

Erfassung von Wissen über Informationsrecherchen: Konzeptuelle Überlegungen und empirische Befunde*

Nikolas Leichner, Johannes Peter, Anne-Kathrin Mayer und Günter Krampen

Berichtet wird über die Konstruktion eines Multiple-Choice-Tests für die Erhebung des Wissens im Bereich der Recherche und Bewertung von (Fach-)Informationen. Eine erste Form des Testverfahrens wurde in einer ersten Studie bei N = 64 Studierenden und Doktoranden/innen des Fachs Psychologie eingesetzt. Auf Grundlage der Ergebnisse dieser Studie wurde der Test revidiert und in einer Online-Studie bei N = 184 Psychologiestudierenden eingesetzt und psychometrisch evaluiert. Die Ergebnisse führen zur Hypothese, dass der Test Wissen aus einer breiten und fragmentierten Wissensdomäne abbildet. Vor diesem Hintergrund ist das Verfahren insbesondere für die Evaluation von Schulungen verwendbar.

This paper reports about the development of a Multiple Choice instrument to assess knowledge about information retrieval in Psychology students. A preliminary version of the test was used with a sample of N = 64 Psychology students (including Ph.D. students). The test was revised and data of N = 184 students were collected online. The results lead us to the hypothesis that the knowledge captured by the test is part of a broad and fragmented knowledge domain. In view of this fact, the test is especially suitable for the evaluation of information literacy instruction.

Einleitung

Informationskompetenz wird als eine Schlüsselkompetenz angesehen, die in fast allen Bereichen des Lebens von Vorteil ist, insbesondere jedoch im Kontext der Aus- und Weiterbildung und am Arbeitsplatz¹. Der Vergleich verschiedener Informationskompetenz-Konzepte und -Definitionen führt zur Schlussfolgerung, dass in allen Konzepten drei zentrale Aspekte enthalten sind: Das Erkennen eines Informationsbedarfs, das effektive und effiziente Suchen und Bewerten von Informationen sowie die Beachtung ethischer Aspekte, wie etwa das Vermeiden von Plagiaten². Nach dem Definitionsvorschlag der Association of College and Research Libraries (ACRL)³, mit dem zugleich Standards der Informationskompetenz spezifiziert

werden sollen, zeichnet sich eine informationskompetente Person durch fünf Verhaltensweisen aus: Sie

1. bestimmt Art und Umfang der benötigten Informationen.
2. verschafft sich effizienten und effektiven Zugang zu den benötigten Informationen.
3. evaluiert Informationen und ihre Quellen kritisch und integriert die ausgewählten Informationen in ihr Wissen und ihr Wertesystem.
4. nutzt Informationen effektiv, sowohl als Individuum als auch als Gruppenmitglied, um ein bestimmtes Ziel zu erreichen.
5. versteht viele der ökonomischen, rechtlichen und sozialen Streitfragen, die mit der Nutzung von Informationen zusammenhängen und hat Zugang und nutzt die Informationen in einer ethischen und legalen Weise⁴.

Für die Erfassung von Informationskompetenz wird meist auf diese Standards zurückgegriffen, weil das Konzept verhaltensnahe Leistungsindikatoren für jeden der fünf Standards beinhaltet. Mittlerweile wurden die Standards der ACRL an die Besonderheiten des Psychologie-Studiums angepasst und Psychologie-spezifische Standards formuliert⁵. Über Möglichkeiten, Informationskompetenz in dem so definierten Sinne zu erfassen, informiert eine Übersicht⁶. Beispielsweise kann Informationskompetenz durch die Untersuchung des tatsächlichen Verhaltens während der Recherche erhoben werden. Bei solchen Studien zum „information seeking behavior“ kann anhand der aus den Informationskompetenz-Standards abgeleiteten behavioralen Leistungsindikatoren bestimmt

4 Siehe auch die Übersetzung der Definitionen von Homann, Benno: „Standards der Informationskompetenz“. Eine Übersetzung der amerikanischen Standards der ACRL als argumentative Hilfe zur Realisierung der „Teaching Library“, in: Bibliotheksdienst 36(5) (2002) S. 625–638.

5 http://www.ala.org/acrl/standards/psych_info_lit [19. März 2013].

6 Walsh, Andrew: „Information literacy assessment: Where do we start?“, in: Journal of Librarianship and Information Science 41(1) (2009) S. 19–28.

* Diese Studien wurden aus Mitteln des Pakts für Forschung und Innovation des Bundes und der Länder im Wettbewerbsverfahren der Leibniz-Gemeinschaft gefördert.

1 http://infolit.org/?page_id=3172 [6. Juni 2012].

2 Boon, Stuart/ Johnston, Bill/ Webber, Sheila: „A phenomenographic study of English faculty's conceptions of information literacy“, in: Journal of Documentation 63(2) (2007) S. 204–228.

3 <http://www.ala.org/acrl/sites/ala.org.acrl/files/content/standards/standards.pdf> [6. Juni 2012].

werden, welches Verhalten als angemessen gilt. Aus ökonomischen Gründen wird Informationskompetenz jedoch in aller Regel mit weitaus weniger aufwändigen Multiple-Choice-Testverfahren erhoben⁷, welche das Wissen über verschiedene Aspekte der Informationssuche, -bewertung und -nutzung abbilden. Uns sind zwei kommerziell vertriebene Tests bekannt, die Multiple-Choice-Items verwenden⁸, allerdings ist keiner dieser Tests spezifisch auf Psychologiestudierende zugeschnitten. Diese Lücke soll durch die Konstruktion eines neuen Wissenstests gefüllt werden. Dabei gehen wir davon aus, dass das durch den Test erhobene Wissen über Informationsrecherchen einen Indikator für Informationskompetenz darstellt.

Studie 1: Konstruktion des Wissenstests und erste Erprobung

Der Itempool wurde aufgrund in der Literatur auffindbarer Tests⁹ und aufgrund eigener Überlegungen ent-

wickelt. Die Zielsetzung besteht darin, mit den Items die Standards 2 (Suchen) und 3 (Bewerten) der ACRL-Definition von Informationskompetenz¹⁰ abzudecken, da dies die Schwerpunkte eines geplanten Trainings von Informationskompetenz sein sollen. Das Training beschäftigt sich vorrangig mit diesen Aspekten von Informationskompetenz, weil die weiteren Aspekte, insbesondere Standards 4 und 5, eher Aspekte wissenschaftlichen Arbeitens sind¹¹.

Methoden

Die Pilotfassung des Wissenstests besteht aus 35 Items, die mit einer Ausnahme (offenes Antwortformat) ein Multiple-Choice-Format (MC) aufweisen. 33 der Items verfügen über drei Antwortoptionen, ein Item über zwei. Für jede korrekt angekreuzte oder ausgelassene Antwortalternative werden 0,33 Punkte vergeben. Von den Items sind inhaltlich 23 Items dem ACRL-Standard 2 (Suchen) und 12 dem ACRL-Stan-

7 Vgl. beispielsweise Noe, Nancy/ Bishop, Barbara: „Assessing Auburn University Library's Tiger Information Literacy Tutorial (TILT)“, in: Reference Services Review 33(2) (2005) S. 173–187.

8 <https://www.projectsails.org> [10. Mai 2013]; Wise, Steven/ Cameron, Lynn/ Yang, Sheng-Ta/ Davis, Susan/ Russell, Javarr: The Information Literacy Test (ILT). Test manual, Harrisonburg, VA 2009.

9 Noe/ Bishop (wie Anm. 7); Ondrusek, Anita/ Dent, Valeda/

Bonadie-Joseph, Ingrid/ Williams, Clay: „A longitudinal study of the development and evaluation of an information literacy test“, in: Reference Services Review 33(4) (2005) S. 388–417.

10 wie Anm. 3.

11 Vgl. Homann, Benno: Standards und Modelle der Informationskompetenz – Kooperationsgrundlagen für bibliothekarische Schulungsaktivitäten, in: Ute Krauß-Leichert (Hrsg.): Teaching Library – Eine Kernaufgabe für Bibliotheken, Frankfurt 2007.



Die Vergangenheit lebendig halten.

In Bibliotheken und Archiven auf der ganzen Welt werden wertvolle Bücher, Zeitungen, Verträge und Manuskripte mit unseren Hightech-Scannern und Softwarelösungen erfasst, aufbereitet und der Forschung und Wissenschaft digital zur Verfügung gestellt. Seit 1961 halten digitale und analoge Speichersysteme von Zeutschel so die Vergangenheit lebendig. Mit einem OS 14000 A0 Großformat-Scanner lassen sich zum Beispiel sehr empfindliche Dokumente wie eine historische Weltkarte aus der Zeit der Mayflower schonend und in bester Qualität digitalisieren. Schreiben Sie uns, wenn Sie mehr darüber wissen möchten: info@zeutschel.de

- 11.-13.09. 2013 Kiel / Deutschland
- 34. Arbeits- und Fortbildungstagung der ASpB 2013
- 16.-18.09. 2013 Berlin / Deutschland
- AGMB-Jahrestagung am Campus Virchow-Klinikum der Berliner Charité - Stand 21
- 24.-26.09. 2013 Stuttgart / Deutschland
- DMS Expo 2013 – Stand 5B25
- 25.-28.09. 2013 Saarbrücken / Deutschland
- 83. Deutscher Archivtag – Stand 17
- 09.-13.10. 2013 Frankfurt am Main Buchmesse – Halle 4.2, Stand P 91.

ZEUTSCHEL – die Zukunft der Vergangenheit.



www.zeutschel.de

dard 3 (Bewerten) zugeordnet. Der Test wurde zusätzlich dichotom ausgewertet, d.h., es wurden nur dann Punkte vergeben, wenn eine Aufgabe vollständig richtig beantwortet wurde (also alle Antwortoptionen korrekt angekreuzt oder ausgelassen wurden). Die Befundmuster änderten sich durch diese strengere Auswertung nicht wesentlich, daher wird im Folgenden ausschließlich über die beschriebene Auswertung mit Teilpunkten berichtet.

Der Test wurde von $N = 64$ Studierenden und Doktoranden/innen der Psychologie bearbeitet. Davon waren $n = 22$ Studierende im ersten Jahr ($M = 21.77$ Jahre, $SD = 2.42$; 77% weiblich), $n = 21$ im dritten oder vierten Jahr ihres Studiums (fortgeschrittene Studierende; $M = 23.90$ Jahre, $SD = 1.90$; 86% weiblich) und $n = 21$ Doktorand/innen ($M = 28.48$ Jahre, $SD = 2.90$; 71% weiblich). Für die Studienteilnahme wurden 20 € erstattet.

Die Datenerhebungen fanden in den Räumen des Leibniz-Zentrums ZPID statt. Die Proband/innen füllten zuerst einen kurzen demografischen Fragebogen aus. Danach bearbeiteten sie den beschriebenen Wissenstest, einen Test zum psychologischen Fachwissen und einen Fragebogen zu epistemologischen Überzeugungen. Abschließend wurden die Proband/innen (mit Ausnahme der Doktoranden/innen, die nur die Tests und Fragebögen bearbeiteten) mit einer standardisierten Literaturrechercheaufgabe konfrontiert. Die Rechercheaufgabe wurde online unter standardisierten Bedingungen mit Zugang zu Fachdatenbanken,

zum Internet und zum elektronischen Bibliothekskatalog bearbeitet. Zusätzlich standen zwei thematisch einschlägige Fachbücher als Nachschlagewerke zur Verfügung. Die Aufgabe bestand aus drei Teilen, für deren Bearbeitung insgesamt maximal 30 Minuten vorgesehen waren. Es sollten (1) drei aktuelle Meta-Analysen zur prädiktiven Validität von Assessment-Center-Verfahren (AC; davon zwei Studien aus dem angloamerikanischen Raum und eine Studie aus dem deutschsprachigen Raum), (2) ein Buch oder Buchkapitel zum Thema AC und (3) eine empirische Studie, die über ein AC mit BWL-Studierenden berichtet, gefunden werden. Nur das Ergebnis der Suche nach Studien aus dem angloamerikanischen Raum wurde analysiert, da die übrigen Aufgaben als Füll-Aufgaben konzipiert waren.

Ergebnisse

Nach Ausschluss eines Items, das von allen Proband/innen korrekt beantwortet wurde und damit

keine Varianz aufweist, wurde eine exploratorische Faktorenanalyse (Hauptkomponentenanalyse) mit den verbleibenden 34 Items berechnet. Ihre Befunde zeigen, dass sich die Korrelationsstruktur am besten mit einem Faktor erklären lässt, der 15.46 Prozent der Varianz aufklärt und auf dem 22 Items substantielle Ladungen ($a_{(ij)} > .20$) aufweisen. Inhaltlich sind 14 dieser Items dem Standard 2 (Suche von Informationen) und 8 dem Standard 3 (Bewerten von Informationen) zugeordnet. Berechnet wurden ein Gesamtskalen-Wert und getrennte Subskalen-Werte für die Standards 2 und 3. Die Subskalen-Werte korrelieren signifikant ($r = .60$, $p < .01$). Sowohl die Gesamtskala als auch die Subskalen weisen akzeptable psychometrische Kennwerte für ihre Reliabilität auf (siehe Tabelle 1). Das hier berichtete Maß Cronbach's Alpha (α) misst, wie stark die verschiedenen Items des Tests miteinander korrelieren. In der Literatur gelten Werte von .70 und darüber als angemessen, wenn der Test zur Unterscheidung von Personen verwendet wird. Bei anderen Verwendungen des Tests, beispielsweise Vergleichen zwischen Gruppen, sind auch niedrigere Werte akzeptabel¹².

Tabelle 1: Psychometrische Kennwerte der drei Skalen des Wissenstests in Studie 1 mit $N=64$ Probanden.

Skala des Wissenstests	k	$M (SD)$	α	$r_{xx'}^*$
Gesamtskala	22	.64 (.13)	.82	.78
Standard 2 (Suche)	14	.68 (.13)	.73	.65
Standard 3 (Bewertung)	8	.58 (.17)	.73	.73

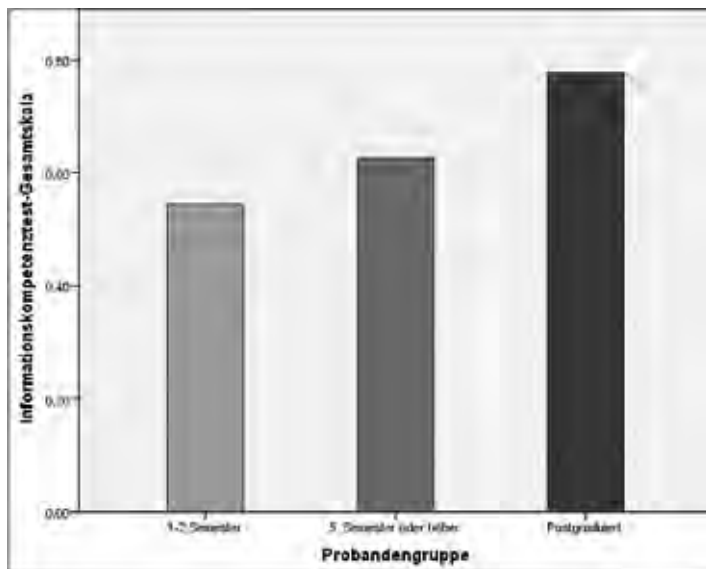
Anmerkungen: k = Anzahl der Items; M = Mittelwert; SD = Standardabweichung; α = Cronbachs Alpha; $r_{xx'}^*$ = Split-Half-Reliabilität (Spearman-Brown-Korrektur)

Zur Prüfung von Unterschieden zwischen den Subgruppen der Studienanfänger, fortgeschrittenen Studierenden und Doktoranden/innen wurden für die drei Skalen separate Varianzanalysen berechnet (siehe Tabelle 2). Die Gruppenunterschiede auf der Gesamtskala werden in Grafik 1 dargestellt. Jede dieser Varianzanalysen belegt signifikante ($p < .01$) Unterschiede zwischen den Gruppen: Für die Gesamtskala $F(2,61) = 44.50$, für die Skala Suchen $F(2,61) = 27.10$ und für die Skala Bewerten $F(2,61) = 28.29$. Geplante Kontraste zeigen, dass sich auf der Gesamtskala und der Skala Suchen alle drei Gruppen statis-

¹² Vgl. Cortina, Jose: „What is coefficient alpha? An examination of theory and applications“, in: Journal of Applied Psychology 78(1) (1993) S. 98–104.; Schmitt, Neal: „Uses and abuses of coefficient alpha“, in: Psychological Assessment 8(4) (1996) S. 350–353.

Tabelle 2: Deskriptive Statistiken der drei Skalen des Wissenstests für die drei Probandengruppen.

Skala des Wissenstests	Studienanfänger (<i>n</i> = 22) <i>M</i> (<i>SD</i>)	Fortgeschrittene Studierende (<i>n</i> = 21) <i>M</i> (<i>SD</i>)	Doktoranden/innen (<i>n</i> = 21) <i>M</i> (<i>SD</i>)
Gesamtskala	.54 (.09)	.62 (.10)	.78 (.07)
Standard 2 (Suche)	.59 (.11)	.68 (.09)	.79 (.06)
Standard 3 (Bewertung)	.47 (.13)	.53 (.13)	.74 (.11)



$p < .05$) höhere Werte erzielen (siehe Tabelle 3). Auf der Subskala Bewerten wird der Unterschied zwischen den Gruppen statistisch nicht signifikant ($t[38] = .97, p > .15$).

tisch bedeutsam voneinander unterscheiden. Auf der Skala Bewerten erweisen sich nur die Unterschiede zwischen Studienanfängern und Doktoranden/innen sowie zwischen fortgeschrittenen Studierenden und Doktoranden/innen signifikant; der Unterschied zwischen Studienanfängern und fortgeschrittenen Studierenden erreicht nicht die kritische Signifikanzgrenze, weist aber in die erwartete Richtung.

In den Analysen der Zusammenhänge zwischen den Leistungen im Wissenstest und denen in der Recherche-Aufgabe mussten die Daten dreier Proband/innen unberücksichtigt bleiben, weil während der Recherche Computerprobleme auftraten. Die Leistungen in der standardisierten Informationsrechercheaufgabe wurde nach einer Musterlösung ausgewertet, bei der die Anforderung darin bestand, zumindest eine von vier in Fachdatenbanken (PsycINFO, PSYNDEX^{plus}) dokumentierten Publikationen zu Metaanalysen zur Validität von Assessment-Center-Verfahren zu identifizieren. Von den 40 Proband/innen gelang dies $n = 29$ (72,5%). Vergleiche dieser Gruppe mit der Gruppe der Proband/innen, die die Rechercheaufgabe nicht lösten, zeigen, dass die in der Rechercheaufgabe erfolgreichen Proband/innen auf der Gesamtskala des Wissenstests ($t[38] = 2.18, p < .05$) und auf der Skala Suchen ($t[38] = 2.35$,

Tabelle 3: Mittelwerte und Standardabweichungen der drei Skalen des Wissenstests für in der Rechercheaufgabe erfolgreiche vs. nicht erfolgreiche Probanden.

Rechercheaufgabe	Wissenstest Gesamtskala <i>M</i> (<i>SD</i>)	Wissenstest Standard 2 (Suche) <i>M</i> (<i>SD</i>)	Wissenstest Standard 3 (Bewertung) <i>M</i> (<i>SD</i>)
Erfolgreich (<i>n</i> = 29)	.60 (.09)	.65 (.11)	.51 (.13)
Nicht erfolgreich (<i>n</i> = 11)	.52 (.10)	.56 (.10)	.46 (.13)

Diskussion der Befunde von Studie 1

Die ersten Befunde zu psychometrischen Qualität des Wissenstests zur Informationsrecherche sind ermutigend. Empirisch belegt werden konnte eine hinreichende Reliabilität der Gesamtskala sowie der Subskalen zum Suchen und Bewerten von Informationen. Für die differentielle Validität des Tests sprechen die Testunterschiede zwischen Studienanfängern, fortgeschrittenen Studierenden und Doktoranden/innen der Psychologie. Zudem weisen die Testgesamtskala und die Subskala Suchen sinnhafte Bezüge zu den Leistungen in einer stan-

dardisierten Rechercheaufgabe auf, was für ihre konvergente Validität spricht¹³.

Studie 2: Experteneinschätzungen der Items des Wissenstests

Zielsetzung von Studie 2 war es, über eine Expertenbefragung zu kriterienbezogenen Einschätzungen der Schwierigkeit der in Studie 1 verwendeten 35 Items des Wissenstests zu gelangen, indem diese vier Kompetenzstufen zugeordnet wurden.

Methoden

Den Expert/innen wurde der Fragebogen in der Form zugesandt, wie er auch den Proband/innen in Studie 1 vorgelegt wurde. Gebeten wurde aber darum, jedes Item einer der von vier Schwierigkeits- bzw. Kompetenzstufen zuzuordnen; hierzu wurden den Expert/innen kurze schriftliche Beschreibungen der vier Kompetenzstufen vorgelegt. Weiterhin wurden die Expert/innen um inhaltliche Anmerkungen gebeten, für den Fall, dass ihnen ungenau formulierte Items oder Lücken auffallen.

Es wurden 12 Expert/innen postalisch kontaktiert, von denen acht in der Dokumentationsabteilung des ZPID arbeiteten. Die in der Dokumentationsabteilung beschäftigten Personen sind für die Inhalte der Psychologie-Fachdatenbank PSYNDEX^{plus} verantwortlich und besitzen daher Erfahrung in der Literaturrecherche sowie weitreichende Kenntnisse über den Aufbau von Fachdatenbanken. Weiterhin wurden vier erfahrene Fachwissenschaftler/innen, die intensiv mit dem Thema Recherche und Bewertung von Fachliteratur vertraut sind, angeschrieben. Von den 12 versendeten Fragebögen wurden 10 zurück erhalten.

Ergebnisse

Bei der Auswertung zeigte sich, dass die Kompetenzstufenzuordnungen der Items durch die Expert/innen sehr unterschiedlich ausfielen. Die Übereinstimmung zwischen den Expert/innen wurde mit dem Maß Cohens Kappa ermittelt. Die so ermittelten Werte lagen zwischen $\kappa = .10$ und $\kappa = .20$ – und damit im geringfügigen Bereich¹⁴. Eine eindeutige Zuordnung von Items zu Kompetenzstufen ist auf Grundlage derart heterogener Urteile nicht möglich. Entsprechend können die Experteneinschätzungen nicht (wie ursprünglich geplant) mit den empirischen Schwierigkeitsindizes

der Items aus Studie 1 verglichen werden. Dass die Expert/innen die Schwierigkeiten der Items so unterschiedlich beurteilen, könnte darauf zurückgehen, dass die Expert/innen kein einheitliches Verständnis des Konzepts Informationskompetenz besitzen. Eine Alternativerklärung könnte sein, dass die schriftlichen Beschreibungen der Kompetenzstufen nicht ausreichten, um bei den Expert/innen ein einheitliches Verständnis der vier Schwierigkeitsstufen zu erzeugen. Die Kommentare zum Inhalt der Items wurden verwendet, um vier Items inhaltlich bzw. sprachlich zu optimieren.

Studie 3: Online-Erhebung

Die Online-Studie wurde durchgeführt, um a) für die revidierte Version des Tests Itemkennwerte zu ermitteln und b) die Eignung des Verfahrens für ein Online-Erhebungsformat sicherzustellen, welches in der Evaluationsstudie des geplanten Trainings zu Informationskompetenzen zum Einsatz kommen sollte. Zudem wurde erwartet, dass sich wie in der Studie 1 akzeptable psychometrische Kennwerte ergeben und dass sich erneut Gruppenunterschiede in Abhängigkeit von der bisherigen Studiendauer beobachten lassen.

Der in der Online-Studie verwendete Fragebogen enthielt wiederum 35 Items. Diese 35 Items wurden in drei Gruppen eingeteilt: fachunspezifische, adaptierbare fachspezifische und nicht-adaptierbare fach- bzw. ortsspezifische Items.

(1) Fachspezifische Items ($k = 22$) beziehen sich auf grundlegende Aspekte der Informationsrecherche und Bewertung (vorrangig) im Internet. Sie sind damit universell, d.h. nicht nur bei Psychologiestudierenden, sondern auch bei Studierenden anderer Fächer sowie unter Umständen bei Zielgruppen außerhalb des Hochschulbereichs einsetzbar.

Beispiel für ein fachunspezifisches Item: Welche Option ist am sinnvollsten, wenn Literatur nicht in Ihrer (Universitäts)bibliothek zu finden ist?

- Bei Bibliotheken in benachbarten Städten anfragen, im Zweifelsfall dort hinfahren.
- Die Literatur per Fernleihe bestellen.
- Es gibt keine Alternative zum Kauf des Buchs/der Zeitschrift.

(2) Psychologiespezifische Items mit Platzhaltern ($k = 6$) sind so formuliert, dass sie in der Regel nur von Proband/innen mit Psychologiestudium beantwortet werden können, allerdings sind die psychologiespezifischen Inhalte austauschbar, womit die Items an andere Fachbereiche angepasst werden können.

Beispiel für ein psychologiespezifisches Item mit

¹³ Vgl. Leichner, Nikolas/ Peter, Johannes/ Mayer, Anne-Kathrin/ Krampen, Günter: „Assessing information literacy among German psychology students“, in: Reference Services Review (im Druck).

¹⁴ Vgl. Landis, J. Richard/ Koch, Gary: „The measurement of observer agreement for categorical data“, in: Biometrics 33(1) (1977) S. 159–174.

Platzhalter Item: Welches sind wichtige Fachdatenbanken für die Psychologie?

- PsycINFO
- PSYINDEX
- PsycBASE

Hier können beispielsweise die Namen der Psychologie-Fachdatenbanken durch Namen von Datenbanken des jeweiligen Fachgebiets ersetzt werden.

(3) Nicht-adaptierbare fach- bzw. ortsspezifische Items ($k = 7$) sind enthalten, da nur so spezifisches Wissen erhoben werden kann.

Beispiel für ein fach- bzw. ortsspezifisches Item: Welche Aussagen über den elektronischen Bibliothekskatalog sind korrekt?

- Zeigt an, ob ein Buch gerade ausgeliehen ist.
- Zeigt an, ob die Bibliothek eine bestimmte Zeitschrift abonniert hat.
- Bietet die Möglichkeit, nicht vorhandene oder ausgeliehene Bücher per Fernleihe zu bestellen.

Sollte der Fragebogen nicht bei Psychologiestudierenden eingesetzt werden, ist es dadurch möglich, nur die relevanten fachunspezifischen Items zu verwenden oder zudem die spezifischen Items anzupassen und ebenfalls zu verwenden.

Aus Studie 1 wurden 19 Items übernommen. Da nun entschieden war, dass der Test primär für die Evaluation von Trainings zur Förderung von Informationskompetenzen eingesetzt werden sollte, wurden diejenigen Items aus Studie 1 nicht berücksichtigt, deren Inhalt im geplanten Training nicht abgebildet wird. Zudem wurden neue Items entwickelt, die auf den Inhalt des geplanten Informationskompetenztrainings abgestimmt sind. Weiterhin wurden vier Vorstudienitems verändert, da sie nach Expertenmeinung inhaltlich ungenau formuliert waren (s. Abschnitt 3.2) oder sich in der Zwischenzeit Änderungen am elektronischen Bibliothekskatalog vor Ort ereignet hatten, die die ursprünglichen Lösungsoptionen einzelner Items unzutreffend machten.

Methode

Der in der Online-Studie eingesetzte Wissenstest besteht aus $k = 35$ Items, die mit einer Ausnahme (offene Antwort) ein Multiple-Choice-Format aufweisen. Die Items besitzen drei Antwortoptionen und sind so skaliert, dass der mögliche Punktwert von 0 bis 1 reicht. Das bedeutet, für jede korrekt angekreuzte oder ausgelassene Antwortalternative werden 0.33 Punkte vergeben. Von den Items sind 22 der Skala Suchen und 13 der Skala Bewerten zugeordnet. $N = 184$ Psychologie-Studierende (Hauptfach) aus

allen Fachsemestern und Studiengängen (BA, MA und Diplom) bearbeiteten den Test und die anderen in Studie 1 verwendeten Instrumente online über die Befragungssoftware „Unipark“. Von den $N = 184$ Proband/innen waren $n = 8$ (4,4%) im Diplomstudiengang, $n = 145$ (79,0%) im BA-Studiengang und $n = 30$ (16,6%) im MA-Studiengang eingeschrieben. Die meisten Probanden waren weiblich ($n = 157$; entsprechend 85,3%). Die Proband/innen erhielten € 10 für das Ausfüllen der Instrumente.

Ergebnisse

Eine Hauptkomponentenanalyse weist auf die Existenz von drei Faktoren hin, die zusammen ca. 20% der Varianz aufklären. Da alle Items für die Evaluation des Trainings relevant sind, wurden keine Items aufgrund der Faktorenanalyse oder aufgrund psychometrischer Kriterien aus dem Test entfernt. Es wurden eine aus allen $k = 35$ Items bestehende Gesamtskala und Subskalen mit $k = 22$ bzw. $k = 13$ Items gebildet. Die beiden Subskalen korrelieren signifikant miteinander ($r = .28, p < .01$).

Die interne Konsistenz der Skalen liegt in dieser Studie wesentlich unter den Werten in Studie 1 (siehe Tabelle 4).

Tabelle 4: Psychometrische Kennwerte für die drei Skalen des Wissenstests, $N = 184$.

Skala des Wissenstests	k	$M (SD)$	α	r_{xx}^*
Gesamtskala	35	.62 (.05)	.49	.45
Standard 2 (Suche)	22	.62 (.05)	.22	.32
Standard 3 (Bewertung)	13	.61 (.10)	.48	.42

Ebenso wie in Studie 1 zeigen sich jedoch Unterschiede zwischen den Probandengruppen auf allen drei Skalen (siehe Tabelle 5).

Tabelle 5: Deskriptive Statistiken der drei Skalen des Wissenstests für die drei Probandengruppen.

Skala des Wissenstests	1. BA 1.-3. Sem. ($n=74$) $M (SD)$	2. BA ab 4. Sem. ($n=72$) $M (SD)$	3. MA und Diplom ($n=38$) $M (SD)$
Gesamtskala	.60 (.05)	.63 (.06)	.65 (.05)
Standard 2 (Suche)	.60 (.06)	.63 (.06)	.65 (.04)
Standard 3 (Bewertung)	.58 (.09)	.62 (.09)	.64 (.10)

Die Gruppenunterschiede erweisen sich in Varianzanalysen als signifikant. Dies gilt sowohl für die Gesamtskala ($F[2,181]=14.90, p < .01$) als auch für die Skala Suchen ($F[2,181] = 13.78, p < .01$) und für die Skala Bewerten ($F[2,181] = 5.46, p < .01$). Post-Hoc-Tests mit Bonferroni-Korrektur zeigen auf der Gesamtskala und der Skala Suchen signifikante Unterschiede zwischen den Gruppen – mit Ausnahme der Differenz zwischen zweiter und dritter Gruppe. Auf der Bewerten-Skala zeigen sich nur zwischen der ersten und der dritten Gruppe signifikante Unterschiede.

Diskussion

Verglichen mit den Ergebnissen von Studie 1 sind die Befunde von Studie 3 enttäuschend. Es wurden zwar erwartungskonforme Unterschiede zwischen den Probandengruppen gefunden, die allerdings nicht so ausgeprägt sind wie erwartet. Die psychometrischen Kennwerte (Skalenreliabilitäten und Trennschärfen) liegen zudem wesentlich niedriger.

Mit Blick auf mögliche Erklärungen dieser Befunde sind zwei Unterschiede zwischen Studie 3 und Studie 1 besonders auffällig: Zum einen wurde in Studie 3 der Test online, d.h. in unkontrollierten Umgebungen, bearbeitet. Zum zweiten war die Probandenstichprobe in der Online-Erhebung homogener: Im Gegensatz zu Studie 1 wurden nicht Extremgruppen (Studienanfänger und Fortgeschrittene) rekrutiert und überdies fehlt die Teilgruppe der Doktoranden. Es ist zu vermuten, dass die Proband/innen beim Ausfüllen des Tests unter unkontrollierten Umgebungen unkonzentrierter sind, wodurch der Messfehler ansteigt. Der erhöhte Messfehler kann die Unterschiede zwischen den Probandengruppen verzerrt haben, außerdem führt ein hoher Messfehler möglicherweise zu inakzeptablen psychometrischen Kennwerten.

Die Homogenität der Stichprobe kann ebenfalls Ursache für die enttäuschenden Befunde sein. Dadurch sinkt die Varianz zwischen den Proband/innen – damit werden die Gruppenunterschiede undeutlicher. Zudem ist die niedrige Varianz eine wahrscheinliche Ursache für die niedrige interne Konsistenz, da Korrelationen zwischen Items Varianz erfordern. Um diese Annahme zu prüfen wurde die interne Konsistenz für die einzelnen Probandengruppen in Studie 1 bestimmt. Die interne Konsistenz für die Gesamtskala lag in der Gruppe der Studienanfänger bei $\alpha = .63$, in der Gruppe der fortgeschrittenen Studierenden bei $\alpha = .66$ und in der Gruppe der Doktoranden bei $\alpha = .41$. Dies spricht dafür, dass die in Studie 1 gefundene befriedigende interne Konsistenz zum Teil auf eine hohe

Varianz zwischen den Proband/innen zurückzuführen ist.

Zuletzt wurde die Vermutung geprüft, dass die 16 Items, die nicht in Studie 1 eingesetzt wurden, für die unerwarteten Ergebnisse verantwortlich sind. Dies ist jedoch nicht der Fall: Werden nur die in Studie 1 eingesetzten Items ausgewertet, liegt die interne Konsistenz der Gesamtskala bei $\alpha = .50$ – und damit nicht höher als der Wert für alle Items. Eine Verzerrung der Ergebnisse durch die neuen Items ist daher unwahrscheinlich.

Gesamtdiskussion

Die aufgrund der Befunde von Studie 1 gezogene Schlussfolgerung, dass es sich bei Wissen über Informationsrecherche um ein eindimensionales Konstrukt handelt, das mithilfe eines Wissenstests erhoben werden kann, lässt sich aufgrund der Ergebnisse von Studie 3 nicht aufrecht erhalten. In dieser Studie erwies es sich als nicht möglich, aus den Items des Wissenstests eine Skala mit akzeptabler Messgenauigkeit zu bilden. Auch die Konstruktion von reliablen Subskalen zu den Wissensfacetten „Suche“ und „Bewertung“ ist nicht gelungen.

Mögliche Erklärungen für diese auf den ersten Blick enttäuschenden Befundmuster lassen sich auf empirischer wie auch auf konzeptueller Ebene lokalisieren. Auf *empirischer Ebene* kann auf die vergleichsweise hohe Homogenität der Online-Erhebungsstichprobe (Studie 3) sowie die unkontrollierten Erhebungsbedingungen verwiesen werden; beide Faktoren erschweren das Erreichen hoher Reliabilitätskennwerte. Demgegenüber zeigte sich, dass der Test durchaus akzeptable psychometrische Kennwerte erreichen kann, wenn die Probandenstichprobe hinreichend leistungsheterogen ist und unter standardisierten Bedingungen untersucht wird, wie dies in Studie 1 der Fall war.

Auf *konzeptueller Ebene* erscheint es fraglich, ob es sich bei dem Wissen über Informationsrecherche um ein einheitliches Konstrukt handelt. Die Annahme der Eindimensionalität wird nicht nur durch die Ergebnisse der beiden Fragebogenstudien, sondern auch durch die Ergebnisse der Expertenbefragung (Studie 2) in Frage gestellt. In dieser Befragung zeigte sich, dass die Expertenurteile über die Kompetenzstufenzuordnungen sehr uneinheitlich ausfielen. Dies lässt die Vermutung zu, dass selbst die Expertengruppe kein einheitliches Verständnis des Konstrukts besitzt. Diese Überlegungen führen zu der Schlussfolgerung, dass es sich bei dem durch den Test erfassten deklarativen Wissen über Informationsrecherche und -bewertung um Wissen aus einer sehr breiten und inhalt-

lich heterogenen Domäne handelt. Möglicherweise handelt es sich um eine Art „fragmentarisches Wissen“ – etwa im Sinne der *knowledge in pieces*-Theorie¹⁵. Gegenstand der *knowledge in pieces*-Theorie ist die Natur von mechanischem Verständnis (oder anderen physikalischen Phänomenen) bei Schülern. Postuliert wird, dass bereits Grundschul Kinder über ein Vorwissen aus dem Bereich der Mechanik verfügen. Dieses Wissen liegt jedoch nicht in strukturierter Form vor, sondern beinhaltet lediglich Wissensfragmente, die einzelne Phänomene erklären können, aber kein konsistentes Gesamtbild im Sinne eines Wissenssystems ergeben, aus dem sich bspw. Hypothesen über unbekannte Problemstellungen ableiten ließen. Diese Theorie wurde mittlerweile auf andere Wissensbereiche, z.B. Mathematik¹⁶, erfolgreich angewendet. Im Laufe der Entwicklung bzw. weiteren Schulbildung werden nicht nur neue Wissensfragmente hinzugefügt, sondern auch die bestehenden Fragmente besser miteinander verknüpft. Ähnlich kann man sich möglicherweise die Entwicklung von Wissen über Informationsrecherche bei Studierenden vorstellen. Einzelne Elemente des im Test erfragten Wissens werden vermutlich in informellen Lern-Settings mehr oder weniger zufällig oder in formellen Lern-Settings eher nebenbei erworben, z.B. durch die Notwendigkeit, für ein Referat eine bestimmte Rechentechnik anzuwenden. Die gesamte Wissensdomäne ist allerdings sehr breit und kann, im Gegensatz zu mathematischen oder naturwissenschaftlichen Wissensdomänen, nicht durch übergeordnete abstrakte Prinzipien systematisiert werden. Zudem befindet sich das Wissen in einem teils rapiden und unvorhersehbaren Wandel (da sich z.B. relevante Technologien und die damit verbundenen Rechercheoptionen schnell weiterentwickeln). Dies kann erklären, warum das Wissen auch bei „Expert/innen“ (der Doktorandenstichprobe in Studie 1 bzw. den Expert/innen in Studie 2) in fragmentierter Form vorliegt. Mit anderen Worten: Expert/innen besitzen zwar mehr Wissen und erzielen in dem Test höhere Punktwerte, da sie mehr Erfahrungen mit verschiedenen Aspekten der Informationssuche und -bewertung haben. Gleichwohl bleibt das (Detail-)Wissen teilweise unverbunden, da es nicht durch erlernbare Regeln systematisiert ist. Psychometrisch impliziert dies, dass mit der Lösungswahrscheinlichkeit eines Items nicht

¹⁵ diSessa, Andrea: „Toward an epistemology of physics“, in: *Cognition and Instruction* 10(2-3) (1993) S. 105–225.

¹⁶ Izsak, Andrew: „You have to count the squares“: Applying knowledge in pieces to learning rectangular area“, in: *Journal of the Learning Sciences* 14(3) (2005) S. 361–403.



BÜCHER SCHÜTZEN – WERTE ERHALTEN!



PROFI-BUCHSCHUTZFOLIEN VON NESCHEN

- ✓ erhöhen die Lebensdauer von Büchern
- ✓ schützen vor Verschmutzung
- ✓ auch mit reduzierter Anfangsklebkraft erhältlich
- ✓ lösungsmittelfreie alterungsbeständige Acrylatkleber

Fragen? Sie erreichen uns unter 05722-207-0 oder Sie senden uns eine E-Mail an neschen@neschen.de.

NESCHEN AG

Hans-Neschen-Str. 1 | D-31675 Bückeburg
Telefon: 0 57 22/2 07-0 | Fax: 0 57 22/2 07-197
E-Mail: neschen@neschen.de

www.neschen.com

zwangsläufig die Lösungswahrscheinlichkeit für die anderen Items steigt, so dass die interne Konsistenz des Tests niedrig bleibt. Eine ähnliche Argumentation findet sich auch in der Literatur¹⁷.

Welche Implikationen besitzen diese Überlegungen für die Erfassung von Wissen über Informationsrecherchen und für die künftige Nutzung des Wissenstests? Sofern es sich bei dem Wissen über Informationsrecherche und -bewertung tatsächlich um heterogenes, fragmentiertes Wissen im oben erläuterten Sinn handelt, ist eine akzeptable interne Konsistenz zumindest in relativ leistungshomogenen Stichproben nicht zu erwarten. Gleichwohl bildet der Test-Gesamtscore interindividuelle Unterschiede in der Menge vorliegenden Wissens ab, wie die in zwei Studien ermittelten systematischen Unterschiede zwischen Studienanfängern und Fortgeschrittenen belegen.

Da der Test mit Blick auf die Inhalte des Trainings für Informationskompetenz entwickelt wurde, kann ihm zudem curriculare Validität¹⁸ zugesprochen werden. Der Test kann also für die Evaluation des Trainings verwendet werden. Dies bedeutet jedoch, dass es sich um einen informellen Test¹⁹ handelt, der für die Evaluation des Trainings konzipiert wurde. Ob der Test für die allgemeine Erhebung von Informationskompetenz einsetzbar ist, bleibt zunächst fraglich. Die Gruppenunterschiede weisen jedoch darauf hin, dass der Inhalt des Tests über das zu konzipierende Training hinaus Geltung besitzen könnte.

Mit Blick auf die gewählte Operationalisierung ist schließlich zu betonen, dass der Test lediglich *einen* Indikator für Informationskompetenz (nämlich deklaratives Wissen über die Informationssuche und -bewertung) erfasst. Wie oben erwähnt, sollte Informationskompetenz im Zuge einer multimodalen Konstrukterfassung ergänzend durch behaviorale Maße, wie z.B. die Leistungen in Recherche- und Bewertungsaufgaben, abgebildet werden. Die Korrelation zwischen dem Testwert und der Leistung in der Rechercheaufgabe, die sich in der Studie 1 ergab, weist darauf hin, dass beide Verfahren als Indikatoren einer Fertigkeit angesehen werden können. Ergänzend könnte erwogen werden, die durch den Test erfasste breite Wissensdomäne durch weitere Methoden wie Essays oder Strukturdarstellungen (z.B. „concept

maps“) zu erfassen²⁰. Dass eine solche Erhebung von Wissensstrukturen jedoch nicht nur sehr aufwändig, sondern auch keinesfalls unproblematisch ist, zeigt eine Studie, die belegt, dass die Erfassung von Wissensstrukturen durch „concept maps“ und durch Fragen mit offenen Antwortmöglichkeiten nicht äquivalent ist²¹. ■



**Nikolas Leichner, Johannes Peter,
Anne-Kathrin Mayer und Günter Krampen**

Leibniz-Zentrum für Psychologische Information
und Dokumentation (ZPID)

Universität Trier

54286 Trier

leichner@zpid.de

17 Schneider, Michael/ Rittle-Johnson, Bethany/ Star, Jon: „Relations among conceptual knowledge, procedural knowledge, and procedural flexibility in two samples differing in prior knowledge“, in: *Developmental Psychology* 47(6) (2011) S. 1525–1538.

18 Vgl. Hartig, Johannes/ Frey, Andreas/ Jude, Nina: Validität, in: Hefried Moosbrugger/ Augustin Kelava (Hrsg.): *Testtheorie und Fragebogenkonstruktion*, Berlin/ Heidelberg 2012.

19 Vgl. Kleber, Eduard Werner: *Tests in der Schule*, München 1979.

20 Vgl. Mandl, Heinz/ Fischer, Frank: *Wissen sichtbar machen. Wissensmanagement mit Mapping-Techniken*, Göttingen 2000.

21 Conradt, Catherine/ Bogner, Franz: „Knowledge presented in concept maps: correlations with conventional cognitive knowledge tests“, in: *Educational Studies* 38(3) (2012) S. 341–354.