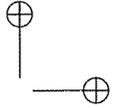
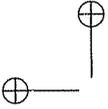


Punter, J.F.,Litzenberger, M., Grambs, T., Jirasko, M. & Spiel, C. (2006). Zur Auswahl von Studierenden mittels lehrzielorientierter Wissensprüfung. In B. Gula, R. Alexandrowicz, S. Strauß, E. Brunner, B. Jenull-Schiefer, & O. Vitouch (Hrsg.). Perspektiven psychologischer Forschung in Österreich. Proceedings zur 7. Wissenschaftlichen Tagung der Österreichischen Gesellschaft für Psychologie (S. 472 - 479). Lengerich: Pabst.



Zur Auswahl von Psychologie-Studierenden mittels lehrzielorientierter Wissensprüfung

Joachim Fritz Punter, Margarete Litzenberger, Timo Gnambs,
Marco Jirasko & Christiane Spiel

An der Fakultät für Psychologie der Universität Wien wurde auf Basis der Ringvorlesung „Psychologie als Wissenschaft“ ein Auswahlverfahren für künftige Psychologie-Studierende eingerichtet. Zur Durchführung der Prüfungen wurde eine lehrzielorientierte, EDV-unterstützte Prüfung mit Multiple-Choice-Antwortformat eingeführt. Das Erarbeiten von Richtlinien zur Gestaltung von Multiple-Choice-Aufgaben, die über die Rekognition von Wissen im Sinne der Taxonomie der kognitiven Lernziele (Bloom, 1956) hinausgehen, wird veranschaulicht sowie das Procedere der Wissensprüfung dargestellt. Das Ziel der Evaluation der Prüfungen besteht darin, Hinweise zur Verbesserung der Fragenqualität zu ermitteln bzw. die Richtlinien zur Erstellung von Multiple-Choice-Items zu adaptieren.

Schlüsselwörter: Multiple-Choice, Psychologie-Studium, Auswahlverfahren

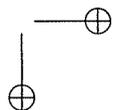
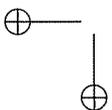
1 Einleitung

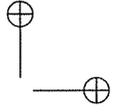
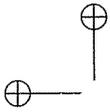
Als zukunftsorientierte Maßnahme zur Handhabung der enormen Prüfungsbelastung einerseits sowie als Sofortmaßnahme zur Bewältigung des möglichen Ansturms ausländischer Studierender aus dem EU-Raum¹ wurde an der Fakultät für Psychologie der Universität Wien das Projekt zur Umstellung der Prüfung zur Vorlesung *Psychologie als Wissenschaft* auf das Multiple-Choice-Antwortformat initiiert. Bei der Ringvorlesung *Psychologie als Wissenschaft* handelt es sich um eine einführende Lehrveranstaltung für das Diplomstudium Psychologie, welche Grundwissen über die verschiedenen Prüfungsfächer vermittelt und zu Beginn jedes Wintersemesters in Form einer Blocklehrveranstaltung angeboten wird.

Die positive Absolvierung der *Psychologie als Wissenschaft* (gliedert sich in 2 Teile zu je 1 SWS) ist laut Studienplan die Voraussetzung für die Teilnahme an allen Lehrveranstaltungen mit immanenten Prüfungscharakter (Übungen, Proseminare). Im Studienjahr 2005/06 ist an die Prüfung zur Ringvorlesung zusätzlich ein Auswahlverfahren² gekoppelt.

1 Im Zusammenhang mit dem Urteil des Gerichtshofes der Europäischen Gemeinschaften vom 07.07.2005 zu den „Voraussetzungen des Zugangs zum Hochschulstudium – Diskriminierung“.

2 Verordnung des Rektorats der Universität Wien vom 08.09.2005 (Mitteilungsblatt, Studienjahr 2004/2005, ausgegeben am 08.09.2005, 39. Stück).





In der Vergangenheit wurden etwa 1500 Prüfungen (mit freiem Antwortformat) pro Studienjahr abgenommen. Die investierte Arbeitszeit für deren Abwicklung betrug etwa 1000 Stunden. Vor diesem Hintergrund ergaben sich folgende Projektziele:

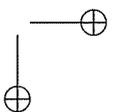
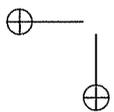
1. Die Erarbeitung von Grundlagen für die Erstellung von computerisiert auswertbaren Multiple-Choice-Prüfungsfragen, die das reine „Wiedererkennen von Faktenwissen“ übersteigen (im Sinne der Taxonomie von Bloom, 1956) durch Aufarbeitung und Dokumentation der einschlägigen Literatur.
2. Die Ermittlung der technischen und administrativen Notwendigkeiten für die Durchführung derartiger Prüfungen.
3. Die Erstellung eines ausreichend umfangreichen Fragenpools für die Prüfung *Psychologie als Wissenschaft* in hoher Qualität, um die Prüfungsdurchführung für zumindest ein Jahr garantieren zu können.
4. Vorbereitung und Durchführung der Prüfungen sowie deren Evaluation.

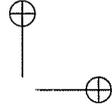
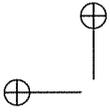
2 Vorgehensweise

2.1 Taxonomie von Bloom

Bloom (1956) unterscheidet in seiner Taxonomie drei Bereiche (kognitiv, affektiv und psychomotorisch), denen er eine Reihe von hierarchisch gegliederten Lernzielen zuordnet. Die kognitiven Lernziele gliedern sich in:

1. Wissen: Umfasst die Fähigkeit, Informationen aus dem Gedächtnis wiederzugeben und stellt die niedrigste Stufe der Lernziele auf der kognitiven Ebene dar.
2. Verständnis: Umfasst die Fähigkeit, die Bedeutung des Erlernten zu erfassen ohne es notwendigerweise auf andere Bereiche zu übertragen oder Implikationen daraus abzuleiten.
3. Verständnis: Umfasst die Fähigkeit, die Bedeutung des Erlernten zu erfassen ohne es notwendigerweise auf andere Bereiche zu übertragen oder Implikationen daraus abzuleiten.
4. Anwendung: Umfasst die Fähigkeit, Erlerntes (Regeln, Methoden) in neuen Situationen anzuwenden.
5. Analyse: Umfasst die Fähigkeit, Material in seine Bestandteile aufzubrechen.
6. Synthese: Umfasst die Fähigkeit, aus Teilen, Elementen etc. etwas neues Ganzes zu formen.
7. Beurteilung: Umfasst die Fähigkeit, den Wert bzw. die Güte eines vorgegebenen Materials für eine bestimmte Situation einzuschätzen.





2.2 Typen von Multiple-Choice-Items

Multiple-Choice-Items können in unterschiedlichster Gestalt Verwendung finden (Albanese, 1993; Haladyna & Downing, 1989a; Haladyna & Downing, 1989b; Haladyna, 1992). Die gebräuchlichste Form stellen *klassische Multiple-Choice-Items* dar, die aus einem Itemstamm und mehreren Antwortoptionen bestehen. *Kontextbezogene Multiple-Choice-Items* weisen einen gemeinsamen Stimulustext auf, dem nachfolgend eine Reihe einzelner Multiple-Choice-Aufgaben folgen. Diese Itemformate sind gebräuchlich und testtheoretisch bereits gut validiert, sodass sie in standardisierten Tests sinnvoll Anwendung finden können. Eine weitere Differenzierungsmöglichkeit kann anhand der Anzahl richtiger Antwortoptionen vorgenommen werden. *Multiple-Choice-Items mit Mehrfachwahl* enthalten nicht, wie klassische Multiple-Choice-Items, nur eine richtige Antwort sondern mehrere. Einen Sonderfall stellen *erweiterte Zuordnungsaufgaben* dar. Diese bestehen aus einer umfangreichen Anzahl von Antwortoptionen, die einer Reihe von Items zugeordnet werden sollen.

2.3 Konstruktion der Items

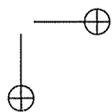
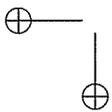
Basierend auf Forschungsergebnissen einschlägiger Literatur wurden zunächst Richtlinien zur Itemkonstruktion ausgearbeitet (Litzenberger, Gnams & Punter, 2005). Dieser Katalog enthielt praktische Empfehlungen sowohl zur formalen wie auch inhaltlichen Konstruktion von Multiple-Choice-Items zur Wissensprüfung.

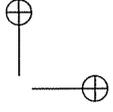
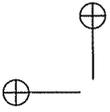
Im nächsten Schritt wurden die bislang impliziten inhaltlichen Lernziele von den PrüfungsfachkoordinatorInnen explizit ausformuliert³ und als Grundlage für die inhaltliche Ausarbeitung der Prüfungsfragen herangezogen. Realisiert wurde für die Prüfung ein Multiple-Choice-Itemformat mit fünf Antwortoptionen mit eins bis vier richtigen Antwortalternativen.

Ausgehend von diesen Richtlinien wurde zur Verwaltung des Itempools eine eigene Datenbank entwickelt, in der jedes Item nach Prüfungsfach, inhaltlichem Lernziel und der Zuordnung nach Bloom (1956) kategorisiert wurde. Items, die aufgrund ihrer Formulierung einen Hinweisreiz zur Lösung eines anderen Items liefern könnten, wurden in der Datenbank explizit ausgewiesen, sodass diese bei einer zufälligen (computerbasierten) Itemauswahl nicht in einer Prüfung gemeinsam vorgegeben wurden. Die Prüfungsbögen wurden aus der Datenbank generiert werden, wobei eine zufällige Auswahl (bei Vorgabe des Verhältnisses der Kategorien nach Bloom sowie der Anzahl der Items pro Prüfungsfach) oder eine händische Auswahl möglich ist. Die Erstellung von Parallelformen (durch Permutation der Antwortoptionen, der Items sowie der Prüfungsfächer) ist problemlos möglich. Für jeden Test ist zudem der Export des Lösungsvektors in eine eigene Datei vorgesehen.

In Zusammenarbeit mit dem Zentralen Informatikdienst der Universität Wien wurde ein Hochgeschwindigkeitsscanner samt entsprechender Software angeschafft. Von

³ Auf Basis des Skriptums „Psychologie als Wissenschaft“ (Vanecek, Kastner-Koller, Deimann & Toyfl, 2002), welches dem gleichnamigen Buch (Fakultät für Psychologie, Kastner-Koller & Deimann, 2006) zugrunde liegt.





jedem eingelesenen Antwortbogen wird eine Grafikdatei gespeichert und die Ankreuzungen werden in eine Textdatei geschrieben. Die Einstellung der Schwellenwerte, das heißt ab welchem Prozentwert an Füllung eines Kästchens dieses als leer, nur verschmutzt, angekreuzt bzw. vollständig ausgestrichen interpretiert wird, zählt zu den problematischen Einstellungen und entscheidet über die spätere Handarbeit. Nach mehreren Testläufen und Verbesserung der Instruktion konnte die Fehlerrate auf ein Minimum gesenkt werden, so dass nur noch einzelne Extremfälle von Hand nachgebessert werden mussten. Die weitere Auswertung erfolgt über eine entsprechende SPSS-Syntax-Datei, welche die Dateien mit den Lösungsvektoren sowie die von der Software des Belegscanners erzeugte Textdatei mit den Antwortvektoren einliest und schließlich die Punkte wie auch die Noten automatisiert berechnet. Aus den vorliegenden Daten (Scan des Antwortbogens, Ankreuzungen der Kandidaten und Lösungsvektoren) wird zudem für jede Person ein PDF mit einer individuellen Ergebnismeldung erzeugt und den Studierenden via APIS (internetbasiertes Anmelde-, Prüfungs- und Informationssystem der Fakultät für Psychologie) aus Gründen der Transparenz zur Verfügung gestellt.

3 Ergebnisse

Die Prüfungen zu *Psychologie als Wissenschaft I* ($n = 1080$) und *Psychologie als Wissenschaft II* ($n = 942$) haben zu verschiedenen Terminen stattgefunden. Es wurden jeweils 42 Items in drei Parallelgruppen vorgegeben.

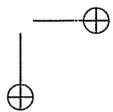
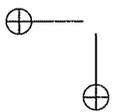
Um Aufschluss über die Qualität der Items bzw. der Distraktoren zu erhalten, wurden die relativen Häufigkeiten der gewählten Antwortalternativen pro Item getrennt für drei verschiedene Leistungsgruppen berechnet. Ein Beispiel:

Beispiel

Sie führen folgenden Versuch durch: Sie zeigen einem Kind zwei gleiche mit Wasser gefüllte Gläser A und A1. Anschließend schütten Sie das Wasser aus A1 in ein niedrigeres, breiteres Glas B um und fragen das Kind, ob in A oder B mehr Wasser sei. Welche Antwort(en) des Kindes lässt bzw. lassen darauf schließen, dass es sich im Stadium der konkret-logischen Operationen nach Piaget befindet?

- a) „In B ist weniger Wasser als in A.“
- b) „In B ist mehr Wasser als in A.“
- c) „In beiden Gläsern ist gleich viel Wasser.“
- d) „Wenn man das Wasser zurückschüttet, sieht es wieder genauso aus.“
- e) Das Kind antwortet nicht.

Antwortalternative „e“ wird praktisch nie gewählt (s. Tabelle 1). Nach den Richtlinien zur Gestaltung von Multiple-Choice-Items, sollte der Distraktor „keine Alternative richtig“ (sinngemäß Antwortalternative „e“) möglichst vermieden werden. Der



Distraktor „e“ kann logisch ausgeschlossen werden, insbesondere wenn eine der Alternativen „a“ bis „d“ gewählt wurde. Dementsprechend wird Distraktor „e“ gegen eine inhaltliche Antwort ausgetauscht: „In A ist mehr Wasser, weil Glas A höher ist.“

Tab. 1. Relative Häufigkeiten der gewählten Antwortalternativen für drei Leistungsgruppen

Leistungsgruppe	Item gesamt	Antwortoptionen				
		a	b	c	d	e
hoch	.8328	.0137	.0000	.9727	.8498	.0000
mittel	.6336	.0870	.0162	.8482	.7146	.0020
niedrig	.3959	.1741	.0512	.6758	.5392	.0000
gesamt	.6231	.0907	.0213	.8352	.7037	.0009

Um Items zu identifizieren, die unterschiedliche Lösungswahrscheinlichkeiten für verschiedene Gruppen aufweisen, wurden Analysen der odds ratio nach Mantel-Haenszel durchgeführt. Von den 84 Items werden lediglich drei Items als stark und weitere sieben als leicht biased eingestuft (das heißt, die Zugehörigkeit zu einer Parallelgruppe begünstigte deren Lösung). Bezüglich des Geschlechts finden sich drei Items, die primär von Männern häufiger gelöst wurden. Die Nationalität hat für die Lösung der Prüfungsaufgaben lediglich eine untergeordnete Rolle gespielt. Studierende aus Deutschland weisen weitgehend vergleichbare Lösungswahrscheinlichkeiten auf wie Studierende aus Österreich. Problematisch stellte sich vielmehr die Sprachzugehörigkeit dar. So wiesen Personen, deren Muttersprache nicht deutsch ist, bei zahlreichen Items divergierende Lösungswahrscheinlichkeiten auf.

Statistisch signifikante Unterschiede zwischen den einzelnen Prüfungsfächern lassen sich mit $\chi^2(11) = 15.28$, $p = .17$ nicht belegen (vgl. Tabelle 2). Dies konnte allerdings aufgrund des sehr geringen Stichprobenumfangs (lediglich 7 Items pro Fachbereich) auch eher nicht erwartet werden.

Tab. 2. Mittlere Itemschwierigkeiten (Median) nach Fachbereich und Leistungsgruppen

Prüfungsfach	Leistungsgruppen			gesamt
	obere 27 %	mittlere 46 %	untere 27 %	
Allgemeine Psychologie	0.89	0.69	0.35	0.65
Methodenlehre	0.70	0.47	0.25	0.47
Entwicklungspsychologie	0.85	0.63	0.30	0.62
Differentielle Psychologie	0.81	0.61	0.29	0.59
Biologische Psychologie	0.73	0.44	0.16	0.44
Sozialpsychologie	0.96	0.87	0.56	0.83
Einführung in die Psychologie	0.60	0.36	0.14	0.36
Psychologische Diagnostik	0.76	0.52	0.21	0.57
Klinische Psychologie	0.88	0.61	0.17	0.63
Wirtschaftspsychologie	0.88	0.55	0.16	0.59
Bildungspsychologie	0.89	0.69	0.32	0.70
Forschungsmethoden und Evaluation	0.82	0.35	0.16	0.45
Kruskal-Wallis Test	$\chi^2(11) = 13.79,$ $p = .25$	$\chi^2(11) = 6.72,$ $p = .12$	$\chi^2(11) = 16.48,$ $p = .13$	$\chi^2(11) = 15.28,$ $p = .17$

Alle Items wurden in Anlehnung an die Lernziel-Taxonomie von Bloom (1956) entwickelt und entsprechend für den Test selektiert. So wurden für jedes Prüfungsfach vier Items der Wissensebene und drei Items höherer Lernziele gewählt. Ein Vergleich der Itemschwierigkeiten von Items der Kategorie Wissen und Items höherer Lernziele offenbarte mit $\bar{x}_{\text{Wissen}} = 0.53$ bzw. $\bar{x}_{\text{höhere Lernziele}} = 0.46$ zunächst nur marginale Unterschiede hinsichtlich der Itemgruppen ($U = 665, p = .07$). Wissensitems schienen zudem hypothesenkonträr schwieriger zu sein als Items, die höhere Lernziele prüfen sollten. Erst unter Berücksichtigung der jeweiligen Prüfungsfächer ergab sich eine interessante Wechselwirkung. So waren bei den meisten Prüfungsfächern wie erwartet jene Items, die reines Faktenwissen prüfen, leichter als Items höherer Lernziele ($\bar{x}_{\text{Wissen}} = 0.58, \bar{x}_{\text{höhere Lernziele}} = 0.42$). Lediglich Items aus der Methodenlehre, Differentiellen Psychologie und Psychologischen Diagnostik, also primär statistisch-methodische Prüfungsfragen, wiesen einen umgekehrten Zusammenhang auf ($\bar{x}_{\text{Wissen}} = 0.39, \bar{x}_{\text{höhere Lernziele}} = 0.55$). In jenen Fächern schienen anwendungsorientierte Fragestellungen einfacher zu beantworten als reine Wissensfragen. Die Resultate einer parameterfreien Rang-Varianzanalyse bestätigen dieses Ergebnis (vgl. Tabelle 3 und Abbildung 1).

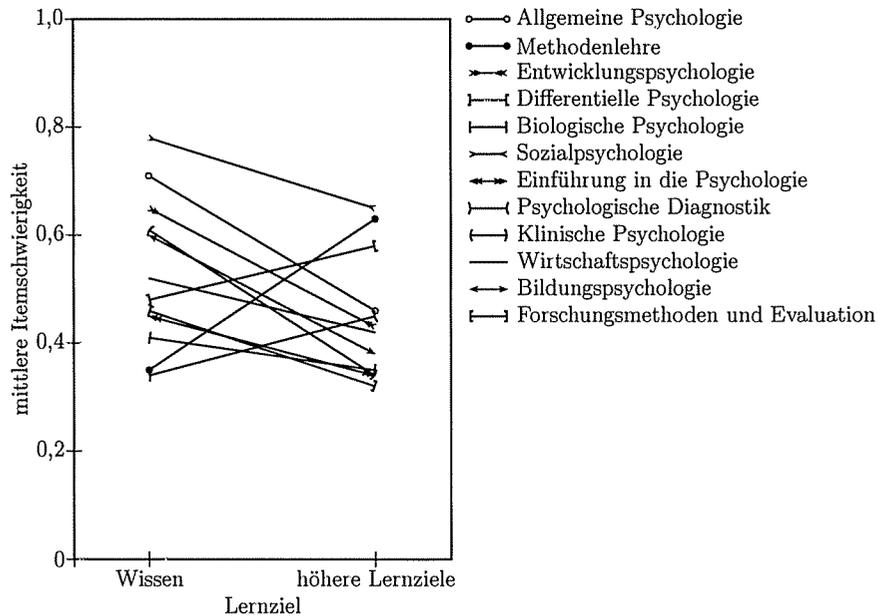


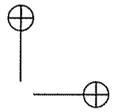
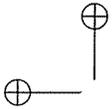
Abb. 1. Mittlere Itemschwierigkeiten nach Lernzielen und Prüfungsfach

Tab. 3. Rang-Varianzanalyse nach Kubinger (1986) der Itemschwierigkeiten nach Prüfungsfach und Lernzielen

	χ^2	df	p
Prüfungsfach	21.18	11	.03
Lernziel	4.94	1	.02
Prüfungsfach × Lernziel	21.76	11	.03

4 Diskussion

Die Umstellung der Prüfung zur Ringvorlesung *Psychologie als Wissenschaft* auf das Multiple-Choice-Antwortformat hat zu einer wesentlichen Entlastung der wissenschaftlichen MitarbeiterInnen der Fakultät für Psychologie geführt. Das verrechnungssichere Itemformat trägt wesentlich zur Objektivität bei. Die Evaluation der bisherigen Prüfungen insbesondere auch auf Itemebene ist notwendig, um die Fragenqualität beurteilen zu können und um gegebenenfalls Neuformulierungen des Itemstamms und der Distraktoren vorzunehmen. Weiterführende Studien zur Prognose des Studienerfolgs anhand der Prüfungsergebnisse sind in Planung.



Literatur

- Albanese, M. (1993). Type K and other complex multiple-choice items: An analysis of research and item properties. *Educational Measurement: Issues and Practice*, 12, 28–33.
- Bloom, B. S. (1956). *Taxonomy of educational objectives*. New York: McKay.
- Fakultät für Psychologie, Kastner-Koller, U. & Deimann, P. (Hrsg.). (2006). *Psychologie als Wissenschaft*. Wien: WUV Universitätsverlag.
- Haladyna, T. M. (1992). The effectiveness of several multiple-choice formats. *Applied Measurement in Education*, 5, 73–88.
- Haladyna, T. M. & Downing, S. M. (1989a). A taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education*, 2, 37–50.
- Haladyna, T. M. & Downing, S. M. (1989b). Validity of a taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education*, 2, 51–78.
- Kubinger, K. D. (1986). A note on non-parametric tests for the interaction in two-way layouts. *Biometrical Journal*, 28, 67–72.
- Litzenberger, M., Gnambs, T. & Punter, J. F. (2005). *Richtlinien zur Konstruktion von Multiple-Choice-Items*. Unveröffentlichtes Manuskript, Universität Wien.
- Vanecek, E., Kastner-Koller, U., Deimann, P. & Toyfl, M. (Hrsg.). (2002). *Psychologie als Wissenschaft: Skriptum zur Eingangsphase für das Studium der Psychologie*. Wien: Institut für Psychologie der Universität Wien.

