 %	0	2,42 %	0	2,53 %	0
dichotome Daten TK < 3	4,84 %	0	2,36 %	0	2,56 %	0
5 Antwortkategorien	**sign. Ergebnisse**	**Abbrüche**	**sign. Ergebnisse**	**Abbrüche**	**sign. Ergebnisse**	**Abbrüche**
mehrkategorielle Daten	5,06 %	0	5,26 %	0	5,82 %	0
dichotome Daten TK < 1	5,07 %	0	15,45 %	0	15,34 %	0
dichotome Daten TK < 2	5,24 %	0	2,55 %	0	2,66 %	0
dichotome Daten TK < 4	5,76 %	0	2,62 %	0	2,44 %	0
6 Antwortkategorien	**sign. Ergebnisse**	**Abbrüche**	**sign. Ergebnisse**	**Abbrüche**	**sign. Ergebnisse**	**Abbrüche**
mehrkategorielle Daten	4,76 %	0	5,71 %	0	6,46 %	0
dichotome Daten TK < 2	5,16 %	0	3,37 %	0	3,36 %	0
dichotome Daten TK < 3	4,89 %	0	2,27 %	0	2,41 %	0
dichotome Daten TK < 4	5,3 %	0	2,58 %	0	2,63 %	0

Wie man Tabelle 13 und 14 entnehmen kann, liegen die signifikanten Ergebnisse der M_2^*-Statistik durchgehend zwischen 4,79 % und 5,76 % – ein sehr einheitliches Ergebnis, welches impliziert, dass es keine Unterschiede in der Modellgeltung zwischen den Modellen für mehrkategorielle und dichotome Daten gibt.

Die Ergebnisse der χ^2-Statistik und der G^2-Statistik ähneln sich sehr. Bei einer Stichprobengröße von 1000 liefern die beiden Statistiken sehr kontroverse Ergebnisse. Das Risiko 1. Art, das bei der Modellierung der mehrkategoriellen Daten bei 5 % liegen sollte, kann in keinem Simulationsszenarium eingehalten werden. Die Anzahl signifikanter Ergebnisse bei der mehrkategoriellen Modellierung schwankt zwischen 6 % und 13 %. Die Anzahl der signifikanten Ergebnisse bei den oberen Teilungskriterien hingegen liegt durchgehend um die 2,5 %. Im Unterschied dazu liegt der Anteil signifikanter Ergebnisse bei den unteren Teilungskriterien zwischen 27,22 % und 70,04 %.

Bei einer Stichprobengröße von 2000 nähern sich die Prozentzahlen einander an und liegen bis auf ein paar Ausnahmen unter 6 %. Auffallend ist auch hier, dass die Anzahl signifikanter Ergebnisse bei den Modellen für dichotomes Antwortformat geringer ist als bei den Modellen für die ursprünglichen Daten.

4.3 Simulation zum Modell für kontinuierliche Antwortskalen nach Müller

In den Tabellen 15–21 sind die Ergebnisse der Simulationsstudie zum Modell für kontinuierliche Antwortskalen nach Müller (CRSM) angeführt. Die kontinuierlichen Daten, die Werte zwischen 0 und 1 annehmen konnten, wurden einmal gemäß ihrem Mittelwert, einmal gemäß den empirischen und einmal gemäß den theoretischen Quartilen unterteilt. Das heißt, bei dieser Simulationsstudie wurde nicht nur dichotomisiert, sondern auch mehrkategoriell unterteilt. Einschränkungen liefert diese Simulationsstudie dahingehend, dass die Modellgeltung des CRSM nicht überprüfbar war. Modellgeltungstests für das CRSM wären zwar grundsätzlich denkbar, wurden aber noch nicht implementiert. Die Modellgeltung der Modelle für mehrkategorielle (PCM) und dichotome Antwortformate (Rasch-Modell) wurde mittels Likelihood-Ratio-Test und Fit-Indizes überprüft. In Tabelle 15 sind deskriptive Statistiken zu den Korrelationen der Personenparameterschätzungen des CRSM und des PCM beziehungsweise des Rasch-Modells angeführt.

Tabelle 15: Simulation zum CRSM - Deskriptive Statistiken zu den Korrelationen der Personenparameterschätzungen (N=1000)

STICHPROBENGRÖSSE 1000 & 20 AUFGABEN

	Dichotom: Mittelwert	Mehrkategoriell: empirische Quartile	Mehrkategoriell: theoretische Quartile
Minimum	0,8802	0,9694	0,9727
25%-Quantil	0,9003	0,9755	0,9769
Median	0,9041	0,9765	0,9778
Mittelwert	0,9039	0,9764	0,9778
75%-Quantil	0,9078	0,9774	0,9787
Maximum	0,9232	0,9811	0,9826
Standardabweichung	0,00556415	0,00143556	0,00134666

Vergleicht man die Ergebnisse aus Tabelle 15 mit denen aus Tabelle 2 (Korrelationen des PCM), weisen die Korrelationen der Personenparameterschätzungen des CRSM mit dem Rasch-Modell die niedrigsten Korrelationskoeffizienten mit der größten Streuung auf. Betrachtet man die Korrelationen der Personenparameter zwischen dem CRSM und den PCMs, steigen diese deutlich an. Zwischen den zwei Partial-Credit-Modellen, bei welchen eines gemäß der empirischen und das andere gemäß der theoretischen Quartile unterteilt wurde, sind keine großen Unterschiede ersichtlich. Die Korrelationskoeffizienten beider Modelle bewegen sich in einem Bereich von circa 0,97 bis 0,98 und weisen demnach auch eine geringe Streuung auf.

In Tabelle 16 sind die signifikanten Ergebnisse des Likelihood-Ratio-Tests angegeben. Auch hier kam es zu keinen Abbrüchen, das heißt, für alle 10000 Durchläufe konnte der LRT berechnet werden.

Tabelle 16: Simulation zum CRSM - Likelihood-Ratio-Test (N=1000)

STICHPROBENGRÖSSE 1000 & 20 AUFGABEN

	Signifikante Ergebnisse	**Abbrüche**
Dichotom: Mittelwert	5 %	0
Mehrkategoriell: Empirische Quartile	5,84 %	0
Mehrkategoriell: Theoretische Quartile	5,36 %	0

Bei allen Teilungskriterien zeigte der LRT in 5 % – 5,84 % der Durchläufe eine Modellverletzung auf. Vergleicht man dazu Tabelle 5 (LRT zur Simulation des PCM), ergeben sich weitaus höhere Prozentzahlen. Nimmt man das Simulationsszenario mit einer Stichprobengröße von 1000 und 20 Aufgaben und betrachtet nur die mittleren Teilungskriterien, ergeben sich trotzdem 12,14 % bis 46,2 % signifikante Ergebnisse.

Tabelle 17: Simulation zum CRSM - Item- & Personenfit-Indizes (N=1000)

STICHPROBENGRÖSSE 1000 & 20 AUFGABEN

	Anzahl nicht modellkonformer Aufgaben				
	0	**1**			
Mittelwert	99,98 %	0,02 %			
Empirische Quartile	100 %	0			
Theoretische Quartile	98,42 %	1,58 %			
	Anzahl nicht modellkonformer Personen				
	0	**{1-5}**	**{6-10}**	**{11-15}**	**{16-21}**
Mittelwert	2,15 %	79,71 %	17,98 %	0,16 %	0
Empirische Quartile	0	12,16 %	62,54 %	23,82 %	1,48 %
Theoretische Quartile	80,09 %	19,91 %	0	0	0

Bei der Überprüfung des Rasch-Modells (durch die Personenfit-Indizes), welches zur Modellierung der dichotomen Daten (Teilungskriterium: Mittelwert) diente, wurde in 2,15 % der Testdurchläufe keine Person als auffällig klassifiziert. In 79,71 % der Fälle wurde(n) eine bis fünf Personen, in 17,98 % der Durchläufe 6 bis 10 Personen und in 0,16 % der Durchläufe 11 bis 15 Personen als nicht modellkonform klassifiziert.

5 Diskussion

Im Zuge dieser Arbeit wurden die Auswirkungen einer nachträglichen Dichotomisierung auf die Personenparameterschätzung und die Modellgeltung im Rahmen der Item-Response-Theorie von der empirischen Seite betrachtet.

Die Vergleichbarkeit der Personenparameterschätzungen wurde mithilfe von Korrelationen zwischen den Schätzungen der ursprünglichen und nachträglich zusammengelegten Daten untersucht. Es wurde sowohl die Gleichheit der Personenparameterschätzungen (Produkt-Moment-Korrelation) als auch die Rangreihung (Rangkorrelation) ebendieser betrachtet. Es zeigte sich, dass sich die Korrelationskoeffizienten der Produkt-Moment-Korrelation und der Rangkorrelation nur minimal unterscheiden. Diese waren zwar durchgehend über 0,9, erreichten aber einen Korrelationskoeffizienten von eins nicht. Geht man von der Äquivalenzannahme zweier Modelle aus, ist dieses Ergebnis nicht zufriedenstellend, da in diesem Fall das Ergebnis von eins für den Korrelationskoeffizienten wünschenswert gewesen wäre.

Dies impliziert, dass ein nachträgliches Dichotomisieren mehrkategorieller Daten die Personenparameterschätzung verzerrt. Vergleicht man die deskriptiven Statistiken zu den Korrelationskoeffizienten des PCM (Tabelle 2) mit denen des GRM (Tabelle 11), erkennt man, dass die Korrelationskoeffizienten des GRM über die Teilungskriterien hinweg konstanter bleiben und höher sind. Dennoch wären auch beim GRM höhere Korrelationskoeffizienten zu erwarten gewesen.

An dieser Stelle sei noch angemerkt, dass die Personenparameter berechnet, aber nicht gespeichert wurden und daher die Rangkorrelationen nur für das PCM und GRM (N=1000) berechnet wurden.

Die Modellgeltung wurde bei den Modellen der Rasch-Familie mit dem Likelihood-Ratio-Test und den Fit-Indizes geprüft. Wie in Tabelle 4 und 5 (Ergebnisse des LRT) angeführt, bewegt sich der Anteil signifikanter Ergebnisse bei der mehrkategoriellen Modellierung zwischen 4,09 % und 5,45 %. Der Anteil signifikanter Ergebnisse bei der dichotomen Modellierung fiel weit höher aus. Bei den mittigen Teilungskriterien lag dieser zwischen 12,14 % und 46,2 % bei einer Stichprobengröße von 1000 und zwischen 32,28 % und 83,4 % bei einer Stichprobengröße von 2000. Bei den unteren und oberen Teilungskriterien kam es in den meisten Fällen zu 98 % – 100 % signifikanten Ergebnissen.

Dies legt die Hypothese nahe, dass die künstliche Teilung des Datensatzes dazu geführt hat, dass der Test für viele Personen zu schwer oder zu leicht war und dass dieser möglicherweise in den oberen und unteren Fähigkeitsbereichen nicht gut genug differenziert.

Im Folgenden werden die Person-Item-Maps für einen Datensatz (1000 Personen, 20 Aufgaben und vier Antwortmöglichkeiten) ausgegeben. An dieser Stelle sei erwähnt, dass die folgenden Darstellungen nicht für alle Datensätze gelten, da zufällig ein Datensatz ausgewählt wurde. Dies soll rein zur Darstellung der oben angeführten Problematik dienen. Der LRT zeigte für die mehrkategorielle Modellierung mit dem Partial-Credit-Modell und die dichotome Modellierung mit dem Rasch-Modell (mittleres Teilungskriterium: $TK < 2$) keine Modellverletzung auf. Wie in den meisten Fällen, kam es jedoch bei dem unteren und oberen Teilungskriterium zu einem signifikanten Ergebnis.

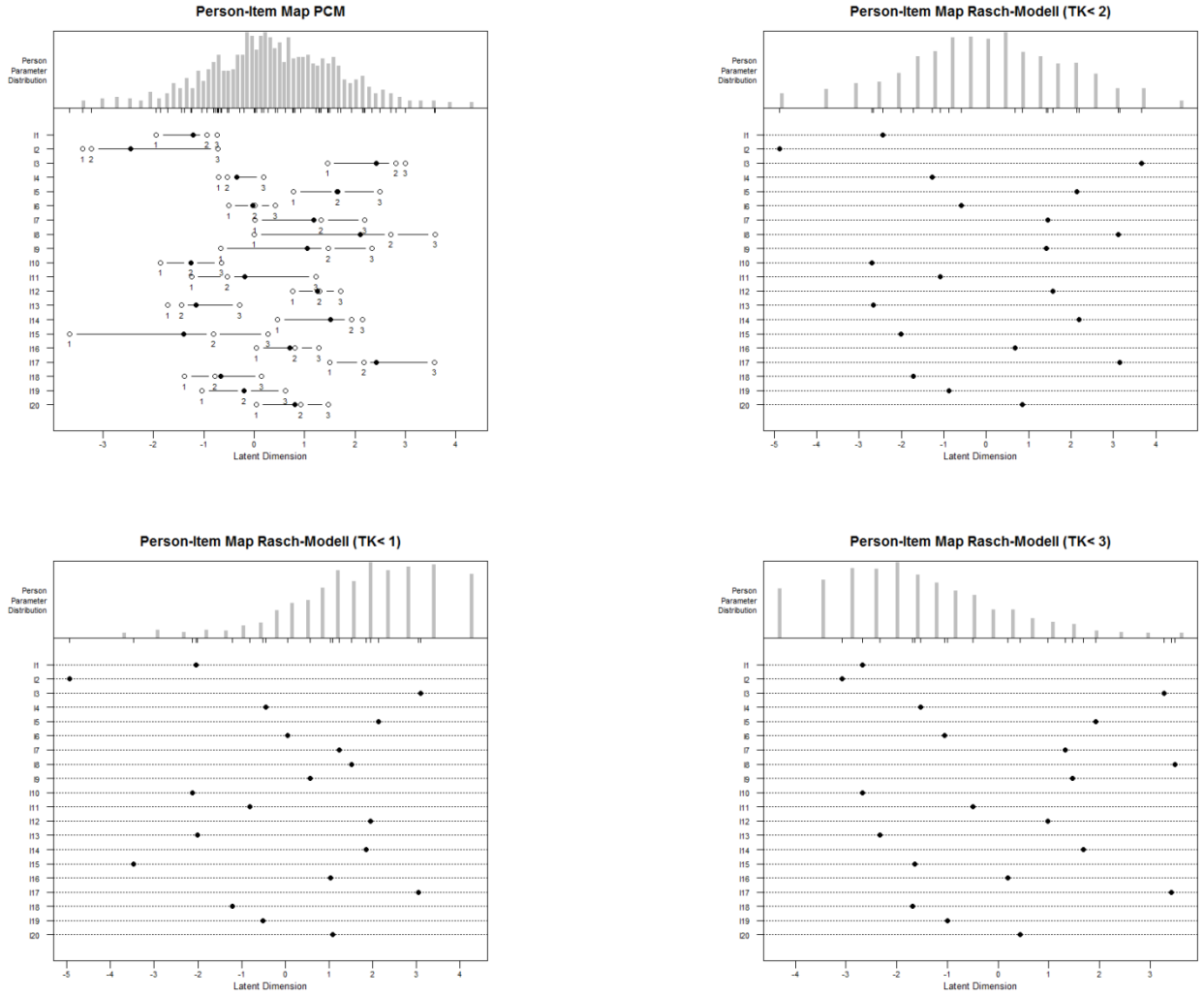


Abbildung 5: Simulation zum PCM - Person-Item-Maps (N=1000, 20 Aufgaben & 4 Kategorien)

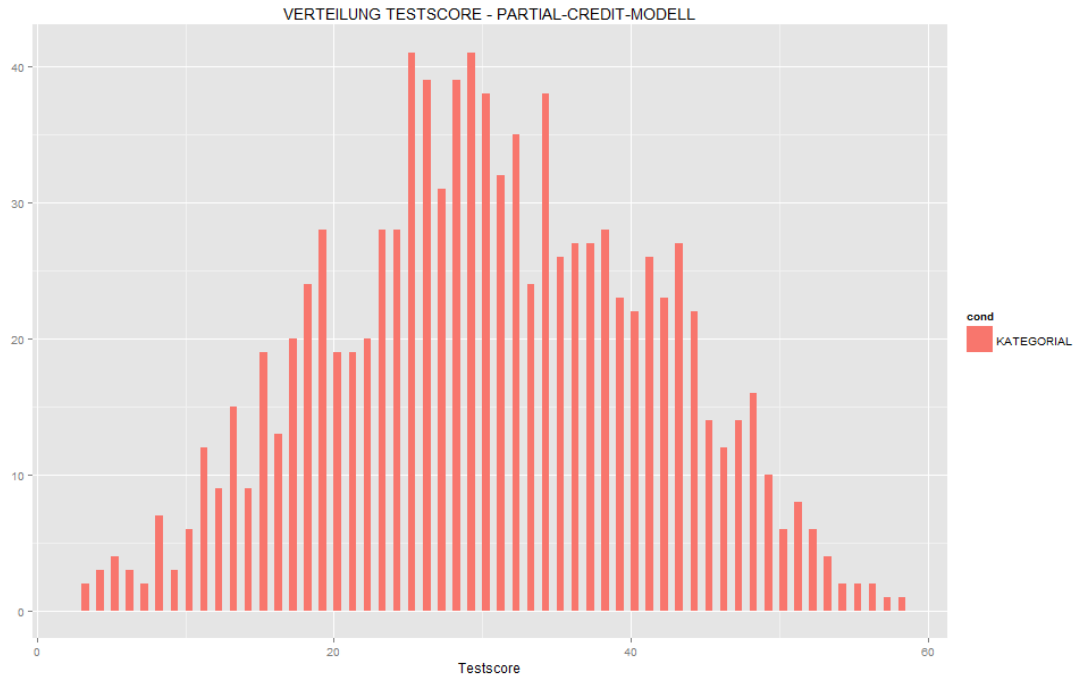
Im oberen Teil jeder Graphik ist die Verteilung der Personenparameter angegeben. Die Personenparameter des PCM und des Rasch-Modells (*TK < 2*) sind gleichmäßig verteilt. Die meisten Personen haben einen mittleren Testscore erzielt, die Enden (unteren und oberen Fähigkeitsbereiche) sind dünner besetzt. Bei den unteren beiden Graphiken (*TK < 1* und *TK < 3*) sieht man deutlich eine links- und rechtsschiefe Verteilung. Bei dem unteren Teilungskriterium haben sehr viele Personen einen hohen Testscore erzielt, das heißt, der Test war zu leicht und differenziert zwischen den oberen Fähigkeitsbereichen nicht ausreichend. Ein ähnliches Bild in die komplementäre Richtung erkennt man bei dem oberen Teilungskriterium.

Zur besseren Veranschaulichung sind in den folgenden Graphiken (Abbildung 6) die Verteilungen der Testscores angeführt. Zum Vergleich sind die drei Modellierungen des Rasch-Modells in einer Graphik dargestellt. Auf der *X*-Achse sind die Testscores angeführt, im Falle des Rasch-Modells ist das die Summe der Aufgaben, denen eine Person zugestimmt hat. Beim Partial-Credit-Modell ist der maximale Testscore 60, da es vier Kategorien (0, 1, 2, 3) und 20 Aufgaben gibt.

Zudem wurden die Testscores der dichotomen und mehrkategoriellen Datenmatrizen des Modells für kontinuierliche Antwortskalen berechnet und ebenfalls graphisch dargestellt. Alle Ergebnisse des LRT für die nachträglich dichotomisierten und mehrkategoriell unterteilten Daten waren nicht signifikant.

Vergleicht man die Testscore-Verteilungen der Simulation zum PCM mit denen der Simulation zum CRSM, sind diese beim CRSM (Abbildung 7) etwas steiler und die Enden sind nicht besetzt. Daraus folgt, dass es mehr Personen gibt, die einen mittleren Testscore erzielt haben und keine Personen, die keine oder alle Aufgaben lösen konnten. Auch hier sei betont, dass es sich lediglich um eine von 10000 Datenmatrizen handelt und diese Ergebnisse nur der Veranschaulichung dienen.

Simulation zum PCM - Mehrkategorielle Modellierung



Simulation zum PCM - Dichotome Modellierung

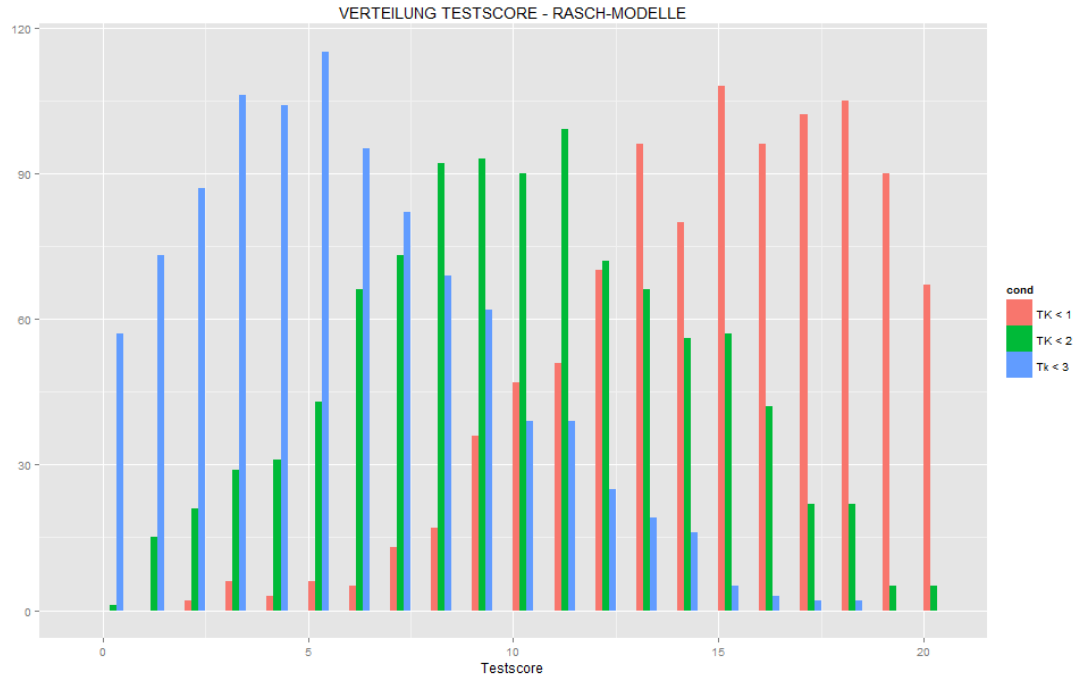
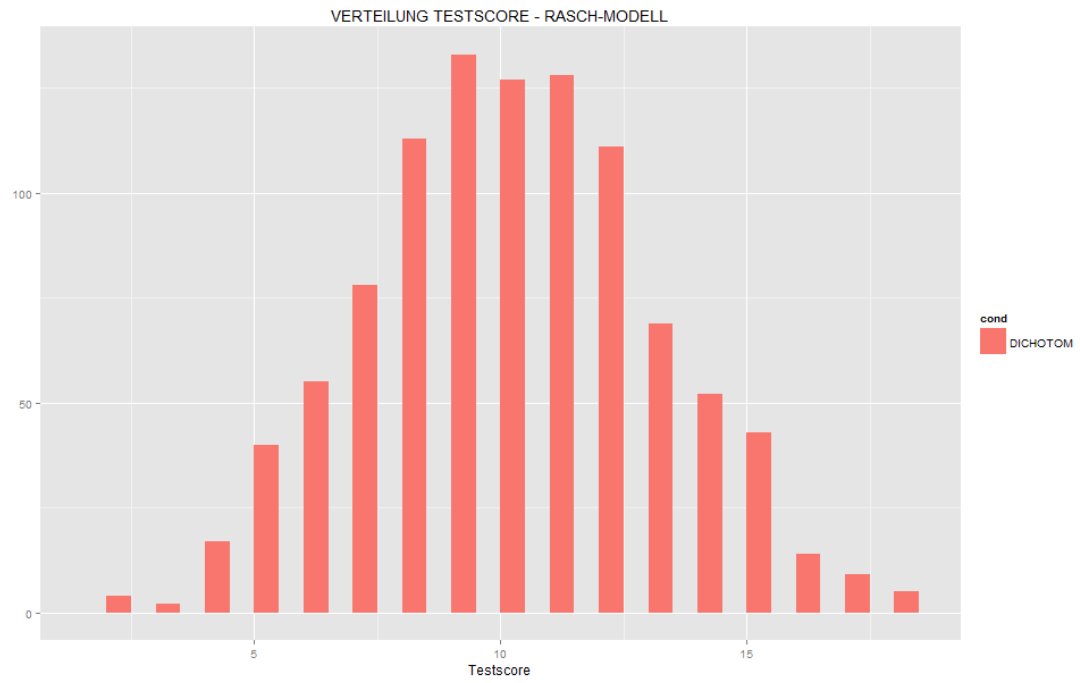


Abbildung 6: Simulation zum PCM - Verteilungen der Testscores

Simulation zum CRSM - Dichotome Modellierung



Simulation zum CRSM - Mehrkategorielle Modellierung

Abbildung 7: Simulation zum CRSM - Verteilungen der Testscores

Vergleicht man die Fit-Indizes der dichotomen und mehrkategoriellen Modellierung des PCM, ergeben sich mehr signifikante Ergebnisse bei der dichotomen Modellierung. Doch auch die 5 % „erlaubten“ signifikanten Ergebnisse werden bei den mehrkategoriellen Daten, die gemäß dem PCM generiert wurden, nicht eingehalten. Eine Erklärung dafür könnte sein, dass die beobachtete Antwort immer ganzzahlig und die erwartete Antwort meist eine Dezimalzahl ist. Den Residuen-basierten Maßen liegt bekanntlich die Differenz der beobachteten und erwarteten Antworten zugrunde. Durch diesen Umstand liegt die Vermutung nahe, dass die Fit-Indizes Modellabweichungen eher überschätzen (Rost, 2004).

Eine weitere mögliche Erklärung wäre, dass andere Fit-Indizes geeigneter gewesen wären. Wie in Kapitel 2.4 (Residuen-basierte Personen und Itemfit-Indizes) schon erwähnt, ist das vorgestellte Fitmaß nur eines mehrerer Fitmaße. Einen guten Überblick zu den verschiedenen Fit-Indizes liefern Glas und Verhelst (1995 a und b). Zur Frage der Wahl des geeigneten Fitmaßes schrieben Christensen, Kreiner und Mesbah (2013, S. 101): *„We would, of course, always search for the most powerful statistics, but there are few studies comparing the power of fit statistics in a systematic and comprehensive way. For what it is worth, we can inform the reader that we have performed a number of (unpublished) Monte Carlo Studies without finding indisputable evidence that some of the fit statistics are more powerful than the others to a degree that makes a difference in practice.“* Zukünftige Simulationsstudien könnten der Frage des am besten geeigneten Fit-Maßes nachgehen und neuere Entwicklungen vergleichen, wie zum Beispiel die *Bootstrap Item Fit Indices* von Wolfe und McGill (2011).

Die in dieser Studie verwendeten Fit-Indizes leisten möglicherweise einen guten Beitrag zur Identifizierung nicht modellkonformer Aufgaben und Personen bei der Überprüfung eines neu entwickelten Tests. Zur Überprüfung der Modellgeltung in einer Simulationsstudie sollte auf diese allerdings in Zukunft verzichtet werden.

Die Ergebnisse zur Überprüfung der Modellgeltung des GRM bestätigen die Ergebnisse von Cai & Hansen (2013) und Maydeu-Olivares & Joe (2006). Das Risiko 1. Art für die mehrkategorielle Modellierung wird nur von der M_2^*-Statistik eingehalten. Die χ^2-Statistik und die G^2-Statistik scheinen nicht dazu geeignet, die Modellgeltung von Modellen mit spärlich besetzten Kontingenztabellen zu prüfen. Betrachtet man die Ergebnisse der M_2^*-Statistik bei der dichotomen Modellierung, scheint ein

nachträgliches Zusammenlegen von Antwortmöglichkeiten keinen Einfluss auf die Modellgeltung zu haben.

Zusammenfassend lässt sich sagen, dass ein nachträgliches Dichotomisieren mehrkategorieller Daten der Rasch-Familie Auswirkungen auf die Parameterschätzung und die Modellgeltung hat. Diese scheinen geringer bei dem Modell für kontinuierliche Antwortskalen von Müller, sind aber in Anbetracht der Verzerrung der Personenparameterschätzungen nicht vertretbar. Das Graded-Response-Modell ist wie erwartet weniger „empfindlich“ gegenüber einer nachträglichen Dichotomisierung. Dennoch sollte auch hier auf ein nachträgliches Zusammenlegen von Antwortmöglichkeiten verzichtet werden, da es zu Verzerrungen bei der Personenparameterschätzung kommen kann. Für die Praxis wird angeraten, mögliche Probleme - zum Beispiel aufgrund einer zu geringen Antwortdichte - bei der Parameterschätzung zu vermeiden, in dem man Pretests vorgibt und die Stichprobengröße adäquat wählt.

Eine Limitation der vorliegenden Studie ist sicherlich die Nicht-Überprüfbarkeit der Modellgeltung des Modells für kontinuierliche Antwortskalen von Müller. Zukünftige Forschungsarbeiten könnten sich mit der Implementierung eines Modelltests beschäftigen. In weiteren Arbeiten wäre es interessant, eine Möglichkeit zu schaffen, die verschiedenen Modelle und Modelltests zu vergleichen. In der vorliegenden Arbeit wurden mehrere Simulationsszenarien zu drei Modellen (PCM, GRM und CRSM) durchgeführt. Die resultierten Ergebnisse können nicht auf andere Modelle oder Simulationsszenarien generalisiert werden. Im Prinzip gäbe es beliebig viele Simualtionsszenarien und eine Reihe von weiteren Modellen und Modelltests, die im Bezug auf eine nachträgliche Dichotomisierung verglichen werden könnten. Interessant wäre es beispielsweise noch, die Auswirkungen einer nachträglichen Dichotomisierung bei dem Modell für kontinuierliche Antwortskalen von Samejima (1973) zu untersuchen. Diese Maßnahmen würden nicht nur zusätzliche wissenschaftliche Maßstäbe setzen, sondern auch ein Bewusstsein für die Auswirkungen einer Datenmanipulation schaffen.

6 Verzeichnisse

6.1 Literaturverzeichnis

Andersen, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika, 38*, 123-140.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43(4)*. 357-374.

Andrich, D. (1995). Models for measurement, precision, and the nondichotomization of graded responses. *Psychometrika. 60*, 7-26.

Cai, L. & Hansen, M. (2013). Limited-information Goodness-of-fit Testing of Hierarchical Item Factor Models. *Br J Math Stat Psychol., 66(2)*, 245-276.

Chalmers, P. (2012). *mirt: Multidimensional Item Response*. R package version 1.10. https://cran.r-project.org/web/packages/mirt/index.html

Christensen, K., B. Kreiner, S. & Mesbah, M. (2013). *Rasch Models in Health*. Great Britain: ISTE Ltd and John Wiley & Sons, Inc.

Fischer, G. H. (1974). *Einführung in die Theorie Psychologischer Tests Grundlagen und Anwendungen*. Bern: Hans Huber.

Glas, C. A. W. & Verhelst N. D. (1995a). Testing the Rasch model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch Models: foundations, recent developements, and applications* (pp. 69-95). New York: Springer.

Glas, C. A. W. & Verhelst N. D. (1995b). Testing the fit for polytomous Rasch model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch Models: foundations, recent developements, and applications* (pp. 325-352). New York: Springer.

Hohensinn, C. (2013). *pcIRT: IRT models for polytomous and continuous item responses*. R package version 0.1. https://cran.r-project.org/web/packages/pcIRT/index.html.

Jansen, P. G. W. & Roskam, E. E. (1986). Latent trait models and dichotomization of graded responses. *Psychometrika, 50*, 69-91.

Koller, I., Alexandrowicz, R. & Hatzinger, R. (2012). *Das Rasch-Modell in der Praxis: Eine Einführung mit eRm*. Wien: facultas.wuv.

Mair, P., Hatzinger, R. & Maier M. J. (2014). *eRm: Extended Rasch Modeling*. R package version 0.15-4.http://erm.r-forge.r-project.org/

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149-174.

Maydeu-Olivares A. (2015). Evaluating the Fit of IRT Models. In Reise, S. P. & Dennis A. R., *Handbook of Item Response Theory Modeling: Applications to typical performance assessment* (pp 111-127). Routledge: New York

Maydeu-Olivares, A. & Cai, L. (2006). A cautionary note on using G^2(dif) to assess relative model fit in categorical data analysis. *Multivariate Behavioral Research, 41(1)*, 55-64.

Maydeu-Olivares, A. & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contigency tables. *Psychometrika, 71*, 713-732.

Müller, H. (1987). A Rasch model for continuous ratings. *Psychometrika, 52*, 165-181.

Müller, H. (1995). Threshold discriminations, nondichotomization, and precision in a Rasch model for continuous responses: A comment on a debate between Roskam and Andrich (Arbeiten aus dem Insitut für Psychologie, Heft 4, 1995). Frankfurt am Main: Institut für Psychologie der Universität.

Müller, H. (1999). *Probabilistische Testmodelle für diskrete und kontinuierliche Ratingskalen: Einführung in die Item-Response-Theorie für abgestufte und kontinuierliche Items*. Bern: Verlag Hans Huber.

Nering, M. & Ostini R. (2010). *Handbook of polytomous item response theory models*. New York Routledge.

Rasch, D. & Guiard, V. (2004). The robustness of parametric statistical methods. *Psychology Science*, 46, 175-208.

Reif, M. (2014). *PP: Estimation of person parameters for the 1,2,3,4-PL model and the GPCM*. R package version 0.5.3. https://cran.r-project.org/web/packages/pcIRT/index.html

Roskam, E. E. (1995). Graded responses and joining categories: A rejoinder to Andrich's "models for measurement, precision, and nondichotomization of graded responses". *Psychometrika, 60*, 27-35.

Roskam, E. E. & Jansen, P. G. W. (1989). Conditions for Rasch-dichotomizability of the unidimenional polytomous Rasch model. *Psychometrika, 54*, 317-332.

R Development Core Team (2012). R: A language and enviroment for statistical computing. Vienna: R Foundation for Statistical Computing. http://www.R-project.org.

Rost, J. (1999). Was ist aus dem Rasch-Modell geworden? *Psychologische Rundschau, 50*, 140-156.

Rost, J. (2004). *Lehrbuch Testtheorie-Testkonstruktion*. Bern: Verlag Hans Huber.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement, 34 (4)*.

Samejima, F. (1973). Homogeneous case of the continuous response model. *Psychometrika, 38*, 203-209.

Strobl, C. (2012). *Das Rasch-Modell: Eine verständliche Einführung für Studium und Praxis*. München: Rainer Hampp Verlag.

Tierney, L., Rossini, A. J., Li, N. & Sevcikova, H. (2012). *snow: Simple Network of Workstations*. R package version 0.3-10. http://CRAN.R-project.org/package=snow.

Thissen, D., & Steinberg, L. (1997). A response model for multiple-choice items. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 51–66). New York: Springer Verlag.

Van Wyke, J. & Andrich, D. (2006). *A typology of polytomously scored mathematic items disclosed by the Rasch model: Implication for constructing a continuum of achievement* (Research Paper No. 2). Australian Research Counsil Linkage Grant LP0454080

Wright, B. D., & Masters, G. N. (1990). Computation of OUTFIT and INFIT Statistics. *Rasch Measurement Transactions*, 3, 84-85.

Wolfe, E. W., & McGill, M. T. (2011). *Comparison of Asymptotic and Bootstrap Item Fit Indices in Identifying Misfit to the Rasch Model*. National Conference on Measurement in Education, New Orleans, LA. URL: http://images.pearsonassessments.com/images/PDF/NCME-Asymptotic-Bootstrap-Indices.pdf (06.11.2015)

6.2 Tabellenverzeichnis

6.3 Abbildungsverzeichnis

6.4 R-Code

```
#++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
# SIMULATION ZUM PARTIAL-CREDIT-MODELL
#++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
#----------------------------------------------------------------------------------------------
# DICHOTOMISIEREN VON PCM-DATEN (N=1000)
#----------------------------------------------------------------------------------------------
# PACKAGES
#----------------------------------------------------------------------------------------------
library(PP)
library(eRm)
library(snow)
#----------------------------------------------------------------------------------------------
# DATENGENERIERUNG
#----------------------------------------------------------------------------------------------
create_data <- function(thres=THRES,size=n,items=it,wh=0,kat=0){
  v <- 1:wh
  sl <- rep(1,it)
  THRESx <- rbind(0,THRES)
  THETA <- rnorm(n)
  for(i in 1:wh){
    simdat_gpcm <- sim_gpcm(thres = THRESx,alpha = sl,theta = THETA)

write.table(simdat_gpcm,file=paste("~/Simulation_2015_PCM/results/pcm_",n,"_",it,"_",
            kat,"_daten","/",v[i],sep=""))
  }
}
#----------------------------------------------------------------------------------------------
# DICHOTOMISIERUNGSFUNKTION
#----------------------------------------------------------------------------------------------
dich.func.pcm <-function(g,n,it,kat){
  require(eRm)
#----------------------------------------------------------------------------------------------
  # Dichotomisierungsfunktion
  dich.pcm <- function(m){
    res <- NULL
    if(((max(m)+1)%%2)==0) {
      if(max(m)==3){
        dich.pcm.u <- ifelse(m < min(m)+1,0,1)
        dich.pcm.m <- ifelse(m <(max(m)+1)/2,0,1)
        dich.pcm.o <- ifelse(m < max(m),0,1)
        res <- list(dich.pcm.u,dich.pcm.m,dich.pcm.o)}
      if(max(m)==5){
        dich.pcm.u <- ifelse(m < min(m)+2,0,1)
        dich.pcm.m <- ifelse(m <(max(m)+1)/2,0,1)
        dich.pcm.o <- ifelse(m < max(m)-1,0,1)
        res <- list(dich.pcm.u,dich.pcm.m,dich.pcm.o)}
    }
    if(((max(m)+1)%%2)!=0) {
      if(max(m)==4){
        dich.pcm.u <- ifelse(m < min(m)+1,0,1)
        dich.pcm.m <- ifelse(m < max(m)/2,0,1)
        dich.pcm.o <- ifelse(m < max(m),0,1)
        res <- list(dich.pcm.u,dich.pcm.m,dich.pcm.o)}
      if(max(m)==6){
        dich.pcm.u <- ifelse(m < min(m)+2,0,1)
        dich.pcm.m <- ifelse(m < max(m)/2,0,1)
        dich.pcm.o <- ifelse(m < max(m)-1,0,1)
        res <- list(dich.pcm.u,dich.pcm.m,dich.pcm.o)}
    }
    return(res)
  }
```

```
#-------------------------------------------------------------------------------------------
  erg <-read.table(paste("~/Simulation_2015_PCM/results/pcm_",
                   n,"_",it,"_",kat,"_daten/",g,sep=""))
  rmprobe <- dich.pcm(erg)

  # Modellberechnung + Personenparameter PCM
  estpcm <- try(PCM(erg))
  perspcm <- try(person.parameter(estpcm))

  # Modellberechnung + Personenparameter RM nach Teilungskriterium
  estdich_u <- try(RM(rmprobe[[1]]))
  persdich_u <- try(person.parameter(estdich_u))
  estdich_m <- try(RM(rmprobe[[2]]))
  persdich_m <- try(person.parameter(estdich_m))
  estdich_o <- try(RM(rmprobe[[3]]))
  persdich_o <- try(person.parameter(estdich_o))

  # Korrelationen
  perspcmdat <- as.numeric(try(unlist(perspcm[[13]][1])))
  persdichdat_u <- as.numeric(try(unlist(persdich_u[[13]][1])))
  persdichdat_m <- as.numeric(try(unlist(persdich_m[[13]][1])))
  persdichdat_o <- as.numeric(try(unlist(persdich_o[[13]][1])))

  a_u <- as.numeric(try(cor(perspcmdat, persdichdat_u,use="na.or.complete")))
  a_m <- as.numeric(try(cor(perspcmdat, persdichdat_m,use="na.or.complete")))
  a_o <- as.numeric(try(cor(perspcmdat, persdichdat_o,use="na.or.complete")))

  # Likelihood-Ratio
  lrtpcm <- try(LRtest(estpcm, splitcr="median"))
  b <- as.numeric(try(lrtpcm$pvalue))

  lrtdich_u <- try(LRtest(estdich_u, splitcr="median"))
  b_u <- as.numeric(try(lrtdich_u$pvalue))

  lrtdich_m <- try(LRtest(estdich_m, splitcr="median"))
  b_m <- as.numeric(try(lrtdich_m$pvalue))

  lrtdich_o <- try(LRtest(estdich_o, splitcr="median"))
  b_o <- as.numeric(try(lrtdich_o$pvalue))

  # Itemfit
  itemfit <- eRm::itemfit
  IFIpcm <- try(itemfit(perspcm))
  IFIdich_u <- try(itemfit(persdich_u))
  IFIdich_m <- try(itemfit(persdich_m))
  IFIdich_o <- try(itemfit(persdich_o))

  alpha <- 0.05
  ipvalvecpcm <- try(1-pchisq(IFIpcm$i.fit, IFIpcm$i.df))
  ipvalsigpcm <- try(which(ipvalvecpcm < alpha))
  inumberpcm <- try(length(ipvalsigpcm))
  d<-inumberpcm

  ipvalvecdich_u <- try(1-pchisq(IFIdich_u$i.fit, IFIdich_u$i.df))
  ipvalsigdich_u <- try(which(ipvalvecdich_u < alpha))
  inumberdich_u <- try(length(ipvalsigdich_u))
  d_u<-inumberdich_u

  ipvalvecdich_m <- try(1-pchisq(IFIdich_m$i.fit, IFIdich_m$i.df))
  ipvalsigdich_m <- try(which(ipvalvecdich_m < alpha))
  inumberdich_m <- try(length(ipvalsigdich_m))
  d_m<-inumberdich_m

  ipvalvecdich_o <- try(1-pchisq(IFIdich_o$i.fit, IFIdich_o$i.df))
  ipvalsigdich_o <- try(which(ipvalvecdich_o < alpha))
  inumberdich_o <- try(length(ipvalsigdich_o))
  d_o<-inumberdich_o

  # Personenfit
  personfit <- eRm::personfit
  PFIpcm <- try(personfit(perspcm))
  PFIdich_u <- try(personfit(persdich_u))
  PFIdich_m <- try(personfit(persdich_m))
  PFIdich_o <- try(personfit(persdich_o))
```

```
  ppvalvecpcm <- try(1-pchisq(PFIpcm$p.fit, PFIpcm$p.df))
  ppvalsigpcm <- try(which(ppvalvecpcm < alpha))
  pnumberpcm <- try(length(ppvalsigpcm))
  e<-pnumberpcm

  ppvalvecdich_u <- try(1-pchisq(PFIdich_u$p.fit, PFIdich_u$p.df))
  ppvalsigdich_u <- try(which(ppvalvecdich_u < alpha))
  pnumberdich_u <- try(length(ppvalsigdich_u))
  e_u<-pnumberdich_u

  ppvalvecdich_m <- try(1-pchisq(PFIdich_m$p.fit, PFIdich_m$p.df))
  ppvalsigdich_m <- try(which(ppvalvecdich_m < alpha))
  pnumberdich_m <- try(length(ppvalsigdich_m))
  e_m<-pnumberdich_m

  ppvalvecdich_o <- try(1-pchisq(PFIdich_o$p.fit, PFIdich_o$p.df))
  ppvalsigdich_o <- try(which(ppvalvecdich_o < alpha))
  pnumberdich_o <- try(length(ppvalsigdich_o))
  e_o<-pnumberdich_o

  inf.krit.pcm <- try(IC(perspcm))
  aicpcm <- try(inf.krit.pcm[[1]][7])
  bicpcm <- try(inf.krit.pcm[[1]][10])
  f<-aicpcm; h<-bicpcm

  inf.krit.dich_u <- try(IC(persdich_u))
  aicdich_u <- try(inf.krit.dich_u[[1]][7])
  bicdich_u <- try(inf.krit.dich_u[[1]][10])
  f_u<-aicdich_u; h_u<-bicdich_u

  inf.krit.dich_m <- try(IC(persdich_m))
  aicdich_m <- try(inf.krit.dich_m[[1]][7])
  bicdich_m <- try(inf.krit.dich_m[[1]][10])
  f_m<-aicdich_m; h_m<-bicdich_m

  inf.krit.dich_o <- try(IC(persdich_o))
  aicdich_o <- try(inf.krit.dich_o[[1]][7])
  bicdich_o <- try(inf.krit.dich_o[[1]][10])
  f_o<-aicdich_o; h_o<-bicdich_o

  dt.pcm.dich <- c(a_u,a_m,a_o,b,b_u,b_m,b_o,d,d_u,d_m,d_o,
                   e,e_u,e_m,e_o,f,f_u,f_m,f_o,h,h_u,h_m,h_o)

  write.table(dt.pcm.dich, file =paste("~/Simulation_2015_PCM/
              results/pcm_",n,"_",it,"_",kat,"_res/",g,sep=""))
}

#-------------------------------------------------------------------------------------------
# VERZEICHNISERSTELLUNG
#-------------------------------------------------------------------------------------------
dir.create("~/Simulation_2015_PCM/results/pcm_1000_20_4_daten")
dir.create("~/Simulation_2015_PCM/results/pcm_1000_20_4_res")
dir.create("~/Simulation_2015_PCM/results/pcm_1000_20_5_daten")
dir.create("~/Simulation_2015_PCM/results/pcm_1000_20_5_res")
dir.create("~/Simulation_2015_PCM/results/pcm_1000_20_6_daten")
dir.create("~/Simulation_2015_PCM/results/pcm_1000_20_6_res")

#-------------------------------------------------------------------------------------------
# DATENGENERIERUNG & SIMULATION 10000
#-------------------------------------------------------------------------------------------
# Stichprobengröße 1000; 20 Items; 4 Kategorien
THRES <- matrix(c(-2,-1.23,-1,-4.2,-3.48,-1.11,1.2,2.5,3,-1,-0.8,-0.2,0.5,1.3,2,
                  -0.8,-0.4,0.2,-0.2,1,1.9,-0.2,2.4,3,-1,1.2,2,-2.1,-1.6,-0.88,
                  -1.5,-1,1,0.4,0.9,1.3,-2.12,-1.8,-0.5,0.2,1.2,2,-3.23,-1.2,0,
                  -0.2,0.4,1,1.2,2,3.12,-1.8,-1,-0.2,-1.2,-0.6,0.3,-0.33,0.66,1),nrow=3)
n <- 1000 ; it <- 20 ; wh <- 10000

create_data(thres=THRES,size=n,items=it,wh=10000,kat=4)

daten <- list.files("~/Simulation_2015_PCM/results/pcm_1000_20_4_daten")
kern <- makeSOCKcluster(rep("localhost", 4))
parLapply(kern, daten, fun=dich.func.pcm,n=1000,it=20,kat=4)
stopCluster(kern)
#-------------------------------------------------------------------------------------------
# Stichprobengröße 1000; 20 Items; 5 Kategorien
```

```
THRES <-matrix(c(-2.4,-1.78,-1.23,-0.2,-3.7,-3.48,-2.11,-1,1.2,2.5,3,3.5,-1.9,-1.2,-
0.9,0.3,0.5,1.3,2,2.4,-0.8,-0.4,0.2,0.4,-0.2,1,1.9,2.3,-0.2,2.4,3,3.3,-2,-1.2, -0.8,-
0.2,-2.3,-1.7,-0.9,-0.6,-1.4,-1,-0.23,0.2,-4.2,-4,-3.31,-2.19,2.2,2.9,3,3.5,-1.7,-1.2,-
0.7,0.2,1,1.8,2.23,2.9,-0.6,-0.1,0.5,1,1,2,3,4,-2,-1.2,-0.8,0.1,-2.6,-2.1,-1.3,-0.9,-
1.5,-1,1,1.5),nrow=4)
n <- 1000 ; it <- 20 ; wh <- 10000

create_data(thres=THRES,size=n,items=it,wh=10000,kat=5)

daten <- list.files("~/Simulation_2015_PCM/results/pcm_1000_20_5_daten")
kern <- makeSOCKcluster(rep("localhost", 4))
system.time(parLapply(kern, daten, fun=dich.func.pcm,n=1000,it=20,kat=5))
stopCluster(kern)

#-------------------------------------------------------------------------------------------
# Stichprobengröße 1000; 20 Items; 6 Kategorien
THRES <- matrix(c(-1.4,-0.78,-0.23,0,0.2,-3.7,-3.48,-2.11,-1,1,-1,1.2,2.5,3,3.5,-3.3,-
2.9,-2.2,-1.9,1.3,0.5,1.3,2,2.4,3,-1.7,-0.8,-0.4,0.2,0.4,-1,-0.2,1,1.9,2.3,-0.2,2.4,3,
3.3,4,-2.5,-1.7,-1.3,-0.7,0.6,-2.3,-1.7,-0.9,-0.6,-0.2,-1.4,-1,-0.23,0.2,1.3,-3.2,-3,-
2.31,-1.19,-1,1.2,2.2,2.9,3,3.5,-2.8,-2.2,-2,-1.8,-1.2,0.6,0.9,1.7,2.23,2.6,
-1.23,-0.78,-0.23,0.2,0,-3.27,-3,-2.11,-1,0.1,0.3,1.8,2.5,3,3.5,-3.3,-2.9,-2.2,-
1.9,1.3,0.5,1.32,2,2.4,3.2),nrow=5)
n <- 1000 ; it <- 20 ; wh <- 10000

create_data(thres=THRES,size=n,items=it,wh=10000,kat=6)

daten <- list.files("~/Simulation_2015_PCM/results/pcm_1000_20_6_daten")
kern <- makeSOCKcluster(rep("localhost", 4))
parLapply(kern, daten, fun=dich.func.pcm,n=1000,it=20,kat=6)
stopCluster(kern)

#-------------------------------------------------------------------------------------------
# ERSTELLUNG DATENMATRIX
#-------------------------------------------------------------------------------------------
# Stichprobengröße 1000; 20 Items; 4 Kategorien
filenames <- list.files(path="~/Simulation_2015_PCM/results/pcm_1000_20_4_res")
res <- do.call("cbind",lapply(paste("~/Simulation_2015_PCM/results/pcm_1000_20_4_res/",
filenames,sep=""),read.table))
res <- t(res)
colnames(res) <-
c("KorrelationPp_u","KorrelationPp_m","KorrelationPp_o","LRTPCMpvalue","LRTRMpvalue_u",
"LRTRMpvalue_m","LRTRMpvalue_o","ItemFitPCM","ItemFitRM_u","ItemFitRM_m","ItemFitRM_o",
"PersonenFitPCM","PersonenFitRM_u","PersonenFitRM_m","PersonenFitRM_o","AICPCM","AICRM_u
","AICRM_m","AICRM_o","BICPCM","BICRM_u","BICRM_m","BICRM_o")
rownames(res) <- NULL
write.table(res, file=paste("~/Simulation_2015_PCM/results/pcm_1000_20_4_table",sep=""))

#-------------------------------------------------------------------------------------------
# Stichprobengröße 1000; 20 Items; 5 Kategorien
filenames <- list.files(path="~/Simulation_2015_PCM/results/pcm_1000_20_5_res")
res <- do.call("cbind",lapply(paste("~/Simulation_2015_PCM/results/pcm_1000_20_5_res/",
               filenames,sep=""), read.table))
res <- t(res)
colnames(res) <-
c("KorrelationPp_u","KorrelationPp_m","KorrelationPp_o","LRTPCMpvalue","LRTRMpvalue_u",
"LRTRMpvalue_m","LRTRMpvalue_o","ItemFitPCM","ItemFitRM_u","ItemFitRM_m","ItemFitRM_o",
"PersonenFitPCM","PersonenFitRM_u","PersonenFitRM_m","PersonenFitRM_o","AICPCM","AICRM_u
","AICRM_m","AICRM_o","BICPCM","BICRM_u","BICRM_m","BICRM_o")
rownames(res) <- NULL
write.table(res, file=paste("~/Simulation_2015_PCM/results/pcm_1000_20_5_table",sep=""))

#-------------------------------------------------------------------------------------------
# Stichprobengröße 1000; 20 Items; 6 Kategorien
filenames <- list.files(path="~/Simulation_2015_PCM/results/pcm_1000_20_6_res")
res <- do.call("cbind",lapply(paste("~/Simulation_2015_PCM/results/pcm_1000_20_6_res/",
filenames,sep=""), read.table))
res <- t(res)
colnames(res) <-
c("KorrelationPp_u","KorrelationPp_m","KorrelationPp_o","LRTPCMpvalue","LRTRMpvalue_u",
"LRTRMpvalue_m","LRTRMpvalue_o","ItemFitPCM","ItemFitRM_u","ItemFitRM_m","ItemFitRM_o",
"PersonenFitPCM","PersonenFitRM_u","PersonenFitRM_m","PersonenFitRM_o","AICPCM","AICRM_u
","AICRM_m","AICRM_o","BICPCM","BICRM_u","BICRM_m","BICRM_o")
rownames(res) <- NULL
```

```
#-------------------------------------------------------------------------------------------
# DICHOTOMISIEREN VON PCM-DATEN (N=2000)
#-------------------------------------------------------------------------------------------
# PACKAGES
#-------------------------------------------------------------------------------------------
library(PP)
library(eRm)
library(snow)
#-------------------------------------------------------------------------------------------
# DATENGENERIERUNG
#-------------------------------------------------------------------------------------------
create_data <- function(thres=THRES,size=n,items=it,wh=0,kat=0){
  v <- 1:wh
  sl <- rep(1,it)
  THRESx <- rbind(0,THRES)
  THETA <- rnorm(n)
  for(i in 1:wh){
    simdat_gpcm <- sim_gpcm(thres = THRESx,alpha = sl,theta = THETA)

write.table(simdat_gpcm,file=paste("~/Simulation_2015_PCM/results/pcm_",n,"_",it,"_",kat
,"_daten","/",v[i],sep=""))
  }
}
#-------------------------------------------------------------------------------------------
# DICHOTOMISIERUNGSFUNKTION
#-------------------------------------------------------------------------------------------
dich.func.pcm <-function(g,n,it,kat){
  require(eRm)
#-------------------------------------------------------------------------------------------
  # Dichotomisierungsfunktion
  dich.pcm <- function(m){
    res <- NULL
    if(((max(m)+1)%%2)==0) {
      if(max(m)==3){
        dich.pcm.u <- ifelse(m < min(m)+1,0,1)
        dich.pcm.m <- ifelse(m <(max(m)+1)/2,0,1)
        dich.pcm.o <- ifelse(m < max(m),0,1)
        res <- list(dich.pcm.u,dich.pcm.m,dich.pcm.o)}
      if(max(m)==5){
        dich.pcm.u <- ifelse(m < min(m)+2,0,1)
        dich.pcm.m <- ifelse(m <(max(m)+1)/2,0,1)
        dich.pcm.o <- ifelse(m < max(m)-1,0,1)
        res <- list(dich.pcm.u,dich.pcm.m,dich.pcm.o)}
    }
    if(((max(m)+1)%%2)!=0) {
      if(max(m)==4){
        dich.pcm.u <- ifelse(m < min(m)+1,0,1)
        dich.pcm.m <- ifelse(m < max(m)/2,0,1)
        dich.pcm.o <- ifelse(m < max(m),0,1)
        res <- list(dich.pcm.u,dich.pcm.m,dich.pcm.o)}
      if(max(m)==6){
        dich.pcm.u <- ifelse(m < min(m)+2,0,1)
        dich.pcm.m <- ifelse(m < max(m)/2,0,1)
        dich.pcm.o <- ifelse(m < max(m)-1,0,1)
        res <- list(dich.pcm.u,dich.pcm.m,dich.pcm.o)}
    }
    return(res)
  }
#-------------------------------------------------------------------------------------------
  erg <-read.table(paste("~/Simulation_2015_PCM/results/pcm_",
                  n,"_",it,"_",kat,"_daten/",g,sep=""))
  rmprobe <- dich.pcm(erg)

  # Modellberechnung + Personenparameter PCM
  estpcm <- try(PCM(erg))
  perspcm <- try(person.parameter(estpcm))

  # Modellberechnung + Personenparameter RM nach Teilungskriterium
  estdich_u <- try(RM(rmprobe[[1]]))
  persdich_u <- try(person.parameter(estdich_u))
  estdich_m <- try(RM(rmprobe[[2]]))
  persdich_m <- try(person.parameter(estdich_m))
  estdich_o <- try(RM(rmprobe[[3]]))
  persdich_o <- try(person.parameter(estdich_o))
```

```
# Korrelationen
perspcmdat <- as.numeric(try(unlist(perspcm[[13]][1])))
persdichdat_u <- as.numeric(try(unlist(persdich_u[[13]][1])))
persdichdat_m <- as.numeric(try(unlist(persdich_m[[13]][1])))
persdichdat_o <- as.numeric(try(unlist(persdich_o[[13]][1])))

a_u <- as.numeric(try(cor(perspcmdat, persdichdat_u,use="na.or.complete")))
a_m <- as.numeric(try(cor(perspcmdat, persdichdat_m,use="na.or.complete")))
a_o <- as.numeric(try(cor(perspcmdat, persdichdat_o,use="na.or.complete")))

# Likelihood-Ratio
lrtpcm <- try(LRtest(estpcm, splitcr="median"))
b <- as.numeric(try(lrtpcm$pvalue))

lrtdich_u <- try(LRtest(estdich_u, splitcr="median"))
b_u <- as.numeric(try(lrtdich_u$pvalue))

lrtdich_m <- try(LRtest(estdich_m, splitcr="median"))
b_m <- as.numeric(try(lrtdich_m$pvalue))

lrtdich_o <- try(LRtest(estdich_o, splitcr="median"))
b_o <- as.numeric(try(lrtdich_o$pvalue))

# Itemfit
itemfit <- eRm::itemfit
IFIpcm <- try(itemfit(perspcm))
IFIdich_u <- try(itemfit(persdich_u))
IFIdich_m <- try(itemfit(persdich_m))
IFIdich_o <- try(itemfit(persdich_o))

alpha <- 0.05
ipvalvecpcm <- try(1-pchisq(IFIpcm$i.fit, IFIpcm$i.df))
ipvalsigpcm <- try(which(ipvalvecpcm < alpha))
inumberpcm <- try(length(ipvalsigpcm))
d<-inumberpcm

ipvalvecdich_u <- try(1-pchisq(IFIdich_u$i.fit, IFIdich_u$i.df))
ipvalsigdich_u <- try(which(ipvalvecdich_u < alpha))
inumberdich_u <- try(length(ipvalsigdich_u))
d_u<-inumberdich_u

ipvalvecdich_m <- try(1-pchisq(IFIdich_m$i.fit, IFIdich_m$i.df))
ipvalsigdich_m <- try(which(ipvalvecdich_m < alpha))
inumberdich_m <- try(length(ipvalsigdich_m))
d_m<-inumberdich_m

ipvalvecdich_o <- try(1-pchisq(IFIdich_o$i.fit, IFIdich_o$i.df))
ipvalsigdich_o <- try(which(ipvalvecdich_o < alpha))
inumberdich_o <- try(length(ipvalsigdich_o))
d_o<-inumberdich_o

# Personenfit
personfit <- eRm::personfit
PFIpcm <- try(personfit(perspcm))
PFIdich_u <- try(personfit(persdich_u))
PFIdich_m <- try(personfit(persdich_m))
PFIdich_o <- try(personfit(persdich_o))

ppvalvecpcm <- try(1-pchisq(PFIpcm$p.fit, PFIpcm$p.df))
ppvalsigpcm <- try(which(ppvalvecpcm < alpha))
pnumberpcm <- try(length(ppvalsigpcm))
e<-pnumberpcm

ppvalvecdich_u <- try(1-pchisq(PFIdich_u$p.fit, PFIdich_u$p.df))
ppvalsigdich_u <- try(which(ppvalvecdich_u < alpha))
pnumberdich_u <- try(length(ppvalsigdich_u))
e_u<-pnumberdich_u

ppvalvecdich_m <- try(1-pchisq(PFIdich_m$p.fit, PFIdich_m$p.df))
ppvalsigdich_m <- try(which(ppvalvecdich_m < alpha))
pnumberdich_m <- try(length(ppvalsigdich_m))
e_m<-pnumberdich_m

ppvalvecdich_o <- try(1-pchisq(PFIdich_o$p.fit, PFIdich_o$p.df))
ppvalsigdich_o <- try(which(ppvalvecdich_o < alpha))
```

```
  pnumberdich_o <- try(length(ppvalsigdich_o))
  e_o<-pnumberdich_o

  inf.krit.pcm <- try(IC(perspcm))
  aicpcm <- try(inf.krit.pcm[[1]][7])
  bicpcm <- try(inf.krit.pcm[[1]][10])
  f<-aicpcm; h<-bicpcm

  inf.krit.dich_u <- try(IC(persdich_u))
  aicdich_u <- try(inf.krit.dich_u[[1]][7])
  bicdich_u <- try(inf.krit.dich_u[[1]][10])
  f_u<-aicdich_u; h_u<-bicdich_u

  inf.krit.dich_m <- try(IC(persdich_m))
  aicdich_m <- try(inf.krit.dich_m[[1]][7])
  bicdich_m <- try(inf.krit.dich_m[[1]][10])
  f_m<-aicdich_m; h_m<-bicdich_m

  inf.krit.dich_o <- try(IC(persdich_o))
  aicdich_o <- try(inf.krit.dich_o[[1]][7])
  bicdich_o <- try(inf.krit.dich_o[[1]][10])
  f_o<-aicdich_o; h_o<-bicdich_o

  dt.pcm.dich <- c(a_u,a_m,a_o,b,b_u,b_m,b_o,d,d_u,d_m,d_o,
                   e,e_u,e_m,e_o,f,f_u,f_m,f_o,h,h_u,h_m,h_o)

  write.table(dt.pcm.dich, file =paste("~/Simulation_2015_PCM/
              results/pcm_",n,"_",it,"_",kat,"_res/",g,sep=""))
}

#-------------------------------------------------------------------------------------------
# VERZEICHNISERSTELLUNG
#-------------------------------------------------------------------------------------------
dir.create("~/Simulation_2015_PCM/results/pcm_2000_20_4_daten")
dir.create("~/Simulation_2015_PCM/results/pcm_2000_20_4_res")
dir.create("~/Simulation_2015_PCM/results/pcm_2000_20_5_daten")
dir.create("~/Simulation_2015_PCM/results/pcm_2000_20_5_res")
dir.create("~/Simulation_2015_PCM/results/pcm_2000_20_6_daten")
dir.create("~/Simulation_2015_PCM/results/pcm_2000_20_6_res")

#-------------------------------------------------------------------------------------------
# DATENGENERIERUNG & SIMULATION 10000
#-------------------------------------------------------------------------------------------
# Stichprobengröße 2000; 20 Items; 4 Kategorien
THRES <- matrix(c(-2,-1.23,-1,-4.2,-3.48,-1.11,1.2,2.5,3,-1,-0.8,-0.2,0.5,1.3,2,
                  -0.8,-0.4,0.2,-0.2,1,1.9,-0.2,2.4,3,-1,1.2,2,-2.1,-1.6,-0.88,
                  -1.5,-1,1,0.4,0.9,1.3,-2.12,-1.8,-0.5,0.2,1.2,2,-3.23,-1.2,0,
                  -0.2,0.4,1,1.2,2,3.12,-1.8,-1,-0.2,-1.2,-0.6,0.3,-0.33,0.66,1),nrow=3)
n <- 2000 ; it <- 20 ; wh <- 10000

create_data(thres=THRES,size=n,items=it,wh=10000,kat=4)

daten <- list.files("~/Simulation_2015_PCM/results/pcm_2000_20_4_daten")
kern <- makeSOCKcluster(rep("localhost", 4))
parLapply(kern, daten, fun=dich.func.pcm,n=2000,it=20,kat=4)
stopCluster(kern)
#-------------------------------------------------------------------------------------------
# Stichprobengröße 2000; 20 Items; 5 Kategorien
THRES <-matrix(c(-2.4,-1.78,-1.23,-0.2,-3.7,-3.48,-2.11,-1,1.2,2.5,3,3.5,-1.9,-1.2,-
0.9,0.3,0.5,1.3,2,2.4,-0.8,-0.4,0.2,0.4,-0.2,1,1.9,2.3,-0.2,2.4,3,3.3,-2,-1.2,-0.8,-
0.2,-2.3,-1.7,-0.9,-0.6,-1.4,-1,-0.23,0.2,-4.2,-4,-3.31,-2.19,2.2,2.9,3,3.5,-1.7,-1.2,-
0.7,0.2,1,1.8,2.23,2.9,-0.6,-0.1,0.5,1,1,2,3,4,-2,-1.2,-0.8,0.1,-2.6,-2.1,-1.3,-0.9,-
1.5,-1,1,1.5),nrow=4)
n <- 2000 ; it <- 20 ; wh <- 10000

create_data(thres=THRES,size=n,items=it,wh=10000,kat=5)

daten <- list.files("~/Simulation_2015_PCM/results/pcm_2000_20_5_daten")
kern <- makeSOCKcluster(rep("localhost", 4))
system.time(parLapply(kern, daten, fun=dich.func.pcm,n=2000,it=20,kat=5))
stopCluster(kern)

#-------------------------------------------------------------------------------------------
# Stichprobengröße 2000; 20 Items; 6 Kategorien
```

```
THRES <- matrix(c(-1.4,-0.78,-0.23,0,0.2,-3.7,-3.48,-2.11,-1,1,-1,1.2,2.5,3,3.5,-3.3,-
2.9,-2.2,-1.9,1.3,0.5,1.3,2,2.4,3,-1.7,-0.8,-0.4,0.2,0.4,-1,-0.2,1,1.9,2.3,-
0.2,2.4,3,3.3,4,-2.5,-1.7,-1.3,-0.7,0.6,-2.3,-1.7,-0.9,-0.6,-0.2,-1.4,-1,-0.23,0.2,1.3,-
3.2,-3,-2.31,-1.19,-1,1.2,2.2,2.9,3,3.5,-2.8,-2.2,-2,-1.8,-1.2,0.6,0.9,1.7,2.23,2.6,
-1.23,-0.78,-0.23,0.2,0,-3.27,-3,-2.11,-1,0.1,0.3,1.8,2.5,3,3.5,-3.3,-2.9,-2.2,-
1.9,1.3,0.5,1.32,2,2.4,3.2),nrow=5)
n <- 2000 ; it <- 20 ; wh <- 10000

create_data(thres=THRES,size=n,items=it,wh=10000,kat=6)

daten <- list.files("~/Simulation_2015_PCM/results/pcm_2000_20_6_daten")
kern <- makeSOCKcluster(rep("localhost", 4))
parLapply(kern, daten, fun=dich.func.pcm,n=2000,it=20,kat=6)
stopCluster(kern)

#-----------------------------------------------------------------------------------------
# ERSTELLUNG DATENMATRIX
#-----------------------------------------------------------------------------------------
# Stichprobengröße 2000; 20 Items; 4 Kategorien
filenames <- list.files(path="~/Simulation_2015_PCM/results/pcm_2000_20_4_res")
res <-
do.call("cbind",lapply(paste("~/Simulation_2015_PCM/results/pcm_2000_20_4_res/",filename
s,sep=""), read.table))
res <- t(res)
colnames(res) <-
c("KorrelationPp_u","KorrelationPp_m","KorrelationPp_o","LRTPCMpvalue","LRTRMpvalue_u",
"LRTRMpvalue_m","LRTRMpvalue_o","ItemFitPCM","ItemFitRM_u","ItemFitRM_m","ItemFitRM_o",
"PersonenFitPCM","PersonenFitRM_u","PersonenFitRM_m","PersonenFitRM_o","AICPCM","AICRM_u
","AICRM_m","AICRM_o","BICPCM","BICRM_u","BICRM_m","BICRM_o")
rownames(res) <- NULL
write.table(res, file
=paste("~/Simulation_2015_PCM/results/pcm_2000_20_4_table",sep=""))

#-----------------------------------------------------------------------------------------
# Stichprobengröße 2000; 20 Items; 5 Kategorien
filenames <- list.files(path="~/Simulation_2015_PCM/results/pcm_2000_20_5_res")
res <-
do.call("cbind",lapply(paste("~/Simulation_2015_PCM/results/pcm_2000_20_5_res/",filename
s,sep=""), read.table))
res <- t(res)
colnames(res) <-
c("KorrelationPp_u","KorrelationPp_m","KorrelationPp_o","LRTPCMpvalue","LRTRMpvalue_u",
"LRTRMpvalue_m","LRTRMpvalue_o","ItemFitPCM","ItemFitRM_u","ItemFitRM_m","ItemFitRM_o",
"PersonenFitPCM","PersonenFitRM_u","PersonenFitRM_m","PersonenFitRM_o","AICPCM","AICRM_u
","AICRM_m","AICRM_o","BICPCM","BICRM_u","BICRM_m","BICRM_o")
rownames(res) <- NULL
write.table(res, file
=paste("~/Simulation_2015_PCM/results/pcm_2000_20_5_table",sep=""))

#-----------------------------------------------------------------------------------------
# Stichprobengröße 2000; 20 Items; 6 Kategorien
filenames <- list.files(path="~/Simulation_2015_PCM/results/pcm_2000_20_6_res")
res <-
do.call("cbind",lapply(paste("~/Simulation_2015_PCM/results/pcm_2000_20_6_res/",filename
s,sep=""), read.table))
res <- t(res)
colnames(res) <-
c("KorrelationPp_u","KorrelationPp_m","KorrelationPp_o","LRTPCMpvalue","LRTRMpvalue_u",
"LRTRMpvalue_m","LRTRMpvalue_o","ItemFitPCM","ItemFitRM_u","ItemFitRM_m","ItemFitRM_o",
"PersonenFitPCM","PersonenFitRM_u","PersonenFitRM_m","PersonenFitRM_o","AICPCM","AICRM_u
","AICRM_m","AICRM_o","BICPCM","BICRM_u","BICRM_m","BICRM_o")
rownames(res) <- NULL
```

```
#+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
# SIMULATION ZUM GRADED-RESPONSE-MODELL
#+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
#-----------------------------------------------------------------------------------------------
# DICHOTOMISIEREN VON GRM-DATEN (N=1000)
#-----------------------------------------------------------------------------------------------
#-----------------------------------------------------------------------------------------------
# PACKAGES
#-----------------------------------------------------------------------------------------------
library(mirt)
library(snow)
#-----------------------------------------------------------------------------------------------
# DATENGENERIERUNG
#-----------------------------------------------------------------------------------------------
create.data <- function(a,d,n=0,wh=0,kat=0,it=0){
  v <- 1:wh
  for(i in 1:wh){
    poly_grm <- simdata(a, d, n, itemtype = 'graded')

write.table(poly_grm,file=paste("~/Simulation_2015_GRM/results/grm_",n,"_",it,"_",kat,"_
daten","/",v[i],sep=""))
  }
}
#-----------------------------------------------------------------------------------------------
# DICHOTOMISIERUNGSFUNKTION
#-----------------------------------------------------------------------------------------------
dich.func.grm <-function(g,n,it,kat){
  require(mirt)

#-----------------------------------------------------------------------------------------------
  # Dichotomisierungsfunktion
  dich.grm <- function(m){
    res <- NULL
    if(((max(m))%%2)==0) {
      if(max(m)==4){
        dich.grm.u <- ifelse(m < min(m)+1,0,1)
        dich.grm.m <- ifelse(m <(max(m)/2+1),0,1)
        dich.grm.o <- ifelse(m < max(m),0,1)
        res <- list(dich.grm.u,dich.grm.m,dich.grm.o)}
      if(max(m)==6){
        dich.grm.u <- ifelse(m < min(m)+2,0,1)
        dich.grm.m <- ifelse(m < (max(m)/2+1),0,1)
        dich.grm.o <- ifelse(m < max(m)-1,0,1)
        res <- list(dich.grm.u,dich.grm.m,dich.grm.o)}
    }
    if(((max(m))%%2)!=0) {
      if(max(m)==5){
        dich.grm.u <- ifelse(m < min(m)+1,0,1)
        dich.grm.m <- ifelse(m < (max(m)+1)/2,0,1)
        dich.grm.o <- ifelse(m < max(m),0,1)
        res <- list(dich.grm.u,dich.grm.m,dich.grm.o)}
    }
    return(res)
  }
#-----------------------------------------------------------------------------------------------
  polygrm <-read.table(paste("~/Simulation_2015_GRM/results/grm_",n,"_",it,"_",kat,
                        "_daten/",g,sep=""))
  grmprobe <- dich.grm(polygrm)

  # Modellberechnung und Personenparameterberechnung GRM mehrkat.
  estgrm <- try(mirt(polygrm, model=1, itemtype="graded"))
  persgrm <- try(fscores(estgrm, full.scores=TRUE))
  #Konnte der EM Algotrithmus konvergieren: Ja=1 & Nein=0
  conv <- estgrm@converge

  # Modellberechnung und Personenparameterberechnung GRM dichotom
  estgrmdich_u <- try(mirt(grmprobe[[1]],model=1,itemtype="graded"))
  persgrmdich_u <- try(fscores(estgrmdich_u, full.scores=TRUE))
  conv_u <- estgrmdich_u@converge

  estgrmdich_m <- try(mirt(grmprobe[[2]],model=1,itemtype="graded"))
  persgrmdich_m <- try(fscores(estgrmdich_m, full.scores=TRUE))
  conv_m <- estgrmdich_m@converge

  estgrmdich_o <- try(mirt(grmprobe[[3]],model=1,itemtype="graded"))
```

```
  persgrmdich_o <- try(fscores(estgrmdich_o, full.scores=TRUE))
  conv_o <- estgrmdich_o@converge

  # Korrelationen
  a_u <- as.numeric(try(cor(persgrm, persgrmdich_u,use="na.or.complete")))
  a_m <- as.numeric(try(cor(persgrm, persgrmdich_m,use="na.or.complete")))
  a_o <- as.numeric(try(cor(persgrm, persgrmdich_o,use="na.or.complete")))

  # Berechnung der Modelfit Statistiken M2, G2 und X2
  # M2
  alpha <- 0.05
  M2_grm <- M2(estgrm)
  M2_pgrm <- M2_grm[,3]

  M2_grmdich_u <- M2(estgrmdich_u)
  M2_pgrmdich_u <- M2_grmdich_u[,3]

  M2_grmdich_m <- M2(estgrmdich_m)
  M2_pgrmdich_m <- M2_grmdich_m[,3]

  M2_grmdich_o <- M2(estgrmdich_o)
  M2_pgrmdich_o <- M2_grmdich_o[,3]

  # G2 und X2 Statistiken -> benötigen Summenbildung von factor scores
  scores_poly <- fscores(estgrm, method = 'EAPsum')
  scores_dich_u <- fscores(estgrmdich_u, method = 'EAPsum')
  scores_dich_m <- fscores(estgrmdich_m, method = 'EAPsum')
  scores_dich_o <- fscores(estgrmdich_o, method = 'EAPsum')

  fitgrm <- attr(scores_poly, 'fit')
  X2_pgrm <- fitgrm[[3]]
  G2_pgrm <- fitgrm[[5]]

  fitgrmdich_u <- attr(scores_dich_u, 'fit')
  X2_pgrmdich_u <- fitgrmdich_u[[3]]
  G2_pgrmdich_u <- fitgrmdich_u[[5]]

  fitgrmdich_m <- attr(scores_dich_m, 'fit')
  X2_pgrmdich_m <- fitgrmdich_m[[3]]
  G2_pgrmdich_m <- fitgrmdich_m[[5]]

  fitgrmdich_o <- attr(scores_dich_o, 'fit')
  X2_pgrmdich_o <- fitgrmdich_o[[3]]
  G2_pgrmdich_o <- fitgrmdich_o[[5]]

  # AIC und BIC
  AIC_grm <- estgrm@AIC
  BIC_grm <- estgrm@BIC

  AIC_grm_u <- estgrmdich_u@AIC
  BIC_grm_u <- estgrmdich_u@BIC

  AIC_grm_m <- estgrmdich_m@AIC
  BIC_grm_m <- estgrmdich_m@BIC

  AIC_grm_o <- estgrmdich_o@AIC
  BIC_grm_o <- estgrmdich_o@BIC

  dt.grm.dich <- c(conv,conv_u,conv_m,conv_o,
                   a_u,a_m,a_o,M2_pgrm,M2_pgrmdich_u,M2_pgrmdich_m,M2_pgrmdich_o,
                   X2_pgrm,X2_pgrmdich_u,X2_pgrmdich_m,X2_pgrmdich_o,G2_pgrm,
                   G2_pgrmdich_u,G2_pgrmdich_m,G2_pgrmdich_o,AIC_grm,AIC_grm_u,
                   AIC_grm_m,AIC_grm_o,BIC_grm,BIC_grm_u,BIC_grm_m,BIC_grm_o)

  write.table(dt.grm.dich, file =paste("~/Simulation_2015_GRM/results/
              grm_",n,"_",it,"_",kat,"_res/",g,sep=""))
}
#-------------------------------------------------------------------------------------
# VERZEICHNISERSTELLUNG & DATENGENERIERUNG
#-------------------------------------------------------------------------------------
dir.create("~/Simulation_2015_GRM/results/grm_1000_20_4_daten")
dir.create("~/Simulation_2015_GRM/results/grm_1000_20_4_res")
dir.create("~/Simulation_2015_GRM/results/grm_1000_20_5_daten")
dir.create("~/Simulation_2015_GRM/results/grm_1000_20_5_res")
dir.create("~/Simulation_2015_GRM/results/grm_1000_20_6_daten")
```

```
dir.create("~/Simulation_2015_GRM/results/grm_1000_20_6_res")

# Stichprobengröße 1000; 20 Items; 4 Kategorien
a <- matrix(rlnorm(20, .2, .3))
diffs <- t(apply(matrix(runif(20*3, .3, 1), 20), 1, cumsum))
diffs <- -(diffs - rowMeans(diffs))
d <- diffs + rnorm(20)
system.time(create.data(a=a,d=d,n=1000,wh=10000,kat=4,it=20))

# Stichprobengröße 1000; 20 Items; 5 Kategorien
a <- matrix(rlnorm(20, .2, .3))
diffs <- t(apply(matrix(runif(20*4, .3, 1), 20), 1, cumsum))
diffs <- -(diffs - rowMeans(diffs))
d <- diffs + rnorm(20)
system.time(create.data(a=a,d=d,n=1000,wh=10000,kat=5,it=20))

# Stichprobengröße 1000; 20 Items; 6 Kategorien
a <- matrix(rlnorm(20, .2, .3))
diffs <- t(apply(matrix(runif(20*5, .3, 1), 20), 1, cumsum))
diffs <- -(diffs - rowMeans(diffs))
d <- diffs + rnorm(20)
system.time(create.data(a=a,d=d,n=1000,wh=10000,kat=6,it=20))
#-------------------------------------------------------------------------------------
# SIMULATION
#-------------------------------------------------------------------------------------
# Stichprobengröße 1000; 20 Items; 4 Kategorien
daten <- list.files("~/Simulation_2015_GRM/results/grm_1000_20_4_daten")
kern <- makeSOCKcluster(rep("localhost", 4))
system.time(parLapply(kern,daten,fun=dich.func.grm,n=1000,it=20,kat=4))
stopCluster(kern)

# Stichprobengröße 1000; 20 Items; 5 Kategorien
daten <- list.files("~/Simulation_2015_GRM/results/grm_1000_20_5_daten")
kern <- makeSOCKcluster(rep("localhost", 4))
system.time(parLapply(kern,daten,fun=dich.func.grm,n=1000,it=20,kat=5))
stopCluster(kern)

# Stichprobengröße 1000; 20 Items; 6 Kategorien
daten <- list.files("~/Simulation_2015_GRM/results/grm_1000_20_6_daten")
kern <- makeSOCKcluster(rep("localhost", 4))
system.time(parLapply(kern,daten,fun=dich.func.grm,n=1000,it=20,kat=6))
stopCluster(kern)
#-------------------------------------------------------------------------------------
# ERSTELLUNG DATENMATRIX
#-------------------------------------------------------------------------------------
# Stichprobengröße 1000; 20 Items; 4 Kategorien
filenames <- list.files(path="~/Simulation_2015_GRM/results/grm_1000_20_4_res")
res <- do.call("cbind",lapply(paste("~/Simulation_2015_GRM/results/grm_1000_20_4_res/",
filenames,sep=""), read.table))
res <- t(res)
colnames(res) <- c("ConvGRM","ConvGRM_u","ConvGRM_m","ConvGRM_o",
                   "KorrelationPp_u","KorrelationPp_m","KorrelationPp_o",
                   "M2GRM","M2GRM_u","M2GRM_m","M2GRM_o",
                   "X2GRM","X2GRM_u","X2GRM_m","X2GRM_o",
                   "G2GRM","G2GRM_u","G2GRM_m","G2GRM_o",
                   "AICGRM","AICGRM_u","AICGRM_m","AICGRM_o",
                   "BICGRM","BICGRM_u","BICGRM_m","BICGRM_o")
rownames(res) <- NULL
write.table(res, file
=paste("~/Simulation_2015_GRM/results/grm_1000_20_4_table",sep=""))

#-------------------------------------------------------------------------------------
# Stichprobengröße 1000; 20 Items; 5 Kategorien
filenames <- list.files(path="~/Simulation_2015_GRM/results/grm_1000_20_5_res")
res <- do.call("cbind",lapply(paste("~/Simulation_2015_GRM/results/grm_1000_20_5_res/",
filenames,sep=""), read.table))
res <- t(res)
colnames(res) <- c("ConvGRM","ConvGRM_u","ConvGRM_m","ConvGRM_o",
                   "KorrelationPp_u","KorrelationPp_m","KorrelationPp_o",
                   "M2GRM","M2GRM_u","M2GRM_m","M2GRM_o",
                   "X2GRM","X2GRM_u","X2GRM_m","X2GRM_o",
                   "G2GRM","G2GRM_u","G2GRM_m","G2GRM_o",
                   "AICGRM","AICGRM_u","AICGRM_m","AICGRM_o",
                   "BICGRM","BICGRM_u","BICGRM_m","BICGRM_o")
rownames(res) <- NULL
```

```
write.table(res, file
=paste("~/Simulation_2015_GRM/results/grm_1000_20_5_table",sep=""))

#-------------------------------------------------------------------------------------------
# Stichprobengröße 1000; 20 Items; 6 Kategorien
filenames <- list.files(path="~/Simulation_2015_GRM/results/grm_1000_20_6_res")
res <- do.call("cbind",lapply(paste("~/Simulation_2015_GRM/results/grm_1000_20_6_res/",
filenames,sep=""), read.table))
res <- t(res)
colnames(res) <- c("ConvGRM","ConvGRM_u","ConvGRM_m","ConvGRM_o",
                   "KorrelationPp_u","KorrelationPp_m","KorrelationPp_o",
                   "M2GRM","M2GRM_u","M2GRM_m","M2GRM_o",
                   "X2GRM","X2GRM_u","X2GRM_m","X2GRM_o",
                   "G2GRM","G2GRM_u","G2GRM_m","G2GRM_o",
                   "AICGRM","AICGRM_u","AICGRM_m","AICGRM_o",
                   "BICGRM","BICGRM_u","BICGRM_m","BICGRM_o")
rownames(res) <- NULL
write.table(res, file
=paste("~/Simulation_2015_GRM/results/grm_1000_20_6_table",sep=""))

#-------------------------------------------------------------------------------------------
```

```
#-----------------------------------------------------------------------------------------
# DICHOTOMISIEREN VON GRM-DATEN (N=2000)
#-----------------------------------------------------------------------------------------
# PACKAGES
#-----------------------------------------------------------------------------------------
library(mirt)
library(snow)
#-----------------------------------------------------------------------------------------
# DATENGENERIERUNG
#-----------------------------------------------------------------------------------------
create.data <- function(a,d,n=0,wh=0,kat=0,it=0){
  v <- 1:wh
  for(i in 1:wh){
    poly_grm <- simdata(a, d, n, itemtype = 'graded')

write.table(poly_grm,file=paste("~/Simulation_2015_GRM/results/grm_",n,"_",it,"_",kat,"_
daten","/",v[i],sep=""))
  }
}
#-----------------------------------------------------------------------------------------
# DICHOTOMISIERUNGSFUNKTION
#-----------------------------------------------------------------------------------------

dich.func.grm <-function(g,n,it,kat){
  require(mirt)

#-----------------------------------------------------------------------------------------
  # Dichotomisierungsfunktion
  dich.grm <- function(m){
    res <- NULL
    if(((max(m))%%2)==0) {
      if(max(m)==4){
        dich.grm.u <- ifelse(m < min(m)+1,0,1)
        dich.grm.m <- ifelse(m <(max(m)/2+1),0,1)
        dich.grm.o <- ifelse(m < max(m),0,1)
        res <- list(dich.grm.u,dich.grm.m,dich.grm.o)}
      if(max(m)==6){
        dich.grm.u <- ifelse(m < min(m)+2,0,1)
        dich.grm.m <- ifelse(m < (max(m)/2+1),0,1)
        dich.grm.o <- ifelse(m < max(m)-1,0,1)
        res <- list(dich.grm.u,dich.grm.m,dich.grm.o)}
    }
    if(((max(m))%%2)!=0) {
      if(max(m)==5){
        dich.grm.u <- ifelse(m < min(m)+1,0,1)
        dich.grm.m <- ifelse(m < (max(m)+1)/2,0,1)
        dich.grm.o <- ifelse(m < max(m),0,1)
        res <- list(dich.grm.u,dich.grm.m,dich.grm.o)}
    }
    return(res)
  }
#-----------------------------------------------------------------------------------------
  polygrm <-read.table(paste("~/Simulation_2015_GRM/results/grm_",n,"_",it,"_",kat,
                        "_daten/",g,sep=""))
  grmprobe <- dich.grm(polygrm)

  # Modellberechnung und Personenparameterberechnung GRM mehrkat.
  estgrm <- try(mirt(polygrm, model=1, itemtype="graded"))
  persgrm <- try(fscores(estgrm, full.scores=TRUE))
  #Konnte der EM Algotrithmus konvergieren: Ja=1 & Nein=0
  conv <- estgrm@converge

  # Modellberechnung und Personenparameterberechnung GRM dichotom
  estgrmdich_u <- try(mirt(grmprobe[[1]],model=1,itemtype="graded"))
  persgrmdich_u <- try(fscores(estgrmdich_u, full.scores=TRUE))
  conv_u <- estgrmdich_u@converge

  estgrmdich_m <- try(mirt(grmprobe[[2]],model=1,itemtype="graded"))
  persgrmdich_m <- try(fscores(estgrmdich_m, full.scores=TRUE))
  conv_m <- estgrmdich_m@converge

  estgrmdich_o <- try(mirt(grmprobe[[3]],model=1,itemtype="graded"))
  persgrmdich_o <- try(fscores(estgrmdich_o, full.scores=TRUE))
  conv_o <- estgrmdich_o@converge
```

```
  # Korrelationen
  a_u <- as.numeric(try(cor(persgrm, persgrmdich_u,use="na.or.complete")))
  a_m <- as.numeric(try(cor(persgrm, persgrmdich_m,use="na.or.complete")))
  a_o <- as.numeric(try(cor(persgrm, persgrmdich_o,use="na.or.complete")))

  # Berechnung der Modelfit Statistiken M2, G2 und X2
  # M2
  alpha <- 0.05
  M2_grm <- M2(estgrm)
  M2_pgrm <- M2_grm[,3]

  M2_grmdich_u <- M2(estgrmdich_u)
  M2_pgrmdich_u <- M2_grmdich_u[,3]

  M2_grmdich_m <- M2(estgrmdich_m)
  M2_pgrmdich_m <- M2_grmdich_m[,3]

  M2_grmdich_o <- M2(estgrmdich_o)
  M2_pgrmdich_o <- M2_grmdich_o[,3]

  # G2 und X2 Statistiken -> benötigen Summenbildung von factor scores
  scores_poly <- fscores(estgrm, method = 'EAPsum')
  scores_dich_u <- fscores(estgrmdich_u, method = 'EAPsum')
  scores_dich_m <- fscores(estgrmdich_m, method = 'EAPsum')
  scores_dich_o <- fscores(estgrmdich_o, method = 'EAPsum')

  fitgrm <- attr(scores_poly, 'fit')
  X2_pgrm <- fitgrm[[3]]
  G2_pgrm <- fitgrm[[5]]

  fitgrmdich_u <- attr(scores_dich_u, 'fit')
  X2_pgrmdich_u <- fitgrmdich_u[[3]]
  G2_pgrmdich_u <- fitgrmdich_u[[5]]

  fitgrmdich_m <- attr(scores_dich_m, 'fit')
  X2_pgrmdich_m <- fitgrmdich_m[[3]]
  G2_pgrmdich_m <- fitgrmdich_m[[5]]

  fitgrmdich_o <- attr(scores_dich_o, 'fit')
  X2_pgrmdich_o <- fitgrmdich_o[[3]]
  G2_pgrmdich_o <- fitgrmdich_o[[5]]

  # AIC und BIC
  AIC_grm <- estgrm@AIC
  BIC_grm <- estgrm@BIC

  AIC_grm_u <- estgrmdich_u@AIC
  BIC_grm_u <- estgrmdich_u@BIC

  AIC_grm_m <- estgrmdich_m@AIC
  BIC_grm_m <- estgrmdich_m@BIC

  AIC_grm_o <- estgrmdich_o@AIC
  BIC_grm_o <- estgrmdich_o@BIC

  dt.grm.dich <- c(conv,conv_u,conv_m,conv_o,
                   a_u,a_m,a_o,M2_pgrm,M2_pgrmdich_u,M2_pgrmdich_m,M2_pgrmdich_o,
                   X2_pgrm,X2_pgrmdich_u,X2_pgrmdich_m,X2_pgrmdich_o,G2_pgrm,
                   G2_pgrmdich_u,G2_pgrmdich_m,G2_pgrmdich_o,AIC_grm,AIC_grm_u,
                   AIC_grm_m,AIC_grm_o,BIC_grm,BIC_grm_u,BIC_grm_m,BIC_grm_o)

  write.table(dt.grm.dich, file =paste("~/Simulation_2015_GRM/results/
              grm_",n,"_",it,"_",kat,"_res/",g,sep=""))
}
#-------------------------------------------------------------------------------------
# VERZEICHNISERSTELLUNG & DATENGENERIERUNG
#-------------------------------------------------------------------------------------
dir.create("~/Simulation_2015_GRM/results/grm_2000_20_4_daten")
dir.create("~/Simulation_2015_GRM/results/grm_2000_20_4_res")
dir.create("~/Simulation_2015_GRM/results/grm_2000_20_5_daten")
dir.create("~/Simulation_2015_GRM/results/grm_2000_20_5_res")
dir.create("~/Simulation_2015_GRM/results/grm_2000_20_6_daten")
dir.create("~/Simulation_2015_GRM/results/grm_2000_20_6_res")

# Stichprobengröße 2000; 20 Items; 4 Kategorien
```

```
a <- matrix(rlnorm(20, .2, .3))
diffs <- t(apply(matrix(runif(20*3, .3, 1), 20), 1, cumsum))
diffs <- -(diffs - rowMeans(diffs))
d <- diffs + rnorm(20)
system.time(create.data(a=a,d=d,n=2000,wh=10000,kat=4,it=20))

# Stichprobengröße 2000; 20 Items; 5 Kategorien
a <- matrix(rlnorm(20, .2, .3))
diffs <- t(apply(matrix(runif(20*4, .3, 1), 20), 1, cumsum))
diffs <- -(diffs - rowMeans(diffs))
d <- diffs + rnorm(20)
system.time(create.data(a=a,d=d,n=2000,wh=10000,kat=5,it=20))

# Stichprobengröße 2000; 20 Items; 6 Kategorien
a <- matrix(rlnorm(20, .2, .3))
diffs <- t(apply(matrix(runif(20*5, .3, 1), 20), 1, cumsum))
diffs <- -(diffs - rowMeans(diffs))
d <- diffs + rnorm(20)
system.time(create.data(a=a,d=d,n=2000,wh=10000,kat=6,it=20))
#-----------------------------------------------------------------------------------------
# SIMULATION
#-----------------------------------------------------------------------------------------
# Stichprobengröße 2000; 20 Items; 4 Kategorien
daten <- list.files("~/Simulation_2015_GRM/results/grm_2000_20_4_daten")
kern <- makeSOCKcluster(rep("localhost", 4))
system.time(parLapply(kern,daten,fun=dich.func.grm,n=2000,it=20,kat=4))
stopCluster(kern)

# Stichprobengröße 2000; 20 Items; 5 Kategorien
daten <- list.files("~/Simulation_2015_GRM/results/grm_2000_20_5_daten")
kern <- makeSOCKcluster(rep("localhost", 4))
system.time(parLapply(kern,daten,fun=dich.func.grm,n=2000,it=20,kat=5))
stopCluster(kern)

# Stichprobengröße 2000; 20 Items; 6 Kategorien
daten <- list.files("~/Simulation_2015_GRM/results/grm_2000_20_6_daten")
kern <- makeSOCKcluster(rep("localhost", 4))
system.time(parLapply(kern,daten,fun=dich.func.grm,n=2000,it=20,kat=6))
stopCluster(kern)
#-----------------------------------------------------------------------------------------
# ERSTELLUNG DATENMATRIX
#-----------------------------------------------------------------------------------------
# Stichprobengröße 2000; 20 Items; 4 Kategorien
filenames <- list.files(path="~/Simulation_2015_GRM/results/grm_2000_20_4_res")
res <- do.call("cbind",lapply(paste("~/Simulation_2015_GRM/results/grm_2000_20_4_res/",
filenames,sep=""), read.table))
res <- t(res)
colnames(res) <- c("ConvGRM","ConvGRM_u","ConvGRM_m","ConvGRM_o",
                   "KorrelationPp_u","KorrelationPp_m","KorrelationPp_o",
                   "M2GRM","M2GRM_u","M2GRM_m","M2GRM_o",
                   "X2GRM","X2GRM_u","X2GRM_m","X2GRM_o",
                   "G2GRM","G2GRM_u","G2GRM_m","G2GRM_o",
                   "AICGRM","AICGRM_u","AICGRM_m","AICGRM_o",
                   "BICGRM","BICGRM_u","BICGRM_m","BICGRM_o")
rownames(res) <- NULL
write.table(res, file
=paste("~/Simulation_2015_GRM/results/grm_2000_20_4_table",sep=""))

#-----------------------------------------------------------------------------------------
# Stichprobengröße 2000; 20 Items; 5 Kategorien
filenames <- list.files(path="~/Simulation_2015_GRM/results/grm_2000_20_5_res")
res <- do.call("cbind",lapply(paste("~/Simulation_2015_GRM/results/grm_2000_20_5_res/",
filenames,sep=""), read.table))
res <- t(res)
colnames(res) <- c("ConvGRM","ConvGRM_u","ConvGRM_m","ConvGRM_o",
                   "KorrelationPp_u","KorrelationPp_m","KorrelationPp_o",
                   "M2GRM","M2GRM_u","M2GRM_m","M2GRM_o",
                   "X2GRM","X2GRM_u","X2GRM_m","X2GRM_o",
                   "G2GRM","G2GRM_u","G2GRM_m","G2GRM_o",
                   "AICGRM","AICGRM_u","AICGRM_m","AICGRM_o",
                   "BICGRM","BICGRM_u","BICGRM_m","BICGRM_o")
rownames(res) <- NULL
write.table(res, file
=paste("~/Simulation_2015_GRM/results/grm_2000_20_5_table",sep=""))
```

```
#-------------------------------------------------------------------------------------------
# Stichprobengröße 2000; 20 Items; 6 Kategorien
filenames <- list.files(path="~/Simulation_2015_GRM/results/grm_2000_20_6_res")
res <- do.call("cbind",lapply(paste("~/Simulation_2015_GRM/results/grm_2000_20_6_res/",
filenames,sep=""), read.table))
res <- t(res)
colnames(res) <- c("ConvGRM","ConvGRM_u","ConvGRM_m","ConvGRM_o",
                   "KorrelationPp_u","KorrelationPp_m","KorrelationPp_o",
                   "M2GRM","M2GRM_u","M2GRM_m","M2GRM_o",
                   "X2GRM","X2GRM_u","X2GRM_m","X2GRM_o",
                   "G2GRM","G2GRM_u","G2GRM_m","G2GRM_o",
                   "AICGRM","AICGRM_u","AICGRM_m","AICGRM_o",
                   "BICGRM","BICGRM_u","BICGRM_m","BICGRM_o")
rownames(res) <- NULL
write.table(res, file
=paste("~/Simulation_2015_GRM/results/grm_2000_20_6_table",sep=""))

#-------------------------------------------------------------------------------------------
```

```
#+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
# SIMULATION ZUM KONTINUIERLICHEN MODELL NACH MÜLLER
#+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
#-----------------------------------------------------------------------------------------------
# DICHOTOMISIEREN VON CRSM-DATEN (N=1000)
#-----------------------------------------------------------------------------------------------
# PACKAGES
#-----------------------------------------------------------------------------------------------
library(pcIRT)
library(eRm)
library(snow)

#-----------------------------------------------------------------------------------------------
# DATENGENERIERUNG
#-----------------------------------------------------------------------------------------------
create.data <- function(g,n){
  library(pcIRT)
  item_p <- seq(-1.5,1.5,length.out=20)
  dis_p <- 5
  pp <- rnorm(n, 0,1)
  res <- simCRSM(item_p, dis_p, pp)
  simdat_crsm <- res[[1]]

write.table(simdat_crsm,file=paste("~/Simulation_2015_CRSM/results/crsm_",n,"_daten","/"
,g,sep=""))
}

#-----------------------------------------------------------------------------------------------
# DICHOTOMISIERUNGSFUNKTION
#-----------------------------------------------------------------------------------------------
dich.func.crsm <-function(g,n){
  require(eRm)
  require(pcIRT)

  dich.crsm <- function(m){
    res1<- NULL
    temp <- mean(m)
    res1 <- ifelse(m<temp,0,1)
    return(res1)
  }

  poly.crsm.1 <- function(m){
    res2 <- NULL
    u <-quantile(m)[[2]]
    mi<-quantile(m)[[3]]
    o <-quantile(m)[[4]]

    m[m>=o & m < 1 ] <- 3
    m[m>=mi & m < o ] <- 2
    m[m>=u & m < mi] <- 1
    m[m<u] <- 0

    res2 <- m
    return(res2)
  }

  poly.crsm.2 <- function(m){
    res3 <- NULL
    u <-0.25
    mi<-0.5
    o <-0.75

    m[m>=o & m < 1 ] <- 3
    m[m>=mi & m < o ] <- 2
    m[m>=u & m < mi] <- 1
    m[m<u] <- 0

    res3 <- m
    return(res3)
  }

  erg <-read.table(paste("~/Simulation_2015_CRSM/results/crsm_",n,"_daten/",g,sep=""))
  erg <-as.matrix(erg)

  cr.dich <- dich.crsm(erg)
```

```
write.table(cr.dich,file=paste("~/Simulation_2015_CRSM/results/crsm_",n,"_nr/
            dich_daten/",g,sep=""))
cr.poly.1 <- poly.crsm.1(erg)
write.table(cr.poly.1,file=paste("~/Simulation_2015_CRSM/results/crsm_",n,"_nr/
            poly1_daten/",g,sep=""))
cr.poly.2 <- poly.crsm.2(erg)
write.table(cr.poly.2,file=paste("~/Simulation_2015_CRSM/results/crsm_",n,"_nr/
            poly2_daten/",g,sep=""))

# Modellberechnung + Personenparameter nach Teilungskriterium
estcrsm <- CRSM(erg, min=0, max=1)
perscrsmorg <- person_par(estcrsm)
save(perscrsmorg,file=paste("~/Simulation_2015_CRSM/results/crsm_",n,"_nr/
     crsm_pp/",g,".rda",sep=""))
perscrsm <- perscrsmorg[[2]][,2]

estdich <- RM(cr.dich)
persdich <- person.parameter(estdich)
save(persdich,file=paste("~/Simulation_2015_CRSM/results/crsm_",n,"_nr/
     dich_pp/",g,".rda",sep=""))

estpcm.1 <- PCM(cr.poly.1)
perspcm.1 <- person.parameter(estpcm.1)
save(perspcm.1,file=paste("~/Simulation_2015_CRSM/results/crsm_",n,"_nr/
     poly1_pp/",g,".rda",sep=""))

estpcm.2 <- PCM(cr.poly.2)
perspcm.2 <- person.parameter(estpcm.2)
save(perspcm.2,file=paste("~/Simulation_2015_CRSM/results/crsm_",n,"_nr/
     poly2_pp/",g,".rda",sep=""))

persdichdat <- as.numeric(unlist(persdich[[13]][1]))
perspcmdat.1 <- as.numeric(unlist(perspcm.1[[13]][1]))
perspcmdat.2 <- as.numeric(unlist(perspcm.2[[13]][1]))

# Korrelationen
a_p.1d<-cor(perspcmdat.1, persdichdat,use="na.or.complete")
a_p.2d<-cor(perspcmdat.2, persdichdat,use="na.or.complete")
a_p.12<-cor(perspcmdat.1, perspcmdat.2,use="na.or.complete")
a_dc<-cor(persdichdat, perscrsm,use="na.or.complete")
a_p.1c<-cor(perspcmdat.1, perscrsm,use="na.or.complete")
a_p.2c<-cor(perspcmdat.2, perscrsm,use="na.or.complete")

# Likelihood-Ratio
lrtdich <- try(LRtest(estdich, splitcr="median"))
lv_d <- as.numeric(try(lrtdich$LR))
lp_d <- as.numeric(try(lrtdich$pvalue))

lrtpcm.1 <- try(LRtest(estpcm.1, splitcr="median"))
lv_p.1 <- as.numeric(try(lrtpcm.1$LR))
lp_p.1 <- as.numeric(try(lrtpcm.1$pvalue))

lrtpcm.2 <- try(LRtest(estpcm.2, splitcr="median"))
lv_p.2 <- as.numeric(try(lrtpcm.2$LR))
lp_p.2 <- as.numeric(try(lrtpcm.2$pvalue))

# Itemfit & Personenfit
itemfit <- eRm::itemfit
IFIdich <- itemfit(persdich)
IFIpcm.1 <- itemfit(perspcm.1)
IFIpcm.2 <- itemfit(perspcm.2)

personfit <- eRm::personfit
PFIdich <- personfit(persdich)
PFIpcm.1 <- personfit(perspcm.1)
PFIpcm.2 <- personfit(perspcm.2)

alpha <- 0.05
ipvalvecdich <- 1-pchisq(IFIdich$i.fit, IFIdich$i.df)
ipvalsigdich <- which(ipvalvecdich < alpha)
inumberdich <- length(ipvalsigdich)
i.d<-inumberdich

ipvalvecpcm.1 <- 1-pchisq(IFIpcm.1$i.fit, IFIpcm.1$i.df)
ipvalsigpcm.1 <- which(ipvalvecpcm.1 < alpha)
```

```
  inumberpcm.1 <- length(ipvalsigpcm.1)
  i.p.1<-inumberpcm.1

  ipvalvecpcm.2 <- 1-pchisq(IFIpcm.2$i.fit, IFIpcm.2$i.df)
  ipvalsigpcm.2 <- which(ipvalvecpcm.2 < alpha)
  inumberpcm.2 <- length(ipvalsigpcm.2)
  i.p.2<-inumberpcm.2

  ppvalvecdich <- 1-pchisq(PFIdich$p.fit, PFIdich$p.df)
  ppvalsigdich <- which(ppvalvecdich < alpha)
  pnumberdich <- length(ppvalsigdich)
  p.d<-pnumberdich

  ppvalvecpcm.1 <- 1-pchisq(PFIpcm.1$p.fit, PFIpcm.1$p.df)
  ppvalsigpcm.1 <- which(ppvalvecpcm.1 < alpha)
  pnumberpcm.1 <- length(ppvalsigpcm.1)
  p.p.1<-pnumberpcm.1

  ppvalvecpcm.2 <- 1-pchisq(PFIpcm.2$p.fit, PFIpcm.2$p.df)
  ppvalsigpcm.2 <- which(ppvalvecpcm.2 < alpha)
  pnumberpcm.2 <- length(ppvalsigpcm.2)
  p.p.2<-pnumberpcm.2

  dt.crsm.dich <- c(a_p.1d,a_p.2d,a_p.12,a_dc,a_p.1c,a_p.2c,lv_d,
                  lp_d,lv_p.1,lp_p.1,lv_p.2,lp_p.2,i.d,i.p.1,i.p.2,p.d,p.p.1,p.p.2)

  write.table(dt.crsm.dich, file =paste("~/Simulation_2015_CRSM/results/
              crsm_",n,"_res/",g,sep=""))
  gc()
}

#-----------------------------------------------------------------------------------------
# VERZEICHNISERSTELLUNG & DATENGENERIERUNG
#-----------------------------------------------------------------------------------------
dir.create("~/Simulation_2015_CRSM/results/crsm_1000_daten")
dir.create("~/Simulation_2015_CRSM/results/crsm_1000_res")
dir.create("~/Simulation_2015_CRSM/results/crsm_1000_nr")

dir.create("~/Simulation_2015_CRSM/results/crsm_1000_nr/dich_daten")
dir.create("~/Simulation_2015_CRSM/results/crsm_1000_nr/poly1_daten")
dir.create("~/Simulation_2015_CRSM/results/crsm_1000_nr/poly2_daten")
dir.create("~/Simulation_2015_CRSM/results/crsm_1000_nr/crsm_pp")
dir.create("~/Simulation_2015_CRSM/results/crsm_1000_nr/dich_pp")
dir.create("~/Simulation_2015_CRSM/results/crsm_1000_nr/poly1_pp")
dir.create("~/Simulation_2015_CRSM/results/crsm_1000_nr/poly2_pp")

# Stichprobengröße 1000; 20 Items
indizes <- as.character(rep(1:10000))
kern <- makeSOCKcluster(rep("localhost", 4))
parLapply(kern,indizes, fun=create.data,n=1000)
stopCluster(kern)

#-----------------------------------------------------------------------------------------
# SIMULATION
#-----------------------------------------------------------------------------------------
# Stichprobengröße 1000; 20 Items
# daten <- list.files("~/Simulation_2015_CRSM/results/crsm_1000_daten")
x <- list.files("~/Simulation_2015_CRSM/results/crsm_1000_daten")
y <- list.files(path="~/Simulation_2015_CRSM/results/crsm_1000_res")
daten <- x[!(x %in% y)]

kern <- makeSOCKcluster(rep("localhost", 4))
parLapply(kern,daten,fun=dich.func.crsm,n=1000)
stopCluster(kern)

#-----------------------------------------------------------------------------------------
# ERSTELLUNG DATENMATRIX
#-----------------------------------------------------------------------------------------
# Stichprobengröße 1000; 20 Items
filenames <- list.files(path="~/Simulation_2015_CRSM/results/crsm_1000_res")
res <-
do.call("cbind",lapply(paste("~/Simulation_2015_CRSM/results/crsm_1000_res/",filenames,s
ep=""),read.table))
res <- t(res)
```

```
colnames(res) <-
c("KorrelationPp_p.1d","KorrelationPp_p.2d","KorrelationPp_p.12","KorrelationPp_dc",
  "KorrelationPp_p.1c","KorrelationPp_p.2c","LRTvalue.RM","LRTpvalue.RM",
  "LRTvalue.PCM.1","LRTpvalue.PCM.1","LRTvalue.PCM.2","LRTpvalue.PCM.2",
  "ItemFitRM","ItemFitPCM.1","ItemFitPCM.2","PersonenFitRM","PersonenFitPCM.1",
  "PersonenFitPCM.2")
rownames(res) <- NULL
write.table(res, file =paste("~/Simulation_2015_CRSM/results/crsm_1000_table",sep=""))
```

```
#++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
# AUSWERTUNG ZUM PARTIAL-CREDIT-MODELL
#++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
#----------------------------------------------------------------------------------------------
# Auswertungsskript Dichotomisierung PCM-Daten (n=1000)
#----------------------------------------------------------------------------------------------
# Korrelationen
#----------------------------------------------------------------------------------------------
pcm_1000_20_4_table <-
read.table("C:/Isabell/Uni/Psychologie/Diplomarbeit_3/Auswertung_04_2015/Daten/pcm_1000_
20_4_table.txt",sep="")

korr_1000_20_4 <- cbind(rbind("4 Kat.; TK < 1;"=summary(pcm_1000_20_4_table[,1]),"4
Kat.; TK < 2;"=c(summary(pcm_1000_20_4_table[,2]),0),"4 Kat.; TK < 3;
"=c(summary(pcm_1000_20_4_table[,3]),0)),"sd"=c(sd(pcm_1000_20_4_table[,1],na.rm=TRUE),
sd(pcm_1000_20_4_table[,2], na.rm=TRUE),sd(pcm_1000_20_4_table[,3])))

pcm_1000_20_5_table <-
read.table("C:/Isabell/Uni/Psychologie/Diplomarbeit_3/Auswertung_04_2015/Daten/pcm_1000_
20_5_table.txt",sep="")

korr_1000_20_5 <- cbind(rbind("5 Kat.; TK < 1;"=summary(pcm_1000_20_5_table[,1]),"5
Kat.; TK < 2;"=summary(pcm_1000_20_5_table[,2]),"5 Kat.; TK <
4;"=summary(pcm_1000_20_5_table[,3])),"sd"=c(sd(pcm_1000_20_5_table[,1],na.rm=TRUE),sd(p
cm_1000_20_5_table[,2],na.rm=TRUE),sd(pcm_1000_20_5_table[,3],na.rm=TRUE)))

pcm_1000_20_6_table <-
read.table("C:/Isabell/Uni/Psychologie/Diplomarbeit_3/Auswertung_04_2015/Daten/pcm_1000_
20_6_table.txt",sep="")

korr_1000_20_6 <- cbind(rbind("6 Kat.; TK < 2;"=c(summary(pcm_1000_20_6_table[,1]),0),"6
Kat.; TK < 3;"=c(summary(pcm_1000_20_6_table[,2]),0),"6 Kat.; TK <
4;"=summary(pcm_1000_20_6_table[,3])),"sd"=c(sd(pcm_1000_20_6_table[,1],na.rm=TRUE),sd(p
cm_1000_20_6_table[,2]),sd(pcm_1000_20_6_table[,3],na.rm=TRUE)))

#----------------------------------------------------------------------------------------------
# LRT
#----------------------------------------------------------------------------------------------
lrt_1000_20_4 <- rbind(c(sum(pcm_1000_20_4_table[,4] < 0.05, na.rm=TRUE),
sum(is.na(pcm_1000_20_4_table[,4]))),c(sum(pcm_1000_20_4_table[,5] < 0.05, na.rm=TRUE),
sum(is.na(pcm_1000_20_4_table[,5]))),c(sum(pcm_1000_20_4_table[,6] < 0.05, na.rm=TRUE),
sum(is.na(pcm_1000_20_4_table[,6]))),c(sum(pcm_1000_20_4_table[,7] < 0.05, na.rm=TRUE),
sum(is.na(pcm_1000_20_4_table[,7]))))
colnames(lrt_1000_20_4) <- c("sign. P-Werte","Anzahl-Abbr\"uche")

rownames(lrt_1000_20_4) <- c("mehrkateg. Daten","dich. Daten TK < 1","dich. Daten TK <
2","dich. Daten TK < 3")

lrt_1000_20_5 <- rbind(c(sum(pcm_1000_20_5_table[,4] < 0.05, na.rm=TRUE),
sum(is.na(pcm_1000_20_5_table[,4]))),c(sum(pcm_1000_20_5_table[,5] < 0.05, na.rm=TRUE),
sum(is.na(pcm_1000_20_5_table[,5]))),c(sum(pcm_1000_20_5_table[,6] < 0.05, na.rm=TRUE),
sum(is.na(pcm_1000_20_5_table[,6]))),c(sum(pcm_1000_20_5_table[,7] < 0.05, na.rm=TRUE),
sum(is.na(pcm_1000_20_5_table[,7]))))

colnames(lrt_1000_20_5) <- c("sign. P-Werte","Anzahl-Abbr\"uche")

rownames(lrt_1000_20_5) <- c("mehrkateg. Daten","dich. Daten TK < 1","dich. Daten TK <
2","dich. Daten TK < 4")

lrt_1000_20_6 <- rbind(c(sum(pcm_1000_20_6_table[,4] < 0.05, na.rm=TRUE),
sum(is.na(pcm_1000_20_6_table[,4]))),c(sum(pcm_1000_20_6_table[,5] < 0.05, na.rm=TRUE),
sum(is.na(pcm_1000_20_6_table[,5]))),c(sum(pcm_1000_20_6_table[,6] < 0.05, na.rm=TRUE),
sum(is.na(pcm_1000_20_6_table[,6]))),c(sum(pcm_1000_20_6_table[,7] < 0.05, na.rm=TRUE),
sum(is.na(pcm_1000_20_6_table[,7]))))

colnames(lrt_1000_20_6) <- c("sign. P-Werte","Anzahl-Abbr\"uche")

rownames(lrt_1000_20_6) <- c("mehrkateg. Daten","dich. Daten TK < 2","dich. Daten TK <
3","dich. Daten TK < 4")

#----------------------------------------------------------------------------------------------
# Item Fit-Indizes
#----------------------------------------------------------------------------------------------
ifit_1000_20_4 <- rbind("mehrkateg.
```

```
Daten"=c(table(pcm_1000_20_4_table[,8]),0,0,0,0,0),"dich. Daten TK <
1;"=c(table(pcm_1000_20_4_table[,9]),0),"dich. Daten TK <
2;"=c(table(pcm_1000_20_4_table[,10])),"dich. Daten TK <
3;"=c(table(pcm_1000_20_4_table[,11]),0))
colnames(ifit_1000_20_4) <- c("0","1","2","3","4","5","6","7","8")

ifit_1000_20_5 <- rbind("mehrkateg.
Daten"=c(table(pcm_1000_20_5_table[,8]),0,0,0),"dich. Daten TK <
2;"=table(pcm_1000_20_5_table[,9]),"dich. Daten TK <
3;"=table(pcm_1000_20_5_table[,10]),"dich. Daten TK <
4;"=table(pcm_1000_20_5_table[,11]))

colnames(ifit_1000_20_5) <- c("0","1","2","3","4","5","6","7")

ifit_1000_20_6 <-
rbind("mehrkateg.Daten"=c(table(pcm_1000_20_6_table[,8]),0,0,0,0),"dich. Daten TK <
2;"=table(pcm_1000_20_6_table[,9]),"dich. Daten TK <
3;"=c(table(pcm_1000_20_6_table[,10]),0),"dich. Daten TK <
4;"=table(pcm_1000_20_6_table[,11]))

colnames(ifit_1000_20_6) <- c("0","1","2","3","4","5","6","7")

#-------------------------------------------------------------------------------------
# Personen Fit-Indizes
#-------------------------------------------------------------------------------------
## 1000; 20 & 4
pfit_1000_20_4 <- rbind("mehrkateg. Daten"=table(cut(pcm_1000_20_4_table[,12],
breaks=c(0,20,40,60,80,100,120,140), include.lowest=FALSE)),"dich. Daten TK <
1;"=table(cut(pcm_1000_20_4_table[,13], breaks=c(0,20,40,60,80,100,120,140),
include.lowest=FALSE)),"dich. Daten TK < 2;"=table(cut(pcm_1000_20_4_table[,14],
breaks=c(0,20,40,60,80,100,120,140), include.lowest=FALSE)),"dich. Daten TK <
3;"=table(cut(pcm_1000_20_4_table[,15], breaks=c(0,20,40,60,80,100,120,140),
include.lowest=FALSE)))

pfit_1000_20_4 <- cbind(rbind(sum(pcm_1000_20_4_table[,12] == 0),
sum(pcm_1000_20_4_table[,13] == 0),sum(pcm_1000_20_4_table[,14] == 0),
sum(pcm_1000_20_4_table[,15] == 0)),pfit_1000_20_4)

colnames(pfit_1000_20_4) <-
c("0","[1,20]","[21,40]","[41,60]","[61,80]","[81,100]","[101,120]","[121,140]")

## 1000; 20 & 5
pfit_1000_20_5 <- rbind("mehrkateg. Daten"=table(cut(pcm_1000_20_5_table[,12],
breaks=c(0,20,40,60,80,100,120,140), include.lowest=FALSE)),"dich. Daten TK <
1;"=table(cut(pcm_1000_20_5_table[,13], breaks=c(0,20,40,60,80,100,120,140),
include.lowest=FALSE)),"dich. Daten TK < 2;"=table(cut(pcm_1000_20_5_table[,14],
breaks=c(0,20,40,60,80,100,120,140), include.lowest=FALSE)),"dich. Daten TK <
4;"=table(cut(pcm_1000_20_5_table[,15], breaks=c(0,20,40,60,80,100,120,140),
include.lowest=FALSE)))

pfit_1000_20_5 <- cbind(rbind(sum(pcm_1000_20_5_table[,12] == 0),
sum(pcm_1000_20_5_table[,13] == 0),sum(pcm_1000_20_5_table[,14] == 0),
sum(pcm_1000_20_5_table[,15] == 0)),pfit_1000_20_5)

colnames(pfit_1000_20_5) <-
c("0","[1,20]","[21,40]","[41,60]","[61,80]","[81,100]","[101,120]","[121,140]")

## 1000; 20 & 6
pfit_1000_20_6 <- rbind("mehrkateg. Daten"=table(cut(pcm_1000_20_6_table[,12],
breaks=c(0,20,40,60,80,100,120,140), include.lowest=FALSE)),"dich. Daten TK <
2;"=table(cut(pcm_1000_20_6_table[,13], breaks=c(0,20,40,60,80,100,120,140),
include.lowest=FALSE)),"dich. Daten TK < 3;"=table(cut(pcm_1000_20_6_table[,14],
breaks=c(0,20,40,60,80,100,120,140), include.lowest=FALSE)),"dich. Daten TK <
4;"=table(cut(pcm_1000_20_6_table[,15], breaks=c(0,20,40,60,80,100,120,140),
include.lowest=FALSE)))


pfit_1000_20_6 <- cbind(rbind(sum(pcm_1000_20_6_table[,12] == 0),
sum(pcm_1000_20_6_table[,13] == 0),sum(pcm_1000_20_6_table[,14] == 0),
sum(pcm_1000_20_6_table[,15] == 0)),pfit_1000_20_6)
```

```
colnames(pfit_1000_20_6) <-
c("0","[1,20]","[21,40]","[41,60]","[61,80]","[81,100]","[101,120]","[121,140]")

#-----------------------------------------------------------------------------------------
# Informationskriterien
#-----------------------------------------------------------------------------------------
aic_1000_20_4 <- cbind("mehrkateg. Daten"=pcm_1000_20_4_table[,16],"dich. Daten TK <
1"=pcm_1000_20_4_table[,17],"dich. Daten TK < 2"=pcm_1000_20_4_table[,18],"dich. Daten
TK < 3"=pcm_1000_20_4_table[,19])

table(apply(aic_1000_20_4, 1, which.min))

bic_1000_20_4 <- cbind("mehrkateg. Daten"=pcm_1000_20_4_table[,20],"dich. Daten TK <
1"=pcm_1000_20_4_table[,21],"dich. Daten TK < 2"=pcm_1000_20_4_table[,22],"dich. Daten
TK < 3"=pcm_1000_20_4_table[,23])

table(apply(bic_1000_20_4, 1, which.min))

aic_1000_20_5 <- cbind("mehrkateg. Daten"=pcm_1000_20_5_table[,16],"dich. Daten TK <
1"=pcm_1000_20_5_table[,17],"dich. Daten TK < 2"=pcm_1000_20_5_table[,18],"dich. Daten
TK < 4"=pcm_1000_20_5_table[,19])

table(apply(aic_1000_20_5, 1, which.min))

bic_1000_20_5 <- cbind("mehrkateg. Daten"=pcm_1000_20_5_table[,20],"dich. Daten TK <
1"=pcm_1000_20_5_table[,21],"dich. Daten TK < 2"=pcm_1000_20_5_table[,22],"dich. Daten
TK < 4"=pcm_1000_20_5_table[,23])

table(apply(bic_1000_20_5, 1, which.min))

aic_1000_20_6 <- cbind("mehrkateg. Daten"=pcm_1000_20_6_table[,16],"dich. Daten TK <
2"=pcm_1000_20_6_table[,17],"dich. Daten TK < 3"=pcm_1000_20_6_table[,18],"dich. Daten
TK < 4"=pcm_1000_20_6_table[,19])

table(apply(aic_1000_20_6, 1, which.min))

bic_1000_20_6 <- cbind("mehrkateg. Daten"=pcm_1000_20_6_table[,20],"dich. Daten TK <
2"=pcm_1000_20_6_table[,21],"dich. Daten TK < 3"=pcm_1000_20_6_table[,22],"dich. Daten
TK < 4"=pcm_1000_20_6_table[,23])

table(apply(bic_1000_20_6, 1, which.min))

#-----------------------------------------------------------------------------------------
# Auswertungsskript Dichotomisierung PCM-Daten (N=2000)
#-----------------------------------------------------------------------------------------
filenames <-
list.files(path="C:/Isabell/Uni/Psychologie/Diplomarbeit_3/Auswertung_04_2015/Daten/pcm_
2000_20_4_res")
res <-
do.call("cbind",lapply(paste("C:/Isabell/Uni/Psychologie/Diplomarbeit_3/Auswertung_04_20
15/Daten/pcm_2000_20_4_res/",filenames,sep=""), read.table))

res <- t(res)

colnames(res) <-
c("KorrelationPp_u","KorrelationPp_m","KorrelationPp_o","LRTPCMpvalue","LRTRMpvalue_u","
LRTRMpvalue_m","LRTRMpvalue_o","ItemFitPCM","ItemFitRM_u","ItemFitRM_m","ItemFitRM_o",
"PersonenFitPCM","PersonenFitRM_u","PersonenFitRM_m","PersonenFitRM_o","AICPCM","AICRM_u
","AICRM_m","AICRM_o","BICPCM","BICRM_u","BICRM_m","BICRM_o")

rownames(res) <- NULL

write.table(res, file
=paste("C:/Isabell/Uni/Psychologie/Diplomarbeit_3/Auswertung_04_2015/pcm_2000_20_4_table
",sep=""))

#-----------------------------------------------------------------------------------------
# Korrelationen
#-----------------------------------------------------------------------------------------
pcm_2000_20_4_table <-
read.table("C:/Isabell/Uni/Psychologie/Diplomarbeit_3/Auswertung_04_2015/Daten/pcm_2000_
20_4_table",sep="")
korr_2000_20_4 <- cbind(rbind("4 Kat.; TK < 1;"=summary(pcm_2000_20_4_table[,1]),"4
Kat.; TK < 2;"=c(summary(pcm_2000_20_4_table[,2]),0),"4 Kat.; TK <
```

```
3;"=c(summary(pcm_2000_20_4_table[,3]),0)),"sd"=c(sd(pcm_2000_20_4_table[,1],na.rm=TRUE)
,sd(pcm_2000_20_4_table[,2], na.rm=TRUE),sd(pcm_2000_20_4_table[,3])))

pcm_2000_20_5_table <-
read.table("C:/Isabell/Uni/Psychologie/Diplomarbeit_3/Auswertung_04_2015/Daten/pcm_2000_
20_5_table.txt",sep="")

korr_2000_20_5 <- cbind(rbind("5 Kat.; TK < 1;"=summary(pcm_2000_20_5_table[,1]),"5
Kat.; TK < 2;"=summary(pcm_2000_20_5_table[,2]),"5 Kat.; TK <
4;"=c(summary(pcm_2000_20_5_table[,3]),0)),"sd"=c(sd(pcm_2000_20_5_table[,1],na.rm=TRUE)
,sd(pcm_2000_20_5_table[,2],na.rm=TRUE),sd(pcm_2000_20_5_table[,3],na.rm=TRUE)))

pcm_2000_20_6_table <-
read.table("C:/Isabell/Uni/Psychologie/Diplomarbeit_3/Auswertung_04_2015/Daten/pcm_2000_
20_6_table.txt",sep="")

korr_2000_20_6 <- cbind(rbind("6 Kat.; TK < 2;"=summary(pcm_2000_20_6_table[,1]),"6
Kat.; TK < 3;"=summary(pcm_2000_20_6_table[,2]),"6 Kat.; TK <
4;"=summary(pcm_2000_20_6_table[,3])),"sd"=c(sd(pcm_2000_20_6_table[,1],na.rm=TRUE),
sd(pcm_2000_20_6_table[,2]),sd(pcm_2000_20_6_table[,3],na.rm=TRUE)))

#-------------------------------------------------------------------------------------
# LRT
#-------------------------------------------------------------------------------------
lrt_2000_20_4 <- rbind(c(sum(pcm_2000_20_4_table[,4] < 0.05, na.rm=TRUE),
sum(is.na(pcm_2000_20_4_table[,4]))),c(sum(pcm_2000_20_4_table[,5] < 0.05, na.rm=TRUE),
sum(is.na(pcm_2000_20_4_table[,5]))),c(sum(pcm_2000_20_4_table[,6] < 0.05, na.rm=TRUE),
sum(is.na(pcm_2000_20_4_table[,6]))),c(sum(pcm_2000_20_4_table[,7] < 0.05, na.rm=TRUE),
sum(is.na(pcm_2000_20_4_table[,7]))))

colnames(lrt_2000_20_4) <- c("sign. P-Werte","Anzahl-Abbr\"uche")

rownames(lrt_2000_20_4) <- c("mehrkateg. Daten","dich. Daten TK < 1","dich. Daten TK <
2","dich. Daten TK < 3")

lrt_2000_20_5 <- rbind(c(sum(pcm_2000_20_5_table[,4] < 0.05, na.rm=TRUE),
sum(is.na(pcm_2000_20_5_table[,4]))),c(sum(pcm_2000_20_5_table[,5] < 0.05, na.rm=TRUE),
sum(is.na(pcm_2000_20_5_table[,5]))),c(sum(pcm_2000_20_5_table[,6] < 0.05, na.rm=TRUE),
sum(is.na(pcm_2000_20_5_table[,6]))),c(sum(pcm_2000_20_5_table[,7] < 0.05, na.rm=TRUE),
sum(is.na(pcm_2000_20_5_table[,7]))))

colnames(lrt_2000_20_5) <- c("sign. P-Werte","Anzahl-Abbr\"uche")

rownames(lrt_2000_20_5) <- c("mehrkateg. Daten","dich. Daten TK < 1","dich. Daten TK <
2","dich. Daten TK < 4")

lrt_2000_20_6 <- rbind(c(sum(pcm_2000_20_6_table[,4] < 0.05, na.rm=TRUE),
sum(is.na(pcm_2000_20_6_table[,4]))),c(sum(pcm_2000_20_6_table[,5] < 0.05, na.rm=TRUE),
sum(is.na(pcm_2000_20_6_table[,5]))),c(sum(pcm_2000_20_6_table[,6] < 0.05, na.rm=TRUE),
sum(is.na(pcm_2000_20_6_table[,6]))),c(sum(pcm_2000_20_6_table[,7] < 0.05, na.rm=TRUE),
sum(is.na(pcm_2000_20_6_table[,7]))))

colnames(lrt_2000_20_6) <- c("sign. P-Werte","Anzahl-Abbr\"uche")

rownames(lrt_2000_20_6) <- c("mehrkateg. Daten","dich. Daten TK < 2","dich. Daten TK <
3","dich. Daten TK < 4")

#-------------------------------------------------------------------------------------
# Item Fit-Indizes
#-------------------------------------------------------------------------------------
ifit_2000_20_4 <- rbind("mehrkateg.
Daten"=c(table(pcm_2000_20_4_table[,8]),0,0,0,0,0),"dich. Daten TK <
1;"=c(table(pcm_2000_20_4_table[,9]),0),"dich. Daten TK <
2;"=c(table(pcm_2000_20_4_table[,10])),"dich. Daten TK <
3;"=c(table(pcm_2000_20_4_table[,11]),0))
colnames(ifit_2000_20_4) <- c("0","1","2","3","4","5","6","7","8")

ifit_2000_20_5 <- rbind("mehrkateg.
Daten"=c(table(pcm_2000_20_5_table[,8]),0,0,0,0),"dich. Daten TK <
2;"=table(pcm_2000_20_5_table[,9]),"dich. Daten TK <
3;"=table(pcm_2000_20_5_table[,10]),"dich. Daten TK <
4;"=table(pcm_2000_20_5_table[,11]))

colnames(ifit_2000_20_5) <- c("0","1","2","3","4","5","6","7")
```

```
ifit_2000_20_6 <- rbind("mehrkateg.
Daten"=c(table(pcm_2000_20_6_table[,8]),0,0,0,0),"dich. Daten TK <
2;"=table(pcm_2000_20_6_table[,9]),"dich. Daten TK <
3;"=table(pcm_2000_20_6_table[,10]),"dich. Daten TK <
4;"=table(pcm_2000_20_6_table[,11]))

colnames(ifit_2000_20_6) <- c("0","1","2","3","4","5","6","7")

#-------------------------------------------------------------------------------------
# Personen Fit-Indizes
#-------------------------------------------------------------------------------------
## 2000; 20 & 4
pfit_2000_20_4 <- rbind("mehrkateg. Daten"=table(cut(pcm_2000_20_4_table[,12],
breaks=c(0,36,72,108,144,180,216,253), include.lowest=FALSE)),"dich. Daten TK <
1;"=table(cut(pcm_2000_20_4_table[,13], breaks=c(0,36,72,108,144,180,216,253),
include.lowest=FALSE)),"dich. Daten TK < 2;"=table(cut(pcm_2000_20_4_table[,14],
breaks=c(0,36,72,108,144,180,216,253), include.lowest=FALSE)),"dich. Daten TK <
3;"=table(cut(pcm_2000_20_4_table[,15], breaks=c(0,36,72,108,144,180,216,253),
include.lowest=FALSE)))

pfit_2000_20_4 <- cbind(rbind(sum(pcm_2000_20_4_table[,12] ==
0),sum(pcm_2000_20_4_table[,13] == 0),sum(pcm_2000_20_4_table[,14] ==
0),sum(pcm_2000_20_4_table[,15] == 0)),pfit_2000_20_4)

colnames(pfit_2000_20_4) <-
c("0","[1,36]","[37,72]","[73,108]","[109,144]","[145,180]","[181,216]","[217,253]")

## 2000; 20 & 5

pfit_2000_20_5 <- rbind("mehrkateg. Daten"=table(cut(pcm_2000_20_5_table[,12],
breaks=c(0,36,72,108,144,180,216,253), include.lowest=FALSE)),"dich. Daten TK <
1;"=table(cut(pcm_2000_20_5_table[,13], breaks=c(0,36,72,108,144,180,216,253),
include.lowest=FALSE)),"dich. Daten TK < 2;"=table(cut(pcm_2000_20_5_table[,14],
breaks=c(0,36,72,108,144,180,216,253), include.lowest=FALSE)),"dich. Daten TK <
4;"=table(cut(pcm_2000_20_5_table[,15], breaks=c(0,36,72,108,144,180,216,253),
include.lowest=FALSE)))

pfit_2000_20_5 <- cbind(rbind(sum(pcm_2000_20_5_table[,12] ==
0),sum(pcm_2000_20_5_table[,13] == 0),sum(pcm_2000_20_5_table[,14] ==
0),sum(pcm_2000_20_5_table[,15] == 0)),pfit_2000_20_5)

colnames(pfit_2000_20_5) <-
c("0","[1,36]","[37,72]","[73,108]","[109,144]","[145,180]","[181,216]","[217,253]")

## 2000; 20 & 6

pfit_2000_20_6 <- rbind("mehrkateg. Daten"=table(cut(pcm_2000_20_6_table[,12],
breaks=c(0,36,72,108,144,180,216,253), include.lowest=FALSE)),"dich. Daten TK <
2;"=table(cut(pcm_2000_20_6_table[,13], breaks=c(0,36,72,108,144,180,216,253),
include.lowest=FALSE)),"dich. Daten TK < 3;"=table(cut(pcm_2000_20_6_table[,14],
breaks=c(0,36,72,108,144,180,216,253), include.lowest=FALSE)),"dich. Daten TK <
4;"=table(cut(pcm_2000_20_6_table[,15], breaks=c(0,36,72,108,144,180,216,253),
include.lowest=FALSE)))

pfit_2000_20_6 <- cbind(rbind(sum(pcm_2000_20_6_table[,12] ==
0),sum(pcm_2000_20_6_table[,13] == 0),sum(pcm_2000_20_6_table[,14] ==
0),sum(pcm_2000_20_6_table[,15] == 0)),pfit_2000_20_6)

colnames(pfit_2000_20_6) <-
c("0","[1,36]","[37,72]","[73,108]","[109,144]","[145,180]","[181,216]","[217,253]")

#-------------------------------------------------------------------------------------
# Informationskriterien
#-------------------------------------------------------------------------------------
aic_2000_20_4 <- cbind("mehrkateg. Daten"=pcm_2000_20_4_table[,16],"dich. Daten TK <
1"=pcm_2000_20_4_table[,17],"dich. Daten TK < 2"=pcm_2000_20_4_table[,18],"dich. Daten
TK < 3"=pcm_2000_20_4_table[,19])
table(apply(aic_2000_20_4, 1, which.min))
```

```
bic_2000_20_4 <- cbind("mehrkateg. Daten"=pcm_2000_20_4_table[,20],"dich. Daten TK <
1"=pcm_2000_20_4_table[,21],"dich. Daten TK < 2"=pcm_2000_20_4_table[,22],"dich. Daten
TK < 3"=pcm_2000_20_4_table[,23])

table(apply(bic_2000_20_4, 1, which.min))

aic_2000_20_5 <- cbind("mehrkateg. Daten"=pcm_2000_20_5_table[,16],"dich. Daten TK <
1"=pcm_2000_20_5_table[,17],"dich. Daten TK < 2"=pcm_2000_20_5_table[,18],"dich. Daten
TK < 4"=pcm_2000_20_5_table[,19])

table(apply(aic_2000_20_5, 1, which.min))

bic_2000_20_5 <- cbind("mehrkateg. Daten"=pcm_2000_20_5_table[,20],"dich. Daten TK <
1"=pcm_2000_20_5_table[,21],"dich. Daten TK < 2"=pcm_2000_20_5_table[,22],"dich. Daten
TK < 4"=pcm_2000_20_5_table[,23])

table(apply(bic_2000_20_5, 1, which.min))

aic_2000_20_6 <- cbind("mehrkateg. Daten"=pcm_2000_20_6_table[,16],"dich. Daten TK <
2"=pcm_2000_20_6_table[,17],"dich. Daten TK < 3"=pcm_2000_20_6_table[,18],"dich. Daten
TK < 4"=pcm_2000_20_6_table[,19])

table(apply(aic_2000_20_6, 1, which.min))

bic_2000_20_6 <- cbind("mehrkateg. Daten"=pcm_2000_20_6_table[,20],"dich. Daten TK <
2"=pcm_2000_20_6_table[,21],"dich. Daten TK < 3"=pcm_2000_20_6_table[,22],"dich. Daten
TK < 4"=pcm_2000_20_6_table[,23])

table(apply(bic_2000_20_6, 1, which.min))
```

```
#+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
# AUSWERTUNG ZUM GRADED-RESPONSE-MODELL
#+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
#---------------------------------------------------------------------------------------------
# Auswertungsskript Dichotomisierung GRM-Daten (N=1000)
#---------------------------------------------------------------------------------------------
#---------------------------------------------------------------------------------------------
# Korrelationen
#---------------------------------------------------------------------------------------------
grm_1000_20_4_table <-
read.table("C:/Isabell/Uni/Psychologie/Diplomarbeit_3/Auswertung_04_2015/Daten/grm_1000_
20_4_table.txt",sep="")

korr_1000_20_4 <- cbind(rbind("4 Kat.; TK < 1;"=summary(grm_1000_20_4_table[,5]),"4
Kat.; TK < 2;"=summary(grm_1000_20_4_table[,6]),"4 Kat.; TK <
3;"=summary(grm_1000_20_4_table[,7])),"sd"=c(sd(grm_1000_20_4_table[,5],na.rm=TRUE),sd(g
rm_1000_20_4_table[,6], na.rm=TRUE),sd(grm_1000_20_4_table[,7])))

grm_1000_20_5_table <-
read.table("C:/Isabell/Uni/Psychologie/Diplomarbeit_3/Auswertung_04_2015/Daten/grm_1000_
20_5_table.txt",sep="")

korr_1000_20_5 <- cbind(rbind("5 Kat.; TK < 1;"=summary(grm_1000_20_5_table[,5]),"5
Kat.; TK < 2;"=summary(grm_1000_20_5_table[,6]),"5 Kat.; TK <
4;"=summary(grm_1000_20_5_table[,7])),"sd"=c(sd(grm_1000_20_5_table[,5],na.rm=TRUE),
sd(grm_1000_20_5_table[,6],na.rm=TRUE),sd(grm_1000_20_5_table[,7],na.rm=TRUE)))

grm_1000_20_6_table <-
read.table("C:/Isabell/Uni/Psychologie/Diplomarbeit_3/Auswertung_04_2015/Daten/grm_1000_
20_6_table.txt",sep="")

korr_1000_20_6 <- cbind(rbind("6 Kat.; TK < 2;"=summary(grm_1000_20_6_table[,5]),"6
Kat.; TK < 3;"=summary(grm_1000_20_6_table[,6]),"6 Kat.; TK <
4;"=summary(grm_1000_20_6_table[,7])),"sd"=c(sd(grm_1000_20_6_table[,5],na.rm=TRUE),
sd(grm_1000_20_6_table[,6]),sd(grm_1000_20_6_table[,7],na.rm=TRUE)))

#---------------------------------------------------------------------------------------------
# Konvergenzkriterium
#---------------------------------------------------------------------------------------------
ConvGRM_4 <- sum(grm_1000_20_4_table[,1]==0)
ConvGRM_u_4 <- sum(grm_1000_20_4_table[,2]==0)
ConvGRM_m_4 <- sum(grm_1000_20_4_table[,3]==0)
ConvGRM_o_4 <- sum(grm_1000_20_4_table[,4]==0)

ConvGRM_5 <- sum(grm_1000_20_5_table[,1]==0)
ConvGRM_u_5 <- sum(grm_1000_20_5_table[,2]==0)
ConvGRM_m_5 <- sum(grm_1000_20_5_table[,3]==0)
ConvGRM_o_5 <- sum(grm_1000_20_5_table[,4]==0)

ConvGRM_6 <- sum(grm_1000_20_6_table[,1]==0)
ConvGRM_u_6 <- sum(grm_1000_20_6_table[,2]==0)
ConvGRM_m_6 <- sum(grm_1000_20_6_table[,3]==0)
ConvGRM_o_6 <- sum(grm_1000_20_6_table[,4]==0)

#---------------------------------------------------------------------------------------------
# Modellgelstungstests
#---------------------------------------------------------------------------------------------
# M2
M2_1000_20_4 <- rbind(c(sum(grm_1000_20_4_table[,8] < 0.05, na.rm=TRUE),
sum(is.na(grm_1000_20_4_table[,8]))),c(sum(grm_1000_20_4_table[,9] < 0.05, na.rm=TRUE),
sum(is.na(grm_1000_20_4_table[,9]))),c(sum(grm_1000_20_4_table[,10] < 0.05, na.rm=TRUE),
sum(is.na(grm_1000_20_4_table[,10]))),c(sum(grm_1000_20_4_table[,11] < 0.05,
na.rm=TRUE), sum(is.na(grm_1000_20_4_table[,11]))))

colnames(M2_1000_20_4) <- c("sign. P-Werte","Anzahl-Abbr\"uche")

rownames(M2_1000_20_4) <- c("mehrkateg. Daten","dich. Daten TK < 1","dich. Daten TK <
2","dich. Daten TK < 3")

M2_1000_20_5 <- rbind(c(sum(grm_1000_20_5_table[,8] < 0.05, na.rm=TRUE),
sum(is.na(grm_1000_20_5_table[,8]))),c(sum(grm_1000_20_5_table[,9] < 0.05, na.rm=TRUE),
sum(is.na(grm_1000_20_5_table[,9]))),c(sum(grm_1000_20_5_table[,10] < 0.05, na.rm=TRUE),
sum(is.na(grm_1000_20_5_table[,10]))),c(sum(grm_1000_20_5_table[,11] < 0.05,
na.rm=TRUE), sum(is.na(grm_1000_20_5_table[,11]))))
```

```
colnames(M2_1000_20_5) <- c("sign. P-Werte","Anzahl-Abbr\"uche")

rownames(M2_1000_20_5) <- c("mehrkateg. Daten","dich. Daten TK < 1","dich. Daten TK <
2","dich. Daten TK < 3")

M2_1000_20_6 <- rbind(c(sum(grm_1000_20_6_table[,8] < 0.05, na.rm=TRUE),
sum(is.na(grm_1000_20_6_table[,8]))),c(sum(grm_1000_20_6_table[,9] < 0.05, na.rm=TRUE),
sum(is.na(grm_1000_20_6_table[,9]))),c(sum(grm_1000_20_6_table[,10] < 0.05, na.rm=TRUE),
sum(is.na(grm_1000_20_6_table[,10]))),c(sum(grm_1000_20_6_table[,11] < 0.05,
na.rm=TRUE), sum(is.na(grm_1000_20_6_table[,11]))))

colnames(M2_1000_20_6) <- c("sign. P-Werte","Anzahl-Abbr\"uche")

rownames(M2_1000_20_6) <- c("mehrkateg. Daten","dich. Daten TK < 1","dich. Daten TK <
2","dich. Daten TK < 3")

# X2

X2_1000_20_4 <- rbind(c(sum(grm_1000_20_4_table[,12] < 0.05, na.rm=TRUE),
sum(is.na(grm_1000_20_4_table[,12]))),c(sum(grm_1000_20_4_table[,13] < 0.05,
na.rm=TRUE), sum(is.na(grm_1000_20_4_table[,13]))),c(sum(grm_1000_20_4_table[,14] <
0.05, na.rm=TRUE), sum(is.na(grm_1000_20_4_table[,14]))),c(sum(grm_1000_20_4_table[,15]
< 0.05, na.rm=TRUE), sum(is.na(grm_1000_20_4_table[,15]))))

colnames(X2_1000_20_4) <- c("sign. P-Werte","Anzahl-Abbr\"uche")

rownames(X2_1000_20_4) <- c("mehrkateg. Daten","dich. Daten TK < 1","dich. Daten TK <
2","dich. Daten TK < 3")

X2_1000_20_5 <- rbind(c(sum(grm_1000_20_5_table[,12] < 0.05, na.rm=TRUE),
sum(is.na(grm_1000_20_5_table[,12]))),c(sum(grm_1000_20_5_table[,13] < 0.05,
na.rm=TRUE), sum(is.na(grm_1000_20_5_table[,13]))),c(sum(grm_1000_20_5_table[,14] <
0.05, na.rm=TRUE), sum(is.na(grm_1000_20_5_table[,14]))),c(sum(grm_1000_20_5_table[,15]
< 0.05, na.rm=TRUE), sum(is.na(grm_1000_20_5_table[,15]))))

colnames(X2_1000_20_5) <- c("sign. P-Werte","Anzahl-Abbr\"uche")

rownames(X2_1000_20_5) <- c("mehrkateg. Daten","dich. Daten TK < 1","dich. Daten TK <
2","dich. Daten TK < 3")

X2_1000_20_6 <- rbind(c(sum(grm_1000_20_6_table[,12] < 0.05, na.rm=TRUE),
sum(is.na(grm_1000_20_6_table[,12]))),c(sum(grm_1000_20_6_table[,13] < 0.05,
na.rm=TRUE), sum(is.na(grm_1000_20_6_table[,13]))),c(sum(grm_1000_20_6_table[,14] <
0.05, na.rm=TRUE), sum(is.na(grm_1000_20_6_table[,14]))),c(sum(grm_1000_20_6_table[,15]
< 0.05, na.rm=TRUE), sum(is.na(grm_1000_20_6_table[,15]))))

colnames(X2_1000_20_6) <- c("sign. P-Werte","Anzahl-Abbr\"uche")

rownames(X2_1000_20_6) <- c("mehrkateg. Daten","dich. Daten TK < 1","dich. Daten TK <
2","dich. Daten TK < 3")

# G2

G2_1000_20_4 <- rbind(c(sum(grm_1000_20_4_table[,16] < 0.05, na.rm=TRUE),
sum(is.na(grm_1000_20_4_table[,16]))),c(sum(grm_1000_20_4_table[,17] < 0.05,
na.rm=TRUE), sum(is.na(grm_1000_20_4_table[,17]))),c(sum(grm_1000_20_4_table[,18] <
0.05, na.rm=TRUE), sum(is.na(grm_1000_20_4_table[,18]))),c(sum(grm_1000_20_4_table[,19]
< 0.05, na.rm=TRUE), sum(is.na(grm_1000_20_4_table[,19]))))

colnames(G2_1000_20_4) <- c("sign. P-Werte","Anzahl-Abbr\"uche")

rownames(G2_1000_20_4) <- c("mehrkateg. Daten","dich. Daten TK < 1","dich. Daten TK <
2","dich. Daten TK < 3")

G2_1000_20_5 <- rbind(c(sum(grm_1000_20_5_table[,16] < 0.05, na.rm=TRUE),
sum(is.na(grm_1000_20_5_table[,16]))),c(sum(grm_1000_20_5_table[,17] < 0.05,
na.rm=TRUE), sum(is.na(grm_1000_20_5_table[,17]))),c(sum(grm_1000_20_5_table[,18] <
0.05, na.rm=TRUE), sum(is.na(grm_1000_20_5_table[,18]))),c(sum(grm_1000_20_5_table[,19]
< 0.05, na.rm=TRUE), sum(is.na(grm_1000_20_5_table[,19]))))

colnames(G2_1000_20_5) <- c("sign. P-Werte","Anzahl-Abbr\"uche")
```

```
rownames(G2_1000_20_5) <- c("mehrkateg. Daten","dich. Daten TK < 1","dich. Daten TK <
2","dich. Daten TK < 3")

G2_1000_20_6 <- rbind(c(sum(grm_1000_20_6_table[,16] < 0.05, na.rm=TRUE),
sum(is.na(grm_1000_20_6_table[,16]))),c(sum(grm_1000_20_6_table[,17] < 0.05,
na.rm=TRUE), sum(is.na(grm_1000_20_6_table[,17]))),c(sum(grm_1000_20_6_table[,18] <
0.05, na.rm=TRUE), sum(is.na(grm_1000_20_6_table[,18]))),c(sum(grm_1000_20_6_table[,19]
< 0.05, na.rm=TRUE), sum(is.na(grm_1000_20_6_table[,19]))))

colnames(G2_1000_20_6) <- c("sign. P-Werte","Anzahl-Abbr\"uche")

rownames(G2_1000_20_6) <- c("mehrkateg. Daten","dich. Daten TK < 1","dich. Daten TK <
2","dich. Daten TK < 3")

#-------------------------------------------------------------------------------------
# Informationskriterien
#-------------------------------------------------------------------------------------
aic_1000_20_4 <- cbind("mehrkateg. Daten"=grm_1000_20_4_table[,20],"dich. Daten TK <
1"=grm_1000_20_4_table[,21],"dich. Daten TK < 2"=grm_1000_20_4_table[,22],"dich. Daten
TK < 3"=grm_1000_20_4_table[,23])

table(apply(aic_1000_20_4, 1, which.min))

bic_1000_20_4 <- cbind("mehrkateg. Daten"=grm_1000_20_4_table[,24],"dich. Daten TK <
1"=grm_1000_20_4_table[,25],"dich. Daten TK < 2"=grm_1000_20_4_table[,26],"dich. Daten
TK < 3"=grm_1000_20_4_table[,27])

table(apply(bic_1000_20_4, 1, which.min))

aic_1000_20_5 <- cbind("mehrkateg. Daten"=grm_1000_20_5_table[,20],"dich. Daten TK <
1"=grm_1000_20_5_table[,21],"dich. Daten TK < 2"=grm_1000_20_5_table[,22],"dich. Daten
TK < 4"=grm_1000_20_5_table[,23])

table(apply(aic_1000_20_5, 1, which.min))

bic_1000_20_5 <- cbind("mehrkateg. Daten"=grm_1000_20_5_table[,24],"dich. Daten TK <
1"=grm_1000_20_5_table[,25],"dich. Daten TK < 2"=grm_1000_20_5_table[,26],"dich. Daten
TK < 4"=grm_1000_20_5_table[,27])

table(apply(bic_1000_20_5, 1, which.min))

aic_1000_20_6 <- cbind("mehrkateg. Daten"=grm_1000_20_6_table[,20],"dich. Daten TK <
2"=grm_1000_20_6_table[,21],"dich. Daten TK < 3"=grm_1000_20_6_table[,22],"dich. Daten
TK < 4"=grm_1000_20_6_table[,23])

table(apply(aic_1000_20_6, 1, which.min))

bic_1000_20_6 <- cbind("mehrkateg. Daten"=grm_1000_20_6_table[,24],"dich. Daten TK <
2"=grm_1000_20_6_table[,25],"dich. Daten TK < 3"=grm_1000_20_6_table[,26],"dich. Daten
TK < 4"=grm_1000_20_6_table[,27])

table(apply(bic_1000_20_6, 1, which.min))

#-------------------------------------------------------------------------------------
# Auswertungsskript Dichotomisierung GRM-Daten (N=2000)
#-------------------------------------------------------------------------------------
#-------------------------------------------------------------------------------------
# Korrelationen
#-------------------------------------------------------------------------------------
grm_2000_20_4_table <-
read.table("C:/Isabell/Uni/Psychologie/Diplomarbeit_3/Auswertung_04_2015/Daten/grm_2000_
20_4_table",sep="")

korr_2000_20_4 <- cbind(rbind("4 Kat.; TK < 1;"=summary(grm_2000_20_4_table[,5]),"4
Kat.; TK < 2;"=summary(grm_2000_20_4_table[,6]),"4 Kat.; TK <
3;"=summary(grm_2000_20_4_table[,7])),"sd"=c(sd(grm_2000_20_4_table[,5],na.rm=TRUE),sd(g
rm_2000_20_4_table[,6], na.rm=TRUE),sd(grm_2000_20_4_table[,7])))

grm_2000_20_5_table <-
read.table("C:/Isabell/Uni/Psychologie/Diplomarbeit_3/Auswertung_04_2015/Daten/grm_2000_
20_5_table",sep="")

korr_2000_20_5 <- cbind(rbind("5 Kat.; TK < 1;"=summary(grm_2000_20_5_table[,5]),"5
Kat.; TK < 2;"=summary(grm_2000_20_5_table[,6]),"5 Kat.; TK <
```

```
4;"=summary(grm_2000_20_5_table[,7])),"sd"=c(sd(grm_2000_20_5_table[,5],na.rm=TRUE),
sd(grm_2000_20_5_table[,6],na.rm=TRUE),sd(grm_2000_20_5_table[,7],na.rm=TRUE)))

grm_2000_20_6_table <-
read.table("C:/Isabell/Uni/Psychologie/Diplomarbeit_3/Auswertung_04_2015/Daten/grm_2000_
20_6_table",sep="")

korr_2000_20_6 <- cbind(rbind("6 Kat.; TK < 2;"=summary(grm_2000_20_6_table[,5]),"6
Kat.; TK < 3;"=summary(grm_2000_20_6_table[,6]),"6 Kat.; TK <
4;"=summary(grm_2000_20_6_table[,7])),"sd"=c(sd(grm_2000_20_6_table[,5],na.rm=TRUE),
sd(grm_2000_20_6_table[,6]),sd(grm_2000_20_6_table[,7],na.rm=TRUE)))

#-------------------------------------------------------------------------------------
# Konvergenzkriterium
#-------------------------------------------------------------------------------------
ConvGRM_4 <- sum(grm_2000_20_4_table[,1]==0)
ConvGRM_u_4 <- sum(grm_2000_20_4_table[,2]==0)
ConvGRM_m_4 <- sum(grm_2000_20_4_table[,3]==0)
ConvGRM_o_4 <- sum(grm_2000_20_4_table[,4]==0)

ConvGRM_5 <- sum(grm_2000_20_5_table[,1]==0)
ConvGRM_u_5 <- sum(grm_2000_20_5_table[,2]==0)
ConvGRM_m_5 <- sum(grm_2000_20_5_table[,3]==0)
ConvGRM_o_5 <- sum(grm_2000_20_5_table[,4]==0)

ConvGRM_6 <- sum(grm_2000_20_6_table[,1]==0)
ConvGRM_u_6 <- sum(grm_2000_20_6_table[,2]==0)
ConvGRM_m_6 <- sum(grm_2000_20_6_table[,3]==0)
ConvGRM_o_6 <- sum(grm_2000_20_6_table[,4]==0)

#-------------------------------------------------------------------------------------
# Modellgelstungstests
#-------------------------------------------------------------------------------------
# M2
M2_2000_20_4 <- rbind(c(sum(grm_2000_20_4_table[,8] < 0.05, na.rm=TRUE),
sum(is.na(grm_2000_20_4_table[,8]))),c(sum(grm_2000_20_4_table[,9] < 0.05, na.rm=TRUE),
sum(is.na(grm_2000_20_4_table[,9]))),c(sum(grm_2000_20_4_table[,10] < 0.05, na.rm=TRUE),
sum(is.na(grm_2000_20_4_table[,10]))),c(sum(grm_2000_20_4_table[,11] < 0.05,
na.rm=TRUE), sum(is.na(grm_2000_20_4_table[,11]))))

colnames(M2_2000_20_4) <- c("sign. P-Werte","Anzahl-Abbr\"uche")

rownames(M2_2000_20_4) <- c("mehrkateg. Daten","dich. Daten TK < 1","dich. Daten TK <
2","dich. Daten TK < 3")

M2_2000_20_5 <- rbind(c(sum(grm_2000_20_5_table[,8] < 0.05, na.rm=TRUE),
sum(is.na(grm_2000_20_5_table[,8]))),c(sum(grm_2000_20_5_table[,9] < 0.05, na.rm=TRUE),
sum(is.na(grm_2000_20_5_table[,9]))),c(sum(grm_2000_20_5_table[,10] < 0.05, na.rm=TRUE),
sum(is.na(grm_2000_20_5_table[,10]))),c(sum(grm_2000_20_5_table[,11] < 0.05,
na.rm=TRUE), sum(is.na(grm_2000_20_5_table[,11]))))

colnames(M2_2000_20_5) <- c("sign. P-Werte","Anzahl-Abbr\"uche")

rownames(M2_2000_20_5) <- c("mehrkateg. Daten","dich. Daten TK < 1","dich. Daten TK <
2","dich. Daten TK < 3")

M2_2000_20_6 <- rbind(c(sum(grm_2000_20_6_table[,8] < 0.05, na.rm=TRUE),
sum(is.na(grm_2000_20_6_table[,8]))),c(sum(grm_2000_20_6_table[,9] < 0.05, na.rm=TRUE),
sum(is.na(grm_2000_20_6_table[,9]))),c(sum(grm_2000_20_6_table[,10] < 0.05, na.rm=TRUE),
sum(is.na(grm_2000_20_6_table[,10]))),c(sum(grm_2000_20_6_table[,11] < 0.05,
na.rm=TRUE), sum(is.na(grm_2000_20_6_table[,11]))))

colnames(M2_2000_20_6) <- c("sign. P-Werte","Anzahl-Abbr\"uche")

rownames(M2_2000_20_6) <- c("mehrkateg. Daten","dich. Daten TK < 1","dich. Daten TK <
2","dich. Daten TK < 3")

# X2

X2_2000_20_4 <- rbind(c(sum(grm_2000_20_4_table[,12] < 0.05, na.rm=TRUE),
sum(is.na(grm_2000_20_4_table[,12]))),c(sum(grm_2000_20_4_table[,13] < 0.05,
na.rm=TRUE), sum(is.na(grm_2000_20_4_table[,13]))),c(sum(grm_2000_20_4_table[,14] <
0.05, na.rm=TRUE), sum(is.na(grm_2000_20_4_table[,14]))),c(sum(grm_2000_20_4_table[,15]
< 0.05, na.rm=TRUE), sum(is.na(grm_2000_20_4_table[,15]))))
```

```
colnames(X2_2000_20_4) <- c("sign. P-Werte","Anzahl-Abbr\"uche")

rownames(X2_2000_20_4) <- c("mehrkateg. Daten","dich. Daten TK < 1","dich. Daten TK <
2","dich. Daten TK < 3")

X2_2000_20_5 <- rbind(c(sum(grm_2000_20_5_table[,12] < 0.05, na.rm=TRUE),
sum(is.na(grm_2000_20_5_table[,12]))),c(sum(grm_2000_20_5_table[,13] < 0.05,
na.rm=TRUE), sum(is.na(grm_2000_20_5_table[,13]))),c(sum(grm_2000_20_5_table[,14] <
0.05, na.rm=TRUE), sum(is.na(grm_2000_20_5_table[,14]))),c(sum(grm_2000_20_5_table[,15]
< 0.05, na.rm=TRUE), sum(is.na(grm_2000_20_5_table[,15]))))

colnames(X2_2000_20_5) <- c("sign. P-Werte","Anzahl-Abbr\"uche")

rownames(X2_2000_20_5) <- c("mehrkateg. Daten","dich. Daten TK < 1","dich. Daten TK <
2","dich. Daten TK < 3")

X2_2000_20_6 <- rbind(c(sum(grm_2000_20_6_table[,12] < 0.05, na.rm=TRUE),
sum(is.na(grm_2000_20_6_table[,12]))),c(sum(grm_2000_20_6_table[,13] < 0.05,
na.rm=TRUE), sum(is.na(grm_2000_20_6_table[,13]))),c(sum(grm_2000_20_6_table[,14] <
0.05, na.rm=TRUE), sum(is.na(grm_2000_20_6_table[,14]))),c(sum(grm_2000_20_6_table[,15]
< 0.05, na.rm=TRUE), sum(is.na(grm_2000_20_6_table[,15]))))

colnames(X2_2000_20_6) <- c("sign. P-Werte","Anzahl-Abbr\"uche")

rownames(X2_2000_20_6) <- c("mehrkateg. Daten","dich. Daten TK < 1","dich. Daten TK <
2","dich. Daten TK < 3")

# G2

G2_2000_20_4 <- rbind(c(sum(grm_2000_20_4_table[,16] < 0.05, na.rm=TRUE),
sum(is.na(grm_2000_20_4_table[,16]))),c(sum(grm_2000_20_4_table[,17] < 0.05,
na.rm=TRUE), sum(is.na(grm_2000_20_4_table[,17]))),c(sum(grm_2000_20_4_table[,18] <
0.05, na.rm=TRUE), sum(is.na(grm_2000_20_4_table[,18]))),c(sum(grm_2000_20_4_table[,19]
< 0.05, na.rm=TRUE), sum(is.na(grm_2000_20_4_table[,19]))))

colnames(G2_2000_20_4) <- c("sign. P-Werte","Anzahl-Abbr\"uche")

rownames(G2_2000_20_4) <- c("mehrkateg. Daten","dich. Daten TK < 1","dich. Daten TK <
2","dich. Daten TK < 3")

G2_2000_20_5 <- rbind(c(sum(grm_2000_20_5_table[,16] < 0.05, na.rm=TRUE),
sum(is.na(grm_2000_20_5_table[,16]))),c(sum(grm_2000_20_5_table[,17] < 0.05,
na.rm=TRUE), sum(is.na(grm_2000_20_5_table[,17]))),c(sum(grm_2000_20_5_table[,18] <
0.05, na.rm=TRUE), sum(is.na(grm_2000_20_5_table[,18]))),c(sum(grm_2000_20_5_table[,19]
< 0.05, na.rm=TRUE), sum(is.na(grm_2000_20_5_table[,19]))))

colnames(G2_2000_20_5) <- c("sign. P-Werte","Anzahl-Abbr\"uche")

rownames(G2_2000_20_5) <- c("mehrkateg. Daten","dich. Daten TK < 1","dich. Daten TK <
2","dich. Daten TK < 3")

G2_2000_20_6 <- rbind(c(sum(grm_2000_20_6_table[,16] < 0.05, na.rm=TRUE),
sum(is.na(grm_2000_20_6_table[,16]))),c(sum(grm_2000_20_6_table[,17] < 0.05,
na.rm=TRUE), sum(is.na(grm_2000_20_6_table[,17]))),c(sum(grm_2000_20_6_table[,18] <
0.05, na.rm=TRUE), sum(is.na(grm_2000_20_6_table[,18]))),c(sum(grm_2000_20_6_table[,19]
< 0.05, na.rm=TRUE), sum(is.na(grm_2000_20_6_table[,19]))))

colnames(G2_2000_20_6) <- c("sign. P-Werte","Anzahl-Abbr\"uche")

rownames(G2_2000_20_6) <- c("mehrkateg. Daten","dich. Daten TK < 1","dich. Daten TK <
2","dich. Daten TK < 3")

#-------------------------------------------------------------------------------------
# Informationskriterien
#-------------------------------------------------------------------------------------
aic_2000_20_4 <- cbind("mehrkateg. Daten"=grm_2000_20_4_table[,20],"dich. Daten TK <
1"=grm_2000_20_4_table[,21],"dich. Daten TK < 2"=grm_2000_20_4_table[,22],"dich. Daten
TK < 3"=grm_2000_20_4_table[,23])

table(apply(aic_2000_20_4, 1, which.min))
```

```
bic_2000_20_4 <- cbind("mehrkateg. Daten"=grm_2000_20_4_table[,24],"dich. Daten TK <
1"=grm_2000_20_4_table[,25],"dich. Daten TK < 2"=grm_2000_20_4_table[,26],"dich. Daten
TK < 3"=grm_2000_20_4_table[,27])

table(apply(bic_2000_20_4, 1, which.min))

aic_2000_20_5 <- cbind("mehrkateg. Daten"=grm_2000_20_5_table[,20],"dich. Daten TK <
1"=grm_2000_20_5_table[,21],"dich. Daten TK < 2"=grm_2000_20_5_table[,22],"dich. Daten
TK < 4"=grm_2000_20_5_table[,23])

table(apply(aic_2000_20_5, 1, which.min))

bic_2000_20_5 <- cbind("mehrkateg. Daten"=grm_2000_20_5_table[,24],"dich. Daten TK <
1"=grm_2000_20_5_table[,25],"dich. Daten TK < 2"=grm_2000_20_5_table[,26],"dich. Daten
TK < 4"=grm_2000_20_5_table[,27])

table(apply(bic_2000_20_5, 1, which.min))

aic_2000_20_6 <- cbind("mehrkateg. Daten"=grm_2000_20_6_table[,20],"dich. Daten TK <
2"=grm_2000_20_6_table[,21],"dich. Daten TK < 3"=grm_2000_20_6_table[,22],"dich. Daten
TK < 4"=grm_2000_20_6_table[,23])

table(apply(aic_2000_20_6, 1, which.min))

bic_2000_20_6 <- cbind("mehrkateg. Daten"=grm_2000_20_6_table[,24],"dich. Daten TK <
2"=grm_2000_20_6_table[,25],"dich. Daten TK < 3"=grm_2000_20_6_table[,26],"dich. Daten
TK < 4"=grm_2000_20_6_table[,27])

table(apply(bic_2000_20_6, 1, which.min))
```

```
#+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
# AUSWERTUNG ZUM KONTINUIERLICHEN MODELL NACH MÜLLER
#+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
erg.dich.crsm <-
read.table("C:/Isabell/Uni/Psychologie/Diplomarbeit_3/Auswertung_04_2015/Daten/crsm_1000
_table.txt")

### Korrelation der Personenparameter
# Korrelationen der mehrkat. 1 (emp. Quantile) und dichotom Daten
korr_p.1d <- erg.dich.crsm[[1]]
summary(korr_p.1d)
sd(korr_p.1d, na.rm=TRUE)

# Korrelationen der mehrkat. 2 (theoret. Quantile) und dichotom Daten
korr_p.2d <- erg.dich.crsm[[2]]
summary(korr_p.2d)
sd(korr_p.2d, na.rm=TRUE)

# Korrelationen der mehrkat. 1 (emp. Quantile) und mehrkat. 2 (theoret. Quantile) Daten
korr_p.12 <- erg.dich.crsm[[3]]
summary(korr_p.12)
sd(korr_p.12, na.rm=TRUE)

### Korrelation der Personenparameter
# Korrelationen der kontinuierlichen und dichotom Daten
korr_dc <- erg.dich.crsm[[4]]
summary(korr_dc)
sd(korr_dc, na.rm=TRUE)

# Korrelationen der mehrkat. 1 (emp. Quantile) und kontinuierlichen Daten
korr_p.1c <- erg.dich.crsm[[5]]
summary(korr_p.1c)
sd(korr_p.1c, na.rm=TRUE)

# Korrelationen der mehrkat. 2 (emp. Quantile) und kontinuierlichen Daten
korr_p.2c <- erg.dich.crsm[[6]]
summary(korr_p.2c)
sd(korr_p.2c, na.rm=TRUE)

### Likelihood-Ratio Andersen
## signifikante P-Werte
lrtpvRM  <- erg.dich.crsm[[8]]
lrtpvPCM1 <- erg.dich.crsm[[10]]
lrtpvPCM2 <- erg.dich.crsm[[12]]
erg.lrt.pv <- rbind(c(sum(lrtpvRM < 0.05, na.rm=TRUE), sum(is.na(lrtpvRM))),
                    c(sum(lrtpvPCM1 < 0.05, na.rm=TRUE), sum(is.na(lrtpvPCM1))),
                    c(sum(lrtpvPCM2 < 0.05, na.rm=TRUE), sum(is.na(lrtpvPCM2))))
colnames(erg.lrt.pv) <- c("sign. P-Werte","Anzahl-Abbrüche")
rownames(erg.lrt.pv) <- c("dichotome Daten","mehrkateg. Daten (PCM 1)","mehrkateg. Daten
(PCM 2)")
erg.lrt.pv

### Fit-Indices
## Item Fit-Indices -> Anzahl signifikanter P-Werte
ifitrm  <- erg.dich.crsm[[13]]
table(ifitrm)

ifitpcm1 <- erg.dich.crsm[[14]]
table(ifitpcm1)

ifitpcm2 <- erg.dich.crsm[[15]]
table(ifitpcm2)

## Personen Fit-Indices -> Anzahl signifikanter P-Werte
pfitrm  <- erg.dich.crsm[[16]]
table(cut(pfitrm,breaks=c(0,5,10,15,21), include.lowest=FALSE))
sum(pfitrm==0)

pfitpcm1 <- erg.dich.crsm[[17]]
table(cut(pfitpcm1,breaks=c(0,5,10,15,21), include.lowest=FALSE))
sum(pfitpcm1==0)

pfitpcm2 <- erg.dich.crsm[[18]]
table(cut(pfitpcm2,breaks=c(0,5,10,15,21), include.lowest=FALSE))
sum(pfitpcm2==0)
```

7 Lebenslauf

Isabell Romana Baldauf

geboren am 14.01.1990 in Lienz

E-Mail: isabell.b@live.at

(Akademische) Ausbildung

Seit Oktober 2015	Magisterstudium der Statistik
Oktober 2012 - Juni 2015	Bachelorstudium der Statistik
Seit Mai 2015	Masterstudium Psychologie
März 2010 - Mai 2015	Bachelorstudium Psychologie
September 2004 - Juni 2009	HTBLVA Ferlach

Arbeitserfahrung

Seit Oktober 2014	Studienassistentin am Arbeitsbereich Forschungsmethoden der Psychologie an der Universität Wien
März 2013 - September 2013	Beraterin für Kinder- und Jugendliche bei „Rat auf Draht“
April 2012 - Juni 2012	Pflichtpraktikum (240h) am Arbeitsbereich Bildungspsychologie & Evaluation an der Universität Wien