



universität
wien

Diplomarbeit

Titel der Diplomarbeit

Eine experimentelle Studie zur Itemschwierigkeit von
Freiem Antwortformat sowie „1 aus 4“- „1 aus 6“-
und „x aus 5“-MC-Antwortformat

Verfasserin

Katharina Vock

Angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag. rer. nat.)

Wien, im Jänner 2009

Studienkennzahl :	A 298
Studienrichtung:	Psychologie
Betreuer:	Univ.-Prof. Mag. Dr. Klaus D. Kubinger

Diplomarbeit

Titel der Diplomarbeit

Eine experimentelle Studie zur Itemschwierigkeit von
Freiem Antwortformat sowie „1 aus 4“- , „1 aus 6“-
und „x aus 5“-MC-Antwortformat

Verfasserin

Katharina Vock

Angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag. rer. nat.)

Wien, im Jänner 2009

Studienkennzahl :	A 298
Studienrichtung:	Psychologie
Betreuer:	Univ.-Prof. Mag. Dr. Klaus D. Kubinger

Inhaltsverzeichnis

1.	Einführung	2
2.	Zielsetzung der vorliegenden Studie	11
3.	Methode	12
3.1.	Erhebungsinstrument, Entwicklung der Parallel- Items im MC-Antwortformat „1 aus 4“	12
3.2.	Stichprobe	16
3.3.	Testtheoretische Analysen der Items gemäß dem Rasch-Modell	18
3.4.	Hypothese der Studie und methodische Vorgehensweise der interferenzstatistischen Überprüfung der Hypothese	22
4.	Ergebnisse.....	24
4.1.	Ergebnisse der Analysen nach dem Rasch-Modell.....	24
4.1.1.	Graphische Modellkontrollen, Ergebnisse der LRTs – A-priori und A- posteriori sowie Übersicht über die Itemausschlüsse	24
4.1.2.	Zusammenfassende Betrachtung der Ergebnisse der Rasch-Modell- Analysen, Überblick über die Item-Ausschlüsse	52
4.2.	Ergebnisse der Hypothesentestung	53
4.2.1.	Ergebnisse der Varianzanalyse	53
4.2.2.	Ergebnisse des Post-Hoc-Tests	56
4.2.3.	Ergebnisse des t-tests für abhängige Stichproben.....	57
	Diskussion	59
5.	Zusammenfassung und Abstract zur Studie	62
5.1.	Zusammenfassung.....	62
5.2.	Abstract	63
6.	Literaturverzeichnis	65
7.	Anhang.....	67
7.1.	Anhang A: Itemleichtigkeitsparameter der 154 Rasch-Modell-konformen Items des DIC `08	67
7.2.	Anhang B: Testbogendeckblatt.....	70
7.3.	Anhang B: Die Items des Diagnostik-Info-Check `08 – dargestellt pro Skala und Testform	71
8.	Tabellenverzeichnis	100

9.	Abbildungsverzeichnis.....	106
----	----------------------------	-----

Sonderanhang: Lebenslauf der Autorin

1. Einführung

Hinsichtlich der Itemkonstruktion und –vorgabe bei psychologisch-diagnostischen Verfahren spielt neben dem Inhalt der Items auch deren Gestaltungsweise eine große Rolle, sie bestimmt, welche Möglichkeiten der Testperson als Reaktion auf eine Aufgabe zur Verfügung stehen. Dabei unterscheidet man grundsätzlich zwischen freiem (oder „offenem“) und gebundenem Antwortformat (auch „Multiple-Choice-Antwortformat“ genannt, im Folgenden mit „MC“ abgekürzt) (vgl. Seiwald, 2003). Während die Testperson beim freien Antwortformat die an sie gestellte Aufgabe eigenständig „frei“ formulieren muss, stehen ihr beim Multiple-Choice-Antwortformat mehrere Möglichkeiten als Reaktion zur Auswahl, von denen sie die passende (oder im Falle von Leistungstests die richtige) auszuwählen hat. Jedes dieser Antwortformate ist mit spezifischen Vor- und Nachteilen verbunden, die die Qualität des psychologisch-diagnostischen Verfahrens maßgeblich beeinflussen

Als Vorteil des freien Antwortformats wird oft der größere diagnostische Informationsgewinn genannt, der sich aus der Vielfalt der Möglichkeit für die Testperson, auf die Aufgabe zu reagieren, ergibt, während beim MC-Format der Handlungsspielraum auf die Wahl zwischen einigen wenigen Reaktionen eingeschränkt ist. Allerdings ist dieser Zusammenhang nicht zwingend, sondern vor allem bei Individualverfahren gegeben (Kubinger, 2006). Das freie Antwortformat eröffnet lediglich dem fachlich erfahrenen Testleiter im Idealfall den Zugang zu individuellen diagnostischen Informationen. Daraus ergibt sich allerdings auch der gravierendste Nachteil des offenen Antwortformats: Während die Verrechnungssicherheit bei MC-Tests stets gegeben ist, da sich aus den gewählten Antwortalternativen der Testperson eine Lösung zwingend und eindeutig als richtig oder falsch verrechnen lässt, liegt es beim freien Antwortformat immer auch im Ermessen des Testleiters, die sehr vielen möglichen unterschiedlichen Reaktionen der Testperson im Sinne von richtig oder falsch (bei Leistungstests) zu bewerten. Selbst wenn ein umfangreiches Kategorisierungssystem zur Auswertung der freien Antworten zur Verfügung steht, was z.B. für die Untertests 6, 9 und 11 des AID 2 vorliegt, können diese Kategoriensysteme logischerweise niemals erschöpfend sein, sodass dem Testleiter bei der Bewertung der Antwort ein gewisser Handlungsspielraum bleibt, der eine Einschränkung der Reliabilität der Messung mit sich bringt. Die Auswertung von MC-

Tests hat weiters den Vorteil, dass sie bei Vorhandensein einer entsprechend entwickelten Software die Auswertung mittels Computers ermöglicht. Dies erhöht nicht nur die Verrechnungssicherheit, sondern macht die Auswertung auch ökonomischer im Sinne eines geringeren Zeitaufwands. Die Entscheidung für ein Antwortformat sollte aber nie vorrangig von Ökonomieaspekten abhängen (weniger Aufwand für den Testleiter bei Testvorgabe und –auswertung), sondern stets aus fachlichen und inhaltlichen Überlegungen geschehen (vg. Kubinger, 2006). Schließlich geht aus Gedächtnisexperimenten hervor, dass das freie Reproduzieren von zuvor gelernten Wissensinhalten wesentlich schwieriger fällt als lediglich das Wiedererkennen von Lösungen, wie es bei MC-Tests verlangt wird (vgl. z. B. Zimbardo & Gerrig, 1999). Demnach stellt sich die Frage, ob mit den beiden Antwortformaten überhaupt noch dieselbe Fähigkeitsdimension erfasst wird. Jedenfalls schließt Kubinger (2006, S.125) daraus: „Wenn es darum geht, bestimmte (Wissens-) Anforderungen frei und selbständig zu leisten, dann ist das Multiple-Choice-Format verfehlt.“ Ebenfalls die Validität von Leistungstests beeinträchtigen kann die Tatsache, dass es zur Lösung von MC-Aufgaben grundsätzlich mehrere, qualitativ unterschiedliche Strategien gibt, zu einer Lösung zu gelangen. Neben der so genannten „Produktionsstrategie“ (vgl. Kubinger, 2006, S 126), bei der die Testperson die Lösung der Aufgabe zuerst im Kopf frei produziert und anschließend aus den zur Verfügung stehenden Antworten die entsprechende sucht, was zweifellos die am meisten herausfordernde und von Testautoren eigentlich intendierte Strategie ist, kann die Lösung auch nach dem Falsifikationsprinzip erfolgen. Hierbei schließt die Testperson ihr besonders unpassend erscheinende Distraktoren (falsche Antwortalternativen) sukzessive aus, bis die richtige oder eine kleine Menge richtiger bzw. wahrscheinlich richtiger Antwortmöglichkeiten übrig bleibt. Zwischen diesen Möglichkeiten hinsichtlich ihrer Schwierigkeit liegt die „Versuch-Irrtum-Strategie“, bei der die Testperson nacheinander die vorgegebenen Antwortmöglichkeiten auf ihre Richtigkeit überprüft und sich schließlich für die plausibelste entscheidet (Kubinger, 2006).

All das stellt eine Gefahr für die Reliabilität, Validität, sowie auch für die Skalierungsgüte eines Tests dar, spricht doch die Möglichkeit mehrerer Wege zur Lösung gegen die intendierte Eindimensionalität der Merkmalsmessung. (Seiwald, 2003; Kubinger, 2006). Schließlich bleibt der Testperson selbst bei nicht Vorhandensein der geforderten Fähigkeit die Möglichkeit, die richtige aus den

Antwortmöglichkeiten (zufällig) zu erraten. Auf diesen so genannten Rateeffekt von Leistungstests wird im folgenden Kapitel näher eingegangen.

Der Rateeffekt bei Leistungstests im MC-Antwortformat ist eine große, und gleichzeitig in der Praxis oft unterschätzte Gefahr für die Validität und Messgenauigkeit psychologisch-diagnostischer Verfahren. Die Wahrscheinlichkeit, eine Aufgabe zufällig richtig zu beantworten, ohne über die vom psychologisch-diagnostischen Verfahren zu messen beabsichtigte Fähigkeit zu verfügen, wird theoretische Ratewahrscheinlichkeit bzw. „a-priori-Ratewahrscheinlichkeit“ genannt (zum Begriff vgl. Kubinger, 2006). Sie ist bei MC-Tests stets größer als null und umso größer, je weniger Antwortmöglichkeiten im konkreten Fall geboten werden. Bei k Antwortmöglichkeiten und einer richtigen beträgt diese $1/k$, das bedeutet, dass im in der Praxis von Leistungstests gängigen Fall von $k = 5$ Antwortmöglichkeiten, von denen eine als richtig gilt, die anderen vier als Distraktoren (falsche Antwortmöglichkeiten) fungieren, die a-priori-Ratewahrscheinlichkeit bei 20% liegt. Besteht ein Test nun beispielsweise aus zehn solcher Items und wird ein Score von 50% gelöster Aufgaben als noch positiv bewertet, so beträgt die Wahrscheinlichkeit, den Test ohne bzw. mit zumindest sehr geringer Fähigkeit im gefragten Bereich zu bestehen – vorausgesetzt die Testperson sieht von der Möglichkeit ab, bei Unwissenheit oder im Zweifelsfall ehrlicherweise keine der Antwortmöglichkeiten zu wählen - gemäß der Binomialverteilung $0,032 = 3,2\%$! Dies veranschaulicht die massive Unfairness, die durch den Rateeffekt in Selektionssituationen entstehen kann. Jeder 16te , der über keine oder eine sehr geringe entsprechende Fähigkeit verfügt, würde den Test bestehen und somit ausgewählt werden, während jene, die durch Nachdenken zu ihrer Antwortauswahl gelangt sind, je nach Fähigkeit mitunter geringere Chancen hätten. Das ist aus wissenschaftlicher und praktischer Sicht nicht akzeptabel. Hinzu kommt die Tatsache, dass die faktische Ratewahrscheinlichkeit stets über der a-priori berechneten liegt, weil kaum alle Distraktoren gleich „attraktiv“ gestaltet werden und den Testpersonengleich plausibel erscheinen, so dass sich durch das sukzessive Ausscheiden unplausibler Antwortmöglichkeiten die Ratewahrscheinlichkeit eines Items auf bis zu 50% steigen kann (Kubinger, 2006, und Seiwald, 2003).

Man könnte entgegenbringen, dass die Ratewahrscheinlichkeit doch für alle Testpersonen gleichermaßen bestehe, sodass sich daraus keine Unfairness in Selektionssituationen ergebe. Dagegen sprechen allerdings zwei Argumente. Zum einen wären trotzdem die Validität der Messung und damit die Güte der Auswahl von Kandidaten massiv beeinträchtigt. Weiters ist nach Zimmermann (1984) anzunehmen, dass das Rate- oder Risikoverhalten von Testpersonen bei Leistungstests je nach Persönlichkeitseigenschaften sehr unterschiedlich sein kann (Kubinger, 2006).

Wie geht man nun als Autor eines MC-Tests mit dem Problem des Rateeffektes um? Grundsätzlich kann man dem Effekt auf zweierlei Arten begegnen: Schon bei der Konstruktion des Verfahrens durch inhaltliche bzw. formal-gestalterische Mittel, oder im Rahmen der Skalierung durch psychometrische Ansätze der probabilistischen Testtheorie.

Als triviale, den formal- gestalterischen Strategien zuzuordnende, sei die Erhöhung der Anzahl der Distraktoren erwähnt, um die a-priori-Ratewahrscheinlichkeit zu senken. Allerdings kann die Distraktorenzahl nicht beliebig erhöht werden, da zu viele Antwortmöglichkeiten Konzentration, Merkfähigkeit und auch Leistungsmotivation der Testperson beeinträchtigen würden.

Weiters wäre es möglich, die Anzahl der richtigen unter den Antwortmöglichkeiten nicht konstant bei eins zu belassen, beispielsweise bei einem Antwortformat nach dem Prinzip „1 aus 5“ (eine richtige ist unter fünf Antwortmöglichkeiten zu erkennen), sondern auf zwei bzw. mehr zu erhöhen („2 aus 5“) oder auch je nach Item zu variieren. Bei einem Format der Art „x aus 5“ z.B. können von 5 Antwortmöglichkeiten keine, eine, zwei, drei, vier, oder sogar alle fünf richtig sein. Um ein solches Item zu lösen, muss die Testperson alle richtigen Antwortmöglichkeiten wählen, und zusätzlich keine der falschen auswählen. Damit verringert sich die a-priori-Ratewahrscheinlichkeit für ein Item von $1/5$ im besprochenen Beispielfall auf $1/32$. Die Vorgabe von Items dieses Antwortformats bedarf allerdings einer genauen entsprechenden Instruktion an die Testpersonen, am besten anhand eines Beispiels erläutert (Kubinger, 2006). Auch die Einführung einer besonderen Antwortalternative der Art „keine Antwort ist richtig“ oder „ich weiß die Lösung nicht“ wird von Testautoren eingesetzt, um Raten zu verhindern (vgl. z.B. Gittler 1990). Erstere ist allerdings zumindest für mäßig testerfahrene Personen keine

realistische Alternative, bei zweiter könnte die Wirksamkeit wieder von Persönlichkeitsmerkmalen wie Risikoverhalten, Ehrlichkeit, Anspruchsniveau, und vor allem von den Konsequenzen abhängen, die mit den Testergebnissen verbunden sind (Kubinger, 2006). Handelt es sich bei dem psychologisch-diagnostischen Verfahren um einen Test, dessen Vorgabe am PC erfolgt, besteht auch die Möglichkeit, die Antwortalternativen sequentiell anstatt simultan vorzugeben, sodass die Testperson bei jeder aufscheinenden Antwortmöglichkeit sofort entscheiden muss, ob sie richtig oder falsch ist, ohne die nachfolgenden Antworten zu sehen oder der Möglichkeit, zurückzublättern. So verringert sich die Ratewahrscheinlichkeit bei einem Antwortformat der Art „1 aus 5“ von 20% sukzessive für die jeweilige Antwortalternative auf bis zu 6,25 bei der letzten. Allerdings ist diese Art der Vorgabe mit einer längeren Bearbeitungszeit verbunden und birgt die Gefahr von Reaktanz seitens der Testperson, wenn ihr nicht die Möglichkeit gegeben wird, einmal getroffene Entscheidungen zu korrigieren (Kubinger, 2003).

In den letzten Jahren erschienen Richtlinien zur Gestaltung von MC-Tests (Moreno, Martinez und Muniz, 2006, sowie Litzenberger, Gnambs und Punter, 2005), die sich sowohl auf die Formulierung der Fragen, als auch die Konstruktion der Antwortalternativen beziehen. Die Einhaltung der darin vorgeschlagenen Vorgehensweisen kann auch dazu beitragen, die Ratewahrscheinlichkeit gering (auf Höhe der a-priori-Ratewahrscheinlichkeit) zu halten. Moreno, Martinez und Muniz (2006) führen beispielsweise an, dass der Inhalt jeder Antwortalternative unabhängig von den übrigen sein soll. Litzenberger, Gnambs und Punter (2005) verweisen darauf, Deskriptoren wie „immer“, „nie“, „vollkommen“, „absolut“ etc. in Antwortformulierungen zu vermeiden, da sie fast immer ein Hinweis auf eine falsche Antwort sind.

Um Beeinträchtigungen der psychometrischen Qualität von Leistungstests durch das MC-Antwortformat auszuschalten, wurden im Rahmen der probabilistischen Testtheorie, auch „Item-Response“-Theorie (kurz: IRT) genannt, spezielle Modell-Ansätze entwickelt. Der Einsatz dieser Modelle zur Skalierung von Leistungstests berücksichtigt die Möglichkeit, auch ohne gewünschte Fähigkeit die Items durch Raten zu lösen. Die Fähigkeit einer Testperson wird so in Form des Personenparameters quasi „korrigiert“ um die geschätzte Ratewahrscheinlichkeit geschätzt.

Dazu zählen das 3-PL-Modell („dreiparametrisches logistisches Testmodell“ [vgl. z.B. Birnbaum, 1968 und Bond & Fox, 2001] und das so genannte „D+G PL-Modell“ (vgl. z. B. Kubinger & Draxler, 2006). Während beim klassischen Rasch-Modell („1-parametrisches logistisches Testmodell“, kurz: 1-PL-Modell, erstmalige Publikation von Rasch, 1960/1980) neben dem Fähigkeitsparameter bezogen auf die Items nur die Schätzung eines Item- oder Schwierigkeitsparameters¹ vorsieht, wird im 3-PL-Modell zusätzlich ein itemspezifischer Diskriminationsparameter sowie ein itemspezifischer Rateparameter geschätzt. Im Fall des D+G PL-Modells wird neben dem Schwierigkeitsparameter auch ein itemspezifischer Rateparameter berechnet, der Diskriminationsparameter allerdings für alle Items gleich angenommen und nicht geschätzt.

Ebenfalls zur Kontrolle von Rateeffekten können im Rahmen der IRT auch so genannte „person-fit-indices“ berechnet werden (vgl. z.B. Ponocny & Klauer, 2002). Dabei wird das Antwortmuster jeder Testperson daraufhin analysiert, inwieweit es im Hinblick auf das jeweilige Fähigkeitsniveau plausibel im Sinne von glaubwürdig ist. Sollte sich herausstellen, dass eine Testperson viele relativ einfache Aufgaben nicht gelöst hat, viele relativ schwierige hingegen schon, spricht das für einen schlechten „person-fit“ und gibt Anlass zur Annahme, dass die betreffende Person geraten hat. Allerdings deckt diese Schätzung nicht lückenlos jeden Fall von Raten auf, sondern nur jene, die „unglücklich“ raten, d.h. gerade bei jenen Items zufällig richtig raten, die einen höheren Schwierigkeit aufweisen.

Mit der praktischen Anwendung der Modelle der IRT sind aber einige Probleme verbunden: Simulationsstudien von Puchhammer (1989) zufolge fallen Schätzungen der Rateparameter an Stichproben von weniger als $n = 500$ Testpersonen zu ungenau aus (vgl. auch Kubinger, 2005). Außerdem geht beim 3-PL-Modell sowie beim D+G PL-Modell durch die zusätzliche Schätzung von Diskriminations- und Rate-, bzw. Rateparameter gegenüber dem 1PL- Modell die Möglichkeit spezifisch objektiver Vergleiche und somit die Eigenschaft der statistischen Stichprobenabhängigkeit verloren. Das hat zur Folge, dass diese Modelle nicht mehr direkt überprüfbar sind. Lediglich mit Hilfe so genannter „goodness-of-fit“- Indices kann man die Anpassung

¹ In der vorliegenden Arbeit wird von „Itemleichtigkeitsparametern“ gesprochen, da die Polung durch die Codierung so erfolgte, dass höhere Werte für geringere Itemschwierigkeiten stehen und umgekehrt.

der beobachteten Daten an das Modell feststellen (Kubinger, 2003). In der vorliegenden Untersuchung wird aus den eben genannten Gründen auf eine Schätzung von Rateparametern verzichtet. Dem Rateeffekt wird formal begegnet, indem unter anderem ein Antwortformat der Art „x aus 5“ zur Anwendung kommt, und in seiner Schwierigkeit mit herkömmlichen „1 aus 4“, „1 aus 6“ bzw. freiem Antwortformat verglichen wird.

Nur wenige Untersuchungen zu MC-Tests behandeln explizit das Problem des Rateeffektes. Sie beziehen sich entweder auf inhaltliche Aspekte dieses Antwortformats oder auf andere formale Aspekte (vgl. z. B. Skinner, 1999, oder Attali & Bar-Hillel, 2003).

Dabei stellt sich nach den in Abschnitt 2.2. genannten Überlegungen die Frage, ob die theoretisch vorgeschlagenen Methoden zur Verringerung des Rateeffektes auch in der praktischen Anwendung tatsächlich den erwarteten Effekt nach sich ziehen, sodass sich der Aufwand der Testautoren diesbezüglich lohnt.

Kubinger, Holocher-Ertl und Frebort (2006) führten eine Pilotstudie durch, um den Rateeffekt verschiedener (MC)-Antwortformate zu vergleichen. Inhaltlich handelte es sich um einen Mathematik-Kompetenztest („Mathematik – 8. Schulstufe“), der von einer sehr großen Stichprobe von insgesamt $n = 6482$ Schülern bearbeitet wurde. Er bestand aus 81 Rasch-Modell-konformen Items, die in verschiedenen Antwortformaten gestaltet waren: 47 Items waren in freiem Antwortformat konzipiert, 21 Items im MC-Antwortformat „2 aus 5“ (d.h. zwei Lösungen neben drei Distraktoren), sowie 13 Items im MC-Format „1 aus 6“ (d.h. eine Lösung neben fünf Distraktoren). Die durch Schätzungen nach dem Rasch-Modell (1-PL-Modell) resultierenden Schwierigkeitsparameter wurden als abhängige Variable herangezogen, um den (Rate-) Effekt des jeweiligen Antwortformats auf die Aufgabenschwierigkeit zu untersuchen. Interferenzstatistisch wurden die Unterschiede in den Schwierigkeitsparametern mittels einfacher Varianzanalyse geprüft. Sie erbrachte ein knapp nicht signifikantes Ergebnis ($\alpha = 0,05$, $p = 0,058$ $df = 3$), lässt jedoch einen statistischen Trend erkennen, der sich folgendermaßen beschreiben lässt: Items im MC-Format „2 aus 5“ erwiesen sich als nicht wesentlich einfacher als jene im freien Antwortformat. Aufgaben im

Antwortformat „1 aus 6“ hingegen wiesen geringere Schwierigkeiten auf, sowohl als jene im freien, als auch jene im MC- Antwortformat „2 aus 5“. Das spricht für einen Rateeffekt, der bei noch immer hoher a-priori-Ratewahrscheinlichkeit, wie sie das „1 aus 6“- Antwortformat besitzt, zum Tragen kommt. Kritisch zu dieser Studie haben bereits die Autoren selbst angemerkt, dass es für den Nachweis einer sinnvollen Effektgröße bei angemessenem Typ I- und Typ II-Risiko ($p = 0,05$ bzw. $p = 0,20$) einer größeren Anzahl von Items (mind. 44 pro Antwortformat) bedarf, was nicht realisiert werden konnte. Außerdem wurden inhaltlich verschiedene Aufgaben hinsichtlich ihrer Schwierigkeit verglichen, d.h. die Unterschiede in den Itemschwierigkeiten können nicht eindeutig auf das unterschiedliche Antwortformat zurückgeführt werden. So wäre es denkbar, dass bestimmte Aufgaben an sich aufgrund ihres Inhaltes oder ihrer Formulierung schwierig bzw. einfach sind.

Gottschall (2007) verwirklichte zu einer ähnlichen Fragestellung ein experimentelles Design. Er konstruierte zu den Aufgaben des DIC („Diagnostik.-Info-Checks '05“) Entsprechungs-Items, d.h. Aufgaben, die sich nur in ihrem Antwortformat, nicht aber in ihrem Aufgabeninhalt unterscheiden. So existierten ein Großteil der Items in mindestens zwei verschiedenen der drei eingesetzten Antwortformate „1 aus 6“, „x aus 5“ und freiem Antwortformat. Die Aufgaben wurden einer Stichprobe von insgesamt $n = 173$ Psychologie Studierenden bzw. praktizierenden Psychologinnen vorgegeben. Die Teilnehmer der Studie wurden randomisiert einer von zwei Parallelformen des DIC, bestehend aus je 150 Items, zugeteilt, die sie zu bearbeiten hatten. Nach Analysen der Items nach dem Rasch-Modell (1-PL-Modell) wurden nicht modell-konforme Items ausgeschieden. Die geschätzten Schwierigkeitsparameter der modellkonformen Items wurden für einen Schwierigkeitsvergleich der drei eingesetzten Antwortformate herangezogen. Auf die Frage auszuscheidender Items und Modellgültigkeit wird im Rahmen der Diskussion in Abschnitt 6. noch genauer eingegangen. Interferenzstatistisch wurden die Unterschiede jeweils mittels T-tests überprüft, als nominales α wurde 0,05 festgesetzt. Es ließ sich ein signifikanter Unterschied in der Schwierigkeit von Items im MC-Antwortformat „1 aus 6“ sowohl im Vergleich zu solchen im Format „x aus 5“ ($p \leq 0,001$; $df = 10$), als auch im Vergleich zum freien Antwortformat ($p = 0,013$; $df = 3$) feststellen: Items im Antwortformat „1 aus 6“ waren signifikant leichter. Das lässt auf deutliche Rateeffekte beim Einsatz von MC-Items im Format „1 aus 6“ schließen. Zwischen freiem Antwortformat und

Antwortformat „x aus 5“ konnte kein signifikanter Schwierigkeitsunterschied festgestellt werden ($p = 0,234$; $df = 9$).

2. Zielsetzung der vorliegenden Studie

In der vorliegenden Studie soll erstens versucht werden, die Ergebnisse von Gottschall (2007) zu replizieren, d.h. die Schwierigkeit von Items in drei verschiedenen Antwortformaten, nämlich freiem Antwortformat, MC-Format „1 aus 6“ und MC-Format „x aus 5“ soll im Rahmen eines Experimentes miteinander verglichen werden. Dies geschieht anhand der nach dem Rasch-Modell geschätzten Itemleichtigkeitsparameter der Entsprechungsisems, das sind solche Items, die sich nur im Antwortformat, nicht aber im Inhalt voneinander unterscheiden. Sollten sich Schwierigkeitsunterschiede zwischen den Entsprechungsisems zeigen, würde das auf das unterschiedliche Antwortformat zurückgeführt werden können. Weiters gelangten noch zusätzlich Items im MC-Format „1 aus 4“ zum Einsatz, wie sie außerhalb der Psychologischen Diagnostik in vielen gängig eingesetzten Leistungstests zu finden sind. Auf diesem Weg soll zweitens untersucht werden, ob sich durch die Erhöhung der a-priori-Ratewahrscheinlichkeit von $1/6$ auf $1/4$ tatsächlich noch signifikant geringere Itemschwierigkeiten ergeben, was gegen einen Einsatz solcher Antwortformate in Leistungstests sprechen würde.

3. Methode

Zur Umsetzung der in Kapitel 2 beschriebenen Untersuchungsziele wurde, wie bei Gottschall (2007), ein experimentelles Design gewählt. Zu den bereits in drei Antwortformaten bestehenden Items des Diagnostik-Info-Check '05 (Kubinger, 2006, im Folgenden kurz „DIC“ genannt) wurden Entsprechungs-Items im Antwortformat „1 aus 4“ konstruiert. Anschließend wurden diese Entsprechungs-Items zufällig zu fünf Testversionen A bis E zugeteilt, die somit jeweils 104 Items mit gleichem Aufgabeninhalt, nur unterschiedlichem Antwortformat enthielten. Insgesamt 312 Testpersonen bearbeiteten schließlich den DIC, wobei die Zuteilung zu einer der Testformen zufällig vorgenommen wurde. Nach erfolgter Datenerhebung wurden die einzelnen Skalen des DIC einer Analyse nach dem Rasch-Modell unterzogen, um zu gewährleisten, dass zumindest a-posteriori von Rasch-Modell-Konformität des Tests gesprochen werden kann. Aus dem resultierenden Itempool wurden die Itemleichtigkeitsparameter geschätzt und gingen als abhängige Variable in die Untersuchung ein. Die unabhängige Variable stellte das jeweilige Antwortformat dar. Interferenzstatistisch wurden die Unterschiede in den Itemschwierigkeiten mittels einfacher Varianzanalyse auf ihre Signifikanz geprüft. Im Falle eines signifikanten Ergebnisses wurde post hoc der Newman-Keuls-Test angewandt. Um speziell Schwierigkeitsunterschiede zwischen Items im MC-Format „1 aus 6“ und „1 aus 4“ zu vergleichen, wurden die Parameter jener Items, die in diesen beiden Antwortformaten vorkamen und dem Rasch-Modell entsprachen, einem T-test für abhängige Stichproben unterzogen.

3.1. Erhebungsinstrument, Entwicklung der Parallel- Items im MC-Antwortformat „1 aus 4“

Als Erhebungsinstrument diente der DIC. Es handelt sich hierbei um einen Leistungstest, der als Wissenscheck im Fachbereich Psychologische Diagnostik 2005 von Kubinger an der Universität Wien erstellt wurde und im Anhangteil seines Lehrbuches (Kubinger, 2006) erschienen ist. Er richtet sich primär an Psychologiestudierende der Universität Wien, aber auch an Praktiker, die vorwiegend

diagnostisch tätig sind. Der DIC besteht aus sechs a-priori-konzipierten Untertests, die verschiedene Teilbereiche der Psychologischen Diagnostik abdecken:

1. Grundlagenwissen
2. Statistik- und Testtheoriewissen
3. Verfahrensinventar-Kenntnis
4. Theoriewissen zur Differentiellen Psychologie
5. Verfahrensbezogenes Detailwissen
6. Up-To-Date

Die ursprüngliche Form des DIC (Kubinger, 2006) enthielt 105 verschiedene Items in drei verschiedenen Antwortformaten: freiem Antwortformat, MC-Format „x aus 5“, und MC-Format „1 aus 6“.

Gottschall (2007) konstruierte zu den einzelnen Items sogenannte „Entsprechungs-Items“, das sind Items, die sich nur im Antwortformat, nicht aber im Aufgabeninhalt unterscheiden, in einem jeweils anderen Antwortformat, sodass ein und dieselbe Aufgabe in verschiedenen Antwortformaten zum Einsatz kam. Damit sollte sichergestellt werden, dass die ev. gefundenen Itemschwierigkeiten tatsächlich auf das Antwortformat zurückzuführen sind, und nicht vom Iteminhalt abhängen.

Im Rahmen der vorliegenden Arbeit wurden zu einem Großteil der Items von Gottschall (2007) noch Entsprechungs-Items im MC-Format „1 aus 4“ konstruiert, sodass schließlich folgende vier verschiedene Antwortformate in den fünf Testformen A bis E zum Einsatz kamen:

Das freie Antwortformat:

Hier werden die Wissensfragen offen gestellt, die Testperson muss die richtige Antwort selbständig finden und in Stichworten ausformulieren, ohne dass bestimmte Antwortmöglichkeiten vorgegeben sind. Die genauen Items finden sich im Anhang B.

Ein Multiple-Choice Antwortformat der Art „1 aus 6“:

Hier wurden zur selben Frage sechs mögliche Antwortalternativen vorgegeben, von denen nur eine richtig ist, und die von der Testperson ausgewählt werden muss.

Ein Multiple-Choice Format der Art „x aus 5“:

Zu jeder Frage stehen fünf Antwortmöglichkeiten zur Verfügung, wobei die Testperson keinerlei Information darüber erhält, wie viele der Antwortmöglichkeiten richtig sind. Dass keine, eine, zwei, drei, vier, oder alle fünf Möglichkeiten richtig sein können, darüber werden die Testpersonen bei diesem Antwortformat selbstverständlich genau instruiert. Ein Item gilt nur dann als gelöst, wenn alle richtigen Antwortmöglichkeiten gewählt, und darüber hinaus keine der Distraktoren gewählt wird.

Ein Multiple-Choice Format der Art „1 aus 4“

Aus vier verschiedenen Antwortmöglichkeiten muss die eine richtige Alternative gefunden und markiert werden. Zusätzlich wurde ein Item (Item 74) wegen veralteten Inhalts aus sämtlichen Testformen genommen.

Schließlich existieren fünf Parallelversionen des DIC, die Formen A bis E. Die Versionen A und B enthalten Fragen im Format „1 aus 6“, „x aus 5“, sowie im freien Antwortformat, Version E im freien Antwortformat und im Format „1 aus 4“, Versionen C und D enthalten Fragen sämtlicher Antwortformate. Die genaue Zuteilung der Items in den jeweiligen Antwortformaten zu den einzelnen Testformen ist dem Anhang B zu entnehmen. Tabelle 1 liefert eine Übersicht über die Anzahl der Items je Antwortformatkategorie, jeweils dargestellt pro Skala des DIC sowie gesamt. Eine Gleichverteilung aller Antwortformate auf die fünf Testformen war aus inhaltlichen Gründen nicht möglich, da nicht jedes Item in jedes Antwortformat sinnvoll umzuwandeln war.

Tabelle 1: Anzahl der Items je Antwortformatkategorie , dargestellt pro Testform A- E des DIC sowie insgesamt

Antwortformatkategorie	Itemanzahl					
	A	B	C	D	E	gesamt
Freies Antwortformat	66	60	66	54	52	74
Multiple-Choice-Format	18	22	12	7	0	35

„1 aus 6“						
Multiple-Choice-Format „1 aus 4“	0	0	12	30	49	50
Multiple-Choice-Format „x aus 5“	20	22	14	13	3	39
Unterkategorie „0 aus 5“	1	5	-	5	-	
Unterkategorie „1 aus 5“	6	8	4	3	2	
Unterkategorie „2 aus 5“	7	8	5	5	1	
Unterkategorie „3 aus 5“	4	1	3	-	-	
Unterkategorie „4 aus 5“	2	-	2	-	-	
Unterkategorie „5 aus 5“	-	-	-	-	-	
gesamt	104	104	104	104	104	198

Zur Wiederholung sei erwähnt, dass jede dieser Versionen aus denselben inhaltlichen 104 Fragen besteht, sie unterscheiden sich lediglich im Antwortformat.

Tabelle 2 Gibt einen Überblick über die Zahl der inhaltlich verschiedenen Items pro Skala des DIC sowie gesamt.

Tabelle 2: Itemanzahl pro Skala des DIC sowie gesamt

Skala	Itemanzahl
	Versionen A ,B,C,D,E
Grundlagenwissen	30
Statistik- und Testtheorie-Wissen	15
Verfahrensinventar-Kenntnis	14
Theoriewissen zur Differentiellen Psychologie	15
Verfahrensbezogenes Detailwissen	15
Up-To-Date	15
Gesamt	104

Der DIC besteht insgesamt aus 104 solcher Items, von denen 30 der Skala „Grundlagenwissen“ zuzuordnen sind, 15 Items der Skala „Statistik- und Testtheorie-Wissen“, 14 der Skala „Verfahrensinventar-Kenntnis, und je 15 der Skala „Theoriewissen zur Differentiellen Psychologie“, „Verfahrensbezogenes Detailwissen“ und „Up-To-Date“.

3.2. Stichprobe

Der DIC wurde insgesamt $n = 312$ Psychologie Studierenden im Alter von 21 bis 52 Jahren vorgegeben, der Median der Altersverteilung lag bei 25 Jahren. Abbildung 1 gibt die Altersverteilung in der Stichprobe wieder.

Unter den Teilnehmern waren 44 männlich, 268 weiblich. Die ungleiche Geschlechterverteilung in der Stichprobe ist auf den generell höheren Frauenanteil Psychologiestudierender an der Universität Wien zurückzuführen.

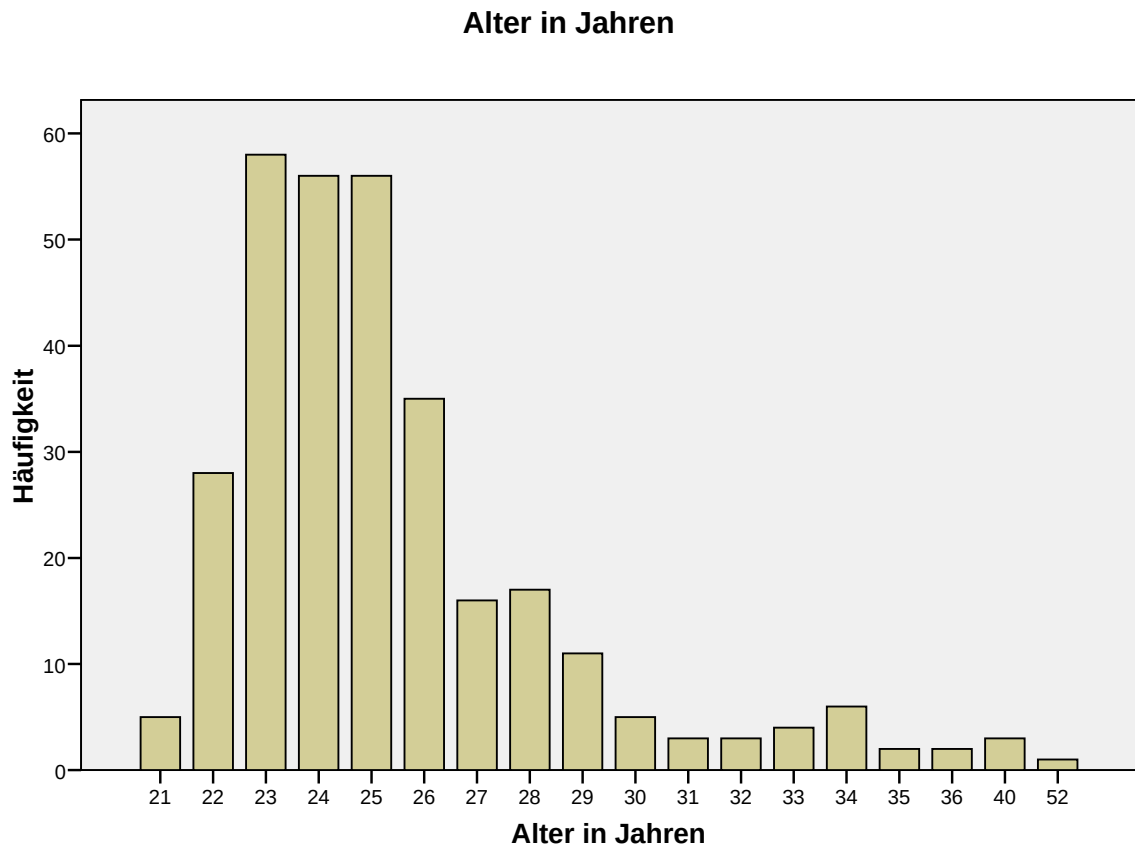


Abbildung 1: Histogramm der Altersverteilung der Testpersonen in der Stichprobe

Da die Bearbeitung des DIC Basiswissen im Bereich der Psychologischen Diagnostik voraussetzt, wurden als Versuchspersonen nur solche herangezogen, die die Grundausbildung in diesem Fach, bestehend aus den beiden Hauptvorlesungen zur *Psychologischen Diagnostik I und II* sowie den begleitenden *Übungen zur Psychologischen Diagnostik I und II* bereits absolviert hatten. Dazu wurde die Population jener Studierenden gewählt, welche sich im aktuellen Semester (WS 06/07 bzw. SS 08) in der Lehrveranstaltung *Demonstrationen zur Psychologischen Diagnostik* befanden, zu deren Teilnahme eine positiv absolvierte Grundausbildung Voraussetzung ist. Zusätzlich wurde der DIC im Wintersemester 2006/07 auch den Teilnehmern des *Forschungsseminars für Fortgeschrittene* unter der Leitung von Univ. Prof. Mag. Dr. Klaus Kubinger vorgegeben, die diese Lehrveranstaltung im Rahmen ihrer Diplomarbeit am Arbeitsbereich für Psychologische Diagnostik zu absolvieren hatten.

Die *Demonstrationen zur Psychologischen Diagnostik* werden pro Semester in sechs bis acht Parallelveranstaltungen zu je 25 Teilnehmern angeboten. Im Wintersemester 2006/7 wurden in diesen Lehrveranstaltungen sowie im Forschungsseminar insgesamt 142 Personen getestet, im Sommersemester 2008 170 Personen. Im Anhang A liefert Tabelle 4 einen Überblick über Anzahl und Prozentsatz der Versuchsteilnehmer pro Lehrveranstaltung.

Von den Testformen kam die Version A insgesamt 44 mal zum Einsatz, Form B 42 mal, 95 mal Version C, 90 mal Version D und 41 mal Version E. Die Zuteilung der Testpersonen zu den Testformen erfolgte randomisiert. Ein balanciertes Design diesbezüglich konnte jedoch nicht erreicht werden; von den ursprünglich gleich vielen ausgehändigten Testbögen pro Testform wurden nicht alle (ernsthaft) bearbeitet. Testbögen, bei denen weniger als 50% der Items bearbeitet wurden sowie solche, die offensichtlich unpassende, ironische Antworten enthielten und den Verdacht nahe legten, dass sie nicht ernsthaft bearbeitet wurden, wurden von der Auswertung ausgeschlossen. In Tabelle 3 sind Anzahl und Prozentsatz der Testpersonen pro Testform angegeben.

Tabelle 3: Überblick über die Anzahl getesteter Studierender pro Testform des DIC

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig Form A (Originalversion)	44	14,1	14,1	14,1
Form B (Parallelversion)	42	13,5	13,5	27,6
Form C (Parallelversion)	95	30,4	30,4	58,0
Form D (Parallelversion)	90	28,8	28,8	86,9
Form E (Parallelversion)	41	13,1	13,1	100,0
Gesamt	312	100,0	100,0	

3.3. Testtheoretische Analysen der Items gemäß dem Rasch-Modell

Da die meisten Items des DIC noch nicht hinsichtlich der Güte ihrer Skalierung untersucht wurden, mussten sie vor der eigentlichen Hypothesenprüfung einer Analyse nach dem Rasch-Modell unterzogen werden. Nur bei Geltung dieses Modells nämlich ist die Summe aller gelösten Items in einem Test ein faires Maß für die erbrachte Leistung darstellt, wie Fischer (1974 bzw.1995) dies in seinem so genannten „Notwendigkeitsbeweis“ feststellt.

Zusammengefasst beschreibt das Modell die Wahrscheinlichkeit, eine Aufgabe zu lösen, in Abhängigkeit von der Fähigkeit einer Person in der gefragten Dimension (Fähigkeitsparameter) und der Schwierigkeit der jeweiligen Aufgabe (Itemschwierigkeitsparameter). Der große Nutzen dieser Funktion liegt darin, dass mit ihr so genannte spezifisch-objektive Vergleiche möglich sind. Das bedeutet, die Fähigkeit einer Person kann eindeutig bestimmt werden, unabhängig davon, welche Items sie bearbeitet hat. Umgekehrt kann die Schwierigkeit jeder Aufgabe bestimmt werden, unabhängig davon, welche Person sie bearbeitet hat. Die Itemschwierigkeitsparameter sind somit stichprobenunabhängig. Dies ist Grundvoraussetzung für die Hypothesenprüfung in der vorliegenden Studie, denn nur so können etwaige Unterschiede der Itemschwierigkeiten paralleler Items (Items, die sich nur im Antwortformat, nicht aber im Aufgabeninhalt unterscheiden) auch eindeutig auf die verschiedenen Antwortformate zurückgeführt werden. Demnach werden nur jene Items zur Hypothesenprüfung herangezogen, welche sich in der vorhergegangenen Prüfung als Rasch-Modell-konform herausstellen.

Die Skalen des DIC betreffen thematisch unterschiedlichen Wissensbereichen der Psychologischen Diagnostik, weshalb die Geltung des Rasch-Modells für jede Skala getrennt überprüft wird. Das konkrete Vorgehen entspricht dabei der von Kubinger und Kubinger (2005) vorgeschlagenen Strategie. Zunächst wird jede Skala einem Likelihood-Ratio-Test von Andersen (im folgenden kurz LRT genannt) unterzogen, und auf Basis dessen Ergebnisses sowie einer graphischen Modellkontrolle und den z-Statistiken des Wald-Tests über die Gültigkeit des Rasch-Modells für den Itempool entschieden. Dazu wird die Gesamtstichprobe hinsichtlich eines bestimmten Kriteriums in zwei Teilstichproben geteilt. Der LRT prüft anschließend interferenzstatistisch, ob die Daten signifikant besser durch die in zwei separaten Teilstichproben geschätzten Itemparameter erklärt werden können als durch eine Schätzung auf Basis der Gesamtstichprobe. Ist dies der Fall, fällt der LRT signifikant aus, der Test misst nicht eindimensional im Sinne von in beiden Teilstichproben dieselbe Eigenschaft; das Rasch-Modell gilt nicht.

Fällt die Prüfung einer Skala auf Modellgültigkeit signifikant aus, wird anschließend mittels eines graphischen Modelltests und eines Wald-Tests ermittelt, welche Items im Speziellen den Annahmen des Rasch-Modells widersprechen. Im graphischen

Modelltest werden die für jede Teilstichprobe geschätzten Itemparameter in einem Koordinatensystem gegeneinander aufgetragen. Gilt das Rasch-Modell, das bedeutet, die Schätzungen der Itemschwierigkeiten sind in beiden Teilstichproben statistisch gleich, liegen die resultierenden Punkte der Items auf einer 45°-Geraden, jede Abweichung von dieser Geraden zeigt eine Abweichung des Items von den Annahmen der Eindimensionalität an.

In der vorliegenden Studie wird zunächst jede Skala einem LRT unterzogen. Fällt dieser signifikant aus, werden die im graphischen Modelltest bzw. aufgrund ihrer z-Statistik auffälligen Items sukzessive einzeln ausgeschlossen, und nach jedem Itemausschluss erneut ein LRT durchgeführt, beides für jedes der vier Teilungskriterien. Dieses Vorgehen wird solange wiederholt, bis jeder LRT hinsichtlich keines der herangezogenen Teilungskriterien mehr signifikant ist. So kann zumindest a-posteriori, das heißt nach Ausschluss einiger Items, Modellgültigkeit erreicht werden. Als Auswertungsprogramm zur Berechnung der LRTs, der Wald-Tests und der graphischen Modellkontrollen diente das Paket „eRm“ (extended Rasch modelling) in R (Poinstingl, Mair & Hatzinger, 2007).

Items, bei denen aufgrund zu extremer Lösungshäufigkeiten in der Stichprobe hinsichtlich keines der vier Teilungskriterien eine Schätzung des Itemparameters möglich war, wurden von der weiteren Betrachtung ausgeschlossen. Das bedeutet nicht, dass sie dem Rasch-Modell widersprechen, vielmehr ist die Überprüfung der Modellgültigkeit nicht möglich. Jene Aufgaben, bei denen die zweimalige Schätzung nicht hinsichtlich aller, aber hinsichtlich zumindest eines Teilungskriteriums möglich ist, werden im endgültigen Itempool beibehalten, sofern sie im graphischen Modelltest keine auffälligen Abweichungen zeigen.

Folgende Kriterien wurden zur Teilung der Gesamtstichprobe herangezogen:

Rohscore (niedriger Rohscore je betroffener Skala des DIC vs. Hoher Rohscore)

Der Rohscore stellt ein internes Teilungskriterium dar, geteilt wurde nach dem Median der Rohwertverteilung in der jeweiligen Skala des DIC.

Alter (bis 24 Jahre vs. Über 24 Jahre):

Hier wurde nach dem externen Kriterium des Medians der Altersverteilung in der Stichprobe (Md = 24) geteilt.

Note (schlechter vs. besser abgeschlossene Grundausbildung in Psychologischer Diagnostik)

Es handelt sich ebenfalls um ein externes Kriterium. Dazu wurden die Studenten der Stichprobe im DIC aufgefordert, ihre jeweiligen Noten der Prüfungen zur Psychologischen Diagnostik I und II sowie der beiden Übungen zur Psychologischen Diagnostik I und II anzugeben. Anschließend wurde pro Person der Mittelwert ihrer Leistungen in diesem Fach errechnet. Das Teilungskriterium stellte schließlich der Median der Notendurchschnittsverteilung (Md = 2,25) dar.

Prüfungsfach der geplanten bzw. bereits begonnenen Diplomarbeit (Psychologische Diagnostik, Differentielle Psychologie, oder Methodenlehre vs. alle anderen angebotenen Fächer des Psychologiestudiums an der Uni Wien)

Im DIC wurde hierzu gefragt, welches Fach die Versuchsperson für ihre Diplomarbeit gewählt hat bzw. (voraussichtlich) zu wählen beabsichtigt.

Die Fächer Psychologische Diagnostik sowie die ihr thematisch sehr nahen Fächer der Differentiellen Psychologie und der Methodenlehre bildeten dabei eine Gruppe, alle übrigen Prüfungsfächer des Diplomstudiums Psychologie die andere Gruppe.

Tabelle 4 veranschaulicht deskriptiv Häufigkeit und den Prozentsatz der Versuchspersonen jeder Teilstichprobe hinsichtlich der drei externen Teilungskriterien.

Tabelle 4: Anzahl und Prozent der Testpersonen pro Teilstichprobe in Bezug auf die Teilungskriterien *Alter*, *Note* und *Prüfungsfach*.

Teilungskriterium		Anzahl (%)
Alter	bis 24	147 (47,3)
	über 24	164 (52,7)
Diagnostik- Grundausbildung: (<i>Note</i>)	schlechter abgeschlossen	183 (60,6)
	besser abgeschlossen	119 (39,4)

Prüfungsfachkategorie:	Psychologische Diagnostik, Differentielle Psychologie und Methodenlehre	50 (18,1)
(Prüfungsfach)	andere Psychologische Prüfungsfächer	227 (81,9)

Der Median der Altersverteilung in der Stichprobe lag bei 24 Jahren; 47,3% der Testpersonen waren 24 Jahre alt oder jünger, 52,7% waren zum Zeitpunkt der Testung jünger als 24 Jahre. Ihre Diagnostik-Grundausbildung haben 39,4% der Testpersonen mit einem Notendurchschnitt von 2,25 oder besser abgeschlossen, die Durchschnittsleistungen von 60,6% lagen darunter. 277 der 312 Studienteilnehmer gaben im DIC ein Prüfungsfach an, in dem Sie ihre Diplomarbeit schreiben bzw. zu schreiben beabsichtigen. Lediglich 18,1% davon nannten Psychologische Diagnostik, Methodenlehre, oder Differentielle Psychologie als gewähltes Prüfungsfach, 81,9% eines der anderen Prüfungsfächer des Diplomstudiums Psychologie.

3.4. Hypothese der Studie und methodische Vorgehensweise der interferenzstatistischen Überprüfung der Hypothese

Folgende Hypothese wurde zur Untersuchung etwaiger Schwierigkeitsunterschiede zwischen den vier Antwortformaten überprüft (nominales $\alpha = 0,05$):

$H0a$: Es besteht kein Mittelwertsunterschied zwischen den Itemleichtigkeitsparametern der Entsprechungs-Items im Format „Freies Antwortformat“, Format „x aus 5“, Format „1 aus 6“ und Format „1 aus 4“.

$$\mu_1 = \mu_2 = \mu_3 = \mu_4$$

$H1a$: Aufgrund von Rateeffekten besteht ein Mittelwertsunterschied zwischen den Itemleichtigkeitsparametern der Entsprechungs-Items im Format „Freies Antwortformat“, Format „x aus 5“, Format „1 aus 6“ und Format „1 aus 4“.

Es existiert mindestens ein Paar (μ_x, μ_y) , für das gilt: $\mu_x \neq \mu_y$

Die Überprüfung der Nullhypothese erfolgte mittels einfacher Varianzanalyse.

Als unabhängige Variable diente dabei das jeweilige Antwortformat der Items, als abhängige Variable die mittels Rasch- Modell geschätzten Itemleichtigkeitsparameter. Zur Berechnung wurde SPSS in der Version 15.00 herangezogen.

Um näher zu untersuchen, zwischen welchen Antwortformaten jeweils signifikante Unterschiede in den Itemleichtigkeitsparametern bestehen, wurden die Ergebnisse der Varianzanalyse post- hoc einer Newman-Keuls-Prozedur unterzogen.

Als Zusatzhypothese wurde die Frage untersucht, ob sich die beiden MC-Antwortformate „1 aus 4“ und „1 aus 6“ signifikant hinsichtlich ihrer Schwierigkeit voneinander unterscheiden.

H0b: Es besteht kein Mittelwertsunterschied zwischen den Leichtigkeitsparametern der Items im Format „1 aus 4“ bzw. die Leichtigkeitsparameter der Items im Format „1 aus 4“ sind signifikant höher als jene im Format „1 aus 6“.

$$\mu_1 \geq \mu_2$$

H1b: Aufgrund der höheren a-priori-Ratewahrscheinlichkeit ist der Mittelwert der Itemleichtigkeitsparameter im MC-Format „1 aus 4“ signifikant höher als im Format „1 aus 6“.

$$\mu_1 > \mu_2$$

Um die Zusatzhypothese der Schwierigkeitsunterschiede zwischen diesen beiden Antwortformaten zu prüfen, wurden jene Entsprechungs-Items, die in beiden MC-Antwortformaten „1 aus 6“ und „1 aus 4“ vorgegeben wurden und dem Rasch-Modell entsprachen, gesondert mittels t-tests bei abhängigen Stichproben untersucht. Die Berechnung erfolgte ebenfalls mit dem Programm SPSS in der Version 15.

4. Ergebnisse

Abschnitt 4.1. widmet sich den Ergebnissen der Analysen nach dem Rasch-Modell, Abschnitt 4.2. den Ergebnissen der eigentlichen Hypothesentestungen der Studie.

4.1. Ergebnisse der Analysen nach dem Rasch-Modell

Im Folgenden werden die Ergebnisse der Rasch-Modellanalysen in Form von Graphischen Modelltests und Likelihood-Ratio-Tests nach Andersen (kurz „LRTs“ genannt), dargestellt. Das α wurde bei 0,01 angesetzt. Weiters werden für jede Skala die einzelnen Itemausschlüsse angeführt, und geordnet nach Antwortformat in Tabelle 9 sowie Tabelle 26 dargestellt.

4.1.1. Graphische Modellkontrollen, Ergebnisse der LRTs – A-priori und A- posteriori sowie Übersicht über die Itemausschlüsse

Um sicherzustellen, dass der endgültige Itempool nur noch Items enthält, deren Leichtigkeitsparameter frei von stichprobenabhängigen Verzerrungen sind, wurden getrennt für jede Skala des DIC LRTs sowie graphische Modellkontrollen durchgeführt, deren Ergebnisse im Folgenden beschrieben sind.

Skala Grundlagenwissen

Die Skala *Grundlagenwissen* umfasste a-priori 65 Items. Bei signifikantem LRT stellten sich im graphischen Modelltest 6 dieser Items in Bezug auf mindestens ein Teilungskriterium als nicht Rasch-Modell- konform heraus; sie wurden somit ausgeschlossen. Eine genaue Übersicht über die einzelnen ausgeschlossenen Items bietet **Fehler! Verweisquelle konnte nicht gefunden werden.** bzw. Anhang B. Ein weiteres Item wies derart extreme Lösungshäufigkeiten auf, dass in keinem einzigen Teilungskriterium eine Parameterschätzungen möglich war, dieses wurde ebenfalls von allen weiteren Berechnungen ausgeschlossen. Die LRTs fielen a-priori in Bezug auf sämtliche Teilungskriterien mit Ausnahme des Kriteriums *Prüfungsfach* signifikant aus; erst nach Ausschluss der sieben Items konnte in Bezug auf alle Teilungskriterien

Modellgültigkeit erzielt werden. So kamen 58 Rasch-Modell-konforme Items in den weiteren Analysen zum Einsatz.

Tabelle 5 bis Tabelle 8 und Abbildungen 2 bis 9 geben die Ergebnisse der LRTs sowie der graphischen Modelltests für die Skala *Grundlagenwissen* je Teilungskriterium wieder.

Teilungskriterium Rohscore

Tabelle 5: Ergebnisse des LRTs der Skala *Grundlagenwissen* a-priori sowie a-posteriori in Bezug auf das Teilungskriterium *Rohscore*

LRT vor Ausschluss von Items (a-priori) :

Andersen $\chi^2 = 117,094$

$\chi^2 (\alpha = 0,01) = 87,166$

df = 59

p-value $\leq 0,001$

LRT nach Ausschluss von 7 Items (a- posteriori):

Andersen $\chi^2 = 77,748$

$\chi^2 (\alpha = 0,01) = 79,843$

df = 53

p-value = 0,015

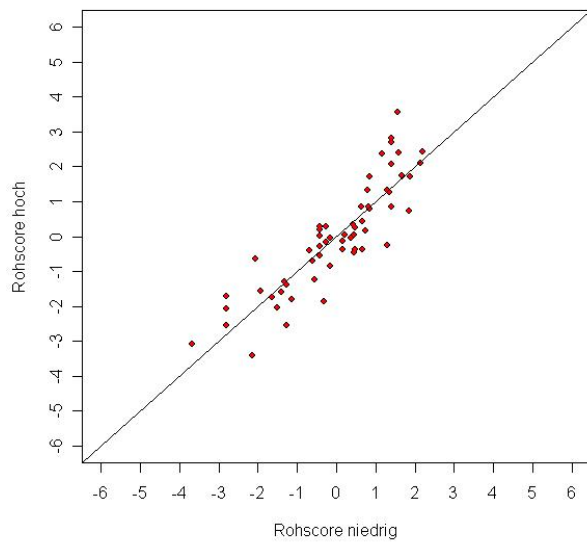


Abbildung 2: Graphischer Modelltest der Skala Grundlagenwissen in Bezug auf das Teilungskriterium Rohscore, a-priori

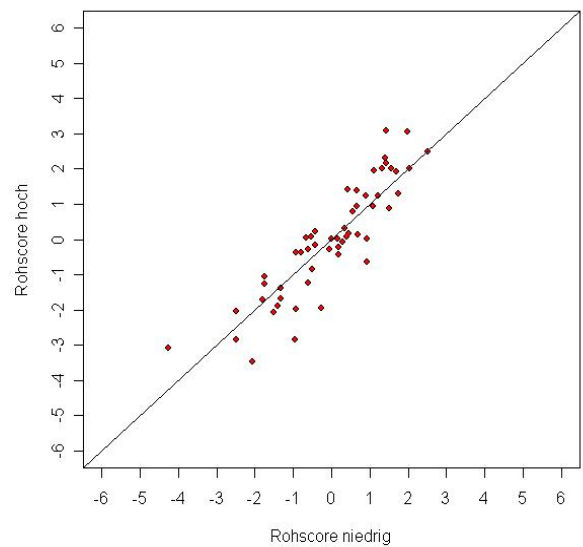


Abbildung 3: Graphischer Modelltest der Skala Grundlagenwissen in Bezug auf das Teilungskriterium Rohscore, a-posteriori

Teilungskriterium Alter

Tabelle 6: Ergebnisse des LRTs der Skala Grundlagenwissen a-priori sowie a-posteriori in Bezug auf das Teilungskriterium *Alter*

LRT vor Ausschluss von Items (a-priori) :

Andersen $\chi^2 = 94,931$

$\chi^2 (\alpha = 0,01) = 88,379$

df = 60

p-value = 0,003

LRT nach Ausschluss von 7 Items (a-posteriori):

Andersen $\chi^2 = 76,142$

$\chi^2 (\alpha = 0,01) = 81,069$

df = 54

p-value = 0,025

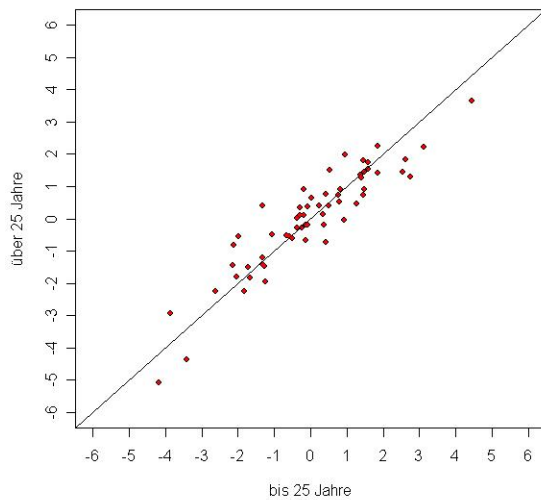


Abbildung 4: Graphischer Modelltest der Skala *Grundlagenwissen* in Bezug auf das Teilungskriterium Alter, a-priori

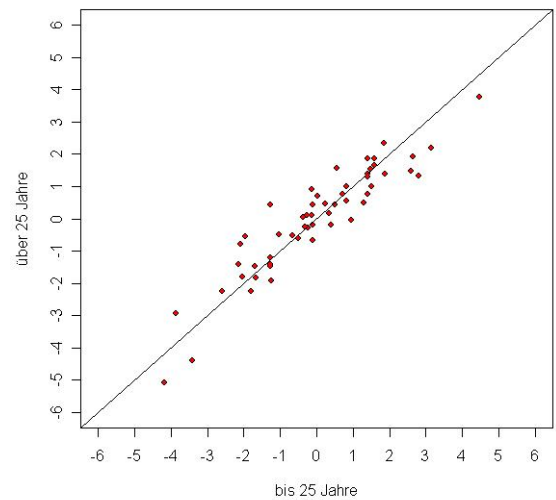


Abbildung 5: Graphischer Modelltest der Skala *Grundlagenwissen* in Bezug auf das Teilungskriterium Alter, a-posteriori

Teilungskriterium Note

Tabelle 7: Ergebnisse des LRTs der Skala *Grundlagenwissen* a-priori sowie a-posteriori in Bezug auf das Teilungskriterium *Note*

LRT vor Ausschluss von Items (a-priori) :

Andersen $\chi^2 = 89,404$

$\chi^2(\alpha = 0,01) = 89,591$

df = 61

p-value = 0,010

LRT nach Ausschluss von 7 Items (a-posteriori):

Andersen $\chi^2 = 67,680$

$\chi^2(\alpha = 0,01) = 82,292$

df = 55

p-value = 0,117

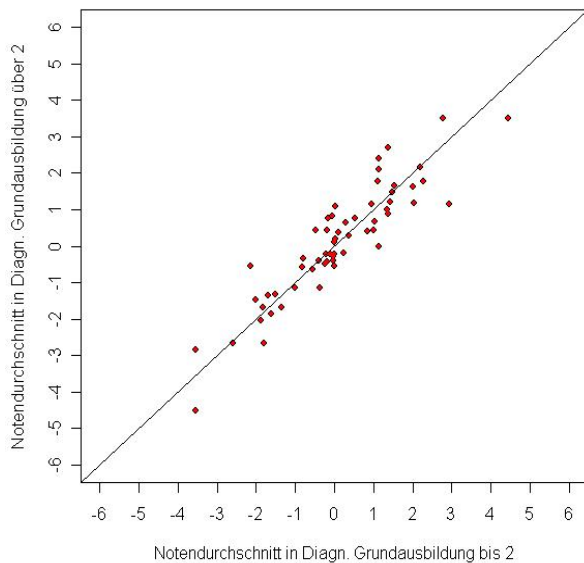


Abbildung 6: Graphischer Modelltest der Skala Grundlagenwissen in Bezug auf das Teilungskriterium Note, a-priori

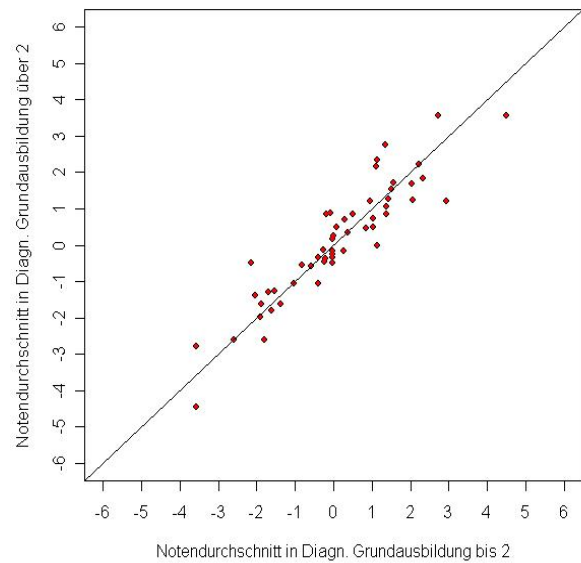


Abbildung 7: Graphischer Modelltest der Skala Grundlagenwissen in Bezug auf das Teilungskriterium Note, a-posteriori

Teilungskriterium Prüfungsfach

Tabelle 8: Ergebnisse des LRTs der Skala Grundlagenwissen a-priori sowie a-posteriori in Bezug auf das Teilungskriterium Prüfungsfach

LRT vor Ausschluss von Items (a-priori) :

Andersen $\chi^2 = 53,904$

$\chi^2 (\alpha = 0,01) = 76,154$

df = 50

p-value = 0,327

LRT nach Ausschluss von 7 Items (a-posteriori):

Andersen $\chi^2 = 49,645$

$\chi^2 (\alpha = 0,01) = 68,710$

df = 44

p-value = 0,258

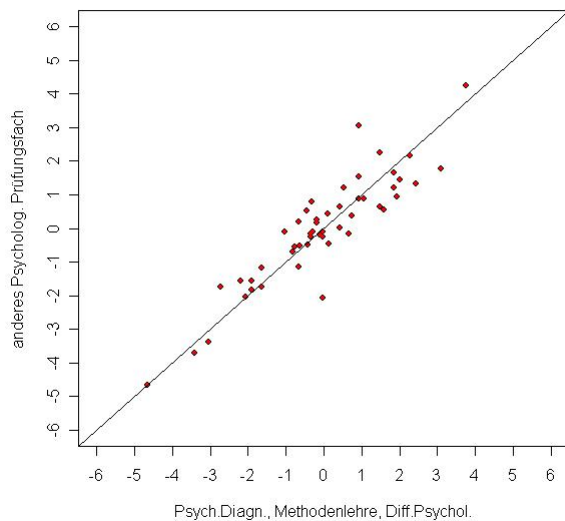


Abbildung 8: Graphischer Modelltest der Skala Grundlagenwissen in Bezug auf das Teilungskriterium Prüfungsfach, a-priori

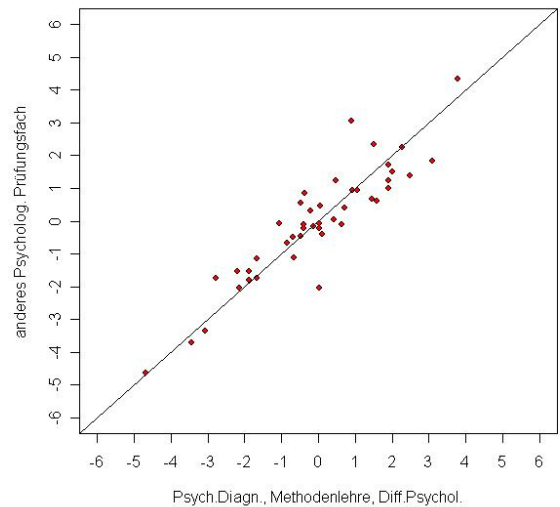


Abbildung 9: Graphischer Modelltest der Skala Grundlagenwissen in Bezug auf das Teilungskriterium Prüfungsfach, a-posteriori

Tabelle 9: Überblick über die Item-Ausschlüsse der Skala Grundlagenwissen: Die grau unterlegten Felder stellen die ausgeschlossenen Items dar, Gedankenstriche stehen für Kombinationen von Aufgabeninhalt und Antwortformat, die in der vorliegenden Untersuchung nicht zum Einsatz kamen.

Itemstamm	Antwortformat			
	frei	„x aus 5“	„1 aus 6“	„1 aus 4“
1	i1F *	-	-	
2	i2F *	i2xa5	-	i2a1a4
8	i8F *	i8xa5	-	i8a1a4
9	-	-	i9a1a6rf *	i9a1a4rf
15	i15FrF *	i15xa5rf	-	-
16	i16F	i16xa5 *	-	i16a1a4
22	i22F	-	i22a1a6 *	i22a1a4
23	i23F *	-	-	-
29	i29F *	-	-	-
30	i30F *	i30xa5	-	i30a1a4
36	i36F *	-	-	-
37	-	i37xa5	i37a1a6 *	i37a1a4
43	-	i43xa5 *	i43a1a6	i43a1a4
44	-	-	i44a1a6 *	i44a1a4
50	-	-	i50a1a6 *	i50a1a4
51	-	i51xa5 *	i51a1a6	i51a1a4
57	i57F *	-	-	-

58		i58xa5	i58a1a6*	i58a1a4
64	i64Fr	i64xa5 *	-	i64a1a4
65	-	i65xa5	i65a1a6 *	i65a1a4
71		i71xa5 *	i71a1a6	i71a1a4
72	i72F *	-	-	-
78	-	i78xa5 *	-	i78a1a4
79	-	i79xa5rf	i79a1a6rf *	i79a1a4
85	-	i85xa5 *	i85a1a6	i85a1a4
86	i86F *	-		
92	i92F	-	i92a1a6 *	i92a1a4
93	i93F *	-	-	-
99	i99F *	-	-	-
100	i100F *	-	-	-

Anmerkung: Die mit * gekennzeichneten Items entstammen der Originalversion des DIC von Kubinger (2006).

Zwei der Itemausschlüsse der Skala *Grundlagenwissen* betrafen Items im freien Antwortformat, einer jene im Format „1 aus 6“ und vier solche im Format „1 aus 4“. Von den insgesamt 7 ausgeschlossenen Items entstammen drei der Originalversion des DIC von Kubinger (Kubinger, 2006), bei vier handelt es sich um im Rahmen von Gottschall (2007) bzw. dieser Arbeit konstruierten Entsprechungs-Items.

Skala Statistik- und Testtheorie-Wissen

Die Skala Statistik- und Testtheorie-Wissen umfasste ursprünglich 38 Items, von denen sämtliche den Modellkontrollen nach dem Rasch-Modell standhielten. Keiner der LRTs fiel a-priori signifikant aus, und sämtliche Items dieser Skala wurden in den endgültigen Itempool aufgenommen.

LRT vor Ausschluss von Items (a-priori) :

Andersen $\chi^2 = 44.133$

$\chi^2(\alpha = 0,01) = 53,486$

df = 32

p-value = 0.075

bis Tabelle 13 und Abbildung 10 bis Abbildung 13 stellen die Ergebnisse der Modelltests je Teilungskriterium dar.

Teilungskriterium Rohscore

Tabelle 10: Ergebnisse des LRTs der Skala Statistik- und Testtheorie-Wissen a-priori in Bezug auf das Teilungskriterium Rohscore

LRT vor Ausschluss von Items (a-priori) :

Andersen $\chi^2 = 44.133$

$\chi^2(\alpha = 0,01) = 53,486$

df = 32

p-value = 0.075

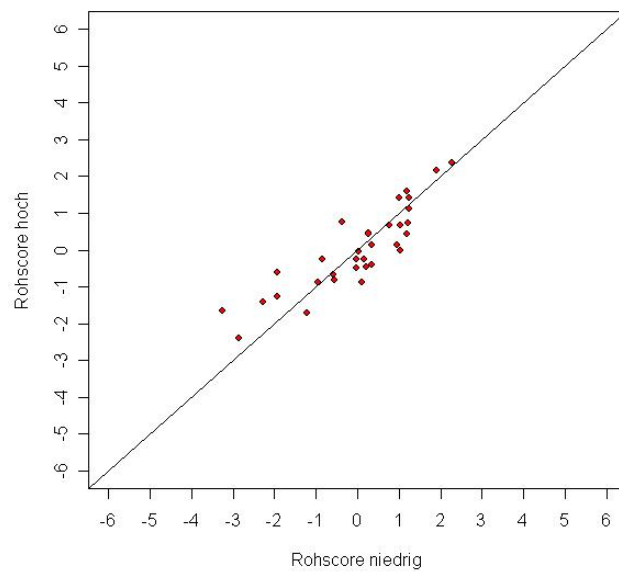


Abbildung 10: Graphischer Modelltest der Skala *Statistik- und Testtheorie-Wissen* in Bezug auf das Teilungskriterium *Rohscore*, a-priori

Teilungskriterium Alter

Tabelle 11: Ergebnisse des LRTs der Skala *Statistik- und Testtheorie-Wissen* a-priori in Bezug auf das Teilungskriterium *Alter*

LRT vor Ausschluss von Items (a-priori) :

Andersen $\chi^2 = 33.335$

$\chi^2(\alpha = 0,01) = 57.342$

df = 35

p-value = 0,549

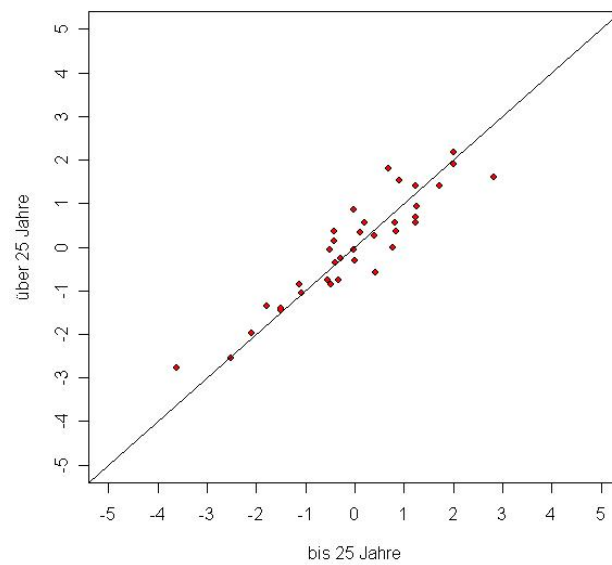


Abbildung 11: Graphischer Modelltest der Skala *Statistik- und Testtheorie-Wissen* in Bezug auf das Teilungskriterium *Alter*, a-priori

Teilungskriterium Note

Tabelle 12: Ergebnisse des LRTs der Skala *Statistik- und Testtheorie-Wissen* a-priori in Bezug auf das Teilungskriterium *Note*

LRT vor Ausschluss von Items (a-priori) :

Andersen $\chi^2 = 31.435$

$\chi^2(\alpha = 0,01) = 57.342$

df = 35

p-value = 0,641

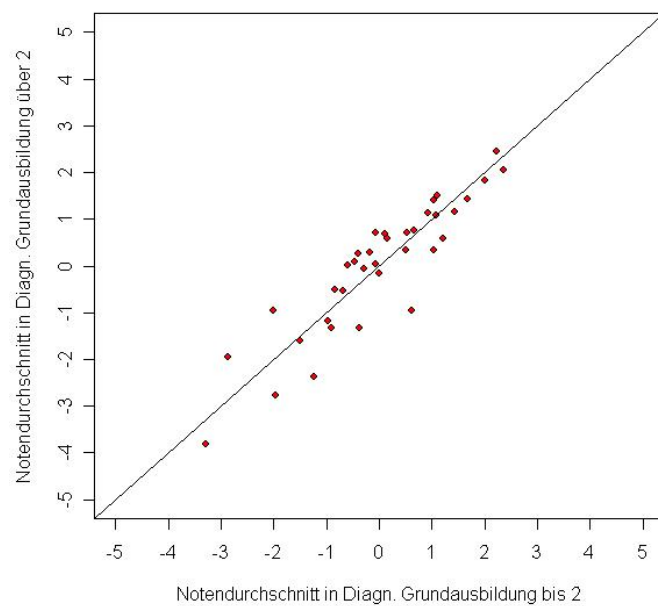


Abbildung 12: Graphischer Modelltest der Skala *Statistik- und Testtheorie-Wissen* in Bezug auf das Teilungskriterium *Note*, a-priori

Teilungskriterium Prüfungsfach

Tabelle 13: Ergebnisse des LRTs der Skala *Statistik- und Testtheorie-Wissen* a-priori in Bezug auf das Teilungskriterium *Prüfungsfach*

LRT vor Ausschluss von Items (a-priori) :
Andersen $\chi^2 = 27,282$
$\chi^2(\alpha = 0,01) = 52,191$
df = 31
p-value = 0,648

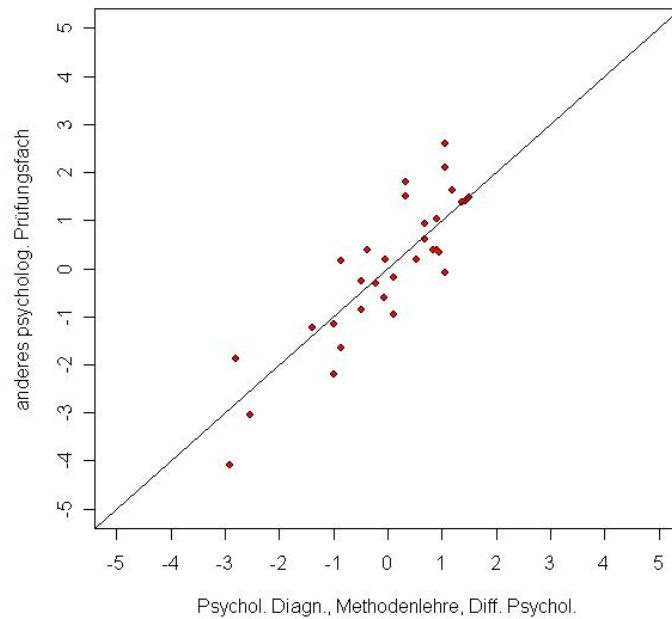


Abbildung 13: Graphischer Modelltest der Skala *Statistik- und Testtheorie-Wissen* in Bezug auf das Teilungskriterium *Prüfungsfach, a-priori*

Skala Verfahrensinventar-Kenntnis

Da keiner der LRTs a-priori signifikant ausfiel, konnte für sämtliche Items Homogenität im Sinne des Rasch-Modells angenommen werden.

Die Ergebnisse der Modelltests je Teilungskriterium sind in Tabelle 14 bis Tabelle 17 bzw. in den Abbildungen 18 bis 21 dargestellt.

Teilungskriterium Rohscore

Tabelle 14: Ergebnisse des LTRs der Skala *Verfahrensinventar-Kenntnis* a-priori in Bezug auf das Teilungskriterium *Rohscore*

LRT vor Ausschluss von Items (a-priori) :

Andersen $\chi^2 = 26.558$

$\chi^2 (\alpha = 0,01) = 44.314$

df = 25

p-value = 0.378

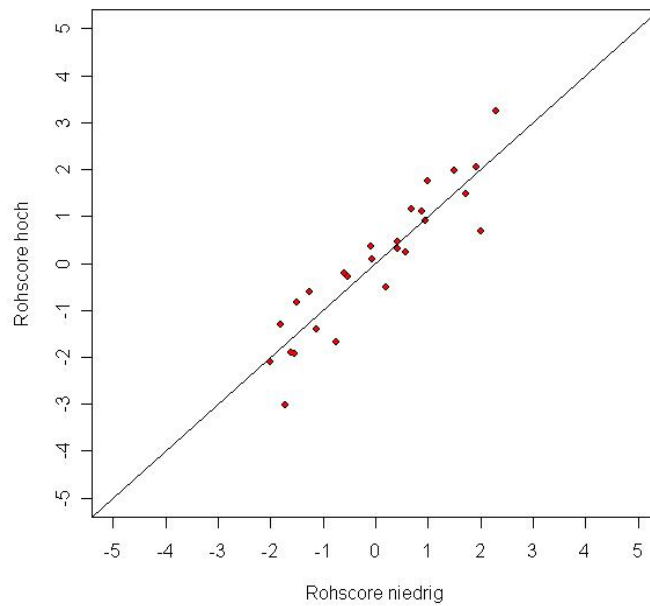


Abbildung 14: Graphischer Modelltest der Skala *Verfahrensinventar-Kenntnis* in Bezug auf das Teilungskriterium *Rohscore*, a-priori

Teilungskriterium Alter

Tabelle 15: Ergebnisse des LTRs der Skala *Verfahrensinventar-Kenntnis* a-priori in Bezug auf das Teilungskriterium *Alter*

LRT vor Ausschluss von Items (a-priori) :

Andersen $\chi^2 = 37.443$

$\chi^2 (\alpha = 0,01) = 46.963$

df = 27

p-value = 0.087

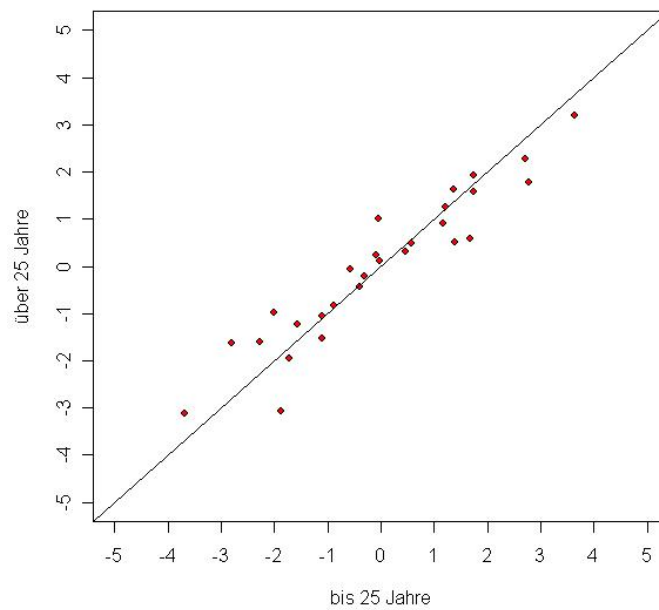


Abbildung 15: Graphischer Modelltest der Skala Verfahrensinventar-Kenntnis in Bezug auf das Teilungskriterium Alter, a-priori

Teilungskriterium Note

Tabelle 16: Ergebnisse des LRTs der Skala Verfahrensinventar- Kenntnis a-priori in Bezug auf das Teilungskriterium Note

LRT vor Ausschluss von Items (a-priori) :

Andersen $\chi^2 = 37.585$

$\chi^2(\alpha = 0,01) = 45.642$

df = 26

P-value = 0.066

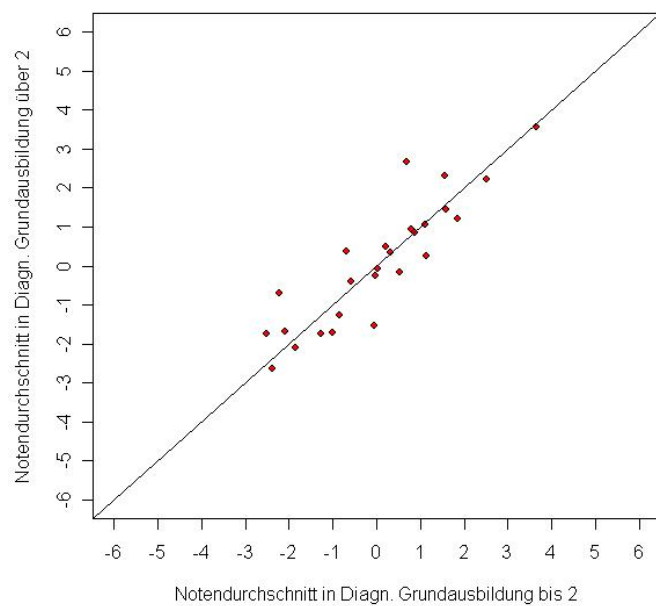


Abbildung 16: Graphischer Modelltest der Skala *Verfahrensinventar-Kenntnis* in Bezug auf das Teilungskriterium *Note*, a-priori

Teilungskriterium Prüfungsfach

Tabelle 17: Ergebnisse des LRTs der Skala *Verfahrensinventar-Kenntnis* a-priori in Bezug auf das Teilungskriterium *Prüfungsfach*

LRT vor Ausschluss von Items (a-priori) :

Andersen $\chi^2 = 25.633$

$\chi^2(\alpha = 0,01) = 41.638$

df = 23

p-value = 0.318

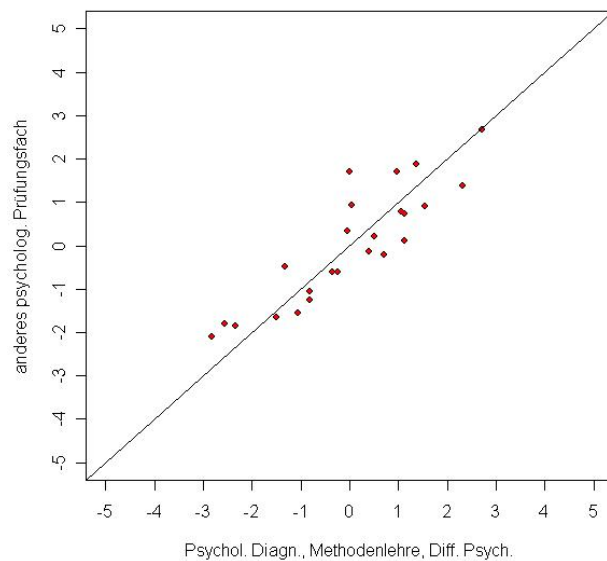


Abbildung 17: Graphischer Modelltest der Skala Verfahrensinventar-Kenntnis in Bezug auf das Teilungskriterium Prüfungsfach, a-priori

Skala Theoriewissen zur Differentiellen Psychologie

In dieser Skala erwiesen sich nach Berechnung der LRTs sämtliche Items als Rasch-Modell-konform, bei einem Item war aufgrund zu extremer Lösungshäufigkeiten keine Schätzung möglich. Es konnten somit 22 Items im endgültigen Itempool erhalten bleiben.

Die Ergebnisse der LRTs je Teilungskriterium sind in Tabelle 18 bis Tabelle 21 dargestellt, die Graphischen Modellkontrollen in den Abbildungen 18 bis 21.

Teilungskriterium Rohscore

Tabelle 18: Ergebnisse des LRTs der Skala Theoriewissen zur Differentiellen Psychologie a-priori in Bezug auf das Teilungskriterium Rohscore

LRT vor Ausschluss von Items (a-priori) :

Andersen $\chi^2 = 14.865$

$$\chi^2(\alpha = 0,01) = 37.566$$

$$df = 20$$

$$p\text{-value} = 0.784$$

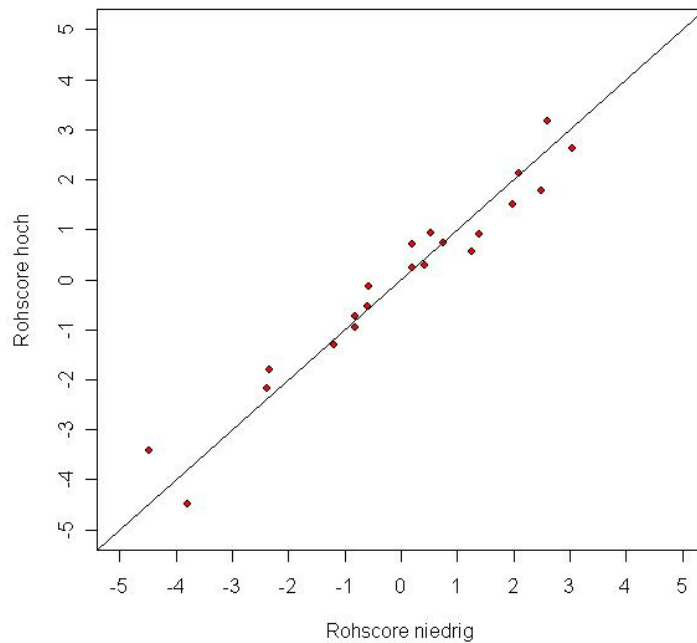


Abbildung 18: Graphischer Modelltest der Skala *Theoriewissen zur Differentiellen Psychologie* in Bezug auf das Teilungskriterium *Rohscore*, a-priori

Teilungskriterium Alter

Tabelle 19: Ergebnisse des LRTs der Skala *Theoriewissen zur Differentiellen Psychologie* a-priori in Bezug auf das Teilungskriterium *Alter*

LRT vor Ausschluss von Items (a-priori) :

$$\text{Andersen } \chi^2 = 24.391$$

$$\chi^2(\alpha = 0,01) = 37.566$$

$$df = 20$$

$$p\text{-value} = 0.226$$

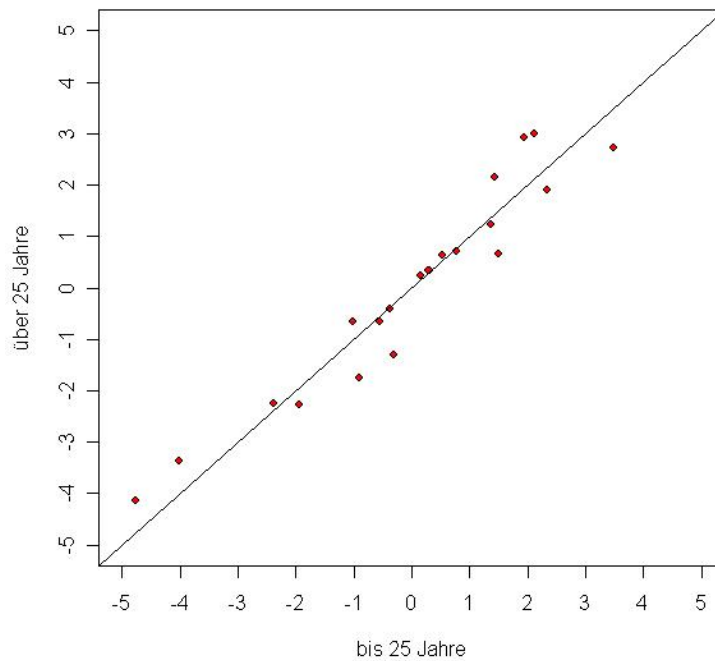


Abbildung 19: Graphischer Modelltest der Skala *Theoriewissen zur Differentiellen Psychologie* in Bezug auf das Teilungskriterium *Alter*, a-priori

Teilungskriterium Note

Tabelle 20: Ergebnisse des LRTs der Skala *Theoriewissen zur Differentiellen Psychologie* a-priori in Bezug auf das Teilungskriterium *Note*

LRT vor Ausschluss von Items (a-priori) :

Andersen $\chi^2 = 20.576$

$\chi^2(\alpha = 0,01) = 37.566$

df = 20

p-value = 0.422

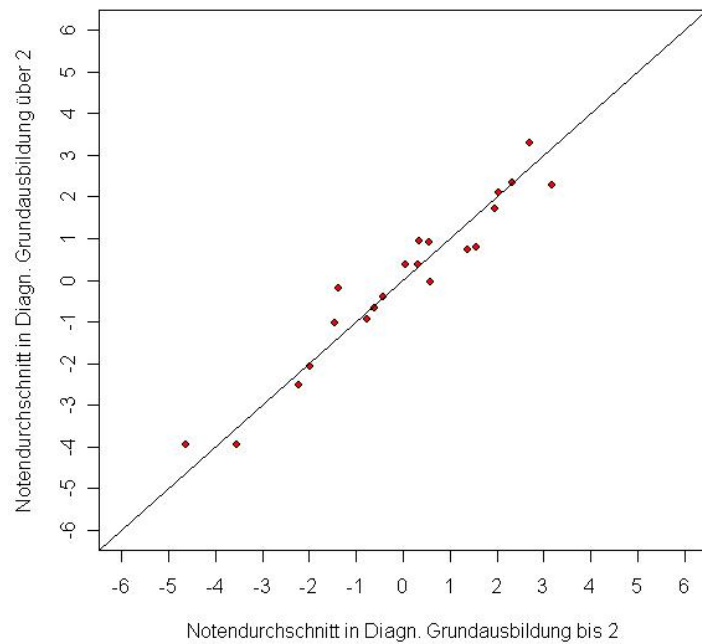


Abbildung 20: Graphischer Modelltest der Skala *Theoriewissen zur Differentiellen Psychologie* in Bezug auf das Teilungskriterium *Note*, a-priori

Teilungskriterium Prüfungsfach

Tabelle 21: Ergebnisse des LRTs der Skala *Theoriewissen zur differentiellen Psychologie* a-priori in Bezug auf das Teilungskriterium *Prüfungsfach*

LRT vor Ausschluss von Items (a-priori) :

Andersen $\chi^2 = 31.587$

$\chi^2 (\alpha = 0,01) = 33.409$

df = 17

p-value = 0.017

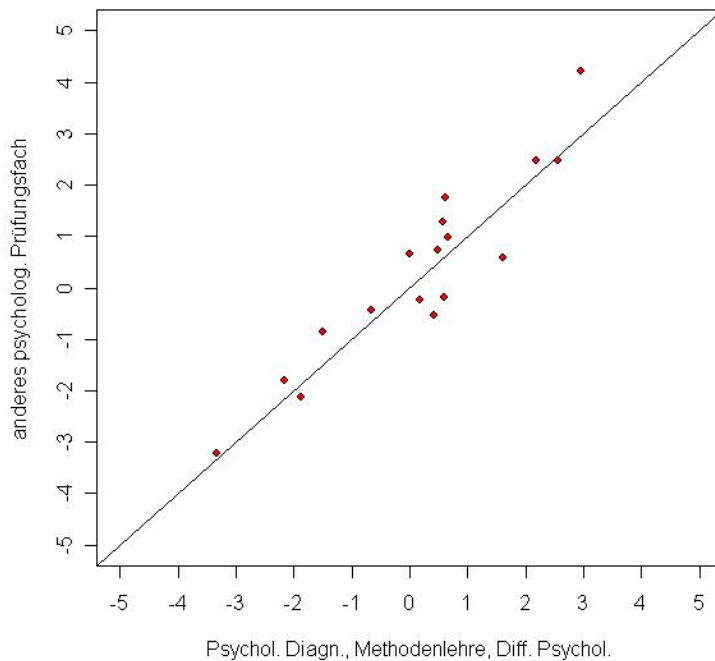


Abbildung 21: Graphischer Modelltest der Skala *Theoriewissen zur Differentiellen Psychologie* in Bezug auf das Teilungskriterium *Prüfungsfach*, a-priori

Skala Verfahrensbezogenes Detailwissen

In dieser Skala stellten sich zwei Items als nicht modellkonform heraus und wurden ausgeschlossen. Es resultieren a- posteriori 18 Items für die Hypothesenprüfung.

Von den LRTs lieferte lediglich jener in Bezug auf das Teilungskriterium *Note* a-priori ein signifikantes Ergebnis.

Die Ergebnisse der LRTs und der graphischen Modellkontrollen a-priori und a-posteriori für die Skala Verfahrensbezogenes Detailwissen finden sich in Tabelle 22 bis Tabelle 25 bzw. in den Abbildungen 22 bis 29, dargestellt pro Teilungskriterium.

Teilungskriterium Rohscore

Tabelle 22: Ergebnisse des LRTs der Skala *Verfahrensbezogenes Detailwissen* a-priori sowie a-posteriori in Bezug auf das Teilungskriterium *Rohscore*

LRT vor Ausschluss von Items (a-priori) :

Andersen $\chi^2 = 29,382$

$\chi^2(\alpha = 0,01) = 34,805$

df = 18

p-value = 0.045

LRT nach Ausschluss von zwei Items (a-posteriori):

Andersen $\chi^2 = 27,548$

$\chi^2(\alpha = 0,01) = 30,578$

df = 15

p-value = 0,025

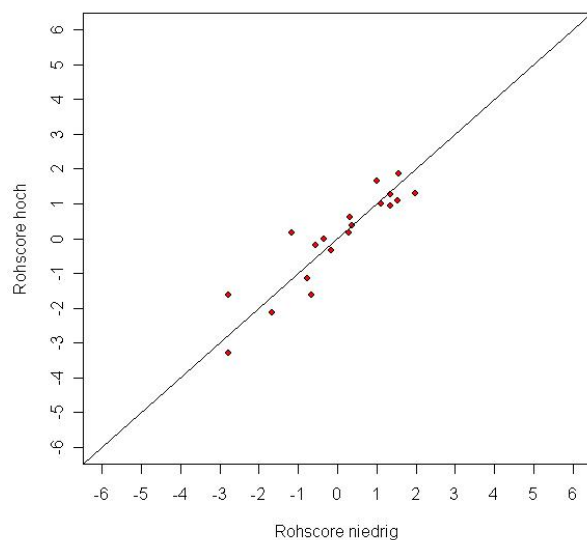


Abbildung 22: Graphischer Modelltest der Skala *Verfahrensbezogenes Detailwissen* in Bezug auf das Teilungskriterium *Rohscore*, a-priori

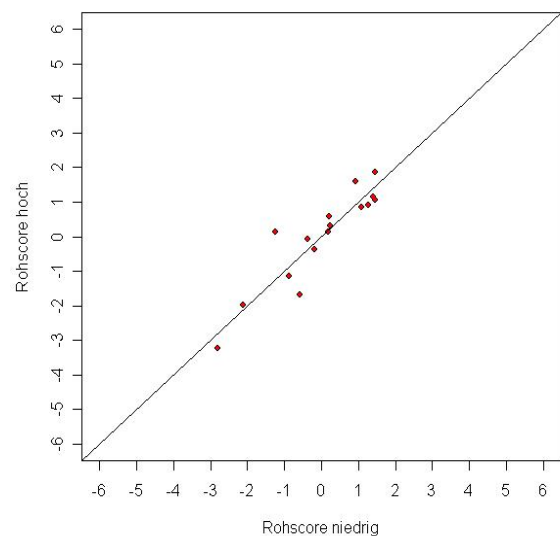


Abbildung 23: Graphischer Modelltest der Skala *Verfahrensbezogenes Detailwissen* in Bezug auf das Teilungskriterium *Rohscore*, a-posteriori

Teilungskriterium Alter

Tabelle 23: Ergebnisse des LRTs der Skala *Verfahrensbezogenes Detailwissen* a-priori sowie a-posteriori in Bezug auf das Teilungskriterium *Alter* LRT vor Ausschluss von Items (a-priori) :

Andersen $\chi^2 = 15.309$

$\chi^2(\alpha = 0,01) = 34.805$

df = 18

p-value = 0.641

LRT nach Ausschluss von zwei Items (a-posteriori):

Andersen $\chi^2 = 14,704$

$\chi^2(\alpha = 0,01) = 31,100$

df = 16

p-value = 0,546

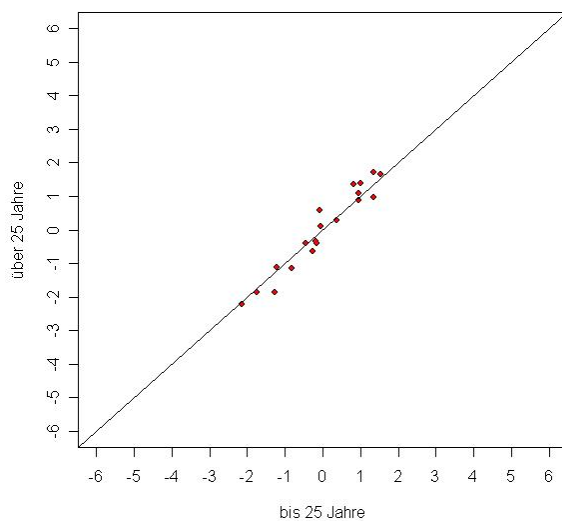


Abbildung 24: Graphischer Modelltest der Skala *Verfahrensbezogenes Detailwissen* in Bezug auf das Teilungskriterium *Alter*, a-priori

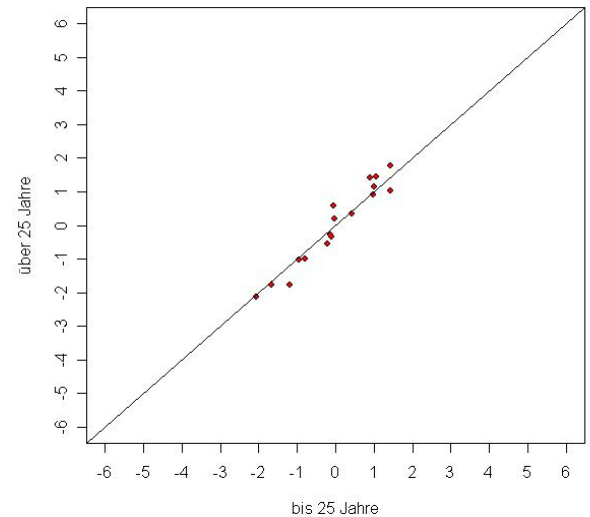


Abbildung 25: Graphischer Modelltest der Skala *Verfahrensbezogenes Detailwissen* in Bezug auf das Teilungskriterium *Alter*, a-posteriori

Teilungskriterium Note

Tabelle 24: Ergebnisse des LRTs der Skala *Verfahrensbezogenes Detailwissen* a-priori sowie a-posteriori in Bezug auf das Teilungskriterium *Note*

LRT vor Ausschluss von Items (a-priori) :

Andersen $\chi^2 = 49.665$

$\chi^2(\alpha = 0,01) = 36.191$

df = 19

p-value $\leq 0,001$

LRT nach Ausschluss von zwei Items (a-posteriori):

Andersen $\chi^2 = 29,780$

$\chi^2(\alpha = 0,01) = 33,409$

df = 17

p-value = 0,028

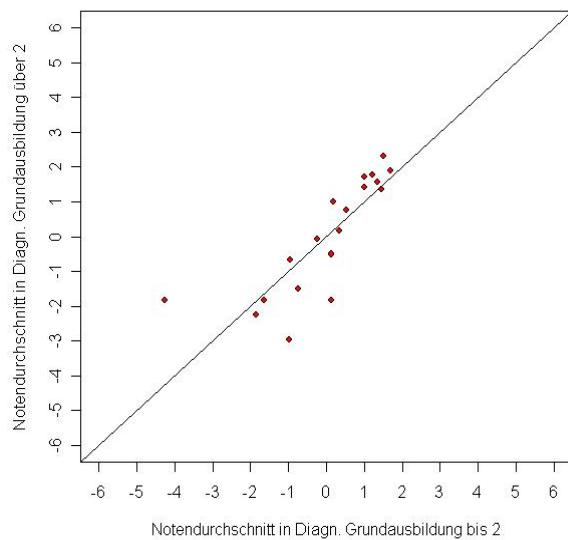


Abbildung 26: Graphischer Modelltest der Skala *Verfahrensbezogenes Detailwissen* in Bezug auf das Teilungskriterium *Note*, a-priori

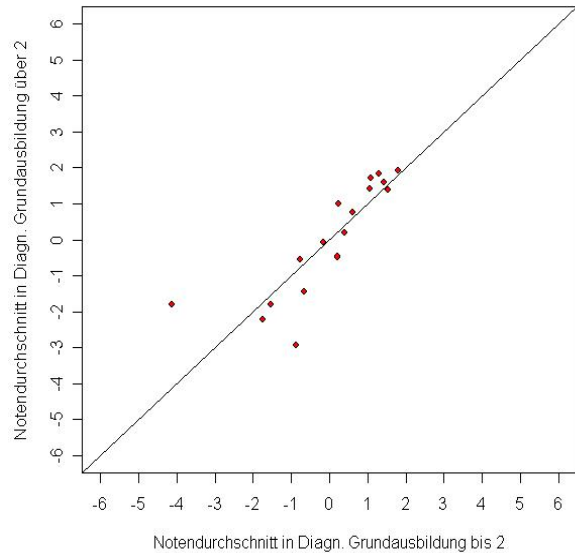


Abbildung 27: Graphischer Modelltest der Skala *Verfahrensbezogenes Detailwissen* in Bezug auf das Teilungskriterium *Note*, a-posteriori

Teilungskriterium Prüfungsfach

Tabelle 25: Ergebnisse des LRTs der Skala *Verfahrensbezogenes Detailwissen* a-priori sowie a-posteriori in Bezug auf das Teilungskriterium *Prüfungsfach*

LRT vor Ausschluss von Items (a-priori) :

Andersen $\chi^2 = 26,786$

$\chi^2(\alpha = 0,01) = 33,409$

df = 17

p-value = 0.061

LRT nach Ausschluss von zwei Items (a-posteriori):

Andersen $\chi^2 = 17,872$

$\chi^2 (\alpha = 0,01) = 30,580$

df = 15

p-value = 0,269

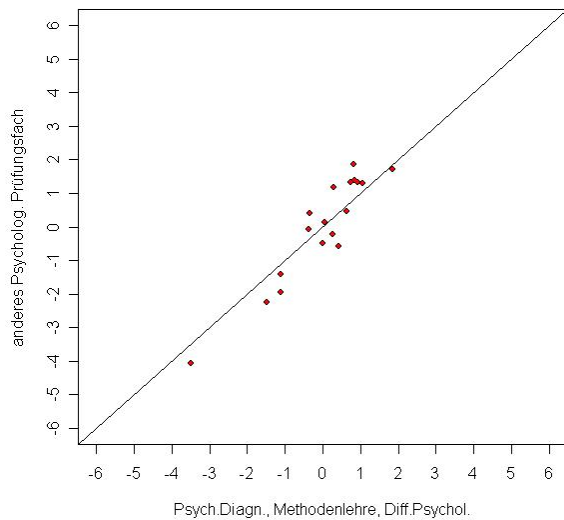


Abbildung 28: Graphischer Modelltest der Skala *Verfahrensbezogenes Detailwissen* in Bezug auf das Teilungskriterium *Prüfungsfach*, a-priori

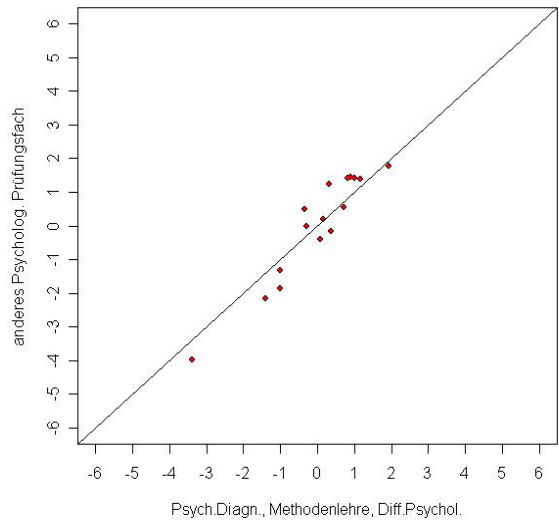


Abbildung 29: Graphischer Modelltest der Skala *Verfahrensbezogenes Detailwissen* in Bezug auf das Teilungskriterium *Prüfungsfach*, a-posteriori

Tabelle 26 stellt die Item- Ausschlüsse der Skala *Verfahrensbezogenes Detailwissen* dar.

(Zum Inhalt der jeweiligen Items siehe [Anhang B.](#))

Tabelle 26: Übersicht über die Item-Ausschlüsse der Skala *Verfahrensbezogenes Detailwissen*: die grau unterlegten Felder stellen die ausgeschlossenen Items dar, Gedankenstriche stehen für Kombinationen von Aufgabeninhalt und Antwortformat, die in der vorliegenden Untersuchung nicht zum Einsatz kamen.

Itemstamm	Antwortformat			
	Frei	„x aus 5“	„1 aus 6“	„1 aus 4“
6	-	i6xa5 *	i6a1a6	i6a1a4
13	i13F	i13xa5 *	-	i13a1a4
20	i20F *	-	-	-
27	-	-	i27a1a6 *	i27a1a4

34	i34F *	-	-	-
41	i41F *	-	-	-
48	i48F *	-	-	-
55	i55F *	-	-	-
62	i62F *	-	-	-
69	i69F *	-	-	-
76	i76F *	-	-	-
83	i83F *	-	-	-
90	i90F *	-	-	-
97	i97F *	-	-	-
104	i104F *	-	-	-

Anmerkung: Die mit * gekennzeichneten Items entstammen der Originalversion des DIC von Kubinger (2006).

Betrachtet man die Skala *Verfahrensbezogenes Detailwissen*, zeigt sich ein Itemausschluss der Kategorie des freien Antwortformats und ein Itemausschluss aus der MC-Antwortformat-Kategorie „1 aus 6“. Eines der zwei ausgeschlossenen Items ist Bestandteil der Originalversion des DIC (Kubinger, 2006).

Skala Up-to-Date

Die Skala *Up-to-Date* umfasste ursprünglich 23 Items. Die LRTs zeigten a-priori in Bezug auf keines der Teilungskriterien signifikante Werte, weshalb kein Item ausgeschlossen wurde.

In Tabelle 27 bis 30 bzw. den Abbildungen 30 bis 33 werden je Teilungskriterium die Ergebnisse der LRTs sowie der graphischen Modellkontrollen wiedergegeben.

Teilungskriterium Rohscore

Tabelle 27: Ergebnisse des LRTs der Skala *Up-To-Date* a-priori in Bezug auf das Teilungskriterium *Rohscore*

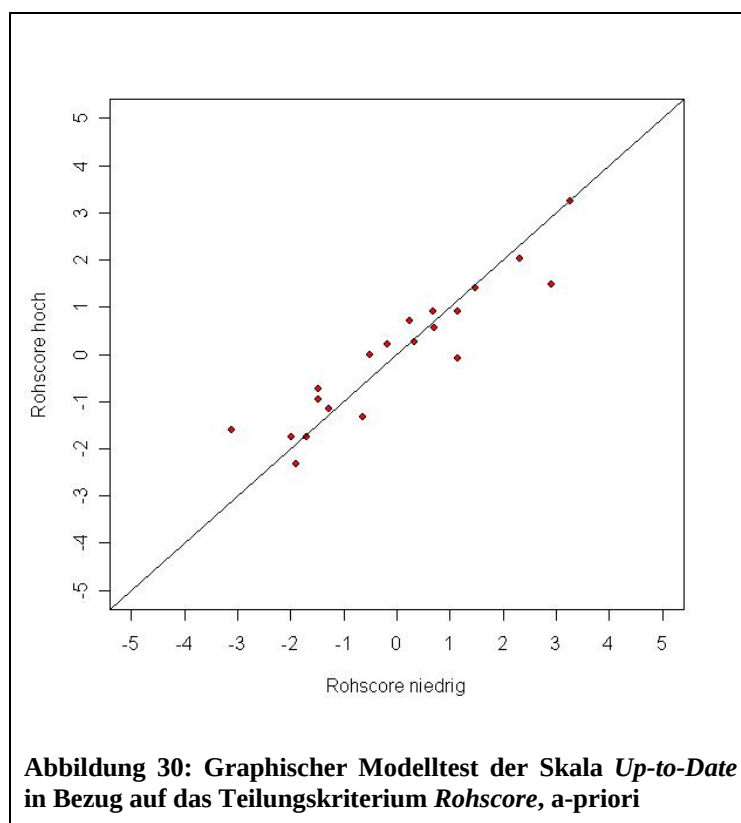
LRT vor Ausschluss von Items (a-priori) :

Andersen $\chi^2 = 30.288$

$\chi^2(\alpha = 0,01) = 36.191$

df = 19

p-value = 0.048



Teilungskriterium Alter

Tabelle 29: Ergebnisse des LRTs der Skala *Up-To-Date* a-priori in Bezug auf das Teilungskriterium *Alter*

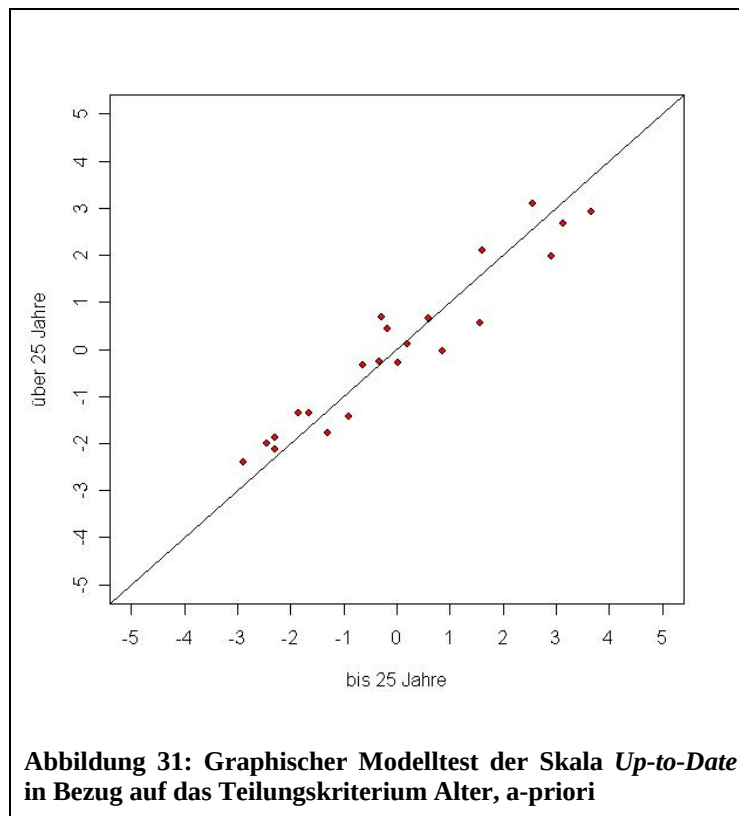
LRT vor Ausschluss von Items (a-priori) :

Andersen $\chi^2 = 31.718$

$\chi^2 (\alpha = 0,01) = 38.932$

df = 21

p-value = 0.063



Teilungskriterium Note

Tabelle 28: Ergebnisse des LRTs der Skala Up-To-Date a-priori in Bezug auf das Teilungskriterium Note

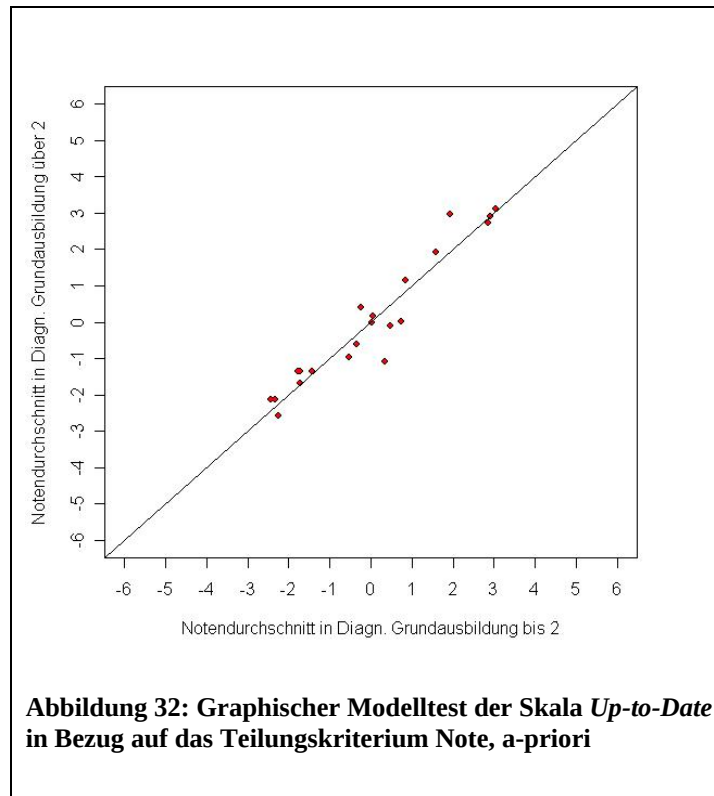
LRT vor Ausschluss von Items (a-priori) :

Andersen $\chi^2 = 17.759$

$\chi^2(\alpha = 0,01) = 37.566$

df = 20

p-value = 0.603



Teilungskriterium Prüfungsfach

Tabelle 29: Ergebnisse des LRTs der Skala Up-To-Date a-priori in Bezug auf das Teilungskriterium Prüfungsfach

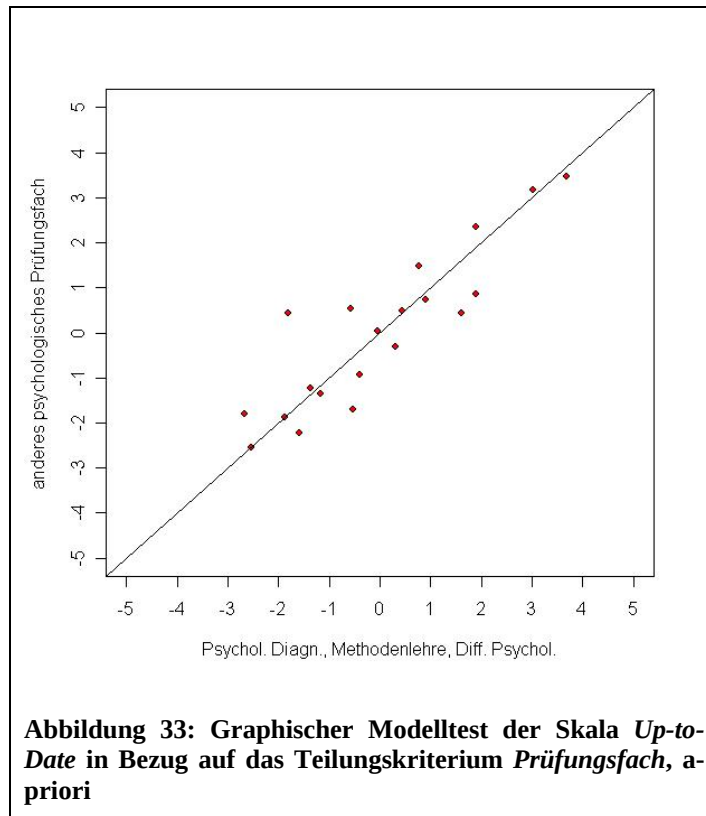
LRT vor Ausschluss von Items (a-priori) :

Andersen $\chi^2 = 28.852$

$\chi^2(\alpha = 0,01) = 36.191$

df = 19

p-value = 0.068



4.1.2. Zusammenfassende Betrachtung der Ergebnisse der Rasch-Modell-Analysen, Überblick über die Item-Ausschlüsse

Insgesamt mussten aus dem ursprünglichen Itempool von 198 Aufgaben 10 aufgrund zu auffälliger Abweichungen in der graphischen Modellkontrolle oder aufgrund zu extremer Lösungshäufigkeiten in der Stichprobe ausgeschieden werden, sodass der endgültige, für die Hypothesenprüfung der Studie eingesetzte Itempool schließlich 188 Aufgaben umfasst.

Tabelle 35 lässt sich die Anzahl der ausgeschiedenen Items pro Skala entnehmen.

Tabelle 30: Anzahl der Items a-priori und a-posteriori sowie Anzahl und Prozentsatz der Itemausschlüsse, dargestellt für jedes der vier Antwortformate. Die Ziffern in (/)* stellen Anzahl und Prozentsatz der Itemausschlüsse dar, die die Originalversion des DIC betreffen/ bzw. die Parallelversionen betreffen)

	Antwortformat				
	frei	„x aus 5“	„1 aus 6“	„1 aus 4“	Gesamt
Anzahl der Items a-	74	39	35	50	198

priori					(104/94)*
Anzahl der Items a-posteriori	71	39	33	45	188 (100/88)*
Anzahl der Itemausschlüsse	3	0	2	5	10 (4/6)*
Itemausschlüsse in %	4	0	6	10	(4/3)*

Von den Aufgaben im freien Antwortformat wurden 4 % ausgeschlossen, 6 % der Items im MC-Format „1 aus 6“, und 10 % der Items im Format „1 aus 4“.

Die geschätzten Itemleichtigkeitsparameter der Rasch-Modell-konformen Items - dargestellt pro Skala des DIC- sind Tabelle 1 im Anhang A zu entnehmen.

4.2. Ergebnisse der Hypothesentestung

Zur Überprüfung der in Kapitel 3.4. spezifizierten Hypothese wurde eine einfache Varianzanalyse durchgeführt, deren Ergebnisse in Abschnitt 5.2.1 beschrieben werden. Abschnitt 5.2.2. gibt die Resultate des Newman-Keuls-Tests wieder.

4.2.1. Ergebnisse der Varianzanalyse

Tabelle 41 bietet einen deskriptiven Überblick über die Itemleichtigkeitsparameter aller 154 Rasch-Modell-konformen Items des DIC in Abhängigkeit vom jeweiligen Antwortformat. Am schwierigsten stellten sich die Aufgaben im MC-Format „x aus 5“ heraus mit einem mittleren Leichtigkeitsparameter von -0,794, gefolgt von jenen im freien Antwortformat mit -0,581. Items, die in einem MC-Format gestaltet waren, bei dem lediglich eine Antwortmöglichkeit gefunden werden musste waren leichter als jene im freien Antwortformat und im MC-Format „x aus 5“, wobei hier die Itemleichtigkeit wie erwartet mit der a-priori-Ratewahrscheinlichkeit steigt. Aufgaben im MC-Format „1 aus 6“ (mittlerer Itemleichtigkeitsparameter = 0,626) erwiesen sich demnach gegenüber jenen im MC-Format „1 aus 4“ (mittlerer Itemleichtigkeitsparameter = 1,141) als schwieriger.

Tabelle 31 : Deskriptive Statistiken der Itemleichtigkeitsparameter je Antwortformat

Antwortformat	N	Mittelwert	Standardabweichung	Standardfehler
freies Antwortformat	71	-,58068	1,728923	,205185
MC-Antwortformat x aus 5	39	-,79446	1,353193	,216684
MC-Antwortformat 1 aus 4	45	1,14120	1,402509	,209074
MC-Antwortformat 1 aus 6	33	,62576	1,380702	,240349
Gesamt	188	-,00111	1,709458	,124675

Um die Unterschiede in den Itemleichtigkeitsparametern zwischen den einzelnen Antwortformat-Kategorien interferenzstatistisch auf ihre Signifikanz zu prüfen, wurde eine einfache Varianzanalyse berechnet. Tabelle 32 gibt die Ergebnisse der Varianzanalyse wieder ($\alpha = 0,05$).

Tabelle 32: Ergebnisse der Varianzanalyse

	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen	120,083	3	40,028	17,274	,000
Innerhalb der Gruppen	426,377	184	2,317		
Gesamt	546,460	187			

Das Ergebnis fällt signifikant aus ($p < 0,001$), die Nullhypothese muss deshalb verworfen werden. Es besteht ein signifikanter Mittelwertsunterschied zwischen den Leichtigkeitsparametern der Items im freien Antwortformat, im MC-Format „x aus 5“, im MC-Format „1 aus 6“ und im MC-Format „1 aus 4“.

Ergänzend zu diesem signifikanten interferenzstatistischen Ergebnis liefert Tabelle 33 einen deskriptiven Überblick über die Itemleichtigkeiten der Entsprechungs-Items gegenübergestellt für die verschiedenen Antwortformate.

Tabelle 33: Itemleichtigkeitsparameter aller 188 Rasch-Modell-konformen Items, die mindestens in zwei Antwortformaten vorkommen (Entsprechungs-Items), für die verschiedenen Antwortformate gegenübergestellt zum Vergleich. Abweichungen, die entgegengesetzt der aufgrund der a-priori-Ratewahrscheinlichkeit erwarteten Richtung ausgefallen sind, sind grau unterlegt.

Itemstamm	Antwortformat			
	frei	„x aus 5“	„1 aus 6“	„1 aus 4“
2	-	-1.622	-	0.857

3	0.752	-	1.340	1.462
4	-	0.304	2.335	3.486
6	-	0.287	-	-0.952
7	-	-	0.416	1.284
8	0.946	0.375	-	1.213
9	-	-	1.524	1.829
10	1.157	1.413	-	1.715
12	1.669	-	2.800	-
13	-0.759	-0.014	-	1.120
14	-	-0.337	3.245	3.393
15	1.965	1.623	-	-
16	-1.978	-0.033	-	1.024
17	-	-0.395	0.624	0.481
18	0.731	-1.451	-	1.499
19	-1.104	-	0.547	0.896
22	-2.232	-	-0.136	-
24	-	-3.817	2.414	2.070
27	-		1.427	0.446
30	0.236	-0.033	-	1.734
31	-	-1.365	-0.234	0.279
32	-	-2.511	-0.325	0.310
35	0.815		0.602	0.149
37	-	0.137	1.200	1.380
38	-	-0.804	0.980	0.856
39	0.236	0.836	-	1.329
40	2.498	-1.462	-	3.713
42	-	0.318	2.046	2.584
43	-	-1.444	0.033	0.469
46	-	-2.034	-1.440	-1.958
47	-2.182	-	0.422	1.056
50	-	-	-1.425	-0.429
51	-	-1.121	-1.763	-1.806
52	-0.568	-2.336	-	2.472
58	-	-1.111	2.689	4.147
60	-3.497	-	-2.242	-1.595
64	-1.253	0.069	-	0.259
65	-	0.137	1.784	2.353
71	-	-0.276	0.672	-
73	-	-0.470	0.383	1.104
78	-	-0.582	-	1.732
79	-	1.416	1.965	2.204
80	-	-1.429	-	0.010
81	1.864	-0.553	-	4.241
84	-2.453	-4.402	-	-

85	-	-2.857	0.347	0.593
87	-	-0.049	0.492	1.228
94	-	-2.764	-1.324	-0.105
95	-	-0.547		1.070
101	-	-1.880	0.101	0.153

Sämtliche Entsprechungs-Items sind im freien Antwortformat schwieriger als im MC-Format „1 aus 4“, in sechs von sieben Fällen ist das Item im freien Antwortformat schwieriger als im MC-Format „1 aus 6“, aber in nur fünf von 11 Fällen schwieriger als im MC-Format „x aus 5“. Vergleicht man die Itemleichtigkeiten der Entsprechungs-Items im Format „x aus 5“ mit den übrigen MC-Antwortformaten so zeigt sich, dass erstere stets (in 19 von 19 Fällen) schwieriger sind als ihre Entsprechungs-Items im Format „1 aus 6“ und jene im Format „1 aus 4“ (in 35 von 35 Fällen). In 25 von 29 Fällen ist das Item im Format „1 aus 6“ schwieriger als sein Entsprechungseitem im Format „1 aus 4“.

4.2.2. Ergebnisse des Post-Hoc-Tests

Um genauer zu untersuchen, zwischen welchen Antwortformaten signifikante Unterschiede in den Itemleichtkeitsparametern bestehen, wurde post-hoc die Newman-Keuls-Prozedur angewandt. Die Ergebnisse sind in Tabelle 34 dargestellt.

Tabelle 34: Ergebnisse der Newman-Keuls-Prozedur

Antwortformat	N	Untergruppe für Alpha = .05.	
		2	1
MC-Antwortformat x aus 5	39	-,79446	
freies Antwortformat	71	-,58068	
MC-Antwortformat 1 aus 6	33		,62576
MC-Antwortformat 1 aus 4	45		1,14120
Signifikanz		,514	,117

Die Mittelwerte für die in homogenen Untergruppen befindlichen Gruppen werden angezeigt.

a Verwendet ein harmonisches Mittel für Stichprobengröße = 43,360.

b Die Gruppengrößen sind nicht identisch. Es wird das harmonische Mittel der Gruppengrößen verwendet. Fehlerniveaus des Typs I sind nicht garantiert.

Wie aus Tabelle 34 ersichtlich lassen sich die vier Antwortformate bezüglich ihrer Itemleichtkeitsparameter in zwei homogene Untergruppen teilen. Jeweils eine Gruppe bilden jene Items im freien Antwortformat und im MC-Format „x aus 5“, sowie jene in den MC-Formaten „1 aus 6“ und „1 aus 4“.

Freies Antwortformat und MC-Antwortformat „x aus 5“ unterscheiden sich nicht signifikant voneinander in ihrer Schwierigkeit, heben sich jedoch beide von den anderen beiden MC-Formaten „1 aus 6“ und „1 aus 4“ ab, die signifikant geringere Schwierigkeiten aufweisen. Der Schwierigkeitsunterschied zwischen MC-Format „1 aus 6“ und „1 aus 4“, ist zwar wie aus Tabelle 41 ersichtlich wird in der erwarteten Richtung vorhanden, bleibt statistisch jedoch über der Signifikanzgrenze ($p=0,117$).

4.2.3. Ergebnisse des t-tests für abhängige Stichproben

Zieht man nun nur diejenigen $n = 29$ Items heran, die sowohl im MC-Format „1 aus 6“ als auch im Format „1 aus 4“ existieren, und den Kontrollen nach dem Rasch-Modell standhalten, und vergleicht sie hinsichtlich der Mittelwerte ihrer Leichtigkeitsparameter miteinander, zeigt sich – wie aus Tabelle 35 ersichtlich - ein deutlicher Unterschied in den mittleren Schwierigkeitsparameter der beiden Antwortformate. Während die Items im Format „1 aus 6“ einen mittleren Leichtigkeitsparameter von 0,627 aufweisen, liegen ihre Entsprechungs-Items im Format „1 aus 4“ mit 0,977 deutlich darüber.

Tabelle 35: Deskriptivstatistiken des Vergleichs der Itemparameter von MC-Items im Format „1 aus 6“ und „1 aus 4“.

		Mittelwert	N	Standardabweichung	Standardfehler des Mittelwertes
Paaren 1	Par1a6	,62628	29	1,382558	,256735
	Par1a4	,97652	29	1,464231	,271901

Eine Berechnung mittels t-tests für abhängige Stichproben belegt die Signifikanz dieses Unterschiedes ($p = 0,001$, $df = 28$), dargestellt in Tabelle 36.

Tabelle 36: Ergebnisse des t-Tests für abhängige Stichproben

	Gepaarte Differenzen					T	df	Sig. (1-seitig)
	Mittelwert	Standardabweichung	Standardfehler des Mittelwerts	95% Konfidenzintervall der Differenz		Mittelwert		Standardabweichung
				Obere	Untere			Standardfehler des Mittelwerts
Par4-Par6	-,350241	,550185	,102167	-,559521	-,140962	-3,428	28	,001

Diskussion

Die vorliegende experimentelle Studie zeigt, dass die Aufgabenschwierigkeit von Leistungstests signifikant vom Antwortformat abhängt. Dies betrifft ebenso die Entscheidung zwischen freiem und MC-Antwortformat, als auch die spezifische Art der Gestaltung des MC-Antwortformats.

Die Schwierigkeit von Items im freien Antwortformat liegt signifikant über der in den beiden MC-Antwortformaten „1 aus 6“ und „1 aus 4“ (nur eine richtige ist aus vier bzw. sechs Antwortmöglichkeiten zu finden). MC-Aufgaben der Art „x aus 5“, bei der jede der fünf Antwortmöglichkeiten unabhängig von den übrigen als richtig oder falsch erkannt werden muss, erwiesen sich demgegenüber als nicht signifikant einfacher als im freien Antwortformat. Sie sind jedoch ebenso signifikant schwieriger als in den beiden anderen MC-Formaten. („1 aus 6“ und „1 aus 4“).

Betrachtet man die Aufgabenschwierigkeit der beiden MC-Formate „1 aus 6“ und „1 aus 4“, im Vergleich, zeigt sich dass die Schwierigkeit bei „1 aus 4“- Items geringer ist als bei jenen im Format „1 aus 6“, dieser ist jedoch statistisch nicht signifikant, was jedoch bei einem angemessenen Risiko 1. und 2. Art an der etwas zu kleinen Stichprobe liegt, wobei allerdings der Effekt mit geschätzten 0,35 der Standardabweichung nicht extrem groß ist.

Vergleicht man gesondert jedoch nur jene Items, die sowohl im MC-Format „1 aus 6“ als auch im Format „1 aus 4“ vorgegeben wurden hinsichtlich ihrer Leichtigkeitparameter mittels t-tests für abhängige Stichproben, zeigt sich ein signifikanter Unterschied in der erwarteten Richtung. Die Items im MC-Format „1 aus 4“ erweisen sich – entsprechend ihrer a-priori-Ratewahrscheinlichkeit - als signifikant leichter als Ihre Entsprechungs-Items im Format „1 aus 6“, was dafür spricht, dass sich der Rateeffekt bei Verringerung der Distraktorenzahl wesentlich schwerwiegender niederschlägt..

Zusammenfassend kann man sagen, dass sich der Rateeffekt (wie erwartet) bei MC-Antwortformaten, die nur das Erkennen einer richtigen unter den (vier bzw. sechs) Antwortmöglichkeiten verlangen, besonders massiv auswirkt, während er beim alternativen MC-Format „x aus 5“ nicht feststellbar ist.

Die Ergebnisse von Gottschall (2007) ließen sich, was den Vergleich der drei Antwortformate „x aus 5“, „1 aus 6“ und freies Antwortformat betrifft, grundsätzlich replizieren.

Dort wie in der vorliegenden Studie zeigt sich, dass bei MC-Antwortformaten der Art „1 aus x“) massive Rateeffekte wirksam werden, die die Schwierigkeit von Items deutlich verringern.

Angesichts der aus der vorliegenden Studie hervorgehenden Tatsache, dass durch die Erhöhung der Anzahl an Distraktoren von drei auf fünf der Rateeffekt zwar verringert, nicht aber gänzlich ausgeschalten werden kann, und außerdem eine Erhöhung der Distraktorenzahl nicht beliebig möglich ist (vgl. Kapitel 2.2.), muss man aus theoretischer und praktischer Sicht dringend vom Einsatz solcher Antwortformate abraten. Bei Leistungstests, besonders in Selektionssituationen würde man sonst schlicht unfair vorgehen. Eine Anweisung, wie sie Moreno, Martinez und Muniz (2006) in ihren Richtlinien zur Gestaltung von MC-Items geben, nicht mehr als drei Antwortmöglichkeiten vorzugeben zugunsten deren Qualität, ist angesichts dieser Ergebnisse sicherlich verfehlt.

Als Alternative hat sich hier wie bei Gottschall (2007) ein MC-Format der Art „x aus 5“ erwiesen, da hier keine Rateeffekte nachgewiesen werden konnten. Einschränkend sei in Frage gestellt, ob ein solches Antwortformat für alle inhaltlichen Aufgabenstellungen sinnvoll zu realisieren ist. So ist es bei Aufgaben, die Verständnis und Transfer gelernter Wissensinhalte verlangen, leichter zu verwirklichen als bei solchen Items, bei denen es um die schlichte Wiedergabe von exaktem Wissen geht, wo schon theoretisch nur eine einzige richtige Antwort existiert. Als Beispiel mögen die Fragen nach der Prävalenzrate einer Erkrankung oder auch nach einem kritischen Wert eines Testergebnisses dienen, welche von „Natur“ aus nur eine einzige Lösung besitzen. In Fällen dieser Art bleibt als einzige faire Alternative die Vorgabe eines offenen Antwortformats. Die Verrechnungssicherheit ist hier ohnedies aufgrund der eindeutigen Identifizierbarkeit der Lösung gegeben. Denkbar wäre natürlich auch, wie bereits in Kapitel 2.2. erwähnt, den Rateeffekt von MC-Items mittels Modellen der IRT zu kontrollieren. Abgesehen von den bereits erwähnten theoretischen Mängeln dieser Ansätze (vgl. Kapitel 2.2.) dürfte dies allerdings am enormen zusätzlichen Aufwand scheitern bzw. an den mangelnden Fähigkeiten von Testautoren scheitern, die nötigen testtheoretischen Analysen durchzuführen.

Für weitere Studien wäre es aber interessant, mit jenen Modellen die Rateeffekte bei MC-Tests eingehender zu untersuchen. Beispielsweise könnte man Schätzungen nach dem 3-PL-Modell oder dem D+G PL-Modell vornehmen, und die resultierenden Rateparameter mit den

a-priori-Ratewahrscheinlichkeiten vergleichen. Denkbar wäre auch, auf diesem Weg zu prüfen, ob sich die Schwierigkeitsparameter z.B. der MC-Formate „1 aus 6“ und „1 aus 4“ gemäß der a-priori-Ratewahrscheinlichkeit vermutet tatsächlich wie 2:3 verhalten. Daraus ließe sich sagen, ob beispielsweise Items mit doppelt so hoher a-Priori-Ratewahrscheinlichkeit auch eine doppelt so hohe faktische Ratewahrscheinlichkeit besitzen. Methodisch sei an dieser Studie kritisiert, dass vor der Untersuchung von Rateeffekten die Daten um jene Items bereinigt wurden, die nicht dem Rasch-Modell (1-PL-Modell) entsprachen. Da dieses Modell keinen Rateparameter vorsieht, könnte man bei a-posteriori Gültigkeit des Modells davon ausgehen, dass eben die ausgeschiedenen Items jene waren, bei denen der Rateeffekt zum tragen kam. Methodisch „sauberer“ müsste man vielmehr ein und dieselben Daten aus einem MC-Test mit verschiedenen Modellen der IRT –mit und ohne Rateparameter– analysieren und anschließend feststellen, bei welchem die Daten die eine bessere Modellanpassung erreichen. Allerdings besteht hier das Problem, dass (bis jetzt) keine geeignete Prüfgröße existiert, um einen etwaigen Unterschied auf statistische Signifikanz zu überprüfen. Und trotzdem bliebe offen, woraus, wenn nicht aus Rateeffekten, sich die deutlichen Schwierigkeitsunterschiede der Items in den verschiedenen Antwortformaten in der vorliegenden Untersuchung ergeben.

Fazit: Der Einsatz von MC-Antwortformat der Art „1 aus x“ (nur eine der gebotenen Antwortmöglichkeiten ist richtig) bei Leistungstests ist messtheoretisch äußerst bedenklich, da die Skalierungsgüte nicht gegeben ist, und besonders in Selektionssituationen unfair. Als Alternative käme, wo es der Aufgabeninhalt erlaubt, ein MC-Format der Art „x aus 5“ in Frage, sofern beim Einsatz eines freien Antwortformates die Verrechnungssicherheit gefährdet scheint.

5. Zusammenfassung und Abstract zur Studie

Abschließend folgt für den überblicksmäßig interessierten Leser eine Zusammenfassung der Studie sowie ein Abstract in englischer Sprache.

5.1. Zusammenfassung

Bei Leistungstests, die im Multiple-Choice-Antwortformat (kurz „MC-Format“) konzipiert sind, besteht grundsätzlich die Möglichkeit, durch bloßes Raten zur Lösung der Aufgabe zu gelangen, ohne über die von Test zu messen beabsichtigte Fähigkeit (in ausreichendem Maß) zu verfügen. Da dies die psychometrische Qualität eines psychologisch-diagnostischen Verfahrens und auch die Fairness des Tests beeinträchtigt, gibt es einige Ansätze in der Psychologie, diesem Rateeffekt zu begegnen. In der vorliegenden Untersuchung wird der Rateeffekt bei verschiedenen Multiple-Choice-Formaten untersucht und zwei konkrete, formal- gestalterische Möglichkeiten zur Verringerung des Rateeffektes auf ihre Wirksamkeit überprüft: Die Erhöhung der Anzahl der Distraktoren und die Erhöhung der möglichen richtigen Antworten unter den Alternativen pro Item. Beide setzen bei der Verringerung der a-priori-Ratewahrscheinlichkeit an und versuchen, diese zu verringern.

Dazu wurden zu bestehenden Aufgaben Entsprechungs-Items konstruiert, das heißt solche, die sich ausschließlich im Antwortformat, nicht aber im eigentlichen Aufgabeninhalt unterscheiden, und zwar in Bezug auf die folgenden vier Antwortformate: Freies Antwortformat, MC-Format „1 aus 6“ (eine richtige ist aus sechs Antwortmöglichkeiten zu wählen), MC-Format „1 aus 4“ (eine richtige ist aus vier Alternativen zu wählen) und MC-Format „x aus 5“ (Von fünf vorgegebenen Antwortmöglichkeiten können keine, eine, zwei, drei, vier, oder sogar alle Möglichkeiten richtig sein, und ein Item gilt nur dann als gelöst, wenn alle richtigen und keine der falschen Möglichkeiten ausgewählt wurde). Diese Items wurden zu fünf verschiedenen Testformen zusammengestellt, von denen jede schließlich je 104 Entsprechungs-Items desselben Aufgabeninhalts enthielt, und einem Experiment 312 Studierenden der Fachrichtung Psychologie zur Bearbeitung vorgegeben. Nach erfolgter Datenerhebung wurden sämtliche Items auf ihre Rasch-Homogenität überprüft und gegebenenfalls ausgeschieden. Aus Daten der verbleibenden Rasch-Modell-konformen Items wurden Itemleichtigkeitsparameter geschätzt, um festzustellen, ob sich die verschiedenen Antwortformate in ihrer Schwierigkeit signifikant voneinander unterscheiden. Zur interferenzstatistischen Überprüfung des Unterschiedes wurde eine einfache Varianzanalyse

durchgeführt, in die das Antwortformat als unabhängige, die Itemleichtigkeitsparameter als abhängige Variable eingingen. Um festzustellen, welche Antwortformate konkret sich von welchen anderen signifikant unterscheiden, wurde post hoc eine Newman-Keuls-Prozedur eingesetzt. Es konnte dabei festgestellt werden, dass das freie Antwortformat und auch das MC-Format „x aus 5“ signifikant niedrigere Itemleichtigkeitsparameter bewirkten als die beiden MC-Formate „1 aus 6“ und „1 aus 4“, was – sofern die Itemkonstruktion gelungen ist - aufgrund des experimentellen Designs nur auf Rateeffekte zurückzuführen ist. Das Format „x aus 5“ stellte sich als nicht signifikant einfacher als das freie Antwortformat heraus, könnte also eine geeignete Alternative zu letzterem im Leistungstests sein. Zwischen den beiden MC-Formaten „1 aus 6“ und „1 aus 4“ bestehen ebenfalls Schwierigkeitsunterschiede in der erwarteten Richtung, sie fallen aber statistisch nicht signifikant aus. Bei einem Vergleich nur jener Items aber, die sowohl im MC-Format „1 aus 6“ als auch im Format „1 aus 4“ vorkommen mittels T-tests für abhängige Stichproben zeigt sich ein signifikanter Unterschied in den Itemleichtigkeitsparametern der beiden Antwortformate; jene Items im Format „1 aus 4“ sind signifikant leichter als ihre Entsprechungs-Items im Format „1 aus 6“. Daraus lässt sich schließen, dass die Erhöhung der Distraktorenzahl pro Item zwar die Aufgabenschwierigkeit signifikant erhöht, der Rateeffekt damit aber nicht gänzlich ausgeschalten werden kann.

Schlagerworte: Antwortformat, Multiple-Choice, Rateeffekt, Itemschwierigkeit, Itemparameter, Rasch-Modell

5.2. Abstract

When performance tests are designed in multiple choice response format there is always a chance to solve the items, even if the testee does not obtain the ability aimed to be tested, just by lucky guessing. This decreases both reliability and fairness of the test. Despite using psychometric models of the item response theory to take the guessing effect into account when scoring a test, there are variable possibilities concerning design of the multiple-choice-formatted items to deal with this problem. In this experimental study item difficulty of free response format and three different multiple choice formats are compared: “1 out of 6” (one correct answer and five distractors), “1 out of 4” (one correct answer and three distractors) and “x out of 5” (either none, one, two, three, four, or even all of the offered answers can be correct, and all of the correct and none of the wrong answers have to be selected to score the

item as mastered. So called parallel items (which differ in response format, but have exactly the same content) were created and assigned to five test forms containing 104 parallel items each. 312 students were randomly selected to work on a test form. Resulting data was analysed using the Rasch-model to establish a scale of the test. A number of items had to be excluded from the item pool. Item difficulty parameters were estimated and compared for each response format. An Anova with these item parameters as the dependent and response format as the independent variable as well as the post hoc Newman-Keuls procedure showed significant difference in item difficulty between free response format and the multiple-choice formats “1 out of 6” and “1 out of 4”. The multiple choice formatted items showed significantly lower item difficulty parameters, which can only be traced back to guessing effects. Items designed in “x out of 5” format did not differ significantly from free response format. Therefore, this format appeared an appropriate alternative to free response format in performance tests. Multiple choice response format “1 out of 4” shows even lower item parameters than the “1 out of 6”- one, as a t-test for paired samples shows.

6. Literaturverzeichnis

- Attali, Y. & Bar-Hillel, M. (2003). Guess Where: The position of Correct Answers in Multiple-Choice Test Items as a Psychometric Variable. *Journal of Educational Measurement*, 40(2), 109-128
- Birnbaum, A. (1968). Some latent trait models and their uses in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Hrsg.), *Statistical theories of mental test scores* (S. 395-479). Reading, MA: Addison-Wesley.
- Bond, G. & Fox, C. (2001). *Applying the Rasch Model – Fundamental Measurement in the Human Sciences*. Mahwah NJ: Lawrence Erlbaum Associates, Inc., USA
- Fischer, G. H. (1995). Derivations of the Rasch Model. In G. H. Fischer & I. W. Molenaar (Hrsg.) *Rasch Models* (S. 15-38). New York: Springer.
- Gittler, G. (1990). *Dreidimensionaler Würfeltest, 3DW*. Weinheim: Beltz Test GmbH.
- Kubinger, K. D. (Hrsg.). (1989). *Moderne Testtheorie – Ein Abriss samt neuesten Beiträgen* (2., verb. Aufl.). Weinheim: Beltz PVU.
- Kubinger, K. D. (2003). Probabilistische Testtheorie. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 415-423). Weinheim: Beltz PVU
- Kubinger, K. D. (2005). Psychological Test Calibration Using the Rasch Model – Some Critical Suggestions on Traditional Approaches. *International Journal of Testing*, 5, 377-394.
- Kubinger, K. D. (2006). *Psychologische Diagnostik – Theorie und Praxis psychologischen Diagnostizierens*. Göttingen: Hogrefe.
- Kubinger, K. D. & Draxler, C. (2006). A comparison of the Rasch model and constrained item response theory models for pertinent psychological test data. In M. von Davier & C. H. Carstensen (Hrsg.), *Multivariate and Mixture Distribution Rasch Models – Extensions and Applications* (S. 295-312). New York: Springer.
- Kubinger, K. D. & Draxler, C. (2007). Probleme bei der Testkonstruktion nach dem Rasch-Modell. *Diagnostica*, 53, 131-143.
- Kubinger, K. D. , Holocher-Ertl, S. & Frebort, M. (2006). Zur testtheoretischen Qualität von Multiple-Choice-Items: 2 richtige aus 5 vs. 1 richtige aus 6 Antwortmöglichkeiten. In B. Gula, R. Alexandrowicz, S. Strauß, E. Brunner, B. Jenull-Schiefer & O. Vitouch (Hrsg.), *Perspektiven psychologischer Forschung in Österreich. Proceedings zur 7.*

- Wissenschaftlichen Tagung der Österreichischen Gesellschaft für Psychologie in Klagenfurt* (S. 459-464). Lengerich: Pabst.
- Litzenberger, M., Gnambs, T. & Punter, J. F. (2005). *Richtlinien zur Konstruktion von Multiple-Choice Items für die Vorlesung „Psychologie als Wissenschaft“ (Ringvorlesung) an der Fakultät für Psychologie*. Unveröff. Forschungsbericht der Fakultät für Psychologie, Wien: Universität Wien.
- Moreno, R., Martinez, R. J. & (2006). New Guidelines for Developing Multiple-Choice Items. *Methodology*, 2, 65-72.
- Ponocny, I. & Klauer, K. C. (2002). Towards identification of unstable personality questionnaire respondents: The use of person fit indices. *Psychologische Beiträge*, 44, 94-107).
- Poinstingl, H., Mair, H. & Hatzinger, R. (dt. Version 2007). *Manual zum Softwarepackage eRm (extended Rasch modelling) – Anwendung des Rasch-Modells (IPL Modell)*. Lengerich: Pabst Science Publishers.
- Puchhammer, M. (1989). Simulationsstudien zur Schätzung der Parameter des Birnbaum-Modells. In K. D. Kubinger (Hrsg.), *Moderne Testtheorie – Ein Abriss samt neuesten Beiträgen* (2., verb. Aufl., S. 259-270). Weinheim: Beltz PVU.
- Seiwald, B. B. (2003). Antwortformat. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 23-28). Weinheim: Beltz PVU.
- Skinner, N. F. (1999). When the Going Gets Tough, The Tough Gets Going: Effects of Order of Item Difficulty on Multiple-Choice Test Performance. *North American Journal of Psychology*, 1(1), 79-82.
- Wagner-Menghin, M. M. (2003). Computerdiagnostik. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 68-82). Weinheim: Beltz PVU.
- Zimbardo, P.G. & Gerrig R. J. (1999). *Psychologie* (7. Aufl.). Berlin: Springer.

7. Anhang

7.1. Anhang A: Itemleichtigkeitsparameter der 154 Rasch-Modell-konformen Items des DIC `08

Anmerkung: Fehler! Verweisquelle konnte nicht gefunden werden. 37 stellt die geschätzten Itemleichtigkeitsparameter aller 188 Rasch-Modell-konformen Items des DIC dar, geordnet nach Skala des DIC. Angegeben sind jeweils die geschätzten Itemleichtigkeitsparameter, die anhand der Gesamtstichprobe $n = 312$ ermittelt wurden und je Skala auf die Summe null normiert wurden, sowie die dazugehörigen Standardschätzfehler (*Standard Error*). Aufgrund der Verrechnung einer Nichtlösung eines Items mit „0“ und einer Lösung mit „1“ sind die resultierenden Itemleichtigkeitsparameter derart zu interpretieren, dass ein hoher Parameter für eine hohe Lösungshäufigkeit und somit eine niedrige Itemschwierigkeit steht, ein niedriger Parameter entsprechend für eine niedrige Lösungshäufigkeit, d.h. für eine hohe Itemschwierigkeit.

Tabelle 37: Itemleichtigkeitsparameter der Aufgaben des DIC, dargestellt pro Skala des DIC

Grundlagenwissen								
Item	Parameter	Std.Error	Item	Parameter	Std.Error	Item	Parameter	Std.Error
i1FrF	-1,254	0.154	i36FrF	,801	0.130	i65a1a6r	1,784	0.388
i2xa5rf	-1,622	0.252	i37a1a6r	1,200	0.202	i65xa5rf	,137	0.330
i2a1a4rf	,857	0.350	i37xa5rf	,137	0.330	i65a1a4r	2,353	0.200
i8FrF	,946	0.562	i37a1a4r	1,380	0.213	i71xa5rf	-,276	0.197
i8xa5rf	,375	0.193	i43xa5rf	-1,444	0.425	i71a1a6r	,672	0.341
i8a1a4rf	1,213	0.511	i43a1a6r	,033	0.330	i72FrF	-4,369	0.505
i9a1a6rf	1,524	0.186	i43a1a4r	,469	0.152	i78xa5rf	-,582	0.174
i9a1a4rf	1,829	0.230	i44a1a6r	-,848	0.255	i78a1a4r	1,732	0.226
i15FrF	1,965	0.231	i50a1a6r	-1,425	0.292	i79a1a6r	1,965	0.231
i15xa5rf	1,623	0.193	i50a1a4r	-,429	0.160	i79xa5rf	1,416	0.384
i16xa5rf	-,033	0.193	i51xa5rf	-1,121	0.390	i79a1a4r	2,204	0.251
i16FrF	-1,978	0.482	i51a1a6r	-1,763	0.449	i85xa5rf	-2,857	0.722
i16a1a4r	1,024	0.204	i51a1a4r	-1,806	0.215	i85a1a6r	,347	0.332
i22FrF	-2,232	0.529	i57FrF	-1,628	0.170	i85a1a4r	,593	0.152
i22a1a6r	-,136	0.194	i58a1a6r	2,689	0.494	i86FrF	-3,640	0.365
i23FrF	-3,202	0.303	i58xa5rf	-1,111	0.376	i92FrF	-2,551	0.600
i29FrF	,469	0.128	i58a1a4r	4,147	0.387	i99FrF	-,130	0.130
i30FrF	,236	0.191	i64xa5rf	,069	0.192	i100FrF	-,463	0.134
i30xa5rf	-,033	0.193	i64FrF	-1,253	0.388			
i30a1a4r	1,734	0.385	i64a1a4r	,259	0.199			

Statistik- und Testtheorie-Wissen								
Item	Parameter	Std.Error	Item	Parameter	Std.Error	Item	Parameter	Std.Error
i3a1a6rf	1,340	0.212	i31xa5rf	-1,365	0.300	i73a1a4r	1,104	0.359
i3FrF	,752	0.200	i31a1a4r	,280	0.368	i80xa5rf	-1,429	0.382
i3a1a4rf	1,461	0.368	i38xa5rf	-,804	0.240	i80a1a4r	,010	0.162
i10xa5rf	1,413	0.214	i38a1a6r	,980	0.365	i87a1a6r	,492	0.350
i10a1a4r	1,715	0.200	i38a1a4r	,856	0.198	i87xa5rf	-,049	0.407
i10FrF	1,157	0.379	i45FrF	-,866	0.165	i87a1a4r	1,228	0.157
i17xa5rf	-,395	0.223	i52FrF	-,568	0.229	i94a1a6r	-1,324	0.492
i17a1a6r	,624	0.374	i52xa5rf	-2,336	0.428	i94xa5rf	-2,764	1.006
i17a1a4r	,481	0.201	i52a1a4r	2,472	0.441	i94a1a4r	-,105	0.165
i24xa5rf	-3,817	0.708	i59FrF	,125	0.138	i101xa5r	-1,880	0.322
i24a1a6r	2,414	0.409	i66FrF	-3,208	0.383	I101A16R	,101	0.212
i24a1a4r	2,070	0.224	i73a1a6r	,383	0.206	I101A14R	,153	0.373
i31a1a6r	-,234	0.218	i73xa5rf	-,470	0.235			

Verfahrensinventar-Kenntnis								
Item	Parameter	Std.Error	Item	Parameter	Std.Error	Item	Parameter	Std.Error
i4xa5rf	,304	0.201	i39xa5rf	,636	0.214	i81a1a4r	4,241	1.019
i4a1a6rf	2,334	0.306	i39a1a4r	1,329	0.415	i88FrF	-,034	0.138
i4a1a4rf	3,486	0.756	i46xa5rf	-2,034	0.274	i95xa5rf	-,547	0.250
i11FrF	-1,017	0.148	i46a1a6r	-1,440	0.405	i95a1a4r	1,069	0.173
i18FrF	,731	0.208	i46a1a4r	-1,958	0.266	i102FrF	-,139	0.138
i18xa5rf	-1,451	0.233	i53FrF	1,502	0.162			
i18a1a4r	1,499	0.429	i60FrF	-3,497	0.462			
i25FrF	,844	0.145	i60a1a6r	-2,242	0.288			
i32xa5rf	-2,511	0.599	i60a1a4r	-1,595	0.443			
i32a1a6r	-,325	0.341	i67FrF	-1,241	0.153			
i32a1a4r	,310	0.160	i81FrF	1,864	0.257			
i39FrF	,236	0.201	i81xa5rf	-,553	0.204			

Theoriewissen zur Differentiellen Psychologie								
Item	Parameter	Std.Error	Item	Parameter	Std.Error	Item	Parameter	Std.Error
i5FrF	-0,991	0.144	i40FrF	2,498	0.338	i68FrF	0,148	0.139
i12FrF	1,669	0.258	i40xa5rf	-1,462	0.220	i75FrF	-2,472	0.190
i12a1a6r	2,800	0.396	i40a1a4r	3,713	0.980	i82FrF	2,152	0.204
i19a1a6r	0,547	0.206	i47FrF	-2,182	0.258	i89FrF	-3,812	0.293
i19FrF	-1,104	0.363	i47a1a4r	1,056	0.393	i96FrF	-0,583	0.139
i19a1a4r	0,896	0.224	i47a1a6r	0,422	0.210	i103FrF	-4,574	0.391
i26FrF	0,006	0.138	i54FrF	-0,791	0.141			

i33Frf	1,900	0.189	i61Frf	0,163	0.139	
--------	-------	-------	--------	-------	-------	--

Verfahrensbezogenes Detailwissen								
Item	Parameter	Std.Error	Item	Parameter	Std.Error	Item	Parameter	Std.Error
i6xa5rf	,287	0.222	i27a1a6r	1,427	0.177	i62Frf	-1,895	0.293
i6a1a4rf	-,952	0.531	i27a1a4r	,446	0.219	i69Frf	1,772	0.139
i13Frf	-,759	0.592	i34Frf	-1,651	0.266	i76Frf	-1,220	0.227
i13xa5rf	-,014	0.236	i41Frf	1,408	0.138	i90Frf	1,424	0.138
i13a1a4r	1,120	0.204	i48Frf	1,248	0.139	i97Frf	,598	0.146
i20Frf	-3,048	0.484	i55Frf	-,183	0.167	i104Frf	-,008	0.161

„Up to Date“								
Item	Parameter	Std.Error	Item	Parameter	Std.Error	Item	Parameter	Std.Error
i7a1a6rf	,416	0.183	i35a1a6r	,602	0.202	i70Frf	-2,009	0.261
i7a1a4rf	1,284	0.207	i35a1a4r	,149	0.392	i77Frf	3,118	0.186
i14xa5rf	-,337	0.241	i42a1a6r	2,046	0.213	i84Frf	-2,453	0.505
i14a1a6r	3,245	0.538	i42xa5rf	,318	0.206	i84xa5rf	-4,402	0.965
i14a1a4r	3,393	0.313	i42a1a4r	2,584	0.458	i91Frf	-2,013	0.261
i21Frf	-1,302	0.203	i49Frf	-1,085	0.190	i98Frf	-,107	0.151
i28Frf	,328	0.143	i56Frf	-1,946	0.255	i105Frf	-1,382	0.208
i35Frf	,815	0.203	i63Frf	-1,263	0.201			

7.2. Anhang B: Testbogendeckblatt

Diagnostik-Info-Check '08

Erläuterung:

Der vorliegende *Diagnostik-Info-Check '08* enthält außer den Skalen

*Grundlagenwissen,
Statistik- und Testtheorie-Wissen,
Verfahrensinventar-Kenntnis,
Theoriewissen zur Differentiellen Psychologie und
Verfahrensbezogenes Detailwissen*

auch eine Skala *Up-to-date*. Die Fragen dieser Skala mögen vordergründig am eigentlichen Zweck vorbeigehen, nämlich das diagnostische Wissen zu prüfen und zur Weiterbildung anzuregen, weil sie hauptsächlich auf das Erscheinungsjahr bzw. die genaue Bezeichnung der aktuellen Version eines einschlägigen psychologisch-diagnostisches Verfahrens abzielen; in einer früheren Studie erwiesen sich aber gerade solche Fragen als wesentliches Indiz für die Bereitschaft von Praktikern zum *Up-dating*. So geht es dabei nur darum, zum Beispiel das Erscheinungsjahr eines Verfahrens größenordnungsmäßig einordnen zu können.

Die Zugehörigkeit zu den einzelnen Skalen muss deshalb nicht speziell gekennzeichnet sein, weil sämtliche Fragen bis jetzt noch nicht entsprechend kalibriert wurden.

Beantworten Sie daher am besten alle Fragen der Reihe nach.

Instruktion:

Manche der Fragen sind im freien Antwortformat verfasst (bitte kurz und prägnant sowie leserlich beantworten), manche im *Multiple-Choice*-Format der Art: 1 der 6 Antwortmöglichkeiten ist richtig (gekennzeichnet mit „①“, um anzuzeigen, nur 1 Antwort ist richtig), manche im *Multiple-Choice*-Format der Art: 1 der 4 Antwortmöglichkeiten ist richtig (ebenfalls gekennzeichnet mit „①“, um anzuzeigen, nur 1 Antwort ist richtig), manche im *Multiple-Choice*-Format der Art: 0, also gar keine, 1 Antwort, 2, 3, 4 oder sogar 5, also alle von 5 Antwortmöglichkeiten ist/sind richtig (gekennzeichnet mit „⊕“, um anzuzeigen, beliebig viele Antworten können richtig sein).

Bitte kreuzen Sie die zutreffende(n) Antwort(en) jeweils deutlich an.

Die Testzeit beträgt ca. 65-70 min., arbeiten Sie bitte zügig und rasch!

Zunächst einige Fragen zu Ihrer Person:

Code:

Geschlecht: m ☐ w ☐

Alter:.....

Ihre jeweiligen Noten im Verlauf der positiv absolvierten Diagnostik-Grundausbildung:

VO I ☐ UE I ☐ VO II ☐ UE II ☐

In welchem Prüfungsfach machen Sie bzw. werden Sie bzw. haben Sie die Diplomarbeit machen/ gemacht?

.....

7.3. Anhang B: Die Items des Diagnostik-Info-Check `08 – dargestellt pro Skala und Testform

Anmerkung: Tabelle 38 veranschaulicht alle 198 Items des Diagnostik-Info-Check`08 – dargestellt pro Skala und Testform A – E. Testform A entspricht der Originalversion des Diagnostik-Info-Check`05, der im Lehrbuch von Kubinger (2006) veröffentlicht wurde. Form B wurde als Parallelversion schon bei Gottschall (2007) konstruiert und vorgegeben. Es handelt sich dabei um dieselben inhaltlichen Items, die sich lediglich im Antwortformat unterscheiden. Die Testformen C, D, und E wurden im Rahmen dieser Arbeit nach demselben Prinzip als Parallelversionen konstruiert. Bei Items im Multiple-Choice-Antwortformat ist (sind) die richtige(n) Lösung(en) unterstrichen. Für die Items im freien Antwortformat ist ein Lösungskatalog in Anhang C zu finden. Weiters sind die Items, welche den Rasch-Modellkontrollen nicht standhielten bzw. zu extreme Lösungshäufigkeiten zeigten und keiner Kontrolle nach dem Rasch-Modell unterzogen werden konnten, zur Kennzeichnung grau unterlegt.

Tabelle 38: Darstellung aller 198 Parallelitens des Diagnostik-Info-Check`08 – dargestellt pro Skala und geordnet nach Testform A/ B/ C/ D/ E. Bei den Items, die im Multiple-Choice-Antwortformat vorgegeben wurden, ist (sind) die richtige(n) Lösung(en) unterstrichen. Weiters sind alle nicht Rasch-Modell-konformen Items bzw. Items, deren Rasch-Homogenität aufgrund zu extremer Lösungshäufigkeiten nicht beurteilt werden konnte, grau unterlegt dargestellt.

Skala Grundlagenwissen

Inhaltl. Item	Testform A	Testform B	Testform C	Testform D	Testform E
1	Welche wissenschaftliche Methode liegt einem psychologischen Test zugrunde?				
2	Die DIN 33430 dient einerseits Personalverantwortlichen bei der Qualitätssicherung und –optimierung von Personalentscheidungen, andererseits	Die DIN 33430 dient einerseits Personalverantwortlichen bei der Qualitätssicherung und –optimierung von Personalentscheidungen, andererseits - dem Schutz der Psychologie vor unlauterer Konkurrenz <u>dem Schutz der Kandidaten vor unsachgemäßer Anwendung von Eignungsverfahren</u>	Die DIN 33430 dient einerseits Personalverantwortlichen bei der Qualitätssicherung und –optimierung von Personalentscheidungen, andererseits	Die DIN 33430 dient einerseits Personalverantwortlichen bei der Qualitätssicherung und –optimierung von Personalentscheidungen, andererseits - dem Schutz der Psychologie vor unlauterer Konkurrenz <u>dem Schutz der Kandidaten vor unsachgemäßer Anwendung von Eignungsverfahren</u>	Die DIN 33430 dient einerseits Personalverantwortlichen bei der Qualitätssicherung und –optimierung von Personalentscheidungen, andererseits - dem Schutz der Psychologie vor unlauterer Konkurrenz <u>dem Schutz der Kandidaten vor unsachgemäßer</u>

		- dem Schutz der Kandidaten in Bezug auf Datenschutz - dem Schutz der Testverlage vor Copyright- Verletzungen <u>dem Schutz der Kandidaten vor missbräuchlicher Anwendung von Eignungsverfahren</u>		- dem Schutz der Kandidaten in Bezug auf Datenschutz - dem Schutz der Testverlage vor Copyright- Verletzungen <u>dem Schutz der Kandidaten vor missbräuchlicher Anwendung von Eignungsverfahren</u>	<u>Anwendung von Eignungsverfahren</u> - dem Schutz der Kandidaten in Bezug auf Datenschutz - dem Schutz der Testverlage vor Copyright- Verletzungen
8	Die Reliabilität eines psychologisch-diagnostischen Verfahrens benötigt man zur	Die Reliabilität eines psychologisch-diagnostischen Verfahrens benötigt man zur <u>Bestimmung des Messfehlers</u> - Bestimmung seiner Nützlichkeit - Bestimmung der Zuverlässigkeit einer Tp - Bestimmung der Wahrscheinlichkeit einer richtigen Prognose <u>Bestimmung des Konfidenzintervalls des wahren Testwerts</u>	Die Reliabilität eines psychologisch-diagnostischen Verfahrens benötigt man zur	Die Reliabilität eines psychologisch-diagnostischen Verfahrens benötigt man zur <u>Bestimmung des Messfehlers</u> - Bestimmung seiner Nützlichkeit - Bestimmung der Zuverlässigkeit einer Tp - Bestimmung der Wahrscheinlichkeit einer richtigen Prognose - Bestimmung des Konfidenzintervalls des wahren Testwerts	Die Reliabilität eines psychologisch-diagnostischen Verfahrens benötigt man zur - Bestimmung seiner Nützlichkeit - Bestimmung der Zuverlässigkeit einer Testperson - Bestimmung der Wahrscheinlichkeit einer richtigen Prognose <u>Bestimmung des Konfidenzintervalls des wahren Testwerts</u>
9	Wie wird üblicherweise die Stabilität eines psychologisch-diagnostischen Verfahrens erfasst? Durch ① <u>Testwiederholung</u> ① Berechnung der konkurrenten Validität ① Paralleltests ① ein Experten-Rating ① die Korrelation mit einem Außenkriterium ① spezifisch objektive Vergleiche	Wie wird üblicherweise die Stabilität eines psychologisch-diagnostischen Verfahrens erfasst? Durch ① <u>Testwiederholung</u> ① Berechnung der konkurrenten Validität ① Paralleltests ① ein Experten-Rating ① die Korrelation mit einem Außenkriterium ① spezifisch objektive Vergleiche	Wie wird üblicherweise die Stabilität eines psychologisch-diagnostischen Verfahrens erfasst? Durch ① <u>Testwiederholung</u> ① Berechnung der konkurrenten Validität ① Paralleltests ① ein Experten-Rating ① die Korrelation mit einem Außenkriterium ① spezifisch objektive Vergleiche	Wie wird üblicherweise die Stabilität eines psychologisch-diagnostischen Verfahrens erfasst? Durch ○ <u>Testwiederholung</u> ○ Paralleltests ○ ein Experten-Rating ○ spezifisch objektive Vergleiche	Wie wird üblicherweise die Stabilität eines psychologisch-diagnostischen Verfahrens erfasst? Durch ○ <u>Testwiederholung</u> ○ Paralleltests ○ ein Experten-Rating ○ spezifisch objektive Vergleiche

15	Wann ist davon zu sprechen, dass ein Test Augenscheinvalidität hat? 	Wann ist davon zu sprechen, dass ein Test Augenscheinvalidität hat? • Dass er nur scheinbar valide ist • <u>Dass es der Tpn augenscheinlich klar ist, was mit ihm zu erfassen beabsichtigt ist</u> • Dass er inhaltliche Gültigkeit besitzt • <u>Dass die Tp seine Messintention durchschaut</u> • Dass für seine Validierung ein Experten-Rating stattgefunden hat	Wann ist davon zu sprechen, dass ein Test Augenscheinvalidität hat? 	Wann ist davon zu sprechen, dass ein Test Augenscheinvalidität hat? • Dass er nur scheinbar valide ist • <u>Dass es der Tpn augenscheinlich klar ist, was mit ihm zu erfassen beabsichtigt ist</u> • Dass er inhaltliche Gültigkeit besitzt • <u>Dass die Tp seine Messintention durchschaut</u> • Dass für seine Validierung ein Experten-Rating stattgefunden hat	Wann ist davon zu sprechen, dass ein Test Augenscheinvalidität hat? • Dass er nur scheinbar valide ist • <u>Dass es der Tpn augenscheinlich klar ist, was mit ihm zu erfassen beabsichtigt ist</u> • Dass er inhaltliche Gültigkeit besitzt • <u>Dass die Tp seine Messintention durchschaut</u> • Dass für seine Validierung ein Experten-Rating stattgefunden hat
16	Was versteht man unter diskriminanter Validität? • Die unterschiedlichen Meinungen der Experten im Zuge eines Experten-Ratings • <u>Dass konstruktferne Tests mit dem interessierenden Test nicht korrelieren</u> • Dass der Test in keinem Faktor hoch lädt • Die Gültigkeit des Tests bei manchen, nicht aber bei allen Personengruppen • Dass konstruktnahe Tests mit dem interessierenden Test hoch korrelieren	Was versteht man unter diskriminanter Validität? 	Was versteht man unter diskriminanter Validität? • Die unterschiedlichen Meinungen der Experten im Zuge eines Experten-Ratings • <u>Dass konstruktferne Tests mit dem interessierenden Test nicht korrelieren</u> • Dass der Test in keinem Faktor hoch lädt • Die Gültigkeit des Tests bei manchen, nicht aber bei allen Personengruppen • Dass konstruktnahe Tests mit dem interessierenden Test hoch korrelieren	Was versteht man unter diskriminanter Validität? - o Die unterschiedlichen Meinungen der Experten im Zuge eines Experten-Ratings - o <u>Dass konstruktferne Tests mit dem interessierenden Test nicht korrelieren</u> - o Die Gültigkeit des Tests bei manchen, nicht aber bei allen Personengruppen - o Dass konstruktnahe Tests mit dem interessierenden Test hoch korrelieren	Was versteht man unter diskriminanter Validität? - o Die unterschiedlichen Meinungen der Experten im Zuge eines Experten-Ratings - o <u>Dass konstruktferne Tests mit dem interessierenden Test nicht korrelieren</u> - o Die Gültigkeit des Tests bei manchen, nicht aber bei allen Personengruppen - o Dass konstruktnahe Tests mit dem interessierenden Test hoch korrelieren

22	Sensitivität beschreibt den Prozentsatz ☐ falsch positiver Diagnosen ☐ falsch negativer Diagnosen ☐ <u>richtig positiver Diagnosen</u> ☐ richtig negativer Diagnosen ☐ richtiger Diagnosen ☐ falscher Diagnosen	Sensitivität beschreibt den Prozentsatz von	Sensitivität beschreibt den Prozentsatz ☐ falsch positiver Diagnosen ☐ falsch negativer Diagnosen ☐ <u>richtig positiver Diagnosen</u> ☐ richtig negativer Diagnosen ☐ richtiger Diagnosen ☐ falscher Diagnosen	Sensitivität beschreibt den Prozentsatz ☐ falsch positiver Diagnosen ☐ <u>richtig positiver Diagnosen</u> ☐ richtig negativer Diagnosen ☐ richtiger Diagnosen	Sensitivität beschreibt den Prozentsatz ☐ falsch positiver Diagnosen ☐ <u>richtig positiver Diagnosen</u> ☐ richtig negativer Diagnosen ☐ richtiger Diagnosen
23	Validierungsversuche über eine sog. MTMM-Matrix prüfen, ob psychologisch-diagnostische Verfahren einerseits sowohl konvergente als auch diskriminante Validität aufweisen, andererseits ob das zu erfassen gesuchte Konstrukt				
29	Zum diagnostischen Prozess zählen: Klärung der Fragestellung, Auswahl der diagnostischen Verfahren, Anwendung und Auswertung der diagnostischen Verfahren, Interpretation und Gutachtenerstellung sowie				
30	Laut DIN 33430 ist die Angemessenheit der Normwerte (Eichtabellen) spätestens nach jeweils wie vielen Jahren zu prüfen?	Laut DIN 33430 ist die Angemessenheit der Normwerte (Eichtabellen) spätestens nach jeweils wie vielen Jahren zu prüfen? ☒ 10 ☐ 5 ☐ 7 ☒ <u>8</u> ☐ jährlich	Laut DIN 33430 ist die Angemessenheit der Normwerte (Eichtabellen) spätestens nach jeweils wie vielen Jahren zu prüfen?	Laut DIN 33430 ist die Angemessenheit der Normwerte (Eichtabellen) spätestens nach jeweils wie vielen Jahren zu prüfen? ☐ 10 ☐ 5 ☐ <u>8</u> ☐ jährlich	Laut DIN 33430 ist die Angemessenheit der Normwerte (Eichtabellen) spätestens nach jeweils wie vielen Jahren zu prüfen? ☐ 10 ☐ 5 ☐ <u>8</u> ☐ jährlich
36	Voraussetzung für adaptives Testen ist				
37	Wenn ein Test mit passenden Items ergänzt wird, erhöht sich regelmäßig die ☐ Objektivität ☐ Validität ☐ Nützlichkeit ☐ <u>Messgenauigkeit</u> ☐ Fairnes ☐ Ökonomie	Wenn ein Test mit passenden Items ergänzt wird, erhöht sich regelmäßig die ☒ <u>Reliabilität</u> ☒ Validität ☒ Ökonomie ☒ <u>Messgenauigkeit</u> ☒ Fairness	Wenn ein Test mit passenden Items ergänzt wird, erhöht sich regelmäßig die ☐ Objektivität ☐ Validität ☐ Nützlichkeit ☐ <u>Messgenauigkeit</u> ☐ Fairnes ☐ Ökonomie	Wenn ein Test mit passenden Items ergänzt wird, erhöht sich regelmäßig die ☐ Validität ☐ Nützlichkeit ☐ <u>Messgenauigkeit</u> ☐ Fairness	Wenn ein Test mit passenden Items ergänzt wird, erhöht sich regelmäßig die ☐ Validität ☐ Nützlichkeit ☐ <u>Messgenauigkeit</u> ☐ Fairness

43	Eine Reduzierung des Rateeffekts beim <i>Multiple-Choice</i>-Format kann man erreichen durch ✱ Anwendung des <i>Rasch</i>-Modells ✱ Anwendung des <i>Rasch</i>-Modells mit Rateparameter ✱ Anwendung des 3-PL Modells ✱ <u>Erhöhung der Anzahl der Distraktoren</u> ✱ <u>Erhöhung der Anzahl richtiger Antwortmöglichkeiten</u>	Eine Reduzierung des Rateeffekts beim <i>Multiple-Choice</i>-Format kann man erreichen durch ① Anwendung des <i>Rasch</i>-Modells ① Anwendung des <i>Rasch</i>-Modells mit Rateparameter ① Anwendung des 3-PL Modells ① Anwendung des 2-PL Modells ① Erhöhung der Reliabilität ① <u>Erhöhung der Anzahl richtiger Antwortmöglichkeiten</u>	Eine Reduzierung des Rateeffekts beim <i>Multiple-Choice</i>-Format kann man erreichen durch ① Anwendung des <i>Rasch</i>-Modells mit Rateparameter ① Anwendung des 3-PL Modells ① Erhöhung der Reliabilität ① <u>Erhöhung der Anzahl richtiger Antwortmöglichkeiten</u>	Eine Reduzierung des Rateeffekts beim <i>Multiple-Choice</i>-Format kann man erreichen durch ① Anwendung des <i>Rasch</i>-Modells mit Rateparameter ① Anwendung des 3-PL Modells ① Erhöhung der Reliabilität ① <u>Erhöhung der Anzahl richtiger Antwortmöglichkeiten</u>	Eine Reduzierung des Rateeffekts beim <i>Multiple-Choice</i>-Format kann man erreichen durch ① Anwendung des <i>Rasch</i>-Modells mit Rateparameter ① Anwendung des 3-PL Modells ① Erhöhung der Reliabilität ① <u>Erhöhung der Anzahl richtiger Antwortmöglichkeiten</u>
44	Unter Stichprobenunabhängigkeit versteht man, dass ① das <i>Rasch</i>-Modell gilt ① eines der <i>Birnbaum</i>-Modelle gilt ① alle Stichproben zum exakt gleichen Ergebnis führen ① <u>die Parameterschätzungen in allen Stichproben statistisch gleich sind</u> ① leistungsstarken und leistungsschwachen Testpersonen die Items gleich schwer fallen ① man für die Eichung keine repräsentativen Stichproben benötigt	Unter Stichprobenunabhängigkeit versteht man, dass ① das <i>Rasch</i>-Modell gilt ① eines der <i>Birnbaum</i>-Modelle gilt ① alle Stichproben zum exakt gleichen Ergebnis führen ① <u>die Parameterschätzungen in allen Stichproben statistisch gleich sind</u> ① leistungsstarken und leistungsschwachen Testpersonen die Items gleich schwer fallen ① man für die Eichung keine repräsentativen Stichproben benötigt	Unter Stichprobenunabhängigkeit versteht man, dass ① das <i>Rasch</i>-Modell gilt ① <u>die Parameterschätzungen in allen Stichproben statistisch gleich sind</u> ① leistungsstarken und leistungsschwachen Testpersonen die Items gleich schwer fallen ① man für die Eichung keine repräsentativen Stichproben benötigt	Unter Stichprobenunabhängigkeit versteht man, dass ① das <i>Rasch</i>-Modell gilt ① <u>die Parameterschätzungen in allen Stichproben statistisch gleich sind</u> ① leistungsstarken und leistungsschwachen Testpersonen die Items gleich schwer fallen ① man für die Eichung keine repräsentativen Stichproben benötigt	Unter Stichprobenunabhängigkeit versteht man, dass ① das <i>Rasch</i>-Modell gilt ① <u>die Parameterschätzungen in allen Stichproben statistisch gleich sind</u> ① leistungsstarken und leistungsschwachen Testpersonen die Items gleich schwer fallen ① man für die Eichung keine repräsentativen Stichproben benötigt

50	Unter Profilinterpretation versteht man, dass ① nicht alle erhaltenen Testwerte interpretiert werden ① nur die über- und unterdurchschnittlichen Testwerte interpretiert werden ① die über- und unterdurchschnittlichen Testwerte besonders hervorgehoben werden ① <u>nur die signifikant von einander abweichenden Testwerte interpretiert werden</u> ① nur die Leistungsspitzen interpretiert werden ① der IQ als typisch für die Fähigkeiten einer Testperson interpretiert wird	Unter Profilinterpretation versteht man, dass ① nicht alle erhaltenen Testwerte interpretiert werden ① nur die über- und unterdurchschnittlichen Testwerte interpretiert werden ① die über- und unterdurchschnittlichen Testwerte besonders hervorgehoben werden ① <u>nur die signifikant von einander abweichenden Testwerte interpretiert werden</u> ① nur die Leistungsspitzen interpretiert werden ① der IQ als typisch für die Fähigkeiten einer Testperson interpretiert wird	Unter Profilinterpretation versteht man, dass ① nicht alle erhaltenen Testwerte interpretiert werden ① nur die über- und unterdurchschnittlichen Testwerte interpretiert werden ① <u>nur die signifikant von einander abweichenden Testwerte interpretiert werden</u> ① der IQ als typisch für die Fähigkeiten einer Testperson interpretiert wird	Unter Profilinterpretation versteht man, dass ① nicht alle erhaltenen Testwerte interpretiert werden ① nur die über- und unterdurchschnittlichen Testwerte interpretiert werden ① <u>nur die signifikant von einander abweichenden Testwerte interpretiert werden</u> ① der IQ als typisch für die Fähigkeiten einer Testperson interpretiert wird	Unter Profilinterpretation versteht man, dass ① nicht alle erhaltenen Testwerte interpretiert werden ① nur die über- und unterdurchschnittlichen Testwerte interpretiert werden ① <u>nur die signifikant von einander abweichenden Testwerte interpretiert werden</u> ① der IQ als typisch für die Fähigkeiten einer Testperson interpretiert wird
51	Die <i>a-priori</i> Ratewahrscheinlichkeit beim <i>Multiple-Choice</i>-Format ist bei durchschnittlich leistungsstarken Testpersonen zumeist ✱ gleich wie ✱ größer als ✱ <u>kleiner als</u> ✱ kleiner, bei einer sehr großen Itemanzahl jedoch größer als ✱ größer, bei einer sehr großen Itemanzahl jedoch kleiner als die faktische Ratewahrscheinlichkeit.	Die <i>a-priori</i> Ratewahrscheinlichkeit beim <i>Multiple-Choice</i>-Format ist bei durchschnittlich leistungsstarken Testpersonen zumeist ① gleich wie ① größer als ① <u>kleiner als</u> ① kleiner, bei einer sehr großen Itemanzahl jedoch größer als ① größer, bei einer sehr großen Itemanzahl jedoch kleiner als ① halb so groß wie die faktische Ratewahrscheinlichkeit.	Die <i>a-priori</i> Ratewahrscheinlichkeit beim <i>Multiple-Choice</i>-Format ist bei durchschnittlich leistungsstarken Testpersonen zumeist ① gleich wie ① größer als ① <u>kleiner als</u> ① kleiner, bei einer sehr großen Itemanzahl jedoch größer als die faktische Ratewahrscheinlichkeit.	Die <i>a-priori</i> Ratewahrscheinlichkeit beim <i>Multiple-Choice</i>-Format ist bei durchschnittlich leistungsstarken Testpersonen zumeist ① gleich wie ① größer als ① <u>kleiner als</u> ① kleiner, bei einer sehr großen Itemanzahl jedoch größer als die faktische Ratewahrscheinlichkeit.	Die <i>a-priori</i> Ratewahrscheinlichkeit beim <i>Multiple-Choice</i>-Format ist bei durchschnittlich leistungsstarken Testpersonen zumeist ① gleich wie ① größer als ① kleiner als ① kleiner, bei einer sehr großen Itemanzahl jedoch größer als die faktische Ratewahrscheinlichkeit.
57	Der Gegensatz von Spezifität ist				

58	Unter dem Testgütekriterium der Fairness versteht man ① die ausreichende Messgenauigkeit des Tests ① dass bei der Interpretation eines Testergebnisses alle Begleitumstände berücksichtigt werden ① <u>dass es zu keiner systematischen Benachteiligung bestimmter Personen kommt</u> ① dass die diagnostische Zielsetzung des eingesetzten Verfahrens der Testperson mitgeteilt wird ① die maximale Nutzung der durch die Testung gewonnenen Informationen ① dass keine geschlechtsspezifischen Eichtabellen benötigt werden	Unter dem Testgütekriterium der Fairness versteht man ✱ die ausreichende Messgenauigkeit des Tests ✱ dass bei der Interpretation eines Testergebnisses alle Begleitumstände berücksichtigt werden ✱ dass die diagnostische Zielsetzung des eingesetzten Verfahrens der Testperson mitgeteilt wird ✱ die maximale Nutzung der durch die Testung gewonnenen Informationen ✱ dass keine geschlechtsspezifischen Eichtabellen benötigt werden	Unter dem Testgütekriterium der Fairness versteht man ① dass bei der Interpretation eines Testergebnisses alle Begleitumstände berücksichtigt werden ① <u>dass es zu keiner systematischen Benachteiligung bestimmter Personen kommt</u> ① dass die diagnostische Zielsetzung des eingesetzten Verfahrens der Testperson mitgeteilt wird ① dass keine geschlechtsspezifischen Eichtabellen benötigt werden	Unter dem Testgütekriterium der Fairness versteht man ① dass bei der Interpretation eines Testergebnisses alle Begleitumstände berücksichtigt werden ① <u>dass es zu keiner systematischen Benachteiligung bestimmter Personen kommt</u> ① dass die diagnostische Zielsetzung des eingesetzten Verfahrens der Testperson mitgeteilt wird ① dass keine geschlechtsspezifischen Eichtabellen benötigt werden	Unter dem Testgütekriterium der Fairness versteht man ① dass bei der Interpretation eines Testergebnisses alle Begleitumstände berücksichtigt werden ① <u>dass es zu keiner systematischen Benachteiligung bestimmter Personen kommt</u> ① dass die diagnostische Zielsetzung des eingesetzten Verfahrens der Testperson mitgeteilt wird ① dass keine geschlechtsspezifischen Eichtabellen benötigt werden
64	Die innere Konsistenz ✱ entspricht der inhaltlichen Gültigkeit ✱ entspricht der konkurrenten Validität ✱ <u>ist ein besonderes Maß der Reliabilität</u> ✱ <u>bezieht sich auf das Ausmaß, mit dem die einzelnen Items eines Tests korrelieren</u> ✱ bezieht sich auf die Testleiterunabhängigkeit	Was versteht man unter innerer Konsistenz? 	Die innere Konsistenz ✱ entspricht der inhaltlichen Gültigkeit ✱ entspricht der konkurrenten Validität ✱ ist ein besonderes Maß der Reliabilität ✱ bezieht sich auf das Ausmaß, mit dem die einzelnen Items eines Tests korrelieren ✱ bezieht sich auf die Testleiterunabhängigkeit	Die innere Konsistenz ① entspricht der inhaltlichen Gültigkeit ① entspricht der konkurrenten Validität ① <u>ist ein besonderes Maß der Reliabilität</u> ① bezieht sich auf die Testleiterunabhängigkeit	Die innere Konsistenz ① entspricht der inhaltlichen Gültigkeit ① entspricht der konkurrenten Validität ① <u>ist ein besonderes Maß der Reliabilität</u> ① bezieht sich auf die Testleiterunabhängigkeit

65	Unter einem Objektiven Persönlichkeits<i>test</i> versteht man ① einen besonders reliablen und validen Leistungstest ① einen besonders validen Persönlichkeitsfragebogen ① einen Test, der der Testperson ausreichend Einsicht in die Testabsicht gewährt ① <u>ein Verfahren, das Verfälschungen weitgehend unmöglich macht</u> ① einen Persönlichkeitsfragebogen mit einer Lügenskala ① die Messung psychophysiologischer Parameter	Unter einem Objektiven Persönlichkeits<i>test</i> versteht man ✳ einen besonders reliablen und validen Leistungstest ✳ <u>ein Verfahren, dass die persönlichen Verhaltensstile bei Leistungsanforderungen beobachtet</u> ✳ einen Test, der der Testperson ausreichend Einsicht in die Testabsicht gewährt ✳ <u>ein Verfahren, das Verfälschungen weitgehend unmöglich macht</u> ✳ einen Persönlichkeitsfragebogen mit einer Lügenskala	Unter einem Objektiven Persönlichkeits<i>test</i> versteht man ① einen besonders reliablen und validen Leistungstest ① einen besonders validen Persönlichkeitsfragebogen ① <u>ein Verfahren, das Verfälschungen weitgehend unmöglich macht</u> ① die Messung psychophysiologischer Parameter	Unter einem Objektiven Persönlichkeits<i>test</i> versteht man ① einen besonders reliablen und validen Leistungstest ① einen besonders validen Persönlichkeitsfragebogen ① <u>ein Verfahren, das Verfälschungen weitgehend unmöglich macht</u> ① die Messung psychophysiologischer Parameter	Unter einem Objektiven Persönlichkeits<i>test</i> versteht man ① einen besonders reliablen und validen Leistungstest ① einen besonders validen Persönlichkeitsfragebogen ① <u>ein Verfahren, das Verfälschungen weitgehend unmöglich macht</u> ① die Messung psychophysiologischer Parameter
71	Von einem <i>Speed-Test</i> spricht man, wenn ✳ bei bestimmten Testpersonen die Items mit verkürzter Bearbeitungszeit vorgegeben werden ✳ <u>die Testleistung ausschließlich nach der Schnelligkeit beurteilt wird</u> ✳ die Testleistung zusätzlich zur Leistungsgüte auch nach der Schnelligkeit beurteilt wird ✳ ein Test mit sehr schwierigen Items nach einer fixen Vorgabezeit abgebrochen wird ✳ ein Test eine sehr kurze Durchführungsdauer hat	Von einem <i>Speed-Test</i> spricht man, wenn ① bei bestimmten Testpersonen die Items mit verkürzter Bearbeitungszeit vorgegeben werden ① <u>die Testleistung ausschließlich nach der Schnelligkeit beurteilt wird</u> ① die Testleistung zusätzlich zur Leistungsgüte auch nach der Schnelligkeit beurteilt wird ① ein Test mit sehr schwierigen Items nach einer fixen Vorgabezeit abgebrochen wird ① ein Test eine sehr kurze Durchführungsdauer hat ① ein Test nur aus wenigen Items besteht	Von einem <i>Speed-Test</i> spricht man, wenn ✳ bei bestimmten Testpersonen die Items mit verkürzter Bearbeitungszeit vorgegeben werden ✳ <u>die Testleistung ausschließlich nach der Schnelligkeit beurteilt wird</u> ✳ die Testleistung zusätzlich zur Leistungsgüte auch nach der Schnelligkeit beurteilt wird ✳ ein Test mit sehr schwierigen Items nach einer fixen Vorgabezeit abgebrochen wird ✳ ein Test eine sehr kurze Durchführungsdauer hat	Von einem <i>Speed-Test</i> spricht man, wenn ① bei bestimmten Testpersonen die Items mit verkürzter Bearbeitungszeit vorgegeben werden ① <u>die Testleistung ausschließlich nach der Schnelligkeit beurteilt wird</u> ① ein Test mit sehr schwierigen Items nach einer fixen Vorgabezeit abgebrochen wird ① ein Test eine sehr kurze Durchführungsdauer hat	Von einem <i>Speed-Test</i> spricht man, wenn ① bei bestimmten Testpersonen die Items mit verkürzter Bearbeitungszeit vorgegeben werden ① <u>die Testleistung ausschließlich nach der Schnelligkeit beurteilt wird</u> ① ein Test mit sehr schwierigen Items nach einer fixen Vorgabezeit abgebrochen wird ① ein Test eine sehr kurze Durchführungsdauer hat

72	Unter Äquivalenzprüfung zweier Versionen eines psychologisch-diagnostischen Verfahrens versteht man,				
78	Wofür ist die Geltung des <i>Rasch</i>-Modells notwendig? Um ✱ mit hoher Wahrscheinlichkeit Normalverteilung der Testwerte zu erhalten ✱ sicherzustellen, dass die <u>Anzahl gelöster Aufgaben einen fairen Testwert liefert</u> ✱ einen <i>Culture-Fair</i> Test zu erhalten ✱ den Rateeffekt im <i>multiple-choice</i> Format abschätzen zu können ✱ den Standardmessfehler exakt bestimmen zu können	Wofür ist die Geltung des <i>Rasch</i>-Modells notwendig? Um ✱ mit hoher Wahrscheinlichkeit Normalverteilung der Testwerte zu erhalten ✱ <u>sicherzustellen, dass die Anzahl gelöster Aufgaben einen fairen Testwert liefert</u> ✱ einen <i>Culture-Fair</i> Test zu erhalten ✱ den Rateeffekt im <i>multiple-choice</i> Format abschätzen zu können ✱ den Standardmessfehler exakt bestimmen zu können	Wofür ist die Geltung des <i>Rasch</i>-Modells notwendig? Um ✱ mit hoher Wahrscheinlichkeit Normalverteilung der Testwerte zu erhalten ✱ <u>sicherzustellen, dass die Anzahl gelöster Aufgaben einen fairen Testwert liefert</u> ✱ einen <i>Culture-Fair</i> Test zu erhalten ✱ den Rateeffekt im <i>multiple-choice</i> Format abschätzen zu können ✱ den Standardmessfehler exakt bestimmen zu können	Wofür ist die Geltung des <i>Rasch</i>-Modells notwendig? Um ① mit hoher Wahrscheinlichkeit Normalverteilung der Testwerte zu erhalten ① <u>sicherzustellen, dass die Anzahl gelöster Aufgaben einen fairen Testwert liefert</u> ① einen <i>Culture-Fair</i> Test zu erhalten ① den Standardmessfehler exakt bestimmen zu können	Wofür ist die Geltung des <i>Rasch</i>-Modells notwendig? Um ① mit hoher Wahrscheinlichkeit Normalverteilung der Testwerte zu erhalten ① <u>sicherzustellen, dass die Anzahl gelöster Aufgaben einen fairen Testwert liefert</u> ① einen <i>Culture-Fair</i> Test zu erhalten ① den Standardmessfehler exakt bestimmen zu können
79	Von <i>face-validity</i> eines Tests spricht man, wenn ① die Validität des Tests unmittelbar einsichtig ist ① der Test nicht validiert wurde ① <u>das, was der Test messen soll, für die Testperson offensichtlich ist</u> ① die Validität durch ein Experten-<i>Rating</i> bestätigt ist ① die Items eines Tests hohe innere Konsistenz haben ① der Test erfolgreich validiert wurde	Von <i>face-validity</i> eines Tests spricht man, wenn ✱ die Validität des Tests unmittelbar einsichtig ist ✱ der Test nicht validiert wurde ✱ <u>das, was der Test messen soll, für die Testperson offensichtlich ist</u> ✱ die Items eines Tests hohe innere Konsistenz haben ✱ <u>die Messintention durchschaubar ist</u>	Von <i>face-validity</i> eines Tests spricht man, wenn ① die Validität des Tests unmittelbar einsichtig ist ① der Test nicht validiert wurde ① <u>das, was der Test messen soll, für die Testperson offensichtlich ist</u> ① die Validität durch ein Experten-<i>Rating</i> bestätigt ist ① die Items eines Tests hohe innere Konsistenz haben ① der Test erfolgreich validiert wurde	Von <i>face-validity</i> eines Tests spricht man, wenn ① die Validität des Tests unmittelbar einsichtig ist ① der Test nicht validiert wurde ① <u>das, was der Test messen soll, für die Testperson offensichtlich ist</u> ① der Test erfolgreich validiert wurde	Von <i>face-validity</i> eines Tests spricht man, wenn ① die Validität des Tests unmittelbar einsichtig ist ① der Test nicht validiert wurde ① <u>das, was der Test messen soll, für die Testperson offensichtlich ist</u> ① der Test erfolgreich validiert wurde

85	Ein relativ junger Begriff in der Psychologischen Diagnostik ist: ✱ Ökologische Reliabilität ✱ Stanine-Wert ✱ Intelligenzquotient ✱ <u>Berufswahlreife</u> ✱ <u>Posttest</u>	Ein relativ junger Begriff in der Psychologischen Diagnostik ist: ① Ökologische Reliabilität ① Stanine-Wert ① Intelligenzquotient ① <u>Berufswahlreife</u> ① Prozentrang ① Anamnese	Ein relativ junger Begriff in der Psychologischen Diagnostik ist: ① Ökologische Reliabilität ① Intelligenzquotient ① <u>Berufswahlreife</u> ① Anamnese	Ein relativ junger Begriff in der Psychologischen Diagnostik ist: ① Ökologische Reliabilität ① Intelligenzquotient ① <u>Berufswahlreife</u> ① Anamnese	Ein relativ junger Begriff in der Psychologischen Diagnostik ist: ① Ökologische Reliabilität ① Intelligenzquotient ① <u>Berufswahlreife</u> ① Anamnese
86	Was ist unter Sinnhaftigkeitsprüfung von psychologisch-diagnostischen Verfahren zu verstehen?				
92	Die DIN 33430 hat den Rechtsstatus ① eines Gesetzes ① einer Verwaltungsverordnung ① eines fachautorisierten Appells ① <u>der rechtsnächsten Norm</u> ① eines Vereinsstatus' ① einer Vereinbarung nach Verfassungsrecht	Die DIN 33430 hat den Rechtsstatus von	Die DIN 33430 hat den Rechtsstatus ① eines Gesetzes ① einer Verwaltungsverordnung ① eines fachautorisierten Appells ① <u>der rechtsnächsten Norm</u> ① eines Vereinsstatus' ① einer Vereinbarung nach Verfassungsrecht	Die DIN 33430 hat den Rechtsstatus ① einer Verwaltungsverordnung ① eines fachautorisierten Appells ① <u>der rechtsnächsten Norm</u> ① eines Vereinsstatus'	Die DIN 33430 hat den Rechtsstatus ① einer Verwaltungsverordnung ① eines fachautorisierten Appells ① <u>der rechtsnächsten Norm</u> ① eines Vereinsstatus'
93	Die sequentielle Vorgabe der Antwortmöglichkeiten in einem Leistungstest – bei der sich die Testperson sukzessive für „richtig“ oder „falsch“ entscheiden muss -, dient				
99	Das Gütekriterium der Zumutbarkeit eines psychologisch-diagnostischen Verfahrens setzt die energetisch-motivationale und emotionale Beanspruchung einer Testperson in Relation zu				
100	Die Alternative zur Selektionsdiagnostik ist die.....				

Skala Statistik- und Testtheorie-Wissen

Inhaltl. Item	Testform A	Testform B	Testform C	Testform D	Testform E
3	Zumeist werden diejenigen Testleistungen als „durchschnittlich“ bezeichnet, welche einem Prozentrang zwischen 25 und 75 entsprechen; so gesehen gilt ein IQ als durchschnittlich, wenn er zwischen ① 70 und 130 ① 80 und 120 ① 55 und 145 ① <u>90 und 110</u> ① 85 und 115 ① 95 und 105 liegt.	Unter Gegenübertragung versteht man das Phänomen der Übertragung bei	Zumeist werden diejenigen Testleistungen als „durchschnittlich“ bezeichnet, welche einem Prozentrang zwischen 25 und 75 entsprechen; so gesehen gilt ein IQ als durchschnittlich, wenn er zwischen ① 70 und 130 ① 80 und 120 ① 55 und 145 ① <u>90 und 110</u> ① 85 und 115 ① 95 und 105 liegt.	Unter Gegenübertragung versteht man das Phänomen der Übertragung bei	Zumeist werden diejenigen Testleistungen als „durchschnittlich“ bezeichnet, welche einem Prozentrang zwischen 25 und 75 entsprechen; so gesehen gilt ein IQ als durchschnittlich, wenn er zwischen ① 70 und 130 ① 80 und 120 ① <u>90 und 110</u> ① 85 und 115 liegt.
10	Ein erreichter Prozentrang von 90% einer Tp in einem Leistungstest bedeutet, ✱ dass die Tp 90% der Aufgaben gelöst hat ✱ nur 10% der je getesteten Personen bessere Leistungen erzielen ✱ <u>90% der Referenzpopulation schlechtere oder höchstens gleich gute Leistungen erzielen</u> ✱ die Testleistung zu 90% als überdurchschnittlich zu bewerten ist ✱ <u>nur 10% der Referenzpopulation bessere Leistungen erzielen</u>	Ein erreichter Prozentrang von 90% einer Tp in einem Leistungstest bedeutet, 	Ein erreichter Prozentrang von 90% einer Tp in einem Leistungstest bedeutet, ✱ dass die Tp 90% der Aufgaben gelöst hat ✱ nur 10% der je getesteten Personen bessere Leistungen erzielen ✱ <u>90% der Referenzpopulation schlechtere oder höchstens gleich gute Leistungen erzielen</u> ✱ die Testleistung zu 90% als überdurchschnittlich zu bewerten ist ✱ <u>nur 10% der Referenzpopulation bessere Leistungen erzielen</u>	Ein erreichter Prozentrang von 90% einer Tp in einem Leistungstest bedeutet, 	Ein erreichter Prozentrang von 90% einer Tp in einem Leistungstest bedeutet, ① die Tp 90% der Aufgaben gelöst hat ① nur 10% der je getesteten Personen bessere Leistungen erzielen ① <u>90% der Referenzpopulation schlechtere oder höchstens gleich gute Leistungen erzielen</u> ① die Testleistung zu 90% als überdurchschnittlich zu bewerten ist

17	Standardmessfehler der <i>Klassischen Testtheorie</i> und Standardschätzfehler der <i>Probabilistischen Testtheorie</i> unterscheiden sich dadurch, dass ✳ erster immer größer als zweiter ist ✳ erster immer kleiner als zweiter ist ✳ <u>erster für alle Testwerte gleich ist, zweiter je nach Niveau verschieden sein kann</u> ✳ <u>erster mit der Reliabilität zusammenhängt, zweiter nicht</u> ✳ erster von der Validität abhängt, zweiter nicht	Standardmessfehler der <i>Klassischen Testtheorie</i> und Standardschätzfehler der <i>Probabilistischen Testtheorie</i> unterscheiden sich dadurch, dass ① erster immer größer als zweiter ist ① erster immer kleiner als zweiter ist ① <u>erster für alle Testwerte gleich ist, zweiter je nach Niveau verschieden sein kann</u> ① erster schwer, zweiter leicht mit Hand (ohne Rechner) berechnet werden kann ① erster von der Validität abhängt, zweiter nicht ① erster von der Irrtumswahrscheinlichkeit abhängt, zweiter nicht	Standardmessfehler der <i>Klassischen Testtheorie</i> und Standardschätzfehler der <i>Probabilistischen Testtheorie</i> unterscheiden sich dadurch, dass ✳ erster immer größer als zweiter ist ✳ erster immer kleiner als zweiter ist ✳ <u>erster für alle Testwerte gleich ist, zweiter je nach Niveau verschieden sein kann</u> ✳ <u>erster mit der Reliabilität zusammenhängt, zweiter nicht</u> ✳ erster von der Validität abhängt, zweiter nicht	Standardmessfehler der <i>Klassischen Testtheorie</i> und Standardschätzfehler der <i>Probabilistischen Testtheorie</i> unterscheiden sich dadurch, dass ① erster immer größer als zweiter ist ① <u>erster für alle Testwerte gleich ist, zweiter je nach Niveau verschieden sein kann</u> ① erster von der Validität abhängt, zweiter nicht ① erster von der Irrtumswahrscheinlichkeit abhängt, zweiter nicht	Standardmessfehler der <i>Klassischen Testtheorie</i> und Standardschätzfehler der <i>Probabilistischen Testtheorie</i> unterscheiden sich dadurch, dass ① erster immer größer als zweiter ist ① <u>erster für alle Testwerte gleich ist, zweiter je nach Niveau verschieden sein kann</u> ① erster von der Validität abhängt, zweiter nicht ① erster von der Irrtumswahrscheinlichkeit abhängt, zweiter nicht
24	Die herkömmliche Faktorenanalyse kann auf folgende Daten standardmäßig angewendet werden, ohne artifizielle Ergebnisse zu liefern: ✳ Dichotome Daten ✳ <u>Intervallskalierte Daten</u> ✳ <u>T-Werte</u> ✳ <u>Centil-Werte</u> ✳ Stanine-Werte	Die herkömmliche Faktorenanalyse kann auf folgende Daten standardmäßig angewendet werden, ohne artifizielle Ergebnisse zu liefern: ① Dichotome Daten ① Prozenträge ① Schulnoten ① Stanine-Werte ① Rangskalierte Daten ① <u>Intervallskalierte Daten</u>	Die herkömmliche Faktorenanalyse kann auf folgende Daten standardmäßig angewendet werden, ohne artifizielle Ergebnisse zu liefern: ✳ Dichotome Daten ✳ <u>Intervallskalierte Daten</u> ✳ <u>T-Werte</u> ✳ <u>Centil-Werte</u> ✳ Stanine-Werte	Die herkömmliche Faktorenanalyse kann auf folgende Daten standardmäßig angewendet werden, ohne artifizielle Ergebnisse zu liefern: ① Dichotome Daten ① Prozenträge ① Stanine-Werte ① <u>Intervallskalierte Daten</u>	Die herkömmliche Faktorenanalyse kann auf folgende Daten standardmäßig angewendet werden, ohne artifizielle Ergebnisse zu liefern: ① Dichotome Daten ① Prozenträge ① Stanine-Werte ① <u>Intervallskalierte Daten</u>

31	Laut <i>Probabilistischer Testtheorie</i> sind teilweise richtig beantwortete Aufgaben ① immer mit ½ Punkt zu verrechnen ① immer mit 1 Punkt zu verrechnen, wenn eine falsche Antwort 0, eine richtige 2 Punkte zählt ① je nach Anzahl der vorgegebenen Antwortmöglichkeiten mit 1, 2 usw. Punkten zu verrechnen ① <u>empirisch auf die faire Relation zu falschen und richtigen Antworten zu untersuchen</u> ① immer als falsch zu verrechnen ① jedes zweite Mal als richtig, sonst als falsch zu verrechnen	Laut <i>Probabilistischer Testtheorie</i> sind teilweise richtig beantwortete Aufgaben ✱ immer mit ½ Punkt zu verrechnen ✱ immer mit 1 Punkt zu verrechnen, wenn eine falsche Antwort 0, eine richtige 2 Punkte zählt ✱ je nach Anzahl der vorgegebenen Antwortmöglichkeiten mit 1, 2 usw. Punkten zu verrechnen ✱ immer als falsch zu verrechnen ✱ jedes zweite Mal als richtig, sonst als falsch zu verrechnen	Laut <i>Probabilistischer Testtheorie</i> sind teilweise richtig beantwortete Aufgaben ① immer mit ½ Punkt zu verrechnen ① immer mit 1 Punkt zu verrechnen, wenn eine falsche Antwort 0, eine richtige 2 Punkte zählt ① je nach Anzahl der vorgegebenen Antwortmöglichkeiten mit 1, 2 usw. Punkten zu verrechnen ① <u>empirisch auf die faire Relation zu falschen und richtigen Antworten zu untersuchen</u> ① immer als falsch zu verrechnen ① jedes zweite Mal als richtig, sonst als falsch zu verrechnen	Laut <i>Probabilistischer Testtheorie</i> sind teilweise richtig beantwortete Aufgaben ✱ immer mit ½ Punkt zu verrechnen ✱ immer mit 1 Punkt zu verrechnen, wenn eine falsche Antwort 0, eine richtige 2 Punkte zählt ✱ je nach Anzahl der vorgegebenen Antwortmöglichkeiten mit 1, 2 usw. Punkten zu verrechnen ✱ immer als falsch zu verrechnen ✱ jedes zweite Mal als richtig, sonst als falsch zu verrechnen	Laut <i>Probabilistischer Testtheorie</i> sind teilweise richtig beantwortete Aufgaben ① immer mit ½ Punkt zu verrechnen ① immer mit 1 Punkt zu verrechnen, wenn eine falsche Antwort 0, eine richtige 2 Punkte zählt ① <u>empirisch auf die faire Relation zu falschen und richtigen Antworten zu untersuchen</u> ① immer als falsch zu verrechnen
38	Die Schwierigkeit eines Items bestimmt sich aus ✱ der absoluten Lösungshäufigkeit ✱ <u>der relativen Lösungshäufigkeit</u> ✱ der Ratewahrscheinlichkeit ✱ der Länge der Angabe ✱ <u>dem Itemparameter</u>	Die Schwierigkeit eines Items bestimmt sich aus ① der absoluten Lösungshäufigkeit ① der Ratewahrscheinlichkeit ① der Länge der Angabe ① <u>dem Itemparameter</u> ① der Anzahl vorausgehender Items ① dem Personenparameter	Die Schwierigkeit eines Items bestimmt sich aus ✱ der absoluten Lösungshäufigkeit ✱ <u>der relativen Lösungshäufigkeit</u> ✱ der Ratewahrscheinlichkeit ✱ der Länge der Angabe ✱ <u>dem Itemparameter</u>	Die Schwierigkeit eines Items bestimmt sich aus ① der absoluten Lösungshäufigkeit ① der Ratewahrscheinlichkeit ① der Länge der Angabe ① <u>dem Itemparameter</u>	Die Schwierigkeit eines Items bestimmt sich aus ① der absoluten Lösungshäufigkeit ① der Ratewahrscheinlichkeit ① der Länge der Angabe ① <u>dem Itemparameter</u>
45	Warum sollte der Eigenwert eines extrahierten Faktors innerhalb der Faktorenanalyse mindestens den Wert 1 haben?				

52	Der höchste praktisch erreichbare IQ beträgt	Der höchste praktisch erreichbare IQ beträgt ☛ 120 ☛ 125 ☛ 130 ☛ > 145 ☛ ≤ 200	Der höchste praktisch erreichbare IQ beträgt	Der höchste praktisch erreichbare IQ beträgt ☛ 120 ☛ 125 ☛ 130 ☛ > 145 ☛ ≤ 200 ☛	Der höchste praktisch erreichbare IQ beträgt ① 120 ① 125 ① 130 ① <u>145</u>
59	Um einen Unterschied von 1 IQ Punkt als signifikant interpretieren zu können, müsste die Reliabilität eines Tests einen Wert erreichen von				
66	Wie hoch muss die Reliabilität eines psychologisch-diagnostischen Verfahrens mindestens sein, um einem Testwert, der exakt mit dem Mittelwert der Referenzpopulation übereinstimmt, bei 95 % Sicherheit einen Prozentrang zwischen 25 % und 75 % zuschreiben zu können?				
73	Eine repräsentative Normierungsstichprobe ist mit hoher Wahrscheinlichkeit gewährleistet, wenn ① die Anzahl der Testpersonen 2000 nicht unterschreitet ① sich die Testwerte normalverteilen ① das <i>Rasch</i> -Modell gilt ① <u>die Stichprobe zufällig der interessierenden Population entnommen wurde</u> ① die Stichprobe mehr als 5 % der interessierenden Population ausmacht ① das Geschlecht, das Alter und die Bildung in der Stichprobe analog zur Population verteilt ist	Eine repräsentative Normierungsstichprobe ist mit hoher Wahrscheinlichkeit gewährleistet, wenn ☛ die Anzahl der Testpersonen 2000 nicht unterschreitet ☛ sich die Testwerte normalverteilen ☛ das <i>Rasch</i> -Modell gilt ☛ <u>die Stichprobe zufällig der interessierenden Population entnommen wurde</u> ☛ die Stichprobe mehr als 5 % der interessierenden Population ausmacht	Eine repräsentative Normierungsstichprobe ist mit hoher Wahrscheinlichkeit gewährleistet, wenn ① die Anzahl der Testpersonen 2000 nicht unterschreitet ① sich die Testwerte normalverteilen ① das <i>Rasch</i> -Modell gilt ① <u>die Stichprobe zufällig der interessierenden Population entnommen wurde</u> ① die Stichprobe mehr als 5 % der interessierenden Population ausmacht ① das Geschlecht, das Alter und die Bildung in der Stichprobe analog zur Population verteilt ist	Eine repräsentative Normierungsstichprobe ist mit hoher Wahrscheinlichkeit gewährleistet, wenn ☛ die Anzahl der Testpersonen 2000 nicht unterschreitet ☛ sich die Testwerte normalverteilen ☛ das <i>Rasch</i> -Modell gilt ☛ <u>die Stichprobe zufällig der interessierenden Population entnommen wurde</u> ☛ die Stichprobe mehr als 5 % der interessierenden Population ausmacht	Eine repräsentative Normierungsstichprobe ist mit hoher Wahrscheinlichkeit gewährleistet, wenn ① das <i>Rasch</i> -Modell gilt ① <u>die Stichprobe zufällig der interessierenden Population entnommen wurde</u> ① die Stichprobe mehr als 5 % der interessierenden Population ausmacht ① das Geschlecht, das Alter und die Bildung in der Stichprobe analog zur Population verteilt ist

80	Flächentransformation ist ✱ <u>eine Transformation nicht normalverteiler (ungeeichter) Testwerte in normalverteilte (geeichte)</u> ✱ eine Methode zur Berechnung der Prozenträge ✱ eine Transformation der Anzahl gelöster Aufgaben in T-Werte ✱ die Projektion einer bivariaten Normalverteilung auf eine Schnittebene ✱ jede Umwandlung von z-Werten in T-Werte, Wertpunkte etc.	Flächentransformation ist ✱ <u>eine Transformation nicht normalverteiler (ungeeichter) Testwerte in normalverteilte (geeichte)</u> ✱ eine Methode zur Berechnung der Prozenträge ✱ eine Transformation der Anzahl gelöster Aufgaben in T-Werte ✱ die Projektion einer bivariaten Normalverteilung auf eine Schnittebene ✱ jede Umwandlung von z-Werten in T-Werte, Wertpunkte etc.	Flächentransformation ist ① <u>eine Transformation nicht normalverteiler (ungeeichter) Testwerte in normalverteilte (geeichte)</u> ① eine Transformation der Anzahl gelöster Aufgaben in T-Werte ① die Projektion einer bivariaten Normalverteilung auf eine Schnittebene ① jede Umwandlung von z-Werten in T-Werte, Wertpunkte etc.	Flächentransformation ist ① <u>eine Transformation nicht normalverteiler (ungeeichter) Testwerte in normalverteilte (geeichte)</u> ① eine Transformation der Anzahl gelöster Aufgaben in T-Werte ① die Projektion einer bivariaten Normalverteilung auf eine Schnittebene ① jede Umwandlung von z-Werten in T-Werte, Wertpunkte etc.	Flächentransformation ist ① <u>eine Transformation nicht normalverteiler (ungeeichter) Testwerte in normalverteilte (geeichte)</u> ① eine Transformation der Anzahl gelöster Aufgaben in T-Werte ① die Projektion einer bivariaten Normalverteilung auf eine Schnittebene ① jede Umwandlung von z-Werten in T-Werte, Wertpunkte etc.
87	Wie hängen Centil- und Stanine-Werte zusammen? ① Gar nicht ① <u>Die Centil-Werte –1 bis +1 bzw. 9 bis 11 werden zusammengefasst zu Stanine-Werten 1 bzw. 9</u> ① $S = C + 2$ ① Die z-Werte heißen bei Leistungstests Centil-, bei Persönlichkeitsfragebogen Stanine-Werte ① Die Standardabweichung der Centil-Werte wird bei den Stanine-Werten auf 1 gesetzt ① $S = C + 2,5$	Wie hängen Centil- und Stanine-Werte zusammen? ✱ Gar nicht ✱ <u>Die Centil-Werte –1 bis +1 bzw. 9 bis 11 werden zusammengefasst zu Stanine-Werten 1 bzw. 9</u> ✱ $S = C + 2$ ✱ Die z-Werte heißen bei Leistungstests Centil-, bei Persönlichkeitsfragebogen Stanine-Werte ✱ Die Standardabweichung der Centil-Werte wird bei den Stanine-Werten auf 1 gesetzt	Wie hängen Centil- und Stanine-Werte zusammen? ① Gar nicht ① <u>Die Centil-Werte –1 bis +1 bzw. 9 bis 11 werden zusammengefasst zu Stanine-Werten 1 bzw. 9</u> ① Die z-Werte heißen bei Leistungstests Centil-, bei Persönlichkeitsfragebogen Stanine-Werte ① Die Standardabweichung der Centil-Werte wird bei den Stanine-Werten auf 1 gesetzt	Wie hängen Centil- und Stanine-Werte zusammen? ① Gar nicht ① <u>Die Centil-Werte –1 bis +1 bzw. 9 bis 11 werden zusammengefasst zu Stanine-Werten 1 bzw. 9</u> ① Die z-Werte heißen bei Leistungstests Centil-, bei Persönlichkeitsfragebogen Stanine-Werte ① Die Standardabweichung der Centil-Werte wird bei den Stanine-Werten auf 1 gesetzt	Wie hängen Centil- und Stanine-Werte zusammen? ① Gar nicht ① <u>Die Centil-Werte –1 bis +1 bzw. 9 bis 11 werden zusammengefasst zu Stanine-Werten 1 bzw. 9</u> ① Die z-Werte heißen bei Leistungstests Centil-, bei Persönlichkeitsfragebogen Stanine-Werte ① Die Standardabweichung der Centil-Werte wird bei den Stanine-Werten auf 1 gesetzt
94	Ein in einem Leistungstest erreichter Standardwert (Z-Wert) von 90 entspricht einem	Ein in einem Leistungstest erreichter Standardwert (Z-Wert) von 90 entspricht einem	Ein in einem Leistungstest erreichter Standardwert (Z-Wert) von 90 entspricht einem	Ein in einem Leistungstest erreichter Standardwert (Z-Wert) von 90 entspricht einem	Ein in einem Leistungstest erreichter Standardwert (Z-Wert) von 90 entspricht einem

	① <u>T-Wert von 40</u> ① IQ von 95 ① Prozentrang von 45 ① Stanine-Wert von 7 ① Centil-Wert von 2 ① z-Wert von 1	✱ <u>T-Wert von 40</u> ✱ <u>IQ von 85</u> ✱ Prozentrang von 45 ✱ Centil-Wert von 2 ✱ <u>z-Wert von -1</u>	① <u>T-Wert von 40</u> ① IQ von 95 ① Prozentrang von 45 ① Centil-Wert von 2	① <u>T-Wert von 40</u> ① IQ von 95 ① Prozentrang von 45 ① Centil-Wert von 2	① <u>T-Wert von 40</u> ① IQ von 95 ① Prozentrang von 45 ① Centil-Wert von 2
101	Worauf ist bei faktorenanalytisch konstruierten Tests zu achten? ✱ <u>Ob die Rotation rechtwinklig oder schiefwinklig erfolgte</u> ✱ <u>Ob die Testwerte pro Faktor nach den Faktor-Scores verrechnet werden</u> ✱ <u>Auf die Größe der Eigenwerte der Faktorenlösung</u> ✱ <u>Auf das Antwortformat der Items</u> ✱ Auf die Varianz der Variablen	Worauf ist bei faktorenanalytisch konstruierten Tests zu achten? ① Alle Eigenwerte sind kleiner als 1 ① Es bestehen keine linearen Zusammenhänge zwischen den Variablen ① Alle Items weisen mittlere Ladungen in den jeweiligen Faktoren auf ① <u>Die Testwerte pro Faktor werden nach den Faktor-Scores verrechnet</u> ① Die Anzahl der extrahierten Faktoren beträgt nicht mehr als maximal 10 ① Auf die Varianz der Variablen	Worauf ist bei faktorenanalytisch konstruierten Tests zu achten? ✱ <u>Ob die Rotation rechtwinklig oder schiefwinklig erfolgte</u> ✱ <u>Ob die Testwerte pro Faktor nach den Faktor-Scores verrechnet werden</u> ✱ <u>Auf die Größe der Eigenwerte der Faktorenlösung</u> ✱ <u>Auf das Antwortformat der Items</u> ✱ Auf die Varianz der Variablen	Worauf ist bei faktorenanalytisch konstruierten Tests zu achten? ① Alle Eigenwerte sind kleiner als 1 ① Es bestehen keine linearen Zusammenhänge zwischen den Variablen ① Alle Items weisen mittlere Ladungen in den jeweiligen Faktoren auf ① <u>Die Testwerte pro Faktor werden nach den Faktor-Scores verrechnet</u> ① Die Anzahl der extrahierten Faktoren beträgt nicht mehr als maximal 10 ① Auf die Varianz der Variablen	Worauf ist bei faktorenanalytisch konstruierten Tests zu achten? ① Alle Eigenwerte sind kleiner als 1 ① Alle Items weisen mittlere Ladungen in den jeweiligen Faktoren auf ① <u>Die Testwerte pro Faktor werden nach den Faktor-Scores verrechnet</u> ① Auf die Varianz der Variablen

Skala Verfahrensinventar-Kenntnis

Inhaltl. Item	Testform A	Testform B	Testform C	Testform D	Testform E
4	Zur Messung von Leistungsmotivation gibt es u.a. folgende psychologisch-diagnostische Verfahren: ✳ INKA ✳ FKK ✳ <u>LMI</u> ✳ <u>MMG</u> ✳ AMT	Zur Messung von Leistungsmotivation gibt es u.a. folgende psychologisch-diagnostische Verfahren: ① INKA ① FKK ① <u>LMI</u> ① BBT ① AMT ① AVT	Zur Messung von Leistungsmotivation gibt es u.a. folgende psychologisch-diagnostische Verfahren: ✳ INKA ✳ FKK ✳ <u>LMI</u> ✳ <u>MMG</u> ✳ AMT	Zur Messung von Leistungsmotivation gibt es u.a. folgende psychologisch-diagnostische Verfahren: ① INKA ① FKK ① <u>LMI</u> ① BBT ① AMT ① AVT	Zur Messung von Leistungsmotivation gibt es u.a. folgende psychologisch-diagnostische Verfahren: ① INKA ① FKK ① <u>LMI</u> ① AMT
11	Für die Messung welcher Eigenschaft ist der ZVT ein typischer Repräsentant?				
18	In welchem Verfahren kann die „Passung“ zwischen Interessen und den Anforderungen an die angestrebte Berufsausbildung geprüft werden?	In welchem Verfahren kann die „Passung“ zwischen Interessen und den Anforderungen an die angestrebte Berufsausbildung geprüft werden? ✳ GIS ✳ PRF ✳ CFT 20 ✳ BIT II ✳ IST	In welchem Verfahren kann die „Passung“ zwischen Interessen und den Anforderungen an die angestrebte Berufsausbildung geprüft werden?	In welchem Verfahren kann die „Passung“ zwischen Interessen und den Anforderungen an die angestrebte Berufsausbildung geprüft werden? ✳ GIS ✳ PRF ✳ CFT 20 ✳ BIT II ✳ IST	In welchem Verfahren kann die „Passung“ zwischen Interessen und den Anforderungen an die angestrebte Berufsausbildung geprüft werden? ① GIS ① <u>AIST</u> ① CFT 20 ① BIT II
25	Nenne einen Lerntest				

32	Zu psychologisch-diagnostischen Verfahren, die Angst erfassen, zählen: ☛ DSI ☛ ACIL ☛ FAF ☛ ADEM ☛ STAXI	Zu psychologisch-diagnostischen Verfahren, die Angst erfassen, zählen: ① DSI ① ACIL ① FAF ① ADEM ① STAXI ① <u>DAI</u>	Zu psychologisch-diagnostischen Verfahren, die Angst erfassen, zählen: ① ACIL ① ADEM ① STAXI ① <u>DAI</u>	Zu psychologisch-diagnostischen Verfahren, die Angst erfassen, zählen: ① ACIL ① ADEM ① STAXI ① <u>DAI</u>	Zu psychologisch-diagnostischen Verfahren, die Angst erfassen, zählen: ① ACIL ① ADEM ① STAXI ① <u>DAI</u>
39	Das Nachfolgeverfahren des NEO-FFI ist: 	Das Nachfolgeverfahren des NEO-FFI ist: ☛ NEO-R ☛ <u>NEO-PI-R</u> ☛ NEOGV ☛ NEO-VG ☛ B5PO	Das Nachfolgeverfahren des NEO-FFI ist: 	Das Nachfolgeverfahren des NEO-FFI ist: ☛ NEO-R ☛ <u>NEO-PI-R</u> ☛ NEOGV ☛ NEO-VG ☛ B5PO	Das Nachfolgeverfahren des NEO-FFI ist: ① NEO-R ① <u>NEO-PI-R</u> ① NEOGV ① B5PO
46	Von welchen der folgenden berühmten Psychologen stammen psychologisch-diagnostische Verfahren, auf denen aufbauend noch immer aktuelle Versionen verlagsmäßig (im deutschen Sprachraum) vertrieben werden ☛ <u>Raymond B. Cattell</u> ☛ James McKeen Cattell ☛ <u>Hans J. Eysenck</u> ☛ <u>Adolf O. Jäger</u> ☛ Karl Bühler	Von welchen der folgenden berühmten Psychologen stammen psychologisch-diagnostische Verfahren, auf denen aufbauend noch immer aktuelle Versionen verlagsmäßig (im deutschen Sprachraum) vertrieben werden ① Hubert Rohrer ① James McKeen Cattell ① Lee J. Cronbach ① <u>Adolf O. Jäger</u> ① Karl Bühler ① Manfred Amelang	Von welchen der folgenden berühmten Psychologen stammen psychologisch-diagnostische Verfahren, auf denen aufbauend noch immer aktuelle Versionen verlagsmäßig (im deutschen Sprachraum) vertrieben werden ☛ <u>Raymond B. Cattell</u> ☛ James McKeen Cattell ☛ <u>Hans J. Eysenck</u> ☛ <u>Adolf O. Jäger</u> ☛ Karl Bühler	Von welchen der folgenden berühmten Psychologen stammen psychologisch-diagnostische Verfahren, auf denen aufbauend noch immer aktuelle Versionen verlagsmäßig (im deutschen Sprachraum) vertrieben werden ① James McKeen Cattell ① <u>Adolf O. Jäger</u> ① Karl Bühler ① Manfred Amelang	Von welchen der folgenden berühmten Psychologen stammen psychologisch-diagnostische Verfahren, auf denen aufbauend noch immer aktuelle Versionen verlagsmäßig (im deutschen Sprachraum) vertrieben werden ① James McKeen Cattell ① <u>Adolf O. Jäger</u> ① Karl Bühler ① Manfred Amelang
53	Nenne ein psychologisch-diagnostisches Verfahren, dass Extraversion erfasst:				

60	Der TAT wurde nach der motivationspsychologisch begründeten Persönlichkeitstheorie von <i>Murray</i> konstruiert; welcher Persönlichkeitsfragebogen basiert ebenfalls auf diesem Konzept?	Der TAT wurde nach der motivationspsychologisch begründeten Persönlichkeitstheorie von <i>Murray</i> konstruiert; welcher Persönlichkeitsfragebogen basiert ebenfalls auf diesem Konzept? ① Sceno-Test ① LMI ① TIPI ① OLMT ① MMG ① <u>PRF</u>	Der TAT wurde nach der motivationspsychologisch begründeten Persönlichkeitstheorie von <i>Murray</i> konstruiert; welcher Persönlichkeitsfragebogen basiert ebenfalls auf diesem Konzept?	Der TAT wurde nach der motivationspsychologisch begründeten Persönlichkeitstheorie von <i>Murray</i> konstruiert; welcher Persönlichkeitsfragebogen basiert ebenfalls auf diesem Konzept? ① Sceno-Test ① LMI ① TIPI ① OLMT ① MMG ① <u>PRF</u>	Der TAT wurde nach der motivationspsychologisch begründeten Persönlichkeitstheorie von <i>Murray</i> konstruiert; welcher Persönlichkeitsfragebogen basiert ebenfalls auf diesem Konzept? ① LMI ① TIPI ① OLMT ① <u>PRF</u>
67	67. Nenne einen, abgesehen vom AID 2 verlagsmäßig vertriebenen Test, dessen Vorgabe adaptiv erfolgt:				
74	-	-	-	-	-
81	Welche Fähigkeit soll durch den Test <i>Schlauchfiguren</i> gemessen werden?	Welche Fähigkeit soll durch den Test <i>Schlauchfiguren</i> gemessen werden? ✳ <u>Space</u> ✳ Kreativität ✳ <i>Perceptual Speed</i> ✳ <u>Raumvorstellung</u> ✳ <i>Reasoning</i>	Welche Fähigkeit soll durch den Test <i>Schlauchfiguren</i> gemessen werden?	Welche Fähigkeit soll durch den Test <i>Schlauchfiguren</i> gemessen werden? ✳ <u>Space</u> ✳ Kreativität ✳ <i>Perceptual Speed</i> ✳ <u>Raumvorstellung</u> ✳ <i>Reasoning</i>	Welche Fähigkeit soll durch den Test <i>Schlauchfiguren</i> gemessen werden? ① Kreativität ① <i>Perceptual Speed</i> ① <u>Raumvorstellung</u> ① Reasoning
88	Das Verfahren EPP-D beruht auf der Persönlichkeitstheorie von				
95	Zu den <i>Culture-Fair</i> Tests zählt: ✳ <u>SPM</u> ✳ WST ✳ Syllogismen ✳ FKK ✳ FSKN	Zu den <i>Culture-Fair</i> Tests zählt: ✳ <u>SPM</u> ✳ WST ✳ Syllogismen ✳ FKK ✳ FSKN	Zu den <i>Culture-Fair</i> Tests zählt: ① <u>SPM</u> ① WST ① FKK ① FSKN	Zu den <i>Culture-Fair</i> Tests zählt: ① <u>SPM</u> ① WST ① FKK ① FSKN	Zu den <i>Culture-Fair</i> Tests zählt: ① <u>SPM</u> ① WST ① FKK ① FSKN
102	102. Bei welchem Persönlichkeitsfragebogen zum Beispiel ist sowohl Selbst- auch Fremdeinschätzung explizit vorgesehen?				

Skala Theoriewissen zur Differentiellen Psychologie

Inhaltl. Item	Testform A	Testform B	Testform C	Testform D	Testform E
5	Das Gegenteil von selektiver (fokussierender) Aufmerksamkeit ist				
12	Das Gegenteil von „Hoffnung auf Erfolg“ ist	Das Gegenteil von „Hoffnung auf Erfolg“ ist ① „Angst vor Zurückweisung“ ① „Hoffnung auf Verstärkung“ ① „Furcht vor Misserfolg“ ① „Angst vor Erfolg“ ① „Wissen um Erfolg“ ① „Hoffnung auf Misserfolg“	Das Gegenteil von „Hoffnung auf Erfolg“ ist	Das Gegenteil von „Hoffnung auf Erfolg“ ist ① „Angst vor Zurückweisung“ ① „Hoffnung auf Verstärkung“ ① „Furcht vor Misserfolg“ ① „Angst vor Erfolg“ ① „Wissen um Erfolg“ ① „Hoffnung auf Misserfolg“	Das Gegenteil von „Hoffnung auf Erfolg“ ist ① „Hoffnung auf Verstärkung“ ① „Furcht vor Misserfolg“ ① „Angst vor Erfolg“ ① „Hoffnung auf Misserfolg“
19	Das Konstrukt Kontrollüberzeugung (<i>locus of control of reinforcement</i>) betrifft die Einstellung einer Person darüber, ① <u>wer oder was ihr Leben kontrolliert</u> ① wen oder was sie selbst kontrolliert ① wem sie vertrauen kann ① dass sie selbst viel Macht hat ① dass Kontrolle wichtiger als Vertrauen ist ① dass Befriedigungsaufschub manchmal notwendig ist	Das Konstrukt Kontrollüberzeugung (<i>locus of control of reinforcement</i>) betrifft die Einstellung einer Person darüber,	Das Konstrukt Kontrollüberzeugung (<i>locus of control of reinforcement</i>) betrifft die Einstellung einer Person darüber, ① <u>wer oder was ihr Leben kontrolliert</u> ① wen oder was sie selbst kontrolliert ① wem sie vertrauen kann ① dass sie selbst viel Macht hat ① dass Kontrolle wichtiger als Vertrauen ist ① dass Befriedigungsaufschub manchmal notwendig ist	Das Konstrukt Kontrollüberzeugung (<i>locus of control of reinforcement</i>) betrifft die Einstellung einer Person darüber, ① <u>wer oder was ihr Leben kontrolliert</u> ① wen oder was sie selbst kontrolliert ① wem sie vertrauen kann ① dass Befriedigungsaufschub manchmal notwendig ist	Das Konstrukt Kontrollüberzeugung (<i>locus of control of reinforcement</i>) betrifft die Einstellung einer Person darüber, ① <u>wer oder was ihr Leben kontrolliert</u> ① wen oder was sie selbst kontrolliert ① wem sie vertrauen kann ① dass Befriedigungsaufschub manchmal notwendig ist
26	Motivation verhält sich zu Motiv so wie Angst zu				
33	Was versteht man unter <i>Reasoning</i> ?				

40	Die theoretische Konzeption des STAI unterscheidet zwischen Angst als <i>State</i> und Angst als	Die theoretische Konzeption des STAI unterscheidet zwischen Angst als <i>State</i> und Angst als ✱ <i>Power</i> ✱ Furcht ✱ Abwehrmechanismus ✱ <i>Coping</i> ✱ Ressource	Die theoretische Konzeption des STAI unterscheidet zwischen Angst als <i>State</i> und Angst als	Die theoretische Konzeption des STAI unterscheidet zwischen Angst als <i>State</i> und Angst als ✱ <i>Power</i> ✱ Furcht ✱ Abwehrmechanismus ✱ <i>Coping</i> ✱ Ressource	Die theoretische Konzeption des STAI unterscheidet zwischen Angst als <i>State</i> und Angst als ① <u>Trait</u> ① Abwehrmechanismus ① Coping ① Ressource
47	Das konkurrierende Motiv zum Aggressionsmotiv heißt:	Das konkurrierende Motiv zum Aggressionsmotiv heißt: ① <u>Aggressionshemmungsmotiv</u> ① Angstmotiv ① Ärgermotiv ① Aggressions-Frustrations-Motiv ① Ärgerhemmungsmotiv ① Angsthemmungsmotiv	Das konkurrierende Motiv zum Aggressionsmotiv heißt:	Das konkurrierende Motiv zum Aggressionsmotiv heißt: ① <u>Aggressionshemmungsmotiv</u> ① Angstmotiv ① Ärgermotiv ① Aggressions-Frustrations-Motiv ① Ärgerhemmungsmotiv ① Angsthemmungsmotiv	Das konkurrierende Motiv zum Aggressionsmotiv heißt: ① <u>Aggressionshemmungsmotiv</u> ① Angstmotiv ① Ärgermotiv ① Angsthemmungsmotiv
54	Neben Extraversion, Neurotizismus, Gewissenhaftigkeit und Verträglichkeit beinhaltet das <i>Big-Five</i> - Persönlichkeits-modell den Faktor				
61	Die Materialien im <i>Berliner-Intelligenz-Strukturmodell</i> sind verbal (lexikalisch), figural und				
68	Das Gegenteil von Reflexivität ist.....				
75	Das Gegenteil von Repressor ist				
82	Das Gegenteil von <i>Fluid Intelligence</i> ist				
89	Belastbarkeit einer Person, als Konstrukt verstanden, ist.....				
96	Das stabilste psychische/psychologische Merkmal eines Menschen ist allgemein:				
103	Unter Gegenübertragung versteht man das Phänomen der Übertragung bei				

Skala Verfahrensbezogenes Detailwissen

Inhaltl. Item	Testform A	Testform B	Testform C	Testform D	Testform E
6	FAIR und <i>Test d2</i> unterscheiden sich darin, dass ✱ <u>FAIR statt „d“ und „p“ „Kreis“ und „Quadrat“ verwendet</u> ✱ Beim <i>Test d2</i> ein instruktionswidriges Bearbeiten verhindert wird, nicht beim FAIR ✱ Beim FAIR ein <u>instruktionswidriges Bearbeiten verhindert wird, nicht beim <i>Test d2</i></u> ✱ Beim FAIR ausschließlich <u>symmetrische Symbole verwendet werden</u> ✱ Der FAIR ein Computerverfahren ist, der <i>Test d2</i> nicht	FAIR und <i>Test d2</i> unterscheiden sich darin, dass ① Beim <i>Test d2</i> ein instruktionswidriges Bearbeiten verhindert wird, nicht beim FAIR ① Beim FAIR ein <u>instruktionswidriges Bearbeiten verhindert wird, nicht beim <i>Test d2</i></u> ① Der FAIR eigene Eich Tabellen für Österreich beinhaltet, der <i>Test d2</i> nicht ① Der <i>Test d2</i> eine höchstens halb so lange Durchführungszeit beansprucht wie der FAIR ① Der FAIR Konzentration, der <i>Test d2</i> Aufmerksamkeit prüft ① Der FAIR geschlechtsspezifische Eich Tabellen enthält, der <i>Test d2</i> nicht	FAIR und <i>Test d2</i> unterscheiden sich darin, dass ✱ <u>FAIR statt „d“ und „p“ „Kreis“ und „Quadrat“ verwendet</u> ✱ Beim <i>Test d2</i> ein instruktionswidriges Bearbeiten verhindert wird, nicht beim FAIR ✱ Beim FAIR ein <u>instruktionswidriges Bearbeiten verhindert wird, nicht beim <i>Test d2</i></u> ✱ Beim FAIR ausschließlich <u>symmetrische Symbole verwendet werden</u> ✱ Der FAIR ein Computerverfahren ist, der <i>Test d2</i> nicht	FAIR und <i>Test d2</i> unterscheiden sich darin, dass ① Beim <i>Test d2</i> ein instruktionswidriges Bearbeiten verhindert wird, nicht beim FAIR ① Beim FAIR ein <u>instruktionswidriges Bearbeiten verhindert wird, nicht beim <i>Test d2</i></u> ① Der FAIR eigene Eich Tabellen für Österreich beinhaltet, der <i>Test d2</i> nicht ① Der <i>Test d2</i> eine höchstens halb so lange Durchführungszeit beansprucht wie der FAIR ① Der FAIR Konzentration, der <i>Test d2</i> Aufmerksamkeit prüft ① Der FAIR geschlechtsspezifische Eich Tabellen enthält, der <i>Test d2</i> nicht	FAIR und <i>Test d2</i> unterscheiden sich darin, dass ① FAIR statt „d“ und „p“ „Dreieck“ und „Quadrat“ verwendet ① Beim <i>Test d2</i> ein instruktionswidriges Bearbeiten verhindert wird, nicht beim FAIR ① Beim FAIR ausschließlich <u>symmetrische Symbole verwendet werden</u> ① Der FAIR ein Computerverfahren ist, der <i>Test d2</i> nicht
13	Welche von den angeführten Bereichen erfasst das DAI: ✱ <u>Angstauslösung</u> ✱ Angstvermeidungsmanifestation ✱ <u>Angstmanifestation</u> ✱ <u>Angstcopingstrategien</u> ✱ <u>Angststabilisierung</u>	Welche Bereiche außer Angstauslösung, Angstmanifestation und Angstcopingstrategien erfasst das DAI?	Welche von den angeführten Bereichen erfasst das DAI: ✱ <u>Angstauslösung</u> ✱ Angstvermeidungsmanifestation ✱ <u>Angstmanifestation</u> ✱ <u>Angstcopingstrategien</u> ✱ <u>Angststabilisierung</u>	Welche von den angeführten Bereichen erfasst das DAI: ① Angstverdrängung ① <u>Angstauslösung</u> ① Angstvermeidungsmanifestation ① Zustandsangst	Welche von den angeführten Bereichen erfasst das DAI: ① Angstverdrängung ① <u>Angstauslösung</u> ① Angstvermeidungsmanifestation ① Zustandsangst
20	Die Abkürzung IPC bedeutet:				

27	Der überlappende Altersbereich von K-ABC und AID 2 liegt zwischen: ① 2;6 und 7;6 Jahren ① 4;6 und 12;6 Jahren ① <u>6;0 und 12;5 Jahren</u> ① 6;0 und 15;11 Jahren ① 6;0 und 8;6 Jahren ① 7;6 und 12;11 Jahren	Der überlappende Altersbereich von K-ABC und AID 2 liegt zwischen: ① 2;6 und 7;6 Jahren ① 4;6 und 12;6 Jahren ① <u>6;0 und 12;5 Jahren</u> ① 6;0 und 15;11 Jahren ① 6;0 und 8;6 Jahren ① 7;6 und 12;11 Jahren	Der überlappende Altersbereich von K-ABC und AID 2 liegt zwischen: ① 2;6 und 7;6 Jahren ① 4;6 und 12;6 Jahren ① <u>6;0 und 12;5 Jahren</u> ① 6;0 und 15;11 Jahren ① 6;0 und 8;6 Jahren ① 7;6 und 12;11 Jahren	Der überlappende Altersbereich von K-ABC und AID 2 liegt zwischen: ① 4;6 und 12;6 Jahren ① <u>6;0 und 12;5 Jahren</u> ① 6;0 und 8;6 Jahren ① 7;6 und 12;11 Jahren	Der überlappende Altersbereich von K-ABC und AID 2 liegt zwischen: ① 4;6 und 12;6 Jahren ① <u>6;0 und 12;5 Jahren</u> ① 6;0 und 8;6 Jahren ① 7;6 und 12;11 Jahren
34	Neben der qualitativen Interpretation der „Arbeitskurve“ erlaubt im <i>Test d2</i> folgender Testkennwert die Interpretation des Verlaufs der Leistungen über mehrere Teilzeiten:				
41	Neben Intelligenzquantität und einer qualitativen Beurteilung der „Arbeitshaltungen“ können beim AID 2 noch zusätzliche Testkennwerte bestimmt werden. Zum Beispiel:				
48	Wie viele unabhängige Dimensionen liegen laut Faktorenanalyse der deutschen Fassung des 16 PF-R zugrunde ?				
55	Neben dem <i>Zahlenfolgentest</i> beinhaltet der CFT 20 als Zusatztest:				
62	Der FAST erfasst außer Kohäsion der Familienmitglieder:				
69	<i>Wechsler</i> untergliedert seine Testbatterien in zwei Testteile:				
76	Die <i>Paardiagnostik mit dem Gießen-Test</i> führt als wesentliches Testergebnis zu				
83	Das Verfahren B5PO verwendet als Antwortformat				
90	Der Untertest <i>Türkisch</i> ist aus welchem Test?				
97	97. Welche Fähigkeit versuchen Tests zu messen, die das Ordnen von Bildern zu einer sinnvollen Bildgeschichte verlangen?				
104	Neben zum Beispiel <i>Burnout</i> erlaubt das Verfahren AVEM das Diagnostizieren des Personentyps				

Skala Up-To-Date

Inhaltl. Item	Testform A	Testform B	Testform C	Testform D	Testform E
7	Die DIN 33430 stammt aus dem Jahr ⓪ 1950 ⓪ 1962 ⓪ 1976 ⓪ 1994 ⓪ <u>2002</u> ⓪ 2005	Die DIN 33430 stammt aus dem Jahr ⓪ 1950 ⓪ 1962 ⓪ 1976 ⓪ 1994 ⓪ <u>2002</u> ⓪ 2005	Die DIN 33430 stammt aus dem Jahr ⓪ 1950 ⓪ 1962 ⓪ 1976 ⓪ 1994 ⓪ <u>2002</u> ⓪ 2005	Die DIN 33430 stammt aus dem Jahr ⓪ 1950 ⓪ 1976 ⓪ <u>2002</u> ⓪ 2005	Die DIN 33430 stammt aus dem Jahr ⓪ 1950 ⓪ 1976 ⓪ <u>2002</u> ⓪ 2005
14	Einschlägige Psychologie-Verlage, die psychologisch-diagnostische Verfahren vertreiben sind: ✳ Germania Psychologica ✳ <u>Dr. G. Schuhfried GmbH</u> ✳ <u>Harcourt Test Services</u> ✳ Kohlhammer ✳ <u>Hogrefe Verlag</u>	Einschlägige Psychologie-Verlage, die psychologisch-diagnostische Verfahren vertreiben sind: ⓪ Germania Psychologica ⓪ Mercury & Co Verlag ⓪ Psychologie heute ⓪ Verlag für psychologische Tests ⓪ <u>Dr. G. Schuhfried GmbH</u> ⓪ Westermann & de Gruyter	Einschlägige Psychologie-Verlage, die psychologisch-diagnostische Verfahren vertreiben sind: ✳ Germania Psychologica ✳ <u>Dr. G. Schuhfried GmbH</u> ✳ <u>Harcourt Test Services</u> ✳ Kohlhammer ✳ <u>Hogrefe Verlag</u>	Einschlägige Psychologie-Verlage, die psychologisch-diagnostische Verfahren vertreiben sind: ⓪ Germania Psychologica ⓪ Verlag für psychologische Tests ⓪ <u>Dr. G. Schuhfried GmbH</u> ⓪ Westermann & de Gruyter	Einschlägige Psychologie-Verlage, die psychologisch-diagnostische Verfahren vertreiben sind: ⓪ Germania Psychologica ⓪ Verlag für psychologische Tests ⓪ <u>Dr. G. Schuhfried GmbH</u> ⓪ Westermann & de Gruyter
21	Wann erfolgte die letzte Eichung des FPI-R?				
28	Der zuletzt publizierte Test zur Messung von „technischem Verständnis“ ist:				
35	Die aktuelle Fassung der <i>Wechsler</i> -Intelligenz-Testbatterie für Kinder heißt.....	Die aktuelle Fassung der <i>Wechsler</i> -Intelligenz-Testbatterie für Kinder heißt ⓪ HAWIK-R ⓪ HAWIK ⓪ <u>HAWIK-IV</u> ⓪ HAWIE-R ⓪ WIP ⓪ W-ITB	Die aktuelle Fassung der <i>Wechsler</i> -Intelligenz-Testbatterie für Kinder heißt.....	Die aktuelle Fassung der <i>Wechsler</i> -Intelligenz-Testbatterie für Kinder heißt ⓪ HAWIK-R ⓪ HAWIK ⓪ <u>HAWIK-IV</u> ⓪ HAWIE-R ⓪ WIP ⓪ W-ITB	Die aktuelle Fassung der <i>Wechsler</i> -Intelligenz-Testbatterie für Kinder heißt ⓪ HAWIK-R ⓪ <u>HAWIK-IV</u> ⓪ HAWIE-R ⓪ W-ITB

42	Testrezensionen findet man ① <u>im PSYINDEX Tests</u> ① in SYMLOG ① in der <i>Zeitschrift für Diagnostik und Testtheorie</i> ① im RHIA/VERA ① im LEWITE ① im Lehrbuch von Westhoff & Kluck (2003)	Testrezensionen findet man ✱ <u>im PSYINDEX Tests</u> ✱ in SYMLOG ✱ <u>in der Zeitschrift für Differentielle und Diagnostische Psychologie</u> ✱ im RHIA/VERA ✱ im Lehrbuch von Westhoff & Kluck (2003)	Testrezensionen findet man ① <u>im PSYINDEX Tests</u> ① in SYMLOG ① in der <i>Zeitschrift für Diagnostik und Testtheorie</i> ① im RHIA/VERA ① im LEWITE ① im Lehrbuch von Westhoff & Kluck (2003)	Testrezensionen findet man ✱ <u>im PSYINDEX Tests</u> ✱ in SYMLOG ✱ <u>in der Zeitschrift für Differentielle und Diagnostische Psychologie</u> ✱ im RHIA/VERA ✱ im Lehrbuch von Westhoff & Kluck (2003)	Testrezensionen findet man ① <u>im PSYINDEX Tests</u> ① in SYMLOG ① in der <i>Zeitschrift für Diagnostik und Testtheorie</i> ① im Lehrbuch von Westhoff & Kluck (2003)
49	Wann erfolgte die Publikation der Neunormierung des <i>Tests d2</i> ?				
56	Wann erfolgte die Publikation der Neunormierung des <i>Gießen-Tests</i> ?				
63	In welchem Jahr wurde der 16 PF-R in deutscher Fassung publiziert?				
70	Wann wurde der HAWIE-R veröffentlicht?				
77	Die neueste Version zum Intelligenz-Struktur-Test (IST) heißt				

84	Der <i>Implicit Association Test</i> ist der Zielsetzung nach am ehesten vergleichbar mit	Der <i>Implicit Association Test</i> ist der Zielsetzung nach am ehesten vergleichbar mit ✱ den Lerntests <i>sensu Guthke</i> ✱ den sog. „Objektiven Persönlichkeitstests“ <i>sensu R.B.Cattell</i> ✱ <u>den performance-tests-of-personality nach Cronbach</u> ✱ <u>der sog. „experimentalpsychologischen Verhaltensdiagnostik“</u> ✱ <u>den Projektiven Verfahren</u>	Der <i>Implicit Association Test</i> ist der Zielsetzung nach am ehesten vergleichbar mit	Der <i>Implicit Association Test</i> ist der Zielsetzung nach am ehesten vergleichbar mit ✱ den Lerntests <i>sensu Guthke</i> ✱ den sog. „Objektiven Persönlichkeitstests“ <i>sensu R.B.Cattell</i> ✱ <u>den performance-tests-of-personality nach Cronbach</u> ✱ <u>der sog. „experimentalpsychologischen Verhaltensdiagnostik“</u> ✱ <u>den Projektiven Verfahren</u>	Der <i>Implicit Association Test</i> ist der Zielsetzung nach am ehesten vergleichbar mit ✱ den Lerntests <i>sensu Guthke</i> ✱ den sog. „Objektiven Persönlichkeitstests“ <i>sensu R.B.Cattell</i> ✱ <u>den performance-tests-of-personality nach Cronbach</u> ✱ <u>der sog. „experimentalpsychologischen Verhaltensdiagnostik“</u> ✱ <u>den Projektiven Verfahren</u>
91	Ein neues, relativ allgemein verbindliches Signierungssystem für das <i>Rorschach</i> -Form-Deute-Verfahren ist				
98	Das Nachfolgeverfahren des MMPI heißt:				
105	Unter welcher Bezeichnung ist <i>psychologisches Diagnostizieren</i> im Englischen geläufig?				

7.4. Anhang C: Weiterführende Stichprobeneigenschaften

Anmerkung: Die folgenden Tabellen 39bis 42 stellen einige ergänzende Angaben zu den Eigenschaften der Stichprobe dar.

Stichprobendetails zur Variable Alter

Tabelle 39: Deskriptive Statistiken zur Variable Alter

N	Gültig	311
	Fehlend	1
Mittelwert		25,48
Median		25,00
Standardabweichung		3,567
Varianz		12,721
Minimum		21
Maximum		52

Tabbelle 40: Altersverteilung der Testpersonen in der Stichprobe

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	21	5	1,6	1,6	1,6
	22	28	9,0	9,0	10,6
	23	58	18,6	18,6	29,3
	24	56	17,9	18,0	47,3
	25	56	17,9	18,0	65,3
	26	35	11,2	11,3	76,5
	27	16	5,1	5,1	81,7
	28	17	5,4	5,5	87,1
	29	11	3,5	3,5	90,7
	30	5	1,6	1,6	92,3
	31	3	1,0	1,0	93,2
	32	3	1,0	1,0	94,2
	33	4	1,3	1,3	95,5
	34	6	1,9	1,9	97,4
	35	2	,6	,6	98,1
	36	2	,6	,6	98,7
	40	3	1,0	1,0	99,7
	52	1	,3	,3	100,0
	Gesamt	311	99,7	100,0	
Fehlend	999	1	,3		
Gesamt		312	100,0		

Geplantes/Gewähltes Diplomprüfungsfach:

Tabelle 41: Übersicht über die Häufigkeiten der angegebenen Diplomprüfungsfächer

Psychologisches Prüfungsfach		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	Psychologische Diagnostik	28	9,0	10,1	10,1
	Klinische- und Gesundheitspsychologie	70	22,4	25,3	35,4
	Wirtschaftspsychologie	43	13,8	15,5	50,9
	Sozialpsychologie	24	7,7	8,7	59,6
	Bildungspsychologie	16	5,1	5,8	65,3
	Forschungsmethoden und Evaluation	1	,3	,4	65,7
	Allgemeine Psychologie	12	3,8	4,3	70,0
	Biologische Psychologie	7	2,2	2,5	72,6
	Entwicklungspsychologie	54	17,3	19,5	92,1
	Methodenlehre	11	3,5	4,0	96,0
	Differentielle Psychologie	11	3,5	4,0	100,0
	Gesamt	277	88,8	100,0	
Fehlend	999	35	11,2		
Gesamt		312	100,0		

Deskriptive Statistiken zu den angegebenen Noten in den vier Lehrveranstaltungen der Grundausbildung der Psychologischen Diagnostik:

Tabelle 42: Deskriptive Statistiken zu den angegebenen Noten der vier Lehrveranstaltungen der Psychologischen Diagnostik

Lehrveranstaltung	N	Minimum	Maximum	Mittelwert	Standardabweichung	Varianz
Note VO I	303	1	4	3,12	,914	,836
Note VO II	255	1	5	3,11	,876	,767
Note UE I	303	1	4	1,49	,723	,522
Note UE II	301	1	4	1,49	,719	,517
Gültige Werte (Listenweise)	254					

8. Tabellenverzeichnis

Tabelle 1: Anzahl der Items je Antwortformatkategorie , dargestellt pro Testform A- E des DIC sowie insgesamt	14
Tabelle 2: Itemanzahl pro Skala des DIC sowie gesamt	15
Tabelle 3: Überblick über die Anzahl getesteter Studierender pro Testform des DIC ...	18
Tabelle 4: Anzahl und Prozent der Testpersonen pro Teilstichprobe in Bezug auf die Teilungskriterien <i>Alter</i> , <i>Note</i> und <i>Prüfungsfach</i>	21
Tabelle 5: Ergebnisse des LRTs der Skala <i>Grundlagenwissen</i> a-priori sowie a-posteriori in Bezug auf das Teilungskriterium <i>Rohscore</i>	25
Tabelle 6: Ergebnisse des LRTs der Skala <i>Grundlagenwissen</i> a-priori sowie a-posteriori in Bezug auf das Teilungskriterium <i>Alter</i>	26
Tabelle 7: Ergebnisse des LRTs der Skala <i>Grundlagenwissen</i> a-priori sowie a-posteriori in Bezug auf das Teilungskriterium <i>Note</i>	27
Tabelle 8: Ergebnisse des LRTs der Skala <i>Grundlagenwissen</i> a-priori sowie a-posteriori in Bezug auf das Teilungskriterium <i>Prüfungsfach</i>	28
Tabelle 9: Überblick über die Item-Ausschlüsse der Skala <i>Grundlagenwissen</i> : Die grau unterlegten Felder stellen die ausgeschlossenen Items dar, Gedankenstriche stehen für Kombinationen von Aufgabeninhalt und Antwortformat, die in der vorliegenden Untersuchung nicht zum Einsatz kamen.	29
Tabelle 10: Ergebnisse des LRTs der Skala <i>Statistik- und Testtheorie-Wissen</i> a-priori in Bezug auf das Teilungskriterium <i>Rohscore</i>	31
Tabelle 11: Ergebnisse des LRTs der Skala <i>Statistik- und Testtheorie-Wissen</i> a-priori in Bezug auf das Teilungskriterium <i>Alter</i>	32
Tabelle 12: Ergebnisse des LRTs der Skala <i>Statistik- und Testtheorie-Wissen</i> a-priori in Bezug auf das Teilungskriterium <i>Note</i>	33

Tabelle 13: Ergebnisse des LRTs der Skala <i>Statistik- und Testtheorie-Wissen</i> a-priori in Bezug auf das Teilungskriterium <i>Prüfungsfach</i>	34
Tabelle 14: Ergebnisse des LRTs der Skala <i>Verfahrensinventar-Kenntnis</i> a-priori in Bezug auf das Teilungskriterium <i>Rohscore</i>	35
Tabelle 15: Ergebnisse des LRTs der Skala <i>Verfahrensinventar-Kenntnis</i> a-priori in Bezug auf das Teilungskriterium <i>Alter</i>	36
Tabelle 16: Ergebnisse des LRTs der Skala <i>Verfahrensinventar-Kenntnis</i> a-priori in Bezug auf das Teilungskriterium <i>Note</i>	37
Tabelle 17: Ergebnisse des LRTs der Skala <i>Verfahrensinventar-Kenntnis</i> a-priori in Bezug auf das Teilungskriterium <i>Prüfungsfach</i>	38
Tabelle 18: Ergebnisse des LRTs der Skala <i>Theoriewissen zur Differentiellen</i> <i>Psychologie</i> a-priori in Bezug auf das Teilungskriterium <i>Rohscore</i>	39
Tabelle 19: Ergebnisse des LRTs der Skala <i>Theoriewissen zur Differentiellen</i> <i>Psychologie</i> a-priori in Bezug auf das Teilungskriterium <i>Alter</i>	40
Tabelle 20: Ergebnisse des LRTs der Skala <i>Theoriewissen zur Differentiellen</i> <i>Psychologie</i> a-priori in Bezug auf das Teilungskriterium <i>Note</i>	41
Tabelle 21: Ergebnisse des LRTs der Skala <i>Theoriewissen zur differentiellen</i> <i>Psychologie</i> a-priori in Bezug auf das Teilungskriterium <i>Prüfungsfach</i>	42
Tabelle 22: Ergebnisse des LRTs der Skala <i>Verfahrensbezogenes Detailwissen</i> a-priori sowie a-posteriori in Bezug auf das Teilungskriterium <i>Rohscore</i>	44
Tabelle 23: Ergebnisse des LRTs der Skala <i>Verfahrensbezogenes Detailwissen</i> a-priori sowie a-posteriori in Bezug auf das Teilungskriterium <i>Alter</i> LRT vor Ausschluss von Items (a-priori) :	44
Tabelle 24: Ergebnisse des LRTs der Skala <i>Verfahrensbezogenes Detailwissen</i> a-priori sowie a-posteriori in Bezug auf das Teilungskriterium <i>Note</i>	45

Tabelle 25: Ergebnisse des LRTs der Skala <i>Verfahrensbezogenes Detailwissen</i> a-priori sowie a-posteriori in Bezug auf das Teilungskriterium <i>Prüfungsfach</i>	46
Tabelle 26: Übersicht über die Item-Ausschlüsse der Skala <i>Verfahrensbezogenes Detailwissen</i> : die grau unterlegten Felder stellen die ausgeschlossenen Items dar, Gedankenstriche stehen für Kombinationen von Aufgabeninhalt und Antwortformat, die in der vorliegenden Untersuchung nicht zum Einsatz kamen.	47
Tabelle 27: Ergebnisse des LRTs der Skala <i>Up-To-Date</i> a-priori in Bezug auf das Teilungskriterium <i>Rohscore</i>	48
Tabelle 28: Ergebnisse des LRTs der Skala <i>Up-To-Date</i> a-priori in Bezug auf das Teilungskriterium <i>Note</i>	50
Tabelle 29: Ergebnisse des LRTs der Skala <i>Up-To-Date</i> a-priori in Bezug auf das Teilungskriterium <i>Prüfungsfach</i>	51
Tabelle 30: Anzahl der Items a-priori und a-posteriori sowie Anzahl und Prozentsatz der Itemausschlüsse, dargestellt für jedes der vier Antwortformate. Die Ziffern in (/)* stellen Anzahl und Prozentsatz der Itemausschlüsse dar, die die Originalversion des DIC betreffen/ bzw. die Parallelversionen betreffen)	52
Tabelle 31 : Deskriptive Statistiken der Itemleichtigkeitsparameter je Antwortformat	54
Tabelle 32: Ergebnisse der Varianzanalyse	54
Tabelle 33: Itemleichtigkeitsparameter aller 188 Rasch-Modell-konformen Items, die mindestens in zwei Antwortformaten vorkommen (Entsprechungs-Items), für die verschiedenen Antwortformate gegenübergestellt zum Vergleich. Abweichungen, die entgegengesetzt der aufgrund der a-priori-Ratewahrscheinlichkeit erwarteten Richtung ausgefallen sind, sind grau unterlegt.	54
Tabelle 34: Ergebnisse der Newman-Keuls-Prozedur	56
Tabelle 35: Deskriptivstatistiken des Vergleichs der Itemparameter von MC-Items im Format „1 aus 6“ und „1 aus 4“	57

Tabelle 36:Ergebnisse des t-Tests für abhängige Stichproben	58
Tabelle 37: Itemleichtigkeitsparameterder Aufgaben des DIC, dargestellt pro Skala des DIC.....	67
Tabelle 38: Darstellung aller 198 Parallelitems des Diagnostik-Info-Check'08 – dargestellt pro Skala und geordnet nach Testform A/ B/ C/ D/ E. Bei den Items, die im Multiple-Choice-Antwortformat vorgegeben wurden, ist (sind) die richtige(n) Lösung(en) unterstrichen. Weiters sind alle nicht Rasch-Modell-konformen Items bzw. Items, deren Rasch-Homgenität aufgrund zu extremer Lösungshäufigkeiten nicht beurteilt werden konnte, grau unterlegt dargestellt.	71
Tabelle 39: Deskriptive Statistiken zur Variable Alter.....	98
Tabbelle 40: Altersverteilung der Testpersonen in der Stichprobe	98
Tabelle 41:Übersicht über die Häufigkeiten der angegebenen Diplomprüfungsfächer .	99
Tabelle 42:Deskriptive Statistiken zu den angegebenen Noten der vier Lehrveranstaltungen der Psychologischen Diagnostik.....	99
1. Einführung	2
2. Zielsetzung der vorliegenden Studie	11
3. Methode	12
3.1. Erhebungsinstrument, Entwicklung der Parallel- Items im MC-Antwortformat „1 aus 4“	12
3.2. Stichprobe	16
3.3. Testtheoretische Analysen der Items gemäß dem Rasch-Modell	18
3.4. Hypothese der Studie und methodische Vorgehensweise der interferenzstatistischen Überprüfung der Hypothese.....	22
4. Ergebnisse.....	24
4.1. Ergebnisse der Analysen nach dem Rasch-Modell.....	24

4.1.1.	Graphische Modellkontrollen, Ergebnisse der LRTs – A-priori und A-posteriori sowie Übersicht über die Itemausschlüsse.....	24
	Skala Grundlagenwissen	24
	Teilungskriterium Rohscore	25
	Teilungskriterium Alter	26
	Teilungskriterium Note	27
	Teilungskriterium Prüfungsfach.....	28
	Skala Statistik- und Testtheorie-Wissen	30
	Teilungskriterium Rohscore	31
	Teilungskriterium Alter	32
	Teilungskriterium Note	33
	Teilungskriterium Prüfungsfach.....	34
	Skala Verfahrensinventar-Kenntnis	35
	Teilungskriterium Rohscore	35
	Teilungskriterium Alter	36
	Teilungskriterium Note	37
	Teilungskriterium Prüfungsfach.....	38
	Skala Theoriewissen zur Differentiellen Psychologie	39
	Teilungskriterium Rohscore	39
	Teilungskriterium Alter	40
	Teilungskriterium Note	41
	Teilungskriterium Prüfungsfach.....	42
	Skala Verfahrensbezogenes Detailwissen.....	43
	Teilungskriterium Rohscore	43
	Teilungskriterium Alter	44
	Teilungskriterium Note	45
	Teilungskriterium Prüfungsfach.....	46
	Skala Up-to-Date.....	48
	Teilungskriterium Rohscore	48
	Teilungskriterium Alter	49
	Teilungskriterium Note	50
	Teilungskriterium Prüfungsfach.....	51
4.1.2.	Zusammenfassende Betrachtung der Ergebnisse der Rasch-Modell-Analysen, Überblick über die Item-Ausschlüsse	52
4.2.	Ergebnisse der Hypothesentestung.....	53
4.2.1.	Ergebnisse der Varianzanalyse.....	53
4.2.2.	Ergebnisse des Post-Hoc-Tests	56
4.2.3.	Ergebnisse des t-tests für abhängige Stichproben	57
	Diskussion.....	59
5.	Zusammenfassung und Abstract zur Studie.....	62
5.1.	Zusammenfassung	62
5.2.	Abstract	63
6.	Literaturverzeichnis	65
7.	Anhang	67

7.1.	Anhang A: Itemleichtigkeitsparameter der 154 Rasch-Modell-konformen Items des DIC `08	67
7.2.	Anhang B: Testbogendeckblatt	70
7.3.	Anhang B: Die Items des Diagnostik-Info-Check `08 – dargestellt pro Skala und Testform	71
7.4.	Anhang C: Weiterführende Stichprobeneigenschaften	98
8.	Tabellenverzeichnis	100
9.	Abbildungsverzeichnis	106

9. Abbildungsverzeichnis

Abbildung 1: Histogramm der Altersverteilung der Testpersonen in der Stichprobe.....	17
Abbildung 2: Graphischer Modelltest der Skala <i>Grundlagenwissen</i> in Bezug auf das Teilungskriterium <i>Rohscore</i> , a-priori	26
Abbildung 3: Graphischer Modelltest der Skala <i>Grundlagenwissen</i> in Bezug auf das Teilungskriterium <i>Rohscore</i> , a-posteriori	26
Abbildung 4: Graphischer Modelltest der Skala <i>Grundlagenwissen</i> in Bezug auf das Teilungskriterium <i>Alter</i> , a-priori	27
Abbildung 5: Graphischer Modelltest der Skala <i>Grundlagenwissen</i> in Bezug auf das Teilungskriterium <i>Alter</i> , a-posteriori	27
Abbildung 6: Graphischer Modelltest der Skala <i>Grundlagenwissen</i> in Bezug auf das Teilungskriterium <i>Note</i> , a-priori	28
Abbildung 7: Graphischer Modelltest der Skala <i>Grundlagenwissen</i> in Bezug auf das Teilungskriterium <i>Note</i> , a- posteriori	28
Abbildung 8: Graphischer Modelltest der Skala <i>Grundlagenwissen</i> in Bezug auf das Teilungskriterium <i>Prüfungsfach</i> , a-priori	29
Abbildung 9: Graphischer Modelltest der Skala <i>Grundlagenwissen</i> in Bezug auf das Teilungskriterium <i>Prüfungsfach</i> , a- posteriori	29
Abbildung 10: Graphischer Modelltest der Skala <i>Statistik- und Testtheorie-Wissen</i> in Bezug auf das Teilungskriterium <i>Rohscore</i> , a-priori	32
Abbildung 11: Graphischer Modelltest der Skala <i>Statistik- und Testtheorie-Wissen</i> in Bezug auf das Teilungskriterium <i>Alter</i> , a-priori	33
Abbildung 12: Graphischer Modelltest der Skala <i>Statistik- und Testtheorie-Wissen</i> in Bezug auf das Teilungskriterium <i>Note</i> , a-priori	34

Abbildung 13: Graphischer Modelltest der Skala <i>Statistik- und Testtheorie-Wissen</i> in Bezug auf das Teilungskriterium <i>Prüfungsfach</i> , a-priori	35
Abbildung 14: Graphischer Modelltest der Skala <i>Verfahrensinventar-Kenntnis</i> in Bezug auf das Teilungskriterium <i>Rohscore</i> , a-priori.....	36
Abbildung 15: Graphischer Modelltest der Skala <i>Verfahrensinventar-Kenntnis</i> in Bezug auf das Teilungskriterium <i>Alter</i> , a-priori	37
Abbildung 16: Graphischer Modelltest der Skala <i>Verfahrensinventar-Kenntnis</i> in Bezug auf das Teilungskriterium <i>Note</i> , a-priori.....	38
Abbildung 17: Graphischer Modelltest der Skala <i>Verfahrensinventar-Kenntnis</i> in Bezug auf das Teilungskriterium <i>Prüfungsfach</i> , a-priori.....	39
Abbildung 18: Graphischer Modelltest der Skala <i>Theoriewissen zur Differentiellen Psychologie</i> in Bezug auf das Teilungskriterium <i>Rohscore</i> , a-priori	40
Abbildung 19: Graphischer Modelltest der Skala <i>Theoriewissen zur Differentiellen Psychologie</i> in Bezug auf das Teilungskriterium <i>Alter</i> , a-priori	41
Abbildung 20: Graphischer Modelltest der Skala <i>Theoriewissen zur Differentiellen Psychologie</i> in Bezug auf das Teilungskriterium <i>Note</i> , a-priori.....	42
Abbildung 21: Graphischer Modelltest der Skala <i>Theoriewissen zur Differentiellen Psychologie</i> in Bezug auf das Teilungskriterium <i>Prüfungsfach</i> , a-priori.....	43
Abbildung 22: Graphischer Modelltest der Skala <i>Verfahrensbezogenes Detailwissen</i> in Bezug auf das Teilungskriterium <i>Rohscore</i> , a-priori.....	44
Abbildung 23: Graphischer Modelltest der Skala <i>Verfahrensbezogenes Detailwissen</i> in Bezug auf das Teilungskriterium <i>Rohscore</i> , a-posteriori	44
Abbildung 24: Graphischer Modelltest der Skala <i>Verfahrensbezogenes Detailwissen</i> in Bezug auf das Teilungskriterium <i>Alter</i> , a-priori.....	45
Abbildung 25: Graphischer Modelltest der Skala <i>Verfahrensbezogenes Detailwissen</i> in Bezug auf das Teilungskriterium <i>Alter</i> , a-posteriori	45

Abbildung 26: Graphischer Modelltest der Skala <i>Verfahrensbezogenes Detailwissen</i> in Bezug auf das Teilungskriterium <i>Note</i> , a-priori	46
Abbildung 27: Graphischer Modelltest der Skala <i>Verfahrensbezogenes Detailwissen</i> in Bezug auf das Teilungskriterium <i>Note</i> , a-posteriori	46
Abbildung 28: Graphischer Modelltest der Skala <i>Verfahrensbezogenes Detailwissen</i> in Bezug auf das Teilungskriterium <i>Prüfungsfach</i> , a-priori	47
Abbildung 29: Graphischer Modelltest der Skala <i>Verfahrensbezogenes Detailwissen</i> in Bezug auf das Teilungskriterium <i>Prüfungsfach</i> , a-posteriori	47
Abbildung 30: Graphischer Modelltest der Skala <i>Up-to-Date</i> in Bezug auf das Teilungskriterium <i>Rohscore</i> , a-priori	49
Abbildung 31: Graphischer Modelltest der Skala <i>Up-to-Date</i> in Bezug auf das Teilungskriterium <i>Alter</i> , a-priori	50
Abbildung 32: Graphischer Modelltest der Skala <i>Up-to-Date</i> in Bezug auf das Teilungskriterium <i>Note</i> , a-priori	51
Abbildung 33: Graphischer Modelltest der Skala <i>Up-to-Date</i> in Bezug auf das Teilungskriterium <i>Prüfungsfach</i> , a-priori	52

Sonderanhang: Lebenslauf der Autorin



Katharina Vock
Theresiengasse 6
2500 Baden

katharina.vock@univie.ac.at

katharina.vock@aon.at

Tel: 0650/8308968

Ausbildung

1987-1991	Volksschule Baden
1991-1999	AHS Frauengasse, Baden
1999	2 Semester Studium der Medizin (Uni Wien)
2000	Beginn des Studiums der Psychologie (Uni Wien)

Praktika/ Zusatzausbildungen im psychologischen Bereich

2002-2005	Ausbildung zur Sexualpädagogin und Projektarbeit bei „Achtung°Liebe“ der <i>Austrian Medical Students Association</i>
25.7.2005- 9.9.2005	Psychologisches Praktikum am LKH Klagenfurt, <i>Abteilung für Neurologie und Psychiatrie des Kindes- und Jugendalters</i>
Nov. 2005- Mai 2006	Forschungsstipendium der TU Wien <i>Wissenschaftliche Mitarbeit im Bereich „Automatic Recognition Systems“ (ARS) am Institut für Computertechnik</i>
Seit Okt. 2005	Spezialisierung auf den Bereich der psychologischen Diagnostik im Rahmen der Wahlfachausbildung

derzeitige Berufstätigkeit

Seit Okt. 2006 **Studienassistentin** am Arbeitsbereich
Psychologische Diagnostik
Institut für Entwicklungspsychologie und psychologische
Diagnostik, **Uni Wien**
(10h/Wo)

sonstige Berufserfahrung

Juli – Aug. 2000 Buchhandlung	Ferialjob als Buchhändlerin im Verkauf, Zweymüller in Baden
Juli- Aug. 2003	Ferialjob als Messemitarbeiterin bei Mazda Europa
seit Jän. 2003	Tätigkeit als Kindermädchen
2002- 2005	Service- Tätigkeit in Gastgewerbe und Konditorei

Kenntnisse

Sprachen

<i>Deutsch</i>	Muttersprache
<i>Englisch</i>	fundierte Kenntnisse in Wort und Schrift
<i>Französisch</i>	Grundkenntnisse in Wort und Schrift

EDV / Statistik

Grundkenntnisse in Windows 98 sowie Windows XP,
(Office Word, Excel), SPSS, sowie R in Anwendung zu Analysen
nach dem Rasch-Modell