



universität
wien

Diplomarbeit

Entwicklung einer informellen Leistungsüberprüfung zur
Lehrveranstaltung Psychologische Diagnostik I
im Multiple-Choice Format

Verfasser

Simon Lehner

angestrebter Akademischer Grad

Magister der Naturwissenschaften (Mag.rer.nat.)

Wien, im Mai 2009

Studienkennzahl lt. Studienblatt: A298

Studienrichtung lt. Studienblatt: Psychologie

Betreuer: Privatdoz. Mag. Dr. Michaela Wagner-Menghin

Abstract

Deutsch

Für die Lehrveranstaltung Psychologische Diagnostik I im Rahmen des Diplomstudiums Psychologie an der Universität Wien wurde eine informelle Leistungsüberprüfung (ILP) im Multiple-Choice Format erstellt. Die testtheoretische Güte der informellen Leistungsüberprüfung wurde mittels Analysen nach dem Rasch-Modell überprüft. Bei diesen Analysen ergab sich für die Items der ILP gute Modellpassung. Von 22 Items mussten lediglich zwei ausgeschieden werden, da diese die Modellannahmen verletzten. Weiters wurde die inhaltliche Gültigkeit der ILP mittels logistischer Regression am Außenkriterium Prüfungserfolg in der Prüfung zur Psychologischen Diagnostik I überprüft. Es wurde versucht, in diese Regression möglichst viele relevante Einflussfaktoren, wie die verbale Intelligenz der Studierenden, einzubinden. Trotzdem konnte ausschließlich das Ergebnis der ILP einen Teil der Varianz der Prüfungsergebnisse erklären. Im Rahmen der aktuellen Studie wurde ebenfalls die Fähigkeit der Studierenden, ihre Prüfungsergebnisse korrekt einzuschätzen, betrachtet. Hier zeigte sich eine generell gute Passung der Selbsteinschätzung mit der tatsächlich erbrachten Leistung. Zwischen der Selbsteinschätzung bei der ILP und der Selbsteinschätzung bei der Prüfung wurden signifikante Unterschiede gefunden und zwar dahingehend, dass Studierende ihre Leistung bei der ILP eher überschätzten als bei der Prüfung.

English

The goal of the following study was the construction of an informal test of student's knowledge (ILP) in the field of psychological assessment at the Faculty of Psychology at the University of Vienna. The test used a multiple-choice response format. The extent to which the exam adhered to the assumptions of the Rasch model was investigated. The analysis showed a good fit of the Rasch model assumptions, only two out of 22 Items violated the model assumptions and had to be excluded. The predictive validity of the test was investigated via a logistic regression model in which other predictors like verbal intelligence of the students together with the test results were used to predict the results of the exam "psychological assessment 1". Only the test result of the ILP showed a significant influence on the results of the exam. A second subject of interest was the accuracy of the students' self-assessment compared to their grade on the exam "psychological assessment 1" and their grades on the ILP. Results showed a good accuracy of the students' self-assessment compared to the ILP and the exam. Furthermore students predicted their grades on the exam significantly more accurately than their grades on the ILP.

Inhaltsverzeichnis

I. Einleitung	1
II. Rahmenbedingungen und Theoretischer Hintergrund	3
1 Anlass und Rahmenbedingungen	3
2 Theoretischer Hintergrund.....	6
2.1 Verbale Intelligenz	6
2.2 Selbsteinschätzung der Prüfungsleistung	7
2.3 Eigenschaften des Multiple-Choice Antwortformats.....	8
III. Empirischer Teil	10
3 Ziel der Untersuchung	10
4 Methode	12
4.1 Untersuchungsplan	12
4.2 Erhebungsinstrumente	13
4.3 Durchführung der Untersuchung.....	21
4.4 Stichprobe.....	23
4.5 Eingesetzte Statistische Verfahren	25
5 Ergebnisse.....	29
5.1 Analysen nach dem Rasch-Modell.....	29
5.2 Deskriptive Statistiken	46
5.3 Ergebnisse der Logistischen Regression	53
5.4 Selbsteinschätzung der Leistung	62
5.5 Beantwortung der Fragestellungen.....	71
6 Diskussion und Ausblick	73
7 Zusammenfassung	77

IV. Referenzen	80
8 Verzeichnisse	80
8.1 Literatur	80
8.2 Abbildungen	82
8.3 Tabellen	83
9 Anhang	85
9.1 Quellen	85
9.2 Items der ILP	85
9.3 Zusatzfragen zur ILP	91
9.4 Nachrichten an die Studierenden via Blackboard Vista	94
9.5 Einleitende Beschreibung der ILP	95
9.6 Fragebogen Prüfung	96
9.7 Lehrziele zur VO Diagnostik I	98
9.8 Emailaufforderung zur Teilnahme an der ILP	99
9.9 Lebenslauf	101

I. Einleitung

Diese Arbeit ist eingebunden in das PeRaS Projekt von PD Mag. Dr. Michaela Wagner-Menghin. Ziel des Projekts ist die systematische Untersuchung von Prädiktoren für den Studienerfolg. In einem ersten Schritt sollten explorativ mögliche Prädiktoren für den Studienerfolg erhoben werden. Als Kriterium für den Studienerfolg wurde in dieser ersten Phase des Projektes das erfolgreiche Absolvieren der Ausbildung zur Psychologischen Diagnostik von Studierenden herangezogen.

Zielsetzung dieser Diplomarbeit war die Entwicklung einer informellen Leistungsüberprüfung (ILP) im Multiple-Choice Format für die Lehrveranstaltung Psychologische Diagnostik I, die als Self-Assessment für Studierende dienen sollte. Studierende sollten eine Rückmeldung über ihren aktuellen Wissensstand zur Psychologischen Diagnostik erhalten und in der ILP eine Hilfe bei der Vorbereitung zur Prüfung zur Psychologischen Diagnostik I finden.

Die Items dieser Leistungsüberprüfung orientieren sich inhaltlich an der Lehrveranstaltung „Psychologische Diagnostik I“ unter der Leitung von PD Dr. Michaela Wagner-Menghin. Die Aufteilung der Items auf reine Wissensitems und solche die Verständnis und Anwendung des Wissens erfordern, erfolgte nach den Lehrzielen von Bloom (1965). Formal orientierten sich die Items an den Richtlinien zur Konstruktion von Multiple-Choice Items von Litzenberger, Gnambs und Punter (2005).

Der theoretische Teil dieser Arbeit beschäftigt sich einleitend mit dem Anlass für diese Diplomarbeit und den Rahmenbedingungen an der Universität Wien (Kapitel 1). Im Weiteren wird auf den theoretischen Hintergrund (Kapitel 2) eingegangen, insbesondere auf die verbale Intelligenz, die Selbsteinschätzung und das Multiple-Choice Antwortformat.

Der empirische Teil dieser Arbeit beschäftigt sich zuerst mit der Überprüfung der testtheoretischen Güte der Items der ILP. In einem zweiten Schritt soll überprüft werden ob die Ergebnisse dieses Self-Assessments auch mit dem Prüfungserfolg bei der Prüfung zur Psychologischen Diagnostik I zusammenhängen. Sollte kein ausreichender Zusammenhang zwischen der Leistung in der ILP und der Prüfungsleistung bestehen, so würde dies bedeuten, dass die ILP als Mittel zur Leistungsfeststellung für Studierende ungeeignet ist. Im Rahmen dieser Überprüfung werden auch weitere mögliche Einflussfaktoren des

Prüfungserfolgs untersucht, wie Aspekte des Studierverhaltens (Besuch von Lehrveranstaltungen) und die verbale Intelligenz der Studierenden.

Weiters wird das Ergebnis der ILP zusammen mit anderen möglichen Prädiktoren (Besuch von Lehrveranstaltungen, verbale Intelligenz) mittels logistischer Regression auf den Zusammenhang mit dem Prüfungsergebnis hin untersucht.

Ein weiterer Untersuchungsgegenstand ist die Fähigkeit der Studierenden ihre Prüfungsleistung korrekt einzuschätzen, also die Selbsteinschätzung. Dies basiert auf Gelegenheitsbeobachtungen der Lehrveranstaltungsleitung, dass Studierende nach der Prüfung, insbesondere in der Prüfungseinsicht häufig ihre erbrachte Leistung überschätzten.

Abschließend erfolgt noch eine Betrachtung der Fähigkeit zur richtigen Selbsteinschätzung der erbrachten Leistung in der ILP und der Prüfung zur Psychologischen Diagnostik. Dabei wird einerseits betrachtet wie viele Studierende ihre Leistung gut einschätzen konnten und andererseits ob sich die Güte der Selbsteinschätzung zwischen der Prüfung und der ILP unterscheiden.

II. Rahmenbedingungen und Theoretischer Hintergrund

1 Anlass und Rahmenbedingungen

Anlass für diese Diplomarbeit war das PeRaS Projekt von PD Dr. Wagner-Menghin, zum Zeitpunkt dieser Arbeit wissenschaftliche Assistentin und Lektorin am Institut für Psychologische Diagnostik und Entwicklungspsychologie der Fakultät für Psychologie an der Universität Wien.

Für ein besseres Verständnis der Rahmendbedingungen dieser Arbeit wird im Folgenden die Ausbildung zur Psychologischen Diagnostik im Rahmen des Diplomstudiums Psychologie an der Universität Wien vorgestellt. Die Ausbildung ist in zwei Teile gegliedert. Im ersten Teil sind zwei Vorlesungen (Psychologische Diagnostik 1 & 2) und zwei Lehrveranstaltungen mit immanentem Prüfungscharakter (Übungen zur Psychologischen Diagnostik 1&2) zu absolvieren, wobei seitens der Lehrveranstaltungsleitung empfohlen wird diese Lehrveranstaltungen parallel im gleichen Semester zu besuchen. Dies ist allerdings nicht verpflichtend. Der zweite Teil der Ausbildung zur Psychologischen Diagnostik beinhaltet wiederum zwei Lehrveranstaltungen mit immanentem Prüfungscharakter, die „Demonstration psychologisch-diagnostischer Fallbeispiele“ und ein „Praktikum zum Psychologischen Diagnostizieren“. Bei beiden Lehrveranstaltungen ist das positive Abschließen des ersten Teils der Ausbildung Voraussetzung.

Die Prüfung zur Psychologischen Diagnostik 1 ist eine schriftliche Prüfung mit freiem Antwortformat. Die Fragen der Prüfung selbst sollen nicht ausschließlich Wissen abfragen, sondern sollen ebenfalls erfassen ob der Studierende die wesentlichen Grundideen der Psychologischen Diagnostik verstanden hat und in der Lage ist aus dem theoretischen Wissen die für die Praxis relevanten Schlüsse zu ziehen. Es soll bei der Prüfung also ein Schwerpunkt auf höhere Lehrziele wie Verständnis, Anwendung und Beurteilung gelegt werden (Fakultät für Psychologie, Arbeitsbereich Psychologische Diagnostik, 2008). Aus diesem Grund ist auch die Mitnahme sämtlicher Unterlagen wie der Mitschrift oder dem Lehrbuch zur Prüfung gestattet.

Von zahlreichen Studierenden wird diese Prüfung und die Prüfung zur Psychologischen Diagnostik 2 als Hürde im Studium wahrgenommen. Auch scheinen Studierende bei dieser Prüfung ihre erbrachte Leistung oft nicht richtig einschätzen zu können. (M. Wagner-

Menghin, persönliche Mitteilung, 25.11.2008). Eine Hypothese dafür ist, dass beim Verständnis der Prüfungsfragen auch die verbale Intelligenz der Studierenden eine Rolle spielt. Die Fragen zielen darauf ab nicht bloßes Wissen abzufragen, sondern versuchen auch das Verständnis für Zusammenhänge und die praktische Bedeutung des Wissen zu erfassen. Dadurch sind beim Beantworten der Fragen möglicherweise auch das Sprachverständnis, die Ausdrucksfähigkeit und die verbale Intelligenz gefordert. Dies steht im Gegensatz zu vielen anderen Prüfungen aus dem Diplomstudium Psychologie, bei denen meist ausschließlich (auswendig) gelerntes Wissen abgeprüft wird, häufig im Multiple-Choice Antwortformat, bei dem, im Gegensatz zum freien Antwortformat, eher die Fähigkeit zum Wiedererkennen gefordert ist, wohingegen beim freien Antwortformat das selbstständige reproduzieren von Wissen gefragt ist (Kubinger, 2003) .

Ein weiteres Thema, das mit dieser Diplomarbeit, wenn auch nur am Rande, zusammenhängt, ist die mögliche Gestaltung der Prüfung ganz oder teilweise im Multiple-Choice Antwortformat. Momentan ist durch das freie Antwortformat und die inhaltliche Gestaltung der Fragen die Auswertung der Prüfung mit erheblichem Zeitaufwand verbunden. Hier wäre es interessant zu wissen ob es überhaupt möglich ist einen Itempool im Multiple-Choice Antwortformat für den Wissensbereich der psychologischen Diagnostik zu erstellen, der auch testtheoretischen Gütekriterien gerecht wird, insbesondere den Annahmen des Rasch-Modells. In der aktuellen Untersuchung würde die Geltung des Rasch-Modells zumindest für die Möglichkeit der Erstellung eines solchen Itempools sprechen. Zu einer möglichen Gestaltung der Prüfung selbst im Multiple-Choice Antwortformat, ob teilweise oder gesamt, wird in dieser Arbeit nicht näher eingegangen.

Um die mangelnde Fähigkeit der Studierenden, ihr Wissen in Bezug auf die Prüfung zur Psychologischen Diagnostik richtig einzuschätzen, und ihnen darüberhinaus ein Hilfsmittel zur Vorbereitung auf die Prüfung zur Verfügung zu stellen, bietet sich eben eine informelle Leistungsüberprüfung an. Natürlich müsste eine solche Leistungsüberprüfung ausreichend Bezug zu den Anforderungen der Lehrveranstaltung haben, sowohl durch Items die sich stark an den Lehrzielen orientieren, aber auch den Prüfungsfragen in einem gewissen Ausmaß entsprechen.

Aufgrund dieser Überlegungen wurde als Zielsetzung der vorliegenden Diplomarbeit, die Entwicklung einer informellen Leistungsüberprüfung (im Weiteren kurz als ILP

bezeichnet) im Multiple-Choice Format für die Lehrveranstaltung Psychologische Diagnostik I formuliert.

2 Theoretischer Hintergrund

2.1 Verbale Intelligenz

Verbale Intelligenz, wie sie in der aktuellen Untersuchung verstanden wird, basiert auf einer modifizierten Version der Intelligenztheorie von Thurstone und Thurstone (1941) wie sie in der aktuellen Version¹ des Verfahrens Intelligenz-Basis-Funktionen (Blum et al., 2005) Verwendung findet. Thurstone und Thurstone definierten in ihrer Theorie 7 Primärfaktoren der Intelligenz, wobei zwei dem „verbalen Bereich“ zuzuordnen sind: *verbal comprehension* (Verbales Verständnis) und *word fluency* (Verbale Flüssigkeit). Die Verbale Intelligenz, wie sie mittels IBF erhoben wird, entspricht ausschließlich dem Bereich *verbal comprehension*.

Da die ersten nennenswerten Intelligenztests dazu konstruiert wurden den späteren Schulerfolg, beziehungsweise die spätere Bewährung in schulischen Situationen vorherzusagen gibt es zahlreiche Untersuchungen über den Zusammenhang von Intelligenz und Schulerfolg. Diese Untersuchungen zum Vergleich von Punktwerten in Intelligenztests mit Kriterien für den schulischen Erfolg liefern die höchsten Übereinstimmungen der psychologischen Diagnostik überhaupt. Insgesamt besteht eine Tendenz zu höheren Übereinstimmungen verbaler Tests mit schulischem Erfolg im Vergleich zu nonverbalen Skalen (Amelang & Bartussek, 1997).

Für die aktuelle Untersuchung relevant ist allerdings der Zusammenhang von (verbaler) Intelligenz mit dem Erfolg im Studium. Hier zeigen sich deutlich niedrigere, aber immer noch signifikante Zusammenhänge der verschiedenen Intelligenzskalen mit dem Studienerfolg. Gründe dafür sind einerseits eine aufgrund der Probandenstichproben reduzierte Leistungsvariabilität sowie die niedrige psychometrische Qualität von Hochschulnoten (Amelang & Bartussek, 1997).

¹ Version 22.00

2.2 Selbsteinschätzung der Prüfungsleistung

Die Selbsteinschätzung kann hinsichtlich ihrer Korrektheit bewertet werden indem sie mit einer objektiven Leistungsmessung verglichen wird. Bei diesem Vergleich spricht man bei guter Passung von Akkuratess² oder *accuracy* (Braun, 2003). Kommt es zu deutlichen Abweichungen zwischen der objektiven Messung und der Selbsteinschätzung so spricht man von Über- beziehungsweise Unterschätzung.

Braun (2003) sieht folgende Probleme bei mangelnder Passung der Selbsteinschätzung. Überschätzt eine Person ihre Leistung, so besteht die Tendenz Handlungen auszuführen die mit Fehlern behaftet sind, auch wenn diese nicht wahrgenommen werden. Auf die aktuelle Situation der Selbsteinschätzung der Prüfungsleistung umgelegt kann das bedeuten, dass Studierende, die ihr Wissen zur Psychologischen Diagnostik überschätzen, vorzeitig zur Prüfung antreten, obwohl zum erfolgreichen Absolvieren noch weitere Vorbereitung nötig wäre. Diese Studierenden werden möglicherweise nach der Prüfung über ein negatives Ergebnis verwundert sein und mit Unverständnis reagieren.

Unterschätzt eine Person ihre Fähigkeit und traut sich daher Handlungen nicht zu, obwohl sie über die nötigen Kompetenzen verfügt, besteht die Tendenz, Handlungen zu unterlassen welche zum Erreichen von Zielen oder zur Lösung eines Problems nötig wären. Wiederum auf die aktuelle Situation der Selbsteinschätzung der Prüfungsleistung umgelegt bedeutet dies, dass Studierende, die ihr Wissen zur Psychologischen Diagnostik unterschätzen, trotz ausreichendem Wissen nicht zur Prüfung antreten. Dies kann zur Verzögerung des Studienfortschrittes sowie, im schlimmsten Fall, zum Abbruch des Studiums führen.

Es existiert keine einheitliche und anerkannte Theorie zur Selbsteinschätzung. Dafür gibt es viele psychologische Theorien, die sich mit unterschiedlichen Aspekten der Selbsteinschätzung befassen. Braun (2003) nennt hierzu folgende Theorien: die Theorien der Handlungsregulation (Miller, Galanter & Pribram, 1960, Dörner, 1983; Hacker, 1986, 1993; Panzer, 2000), die kognitive Verhaltenstheorie (Lazarus, 1966; Lazarus & Folkman, 1984; Carver & Scheier, 1981,1991), die sozial-kognitiven Lerntheorie (Bandura, 1979; Skinner & Chapman, 1984; Skinner, Chapman & Baltes, 1988b; Flammer, 1990), die

² Im weiteren Verlauf der Arbeit wird die Bezeichnung „gute Einschätzung“ oder „gute Passung“ verwendet.

Motivationspsychologie (Heckhausen, 1984), die Konsistenztheorie (Festinger, 1954, 1957) und die Selbstwert-Theorie (Covington, 1984a; Mummendey, 1990, 2000).

Für die aktuelle Untersuchung relevant ist welche Faktoren Einfluss auf die Genauigkeit der Selbsteinschätzung haben können. Braun (2003) gibt eine Literaturübersicht zu untersuchten Gründen für Über- und Unterschätzung sowie korrekte Selbsteinschätzung. Zwei mögliche Gründe scheinen für die aktuelle Untersuchung relevant: die Aufgabenschwierigkeit und die selektive Wahrnehmung erwünschter Ergebnisse.

Lichtenstein und Fischhoff (1977, 1978, 1980, zitiert nach Braun, 2003) konnten folgende Auswirkungen der Aufgabenschwierigkeit auf die Güte der Selbsteinschätzung zeigen. Bei sehr leichten Aufgaben unterschätzen sich die Versuchspersonen eher und bei sehr schwierigen Aufgaben überschätzen Versuchspersonen eher die von ihnen erbrachte Leistung. Diese Ergebnisse konnten in zahlreichen Studien repliziert werden (vgl. Braun, 2003)

Selektive Wahrnehmung beeinflusst die Güte der Selbsteinschätzung dahingehend, dass Personen die Auftretenswahrscheinlichkeit von erwünschten Ergebnissen signifikant überschätzen und von unerwünschten Ergebnissen signifikant unterschätzen (Irwin, 1953, zitiert nach Braun, 2003).

2.3 Eigenschaften des Multiple-Choice Antwortformats

Beim Multiple-Choice Antwortformat werden, im Gegensatz zum freien Antwortformat, zu einer Aufgabestellung jeweils zwei oder mehrere Antwortmöglichkeiten vorgegeben aus denen die jeweils Richtige auszuwählen ist, beziehungsweise die Richtigen auszuwählen sind.

Die Verwendung des Multiple-Choice Formats für Wissenstests bringt einige spezifische Vor- und Nachteile mit sich. Zu den Vorteilen zählen die Verrechnungssicherheit, die besondere Eignung für Gruppentestungen sowie der geringe Auswertungsaufwand. Nachteile des Multiple-Choice Formats sind der geringere Informationsgewinn gegenüber dem freien Antwortformat, der mehr oder weniger große Rateeffekt sowie dem inhaltlichen Unterschied zwischen „Wiedererkennen“ gegenüber „Reproduzieren“ bei freiem Antwortformat (Kubinger, 2006).

Der für die aktuelle Arbeit entscheidende Vorteil liegt im geringen Auswertungsaufwand des Multiple-Choice Formats. Im konkreten Fall übernimmt die Auswertung der Computer und ermöglicht eine unmittelbare Rückmeldung der Ergebnisse an die Probanden.

Der Rateeffekt soll in der vorliegenden Arbeit durch die Festlegung auf das Antwortformat „x aus 5“ minimiert werden. Gottschall (2007) konnte im Rahmen seiner Diplomarbeit zeigen, dass sich das Antwortformat „x aus 5“ hinsichtlich der Itemschwierigkeit bei Items gleichen Inhalts nicht vom freien Antwortformat unterscheidet. Dies legt einen ausreichend niedrigen Rateeffekt nahe und empfiehlt somit das Antwortformat „x aus 5“ dann einzusetzen, wenn das freie Antwortformat, aus welchen Gründen auch immer, nicht praktikabel erscheint. Da für die Items der ILP das freie Antwortformat aus weiter oben genannten Gründen nicht in Frage kommt, ist also das Antwortformat „x aus 5“ naheliegend.

III. Empirischer Teil

3 Ziel der Untersuchung

Ziel dieser Untersuchung ist es, in einem ersten Schritt, eine informelle Leistungsüberprüfung im Multiple-Choice Format zur Vorlesung „Psychologische Diagnostik 1“ zu entwickeln und anhand der probabilistischen Testtheorie zu analysieren. Dabei interessiert insbesondere die Geltung des Rasch-Modells, da dessen Modelleigenschaften³ die Verwendung des Summenscores als Prädiktor für den Prüfungserfolg in der Prüfung „Psychologische Diagnostik 1“ begünstigt. Aus diesen Überlegungen leitet sich somit die erste Fragestellung ab:

Fragestellung 1:

Gelten die Annahmen des Rasch-Modells für die Items der informellen Leistungsüberprüfung?

Als zweiten Schritt soll die Validität der informellen Leistungsüberprüfung anhand eines externen Kriteriums (Prüfungserfolg in der Prüfung „Psychologische Diagnostik 1“) überprüft werden. Dabei interessiert einerseits die Vorhersagekraft der informellen Leistungsüberprüfung bezüglich dieses Kriteriums, wie auch der Einfluss anderer möglicher Prädiktoren. Aus diesen Zielen ergeben sich zwei weitere Fragestellungen.

Fragestellung 2:

Besteht ein Zusammenhang zwischen den Ergebnissen der informellen Leistungsüberprüfung und den Ergebnissen der Prüfung „Psychologische Diagnostik 1“?

Fragestellung 3:

Welche weiteren Prädiktoren leisten einen Beitrag zur Vorhersage der Ergebnisse der Prüfung „Psychologische Diagnostik 1“?

Eine weitere interessierende Fragestellung betrifft die Selbsteinschätzung der Prüfungsleistung durch die Studierenden. Diese Fragestellung ergab sich insbesondere aus der Gelegenheitsbeobachtung der Lehrveranstaltungsleitung, dass Studierende ihre

³ Eindimensionalität, Summenscore als erschöpfende Statistik, vgl. Kapitel 2.2.1

Leistung bei der Prüfung „Psychologische Diagnostik 1“ häufig falsch einschätzten und zwar in die Richtung dass sie ihre erbrachte Leistung bei der Prüfung eher überschätzten (vgl. Kapitel 1).

Fragestellung 4:

Schätzen Studierende ihre erbrachte Leistung richtig ein oder neigen sie zu Über- oder Unterschätzung der erbrachten Leistung?

Aufgrund dieser Fragestellung war es nur naheliegend auch zu untersuchen ob es Unterschiede in der Selbsteinschätzung zwischen der Prüfung und der ILP gibt.

Fragestellung 5:

Gibt es Unterschiede in der Selbsteinschätzung der Studierenden zwischen der Prüfung „Psychologische Diagnostik 1“ und der ILP?

4 Methode

4.1 Untersuchungsplan

Die Untersuchung erfolgte als Teil des PeRas-Projektes unter Leitung von PD Mag. Dr. Wagner-Menghin am Arbeitsbereich Psychologische Diagnostik der Fakultät für Psychologie der Universität Wien. Sie bediente sich der anfallenden Stichprobe all jener Studierenden, die einerseits im Sommersemester 2008 die Lehrveranstaltung „Übungen zum psychologischen Diagnostizieren 1“ besuchten oder andererseits an der Prüfung zur Lehrveranstaltung „Psychologische Diagnostik 1“ zum 1. Prüfungstermin im Sommersemester 2008 teilnahmen.

Basierend auf den Lehrzielen der Vorlesung psychologische Diagnostik 1 wurde eine Informelle Leistungsüberprüfung (ILP) im Multiple-Choice Format entwickelt. Kapitel 4.1 schildert im Detail das Vorgehen bei der Entwicklung der Items. Die ILP wurde den Studierenden mittels der Lernplattform der Universität Wien, Blackboard Vista, vorgegeben.

Neben einer testtheoretischen Überprüfung der ILP nach dem Rasch-Modell, sollte überprüft werden ob der bei der ILP erreichte Score gemeinsam mit weiteren Variablen mit dem Prüfungserfolg zusammenhängt. Die zusätzlichen Variablen wurden zum Teil via Fragebogen im Anschluss an die ILP und nach der Prüfung erhoben. Dies betraf die Variable Vorlesungsbesuch. Ein anderer Teil der Daten wurde durch die Lehrveranstaltungsleitung zur Verfügung gestellt. Von der Lehrveranstaltungsleitung kamen die Daten zum Übungsbesuch der Studierenden. Als weiterer möglicher Prädiktor kam die Verbale Intelligenz zur Verwendung, erhoben im Rahmen der Diplomarbeitsstudie von Stadler (2009) mittels dem Testverfahren Intelligenz Basis Funktionen (IBF, Blum et al. 2005). Die Daten zur ILP wurden alle im Zeitraum vom 02.06.2008 bis zum 13.06.2008 erhoben. Die Daten des IBF wurden bereits im Zeitraum vom 05.05.2008 bis 15.05.2009 erhoben. Die Prüfung selbst fand am 27.6.2008 statt.

4.2 Erhebungsinstrumente

Zur Datenerhebung wurden folgende Instrumente eingesetzt:

- Informelle Leistungsüberprüfung (ILP, vgl. Kapitel 4.2.1)
- Zusatzfragebogen zur ILP (vgl. Kapitel 4.2.2)
- Fragebogen zur Prüfung (vgl. Kapitel 4.2.3)
- Unterteste zur verbalen Intelligenz des IBF (vgl. Kapitel 4.2.4)

4.2.1 Informelle Leistungsüberprüfung (ILP)

Die Basis dieser Arbeit bildet die Konstruktion und Vorgabe einer informellen Leistungsüberprüfung (ILP) im Multiple-Choice Format. Zielgruppe waren Studierende die im Sommersemester 2008 mit ihrer Ausbildung im Bereich psychologischer Diagnostik begannen, also entweder die Vorlesung „Psychologische Diagnostik 1“ besuchten oder an der Lehrveranstaltung „Übungen zum psychologischen Diagnostizieren 1“ teilnahmen. Die ILP sollte den Studierenden eine Feststellung ihres aktuellen Wissensstandes im Fach ermöglichen, der einerseits nicht von der Lehrveranstaltungsleitung beurteilt wird und es ihnen andererseits ermöglicht noch rechtzeitig vor dem Prüfungstermin fehlendes Wissen anzueignen. Dadurch ergeben sich besondere Anforderungen an die Items der ILP. Einerseits sollten die Items inhaltlich möglichst nahe an den Inhalten der Prüfung und somit an den Lehrzielen der Vorlesung Psychologische Diagnostik 1 liegen. Andererseits sollte eine automatische Auswertung möglich sein um eine unmittelbare, oder zumindest rasche Rückmeldung der Ergebnisse zu ermöglichen und dies ohne nennenswerten Auswertungsaufwand. Des Weiteren sollte sich die Schwierigkeit der Items nicht aufgrund ihrer Gestaltungsweise von der Schwierigkeit der Prüfungsfragen unterscheiden.

Ziel bei der Itemkonstruktion war es zu den Lehrzielen der Vorlesung Psychologische Diagnostik 1 jeweils drei Items pro Lehrziel zu konstruieren. Dadurch sollte gewährleistet werden, dass die ILP inhaltlich möglichst nahe an der Prüfung Psychologische Diagnostik 1 liegt. In Tabelle 1 sind die Lehrziele gelistet, die seitens der Lehrveranstaltungsleitung definiert wurden und in dieser Form auch auf der Homepage⁴ nachzulesen sind.

⁴ Homepage des Arbeitsbereichs „Psychologische Diagnostik“ des Instituts für Entwicklungspsychologie und Psychologische Diagnostik der Universität Wien (Stand Mai 2008).

Tabelle 1 Lehrziele der Lehrveranstaltung Psychologische Diagnostik 1 laut Institutshomepage (Stand Mai 2008)

A	Kenntnis und Fähigkeit zur Umsetzung der einzelnen Schritte des psychologisch-diagnostischen Prozesses inklusive der organisatorischen Konzeption des Vorgehens.
B	Fähigkeit zur (rechnerischen) Transformation verschiedener Normwerte um deren Vergleichbarkeit zu gewährleisten inklusive der Berücksichtigung von Konfidenzintervallen.
C	Kenntnis der Gütekriterien und Verständnis für die Problematik, die mit den Gütekriterien verbunden ist (z. B. unterschiedliche Arten der Validierung inklusive der Vor- und Nachteile dieser Arten)
D	Kenntnis der verschiedenen Arten psychologisch-diagnostischer Verfahren und Fähigkeit, psychologisch-diagnostische Verfahren, kritisch reflektieren zu können.
E	Kenntnis der psychologisch-diagnostischen Verfahren des Testinventars zur Leistungsdiagnostik.
F	Fähigkeit, verschiedenste Gestaltungsweisen (Antwortformate, Power- vs. Speed & Power, Gruppen- vs. Individualverfahren) von psychologisch-diagnostischen Verfahren kritisch reflektieren zu können.
G	Fähigkeit, verschiedenste Erhebungstechniken kritisch reflektieren zu können.
H	Kenntnis testtheoretischer Grundlagen, insbesondere Verständnis für die Umsetzung Klassischer Testtheorie und Item-Response-Theorie in der Psychologischen Diagnostik.

Die Erstellung der Items orientierte sich an den „Richtlinien zur Konstruktion von Multiple-Choice Items“ (Litzenberger, Gnambs & Punter, 2005). Darin werden, analog den Lernzielen von Bloom (1956), fünf Arten von Iteminhalten unterschieden:

- Wissen (z. B. Fachausdrücke und spezifische Fakten kennen)
- Verständnis (z. B. verbale Informationen oder Graphiken interpretieren)
- Anwendung (z. B. Gesetze und Theorien auf praktische Situationen umlegen)
- Analyse (z. B. logische Fehlschlüsse erkennen)
- Beurteilung (z. B. Korrektheit und Interpretationen aufgrund der Datenlage beurteilen)

Es wurde versucht zu jedem in Tabelle 1 gelisteten Lehrziel drei Items zu erstellen. Dabei sollte je ein Item dem Lernziel Wissen, eines den Lernzielen Verständnis und Anwendung und eines den Lernzielen Analyse und Beurteilung zuordenbar sein. In Tabelle 2 ist zu sehen welche Items welchen Iteminhalten zugeordnet sind. Eine Auflistung aller 22

erstellter Items findet sich im Anhang. Für das Lehrziel G wurden keine Items, für das Lehrziel B hingegen ein zusätzliches Wissensitem erstellt (vgl. Tabelle 2).

Das Verhältnis von Wissensitems zu Items höherer Lernziele von 1:2 wurde gewählt, um möglichst nahe an der Aufteilung der Prüfung Psychologische Diagnostik 1 zu sein, bei der auch ein Schwerpunkt auf die höheren Lernziele Verständnis, Anwendung, Analyse und Beurteilung gelegt wird.

Tabelle 2 Zuordnung der einzelnen Items zu den entsprechenden Lehrzielen und Iteminhalten. Die genaue Beschreibung der Lehrziele findet sich in Tabelle 1

Lehrziel	Wissen	Verständnis/Anwendung	Analyse/Beurteilung	Gesamt
A	A1	A2	A3	3
B	B1a;B1b	B2	B3	4
C	C1	C2	C3	3
D	D1	D2	D3	3
E	E1	E2	E3	3
F	F1	F2	F3	3
G	-	-	-	0
H	H1	H2	H3	3
Gesamt	8	7	7	22

Entscheidend für die Qualität von Multiple-Choice Items ist auch die Auswahl geeigneter Distraktoren. Dazu wurden die Prüfungsbögen vergangener Prüfungen zur Psychologischen Diagnostik 1 durchgearbeitet und häufige Fehler der Studierenden gesucht. Sollte ein Fehler in mehreren Prüfungsbögen vorgekommen sein, so wurde versucht aus der dem Fehler zugrunde liegenden Annahme, geeignete Distraktoren für die entsprechende Frage zu entwickeln. Als Beispiel soll hier das Vorgehen vieler Studierender bei der Skizzierung des diagnostischen Prozesses dargestellt werden. Meist wird in Prüfungen eine konkrete psychologische Fragestellung gegeben und die Studierenden sollen ihr weiteres Vorgehen beschreiben. Dabei wird regelmäßig der Fehler begangen, dass weder die Ausgangssituation ausreichend berücksichtigt wird, noch die Verhältnismäßigkeit der eingesetzten Verfahren. Ein besonders häufiges Vorgehen ist es, Tätigkeitsanalysen durchzuführen ohne zu reflektieren ob es nicht schon Anforderungsprofile gibt oder ob eine Tätigkeitsanalyse überhaupt zur Lösung der Fragestellung beiträgt. Ein Item, das diese Überlegungen berücksichtigt ist das Item A2 (vgl. Abbildung 1).

Sie erhalten den Auftrag für ein großes Unternehmen die Lehrlingsauswahl durchzuführen. Es sollen im Schnitt 20 Kandidaten pro Woche begutachtet werden. Wie sieht Ihr unmittelbar nächster Schritt, nach erfolgter Formulierung der Fragestellung, aus?

Sie führen eine Tätigkeitsanalyse durch

Sie klären ob ein Anforderungsprofil erstellt werden muss

Sie lassen die Kandidaten eine Arbeitsprobe ablegen

Sie wählen die einzusetzenden psychologisch diagnostischen Verfahren aus

Sie führen computergestützte Gruppentestungen durch

Abbildung 1 Item A2; die korrekte Antwort ist unterstrichen.

Zusätzlich wurden die Items der ILP von Experten, nämlich Diplomanden des Arbeitsbereichs Psychologische Diagnostik der Universität Wien durchgesehen und kommentiert. Diese Kommentare betrafen sowohl inhaltliche als auch testpsychologische Aspekte und wurden in der fertigen Version der ILP berücksichtigt.

Um die Ratewahrscheinlichkeit der Items möglichst gering und ihre Schwierigkeit somit möglichst nahe am freien Antwortformat zu halten, wurde das Antwortformat „x aus 5“ verwendet. Das Antwortformat „x aus 5“ bedeutet, dass von 5 gegebenen Antwortmöglichkeiten zwischen 0 und 5 richtig sind. Dazu konnte Gottschall (2007) zeigen, dass sich das Antwortformat „x aus 5“ bei gleichen Iteminhalten hinsichtlich der Itemschwierigkeit nicht vom freien Antwortformat unterscheidet, hingegen das bei Multiple-Choice Items beliebte Antwortformat „1 aus 6“ signifikant geringere Itemschwierigkeiten aufweist. Bei der Auswertung des Antwortformats „x aus 5“ werden nur Antworten als richtig verrechnet, bei denen alle richtigen Antwortmöglichkeiten markiert und alle falschen nicht markiert sind. Ein Item wird auch dann als falsch verrechnet wenn eine richtige Antwortmöglichkeit nicht markiert wurde oder nur eine falsche markiert wurde; es gilt also nur die Verrechnung 0 für falsch und 1 für richtig.

Die Items selbst wurden in einer zufälligen Reihenfolge vorgegeben, die zuvor mittels der Software "Cademio - light" (Version 3.20 2006) ermittelt wurde. Jeder Studierende bekam die Items also in der Reihenfolge: C2, B1b, D3, B1a, C3, A2, E1, B3, F2, E2, D1, F3, H2, B2, H1, F1, H3, D2, A3, C1, E3.

Die Position der Antwortalternativen der einzelnen Items wurde bei der Vorgabe durch die Lernplattform zufällig variiert. Dies sollte verhindern, dass die richtige(n) Antwort(e)n immer an derselben Stelle zu finden sind.

4.2.2 Zusatzfragebogen zur ILP

Im Anschluss an die Fragen der ILP wurden noch acht Zusatzfragen vorgegeben. Die Zusatzfragen hatten folgende Inhalte:

- Fortschritt der Prüfungsvorbereitung (Z1)
- Geplanter Prüfungstermin (Z2)
- Bemühen bei Beantwortung der ILP (Z3)
- Vorlesungsbesuch (Z4)
- Interesse an Psychologischer Diagnostik (Z5)
- Anstrengung bei ILP (Z6)
- Pausenzeit (Z7)
- Selbsteinschätzung (Z8)

Der genaue Wortlaut der Fragen und die dazugehörenden Antwortmöglichkeiten werden im Anhang dargestellt.

Mit den Zusatzfragen wurden mehrere Ziele verfolgt. Es sollten zum einen Variablen erfasst werden die für die Lehrveranstaltungsleitung von Interesse sind, wie die Prüfungsvorbereitung, der geplante Prüfungstermin oder das Interesse an Psychologischer Diagnostik. Zum anderen sollten Variablen erhoben werden, die sich auf das Ergebnis der ILP auswirken können wie das Bemühen beziehungsweise die Anstrengung, die Pausenzeit oder der Vorlesungsbesuch.

Das Bemühen schien deshalb relevant, weil ein großer Teil der Studierenden die ILP als „Pflichtübung“ im Rahmen der „Übungen zum Psychologischen Diagnostizieren“ absolvierten. Dadurch war zu befürchten, dass es eine große Anzahl Studierender geben würde, die die Fragen zwar durcharbeiten, aber nicht versuchen würden diese auch richtig zu beantworten. Das Bemühen wurde deshalb auf zwei Arten erfasst, da die Frage nach der Anstrengung auch in den anderen Erhebungen im Rahmen des PeRaS Projekts erhoben wurde. Für etwaige weiterleitende Studien sollte hier auf Vergleichbarkeit geachtet

werden. Da die im Projekt verwendete Frage (Z6) für die ILP nach Einschätzung des Autors wenig geeignet ist, wurde das Bemühen auch auf eine zweite Art (Frage Z3) erfasst.

Die Pausenzeit wurde erfasst, da es keine Zeitvorgaben für die Bearbeitung der ILP gab und es möglich war Pausen zu machen. Um trotzdem eine ungefähre Bearbeitungszeit schätzen zu können wurde es nötig die Länge der durchgeführten Pausen zu kennen. Die so errechnete Bearbeitungszeit wurde auch als Teilungskriterium in den Analysen nach dem Rasch-Modell herangezogen (vgl. Kapitel 5.1.1.4).

Das in Frage Z5 erhobene Interesse an psychologischer Diagnostik fand ebenfalls Verwendung als Teilungskriterium in den Analysen nach dem Rasch-Modell (vgl. Kapitel 5.1.1.5).

Zur Beantwortung der Fragestellungen 3 und 4 (vgl. Kapitel 3) wurde auch die Selbsteinschätzung der erbrachten Leistung erhoben. Die diesbezüglichen Ergebnisse finden sich in Kapitel 5.4.

4.2.3 Fragebogen zur Prüfung

Im Anschluss an die Prüfung zur Psychologischen Diagnostik 1, wurde ein Fragebogen, bestehend aus 6 Items vorgegeben. Die Fragen hatten folgende Inhalte:

- Selbsteinschätzung der erbrachten Leistung (FB1)
- Nummer des aktuellen Prüfungsantritts (FB2)
- Abschluss der Prüfungsvorbereitungen (FB3)
- Teilnahme an der ILP (FB4)
- Vorlesungsbesuch (FB5)
- Benutzung des E-learning Angebotes (FB6)

Der genaue Wortlaut der Fragen und die dazugehörenden Antwortmöglichkeiten werden im Anhang dargestellt.

Mit diesem Fragebogen wurden mehrere Ziele verfolgt. Erstens sollte die Variable Vorlesungsbesuch die bereits im Rahmen der ILP erhoben wurden, auch nach der Prüfung erhoben werden, da ja nicht alle Teilnehmer der Prüfung auch die ILP samt Zusatzfragen bearbeitet hatten. Zweitens sollten mögliche Prädiktoren wie E-learning-Nutzung oder die

Teilnahme an der ILP nochmals erhoben werden, für den Fall dass die Erhebung dieser Variablen über Teilnehmerlisten nicht möglich ist.

Für die späteren Berechnungen hatte vor allem die Frage zum Vorlesungsbesuch Bedeutung. Sie wurde als Prädiktor im Regressionsmodell verwendet (vgl. Kapitel 5)

Zur Beantwortung der Fragestellungen 3 und 4 (vgl. Kapitel 3) wurde auch die Selbsteinschätzung der erbrachten Leistung erhoben. Die diesbezüglichen Ergebnisse finden sich in Kapitel 5.4.

4.2.4 Verbale Intelligenzfunktionen (IBF)

Verwendet wurde die Version 22.00 (November 2005) des Tests Intelligenz-Basis-Funktionen (IBF, Blum et al., 2005). Die Vorgabe erfolgte im Rahmen der Diplomarbeitsstudie von Stadler (2009), als Teil des PeRaS-Projekts (Wagner-Menghin, 2009). Vorgegeben wurden die Untertests „Verbale Intelligenzfunktionen“ und „Numerische Intelligenzfunktionen“. Für diese Arbeit wurden nur die Ergebnisse des Untertests „Verbale Intelligenzfunktionen“ herangezogen.

Die Erhebung der verbalen Intelligenz im Untertest „Verbale Intelligenzfunktionen“ erfolgte durch zwei Itemtypen. Beim ersten Typ handelt es sich um Satzergänzungen, bei denen aus 5 Begriffen der gewählt werden sollte, welcher den Satz richtig ergänzt (vgl. Abb. 2). Beim zweiten Typ handelt es sich um verbale Analogieaufgaben. Dabei werden drei Begriffe präsentiert wobei die ersten beiden in einer semantischen Relation zueinander stehen. Der Proband muss den vierten Begriff aus einer Auswahl von fünf Begriffen so wählen, dass die Relation der ersten beiden Begriffe auch auf die Begriffe drei und vier zutrifft. Ein Beispielitem ist in Abbildung 3 zu sehen.

Anleitung...

Ein zweites Beispiel:

Zu einem Gemälde gehört(e) immer

- ☐ ein Rahmen
- ☐ eine Leinwand
- ☐ ein Maler
- ☐ ein Preis
- ☐ ein Museum

Dieser Satz wird richtig ergänzt durch „ein Maler“. Wählen Sie daher „ein Maler“ aus.

☐ Zurück
 Weiter ☐

Abbildung 2 Beispielaufgabe der Verbalen Intelligenzfunktionen - Satzergänzungen (Blum et al., 2005)

Aufgabe 1
 Zeit:

☐ 1
 ☐ 2
 ☐ 3
 ☐ 4
 ☐ 5
 ☐ 6
 ☐ 7
 ☐ 8
 ☐ 9
 ☐ 10
 ☐ 11

Wort : Satz = Teller : ?

- ☐ Löffel
- ☐ Speise
- ☐ Geschirr
- ☐ Hunger
- ☐ Schüssel

Weiter ☐

Abbildung 3 Beispielaufgabe der Verbalen Intelligenzfunktionen – Analogien (Blum et al., 2005)

Für die aktuelle Untersuchung wurde der Untertest „Verbale Intelligenzfunktionen“ nicht als Computerversion mittels Wiener Testsystem vorgegeben, sondern in einer adaptierten Papierform. Dabei wurden alle 35 Items aus der Standardform vorgegeben.

4.3 Durchführung der Untersuchung

Die Durchführung der Untersuchung erfolgte in drei Phasen. Die erste Phase fand im Rahmen der Diplomarbeitsstudie von Stadler (2009) statt. Neben anderen Verfahren, wurde der Untertest verbale Intelligenz des IBF vorgegeben (vgl. Kapitel 4.3.1). In der zweiten Phase wurde den Studierenden die ILP zum Bearbeiten zur Verfügung gestellt (vgl. Kapitel 4.3.2). Die dritte Phase stellte die Fragebogenerhebung im Anschluss an die Prüfung zur Psychologischen Diagnostik dar (vgl. Kapitel 4.3.3).

4.3.1 Vorgabe des Untertest Verbalen Intelligenzfunktionen (IBF)

Der Untertest Verbale Intelligenzfunktionen wurde den Studierenden als Teil einer Testbatterie im Rahmen der Diplomarbeitsstudie von Stadler (2009) vorgegeben. Die Testbatterie bestand aus den Untertests Verbale Intelligenzfunktionen und Numerische Intelligenzfunktionen, beide aus dem Verfahren Intelligenz-Basis-Funktionen (IBF, Blum et al., 2005), LAsO (Fill-Giordano & Litzenberger, 2005), LEWITE (Wagner-Menghin, 1998), BFI-S (Rammstedt, 1997) und AHA (Kubinger & Ebenhöf, 1996). Die Vorgabe erfolgte während den Lehrveranstaltungen „Übungen zum psychologischen Diagnostizieren 1“, wovon es im Sommersemester 2008 6 Parallelveranstaltungen gab. Die Erhebung erfolgte im Zeitraum vom 5.5.2008 bis zum 15.5.2008. Wie bereits in Kapitel 4.2.4 erwähnt, wurden die beiden Untertests des IBF nicht via Computer sondern in Papierform vorgegeben.

4.3.2 Vorgabe der ILP

Die ILP wurde über die Lernplattform Blackboard Vista vorgegeben. Zugriff zu der Lernplattform haben prinzipiell alle Studierenden der Universität Wien. Die ILP selbst wurde in dem Kurs zur Vorlesung Psychologische Diagnostik bereitgestellt. Die Studierenden konnten sich via Internet selbstständig bei dem Kurs zur Lehrveranstaltung anmelden. Nur angemeldete Studierende konnten an der ILP teilnehmen. Zur Teilnahme wurden die Besucher der Vorlesung, während der Lehrveranstaltung, von der

Lehrveranstaltungsleitung eingeladen. Auch per Email⁵ wurden Einladungen an alle Übungsteilnehmer der Übungen zum psychologischen Diagnostizieren im Sommersemester 2008 verschickt. Der Wortlaut des Emails ist im Anhang zu finden.

Beim Bearbeiten der ILP mussten die Items nicht sofort bearbeitet werden, es konnten Fragen also übersprungen und später bearbeitet werden, wie bei der tatsächlichen Prüfung auch. Um sicher zu gehen dass keine Items vergessen wurden, gab das System Studenten die die ILP abgeben (=beenden) wollten, eine Übersicht über alle Items und ob diese beantwortet wurden.

Sollte bei der Bearbeitung das Browserfenster aus irgendwelchen Gründen geschlossen werden, stellte das System sicher, dass mit der zuletzt beantworteten Frage wieder fortgefahren werden konnte.

Da die Studenten die ILP nicht unter Anleitung oder Aufsicht ausfüllten, gab es, im Gegensatz zur Prüfung selbst, kein Zeitlimit. Damit sollte verhindert werden, dass diverse Ablenkungen, wie beispielsweise Telefonanrufe, nicht die Bearbeitungszeit der Studierenden verkürzen und somit zu unterschiedlichen Voraussetzungen führen hätten können.

Da ein Teil der Studierenden die ILP als Teil der Pflichtveranstaltung „Übungen zum psychologischen Diagnostizieren 1“ absolvierten und sich dafür 30 Minuten Testerfahrung anrechnen lassen konnten, bestand die Gefahr, dass einige Studierende die Items gar nicht „ernsthaft“ ausfüllen würden. Um mangelnde inhaltliche Auseinandersetzung mit den Items zu verhindern wurde eine Mindestbearbeitungszeit von 10 Minuten vorgegeben. Die Auswirkungen dieser Maßnahme, und mögliche alternative Vorgehensweisen werden in Kapitel 6 diskutiert.

Die ILP inklusive den Zusatzfragen (Kapitel 4.2.2) konnte vom 02.06.2008 bis zum 13.06.2008 ausgefüllt werden. Zu Beginn dieses Zeitfensters wurde nochmals per Nachricht über das Nachrichtensystems der Lernplattform auf die ILP hingewiesen. Eine weitere Hinweismeldung wurde am 10.06.2008 als Erinnerung versandt. Nach Ablauf des

⁵ versandt am 31.05.2008

Verfügbarkeitszeitraums konnten die Studierenden ihre Ergebnisse und auch eine, mit der Lehrveranstaltungsleitung erarbeitete, Rückmeldungen⁶ zu den einzelnen Items einsehen.

4.3.3 Vorgabe des Zusatzfragebogens

Der Zusatzfragebogen wurde direkt im Anschluss an die Prüfung zur psychologischen Diagnostik 1 den Studierenden Vorgegeben. Es handelte sich um den 1. Prüfungstermin (27.6.2008) der Vorlesung psychologische Diagnostik 1 des Sommersemesters 2008 (SS2008). Da an diesem Termin zwei Prüfungen in einem Hörsaal stattfanden, mussten die Studierenden der Prüfung zur psychologischen Diagnostik 1 nach Ablauf der Prüfungszeit, aus organisatorischen Gründen, noch 10 Minuten an ihrem Platz sitzen bleiben und durften den Hörsaal nicht verlassen. Der Zusatzfragebogen wurde nun gleichzeitig mit dem Einsammeln der Prüfungsbögen verteilt. Studierende die vor der maximalen Prüfungszeit fertig waren bekamen ihren Fragebogen bei der Abgabe ihres Prüfungsbogens und wurden gebeten die Fragen noch vor Verlassen des Hörsaals auszufüllen. Durch dieses Vorgehen wurde verhindert, dass sich Studierende durch die Befragung bei der Bearbeitung der Prüfung beeinträchtigt fühlten. Durch die 10 Minuten, die die Studierenden nach der Prüfung noch anwesend sein mussten, kam es zu einer sehr hohen Rücklaufquote von 91,5% (vgl. Kapitel 4.4).

4.4 Stichprobe

Die Stichprobe bestand aus drei Teilstichproben, die alle aus der Population von Studierenden der Psychologie an der Universität Wien gezogen wurden. Die erste Teilstichprobe bestand aus Studierenden die an der ILP teilgenommen haben. Diese teilten sich in zwei Gruppen, nämlich einerseits Studierende die im Sommersemester 2008 die Lehrveranstaltung „Übungen zum psychologischen Diagnostizieren 1“ besuchten (n=163; 140 weiblich, 23 männlich), andererseits Personen die die Vorlesung psychologische Diagnostik 1 besuchten oder zumindest planten die Prüfung zur Vorlesung zu Semesterende zu absolvieren (n=33; 29 weiblich, 4 männlich). Die erste Teilstichprobe bestand also insgesamt aus 199 Studierenden (172 weibliche, 27 männliche).

Die zweite Teilstichprobe bestand aus Studierenden die im Rahmen der Erhebung von Stadler (2009) den Untertest „Verbale Intelligenzfunktionen“ des IBF (Blum et al., 2005)

⁶ Sowohl die beiden Nachrichten, als auch die Rückmeldungen finden sich in Anhang

ausgefüllt haben. Diese Teilstichprobe bestand aus insgesamt 196 Studierenden (165 weibliche, 31 männliche), die im Sommersemester 2008 die Lehrveranstaltung „Übungen zum psychologischen Diagnostizieren 1“ besuchten.

Die dritte Teilstichprobe bestand schließlich aus Studierenden die die Prüfung zur psychologischen Diagnostik zum ersten Prüfungstermin am 27.6.2008 absolviert haben. Diese Teilstichprobe bestand aus insgesamt 118 Studierenden (16 weibliche, 102 männliche).

Für die späteren Regressionsanalysen waren insbesondere die Schnittmengen zwischen den einzelnen Teilstichproben relevant. Vertreten in allen 3 Teilstichproben waren nur 66 Studierende (59 weiblich, 7 männlich). Von diesen Personen waren also die Prüfungsergebnisse, die Ergebnisse der ILP und die Daten zur verbalen Intelligenz vorhanden.

4.5 Eingesetzte Statistische Verfahren

Neben diversen deskriptiven Verfahren wurden das Rasch-Modell (vgl. Kapitel 4.5.1), ein Modell der probabilistischen Testtheorie, und die logistische Regression (vgl. Kapitel 4.5.2), eine multivariate Analyseverfahren eingesetzt. Im Folgenden werden die für die aktuelle Untersuchung wesentlichen Aspekte der beiden Modelle vorgestellt und weiterführende Literatur angeführt.

4.5.1 Das Rasch-Modell

In der Literatur wird das Rasch-Modell auch als dichotomes logistisches Testmodell von Rasch“ (1 PL Modell) bezeichnet (Kubinger, 1989). Es ist anwendbar für Tests mit dichotomen Items.

4.5.1.1 Modellbeschreibung

Kubinger (2006) beschreibt das Rasch-Modell folgendermaßen: „Das Rasch-Modell beschreibt die Wahrscheinlichkeit, dass Person v Item i löst („+“), in Abhängigkeit des Personenparameters ξ_v , das ist die (wahre) Fähigkeit von v , und des Itemparameters σ_i , das ist die (wahre) Schwierigkeit von i “. Die Formel dazu findet sich in Abbildung 4.

$$P(+|\xi_v, \sigma_i) = \frac{e^{\xi_v - \sigma_i}}{1 + e^{\xi_v - \sigma_i}}$$

Abbildung 4 Modellgleichung des Rasch-Modells

Das Rasch-Modell weist folgende Charakteristika auf:

- Eindimensionalität
- Lokale stochastische Unabhängigkeit
- Spezifische Objektivität
- Erschöpfende Statistiken

Für diese Untersuchung sind vor allem die „Erschöpfenden Statistiken“ und die Eindimensionalität des Rasch-Modells entscheidend. Die Bezeichnung „Erschöpfende Statistiken“ bedeutet, dass bei Geltung des Rasch-Modells der Summenscore, also die ungewichtete Summe der gelösten Items, alles über das Antwortverhalten der Person aussagt. Eindimensionalität bedeutet, dass bei Geltung des Rasch-Modells alle Items dieselbe latente Fähigkeit messen (Rost, 2004).

Ausführliche Erläuterungen zum Rasch-Modell finden sich bei Fischer (1974a, 1988 und 1995a), Kubinger (1989), Rost (2004) und Bond & Fox (2007).

4.5.1.2 Überprüfung der Modellgeltung

Um die Rasch-Modell-Passung der Daten zu überprüfen wurde der Likelihood-Quotienten-Test (LQT) nach Andersen (1973) angewandt. Die Passung der einzelnen Items wurde mittels des Grafischen Modelltests (Rasch) überprüft. Die *Itemfit*-Werte (*Infit*- und *Outfit*-Werte) wurden ebenfalls zur Überprüfung der Modellpassung der einzelnen Items herangezogen (vgl. Bond, Fox; 2007)

Der Likelihood-Quotienten-Test (LQT) nach Andersen (1973) prüft die Model-Passung der zu analysierenden Daten. Dazu werden die Daten in zwei Personengruppen, anhand eines Teilungskriteriums aufgeteilt (die in der aktuellen Studie verwendeten Teilungskriterien finden sich in Kapitel 5.1) und die Likelihood der beiden Gruppen berechnet. Als Prüfgröße wird daraus ein χ^2 -Wert berechnet welcher mit einem kritischen χ^2 -Wert verglichen werden kann. Dabei gilt: ist der empirische χ^2 -Wert kleiner als der kritische χ^2 -Wert, dann wird die H_0 : „Das Rasch-Modell gilt“ beibehalten (Poinstingl, Mair & Hatzinger, 2007).

Da der LQT lediglich die Passung der Gesamtdaten zum Rasch-Modell überprüft, wird zur Betrachtung der einzelnen Items ein weiteres Verfahren benötigt. In der aktuellen Studie kommen dazu der Grafische Modelltest nach Rasch sowie die *Itemfit*-Werte zum Einsatz.

Der Grafische Modelltest (Rasch) basiert auf den Daten des LQT. Dabei wird der Datensatz analog dem LQT in zwei Gruppen geteilt. In einem Koordinatensystem werden auf der x-Achse die geschätzten Itemschwierigkeiten in Gruppe 1 und auf der y-Achse die geschätzten Itemschwierigkeiten in Gruppe 2 aufgetragen. Die Items selbst haben nun Koordinaten, die den geschätzten Itemschwierigkeiten in den beiden Gruppen entsprechen. Gemäß der Forderung nach „spezifisch objektiven Vergleichen“ im Rasch-Modell, ist bei Modell-Geltung zu erwarten, dass die Items in beiden Gruppen die gleiche Schwierigkeit aufweisen. Für die Grafik bedeutet das, dass die Items möglichst nahe an der 45° Geraden liegen, also die x- und y-Koordinaten möglichst ident sind.

Sobald Items von dieser Geraden abweichen, bedeutet dies, dass sie in einer der beiden Gruppen eine andere relative Schwierigkeit aufweisen als in der anderen Gruppe. Dies

widerspricht den Modellannahmen der spezifischen Objektivität sowie der Eindimensionalität.

Als zusätzliche Überprüfung der Item-Passung werden die *Itemfit-Werte* betrachtet. Wright & Masters (1982) beschreiben zwei *Itemfit-Werte*, den Outfit und den Infit. Outfit stellt die Summe der Quadrate der standardisierten Residuen dar. Als Residuen werden die Differenzen zwischen beobachtetem und erwartetem Wert einer Person j bei Item g bezeichnet. Für den Outfit Wert werden also für jedes Item die Quadrate der Residuen der einzelnen Personen summiert und der Mittelwert gebildet. Infit stellt eine gewichtete Summe der Quadrate dar. Hier werden die Residuen noch mit der Modell-Varianz gewichtet (Bond & Fox, 2007).

Als passenden Wertebereich für die Infit- und Outfit-Werte nennen Bond & Fox (2007) für Tests mit Multiple-Choice Antwortformat den Intervall $[0,8; 1,2]$. Liegen die Werte im Bereich $<0,8$, so spricht man von Overfit, hier liegt zu wenig Varianz vor. Overfit hat keine praktischen Auswirkungen, sondern führt lediglich zu kleineren Standardfehlern und einer überhöhten Reliabilität. Liegen die Werte im Bereich $>1,2$, so spricht man von Underfit, hier liegt zu viel Varianz vor. Underfit führt zu einer verringerten Qualität der Messung. Liegt nur ein Wert (Infit- oder Outfit-Wert) außerhalb des Intervalls, spricht man von Mixed Fit (Bond & Fox, 2007).

4.5.2 Die logistische Regression

Die logistische Regression ist eine multivariate Analysemethode zur Untersuchung diskreter abhängiger Variablen. In der aktuellen Studie wurde die binär logistische Regression eingesetzt, zur Analyse von binären abhängigen Variablen. Diese ermöglicht die Ableitung der Eintrittswahrscheinlichkeit eines Komplementärereignisses in Abhängigkeit von beeinflussenden unabhängigen Variablen, den sogenannten Regressoren. Diese können dabei sowohl kategoriales als auch metrisches Skalenniveau aufweisen (Backhaus, Erichson, Plinke & Weiber, 2006).

Die Modelpassung wird mittels sogenannten Pseudo-R-Quadrat-Statistiken geprüft. Diese versuchen den Anteil der durch das Regressionsmodell erklärten Varianz zu quantifizieren. Hier werden zu jedem Modell das „Cox & Snell R^2 “ sowie das „Nagelkerke R^2 “ betrachtet. Bei diesen Statistiken handelt es sich um Likelihood-Quotienten bei denen die Likelihood des Nullmodells mit der des vollständigen Modells verglichen wird, die Werte liegen dabei

zwischen 0 und 1. Da im Gegensatz zum „Cox & Snell R^2 “ das „Nagelkerke R^2 “ auch den Maximalwert 1 annehmen kann, ist diesem Kriterium bei der Beurteilung der Güte des Modells der Vorzug zu geben (Backhaus et al., 2006).

Zur Beurteilung der Klassifikationsfähigkeit der logistischen Regression empfehlen Backhaus et al. (2006) die Betrachtung der richtigen Zuordnungen aufgrund der Regressionsfunktion. Diese wird mit der zufälligen Zuordnung verglichen. Der Vergleich erfolgt mit zwei Arten von Zufallswahrscheinlichkeiten. Zum einen mit der maximalen Zufallswahrscheinlichkeit, diese entspricht dem Anteil der größten beobachteten Gruppe der Stichprobe. Des Weiteren wird die proportionale Zufallswahrscheinlichkeit betrachtet. Die Formel für diese lautet: $a^2 + (1-a)^2$, wobei a dem Anteil einer der beiden beobachteten Gruppen entspricht (Backhaus et al., 2006).

5 Ergebnisse

Im ersten Abschnitt dieses Kapitels (5.1) werden die Ergebnisse der Analysen nach dem Rasch-Modell dargestellt. Im zweiten Abschnitt dieses Kapitels (5.2) werden die einzelnen Prädiktoren dargestellt und in weiterer Folge die Ergebnisse der Logistischen Regression präsentiert (Kapitel 5.3). Abschließend werden die Hypothesen beantwortet (Kapitel 5.4).

Sämtliche statistischen Auswertungen und Berechnungen wurden mittels der Statistik-Software SPSS 16.0 für Windows⁷ durchgeführt. Für die Analysen nach dem Rasch-Modell wurde die Statistik-Software „R: Version 2.8.0“⁸ verwendet mit dem Programmpaket „eRm: extended Rasch modeling-R package version 0.10.0“⁹.

5.1 Analysen nach dem Rasch-Modell

Um die Rasch-Modell-Passung der Daten zu testen wurde der Likelihood-Quotienten-Test (LQT) nach Andersen (1973) angewandt. Die Passung der einzelnen Items wurde mittels des Grafischen Modelltests (Rasch) überprüft. Die *Itemfit*-Werte (*Infit*- und *Outfit*-Werte) wurden ebenfalls zur Überprüfung der Modellpassung der einzelnen Items herangezogen (Bond, Fox; 2007)

Für die Analysen nach dem Rasch-Modell wurden folgende Teilungskriterien verwendet: Score (Median-Split), Geschlecht, Bearbeitungszeit (Median-Split), „Interesse an dem Fach „Psychologische Diagnostik“ (Median-Split), Vorlesungsbesuch (Median-Split) und Übungsbesuch.

Bei folgenden Teilungskriterien fehlten zu einzelnen Studierenden Informationen: „Interesse an dem Fach „Psychologische Diagnostik“ (1 Person) und Vorlesungsbesuch (2 Personen). Diese Studierenden wurden zufällig auf die beiden Merkmalsausprägungen zugeteilt.

⁷ Release 16.0.1; Nov. 15, 2007

⁸ Okt. 20, 2008

⁹ Juni 28, 2008

5.1.1 Analysen nach dem Rasch-Modell für den Gesamt-Itempool

Im ersten Schritt wurden alle 22 Items einer gemeinsamen Analyse unterzogen (n=199). Bei *Rasch*-Modell-Konformität würde dies für eine, von den Lernzielen „Wissen“, „Verständnis und Anwendung“ und „Analyse und Beurteilung“ unabhängige „psychologisch-diagnostische Gesamtfähigkeit“ sprechen.

Bereits für den Gesamt-Itempool ergab der LQT für die Teilungskriterien Score, Geschlecht, Interesse, Vorlesungsbesuch und Übungsbesuch, bei einem einheitlich festgesetzten Risiko 1. Art von $\alpha=0,01$, keine signifikanten Ergebnisse. Für das Teilungskriterium Bearbeitungszeit ergab der LQT, bei einem einheitlich festgesetzten Risiko 1. Art von $\alpha=0,01$, jedoch ein signifikantes Ergebnis (vgl. Tabelle 8). Nach Ausschluss von 2 Items (A3, C1) ergab der LQT auch für das Teilungskriterium Bearbeitungszeit kein signifikantes Ergebnis mehr. Die Entscheidung diese beiden Items auszuschließen erfolgte nach Betrachtung der *Itemfit*-Werte einerseits und des Grafischen Modelltests für die einzelnen Teilungskriterien andererseits. Die Darstellung der *Itemfit*-Werte erfolgt in Kapitel 5.1.1.1. Eine Übersicht in welchen Teilungskriterien die ausgeschlossenen Items im Grafischen Modelltest deutlich von der 45° Geraden abweichen gibt Tabelle 3. Detaillierte Ausführungen zu den einzelnen Teilungskriterien sind den entsprechenden Unterkapiteln zu entnehmen.

Tabelle 3 Ausgeschiedene Items in Bezug auf die jeweiligen Teilungskriterien, in denen eine zu große Abweichung von der 45° Geraden vorliegt

Item	Score	Geschlecht	Teilungskriterien			
			Bearbeitungszeit	Interesse	VO-Besuch	UE-Besuch
A3	x		x			x
C1	x		x			

5.1.1.1 Darstellung der *Itemfit*-Werte

Für die Items eines Multiple-Choice Wissenstest sollten laut Bond und Fox (2007) die *Infit*- und *Outfit*-Werte im Intervall $[0,8; 1,2]$ liegen. Wie in Tabelle 4 zu sehen, liegen beide Werte des Items A3 über diesem Intervall, hier liegt also „underfit“ vor. Die *Infit*-Werte der Items A1, B1b (0,774) und C1 liegen unter dem Intervall, hier liegt „overfit“ vor. Die *Outfit*-Werte der Items A1, B1b und C1 liegen im gewünschten Intervall. Erläuterungen zu den Begriffen „underfit“ und „overfit“ finden sich in Kapitel 2.2.1.

Nach Ausschluss der Items A3 und C1 liegen die *Infit*-Werte der Items B1b (0,784) und F3 (0,791) unter dem Intervall [0,8; 1,2]. Die *Outfit*-Werte aller Items liegen im Intervall [0,8; 1,2]. Bei den *Infit*-Werten dieser Items liegt als „overfit“ vor. Da die wie in Kapitel 2.2.1 erläutert, keine praktische Auswirkung hat, können die beiden Items im Itempool belassen werden. Die Werte nach Ausschluss der Items sind in Tabelle 5 zu sehen.

Tabelle 4 Infit- und Outfit-Werte vor Ausschluss der Items

Item	Itemfit-Werte		Fit Typ
	Infit	Outfit	
	Quadratisches Mittel	Quadratisches Mittel	
A1	0,797	0,941	Mixed
A2	1,086	1,053	Good Fit
A3	1,289	1,223	Underfit
B1a	1,072	1,021	Good Fit
B1b	0,774	0,863	Mixed
B2	1,030	1,047	Good Fit
B3	1,092	1,042	Good Fit
C1	0,776	0,844	Mixed
C2	0,806	0,979	Good Fit
C3	1,001	1,029	Good Fit
D1	0,878	0,931	Good Fit
D2	1,052	1,031	Good Fit
D3	1,186	1,076	Good Fit
E1	0,935	0,936	Good Fit
E2	0,972	0,938	Good Fit
E3	1,028	1,039	Good Fit
F1	0,878	0,915	Good Fit
F2	0,914	0,992	Good Fit
F3	0,815	0,902	Good Fit
H1	0,960	0,967	Good Fit
H2	0,805	0,894	Good Fit
H3	0,912	0,944	Good Fit

Tabelle 5 Infit- und Outfit-Werte nach Ausschluss der Items A3 und C1

Item	Itemfit-Werte		Fit Typ
	Infit	Outfit	
	Quadratisches Mittel	Quadratisches Mittel	
A1	0,808	0,958	Good Fit
A2	1,096	1,071	Good Fit
B1a	1,103	1,038	Good Fit
B1b	0,784	0,870	Mixed
B2	1,037	1,044	Good Fit
B3	1,157	1,065	Good Fit
C2	0,901	0,983	Good Fit
C3	0,978	1,019	Good Fit
D1	0,852	0,931	Good Fit
D2	1,058	1,031	Good Fit
D3	1,167	1,062	Good Fit
E1	0,944	0,923	Good Fit
E2	0,972	0,941	Good Fit
E3	1,008	1,029	Good Fit
F1	0,881	0,920	Good Fit
F2	0,914	0,980	Good Fit
F3	0,791	0,886	Mixed
H1	0,939	0,962	Good Fit
H2	0,868	0,895	Good Fit
H3	0,938	0,955	Good Fit

5.1.1.2 Teilungskriterium Score

Bei dem Teilungskriterium Score handelt es sich um ein internes Kriterium. Die untersuchte Stichprobe (n=199) wurde in die Gruppen „niedriger Score“ und „hoher Score“ aufgeteilt nach dem Median (Md=9) der Scoreverteilung. Dabei wurden Personen mit dem Median Score der Gruppe „niedriger Score“ zugeordnet.

Sowohl vor Ausscheiden der Items als auch danach zeigt der LQT, bei einem einheitlich festgesetzten Risiko 1. Art von $\alpha=0,01$, kein signifikantes Ergebnis (vgl. Tabelle 6). Beim Grafischen Modelltest kommt es bei zwei Items zu großen Abweichungen von der 45° Geraden (vgl. Abbildung 5). Das Item A3 ist für Personen mit hohem Score relativ

schwieriger und für Personen mit niedrigem Score relativ leichter. Das Item C1 ist für Personen mit hohem Score relativ leichter und für Personen mit niedrigem Score relativ schwieriger.

Nach Ausschluss der Items (A3, C1) sprechen sowohl die Ergebnisse des LQT als auch des Grafischen Modelltests (vgl. Abbildung 6) für gute Modellpassung bezogen auf das Teilungskriterium Score.

Tabelle 6 Ergebnisse des LQT für das Teilungskriterium Score

Likelihood-Quotienten-Test vor Ausschluss der Items:	
LR-value:	38,749
Chi-square df:	21
p-value:	0,011
Likelihood- Quotienten -Test nach Ausschluss der Items:	
LR-value:	18,664
Chi-square df:	19
p-value:	0,479

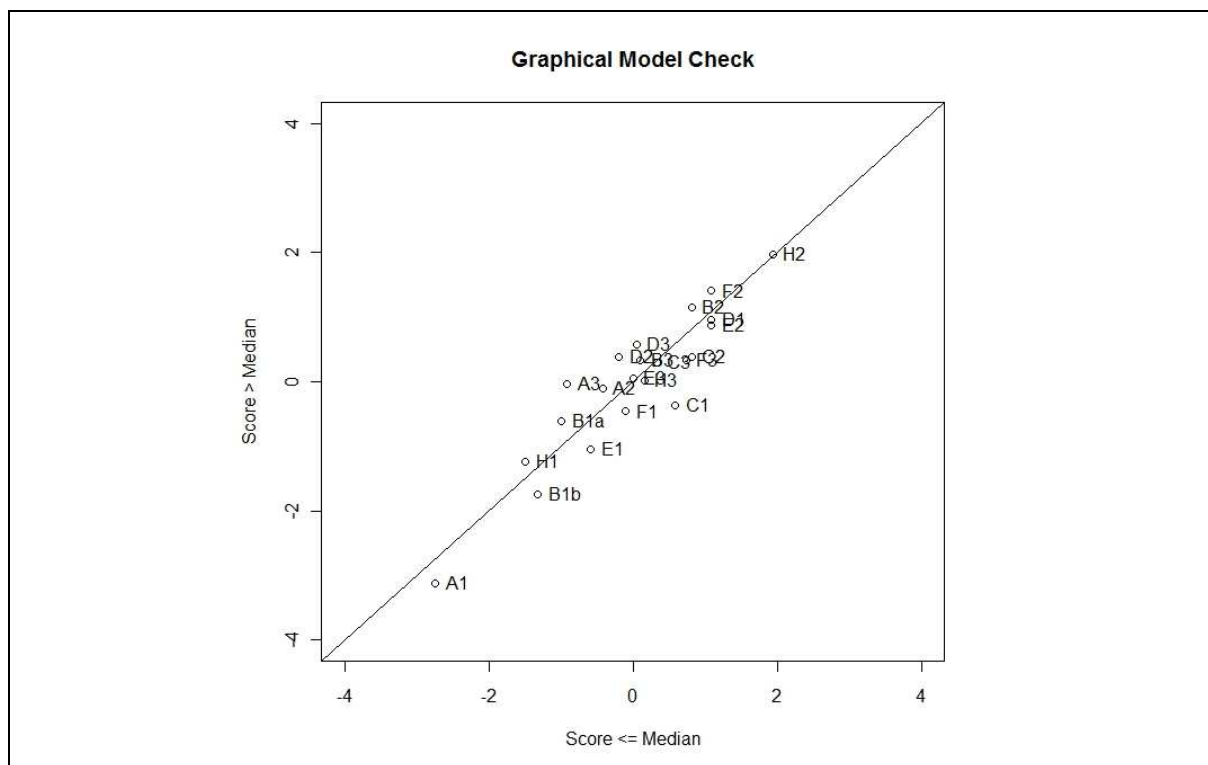


Abbildung 5 Grafischer Modelltest vor dem Ausschluss der Items, Teilungskriterium Score, n=199

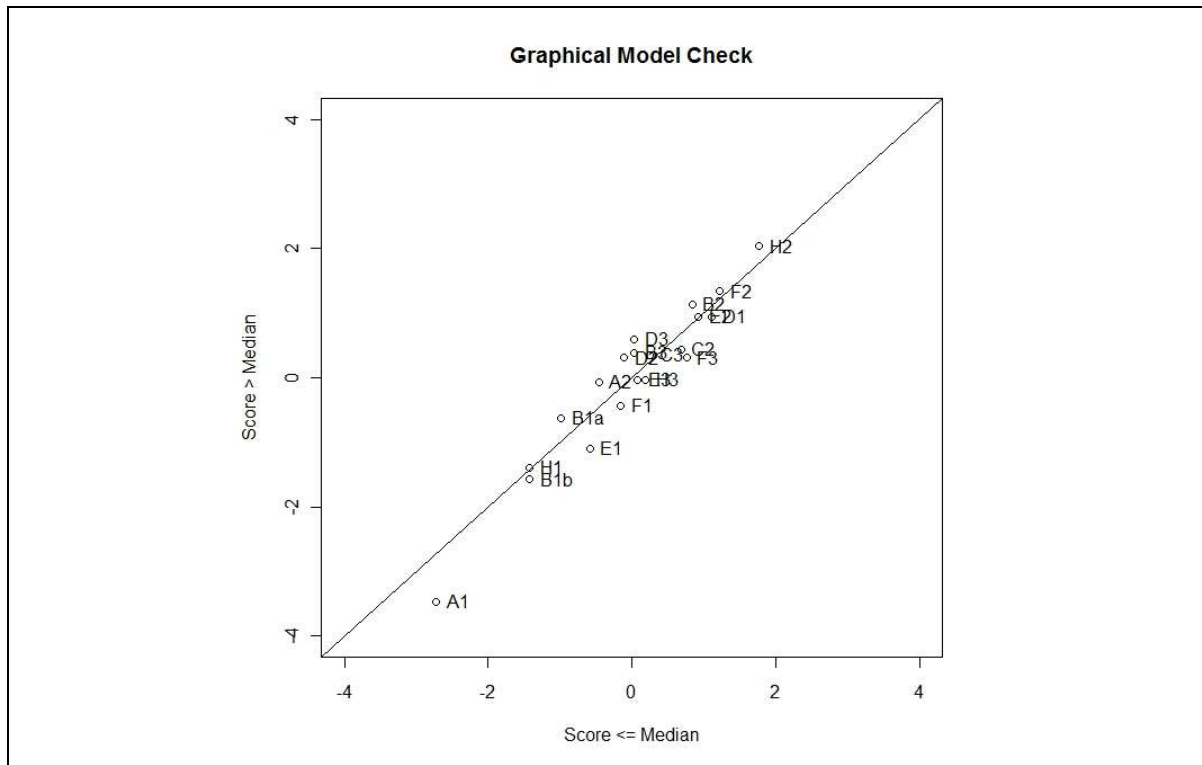


Abbildung 6 Grafischer Modelltest nach dem Ausschluss der Items, Teilungskriterium Score, n=199

5.1.1.3 Teilungskriterium Geschlecht

Bei dem Teilungskriterium Geschlecht handelt es sich um ein externes Kriterium. Die untersuchte Stichprobe (n=199) wurde in die Gruppen weiblich (n=172) und männlich (n=27) aufgeteilt.

Sowohl vor Ausscheiden der Items als auch danach zeigt der LQT, bei einem einheitlich festgesetzten Risiko 1. Art von $\alpha=0,01$, kein signifikantes Ergebnis (vgl. Tabelle 7). Im Grafischen Modelltest zeigt kein Item eine zu große Abweichung von der 45° Gerade, weder vor Ausschluss der Items (vgl. Abbildung 7) noch danach (vgl. Abbildung 8). Somit sprechen sowohl die Ergebnisse des LQT als auch des Grafischen Modelltests für gute Modellpassung bezogen auf das Teilungskriterium Geschlecht.

Tabelle 7 Ergebnisse des LQT für das Teilungskriterium Geschlecht

Likelihood- Quotienten -Test vor Ausschluss der Items:	
LR-value:	15,14
Chi-square df:	21
p-value:	0,816
Likelihood- Quotienten -Test nach Ausschluss der Items:	
LR-value:	14,45
Chi-square df:	19
p-value:	0,757

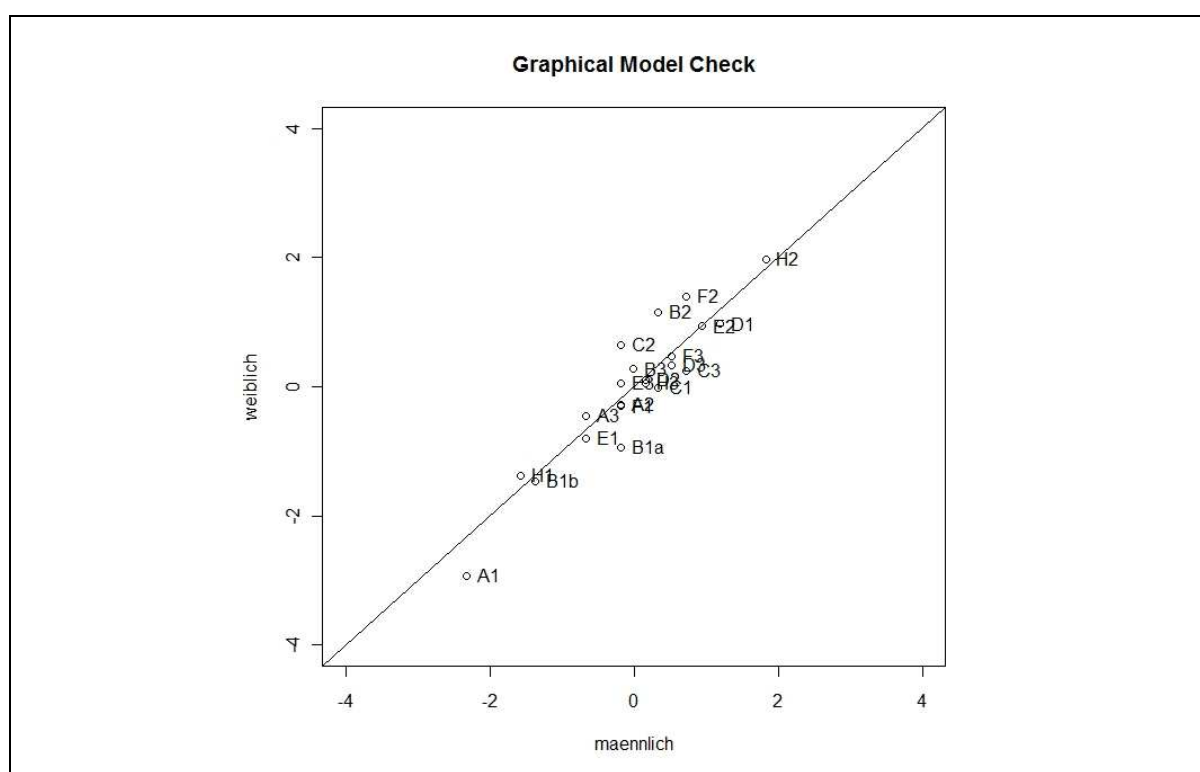


Abbildung 7 Grafischer Modelltest vor dem Ausschluss der Items, Teilungskriterium Geschlecht, n=199

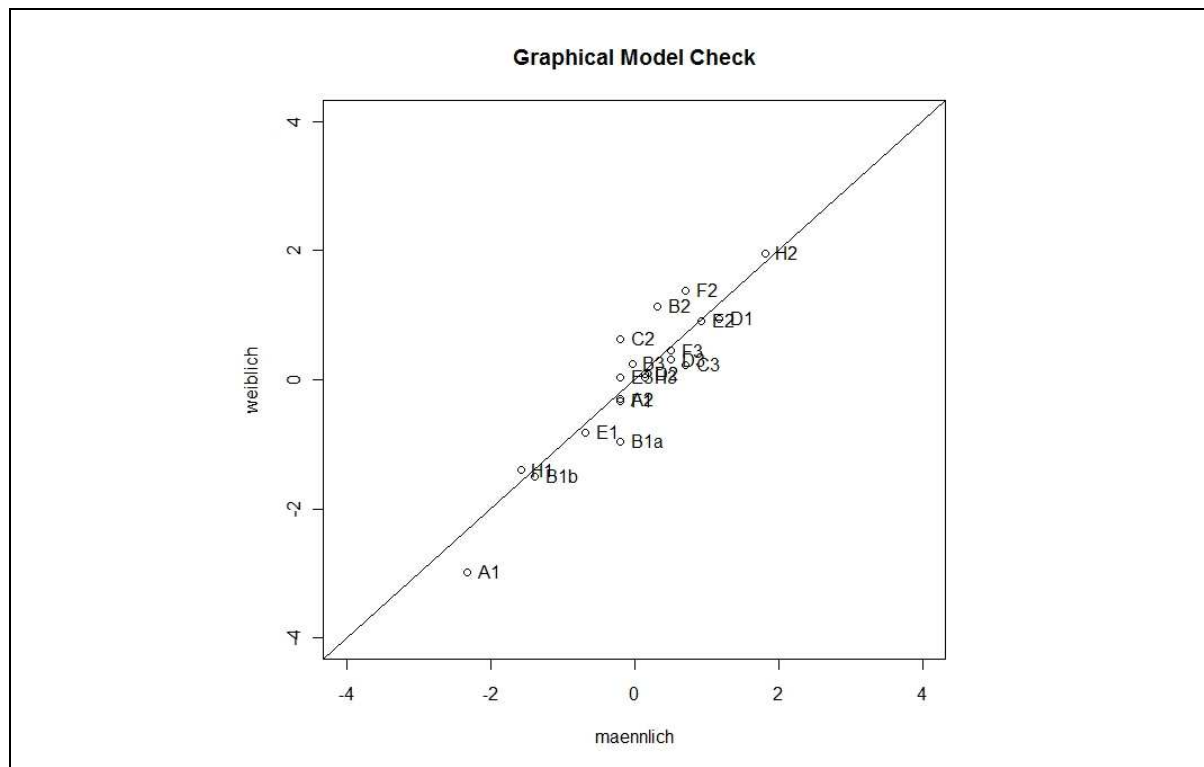


Abbildung 8 Grafischer Modelltest nach dem Ausschluss der Items, Teilungskriterium Geschlecht, n=199

5.1.1.4 Teilungskriterium Bearbeitungszeit

Bei dem Teilungskriterium Bearbeitungszeit handelt es sich um ein internes Kriterium. Die untersuchte Stichprobe (n=199) wurde in die Gruppen „Bearbeitungszeit $\leq$ 44 Minuten“ (n=100) und „Bearbeitungszeit $>$ 44 Minuten“ (n=99) nach dem Median der Bearbeitungszeit der ILP aufgeteilt. Da es den Studierenden bei der Bearbeitung der Items freistand eine Unterbrechung zu machen, wurde die gemessene Bearbeitungszeit um die Pausenzeit korrigiert. Die Pausenzeit wurde mittels Zusatzfrage nach Bearbeitung der Items erhoben (vgl. Kapitel 4.2.2).

Vor Ausscheiden der Items zeigt der LQT, bei einem einheitlich festgesetzten Risiko 1. Art von $\alpha=.01$, ein signifikantes Ergebnis (vgl. Tabelle 8). Erst nach dem Ausscheiden der Items A3 und C1 wird das Ergebnis des LQT nicht mehr signifikant. Beim Grafischen Modelltest kommt es bei 4 Items (A3, B1b, C1) zu deutlichen Abweichungen von der 45° Geraden, sowohl vor Ausschluss der Items (A3, C1; vgl. Abbildung 9) als auch danach (vgl. Abbildung 10). Das Item A3 ist für Personen der Gruppe „Bearbeitungszeit $>$ 44 Minuten“ relativ schwieriger und für Personen der Gruppe „Bearbeitungszeit $\leq$ 44 Minuten“ relativ leichter. Das Item C1 ist für Personen der Gruppe „Bearbeitungszeit $\leq$ 44 Minuten“ relativ schwieriger und für Personen der Gruppe „Bearbeitungszeit $>$ 44

Minuten“ relativ leichter. Das Item B1b ist für Personen der Gruppe „Bearbeitungszeit $\leq$ 44 Minuten“ relativ schwieriger und für Personen der Gruppe „Bearbeitungszeit $>$ 44 Minuten“ relativ leichter.

Nach Ausschluss der Items (A3, C1) sprechen die Ergebnisse des LQT für eine akzeptable Modellpassung bezogen auf das Teilungskriterium Vorlesungsbesuch. Bei dem Item B1b kommt es im Grafischen Modelltest zu einer deutlichen Abweichung von der 45° (vgl. Abbildung 10). Dieses Item ist für Personen der Gruppe „Bearbeitungszeit $\leq$ 44 Minuten“ relativ schwieriger und für Personen der Gruppe „Bearbeitungszeit $>$ 44 Minuten“ relativ leichter. Die Ergebnisse des Grafischen Modelltests sind trotz dieser Abweichung zufriedenstellend.

Tabelle 8 Ergebnisse des LQT für das Teilungskriterium Bearbeitungszeit

Likelihood- Quotienten -Test vor Ausschluss der Items:	
LR-value:	45,803
Chi-square df:	21
p-value:	$<0,01$
Likelihood- Quotienten -Test nach Ausschluss der Items:	
LR-value:	37,715
Chi-square df:	19
p-value:	0,011

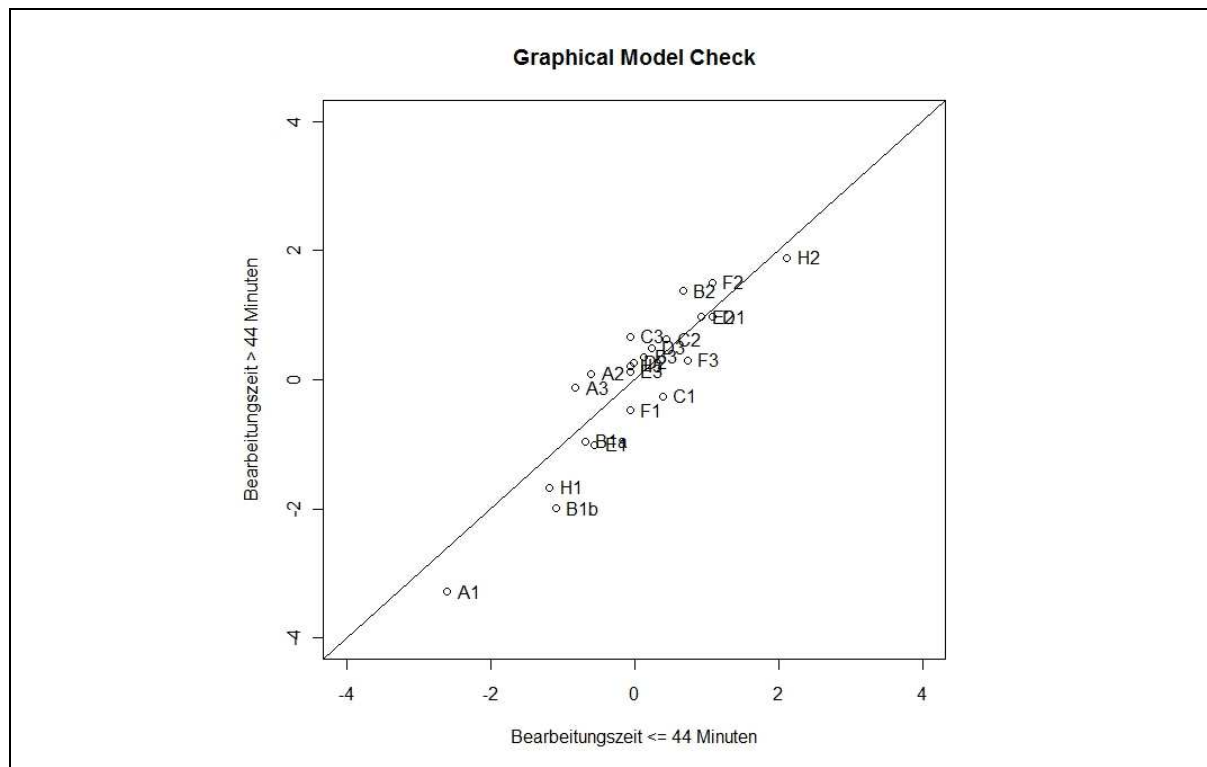


Abbildung 9 Grafischer Modelltest vor dem Ausschluss der Items,
Teilungskriterium Bearbeitungszeit, n=199

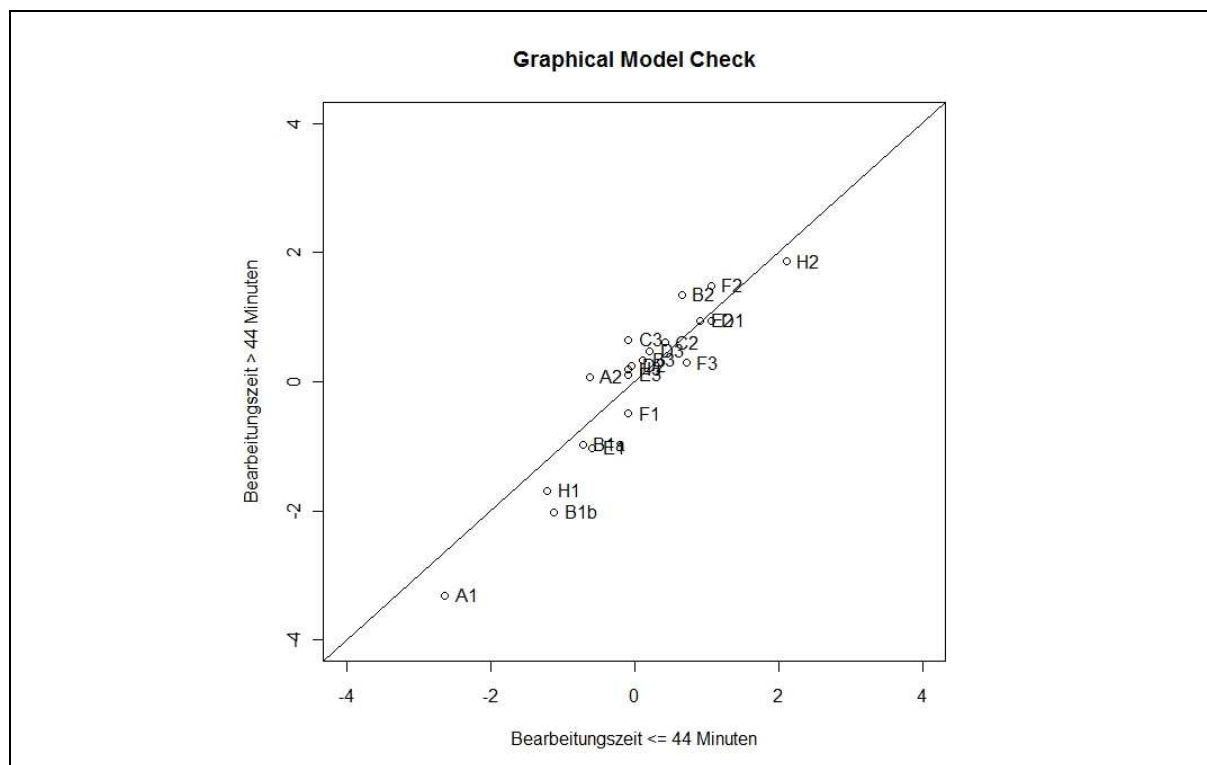


Abbildung 10 Grafischer Modelltest nach dem Ausschluss der Items,
Teilungskriterium Bearbeitungszeit, n=199

5.1.1.5 Teilungskriterium Interesse an dem Fach „Psychologische Diagnostik“

Bei dem Teilungskriterium Interesse handelt es sich um ein externes Kriterium, welches im Rahmen der ILP als Zusatzfrage mittels 5-stufiger Skala erhoben wurde. Die untersuchte Stichprobe (n=199) wurde in die Gruppen „niedriges Interesse“ (n=98) und „hohes Interesse“ (N=101) aufgeteilt nach dem Median (Md=2). Dabei wurden Personen mit dem Median Score der Gruppe „hohes Interesse“ zugeordnet.

Sowohl vor Ausscheiden der Items als auch danach zeigt der LQT, bei einem einheitlich festgesetzten Risiko 1. Art von $\alpha=.01$, kein signifikantes Ergebnis (vgl. Tabelle 9). Im Grafischen Modelltest zeigt kein Item eine zu große Abweichung von der 45° Gerade, weder vor Ausschluss der Items (vgl. Abbildung 11) noch danach (vgl. Abbildung 12). Somit sprechen sowohl die Ergebnisse des LQT als auch des Grafischen Modelltests für gute Modellpassung bezogen auf das Teilungskriterium Interesse.

Tabelle 9 Ergebnisse des LQT für das Teilungskriterium Interesse

Likelihood- Quotienten -Test vor Ausschluss der Items:	
LR-value:	17,322
Chi-square df:	21
p-value:	0,691
Likelihood- Quotienten -Test nach Ausschluss der Items:	
LR-value:	15,958
Chi-square df:	19
p-value:	0,66

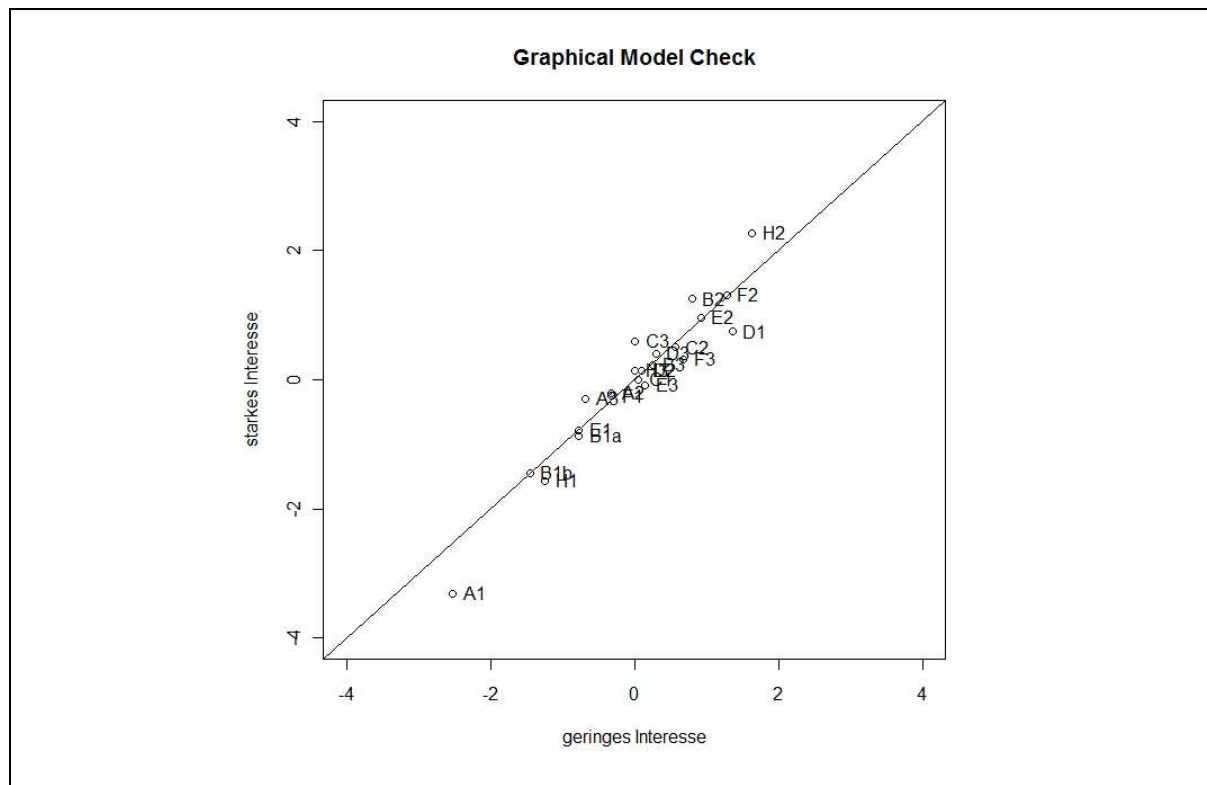


Abbildung 11 Grafischer Modelltest vor dem Ausschluss der Items, Teilungskriterium Interesse, n=199

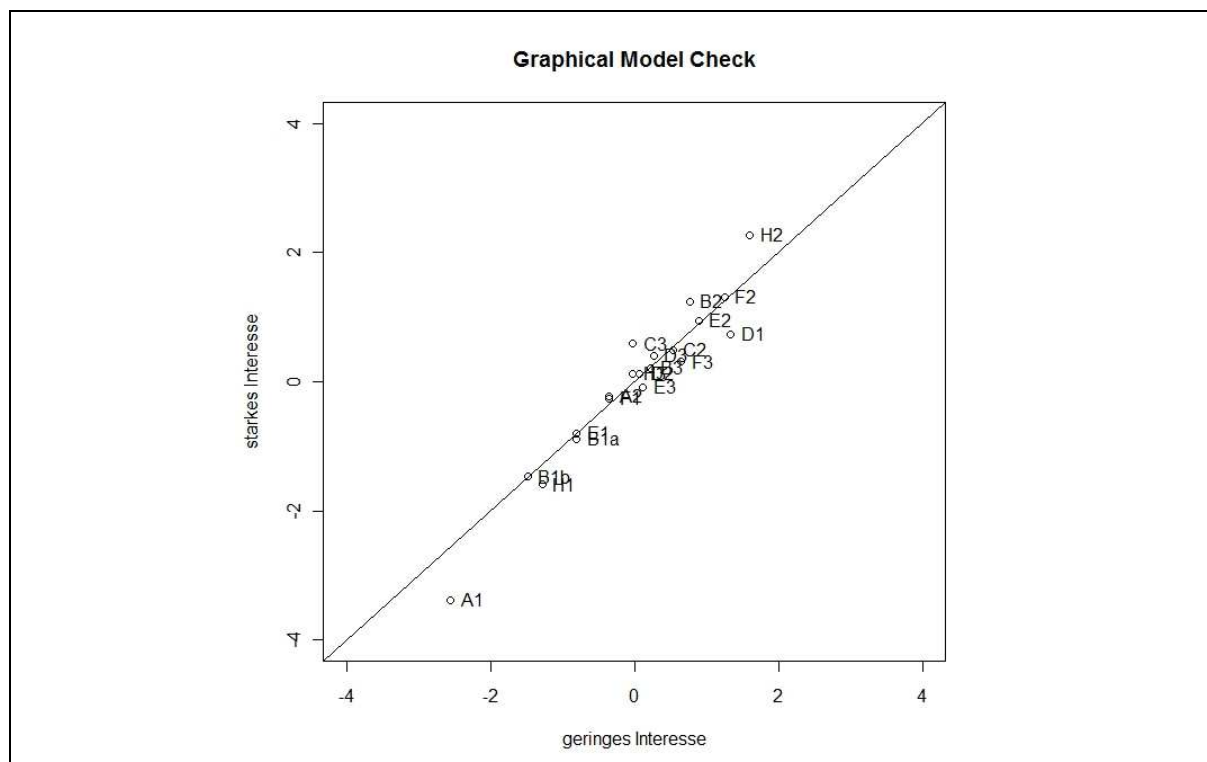


Abbildung 12 Grafischer Modelltest nach dem Ausschluss der Items, Teilungskriterium Interesse, n=199

5.1.1.6 Teilungskriterium Vorlesungsbesuch

Bei dem Teilungskriterium Vorlesungsbesuch handelt es sich um ein externes Kriterium, welches im Rahmen der ILP als Zusatzfrage mittels 5-stufiger Skala erhoben wurde. Die untersuchte Stichprobe (n=199) wurde in die Gruppen „häufiger Besuch“ (n=115) und „seltener Besuch“ (N=84) aufgeteilt nach dem Median (Md=2). Dabei wurden Personen mit dem Median Score der Gruppe „häufiger Besuch“ zugeordnet.

Sowohl vor Ausscheiden der Items als auch danach erbringt der LQT, bei einem einheitlich festgesetzten Risiko 1. Art von $\alpha=0,01$, kein signifikantes Ergebnis (vgl. Tabelle 10). Beim Grafischen Modelltest kommt es bei zwei Items (H1,H2) zu deutlichen Abweichungen von der 45° Geraden, sowohl vor Ausschluss der Items (A3,C1; vgl. Abbildung 13) als auch danach (vgl. Abbildung 14). Das Item H1 ist für Personen der Gruppe „seltener Besuch“ relativ schwieriger und für Personen der Gruppe „häufiger Besuch“ relativ leichter. Das Item H2 ist für Personen der Gruppe „häufiger Besuch“ relativ schwieriger und für Personen der Gruppe „seltener Besuch“ relativ leichter.

Nach Ausschluss der Items (A3, C1) sprechen die Ergebnisse des LQT für gute Modellpassung bezogen auf das Teilungskriterium Vorlesungsbesuch. Die Ergebnisse des Grafischen Modelltests sind trotz Abweichungen der Items H1 und H2 zufriedenstellend.

Tabelle 10 Ergebnisse des LQT für das Teilungskriterium Vorlesungsbesuch

Likelihood- Quotienten -Test vor Ausschluss der Items:	
LR-value:	33,389
Chi-square df:	21
p-value:	0,042
Likelihood- Quotienten -Test nach Ausschluss der Items:	
LR-value:	27,604
Chi-square df:	19
p-value:	0,091

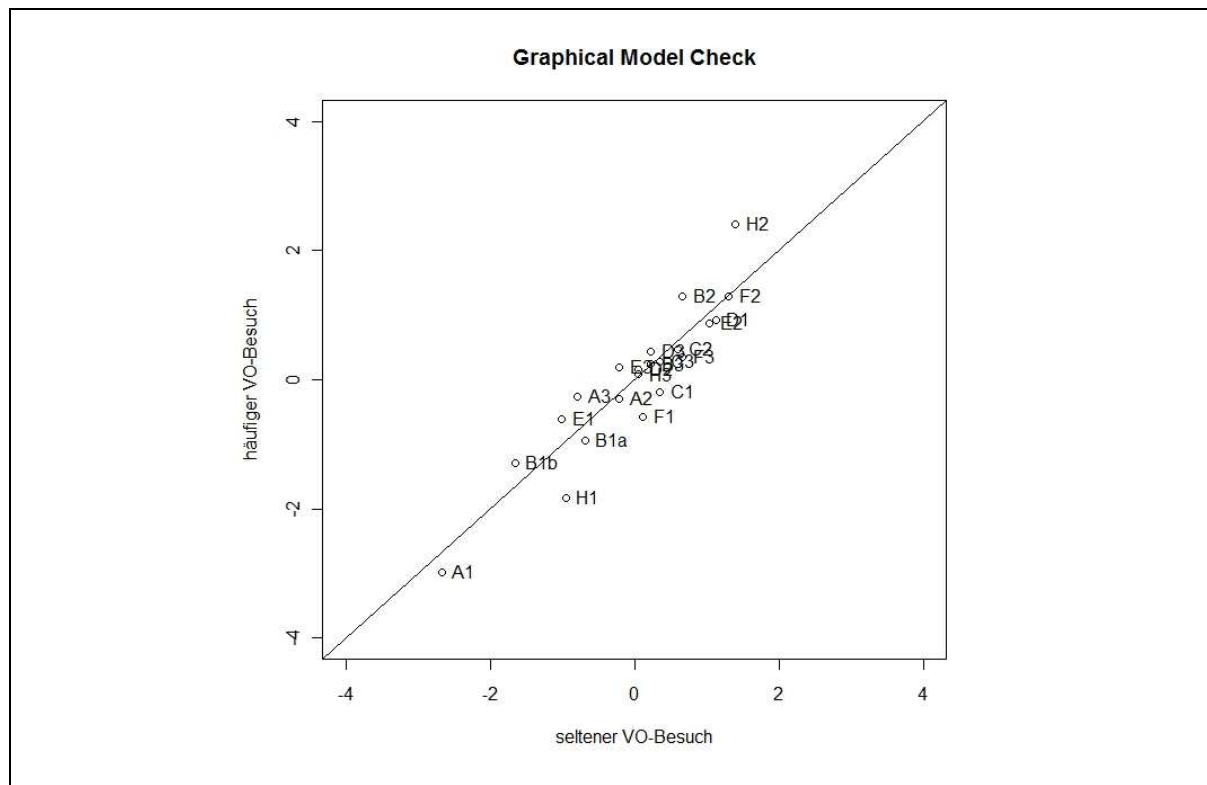


Abbildung 13 Grafischer Modelltest vor dem Ausschluss der Items,
Teilungskriterium Vorlesungsbesuch, n=199

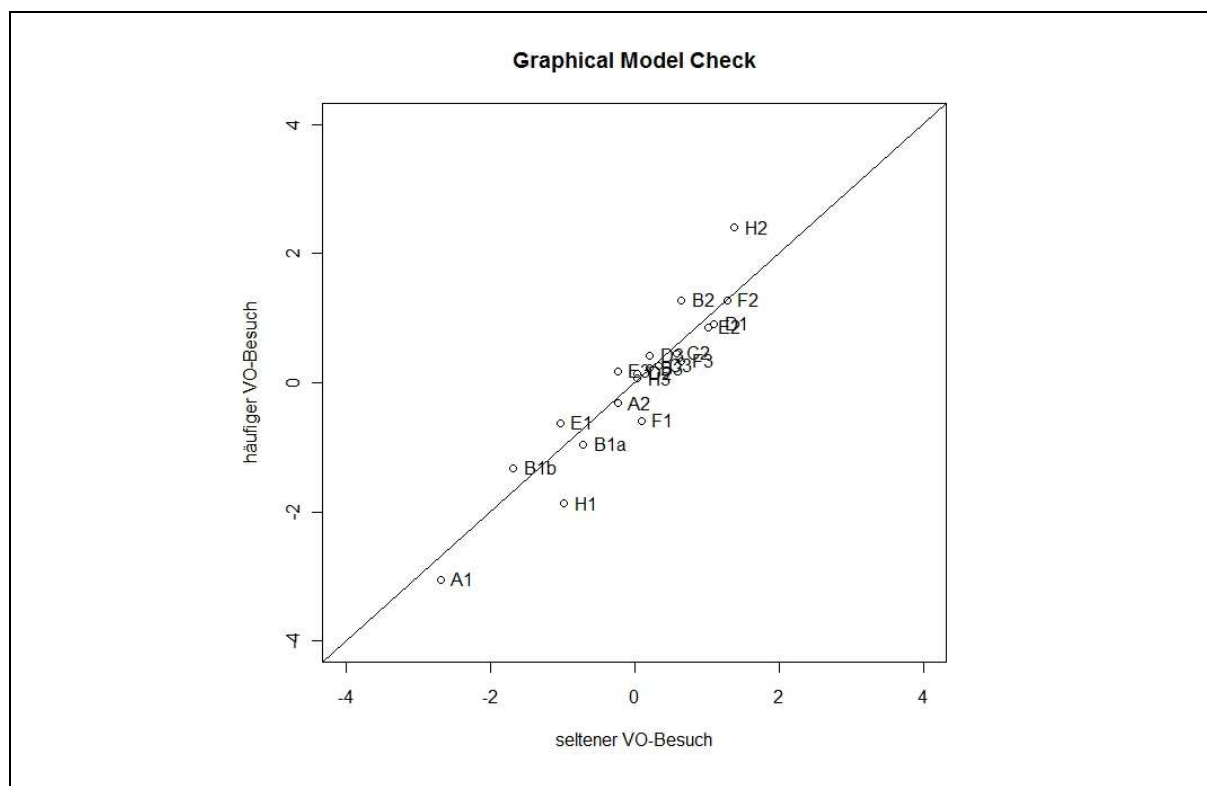


Abbildung 14 Grafischer Modelltest nach dem Ausschluss der Items,
Teilungskriterium Vorlesungsbesuch, n=199

5.1.1.7 Teilungskriterium Übungsbesuch

Bei dem Teilungskriterium Übungsbesuch handelt es sich um ein externes Kriterium. Die untersuchte Stichprobe (n=199) wurde in die Gruppen „Übungsbesuch im laufenden Semester“ (n=162) und „kein Übungsbesuch im laufenden Semester“ (N=34) aufgeteilt.

Sowohl vor Ausscheiden der Items als auch danach erbringt der LQT, bei einem einheitlich festgesetzten Risiko 1. Art von $\alpha=0,01$, kein signifikantes Ergebnis (vgl. Tabelle 11). Beim Grafischen Modelltest kommt es bei einem Item (A3) zu einer großen Abweichung von der 45° Geraden (vgl. Abbildung 15). Das Item A3 ist für Personen welche die Übungen im laufenden Semester besuchten relativ schwieriger und für Personen die Übungen im laufenden Semester nicht besuchten relativ leichter. Nach Ausschluss der Items (A3, C1) sprechen sowohl die Ergebnisse des LQT als auch des Grafischen Modelltests (vgl. Abbildung 16) für gute Modellpassung bezogen auf das Teilungskriterium Übungsbesuch.

Tabelle 11 Ergebnisse des LQT für das Teilungskriterium Übungsbesuch

Likelihood- Quotienten -Test vor Ausschluss der Items:	
LR-value:	27,515
Chi-square df:	21
p-value:	0,154
Likelihood- Quotienten -Test nach Ausschluss der Items:	
LR-value:	10,73
Chi-square df:	19
p-value:	0,933

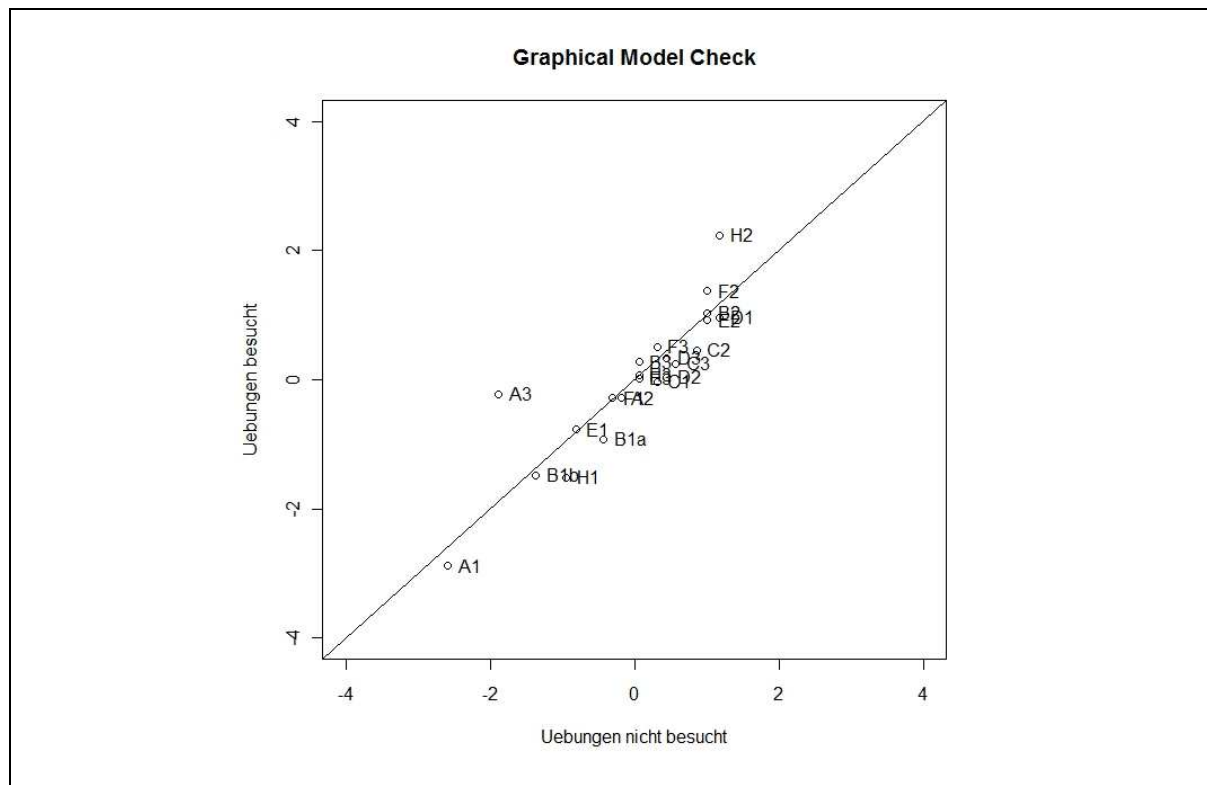


Abbildung 15 Grafischer Modelltest vor dem Ausschluss der Items,
Teilungskriterium Übungsbesuch, n=199

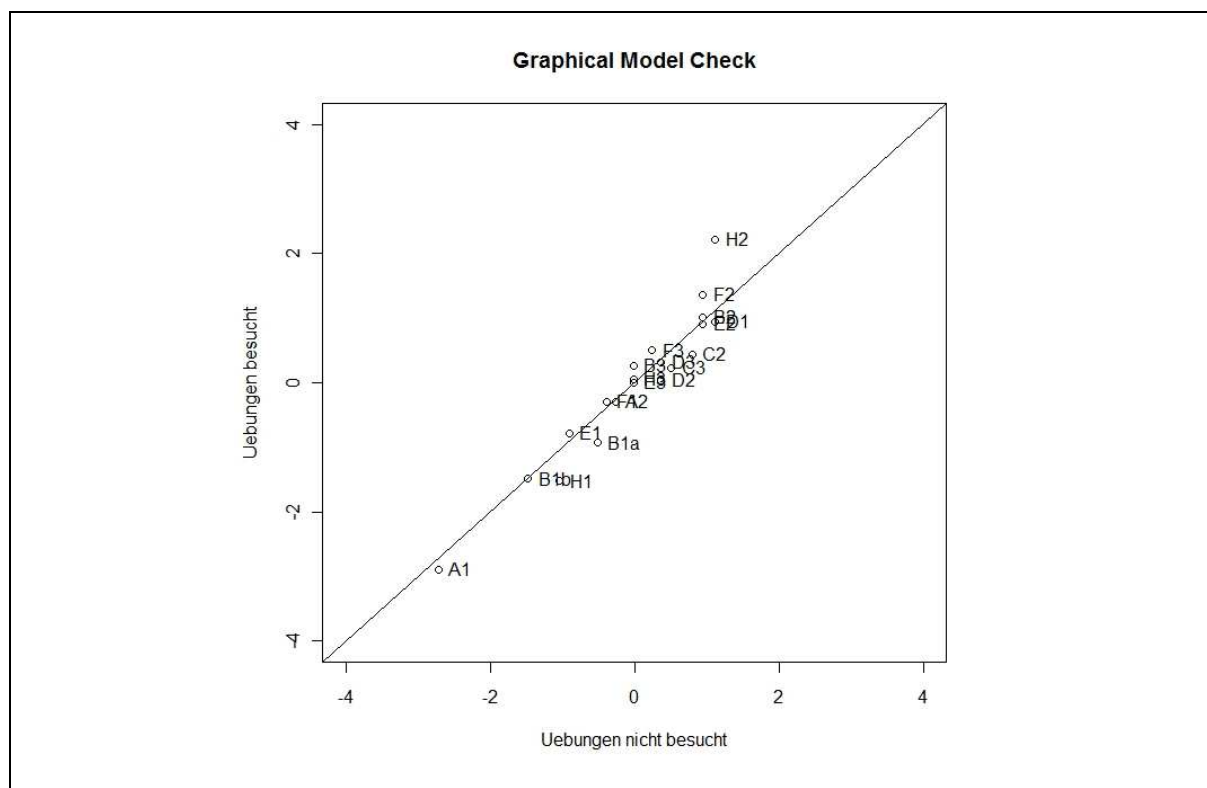


Abbildung 16 Grafischer Modelltest nach dem Ausschluss der Items,
Teilungskriterium Übungsbesuch, n=199

5.1.2 Zusammenfassende Betrachtung der Ergebnisse der Analysen nach dem Rasch-Modell

Bereits nach Ausschluss von zwei (9%) der 22 Items ergaben die Modelltests eine zufriedenstellende Modellpassung. Die Geltung des Rasch-Modells bedeutet, dass die Items der ILP eine Dimension „psychologisch diagnostischen Wissens“ erfassen und nicht drei von einander weitgehend unabhängige Fähigkeiten analog den Lernzielen nach Bloom (1956).

Für die weiteren Berechnungen bedeutet die Geltung des Rasch-Modells, dass der Summenscore ein passendes Maß für die Fähigkeit einer Person ist und es somit gerechtfertigt ist mit dem Summenscore weiter zu arbeiten (vgl. Rost, 2004).

5.2 Deskriptive Statistiken zur Prüfung psychologische Diagnostik 1 und den einzelnen Prädiktoren

Die in Kapitel 2 aufgezählten möglichen Prädiktoren für die Leistung in der Prüfung zur psychologischen Diagnostik 1 werden in diesem Abschnitt deskriptiv statistisch analysiert und mit der Prüfungsleistung korreliert sowie die Korrelationen untereinander werden betrachtet. Eine hohe Korrelation einzelner Prädiktoren untereinander würde auf Multikollinearität hinweisen, und eine somit gegen die Anwendung eines Regressionsmodells sprechen (vgl. Backhaus, Erichson, Plinke & Weiber 2006).

5.2.1 Ergebnisse der Prüfung psychologische Diagnostik 1

Die in dieser Studie verwendeten Prüfungsergebnisse stammen alle vom Prüfungstermin 27.6.2008. Dieser Termin war der 1. Prüfungstermin der Vorlesung psychologische Diagnostik 1 des Sommersemesters 2008 (SS2008). Insgesamt nahmen 118 Studierende an diesem Prüfungstermin teil. Die Verteilung der Prüfungsergebnisse ist Tabelle 12 zu entnehmen.

Tabelle 12 Noten der Prüfung psychologische Diagnostik 1 (Md=4)

	Note	Häufigkeit	Prozent
1	(Sehr Gut)	3	2,5
2	(Gut)	12	10,2
3	(Befriedigend)	33	28
4	(Genügend)	37	31,4
5	(Nicht Genügend)	33	28
Gesamt		118	100

Da Prädiktoren gefunden werden sollten die Hinweise über erfolgreiches und nicht erfolgreiches Abschneiden bei der Prüfung geben sollten, wurden die Ergebnisse der Prüfung dichotomisiert. Hier wurde auf 2 Arten eine Dichotomisierung durchgeführt. Einerseits wurde ein Mediansplit durchgeführt in eine Gruppe erfolgreicher Studierender (Noten 1-3) und in eine Gruppe nicht erfolgreicher Studierender (Note 4-5), im Weiteren wird diese Variable als Prüfungsleistung bezeichnet. Andererseits wurde getrennt in „bestanden“ (Noten 1-4) und „nicht bestanden“ (Note 5), im Weiteren wird diese Variable einfach als „Prüfung bestanden“ bezeichnet. Die Personenanzahlen pro Gruppe sind Tabelle 13 zu entnehmen.

Tabelle 13 Dichotomisierte Prüfungsergebnisse

Mediansplit (=Prüfungsleistung)		
	Personen	Prozent
Erfolgreich (Noten 1-3)	48	40,7
Nicht erfolgreich (Noten 4+5)	70	59,3
Gesamt	118	100
Prüfung bestanden		
	Personen	Prozent
Bestanden (Noten 1-4)	85	72
Nicht Bestanden (Note 5)	33	28
Gesamt	118	100

Folgend wurde jeweils eine logistische Regression beider mit beiden Arten der Dichotomisierung als abhängige Variable gerechnet (siehe dazu Kapitel 5.3).

5.2.2 Ergebnis der ILP

Ein möglicher Prädiktor für die Prüfungsleistung ist die Anzahl gelöster Items der ILP. Da für die Items der ILP Rasch-Modell-Konformität festgestellt werden konnte (vgl. Kapitel 5.1) stellt die Anzahl gelöster Items eine „erschöpfende Statistik“ dar (Kubinger 1989). Für die logistische Regression konnte daher die Anzahl gelöster Items als metrisch skalierte Variable (Ergebnis ILP) einbezogen werden (siehe dazu Kapitel 5.3).

Von den 118 Studierenden welche die Prüfung zum 1. Prüfungstermin absolviert haben, haben 79 (66,9%) zuvor an der ILP teilgenommen (vgl. Tabelle 14). Da nicht alle Prüfungsteilnehmer auch an der ILP teilgenommen haben, sondern eben nur 66,9% wurde für ein Gesamt-Regressionsmodell für alle 118 Prüfungsteilnehmer die binäre Variable „Teilnahme ILP“ als Prädiktor herangezogen.

Tabelle 14 Teilnahme an der ILP unter den Teilnehmern des 1. Prüfungstermins

Teilnahme ILP		
	Personen	Prozent
Teilgenommen	79	66,9
Nicht teilgenommen	39	33,1
Gesamt	118	100

Für die weiteren Berechnungen wurde in Kapitel 5.1 die Items A3 und C1 der ILP ausgeschlossen. Das in Abbildung 17 dargestellte Histogramm bezieht sich daher auf den

bereinigten Itempool von 20 Items. Das erreichte Ergebnis kann also nur im Intervall [0;20] liegen. Der Median für die Anzahl gelöster Items liegt bei 9. Der niedrigste erreichte Wert liegt bei 2, der höchste bei 16 richtig gelösten Items.

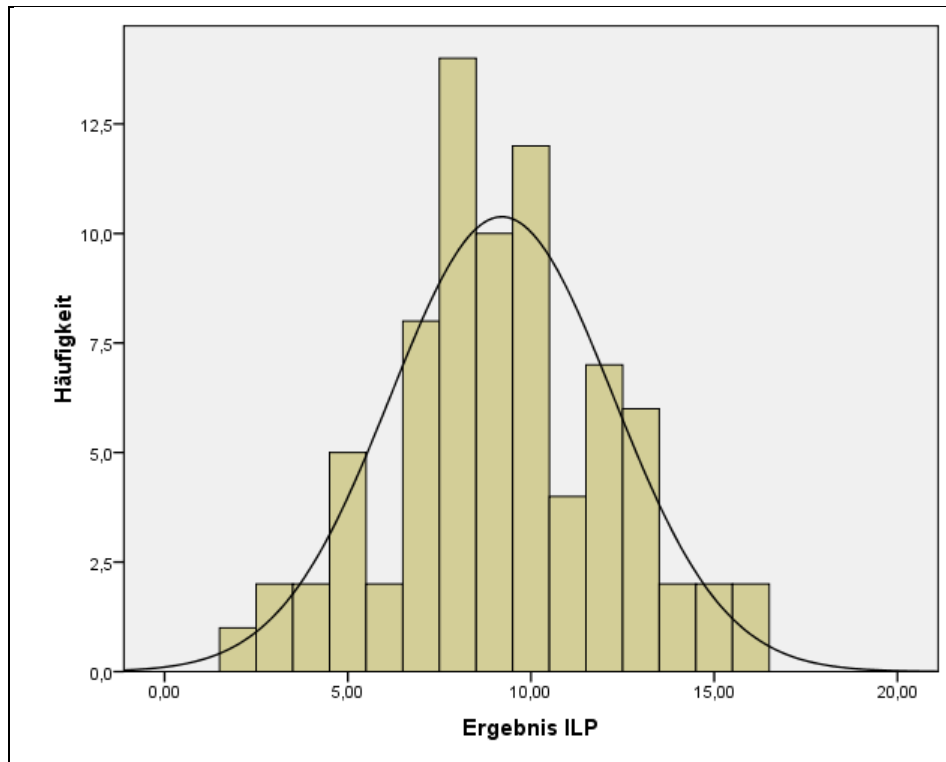


Abbildung 17 Ergebnis ILP. Md=9; n=79

5.2.3 Verbale Intelligenz (erhoben mittels IBF)

Als weiterer möglicher Prädiktor wird die Verbale Intelligenz berücksichtigt, welche mittels Untertest Verbale Intelligenzfunktionen des Verfahrens Intelligenz-Basis-Funktionen (IBF, Blum et al., 2005) im Rahmen des PeRaS Projekts (vgl. Kapitel 1) erhoben wurde. Für die logistische Regression wurde die Verbale Intelligenz in Form von T-Werten und daher als metrisch skalierte Variable einbezogen (siehe dazu Kapitel 5.3).

Von den 118 Studierenden welche die Prüfung zum 1. Prüfungstermin absolviert haben, konnte bei 70 (59,3%) die Verbale Intelligenz erhoben werden (vgl. Tabelle 15).

Tabelle 15 Daten zur Verbale Intelligenz lt. IBF vorhanden unter den Teilnehmern des 1. Prüfungstermins

Verbale Intelligenz lt. IBF		
	Personen	Prozent
Daten vorhanden	70	59,3
Daten nicht Vorhanden	48	40,7
Gesamt	118	100

Die Daten liegen in T-Werten¹⁰ vor. Der Median für Verbale Intelligenz in der betrachteten Stichprobe liegt bei einem T-Wert von Md=60,41. Der niedrigste erreichte T-Wert liegt bei 36,7, der höchste bei 68,07 (vgl. Abbildung 18).

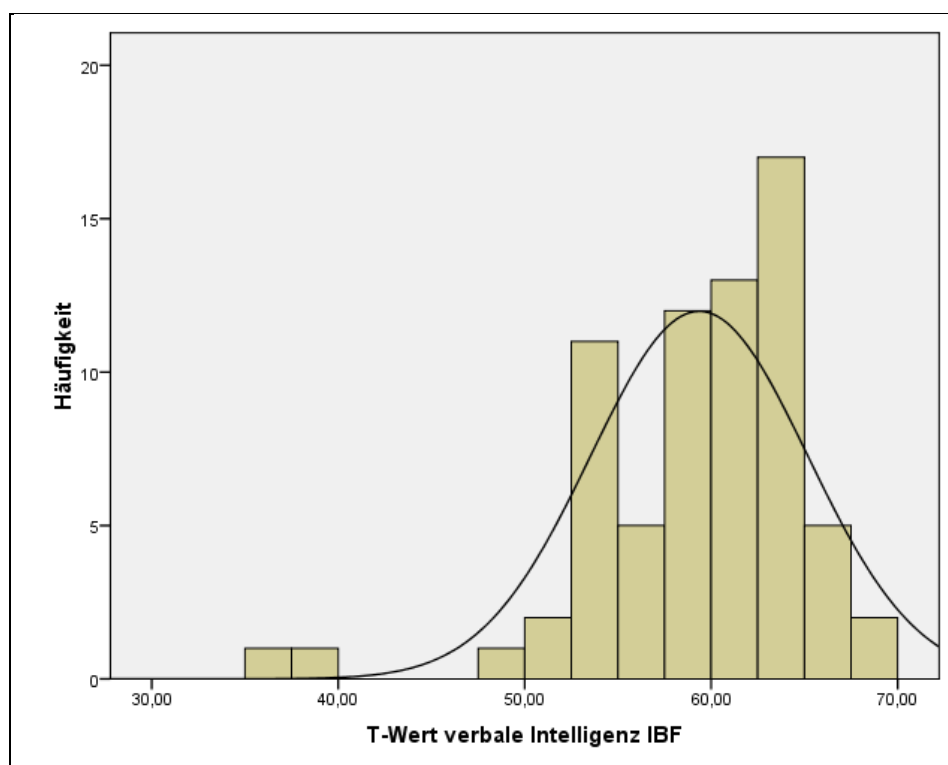


Abbildung 18 Verbale Intelligenz lt. IBF; Md=60,03; n=70

Um Mittelwert und Standardabweichung berechnen zu können wurden die beiden Ausreißer (T-Wert unter 40) ausgeschlossen. Somit verringerte sich der Stichprobenumfang auf n=68. Der nach dem Ausschluss erhobene Mittelwert beträgt M=59,98 und die dazugehörige Standardabweichung beträgt s=4,571. Mittelwert und Standardabweichung der Stichprobe wurden weiters mit dem Mittelwert und der

¹⁰ als Vergleichsstichprobe wurde die Bildungsstufen-Norm herangezogen (EU 4-5: Matura oder höher).

Standardabweichung der Population¹¹ verglichen. Hier zeigte sich, dass der in der Stichprobe erhobene Mittelwert ($M=59,98$) sich signifikant vom Mittelwert der Referenzpopulation unterscheidet und zwar dahingehend, dass der Mittelwert der Stichprobe signifikant größer ist als jener der Referenzpopulation. Die Varianz der Stichprobe wurde mit der Varianz der Referenzpopulation verglichen¹². Für die Stichprobe ergab sich $\chi^2=13,999$. Der kritische Wert lautet $\chi^2(67; \alpha=0,05)=49,162$. Da der kritische Wert höher ist als der empirische wird die Alternativhypothese $H_A: \sigma^2 > \sigma_0^2$ angenommen. Dies bedeutet, dass sich die Varianzen signifikant unterscheiden und zwar dahingehend, dass die Streuung der Testwerte in der Stichprobe niedriger ist als in der Referenzpopulation.

5.2.4 Vorlesungsbesuch

Im Rahmen der Prüfung zur psychologischen Diagnostik 1 wurde auch eine Fragebogenerhebung durchgeführt (vgl. Kapitel 4.2.3). Dabei wurde der mögliche Prädiktor Vorlesungsbesuch erhoben, mittels fünfstufiger Skala, wobei 1 „kein Vorlesungsbesuch“ und 5 „alle Vorlesungseinheiten besucht“ bedeutet. Der exakte Wortlaut der Frage und der dazugehörigen Antwortmöglichkeiten findet sich im Anhang 9.5. Für die logistische Regression wurde der Vorlesungsbesuch als metrisch skalierte Variable (Rangskala) einbezogen (siehe dazu Kapitel 5.3).

108 von insgesamt 118 Studierenden welche die Prüfung zum 1. Prüfungstermin absolviert haben, haben Angaben zum Vorlesungsbesuch im Sommersemester 2008 gemacht (vgl. Tabelle 16).

¹¹ Als Population wird hier die Referenzpopulation von Personen der Bildungsstufen 4-5 herangezogen.

¹² $H_0: \sigma^2 = \sigma_0^2$

Tabelle 16 Angaben zum Vorlesungsbesuch im Sommersemester 2008, erhoben mittels Fragebogen nach dem 1. Prüfungstermin

Vorlesungsbesuch		
	Personen	Prozent
Kein Besuch (1)	31	26,3
1 bis 3 Besuche (2)	9	7,6
4 bis 6 Besuche (3)	9	7,6
7 bis 9 Besuche (4)	35	29,7
alle 10 LV besucht (5)	24	20,3
Keine Angabe	10	8,5
Gesamt	118	100

5.2.5 Übungsbesuch

Der Prädiktor Übungsbesuch konnte mittels der Daten der Lehrveranstaltungsleiter erhoben werden. Für die Logistische Regression wurde der Übungsbesuch als binäre Variable (0=nicht besucht; 1=besucht) einbezogen (siehe dazu Kapitel 5.3).

70 von insgesamt 118 Studierenden welche die Prüfung zum 1. Prüfungstermin absolviert haben, haben parallel zur Vorlesung die Übungen zum psychologischen Diagnostizieren im Sommersemester 2008 besucht (vgl. Tabelle 17).

Tabelle 17 Besuch der Übungen zum psychologischen Diagnostizieren im Sommersemester 2008

Übungsbesuch		
	Personen	Prozent
Übungen besucht	70	59,3
Übungen nicht besucht	48	40,7
Gesamt	118	100

5.2.6 Korrelationen

Als letzter Schritt vor der Berechnung des Regressionsmodells wurden alle Prädiktoren (Regressoren) und die abhängige Variable (Prüfungserfolg) in einer Korrelationsmatrix betrachtet um ungewünschte Korrelationen zwischen den Prädiktoren einerseits und gewünschte Korrelationen zwischen Prädiktoren und der abhängigen Variable andererseits zu erkennen.

Der Prädiktor „Teilnahme an der ILP“ weist signifikante Korrelationen mit den Prädiktoren „Übungsbesuch“ (0,702; vgl. Tabelle 18) und Vorlesungsbesuch auf. Ebenso korrelieren die Prädiktoren Übungsbesuch und Vorlesungsbesuch miteinander (0,437; vgl. Tabelle 18). Die Prädiktoren „Ergebnis ILP“ und Verbale Intelligenz korrelieren ebenfalls signifikant miteinander (0,310; vgl. Tabelle 18).

Mit den beiden Varianten der abhängigen Variablen weisen die Prädiktoren „Teilnahme an der ILP“ (0,252 und 0,204; vgl. Tabelle 18) und „Ergebnis ILP“ (0,350 und 0,324; vgl. Tabelle 18) signifikante Korrelationen auf. Der Prädiktor Übungsbesuch korreliert mit der Variablen Prüfungsleistung signifikant (0,229; vgl. Tabelle 18).

Tabelle 18 Korrelationsmatrix der Prädiktoren und des Studienerfolges. Bei einem einheitlich festgesetzten Risiko 1. Art von $\alpha=.05$ signifikante Ergebnisse sind **fett** markiert. Alle Ergebnisse wurden mittels Spearmans Rangkorrelation gerechnet; 2 seitig

	Prüfungs- Note	Prüfungs- leistung	Prüfung bestanden	Teilnahme ILP	Ergebnis ILP	Verbale Intelligenz	Übungs- besuch	Vorlesungs- besuch
Prüfungs- Note	X	0,885	0,809	0,259	0,389	0,186	0,151	0,074
Prüfungslei- stung	0,885	X	0,516	0,252	0,350	0,176	0,229	0,126
Prüfung bestanden	0,809	0,516	X	0,204	0,324	0,184	0,061	-0,038
Teilnahme ILP	0,259	0,252	0,204	X	-	0,018	0,702	0,374
Ergebnis ILP	0,389	0,350	0,324	-	X	0,280	0,040	0,173
Verbale Intelligenz	0,186	0,176	0,184	0,018	0,280	X	-	-0,104
Übungsbesuch	0,151	0,229	0,061	0,702	0,040	-	X	0,437
Vorlesungs- besuch	0,074	0,126	-0,038	0,374	0,173	-0,104	0,437	X

5.3 Ergebnisse der Logistischen Regression

Insgesamt wurden 6 Regressionsmodelle gerechnet. Nötig war dies, da in der Gesamtstichprobe der Prüfungsteilnehmer ($n=118$) für einige Variablen Daten fehlten. Die Stichprobengröße reduzierte sich somit um 50,8% auf $n=58$. Außerdem besteht in dieser Stichprobe keine Varianz in der Variablen Übungsbesuch, da nur für Personen die die Übung besucht haben auch Daten zur Verbalen Intelligenz vorliegen. Da die Korrelationen aus Kapitel 5.2.6 zeigen dass die Variable Übungsbesuch mit dem Prüfungserfolg korreliert, wurde es nötig zwei Regressionsmodelle zu rechnen. Im ersten Modell mit dem Stichprobenumfang $n=108$ (bei 10 Personen waren nicht alle Daten der Variablen Vorlesungsbesuch vorhanden), wurden die Variablen Vorlesungsbesuch, Übungsbesuch und Teilnahme an der ILP als Regressoren eingebunden. Im zweiten Modell mit dem Stichprobenumfang $n=58$ wurden die Variablen Vorlesungsbesuch, Ergebnis ILP und Verbale Intelligenz als Regressoren eingebunden. Aufgrund der Ergebnisse des 2. Regressionsmodells ($n=58$) wurde auch noch ein 3. Modell ohne die Variable Verbale Intelligenz, dafür aber mit der Variablen Übungsbesuch berechnet. Da die abhängige Variable auf zwei unterschiedliche Arten kodiert wurde, die allerdings beide von Interesse sind (vgl. Kapitel 5.2.1), wurden alle drei Modelle sowohl mit der abhängigen Variablen Prüfungsleistung als auch mit der Variablen „Prüfung bestanden“ berechnet. Insgesamt wurden also sechs Regressionsmodelle berechnet deren Ergebnisse in den folgenden Abschnitten dargestellt werden.

Die Modelle wurden in SPSS mittels schrittweiser Regression berechnet.

5.3.1 Regressionsmodell 1a und 1b

Folgende Prädiktoren wurden für diese Modelle verwendet:

- Vorlesungsbesuch (vgl. Kapitel 5.2.4)
- Übungsbesuch (vgl. Kapitel 5.2.5)
- Teilnahme ILP (vgl. Kapitel 5.2.2)

5.3.1.1 Regressionsmodell 1a

Hier wurde die Stichprobe ($n=108$) anhand der Variablen Prüfungsleistung in die beiden Gruppen „gute Leistung“ und „schlechte Leistung“ geteilt (vgl. Kapitel 5.2.1). Die

logistische Regressionsgleichung wurde erstellt um die Zugehörigkeit zu diesen zwei Gruppen vorherzusagen.

Tabelle 19 Modellfit von Regressionsmodell 1a

-2LL-Wert	Cox & Snell R ²	Nagelkerke R ²
1138,485	0,053	0,072

Als Prüfgrößen für den Modellfit wurden Cox und Snell R² und Nagelkerke R² herangezogen. Für beide Werte gilt lt. Backhaus (2006), dass Werte grösser 0,2 akzeptable und Werte grösser 0,4 gute Modellpassung bedeuten. Wie in Tabelle 19 zu sehen, sind die notwendigen Werte weit unterschritten, was bedeutet dass keine ausreichende Modellpassung vorliegt.

Tabelle 20 Koeffizienten der logistischen Regression Modell 1a; Das Signifikanzniveau wurde einheitlich bei $\alpha=0,05$ festgelegt

In das Modell eingeschlossene Variablen						
	B	Std. Fehler	Wald	df	Signifikant?	Exp(B)
Konstante	-1,216	0,403	9,131	1	Ja	0,296
Teilnahme ILP	1,079	0,466	5,365	1	Ja	2,942
Nicht eingeschlossene Variablen						
Vorlesungsbesuch	-	-	0,021	1	Nein	-
Übungsbesuch	-	-	0,396	1	Nein	-

Tabelle 20 zeigt die Koeffizienten des Regressionsmodells, die aufgrund des schlechten Modellfits allerdings nicht weiter interpretiert werden. Es konnte keiner Vorhersage über die Gruppenzugehörigkeit getroffen werden.

5.3.1.2 Regressionsmodell 1b

Hier wurde die Stichprobe (n=108) anhand der Variablen „Prüfung bestanden“ in die beiden Gruppen „Prüfung positiv absolviert“ und „Prüfung nicht positiv absolviert“ geteilt (vgl. Kapitel 5.2.1). Die logistische Regressionsgleichung wurde erstellt um die Zugehörigkeit zu diesen zwei Gruppen vorherzusagen.

Tabelle 21 Koeffizienten der logistischen Regression Modell 1b; Das Signifikanzniveau wurde einheitlich bei $\alpha=0,05$ festgelegt

In das Modell eingeschlossene Variablen						
	B	Std. Fehler	Wald	df	Signifikant?	Exp(B)
Konstante	0,910	0,213	18,295	1	Ja	2,484
Nicht eingeschlossene Variablen						
Teilnahme ILP	-	-	3,229	1	Nein	-
Vorlesungsbesuch	-	-	0,235	1	Nein	-
Übungsbesuch	-	-	0,026	1	Nein	-

Wie Tabelle 21 zeigt, wurde keiner der Regressoren in das Modell 1b aufgenommen. Es konnte keine Vorhersage über die Gruppenzugehörigkeit getroffen werden.

5.3.2 Regressionsmodell 2a und 2b

Folgende Prädiktoren wurden für diese Modelle verwendet:

- Vorlesungsbesuch (vgl. Kapitel 5.2.4)
- Ergebnis ILP (vgl. Kapitel 5.2.2)
- Verbale Intelligenz (vgl. Kapitel 5.2.3)

5.3.2.1 Regressionsmodell 2a

Hier wurde die Stichprobe ($n=60$) anhand der Variablen Prüfungsleistung in die beiden Gruppen „gute Leistung“ und „schlechte Leistung“ geteilt (vgl. Kapitel 5.2.1). Die logistische Regressionsgleichung wurde erstellt um die Zugehörigkeit zu diesen zwei Gruppen vorherzusagen.

Tabelle 22 Modellfit von Regressionsmodell 2a

-2LL-Wert	Cox & Snell R^2	Nagelkerke R^2
68,096	0,189	0,252

Wie in Tabelle 22 zu sehen ist, liegt der Cox & Snell R^2 Wert 0,189 also knapp unter dem von Backhaus et. al (2006) geforderten Wert von 0,2 und der Nagelkerke R^2 Wert bei 0,252 also in einem akzeptablen Bereich. Insgesamt ist die Modellpassung noch akzeptabel.

Tabelle 23 Koeffizienten der logistischen Regression Modell 2a; Das Signifikanzniveau wurde einheitlich bei $\alpha=0,05$ festgelegt

In das Modell eingeschlossene Variablen						
	B	Std. Fehler	Wald	df	Signifikant?	Exp(B)
Konstante	-3,345	1,141	8,595	1	Ja	0,035
Ergebnis ILP	0,353	0,117	9,177	1	Ja	1,423
Nicht eingeschlossene Variablen						
Vorlesungsbesuch	-	-	1,232	1	Nein	-
Verbale Intelligenz	-	-	0,005	1	Nein	-

Tabelle 23 zeigt die Koeffizienten des Regressionsmodells. In das Modell eingeschlossen wurde lediglich der Prädiktor „Ergebnis ILP“. Die beiden anderen Prädiktoren weisen keinen signifikanten Einfluss auf die abhängige Variable auf.

Tabelle 24 Klassifikationstabelle Modell 2a

Beobachtet	Vorhergesagt		
	Prüfungsleistung		
	schlechte Leistung	gute Leistung	% korrekt
schlechte Leistung	22	8	73,3
gute Leistung	10	20	66,7
Gesamt % korrekt			70,0

Wie in Tabelle 24 zu sehen ist, liegt der Prozentsatz richtig klassifizierter Personen bei 70,0% und somit deutlich über der maximalen Zufallswahrscheinlichkeit von 50% sowie der proportionalen Zufallswahrscheinlichkeit von ebenfalls 50%. Dies entspricht einem Zuwachs von 20,7% richtiger Zuordnungen.

5.3.2.2 Regressionsmodell 2b

Hier wurde die Stichprobe (n=60) anhand der Variablen „Prüfung bestanden“ in die beiden Gruppen „Prüfung positiv absolviert“ und „Prüfung nicht positiv absolviert“ geteilt (vgl. Kapitel 5.2.1). Die logistische Regressionsgleichung wurde erstellt um die Zugehörigkeit zu diesen zwei Gruppen vorherzusagen.

Tabelle 25 Modellfit von Regressionsmodell 2b

-2LL-Wert	Cox & Snell R ²	Nagelkerke R ²
53,396	0,192	0,285

Wie in Tabelle 25 zu sehen ist, liegt der Cox & Snell R² Wert bei 0,192 also knapp unter dem geforderten Wert von 0,2 und der Nagelkerke R² Wert bei 0,285 also in einem akzeptablen Bereich. Insgesamt ist die Modellpassung noch akzeptabel.

Tabelle 26 Koeffizienten der logistischen Regression Modell 2b;

Das Signifikanzniveau wurde einheitlich bei $\alpha=0,05$ festgelegt

In das Modell eingeschlossene Variablen						
	B	Std. Fehler	Wald	df	Signifikant?	Exp(B)
Konstante	-2,493	1,175	4,505	1	Ja	0,083
Ergebnis ILP	0,413	0,137	9,061	1	Ja	1,511
Nicht eingeschlossene Variablen						
Vorlesungsbesuch	-	-	0,002	1	Nein	-
Verbale Intelligenz	-	-	0,620	1	Nein	-

Tabelle 26 zeigt die Koeffizienten des Regressionsmodells. In das Modell eingeschlossen wurde lediglich der Prädiktor „Ergebnis ILP“. Die beiden Prädiktoren Vorlesungsbesuch und Verbale Intelligenz weisen keinen signifikanten Einfluss auf die abhängige Variable auf.

Tabelle 27 Klassifikationstabelle Modell 2b

Beobachtet	Vorhergesagt		
	„Prüfung bestanden“		
	nicht bestanden	bestanden	% korrekt
nicht bestanden	5	10	33,3
bestanden	3	42	93,3
Gesamt % korrekt			78,3

Der Prozentsatz richtig klassifizierter Personen liegt bei 78,3% (vgl. Tab. 27) und somit nur knapp über der maximalen Zufallswahrscheinlichkeit von 75% und etwas deutlicher über der proportionalen Zufallswahrscheinlichkeit von 62,5%. Dies entspricht einem

Zuwachs zwischen 3,3% und 15,8% richtiger Zuordnungen, einmal verglichen mit der maximalen und einmal verglichen mit der proportionalen Zufallswahrscheinlichkeit. Kritisch ist anzumerken, dass die Ausprägung „nicht bestanden“ der abhängige Variable „Prüfung bestanden“ nur 15 Fälle aufweist. Backhaus et al (2006) geben an, dass die Fallzahl pro Gruppe der abhängigen Variable zumindest 25 betragen soll.

5.3.3 Regressionsmodelle 3a und 3b

Folgende Prädiktoren wurden für diese Modelle verwendet:

- Vorlesungsbesuch (vgl. Kapitel 5.2.4)
- Ergebnis ILP (vgl. Kapitel 5.2.2)
- Übungsbesuch (vgl. Kapitel 5.2.5)

5.3.3.1 Regressionsmodell 3a

Hier wurde die Stichprobe (n=73) anhand der Variablen Prüfungsleistung in die beiden Gruppen „gute Leistung“ und „schlechte Leistung“ geteilt (vgl. Kapitel 5.2.1). Die logistische Regressionsgleichung wurde erstellt um die Zugehörigkeit zu diesen zwei Gruppen vorherzusagen.

Tabelle 28 Modellfit von Regressionsmodell 3a

-2LL-Wert	Cox & Snell R^2	Nagelkerke R^2
87,440	0,168	0,224

Wie in Tabelle 28 zu sehen ist, liegt der Cox & Snell R^2 Wert 0,168 also unter den geforderten 0,2 und der Nagelkerke R^2 Wert bei 0,224 also knapp über den geforderten 0,2. Insgesamt ist die Modellpassung hier mangelhaft.

Tabelle 29 Koeffizienten der logistischen Regression Modell 3a;
Das Signifikanzniveau wurde einheitlich bei $\alpha=0,05$ festgelegt

In das Modell eingeschlossene Variablen						
	B	Std. Fehler	Wald	df	Signifikant?	Exp(B)
Konstante	-3,219	1,002	10,312	1	Ja	0,040
Ergebnis ILP	0,326	0,102	1	1	Ja	1,385
Nicht eingeschlossene Variablen						
Vorlesungsbesuch	-	-	0,004	1	Nein	-
Übungsbesuch	-	-	1,258	1	Nein	-

Tabelle 29 zeigt die Koeffizienten des Regressionsmodells. In das Modell eingeschlossen wurde lediglich der Prädiktor „Ergebnis ILP“. Die beiden Prädiktoren Vorlesungsbesuch und Übungsbesuch weisen keinen signifikanten Einfluss auf die abhängige Variable auf.

Tabelle 30 Klassifikationstabelle Modell 3a

Beobachtet	Vorhergesagt		
	Prüfungsleistung		
	schlechte Leistung	gute Leistung	% korrekt
schlechte Leistung	27	12	69,2
gute Leistung	12	22	64,7
Gesamt % korrekt			67,1

Der Prozentsatz richtig klassifizierter Personen liegt bei 67,1% (vgl. Tabelle 30), und somit klar über der maximalen Zufallswahrscheinlichkeit von 53,4% und auch über der proportionalen Zufallswahrscheinlichkeit von 50,23%. Dies entspricht einem Zuwachs zwischen 13,7% und 16,87% richtiger Zuordnungen, einmal verglichen mit der maximalen und einmal verglichen mit der proportionalen Zufallswahrscheinlichkeit.

5.3.3.2 Regressionsmodell 3b

Hier wurde die Stichprobe ($n=73$) anhand der Variablen „Prüfung bestanden“ in die beiden Gruppen „Prüfung positiv absolviert“ und „Prüfung nicht positiv absolviert“ geteilt (vgl. Kapitel 5.2.1). Die logistische Regressionsgleichung wurde erstellt um die Zugehörigkeit zu diesen zwei Gruppen vorherzusagen.

Tabelle 31 Modellfit von Regressionsmodell 3b

-2LL-Wert	Cox & Snell R ²	Nagelkerke R ²
68,216	0,140	0,212

Wie in Tabelle 31 zu sehen ist, liegt der Cox & Snell R² Wert 0,140 also unter den geforderten 0,2 und der Nagelkerke R² Wert bei 0,212 also knapp über den geforderten 0,2. Insgesamt ist die Modellpassung hier mangelhaft.

Tabelle 32 Koeffizienten der logistischen Regression Modell 3b;

Das Signifikanzniveau wurde einheitlich bei $\alpha=0,05$ festgelegt

In das Modell eingeschlossene Variablen						
	B	Std. Fehler	Wald	df	Signifikant?	Exp(B)
Konstante	-1,736	0,979	3,147	1	Ja	0,176
Ergebnis ILP	0,339	0,115	8,707	1	Ja	1,403
Nicht eingeschlossene Variablen						
Vorlesungsbesuch	-	-	2,727	1	Nein	-
Übungsbesuch	-	-	1,001	1	Nein	-

Tabelle 32 zeigt die Koeffizienten des Regressionsmodells. In das Modell eingeschlossen wurde lediglich der Prädiktor „Ergebnis ILP“. Die beiden Prädiktoren Vorlesungsbesuch und Übungsbesuch weisen keinen signifikanten Einfluss auf die abhängige Variable auf.

Tabelle 33 Klassifikationstabelle Modell 3b

Vorhergesagt „Prüfung bestanden“			
Beobachtet	nicht bestanden	bestanden	% korrekt
nicht bestanden	5	12	29,4
bestanden	4	52	92,9
Gesamt % korrekt			78,1

Der Prozentsatz richtig klassifizierter Personen liegt bei 78,1% (vgl. Tabelle 33) und somit nur knapp über der maximalen Zufallswahrscheinlichkeit von 76,7% aber noch deutlich über der proportionalen Zufallswahrscheinlichkeit von 64,26%. Dies entspricht einem Zuwachs zwischen 1,4% und 13,84% richtiger Zuordnungen, einmal verglichen mit der

maximalen und einmal verglichen mit der proportionalen Zufallswahrscheinlichkeit. Auch für Modell 3b ist kritisch anzumerken, dass die Ausprägung „nicht bestanden“ der abhängige Variable „Prüfung bestanden“ nur 17 Fälle Aufweist. Backhaus et al. (2006) geben an, dass die Fallzahl pro Gruppe der abhängigen Variable zumindest 25 betragen soll.

5.3.4 Zusammenfassung

Obwohl die verschiedenen Erhebungszeitpunkte der einzelnen Variablen zu einer großen Anzahl an fehlenden Daten führten, waren die Stichprobengrößen für die Berechnung logistischer Regressionsmodelle noch ausreichend. Backhaus et al. (2006) empfiehlt für eine logistische Regression eine Stichprobengröße von 25 Fällen für jeden Regressor. Außer in den Modellen 2a und 2b ist die Stichprobengröße somit selbst für 3 Regressoren ausreichend.

Ein Grund für die in allen Modellen geringe Modelpassung ist die niedrige Anzahl von Prädiktoren (Regressoren). Neben den verwendeten Prädiktoren haben sicherlich auch weitere Faktoren einen wesentlichen Einfluss auf den Prüfungserfolg, die in dieser Untersuchung nicht erhoben wurden (z.B. Lernstrategien, Vorwissen, Tagesverfassung etc.).

Obwohl die Modelle häufig keinen ausreichenden Modellfit aufwiesen (Modelle 1a und 1b), konnten einige inhaltlich interessante Schlüsse aus den Daten gezogen werden. Es konnte nur der Prädiktor „Ergebnis ILP“ einen Teil der Varianz im Prüfungserfolg erklären und zwar für beide gewählte Formen der Variablen Prüfungserfolg. Dabei zeigte sich dass der minimale Erklärungswert für die Vorhersage ob Studierende die Prüfung positiv absolviert oder nicht (Modelle 2b und 3b) deutlich niedriger ist, nämlich zwischen 1,4% und 3,3% (verglichen mit der maximalen Zufallswahrscheinlichkeit, siehe Kapitel 5.3.2.2 und 5.3.3.2) als der Erklärungswert für die Vorhersage ob ein Student gut oder schlecht abschneidet, nämlich zwischen 13,7 und 20,7 (verglichen mit der maximalen Zufallswahrscheinlichkeit, siehe Kapitel 5.3.2.1 und 5.3.3.1). Obwohl ein Zusammenhang zwischen den Ergebnissen der ILP und dem Prüfungserfolg zu erwarten war, da relevante Einflussfaktoren wie Vorwissen und Vorbereitungszeit nicht erhoben wurden (M. Wagner-Menghin, persönliche Mitteilung, 22.01.2009), ist dieses Ergebnis insofern überraschend, als das Ergebnis der ILP als einziger Prädiktor bereits einen derartigen Zuwachs an Erklärungswert liefert.

Bemerkenswert ist, dass die Verbale Intelligenz, entgegen ursprünglicher Erwartungen (vgl. Kapitel 1) keinen Einfluss auf die Prüfungsleistung hat. Auch ein möglicher Einfluss des parallelen Besuchs der Übungen zur psychologischen Diagnostik (vgl. Kapitel 1) konnte nicht bestätigt werden.

Abschließend müssen auch noch methodisch kritische Anmerkungen an die berechneten Regressionsmodelle gemacht werden. So sind vor allem die Fallzahlen der Ausprägung „nicht bestanden“ der abhängigen Variable „Prüfung bestanden“ in den Modellen 2b und 3b zu gering. Dies mag eine Ursache für den relativ geringen Erklärungswert des Prädiktors „Ergebnis ILP“ für die abhängige Variable „Prüfung bestanden“ verglichen mit der abhängigen Variable Prüfungsleistung sein.

5.4 Selbsteinschätzung der Leistung

Im Anschluss an die ILP sowie nach der Prüfung zur psychologischen Diagnostik wurde mittels Fragebogen die Selbsteinschätzung der soeben erbrachten Leistung erhoben. Im Folgenden werden zuerst die Ergebnisse der Selbsteinschätzung deskriptiv betrachtet (Kapitel 5.4.1 & Kapitel 5.4.2). Danach erfolgt ein Vergleich der Selbsteinschätzung der bei der ILP erbrachten Leistung und der bei der Prüfung zur psychologischen Diagnostik erbrachten Leistung (Kapitel 5.4.3).

5.4.1 Selbsteinschätzung der Leistung der ILP

Im Rahmen des Zusatzfragebogens zur ILP wurde die Selbsteinschätzung der erbrachten Leistung der Studierenden erhoben. Die Studierenden mussten ihre Leistung auf einer 5-stufigen Skala einschätzen, die von „1-Sehr Gut“ bis „5-Nicht Genügend“ reicht. Dabei wurde zu jeder Note auch ein Punktwert zugeordnet um die Einschätzung zu erleichtern. Der Wortlaut der Frage und der möglichen Antworten ist in Abbildung 19 zu sehen.

Obwohl Ihre Leistung nicht beurteilt wird, würden wir gerne wissen wie Sie Ihre soeben erbrachte Leistung einschätzen.

- Sehr Gut, ich glaube zumindest 19 Fragen richtig beantwortet zu haben.
- Gut, ich glaube zumindest 16 Fragen richtig beantwortet zu haben.
- Befriedigend, ich glaube zumindest 13 Fragen richtig beantwortet zu haben.
- Genügend, ich glaube zumindest 10 Fragen richtig beantwortet zu haben.
- Nicht Genügend, ich glaube weniger als 10 Fragen richtig beantwortet zu haben.

Abbildung 19 Zusatzfrage 8 – Selbsteinschätzung

In Tabelle 34 werden die Häufigkeiten der Studierendenantworten dargestellt. Insgesamt lagen die Daten von 196 Studierenden vor. 3 Studierende haben zwar die Fragen der ILP ausgefüllt, jedoch nicht die Zusatzfragen beantwortet.

Tabelle 34 Selbsteinschätzung zur erbrachten Leistung in der ILP (Zusatzfrage 8)

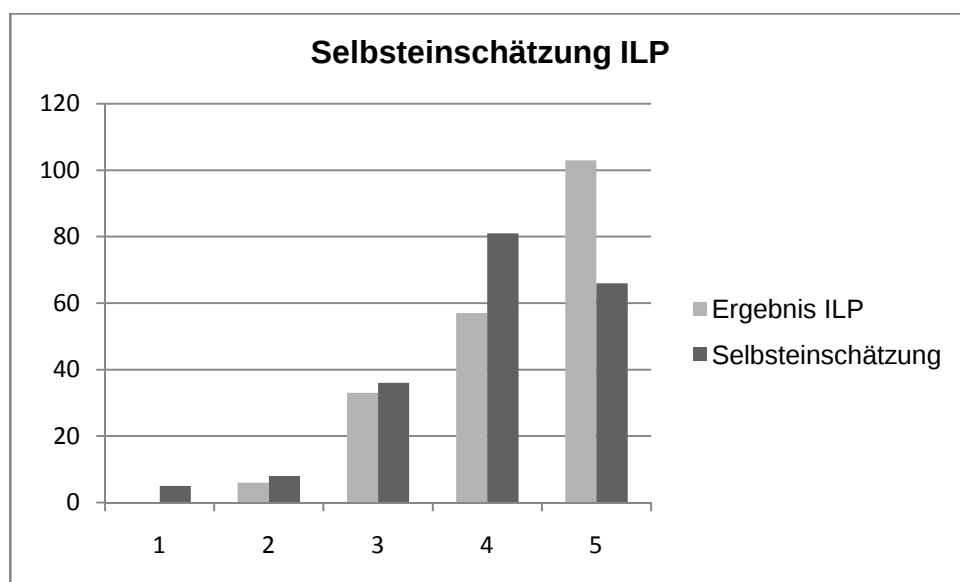
Einschätzung	Häufigkeit	Prozent
1 - Sehr Gut	5	2,6
2 – Gut	8	4,1
3 – Befriedigend	36	18,4
4 – Genügend	81	41,3
5 – Nicht Genügend	66	33,7
Gesamt	196	100

Um die Ergebnisse der ILP mit der Selbsteinschätzung besser vergleichen zu können, wurde das Ergebnis der ILP in eine ebenfalls 5-stufige Notenskala umgewandelt. Dies geschah analog der Punkteangaben der Zusatzfrage 8. 19 oder mehr Punkte entsprechen einem „Sehr gut“, zwischen 18 und 16 Punkte einem „Gut“ usw. Da in den Analysen nach dem Raschmodell allerdings zwei Items ausgeschlossen wurden, musste auch die Punkteanzahl für die Noten angepasst werden. Dabei wurde der Prozentsatz der erforderlichen Punkte bei 22 Items ermittelt und anschließend der entsprechende Punktwert bei 20 Items ermittelt. Die für die Umwandlung verwendeten Punktwerte und die Notenverteilung ist Tabelle 35 zu entnehmen.

Tabelle 35 Ergebnis der ILP in Schulnoten

Note	Punkte	Häufigkeit	Prozent
1 - Sehr Gut	17-20	0	0
2 – Gut	15-16	6	3
3 – Befriedigend	12-14	33	16,6
4 – Genügend	9-11	57	28,6
5 – Nicht Genügend	<8	103	51,8
Gesamt		199	100

Die Selbsteinschätzung soll nun in Relation zum tatsächlichen Ergebnis in der ILP gestellt werden. In einer rein visuellen Betrachtung der Ergebnisse zeigt sich, dass sich diese relativ gut decken, mit einer leichten Tendenz zur Überschätzung der eignen Leistung (vgl. Abbildung 20).

**Abbildung 20** Vergleich Selbsteinschätzung und tatsächlich erbrachte Leistung ILP.

Um die Güte der Selbsteinschätzung darzustellen wurde nun das tatsächliche Ergebnis der ILP in Relation zur Selbsteinschätzung gestellt und ein Abweichungswert berechnet. Dies geschah, indem die erwartete Leistung von der tatsächlichen Leistung¹³ subtrahiert wurde. Die erhaltene Variable ist folgendermaßen zu interpretieren: bei einem Wert von 0 decken sich Einschätzung und tatsächliche Leistung, bei einem negativen Wert wurde die eigene Leistung unterschätzt, bei einem positiven Wert überschätzt. Die erhaltenen Werte liegen im Intervall [-4; 4]. Mit Hilfe dieser Variablen ist es also auch möglich festzustellen ob die

¹³ jeweils als 5-stufige Notenskala

Studierenden ihre Leistung eher über- oder unterschätzten oder die Einschätzung mit der Leistung übereinstimmte.

Für die ILP lag Abweichungswert (Abweichung_ILP) der 196 Studierenden im Intervall [-2; 4]. Der Mittelwert der Einschätzung betrug 0,30 und die Standardabweichung 0,97. Bei Betrachtung der Verteilung der Abweichungswerte (vgl. Tabelle 36) sieht man, dass 42,9% der Studierenden ihre Leistung richtig eingeschätzt haben. Wenn man einen Abweichungswert von -1 bis 1 als guten Wert für die Selbsteinschätzung sieht, dann weisen 88,3% der Studierenden einen guten Wert auf, sie konnten also ihre erbrachte Leistung gut einschätzen.

Tabelle 36 Verteilung der Abweichungswerte für die ILP

Abweichungswert	Häufigkeit	Prozent
-2	3	1,5
-1	34	17,3
0	84	42,9
1	55	28,1
2	17	8,7
3	2	1
4	1	0,5
Gesamt	196	100

5.4.2 Selbsteinschätzung der Prüfungsleistung

Im Anschluss an die Prüfung zur Psychologischen Diagnostik 1 wurde mittels Fragebogen die Selbsteinschätzung der erbrachten Leistung erhoben. Die Studierenden mussten ihre Leistung auf einer 5-stufigen Skala einschätzen, die von „1-Sehr Gut“ bis „5-Nicht Genügend“ reicht. Der Wortlaut der Frage und der möglichen Antworten ist in Abbildung 21 zu sehen.

Wie schätzen Sie Ihre soeben erbrachte Leistung ein?

- Sehr Gut, ich glaube zumindest 19 Punkte erreicht zu haben.
- Gut, ich glaube zumindest 16 Punkte erreicht zu haben.
- Befriedigend, ich glaube zumindest 13 Punkte erreicht zu haben.
- Genügend, ich glaube zumindest 10 Punkte erreicht zu haben.
- Nicht Genügend, ich glaube weniger als 10 Punkte erreicht zu haben.

Abbildung 21 Fragebogen (FB1): Selbsteinschätzung der erbrachten Leistung

In Tabelle 37 werden die Häufigkeiten der Studierendenantworten dargestellt. Insgesamt lagen die Daten von 108 Studierenden vor. 10 Studierende haben zwar an der Prüfung zur Psychologischen Diagnostik 1 teilgenommen, jedoch nicht den Fragebogen ausgefüllt.

Tabelle 37 Selbsteinschätzung der erbrachten Prüfungsleistung (FB1)

Einschätzung	Häufigkeit	Prozent
1 - Sehr Gut	0	0
2 – Gut	7	6,5
3 – Befriedigend	33	30,6
4 – Genügend	59	54,6
5 – Nicht Genügend	9	8,3
Gesamt	108	100

Die entsprechenden Prüfungsergebnisse der Prüfung zur Psychologischen Diagnostik sind in Tabelle 38 dargestellt.

Tabelle 38 Prüfungsergebnisse der Prüfung zur Psychologischen Diagnostik

Note	Häufigkeit	Prozent
1 - Sehr Gut	3	2,5
2 – Gut	12	10,2
3 – Befriedigend	33	28,0
4 – Genügend	37	31,4
5 – Nicht Genügend	33	28,0
Gesamt	118	100

Die Selbsteinschätzung soll nun in Relation zum tatsächlichen Prüfungsergebnis gestellt werden. In einer rein visuellen Betrachtung der Ergebnisse zeigt sich, dass sich diese im

Notenbereich 1 bis 3 relativ gut decken. Auffallend ist, dass viele Studierende ihre Leistung als „Genügend“ einschätzen und kaum welche mit „Nicht genügend“ obwohl diese Gruppen bei der tatsächlichen Leistung annähernd gleich groß sind (vgl. Abbildung 22).



Abbildung 22 Vergleich von Selbsteinschätzung und tatsächlich erbrachter Prüfungsleistung.

Für die Prüfung zur Psychologischen Diagnostik lag der Abweichungswert (Abweichung_PT1) der 108 Studierenden im Intervall $[-3; 2]$. Der Mittelwert der Einschätzung betrug 0,08 und die Standardabweichung 1,21. Bei Betrachtung der Verteilung der Abweichungswerte (vgl. Tabelle 39) sieht man, dass 32,4% der Studierenden ihre Leistung richtig eingeschätzt haben. Wenn man einen Abweichungswert von -1 bis 1 als guten Wert für die Selbsteinschätzung betrachtet, dann weisen 78,7% der Studierenden einen guten Wert auf, sie konnten also ihre erbrachte Leistung gut einschätzen.

Tabelle 39 Verteilung der Abweichungswerte für die ILP

Abweichungswert	Häufigkeit	Prozent
-3	3	2,8
-2	6	5,6
-1	24	22,2
0	35	32,4
1	26	24,1
2	14	13,0
Gesamt	108	100

5.4.3 Vergleich der Selbsteinschätzung bei Prüfung und ILP

Für einen Vergleich der Selbsteinschätzung der erbrachten Leistung bei der ILP mit der Selbsteinschätzung der Prüfungsleistung war eine Teilstichprobe von Studierenden nötig, die sowohl die ILP inklusive Zusatzfragen ausgefüllt haben, als auch die Prüfung absolviert und den Fragebogen im Anschluss beantwortet haben. Diese Teilstichprobe umfasste 73 Studierende.

Der Mittelwert des Abweichungswerts der ILP beträgt in dieser Teilstichprobe 0,49 (vgl. Tabelle 40), die Studierenden überschätzten im Schnitt also ihre Leistung eher. Der Mittelwert des Abweichungswerts der Prüfung beträgt hingegen -0,12, die Studierenden unterschätzten im Schnitt also ihre Leistung eher. Dieser Unterschied ergab mittels T-Test für gepaarte Stichproben einen signifikanten Wert ($t(72, N = 73) = 3,541, p = 0,001$). Der Unterschied geht dabei in die Richtung, dass Studierende ihre Leistung bei der ILP eher überschätzen als ihre Prüfungsleistung.

Tabelle 40 Ergebnisse des Mittelwertvergleiches Abweichung_ILP und Abweichung_PT1

	Mittelwert	Std. Abweichung	t	df	Sig. ($\alpha=0,05$)
Abweichung_PT1	-0,12	1,29			
Abweichung_ILP	0,49	0,94			
Abweichung_PT1- Abweichung_ILP	-0,62	1,49	-3,541	72	.001

Um festzustellen ob sich Studierende, die ihre Leistung bei der ILP über- bzw. unterschätzten, dies entsprechend auch bei der Prüfung taten, wurden die Studierenden entsprechend dem Abweichungswert in Gruppen eingeteilt. Studierende mit einem Abweichungswert von -4 bis -2 wurden der Gruppe „Unterschätzung“, Studierende mit einem Abweichungswert von -1 bis 1 wurden der Gruppe „richtige Einschätzung“ und Studierende mit einem Abweichungswert von 2 bis 4 wurden der Gruppe „Überschätzung“ zugeordnet.

Tabelle 41 zeigt zu welcher Gruppe¹⁴ die Studierenden bei der ILP und der Prüfung zur psychologischen Diagnostik zuzuordnen waren. Dabei weichen die beobachteten

¹⁴ Unterschätzung, korrekte Einschätzung, Überschätzung

Häufigkeiten nicht signifikant¹⁵ von den zu erwartenden Häufigkeiten ab. Nur 2 Studierende überschätzen ihre Leistung sowohl bei der ILP als auch bei der Prüfung. 47 Studierende (64,38%) schätzten ihre Leistung sowohl bei der ILP als auch bei der Prüfung richtig ein. Bei der ILP unterschätzte kein Studierender dieser Teilstichprobe seine Leistung.

Tabelle 41 Kreuztabelle der Gruppenhäufigkeiten Selbsteinschätzung Prüfung x Selbsteinschätzung ILP;
n=73

		Prüfung			Gesamt
		Unterschätzung	korrekte Einschätzung	Überschätzung	
ILP	korrekte E.	9	47	7	63
	Überschätzung	0	8	2	10
Gesamt		9	55	9	73

5.4.4 Zusammenfassung Selbsteinschätzung

Die Betrachtung der Selbsteinschätzung der Studierenden bei der ILP zeigt, dass die Studierenden ihre erbrachte Leistung gut einschätzen können. Lediglich im unteren Leistungsbereich (Noten 4 und 5) gibt es größere Abweichungen zwischen der Selbsteinschätzung und der erbrachten Leistung. Studierende überschätzen hier eher ihre erbrachte Leistung. Insgesamt kann die Selbsteinschätzung der Studierenden allerdings als gut angesehen werden, 88,3% der Studierenden können ihre Leistung auf +/- einen Notengrad genau einschätzen.

Bei der Prüfung zur Psychologischen Diagnostik zeigt sich ein ähnliches Bild. Auch hier weicht im unteren Leistungsbereich (Noten 4 und 5) die Selbsteinschätzung deutlich von der erbrachten Leistung ab. Studierende überschätzen auch hier eher ihre erbrachte Leistung. Die Abweichung der Selbsteinschätzung von der erbrachten Leistung ist hier deutlich stärker als bei der ILP.

Für einen aussagekräftigen Vergleich wurde eine reduzierte Stichprobe verwendet, von Personen die sowohl ILP als auch die Prüfung absolviert und die Fragebögen ausgefüllt

¹⁵ (χ^2 (2, N = 73) = 2,015, p = 0,99)

haben. Hier zeigt sich, dass die Studierenden ihre Leistung bei der ILP eher überschätzen als ihre Leistung bei der Prüfung. Dieser Unterschied ist auch signifikant.

Weiter wurde überprüft ob Studierende, die ihre Leistung bei der ILP über- oder unterschätzten, dies auch bei der Prüfung taten. Hier zeigte sich, dass dies nicht überzufällig oft der Fall war.

5.5 Beantwortung der Fragestellungen

Die Beantwortung der in Kapitel 3 aufgestellten Fragestellungen ergibt sich aus den in den Kapiteln 5.1, 5.2, 5.3 und 5.4 dargestellten Ergebnissen. An dieser Stelle soll nun eine zusammenfassende Beantwortung der Fragestellungen erfolgen.

Fragestellung 1:

Gelten die Annahmen des Rasch-Modells für die Items der informellen Leistungsüberprüfung?

Die Ergebnisse der Analysen nach dem Raschmodell (vgl. Kapitel 5.1) weisen den Items der ILP gute Modelpassung aus. Nur zwei von 22 Items mussten ausgeschlossen werden.

Fragestellung 2:

Besteht ein Zusammenhang zwischen den Ergebnissen der informellen Leistungsüberprüfung und den Ergebnissen der Prüfung „Psychologische Diagnostik 1“?

Sowohl die berechneten Korrelationen aus Kapitel 5.2.6 sowie die Regressionsmodelle aus dem Kapitel 5.3 zeigen einen eindeutigen Zusammenhang zwischen dem Ergebnis in der ILP und dem späteren Prüfungsergebnis.

Fragestellung 3:

Welche weiteren Prädiktoren leisten einen Beitrag zur Vorhersage der Ergebnisse der Prüfung „Psychologische Diagnostik 1“?

Hier zeigt sich, dass keiner der in den Kapiteln 5.2 und 5.3 betrachteten Prädiktoren einen Beitrag zur Vorhersage der Prüfungsleistung erbringt. Nur der Prädiktor Übungsbesuch zeigt eine signifikante Korrelation mit der Prüfungsleistung, jedoch von geringer Stärke. In den Regressionsmodellen besteht dieser Zusammenhang dann nicht mehr.

Fragestellung 4:

Schätzen Studierende ihre erbrachte Leistung richtig ein oder neigen sie zu Über- oder Unterschätzung der erbrachten Leistung?

Die Ergebnisse von Kapitel 5.4 zeigen, dass insbesondere im unteren Leistungsbereich (Noten 4 und 5) eine Tendenz zur Überschätzung der erbrachten Leistung besteht. Dieser

Effekt ist bei der Prüfung zur Psychologischen Diagnostik tendenziell stärker ausgeprägt als bei der ILP. Insgesamt können 88,3% der Studierenden ihre Leistung bei der ILP und 78,7% der Studierenden ihre Leistung bei der Prüfung zur Psychologischen Diagnostik ihre Leistung auf +/- einen Notengrad genau einschätzen.

Fragestellung 5:

Gibt es Unterschiede in der Selbsteinschätzung der Studierenden zwischen der Prüfung „Psychologische Diagnostik 1“ und der ILP?

Es zeigen sich signifikante Unterschiede in der Selbsteinschätzung bei der ILP und bei der Prüfung zur Psychologischen Diagnostik. Bei der ILP überschätzen Studierende ihre erbrachte Leistung eher als bei der Prüfung zur Psychologischen Diagnostik.

6 Diskussion und Ausblick

Die Ergebnisse der Analysen nach dem Rasch-Modell zeigen, dass die Konstruktion eines Rasch-homogenen Itempools im Multiple-Choice Antwortformat für einen Wissenstest zur Psychologischen Diagnostik möglich ist. Von 22 Items mussten 2 Items ausgeschieden werden, was einem Prozentsatz von 9% entspricht. Dieses Ergebnis spricht sicherlich für das Vorgehen bei der Itemkonstruktion. Inhaltlich bewährte es sich analog den Lernzielen nach Bloom (1956) zu jedem gewünschten Lehrziel¹⁶ Items, nach einem vorher festgelegten Schlüssen zu konstruieren.

Im Rahmen der Raschanalysen als Problematisch, hat sich die Verwendung des Teilungskriteriums Bearbeitungszeit herausgestellt. Hier war auch nach dem Ausscheiden der beiden Items das Ergebnis des LQT nicht besonders „gut“ (hoher LR-Wert, vgl. Tabelle 8). Der erhaltene Wert war bei einem zuvor einheitlich festgelegten Risiko 1. Art von $\alpha=.01$ nicht signifikant, jedoch nur „knapp“. Auch war nach Ausschluss der Items das Item B1b im Grafischen Modelltest dahingehend auffällig (vgl. Kapitel 5.1.1.4), dass es für Personen mit geringerer Bearbeitungszeit relativ schwieriger war. Dies begründete sich dadurch, dass es sich bei diesem Item um ein Beispiel handelte das durch Nachschlagen im Lehrbuch einfach zu lösen war. Dies ist daher nicht unbedingt verwunderlich und spricht auch nicht gegen einen Einsatz dieses Items. Eine mögliche Erklärung ist, dass ein gewisser Teil der Studierenden sich nicht die Zeit genommen hat, nachzuschlagen. Es spielte also möglicherweise auch die Leistungsmotivation und/oder die Lösungsstrategie der Studierenden eine Rolle. Da ähnliche Aufgaben häufig Teil von Prüfungsfragen sind, und das Item in den restlichen Kriterien unauffällig ist, wurde es im Itempool belassen. Für einen möglichen weiteren Einsatz der ILP ist jedoch zu überlegen in wie weit man dieses Item, durch ein anderes, das zum Lehrziel passt, ersetzen kann oder es motivierender gestalten kann.

Wie in Kapitel 5.1.1.4 gezeigt, war die Modellpassung für das Teilungskriterium Bearbeitungszeit zwar ausreichend jedoch nicht besonders gut. Eine mögliche Erklärung dafür ist der Zusammenhang zwischen Anstrengung beim Beantworten der Fragen und der für die ILP aufgewandten Zeit. Zur Beantwortung von 22 Items und den Zusatzfragen

¹⁶ Homepage für „Psychologische Diagnostik“ des Instituts für Entwicklungspsychologie und Psychologische Diagnostik der Universität Wien; Stand Mai 2008

benötigten 50% der Studierenden unter 44 Minuten¹⁷. Hier kann man also davon ausgehen, dass ein guter Teil der Studierenden viele Fragen nur oberflächlich und manche eventuell gar nicht beantwortet hat. Ein besseres Vorgehen wäre gewesen keine Mindestbearbeitungszeit vorzugeben, dafür alle Personen, die eine gewisse Bearbeitungszeit unterschreiten, von den Analysen auszuschließen. In der aktuellen Untersuchung hätte dieses Vorgehen allerdings zu mangelhaften Stichprobengrößen geführt, weshalb diese Personen in die Analysen eingeschlossen werden mussten.

Die Ergebnisse der Regressionsanalysen sind aufschlussreich, obwohl die Modellpassung insgesamt mangelhaft ist. Aufgrund der sehr geringen Anzahl an Regressoren und der geringen Stichprobengröße (vgl. Kapitel 5.3.4) war keine besonders gute Modellpassung zu erwarten (M. Wagner-Menghin, persönliche Mitteilung, 22.01.2009). In der aktuellen Studie erklärt ausschließlich der Prädiktor „Ergebnis ILP“ einen Teil der Varianz im Prüfungserfolg. Dies lässt immerhin darauf schließen, dass sich ein Wissenstest im Multiple-Choice Antwortformat zur Leitungsfeststellung, im Rahmen der Prüfungsvorbereitung, für Studierende eignet. Der Prüfungserfolg durch eine Vielzahl, in der aktuellen Studie nicht erhobener Faktoren bedingt. Dadurch ist, ist die Größe der, durch den Prädiktor ILP erklärte Varianz als hoch einzustufen.

Es wurde ja schon eingangs erwähnt, dass Vermutungen bestanden, die verbale Intelligenz hätte Einfluss auf das Abschneiden bei der Prüfung zur Psychologischen Diagnostik (vgl. Kapitel 1). Ein solcher Einfluss konnte in der aktuellen Untersuchung nicht bestätigt werden, weder in den Korrelationsanalysen, noch im Regressionsmodell. Etwas überraschend war hingegen die signifikante Korrelation zwischen dem Ergebnis in der ILP und der verbalen Intelligenz nach IBF (vgl. Kapitel 5.2.6). Eine mögliche Erklärung für diese Korrelation bietet die Gestaltung der Untertests zur verbalen Intelligenz im IBF (vgl. Kapitel 4.2.4). Diese verwenden, wie die Items der ILP, das Multiple-Choice Antwortformat, was eine methodenspezifische Korrelation vermuten lässt. Trotzdem kann ein andersartiger Zusammenhang nicht ausgeschlossen werden und sollte noch genauer untersucht werden, denn für eine mögliche Gestaltung der Prüfung zur Psychologischen Diagnostik im Multiple-Choice Antwortformat ist diese Korrelation bedenklich.

¹⁷ Die deskriptiven Statistiken zur Bearbeitungszeit befinden sich in Anhang x.

Entgegen ursprünglichen Erwartungen konnten auch keine Zusammenhänge zwischen dem parallelen Besuch von Vorlesung und Übung zum Psychologischen Diagnostizieren gefunden. Es wurde zwar eine signifikante Korrelation zwischen dem Übungsbesuch und der Prüfungsleistung von $r=0,229$ festgestellt (vgl. Kapitel 6.2.6), in den Regressionsanalysen stellte sich dieser Zusammenhang jedoch als nicht signifikant heraus (vgl. Kapitel 6.3). Zwischen dem Übungsbesuch und der Variablen „Prüfung bestanden“ konnte auch keine signifikante Korrelation festgestellt werden. Bei der Betrachtung dieser Ergebnisse ist jedoch der geringe Stichprobenumfang zu beachten. Bei einem möglichen Einfluss war eine geringe Effektstärke zu erwarten, da wie in Kapitel 1 erläutert, viele relevante Einflussgrößen für den Prüfungserfolg in der aktuellen Studie nicht erfasst wurden. Um auch geringe Effektstärken festzustellen, ist wohl eine größere Stichprobe, als hier verwendet wurde, nötig. Für eine Folgeuntersuchung, beziehungsweise im Fall der Weiterführung der aktuellen Untersuchung, sollten unbedingt im Vorhinein notwendige Stichprobengrößen festgelegt werden, basierend auf den erwünschten Effektstärken.

Die Betrachtung der Selbsteinschätzung der erbrachten Leistung zeigte, dass ein großer Teil der Studierenden, 88,3% bei der ILP und 78,7% bei der Prüfung zur Psychologischen Diagnostik, ihre erbrachte Leistung auf +/- einen Notengrad genau einschätzen konnten. Die Frage, ob dieser Wert nun als hoch zu bewerten ist, also den Studierenden eine gute Fähigkeit zur Selbsteinschätzung ausspricht, ist freilich nur mit Hilfe von Vergleichen zu anderen Prüfungen zu beantworten.

Die Beobachtung, dass es im unteren Notenbereich (Noten 4 und 5) tendenziell zu einer Überschätzung der eigenen Leistung kommt deckt sich mit den theoretischen Annahmen aus Kapitel 2.2., dass aufgrund selektiver Wahrnehmung, die Auftretenswahrscheinlichkeit erwünschter Ergebnisse, hier das positive Absolvieren der Prüfung /ILP, signifikant überschätzt wird.

Der Vergleich der Selbsteinschätzung bei ILP und Prüfung ergab, dass Studierende ihre Leistung bei der ILP eher überschätzten als bei der Prüfung zur Psychologischen Diagnostik. Studierende konnten also ihre Leistung bei der Prüfung zur Psychologischen Diagnostik signifikant besser einschätzen als bei der ILP. Dies kann mehrere Ursachen haben, die in der aktuellen Studie allerdings nicht erklärend untersucht werden konnten. Eine mögliche Ursache für den Unterschied ist eine gewisse Prüfungsangst bei den Studierenden. Eingehend wurde ja erwähnt, dass Studierende die Prüfung zur

Psychologischen Diagnostik häufig als Hürde wahrnehmen. Dies kann auch bei der Einschätzung der Prüfungsleistung eine Rolle spielen, da möglicherweise einfach mit einer „strengen“ Benotung gerechnet wird, zumindest im Vergleich zur ILP. Eine weitere mögliche Erklärung liegt im unterschiedlichen Antwortformat, insbesondere im Multiple-Choice Antwortformat der ILP. Hier wurde ja das Antwortformat „x aus 5“ verwendet. Hier kann es häufig vorkommen, dass Studierende „teilrichtige“ Antworten geben, also nicht alle korrekten Antwortmöglichkeiten markieren. Dies kann den subjektiven Eindruck bei Studierenden auslösen, mehr Antworten gewusst zu haben, als tatsächlich der Fall war. Für eine Überprüfung dieser Erklärungsmöglichkeiten sind jedoch weitere Untersuchungen nötig.

7 Zusammenfassung

Die vorliegende Arbeit beschäftigte sich mit der Erstellung einer informellen Leistungsüberprüfung (ILP) im Multiple-Choice Antwortformat für Studierende zur Lehrveranstaltung Psychologische Diagnostik 1. Die Fragestellungen bezogen sich auf die testtheoretische Güte der Items der ILP und auf die prognostische Validität hinsichtlich der Prüfung zur Psychologischen Diagnostik 1, sowie dem Einfluss verschiedener anderer Faktoren auf das Prüfungsergebnis. Einerseits wurde überprüft ob die Items der ILP den Annahmen des Rasch-Modells entsprechen. Andererseits wurde der Zusammenhang der Ergebnisse der ILP mit dem Prüfungsergebnis, sowie der Einfluss weiterer Faktoren wie dem regelmäßigen Lehrveranstaltungsbesuch und der verbalen Intelligenz, untersucht. Die Fähigkeit der Studierenden ihre Prüfungsleistung einzuschätzen wurde ebenfalls erhoben.

Die Stichprobe bestand aus drei Teilstichproben, die alle aus der Population von Studierenden der Psychologie an der Universität Wien gezogen wurden. Die erste Teilstichprobe bestand aus Studierenden die an der ILP teilgenommen haben ($n=199$), die zweite aus Studierenden die im Rahmen der Erhebung von Stadler (2009) den Untertest Verbale Intelligenzfunktionen ausgefüllt haben ($n=196$) und die dritte Teilstichprobe bestand aus Studierenden, die die Prüfung zur psychologischen Diagnostik zum ersten Prüfungstermin am 27.6.2008 absolviert haben ($n=118$).

Die Analysen nach dem Rasch-Modell bestätigten für 20 der insgesamt 22 Items der ILP gute Modellpassung. Zur Überprüfung der Modellpassung wurden die Teilungskriterien Score, Geschlecht, Bearbeitungszeit, Interesse an dem Fach „Psychologische Diagnostik“, Vorlesungsbesuch und Übungsbesuch herangezogen. Dies bedeutet, dass es möglich ist einen Itempool im Multiple-Choice Format zur Psychologischen Diagnostik zu entwickeln, der den Kriterien des Rasch-Modells entspricht.

Zur Überprüfung der prognostischen Validität der ILP in Hinblick auf die Leistung bei der Prüfung zur Psychologischen Diagnostik 1 wurden in einem ersten Schritt die Korrelationen zwischen dem Ergebnis der ILP und der Prüfungsleistung erhoben. Hier wurden signifikante Zusammenhänge zwischen dem Ergebnis der ILP und dem Prüfungsergebnis festgestellt (0,350 und 0,324; vgl. Kapitel 5.2.6). In die Analysen wurden auch weitere mögliche Einflussfaktoren auf den Prüfungserfolg erhoben wie der regelmäßige Besuch der Vorlesung, der Besuch der Übungen und die verbale Intelligenz. Hierbei zeigten sich auch signifikante Zusammenhänge zwischen dem Übungsbesuch und

der Prüfungsleistung (0,229; vgl. Kapitel 5.2.6) sowie zwischen dem Ergebnis der ILP und der verbalen Intelligenz (0,280; vgl. Kapitel 5.2.6).

In einem zweiten Schritt wurden das Ergebnis der ILP sowie der Übungsbesuch, der Vorlesungsbesuch und die verbale Intelligenz in mehreren Regressionsmodellen als Regressoren für den Regressanden „Prüfungserfolg“ herangezogen. Da sich die 3 Stichproben nur zu einem geringen Teil überschneiden ($n=66$) wurden mehrere Modelle gerechnet. Über alle Modelle hinweg zeigte sich, dass allein das Ergebnis der ILP einen signifikanten Zusammenhang mit dem Prüfungserfolg aufweist.

Obwohl die Modellpassung der Regressionsmodelle mangelhaft ist, deutet dieses Ergebnis auf eine gewisse Validität der ILP, hinsichtlich der Prüfungsleistung hin. Dieses Ergebnis ist eine Voraussetzung für die Eignung der ILP als Hilfsmittel zur Leistungsbestimmung für Studierende.

Die Untersuchung der Fähigkeit der Studierenden ihre erbrachte Leistung einzuschätzen zeigte dass ein großer Teil der Studierenden, 88,3% bei der ILP und 78,7% bei der Prüfung zur Psychologischen Diagnostik, ihre erbrachte Leistung auf +/- einen Notengrad genau einschätzen konnten. Auch zeigte sich, dass die Studierenden ihre Leistung bei der Prüfung signifikant besser einschätzen konnten als bei der ILP. Bei der ILP tendierten die Studierenden eher dazu ihre Leistung zu überschätzen. Eine mögliche Erklärung liegt in der Aufgabenschwierigkeit der Items der ILP. Wie in Kapitel 2.2 erläutert, tendieren Personen bei schwierigen Aufgaben eher dazu ihre Leistung zu überschätzen. Wie die Analysen der ILP zeigen, war die Schwierigkeit der Fragen hier eher hoch anzusiedeln (vgl. Kapitel 5.2.2).

Zusammenfassend lässt sich die ILP sowohl in Hinblick auf ihre Rasch-Modell Konformität als auch aufgrund der Vorhersagevalidität als zufriedenstellend einstufen. Hauptkritikpunkte an der vorliegenden Studie sind sicher die mangelnde Stichprobengröße bei den Regressionsmodellen, sowie die geringe Anzahl an Regressoren die zur Verfügung standen. Auch für Durchführung der ILP gibt es Verbesserungsmöglichkeiten. Es ist für die Qualität der erhobenen Daten sicher nicht zuträglich die Testung in einem gewissen Maße zur „Verpflichtung“ zu machen. Dies führte sicher dazu, dass einige Studierende die Testung ohne jegliche Anstrengungsbereitschaft absolvierten. Auch die Vorgabe einer „Mindestbearbeitungszeit“ stellte sich als problematisch heraus, da diese Maßnahme die Compliance der Studierenden wohl nicht erhöhte, sondern eher als Bevormundung gesehen

wurde. Auch der Beobachtungszeitraum der Studie war äußerst gering angesetzt. Hier wäre eine längerfristige Betrachtung über mehrere Semester und die Berücksichtigung mehrere Prüfungstermine wünschenswert um die festgestellten Zusammenhänge abzusichern.

IV. Referenzen

8 Verzeichnisse

8.1 Literatur

- Amelang, M., Bartussek, D. (1997). *Differentielle Psychologie und Persönlichkeitsforschung*. Stuttgart: Kohlhammer
- Backhaus, K., Erichson, B., Plinke, W. & Weiber R. (2006). *Multivariate Analysemethoden – Eine anwendungsorientierte Einführung*. Berlin: Springer
- Blum, F., Didi, H.-J., Fay, E., Maichle, U., Trost, G., Wahlen, J.-H. & Gittler, G. (2005). *Intelligenz-Basis-Funktionen (IBF), Version 22.00*. Mödling: Dr.G.Schuhfried Ges.m.b.H.
- Bond, T.G. & Fox, C.M. (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences (Second Editioin)*. Mahwah, New Jersey; London: Erlbaum.
- Braun, M.W. (2003). *Genauigkeit der Selbsteinschätzung beim Erwerb neuer Kompetenzen in Abhängigkeit von Kontrollmeinung, Erfahrung, Selbstaufmerksamkeit, Ängstlichkeit und Geschlecht*. Unveröffentlichte Inauguraldissertation, Universität Bern.
- Fakultät für Psychologie, Arbeitsbereich Psychologische Diagnostik (2008). *Lehrziele Psychologische Diagnostik 1*. Internet: <http://psychologie.univie.ac.at/diagnostik/studium/lehrveranstaltungen/lehrziele/>, zugegriffen am 03.Februar 2008
- Fill-Giordano, R. & Litzenberger, M. (2005). *Lernen Anwenden - systematisch Ordnen (LAsO)*. Bozen: Eigenverlag.
- Gottschall, C.H. (2007). *Über die Itemschwierigkeiten von (Multiple-Choice-) Leistungstests in Abhängigkeit verschiedener Antwortformate – Ein Experiment zur Grundlagenforschung im Rahmen der Psychologischen Diagnostik*. Unveröffentlichte Diplomarbeit, Universität Wien.
- Kovacs, C. (2007). *Zur Vorhersagekraft eines Auswahlverfahrens hinsichtlich des Studienerfolgs im ersten Studienjahr*. Unveröff. Diplomarbeit, Universität Wien.
- Kubinger, K. D. (Hrsg.) (1989). *Moderne Testtheorie - Ein Abriß samt neuesten Beiträgen*. Weinheim: PVU.
- Kubinger, K. D. & Jäger, R. S. (2003). *Schlüsselbegriffe der Psychologischen Diagnostik*. Weinheim: Beltz.
- Kubinger, K. D. (2006). *Psychologische Diagnostik - Theorie und Praxis psychologischen Diagnostizierens*. Göttingen: Hogrefe.
- Kubinger, K. D. & Ebenhöf, J. (1996). *Arbeitshaltungen (AHA)*. Frankfurt a.M.: Swets.
- Litzenberger, M., Gnambs, T. & Punter, J. F. (2005). *Richtlinien zur Konstruktion von Multiple-Choice Items am Beispiel der Vorlesung Psychologie als Wissenschaft (Ringvorlesung) an der Fakultät für Psychologie*. Unveröff. Projektbericht, Universität Wien.
- Owen, W.J. (2006). *The R Guide. Version 2.2*. [Online im Internet]. URL: <http://mathcs.richmond.edu/~wowen/TheRGuide.pdf> [26.09.2008]

- Poinstingl, H., Mair, P. & Hatzinger, R. (2007). *Manual zum Softwarepackage eRM (extended Rasch modeling): Anwendung des Rasch-Modells (1-PL Modell)*. Lengerich, Wien: Pabst Publisher.
- Rammstedt, B. (1997). *Die deutsche Version des Big Five Inventory (BFI). Übersetzung und Validierung eines Fragebogens zur Erfassung des Fünf-Faktoren-Modells der Persönlichkeit*. Unveröffentlichte Diplomarbeit, Universität Bielefeld.
- Rasch, D. & Kubinger, K. D. (2006). *Statistik*. Heidelberg: Spektrum.
- Rasch, D., Kubinger, K. D., Schmidtke, J. & Häusler, J. (2004). *The misuse of asterisks in hypothesis testing*. Psychology Science, Volume 46 (2), p. 227-242.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion* (2. Aufl.). Bern: Huber.
- Thurstone, L. L. & Thurstone T. G. (1941). *Factorial studies of intelligence*. Psychometrics Monographs, 2.
- Wagner-Menghin, M. (1998). *Lexikon-Wissen-Test (LEWITE)*. Wien: Universität, Institut für Psychologie, Arbeitsbereich Psychologische Diagnostik.
- Wright, B.D. & Masters G.N. (1982) *Rating Scale Analysis*. Chicago: Mesa.

8.2 Abbildungen

Abbildung 1 Item A2; die korrekte Antwort ist unterstrichen.	16
Abbildung 2 Beispielaufgabe der Verbalen Intelligenzfunktionen - Satzergänzungen	20
Abbildung 3 Beispielaufgabe der Verbalen Intelligenzfunktionen – Analogien	20
Abbildung 4 Modellgleichung des Rasch-Modells	25
Abbildung 5 Grafischer Modelltest vor dem Ausschluss, Teilungskriterium Score.....	33
Abbildung 6 Grafischer Modelltest nach dem Ausschluss, Teilungskriterium Score.....	34
Abbildung 7 Grafischer Modelltest vor dem Ausschluss, Teilungskriterium Geschlecht	35
Abbildung 8 Grafischer Modelltest nach dem Ausschluss, Teilungskriterium Geschlecht	36
Abbildung 9 Grafischer Modelltest vor dem Ausschluss, Teilungskriterium Bearbeitungszeit	38
Abbildung 10 Grafischer Modelltest nach dem Ausschluss, Teilungskriterium Bearbeitungszeit	38
Abbildung 11 Grafischer Modelltest vor dem Ausschluss, Teilungskriterium Interesse.....	40
Abbildung 12 Grafischer Modelltest nach dem Ausschluss, Teilungskriterium Interesse.....	40
Abbildung 13 Grafischer Modelltest vor dem Ausschluss. Teilungskriterium Vorlesungsbesuch.	42
Abbildung 14 Grafischer Modelltest nach dem Ausschluss, Teilungskriterium Vorlesungsbesuch	42
Abbildung 15 Grafischer Modelltest vor dem Ausschluss, Teilungskriterium Übungsbesuch.....	44
Abbildung 16 Grafischer Modelltest nach dem Ausschluss, Teilungskriterium Übungsbesuch	44
Abbildung 17 Ergebnis ILP. Md=9; n=79	48
Abbildung 18 Verbale Intelligenz lt. IBF; Md=60,03; n=70	49
Abbildung 19 Zusatzfrage 8 – Selbsteinschätzung	63
Abbildung 20 Vergleich Selbsteinschätzung und tatsächlich erbrachte Leistung ILP.....	64
Abbildung 21 Fragebogen (FB1): Selbsteinschätzung der erbrachten Leistung.....	66
Abbildung 22 Vergleich von Selbsteinschätzung und tatsächlich erbrachter Prüfungsleistung.	67

8.3 Tabellen

Tabelle 1 Lehrziele der Lehrveranstaltung Psychologische Diagnostik 1.....	14
Tabelle 2 Zuordnung der Items zu den Lehrzielen und Iteminhalten.....	15
Tabelle 3 Ausgeschiedene Items in Bezug auf die jeweiligen Teilungskriterien	30
Tabelle 4 Infit- und Outfit-Werte vor Ausschluss der Items	31
Tabelle 5 Infit- und Outfit-Werte nach Ausschluss der Items A3 und C1.....	32
Tabelle 6 Ergebnisse des LQT für das Teilungskriterium Score.....	33
Tabelle 7 Ergebnisse des LQT für das Teilungskriterium Geschlecht	35
Tabelle 8 Ergebnisse des LQT für das Teilungskriterium Bearbeitungszeit	37
Tabelle 9 Ergebnisse des LQT für das Teilungskriterium Interesse.....	39
Tabelle 10 Ergebnisse des LQT für das Teilungskriterium Vorlesungsbesuch.....	41
Tabelle 11 Ergebnisse des LQT für das Teilungskriterium Übungsbesuch	43
Tabelle 12 Noten der Prüfung psychologische Diagnostik 1 (Md=4)	46
Tabelle 13 Dichotomisierte Prüfungsergebnisse	47
Tabelle 14 Teilnahme an der ILP unter den Teilnehmern des 1. Prüfungstermins	47
Tabelle 15 Daten zur Verbalen Intelligenz.....	49
Tabelle 16 Angaben zum Vorlesungsbesuch im Sommersemester 2008, erhoben	51
Tabelle 17 Besuch der Übungen zum psychologischen Diagnostizieren Sommersemester 2008... 51	51
Tabelle 18 Korrelationsmatrix der Prädiktoren und des Studienerfolges.....	52
Tabelle 19 Modellfit von Regressionsmodell 1a.....	54
Tabelle 20 Koeffizienten der logistischen Regression Modell 1a	54
Tabelle 21 Koeffizienten der logistischen Regression Modell 1b.....	55
Tabelle 22 Modellfit von Regressionsmodell 2a.....	55
Tabelle 23 Koeffizienten der logistischen Regression Modell 2a	56
Tabelle 24 Klassifikationstabelle Modell 2a	56
Tabelle 25 Modellfit von Regressionsmodell 2b.....	57
Tabelle 26 Koeffizienten der logistischen Regression Modell 2b.....	57
Tabelle 27 Klassifikationstabelle Modell 2b	57
Tabelle 28 Modellfit von Regressionsmodell 3a.....	58
Tabelle 29 Koeffizienten der logistischen Regression Modell 3a	59
Tabelle 30 Klassifikationstabelle Modell 3a	59
Tabelle 31 Modellfit von Regressionsmodell 3b.....	60
Tabelle 32 Koeffizienten der logistischen Regression Modell 3b.....	60
Tabelle 33 Klassifikationstabelle Modell 3b	60
Tabelle 34 Selbsteinschätzung zur erbrachten Leistung in der ILP (Zusatzfrage 8)	63
Tabelle 35 Ergebnis der ILP in Schulnoten.....	64

Tabelle 36 Verteilung der Abweichungswerte für die ILP	65
Tabelle 37 Selbsteinschätzung der erbrachten Prüfungsleistung (FB1)	66
Tabelle 38 Prüfungsergebnisse der Prüfung zur Psychologischen Diagnostik.....	66
Tabelle 39 Verteilung der Abweichungswerte für die ILP	67
Tabelle 40 Ergebnisse des Mittelwertvergleiches Abweichung_ILP und Abweichung_PT1	68
Tabelle 41 Kreuztabelle der Gruppenhäufigkeiten Selbsteinschätzung	69

9 Anhang

9.1 Quellen

Ich habe mich bemüht, sämtliche Inhaber der Bildrechte ausfindig zu machen und ihre Zustimmung zur Verwendung der Bilder in dieser Arbeit eingeholt. Sollte dennoch eine Urheberrechtsverletzung bekannt werden, ersuche ich um Meldung bei mir.

9.2 Items der ILP

Die als korrekt verrechneten Antworten sind unterstrichen. Die Reihenfolge der Items entspricht der Reihenfolge in der sie vorgegeben wurden. Die Reihenfolge der Antwortmöglichkeiten wurde bei der Vorgabe für jeden Studierenden randomisiert, sie entspricht also nicht der hier dargestellten Reihenfolge.

C2

Sie sollen bei einem Test zum Alltagswissen mit freiem Antwortformat, die Verrechnungssicherheit erhöhen, da Studien zeigten, dass diese noch mangelhaft ist. Welche Möglichkeiten stehen Ihnen zur Verfügung?

1. Sie ändern das Antwortformat hin zu Multiple-Choice-Format
2. Sie geben im Manual Antwortbeispiele für den Testleiter
3. Sie stellen den Testleitern ein Computerprogramm als Auswertungshilfe zur Verfügung
4. Sie lassen die Testdurchführung und die Auswertung von unterschiedlichen Personen durchführen
5. Sie geben zukünftig alle Instruktionen schriftlich

B1b

Für so genannte Z-Werte oder Standardwerte gilt $Z = z \cdot x + y$. Wie lauten die Variablen x und y ?

1. $x=100$
2. $y=100$
3. $x=10$
4. $y=10$
5. $x=50$

D3

Für eine ausgeschriebene Stelle werden im Anforderungsprofil folgende Eigenschaften von einem möglichen Kandidaten gefordert: die Fähigkeit zum schlussfolgernden Denken, sowie gute Konzentrationsfähigkeit. Welche der nachstehenden Arten psychologisch diagnostischer Verfahren eignen sich zur Erhebung dieser Konstrukte unter Berücksichtigung des Faktors Ökonomie?

1. Intelligenztestbatterie
2. Arbeitsprobe
3. Persönlichkeitsfragebogen
4. Spezieller Leistungstest
5. Strukturiertes Interview

B1a

Ein in einem Leistungstest erreichter Prozentrang von 64 bedeutet, dass...

1. 64% der Vergleichspopulation ein besseres oder zumindest gleich gutes Ergebnis erreicht haben
2. 36% der Vergleichspopulation ein schlechteres oder maximal gleich gutes Ergebnis erreicht haben
3. Die Testperson 64% der Aufgaben gelöst hat
4. 64% der Vergleichspopulation ein schlechteres oder maximal gleich gutes Ergebnis erreicht haben
5. 36% der Vergleichspopulation ein besseres oder zumindest gleich gutes Ergebnis erreicht haben

C3

Ein Kollege möchte die Validität eines neuen Reasoning Tests (NRT) überprüfen. Dazu bestimmt er die Korrelationen des NRT mit einem gängigen Reasoning Test (SPM) und einem konstruktfernen Test (LEWITE). Welche der nachstehenden Ergebnisse würden zumindest für eine hohe Konstrukt-Validität sprechen?

1. Eine hohe Korrelation zwischen dem NRT und dem SPM
2. Eine hohe Korrelation zwischen dem NRT und dem LEWITE
3. Keine oder eine niedrige Korrelation zwischen dem NRT und dem SPM
4. Eine Korrelation zwischen NRT und LEWITE die bedeutend größer als 0 ist.
5. Eine Korrelation zwischen NRT und SPM die bedeutend größer als 0 ist.

A2

Sie erhalten den Auftrag für ein großes Unternehmen die Lehrlingsauswahl durchzuführen. Es sollen im Schnitt 20 Kandidaten pro Woche begutachtet werden. Wie sieht Ihr unmittelbar nächster Schritt, nach erfolgter Formulierung der Fragestellung, aus?

1. Sie führen eine Tätigkeitsanalyse durch
2. Sie klären ob ein Anforderungsprofil erstellt werden muss
3. Sie lassen die Kandidaten eine Arbeitsprobe ablegen
4. Sie wählen die einzusetzenden psychologisch diagnostischen Verfahren aus
5. Sie führen computergestützte Gruppentestungen durch

E1

Welche der nachstehenden Intelligenz-Testbatterien sind für Schulkinder zwischen 13 und 15 Jahren geeignet?

1. Adaptives Intelligenz Diagnostikum 2 (AID2)
2. Hamburg-Wechsler Intelligenztest für Kinder Dritte Auflage (HAWIK III)
3. Intelligenz-Struktur-Test 2000-R (IST 2000-R)
4. Kaufman Assessment Battery for Children –Deutsche Version (KABC)
5. Prüfsystem für Schul- und Bildungsberatung (PSB 6-13)

B3

2 Kinder werden mit dem gleichen Leistungstest von unterschiedlichen Testleitern getestet. Ein Testleiter gibt für Kind A einen IQ von 100 an, der andere Testleiter bevorzugt T-Werte und gibt für Kind B als Ergebnis einen T-Wert von 43 an. Welche der folgenden Aussagen treffen zu wenn die Reliabilität des Tests 0,86 beträgt?

1. Kind B verfügt über eine höhere Fähigkeit als Kind A
2. Kind A verfügt über eine höhere Fähigkeit als Kind B
3. Kind A verfügt über überdurchschnittliche Fähigkeit
4. Kind A und B unterscheiden sich nicht bezüglich ihrer Fähigkeit
5. Kind B verfügt über unterdurchschnittliche Fähigkeit

F2

Welche Schwierigkeiten treten im Zusammenhang mit „Speed und Power“ Testungen auf?

1. Sie stehen im Widerspruch zur Eindimensionalität von Tests
2. Sie benachteiligen Personen mit besonders niedrigen Fähigkeiten
3. Schlechte Leistungen sind nicht eindeutig zu interpretieren
4. Gute Leistungen sind nicht eindeutig zu interpretieren

5. Sie benachteiligen Personen mit besonders hohen Fähigkeiten

E2

Sie sollen für ein großes Unternehmen die Lehrlingsauswahl durchführen. Als Anforderungen werden gute Konzentrationsfähigkeit sowie zumindest durchschnittliche Fähigkeit zum schlussfolgernden Denken angegeben. Welche der nachstehenden Verfahren würden Sie auswählen, wenn Sie täglich ca. 15 Kandidaten testen sollen, also vor allem Ökonomie in Vorgabe und Auswertung gefordert ist.

1. Inventar Komplexer Aufmerksamkeit (INKA)
2. Intelligenz-Struktur-Test 2000-R (IST 2000-R)
3. Cognitrone
4. Grundintelligenz-Skala 2 (CFT 20-R)
5. Adaptiver Matrizen Test (AMT)

D1

Welche der nachstehenden Aussagen treffen auf einen Test nach Lienert zu?

1. Ein Test hat eine quantitative Aussage über ein Persönlichkeitsmerkmal zum Ziel
2. Er ist ein stets objektives wissenschaftliches Verfahren
3. Er erhebt unter standardisierten Bedingungen Informationen über eine Person
4. Ein Test ist ein wissenschaftliches Routineverfahren
5. Er folgt dem Prinzip der Manipulation, analog dem Experiment

F3

Zur österreichweiten Überprüfung der kognitiven Fertigkeiten von Schülern mit mangelhaften schulischen Leistungen erstellt ein Kollege eine Testbatterie. Um die große Anzahl geforderter Testungen bewältigen zu können, legt er besonderen Wert auf die Ökonomie und gibt ausschließlich Gruppenverfahren vor.

Die Festlegung auf Gruppenverfahren...

1. führt zu hoher Testleiterunabhängigkeit
2. benachteiligt Testpersonen mit geringen kognitiven Fertigkeiten.
3. benachteiligt Testpersonen mit geringer Lesefertigkeit.
4. birgt die Gefahr der Störung der Testpersonen durch die anderen Testpersonen.
5. ermöglicht das Abschreiben der Testpersonen voneinander.

H2

Das Rasch-Modell benötigt man um...

1. Tests verrechnungsfair zu konstruieren
2. die Fähigkeiten zweier Personen unabhängig von den dazu verwendeten Items zu vergleichen
3. die Schwierigkeit zweier Items unabhängig von der verwendeten Stichprobe zu vergleichen.
4. adaptive Leistungstests zu konstruieren.
5. Ein Konfidenzintervall unabhängig von der verwendeten Stichprobe zu bestimmen

B2

Welche Reliabilität muss ein Test aufweisen, dass bei einem T-Wert von 47 eine durchschnittliche Leistung mit 95% Wahrscheinlichkeit, als gegeben angenommen werden kann?

1. $r=0.96$
2. $r=0.90$
3. $r=0.93$
4. $r=0.99$
5. $r=0.87$

A1

An welcher Stelle im psychologisch-diagnostischen Prozess erfolgt üblicherweise die Erstellung, bzw. Festlegung eines Anforderungsprofils?

1. *an erster Stelle im psychologisch-diagnostischen Prozess*
2. *nach der Anwendung und Auswertung der psychologisch diagnostischen Verfahren*
3. *nachdem die Fragestellung geklärt wurde*
4. *nach der Auswahl der psychologisch diagnostischen Verfahren*
5. *vor der Klärung der Fragestellung*

H1

Wovon hängt die Lösungswahrscheinlichkeit eines Items nach dem Rasch-Modell ab?

1. *Von der Lösungshäufigkeit in der Normstichprobe*
2. *Von der Fähigkeit der Person*
3. *Vom Standardschätzfehler*
4. *Von der Schwierigkeit des Items*

5. Von der Ratewahrscheinlichkeit

F1

Welche Charakteristiken weisen Computer-Simulationen auf?

1. Sie eignen sich insbesondere zur Überprüfung intellektueller Fähigkeiten.
2. Sie eignen sich insbesondere zur Überprüfung organisatorischer Fähigkeiten.
3. Sie versuchen das Verhalten der Testperson in möglichst realistischen Aufgaben zu erfassen.
4. Sie sind Voraussetzung für objektive Persönlichkeitstests.
5. Sie sind notwendig um Verrechnungssicherheit zu gewährleisten.

H3

Bei der Analyse der Items eines neuen Raumvorstellungstests liegen die Lösungshäufigkeiten der Items im Intervall $[0,43;0,97]$.

Welche Aussagen können Sie aufgrund dieser Werte treffen?

1. Es fehlen leichte Items, daher misst der Test bei leistungsstarken Personen ungenau.
2. Es fehlen schwierige Items, daher misst der Test bei leistungsstarken Personen ungenau.
3. Es fehlen leichte Items, daher misst der Test bei leistungsschwachen Personen ungenau.
4. Es fehlen schwierige Items, daher misst der Test bei leistungsschwachen Personen ungenau.
5. Diese Itemauswahl benachteiligt leistungsschwache Personen.

D2

Was sind die Vorteile der Verhaltensbeobachtung?

1. Sie ermöglicht die vollständige Erfassung des beobachteten Geschehens
2. Es wird reales Verhalten erfasst
3. Die Übereinstimmung der Urteile verschiedener Beobachter ist gegeben
4. Hohe Auswertungsobjektivität
5. Sie kann als Außenkriterium für Persönlichkeitsfragebögen dienen

A3

In der Praxis eines Psychologen meldet sich telefonisch eine Mutter wegen ihrer 9 jährigen Tochter. Die Mutter vermutet, dass ihre Tochter Legasthenikerin ist. Die Psychologin verfügt über sehr viel Erfahrung im Bereich Legasthenie und vereinbart einen Testtermin mit dem Kind. Dabei gibt sie eine bewährte „Testbatterie“ vor, bestehend aus dem PSB-R 4-6 und eine Lese- und Rechtschreibtest. Wie beurteilen Sie das Vorgehen der Kollegin?

1. Das Verfahren PSB-R 4-6 ist nicht geeignet da keine Profilinterpretation möglich ist.
2. Ein Lese- und Rechtschreibtest ist nicht nötig, falls der AID 2 anstatt des PSB-R 4-6 verwendet wird.
3. Die Vorgabe der Testbatterie ist verfrüht, da die Klärung der Fragestellung noch nicht erfolgt ist.
4. Die Vorgabe der Batterie ist ein passendes Vorgehen, da die Kollegin bereits über sehr viel Erfahrung mit Legasthenie verfügt.
5. Das Vorgehen der Psychologin ist passend, da die Fragestellung ja von der Mutter bereits vorgegeben wurde.

C1

Unter Kriteriums-Validität versteht man...

1. dass eine Testperson erkennt, was der Test misst.
2. die Vorhersagekraft eines Tests.
3. die Übereinstimmung des Tests mit einem anderen Test der dasselbe misst.
4. die Erfüllung gewisser theoriegeleiteter Vorstellungen.
5. dass das Testverhalten selbst das optimale Kriterium darstellt.

E3

Welche Eigenschaften des AID 2 ermöglichen die Diagnose von Teilleistungsstörungen?

1. Der Verzicht auf die Berechnung eines IQ beim AID2
2. Die ökonomische Vorgabe des AID2
3. Dass der AID 2 adaptiv in seiner Vorgabe ist
4. Die Möglichkeit der Profilinterpretation
5. Die mögliche Betrachtung von Leistungshöhen und Tiefen

9.3 Zusatzfragen zur ILP

Zusatzfrage 1

Wie weit ist Ihre Vorbereitung für die Prüfung abgeschlossen?

1. 0-20%
2. 21-40%
3. 41-60%
4. 61-80%
5. 81-100%

Zusatzfrage 2

Wann werden sie die Prüfung voraussichtlich absolvieren?

1. Ende Sommersemester 2008
2. Anfang Wintersemester 2008/2009
3. Im weiteren Verlauf des Wintersemester 2008/2009
4. Nach dem Wintersemester 2008/2009
5. Ich habe die Prüfung bereits positiv absolviert

Zusatzfrage 3

Wie hoch würden Sie ihre Bemühung einschätzen, die vorangegangenen Fragen richtig zu beantworten?

1. sehr hoch. Bei mehr als 16 Aufgaben habe ich mich bemüht die Lösung zu finden.
2. hoch. Bei 12-15 Aufgaben habe ich mich bemüht die Lösung zu finden.
3. mittelmäßig. Bei 8-11 Aufgaben habe ich mich bemüht die Lösung zu finden.
4. niedrig. Ich habe die Fragen und Antworten angesehen, bei weniger als 7 Aufgaben habe ich mich bemüht die Lösung zu finden).
5. sehr niedrig. Ich habe die Fragen und Antworten angesehen und probierhalber auch auf Antworten geklickt, mich dabei aber kaum bemüht die richtige Antwort zu finden.

Zusatzfrage 4

Wie regelmässig haben Sie dieses Semester die Vorlesung zu psychologischen Diagnostik 1 besucht?

1. immer (bisher 8 Einheiten)
2. meistens (6-7 mal)
3. gelegentlich (3-5 mal)
4. selten (1-2 mal)
5. nie
6. Ich habe die Vorlesung zu psychologischen Diagnostik 1 bereits vor dem Sommersemester 2008 besucht.

Zusatzfrage 5

Wie stark ist Ihr Interesse an der psychologischen Diagnostik?

1. *sehr hoch*
2. *hoch*
3. *durchschnittlich*
4. *niedrig*
5. *sehr niedrig*

Zusatzfrage 6

Geben Sie bitte an, wie sehr sie sich bei der heutigen Testung angestrengt haben:

Von meiner gesamt möglichen Leistungskapazität habe ich [Prozent] (0-100) Prozent gegeben.

Zusatzfrage 7

Haben Sie während der Bearbeitung der Fragen eine Unterbrechung oder eine Pause gemacht?

Wenn ja wie lange dauerten die Unterbrechungen insgesamt?

1. *unter 15 min.*
2. *16-30 min*
3. *31-45 min*
4. *46-60 min*
5. *über 60 min.*
6. *Ich habe keine Unterbrechung gemacht.*

Zusatzfrage 8

Obwohl Ihre Leistung nicht beurteilt wird, würden wir gerne wissen wie Sie Ihre soeben erbrachte Leistung einschätzen.

1. *Sehr Gut, ich glaube zumindest 19 Fragen richtig beantwortet zu haben.*
2. *Gut, ich glaube zumindest 16 Fragen richtig beantwortet zu haben.*
3. *Befriedigend, ich glaube zumindest 13 Fragen richtig beantwortet zu haben.*
4. *Genügend, ich glaube zumindest 10 Fragen richtig beantwortet zu haben.*

5. *Nicht Genügend, ich glaube weniger als 10 Fragen richtig beantwortet zu haben.*

9.4 Nachrichten an die Studierenden via Blackboard Vista

Nachricht 1

Betreff: Informelle Leistungsüberprüfung

01. Juni 2008 18:00

Liebe Studentinnen und Studenten! Ab Montag den 02.06.2008 besteht die Möglichkeit einen Selbsttest durchzuführen, der bei der Prüfungsvorbereitung behilflich sein soll. Nach Ablauf der Verfügbarkeit (09.06.2008) können Sie die Ergebnisse mit einem allgemeinen Feedback einsehen. Den Selbsttest finden sie unter "Kursinhalt" oder unter "Tests". Wir hoffen, dass die Bearbeitung der informellen Leistungsüberprüfung für euch hilfreich ist und wünschen weiters viel Erfolg bei der Prüfung!

Nachricht 2

Betreff: Feedback zur informellen Leistungsüberprüfung

10. Juni 2008 21:27

Liebe Studentinnen und Studenten! Da die Verfügbarkeit der informellen Leistungsüberprüfung bis Freitag verlängert wurde, könnt ihr eure Ergebnisse dann ab Freitag den 13.06. ab 12:00 Uhr einsehen. Für alle die die informelle Leistungsüberprüfung also noch nicht ausgefüllt haben, besteht dazu noch die Möglichkeit bis eben Freitagmittag.

9.5 Einleitende Beschreibung der ILP

Die folgende Beschreibung war für die Studierenden zu lesen, bevor sie mit der Bearbeitung der Items der ILP begannen.

Es handelt sich hier um eine informelle Leistungsüberprüfung, die NICHT benotet wird, sondern Ihnen bei der Prüfungsvorbereitung behilflich sein soll.

Bitte arbeiten Sie bei diesem Test so, wie Sie bei der Prüfung vorgehen werden, versuchen Sie also möglichst keine Pausen zu machen.

Sie dürfen alle Unterlagen verwenden, die auch bei der Prüfung erlaubt sind.

Arbeiten Sie alleine. Sie haben die Möglichkeit die Fragen und Ergebnisse nach der Bearbeitung (ab 08.06.) mit Ihren Kollegen zu diskutieren!

Bei den folgenden Fragen sind immer zwischen einer und fünf Antworten richtig (1-5 aus 5).

Die MINDESTBEARBEITUNGSZEIT beträgt 10 Minuten, erst nach Ablauf dieser Zeit, zählt der Versuch für die Testerfahrung.

Viel Erfolg beim Bearbeiten der Fragen!

9.6 Fragebogen Prüfung

Schriftart und Formatierung wurde beibehalten. Der Fragebogen befindet sich auf der nächsten Seite.

Matrikelnummer: _____

Bitte Füllen Sie die nachstehenden Fragen aus nachdem die Prüfungsbögen abgesammelt wurden

Wie schätzen Sie Ihre soeben erbrachte Leistung ein?

- ☐ Sehr Gut, ich glaube zumindest 19 Punkte erreicht zu haben.
- ☐ Gut, ich glaube zumindest 16 Punkte erreicht zu haben.
- ☐ Befriedigend, ich glaube zumindest 13 Punkte erreicht zu haben.
- ☐ Genügend, ich glaube zumindest 10 Punkte erreicht zu haben.
- ☐ Nicht Genügend, ich glaube weniger als 10 Punkte erreicht zu haben.

Der wievielte Antritt zur Prüfung Psychologische Diagnostik 1 ist das?

- ☐ Der erste Antritt.
- ☐ Der zweite Antritt.
- ☐ Der dritte Antritt.

Wie weit konnten Sie Ihre Vorbereitung für die Prüfung abschließen?

- ☐ 0-20%
- ☐ 21-40%
- ☐ 41-60%
- ☐ 61-80%
- ☐ 81-100%

Haben Sie die Informelle Leistungsüberprüfung im Rahmen der VO Psychologische Diagnostik absolviert?

- ☐ Ja, ich habe die Fragen der Informellen Leistungsüberprüfung beantwortet, mir meine Ergebnisse allerdings nicht angesehen.
- ☐ Ja, ich habe die Fragen der Informellen Leistungsüberprüfung beantwortet und auch meine Ergebnissen angesehen.
- ☐ Ja, ich habe die Fragen der Informellen Leistungsüberprüfung beantwortet ,meine Ergebnissen angesehen und mich mit meinen Fehlern auseinandergesetzt.
- ☐ Nein, ich habe die Fragen der Informellen Leistungsüberprüfung nicht beantwortet.

Wie regelmäßig haben sie dieses Semester(SS2008) die Vorlesung zur psychologischen Diagnostik 1 besucht?

- ☐ immer (10 Einheiten)
- ☐ meistens (7-9 mal)
- ☐ gelegentlich (4-6 mal)
- ☐ selten (1-3 mal)
- ☐ nie
- ☐ Ich habe die Vorlesung zur psychologischen Diagnostik 1 bereits vor dem Sommersemester 2008 besucht. Genau im _____ (bitte Semester einfügen)

Falls Sie die Vorlesung dieses Semester (SS2008) absolviert haben, wie sehr haben Sie das E-learning Angebot genutzt?

- ☐ Ich habe das E-learning gar nicht genutzt.
- ☐ Ich habe die zur Verfügung gestellten Materialien angesehen.
- ☐ Ich habe die zur Verfügung gestellten Materialien angesehen und durchgearbeitet ohne Beteiligung im Forum.
- ☐ Ich habe die zur Verfügung gestellten Materialien angesehen und durchgearbeitet und mich auch im Forum beteiligt.

9.7 Lehrziele zur VO Diagnostik I

- A. Kenntnis und Fähigkeit zur Umsetzung der einzelnen Schritte des psychologisch-diagnostischen Prozesses inklusive der organisatorischen Konzeption des Vorgehens
- B. Fähigkeit zur (rechnerischen) Transformation verschiedener Normwerte um deren Vergleichbarkeit zu gewährleisten inklusive der Berücksichtigung von Konfidenzintervallen
- C. Kenntnis der Gütekriterien und Verständnis für die Problematik, die mit den Gütekriterien verbunden ist (z. B. unterschiedliche Arten der Validierung inklusive der Vor- und Nachteile dieser Arten)
- D. Kenntnis der verschiedenen Arten psychologisch-diagnostischer Verfahren und Fähigkeit, psychologisch-diagnostische Verfahren, kritisch reflektieren zu können.
- E. Kenntnis der psychologisch-diagnostischen Verfahren des Testinventars zur Leistungsdiagnostik
- F. Fähigkeit, verschiedenste Gestaltungsweisen (Antwortformate, Power- vs. Speed&Power, Gruppen- vs. Individualverfahren) von psychologisch-diagnostischen Verfahren kritisch reflektieren zu können
- G. Fähigkeit, verschiedenste Erhebungstechniken kritisch reflektieren zu können
- H. Kenntnis testtheoretischer Grundlagen, insbesondere Verständnis für die Umsetzung Klassischer Testtheorie und Item-Response-Theorie in der Psychologischen Diagnostik

9.8 Emailaufforderung zur Teilnahme an der ILP

Diese Email wurde an alle Übungsteilnehmer der Übungen zum psychologischen Diagnostizieren 1 im Sommersemester 2008 am 31.05.2008 versandt. Der Inhalt der Email wurde zur bessern Unterscheidung vom übrigen Text kursiv gehalten.

Liebe Übungsleiterinnen und Übungsleiter!

Liebe Studentinnen und Studenten!

Wie bereits zu Beginn des Semesters bekannt gegeben, stellt eine informelle Leistungsüberprüfung den Abschluss der Testerfahrung im Sommersemester 2008 dar. Hier sollen nochmals die Vorteile für die Studierenden dargestellt, sowie der genaue Ablauf beschrieben werden.

Ablauf:

- 1. Alle jene, die nicht via Blackboard Vista für die VO Psychologische Diagnostik I angemeldet sind, **melden sich [HIER](#) an.***
- 2. Am 02.06.2008 wird die informelle Leistungsüberprüfung freigegeben. Sie ist dann entweder unter dem Menüpunkt „Tests“ oder direkt unter „Kursinhalt“ zu finden.*
- 3. Die informelle Leistungsüberprüfung ist von **02.06.2008 bis 08.06.2008** verfügbar. Danach könnt ihr via Blackboard eure Ergebnisse und ein allgemeines Feedback zu den Fragen einsehen.*

Was ist bei der Bearbeitung zu beachten?

*Ihr könnt **alle Unterlagen** verwenden, die auch bei der Prüfung gestattet sind.*

- Die Fragen sind im Multiple-Choice Format, wobei immer **zwischen 1 und 5 von fünf möglichen Antworten richtig** sind.*
- Es ist wichtig die Items **alleine zu bearbeiten**, um einen Nutzen für die Prüfung daraus zu ziehen. Nachdem ihr den Test ausgefüllt habt, könnt ihr die einzelnen Fragen gerne in Gruppen diskutieren, ihr verfügt nach Ablauf der Vorgabezeit (08.06.08) auch über die richtigen Antworten.*
- Versucht bei der Bearbeitung keine Pausen oder Unterbrechungen zu machen um möglichst nahe an der Prüfungssituation zu sein.*
- Es handelt sich hier um eine informelle Leistungsüberprüfung, die **nicht benotet** wird, sondern euch bei der Prüfungsvorbereitung behilflich sein soll.*

Eure Vorteile:

- Die Fragen der informellen Leistungsüberprüfung wurden in Zusammenarbeit mit der Lehrveranstaltungsleitung erarbeitet und orientieren sich inhaltlich an den vom*

Institut vorgegebenen Lehrzielen. Sie stellen somit eine gute Gelegenheit dar den eigenen Wissenstand an „prüfungsnahe“ Fragen zu testen.

- *Da es sich um eine informelle Leistungsüberprüfung handelt, werden eure Ergebnisse weder benotet, noch von den Projektmitarbeitern oder den Lehrveranstaltungsleitern individuell betrachtet. Der Erkenntnisgewinn für uns (Diplomanden und LVA-Leitung) steckt in den aggregierten Daten der „Gruppe“, im Rahmen der Studie haben wir also kein Interesse an euren Einzelergebnissen.*
- *Da es rechtzeitig vor der Prüfung eine Rückmeldung zur informellen Leistungsüberprüfung geben wird, könnt ihr diese Erkenntnisse noch für eure Prüfungsvorbereitung nutzen. In der Rückmeldung seht ihr einerseits eure Ergebnisse den richtigen Antworten gegenübergestellt, andererseits gibt es zu jeder Frage eine kurze Erläuterung.*

Wir hoffen, dass die Bearbeitung der informellen Leistungsüberprüfung für euch hilfreich ist und wünschen euch weiters viel Erfolg bei der Prüfung!

Michaela Wagner-Menghin (LVA(VO)-Leitung)

Simon Lehner (Diplomand)

9.9 Lebenslauf

Zur Person:

Name: Simon Lehner
Geburtsdatum: 02.07.1981
Geburtsort: Salzburg

Ausbildung:

2007-2008: Ausbildung zum zertifizierten AID2 Testleiter
2001-2009 Diplomstudium Psychologie Universität Wien
1992-1999 BRG Steyr, Michaelerplatz 4400 Steyr,
1991-1992 BRG Adolf-Pichler-Platz 6020 Innsbruck
1987-1991 Volksschule Sierning 4522 Sierning

Berufserfahrung:

2009 Projektmitarbeiter Test- und Beratungsstelle des Arbeitsbereichs Psychologische Diagnostik an der Fakultät für Psychologie der Universität Wien – Projekte: IKM; PIAAC
2008 Testleiter AID 2 Englisch
2008 Absolvierung des 6-Wochen Praktikums bei der Firma ÖBB-Human Resources Operations, Servicebereich - Psychologie & Diagnostik
2007 Student-Mentor Universität Wien - CBM Projekt
2007 Mitarbeiter 9. Arbeitstagung der Fachgruppe für Differentielle Psychologie, Persönlichkeitspsychologie und Psychologische Diagnostik an der Fakultät für Psychologie der Universität Wien
2006-2008 Mitarbeiter Firma Direct Line Wien – Backoffice: Reporting , Controlling, EDV Verwaltung und Personalangelegenheiten
2005 Mitarbeiter Firma Direct Wien - Telefonisches Kundenservice
2004 Errichtung und Betreuung der EDV Infrastruktur für das „Blaue KreuzWien
2003-2004 Mitarbeiter Competence Callcenter AG Wien - Telefonisches Kundenservice
2003 Netzwerk und EDV Betreuung in der Firma Bioinfo Austria
2000-2001 Zivildienstes im Seniorenwohnheim 4540 Bad Hall

Weitere Kenntnisse:

Sprachen: Englisch (verhandlungsfähig); Spanisch (Grundlagen)
Computer: SPSS, R inkl. eRm, Microsoft Office, Crystal Reports/Enterprise, Linux (Ubuntu, Debian), HTML, MYSQL;