

Multimethodale Erfassung von Informationskompetenz bei Psychologiestudierenden

Dissertation zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften des Fachbereichs I der
Universität Trier

vorgelegt von

Nikolas Leichner

Trier, 2015

1. GutachterIn: Prof. Dr. Günter Krampen
2. GutachterIn: Dr. Anne-Kathrin Mayer

Danksagung

Hiermit möchte ich meinem Arbeitgeber, dem Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID), für die Ermöglichung dieser Arbeit sowie meinen Betreuer/innen, Herrn Prof. Dr. Günter Krampen und Frau Dr. Anne-Kathrin Mayer für die wertvolle inhaltliche und konzeptuelle Hilfe danken. Weiterer Dank gilt den Korrekturleser/innen. Besonderer Dank gilt meinen Eltern, die mich während des gesamten Studiums, wie auch während der Anfertigung der Dissertationsschrift, unterstützt haben.

Zusammenfassung

Mit fachlicher Informationskompetenz werden die Fähigkeiten bezeichnet, die notwendig sind, um effektiv Fachinformationen (z.B. Zeitschriftenartikel) finden und bewerten zu können. Informationskompetenz gilt als wichtige Voraussetzung für ein erfolgreiches Studium und Erfolg auf dem Arbeitsmarkt. Ziel dieser Arbeit ist es, Messinstrumente für fachliche Informationskompetenz bei Psychologiestudierenden zu entwickeln. Dabei wird davon ausgegangen, dass Informationskompetenz sowohl deklarative, als auch prozedurale Wissenskomponenten (Hintergrundwissen über Konzepte und die Beherrschung von Strategien und Vorgehensweisen) enthält. Um diese erfassen zu können, wurde ein Multiple-Choice-Test und Rechercheaufgaben entwickelt; beide Instrumente wurden zur Evaluation eines Trainings für Informationskompetenz eingesetzt. Mittlere Korrelationen zwischen den Instrumenten und die Ergebnisse einer Clusteranalyse zeigen, dass mit beiden Instrumenten unterschiedliche Wissenskomponenten erfasst werden.

Inhaltsverzeichnis

| | | |
|-------|--|----|
| 1 | Einleitung | 1 |
| 2 | Entstehung | 4 |
| 2.1 | Entwicklung in den USA | 4 |
| 2.2 | Situation in Deutschland..... | 5 |
| 2.3 | Bildungspolitischer Kontext | 6 |
| 3 | Modelle von Informationskompetenz | 8 |
| 3.1 | Seven Pillars for Information Literacy | 8 |
| 3.2 | Seven Faces of Information Literacy | 9 |
| 3.3 | Prozessmodelle von Informationskompetenz..... | 9 |
| 3.3.1 | Information Search Process (ISP) | 10 |
| 3.3.2 | The Big Six Skills..... | 12 |
| 3.3.3 | Das dynamische Modell der Informationskompetenz (DYMIK)..... | 13 |
| 3.4 | Zusammenfassung | 14 |
| 4 | Standards der Informationskompetenz | 16 |
| 4.1 | Information Literacy Standards for Higher Education..... | 16 |
| 4.2 | Standards der Informationskompetenz für Studierende | 18 |
| 4.3 | Exkurs: "Information Behavior" | 19 |
| 4.4 | Zusammenfassung | 20 |
| 5 | Förderung von Informationskompetenz..... | 21 |
| 5.1 | Inhalte | 21 |
| 5.2 | Lehrveranstaltungen..... | 23 |
| 5.3 | Online-Lehre | 25 |
| 5.4 | Exkurs: Fachliche Recherchen | 26 |

| | | |
|--------|--|----|
| 5.5 | Schlussfolgerung..... | 28 |
| 6 | Kritik..... | 29 |
| 7 | Vergleich mit verwandten Konzepten | 31 |
| 8 | Selbstwirksamkeit | 33 |
| 9 | Kompetenzbegriff | 35 |
| 9.1 | Deklaratives und prozedurales Wissen | 36 |
| 10 | Erfassung von Informationskompetenz im fachlichen Kontext | 39 |
| 10.1 | Standardisierte Tests | 39 |
| 10.2 | Performance Assessments | 41 |
| 10.3 | Schlussfolgerung..... | 43 |
| 11 | Entwicklung von Multiple-Choice-Test und Selbstwirksamkeitsskala | 44 |
| 11.1 | Entwicklung des Multiple-Choice-Tests..... | 44 |
| 11.2 | Entwicklung der Selbstwirksamkeitsskala | 45 |
| 11.3 | Vorstudie | 46 |
| 11.3.1 | Methode..... | 46 |
| 11.3.2 | Ergebnisse | 47 |
| 11.4 | Online-Studie | 49 |
| 11.4.1 | Methode..... | 49 |
| 11.4.2 | Ergebnisse | 50 |
| 11.5 | Diskussion | 52 |
| 12 | Entwicklung der Rechercheaufgaben..... | 57 |
| 13 | Multimethodale Trainingsevaluation..... | 63 |
| 13.1 | Methode | 64 |
| 13.1.1 | Hypothesen | 65 |
| 13.2 | Ergebnisse..... | 66 |

| | | |
|--------|--|-----|
| 13.2.1 | Hypothese 1 | 69 |
| 13.2.2 | Hypothese 2 | 70 |
| 13.2.3 | Hypothese 3 | 70 |
| 13.2.4 | Hypothese 4 | 71 |
| 13.2.5 | Hypothese 5 | 75 |
| 13.3 | Diskussion | 78 |
| 13.3.1 | Hypothesen 1-3 | 79 |
| 13.3.2 | Hypothese 4 | 81 |
| 13.3.3 | Hypothese 5 | 82 |
| 14 | Gesamtdiskussion..... | 84 |
| 15 | Literatur..... | 89 |
| 16 | Anhang..... | 99 |
| 16.1 | Multiple-Choice-Test | 99 |
| 16.2 | Selbstwirksamkeitsskala | 105 |
| 16.3 | Rechercheaufgaben | 106 |
| 16.3.1 | Messzeitpunkt 1 | 106 |
| 16.3.2 | Messzeitpunkt 2 | 106 |
| 16.3.3 | Messzeitpunkt 3 | 106 |
| 16.4 | Fragen zur Prozessbewertung der Rechercheaufgaben..... | 106 |
| 16.5 | Auswertungsschlüssel zu den Rechercheaufgaben..... | 107 |
| 16.5.1 | Bewertung des Ergebnisses..... | 107 |
| 16.5.2 | Bewertung des Vorgehens (Prozesspunktwerte)..... | 108 |

1 Einleitung

Mit dem Begriff Informationskompetenz werden auf einer abstrakten Ebene die Fähigkeiten¹ bezeichnet, die notwendig sind, um einen Informationsbedarf zu erkennen, die benötigten Informationen zu identifizieren, beschaffen und bewerten. Schließlich können die neuen Informationen genutzt werden, um das Problem zu lösen (Association of College and Research Libraries (ACRL), 1989, 2000).

Übersichtsarbeiten zeigen, dass unterschiedliche Definitionen des Konstrukts Informationskompetenz vorliegen und sich (bislang) keine durchgesetzt hat (Andretta, 2005; Ingold, 2005). Dies liegt unter anderem daran, dass, je nach Kontext, unterschiedliche Fähigkeiten und Fertigkeiten für eine kompetente Recherche notwendig sind. So sind für eine fachliche Recherche im Hochschulkontext (im Fach Psychologie beispielsweise die Recherche nach Fachartikeln zu einem bestimmten Therapieverfahren) andere Fähigkeiten und anderes Hintergrundwissen erforderlich, als bei einer alltäglichen Recherche – beispielsweise nach verschiedenen Angeboten für die kommende Urlaubsreise. Daher sollte prinzipiell zwischen alltäglicher Informationskompetenz, wie sie beispielsweise durch das sogenannte „Prager Expertentreffen“ (Übers. d. Verf., vgl. Thompson, 2003) beschrieben wird, und fachlicher Informationskompetenz, welche im Hochschulkontext erforderlich ist, unterschieden werden. Fachliche Informationskompetenz kann als Facette von Informationskompetenz angesehen werden (vgl. Leichner, Peter, Mayer, & Krampen, 2014a). Wichtige Komponenten fachlicher Informationskompetenz sind neben technischen Kompetenzen in der Nutzung von Informationssystemen insbesondere auch disziplinspezifisches Wissen über Publikationsmechanismen im jeweiligen Fach (Hochschulrektorenkonferenz (HRK), 2012). Es liegen mehrere Studien vor, die über den Mangel entsprechender Fähigkeiten bei Studierenden berichten, welche aber auf fragwürdigen Messinstrumenten oder Selbstauskünften beruhen (Franke & Schüller-Zwierlein, 2008; Heinze, 2008; Heinze, Fink, & Wolf, 2009). Zugleich wird betont wie wichtig diese Fähigkeiten für ein erfolgreiches Studium sind (HRK, 2012).

Die vorliegende Arbeit beschäftigt sich vor diesem Hintergrund vorwiegend mit der Frage, wie fachliche Informationskompetenz bei Psychologiestudierenden verlässlich erfasst wer-

¹ Für eine Definition der Begriffe Fähigkeiten und Kompetenzen siehe Abschnitt 9, S. 35.

den kann. Dabei wird angenommen, dass Informationskompetenz sowohl deklarative, als auch prozedurale Wissenskomponenten beinhaltet. Unter deklarativen Komponenten werden konzeptuelles Hintergrundwissen über verschiedene fachliche Informationsquellen und deren Umfang und Funktionen verstanden. Prozedurales Wissen bezieht sich in diesem Zusammenhang auf die Kenntnis und Beherrschung verschiedener Recherchestrategien und Vorgehensweisen. Für diese Wissenskomponenten wurden zwei separate Messinstrumente entwickelt: Ein Test im Multiple-Choice-Format zur Erfassung deklarativer und Rechercheaufgaben zur Erfassung prozeduraler Wissenskomponenten.

In der vorliegenden Arbeit wird in Kapiteln 2 bis 4 die Entstehung des Begriffs Informationskompetenz skizziert sowie Modelle und Standards von Informationskompetenz vorgestellt. Dabei wird deutlich, dass der Begriff im bibliothekarischen Umfeld entstand und meist im Rahmen von Maßnahmen zur Förderung von Recherchefähigkeiten verwendet wird. Kapitel 5 beschäftigt sich mit der Förderung von Informationskompetenz. In diesem Kapitel wird diskutiert, inwieweit sich entsprechende Fähigkeiten überhaupt in Schulungsveranstaltungen vermitteln lassen und welche Probleme bei der Vermittlung der entsprechenden Inhalte bestehen. Eine Bestandsaufnahme zeigt, dass Informationskompetenzschulungen von den wenigsten deutschen Universitäten in die Curricula aufgenommen wurden. Um Studierende zur freiwilligen Teilnahme zu ermutigen, sollten Schulungen so gestaltet werden, dass sie zeiteffizient sind und zeitliche Flexibilität bieten. Hierfür scheint Online-Lehre, oder die Verknüpfung von Online- und Präsenzlehre (sogenanntes "Blended Learning"), besonders gut geeignet. Es folgt in Kapitel 6 eine Beschäftigung mit kritischen Stimmen zum Konzept von Informationskompetenz und dessen Umsetzung und in Kapitel 7 der Vergleich mit den verwandten Kompetenzfeldern Medienkompetenz und "computer literacy". In Kapitel 8 wird auf das Konzept der Selbstwirksamkeit eingegangen, wobei deutlich wird, dass Selbstwirksamkeit nicht nur als Förderer von Leistung, sondern auch als Ergebnis vergangener Leistung verstanden werden kann. Anschließend wird in Kapitel 9 der Kompetenzbegriff und die Unterscheidung von deklarativem und prozeduralem Wissen besprochen; in Kapitel 10 geht es um Möglichkeiten, Informationskompetenz vor dem Hintergrund dieser beiden Wissensarten zu erfassen. Im empirischen Teil wird in Kapitel 11 zunächst über die Entwicklung des Multiple-Choice-Tests und einer Skala zur Erfassung der Selbstwirksamkeit bei der fachlichen Recherche berichtet. Anschließend wird über die Entwicklung der Rechercheaufgaben (Kap. 12)

berichtet. Kapitel 13 behandelt die Evaluation eines Blended Learning-Seminars für Informationskompetenz mithilfe der Rechercheaufgaben und des Multiple-Choice-Tests. Die Befunde sprechen dafür, dass mit den Instrumenten unterschiedliche Fähigkeiten erfasst werden.

2 Entstehung

Interessanterweise betont ein nicht unerheblicher Teil der Publikationen zum Thema Informationskompetenz die Notwendigkeit, diese zu fördern oder präsentiert Konzepte zur Förderung. Der enge Bezug zu Förderungsmaßnahmen zeigt sich auch darin, dass viele Informationskompetenzmodelle mit dem Ziel konzipiert wurden, Schulungen auf eine theoretische Grundlage stellen zu können (z.B. Eisenberg & Berkowitz, 2014; Homann, 2000).

Die Feststellung, dass Hintergrundwissen zur Nutzung der Informationsangebote der Bibliotheken erforderlich ist, ist jedoch nicht neu. So sind Benutzerschulungen durch Bibliothekare in Deutschland im 17. Jahrhundert und in den USA aus dem 19. Jahrhundert belegt (Ingold, 2005). Der Bedarf an Benutzerschulungen stieg im 20. Jahrhundert durch zunehmende Studierendenzahlen und den Einsatz elektronischer Systeme an. Diese Schulungen können als Vorläufer der heutigen Informationskompetenzschulungen durch Bibliothekare angesehen werden (Andretta, 2005; Ingold, 2005).

2.1 Entwicklung in den USA

In den 1960er Jahren begann die eigentliche „Blütezeit“ der Benutzerschulungen durch Bibliothekare (Ingold, 2005). Dazu trugen vor allem steigende Studierendenzahlen und eine wachsende Menge an Publikationen bei (Ingold, 2005). Der Bedarf an Benutzerschulungen stieg in den 1990er Jahren durch die Verfügbarkeit neuer elektronischer Systeme, beispielsweise dem öffentlich zugänglichen Bibliothekskatalog („OPAC“, „Online Public Access Catalogue“), weiter an (Ingold, 2005). In diesen Schulungen, im angloamerikanischen Sprachraum meist als „library instruction“ oder „bibliographic instruction“ bezeichnet, wurden überwiegend Anleitungen vermittelt, wie die Ressourcen der Bibliothek genutzt werden können (Andretta, 2005). Klassischerweise wurden Bibliotheksführungen durchgeführt, sowie die Benutzung des Bibliothekskatalogs und der Fernleihe demonstriert.

Parallel dazu etablierte sich in den USA der Begriff „information literacy“, der zunächst nicht auf die Benutzung von Bibliotheken bezogen wurde, sondern auf die Fähigkeit, mit Informationen im Alltag und am Arbeitsplatz umgehen zu können (Grassian & Kaplowitz, 2001). Den Einzug in das (universitäre) Bibliothekswesen feierte der Begriff spätestens mit dem Schluss-

bericht des “American Library Association Presidential Committee on Information Literacy” im Jahre 1989 (vgl. Balceris, 2011).

Schulungen zur Informationskompetenz (“information literacy instruction”) ersetzen nach und nach die Benutzerschulungen. Hierbei handelte es sich nicht nur um einen Wechsel des Etiketts, sondern um eine Erweiterung der Inhalte als Reaktion auf veränderte Bedingungen (Andretta, 2005). So stieg mit dem Aufkommen des Internets die Anzahl der Informationsquellen; für das Studium relevante Informationen konnten nun nicht mehr ausschließlich durch Nutzung der Bibliotheksressourcen gefunden werden. Daher erfüllte die sogenannte „objektorientierte“ Benutzerschulung (vgl. Homann, 2001, August), in der meist ein System der Bibliothek, z.B. der OPAC, vorgestellt wurde, nicht mehr die Anforderungen (Andretta, 2005). Durch das Aufkommen des Internets standen zudem nicht nur Informationen aus Bibliotheksressourcen, sondern auch teilweise ungesicherte und widersprüchliche Informationen aus dem Web zur Verfügung. Daher mussten Inhalte zum kritischen Denken und zur Bewertung von Informationen aufgenommen werden.

2.2 Situation in Deutschland

Auch in Deutschland fanden bereits in den 1970er Jahren Benutzerschulungen durch Bibliothekare statt. Nachdem das Kursangebot in den 1980er Jahren aus verschiedenen Gründen reduziert worden war, stieg es in den 1990er Jahren wieder an. Hauptgründe waren die Einführung elektronischer Systeme wie OPAC und Datenbanken auf CD-ROM. Wie bei den amerikanischen Pendanten handelte es sich auch hier um objektorientierte Schulungen, in denen ein einzelnes System vorgestellt wurde. Diese Angebote mussten überarbeitet werden, als die Zahl der Informationssysteme anstieg und einzelne Schulungen für jedes System angesichts der neuen Situation unangemessen erschienen. Den neueren Schulungen in Deutschland liegen US-amerikanische Konzepte zur Schulung von Informationskompetenz zugrunde. Breiter gefasste Inhalte – verglichen mit denen der vorher durchgeführten Benutzerschulungen – wurden zusammen mit den neuen Schulungskonzepten übernommen (für eine Übersicht über die Geschichte der Informationskompetenzschulungen in Deutschland siehe Homann, 2003). Dabei wurde in deutschen Fachpublikationen bis in die Mitte der 1990er Jahre die Bezeichnung information literacy verwendet. Die Übersetzung Informationskompetenz wurde zuerst in einigen Fachhochschul-Diplomarbeiten verwendet, bevor sie 1999 von

einer Expertengruppe offiziell eingeführt wurde (Ingold, 2005, p. 31). Damit verzichtet man auf eine wörtliche Übersetzung des Wortes literacy, die Lese- und Schreibfähigkeit lauten müsste, da dies irreführend sein könnte. Auch in anderen europäischen Sprachen wird literacy mit Kompetenz übersetzt, so im Dänischen, Schwedischen und Spanischen (vgl. Ingold, 2005, p. 29). Im Englischen hingegen hat die Bezeichnung literacy ein hinreichend großes Bedeutungsspektrum, sodass eine Anwendung auf Informationen als zulässig angesehen werden kann. Bawden (2001, pp. 221–222) zeigt auf, dass es im Englischen drei Definitionen von literacy gibt. Zum einen ist dies die Lese- und Schreibfähigkeit, zum anderen kann literacy auch darüber hinaus gehende Fähigkeiten bezeichnen, nämlich die Fähigkeit, das Gelesene zu verstehen. Die dritte Definition von literacy ist “being well educated, well-read, versed in the literature [...]” (Clifford, 1984 as cited in Bawden, 2001, p. 221). Die Bezeichnung Kompetenz für die mit information literacy bezeichneten Fähigkeiten hingegen scheint relativ passend, da hiermit ein Cluster an Fähigkeiten bezeichnet wird, das für die Lösung bestimmter Probleme genutzt werden kann; in diesem Fall wären das Probleme bei der Informationssuche (vgl. Abschnitt 9, S. 35).

2.3 Bildungspolitischer Kontext

Die verstärkte Förderung von Informationskompetenz an deutschen Universitäten steht in einem größeren bildungspolitischen Kontext. Zum einen geschieht die Förderung fachlicher Informationskompetenz unter der Annahme, dass Informationskompetenz für den Erfolg im Studium notwendig ist (HRK, 2012; Johnston & Webber, 2003). So konnte Bowles-Terry (2012) zeigen, dass US-amerikanische College-Absolventen, welche nach dem ersten Studienjahr an einer Informationskompetenzschulung teilgenommen hatten, höhere Durchschnittsnoten erzielten als Studierende, die bereits im ersten Studienjahr oder noch nie an einer solchen Schulung teilgenommen hatten. Eine andere Studie zeigte, dass die jeweilige Universitätsbibliothek nutzende Studierende nicht nur höhere Durchschnittsnoten erzielten, sondern auch mit geringerer Wahrscheinlichkeit ihr Studium abbrechen (Soria, Fransen, & Nackerud, 2014). Auch wird Informationskompetenz mit weiteren Anforderungen wie der Bewältigung der Informationsflut oder lebenslangem Lernen in Verbindung gebracht (Ingold, 2005, pp. 38–47). So schreibt Bundy (2001 as cited in Ingold, 2005, p. 38) “information literacy is required because of proliferating information access and resources”. Es wird argumentiert, dass die Fähigkeit, mit Informationen umzugehen, notwendig ist, um alltägliche Prob-

leme zu lösen und Entscheidungen zu treffen. Als Beispiele werden Wahlentscheidungen oder die Entscheidung darüber, ob man ein Kind bekommen möchte, genannt (Hancock, 1993). Unter lebenslangem Lernen wird „alles Lernen während des gesamten Lebens, das der Verbesserung von Wissen, Qualifikationen und Kompetenzen dient“ (Kommission der europäischen Gemeinschaften, 2001, p. 9) verstanden. Informationskompetenz soll Personen darauf vorbereiten, da informationskompetente Personen gelernt haben, wie sie Informationen finden und verwenden können (ACRL, 1989).

3 Modelle von Informationskompetenz

Bei den meisten Modellen handelt es sich um personenbezogene Prozessmodelle, die die Suche nach Informationen aus Sicht des Suchenden abbilden. Ausnahmen sind das kompetenzstufenorientierte Modell "Seven Pillars Model for Information Literacy" der britischen "Society of College, National and University Libraries" (SCONUL, 2011) sowie das konzeptuelle Modell "Seven Faces of Information Literacy" (Bruce, n.d., 1997). Der Vollständigkeit wegen werden zunächst diese beiden Modelle kurz beschrieben. Dann wird auf die aus den USA kommenden prozessbezogenen Modelle "Information Search Process", "The Big Six Skills" sowie das deutsche „Dynamische Modell der Informationskompetenz“ („DYMIK“) eingegangen, da sie aufeinander aufbauen. Für eine genauere Beschreibung und einen Vergleich der Modelle siehe Balceris (2011) oder Ingold (2005).

3.1 Seven Pillars for Information Literacy

Den Kern dieses 1999 entwickelten und 2011 überarbeiteten Modells (SCONUL, 1999, 2011) bilden sieben Fähigkeiten (sogenannte Säulen), die nach Ansicht der SCONUL Teile von Informationskompetenz sind. Hierbei handelt es sich um die folgenden Fähigkeiten:

1. the ability to recognise a need for information;
2. the ability to distinguish ways in which the information 'gap' may be addressed;
3. the ability to construct strategies for locating information;
4. the ability to locate and access information;
5. the ability to compare and evaluate information obtained from different sources;
6. the ability to organise, apply and communicate information to others in ways appropriate to the situation;
7. the ability to synthesise and build upon existing information, contributing to the creation of new knowledge.

(vgl. SCONUL, 1999)

Es wird angenommen, dass sich eine Person – je nach Säule bzw. Fähigkeit – auf verschiedenen Kompetenzniveaus befinden kann. Dabei ist vorgesehen, dass Personen erst ab Erreichen eines gewissen Fähigkeitsgrades als informationskompetent bezeichnet werden (vgl. Balceris, 2011, pp. 71–73). Zudem wird angenommen, dass Personen mit einem niedrigen

Kompetenzniveau lediglich auf den ersten vier Säulen Fähigkeiten aufweisen. Erst bei Erreichen höherer Kompetenzstufen kommen Fähigkeiten auf den weiteren drei Säulen hinzu. Leider liegen für dieses Modell keine empirischen Belege vor (Balceris, 2011, p. 73).

3.2 Seven Faces of Information Literacy

Dieses Modell wurde auf der Basis einer phänomenographischen Untersuchung, an der Berufstätige teilnahmen, von Christine Bruce entwickelt (Bruce, n.d., 1997). Dementsprechend bildet es die Sicht von Nutzern ab – und nicht die normative Sicht von Experten. In dem Modell wurden sieben Konzeptionen von Informationskompetenz identifiziert:

- Informationstechnologie (Nutzung von Informationstechnologie, um auf Informationen zuzugreifen);
- Informationsquellen (Kenntnis und Nutzung von Informationsquellen);
- Informationsprozess (Ausführung eines Prozesses, der mit einer neuen Situation, für die ein Informationsbedarf besteht, beginnt);
- Informationskontrolle (Kontrolle über Information, bspw. durch kompetente Speicherung und Aufbewahrung);
- Wissenskonstruktion (Aufbau einer persönlichen Wissensgrundlage);
- Wissenserweiterung (Gewinnung von neuen Erkenntnissen);
- Weisheit (weise Verwendung von Informationen zum eigenen Vorteil oder zum Vorteil anderer, bspw. Klienten).

(Übersetzung nach Ingold, 2005)

Es fällt auf, dass Informationskompetenz in diesem Modell breiter aufgefasst wird als bei anderen Definitionen. So wird auch der Aspekt des Problemlösens behandelt (Balceris, 2011).

3.3 Prozessmodelle von Informationskompetenz

Wie bereits angedeutet, handelt es sich bei den meisten Modellen von Informationskompetenz um Prozessmodelle, die den Informationsprozess strukturieren. Im Folgenden werden drei Modelle vorgestellt.

3.3.1 Information Search Process (ISP)

Das ISP-Modell wurde von Carol Kuhlthau entwickelt, um den Suchprozess aus Sicht des Nutzers abzubilden und dabei auch seine Gefühle und Motivation zu berücksichtigen (Kuhlthau, 1991, 1993). Das Modell entstand im Rahmen einer Studie mit 26 Schülern einer weiterführenden Schule, die zwei Hausarbeiten anfertigen mussten. Während der Recherche für diese Hausarbeiten dokumentierten die Schüler die Schritte ihrer Recherche. Weiterhin füllten sie einen Fragebogen bezüglich ihrer Wahrnehmung der Bibliotheksangebote aus. Mit sechs der 26 Probanden wurden qualitative Interviews geführt, welche inhaltsanalytisch ausgewertet wurden. Diese Daten waren die Grundlage für ein Modell mit sechs Phasen, welches mit den zugrunde liegenden Theorien von Kelly, Belkin sowie Taylor in Einklang steht.

Von Kellys "personal construct theory" geht die Annahme in das Modell ein, dass sich jede Person ein individuelles Modell der Wirklichkeit erschafft (Kelly, 1955/1986). Dabei entsteht ein System persönlicher Konstrukte – beispielsweise mit Annahmen über andere Personen und deren Verhaltenstendenzen in bestimmten Situationen. Diese Konstrukte helfen, Ereignisse zu antizipieren und darauf zu reagieren. Weiterhin bestimmen die persönlichen Konstrukte auch, in welcher Weise Ereignisse interpretiert werden (Amelang, Bartussek, Stemmler, & Hagemann, 2006, p. 445). Darüber hinaus nimmt Kelly an, dass der Konstruktionsprozess in Phasen verläuft. Kann eine neue Information nicht in das Konstruktesystem aufgenommen werden, entstehen Unsicherheit und Zweifel. Wird die neue Information dennoch nicht verworfen, formuliert die Person eine Hypothese, um die Information in das persönliche System integrieren zu können (Maher, 1969 as cited in Kuhlthau, 1991, p. 21). Das Erstellen einer Hypothese, anstatt die Information zu verwerfen, ist also der entscheidende Schritt im Konstruieren und Umkonstruieren von persönlichen Systemen.

Belkins "anomalous state of knowledge" (Belkin, Oddy, & Brooks, 1982) postuliert, dass die Informationssuche mit einem Problem des Benutzers, in dem Fall einem Mangel an Informationen zu einem Thema, beginnt. Anfangs kann der Benutzer sein Informationsproblem nur vage umreißen. Im Laufe des Prozesses der Informationssuche kann der Benutzer das Problem genauer beschreiben und entsprechend seinen Informationsbedarf genauer spezifizieren (vgl. Kuhlthau, 1991).

Taylor (1968) spezifizierte vier Arten, oder Niveaus, von Informationsbedürfnissen:

1. "visceral": Ein unbewusst vorhandener, aber nicht formulierter Informationsbedarf;
2. "conscious": Ein bewusster, aber noch nicht formulierter Informationsbedarf;
3. "formalized": Eine formale Angabe des Informationsbedarfs;
4. "compromised": Informationsbedarf in einer Form, in der er in ein Informationssystem eingegeben werden kann.

Auch Taylor stellt fest, dass Suchende zu Beginn des Suchprozesses häufig Fragen formulieren, die an ihr Vorwissen anknüpfen, während gegen Ende des Suchprozesses gezielt nach Informationen gesucht werden kann, die eine Wissenslücke füllen (vgl. Kuhlthau, 1991).

Daraus entstanden im ISP-Modell sechs Phasen, in denen jeweils spezifische Gedanken, Gefühle und Handlungen vorherrschen. In der ersten Phase ("initiation") wird der Person der Informationsbedarf bewusst. Die wichtigste Handlung besteht darin, Suchstrategien zu diskutieren. Die Gedanken beschäftigen sich überwiegend mit dem Problem; bei den Gefühlen überwiegen Unsicherheit und Ängstlichkeit. In der zweiten Phase ("selection") wird das Thema, zu dem recherchiert wird, identifiziert. Gefühle von Unsicherheit werden von Optimismus verdrängt, weil ein Anfang gemacht und das Thema identifiziert wurde. Gedanken kreisen um die Wahl und Spezifizierung des Themas. Während der dritten Phase ("exploration") beschäftigt sich die Person genauer mit dem breiteren Themenkomplex, in dem das eigentliche Thema liegt. Während dieser Phase werden allgemeine Informationen zu dem Themenkomplex gesucht, dabei dominieren Gefühle von Verwirrung und Unsicherheit. Viele denken in dieser Phase darüber nach, die Suche abubrechen. Die vierte Phase ("formulation") stellt den Wendepunkt des Suchprozesses dar. Das Thema kann fokussiert werden und ein Gefühl von Zuversicht entsteht. Während der fünften Phase ("gathering") wird Information zu dem nun definierten Thema gesammelt. Gedanken kreisen vor allem um die Beschaffung von Informationen und die Abgrenzung des Themas. Das Gefühl von Zuversicht verstärkt sich. Die sechste Phase ("presentation") wird von Gefühlen der Erleichterung geprägt. Die Information wird gesammelt und organisiert. Hierbei kann die Verwendung der Informationen, bspw. für eine Hausarbeit, vorbereitet werden. In dem Modell ist vorgesehen, dass eine Phase mehrfach durchlaufen oder zu einer Phase nach Evaluation der Ergebnisse zurückgekehrt werden kann.

Zur Validierung des Modells wurden mehrere Studien durchgeführt. In einer Studie mit Schülern einer weiterführenden Schule konnte anhand während der Recherche durchgeführter

Umfragen gezeigt werden, dass die Gedanken der Schüler zu dem Rechercheproblem im Laufe des Suchprozesses klarer wurden. Zudem wurden gegen Ende des Prozesses spezifischere Informationen gesucht. Hierbei musste ein Fragebogen drei Mal (zu Beginn, in der Mitte und zum Ende der Recherche) ausgefüllt werden, sodass der Verlauf beobachtet werden konnte. In einer größeren Studie mit Nutzern von akademischen, öffentlichen und schulischen Bibliotheken wurde derselbe Fragebogen wieder drei Mal während des Rechercheprozesses eingesetzt. Auch hier zeigte sich das zu erwartende Muster von Informationssuche und begleitenden Gedanken, sodass man annehmen kann, dass das Modell nicht nur bei Studierenden, die fachliche Recherchen durchführen, Gültigkeit hat (Kuhlthau, 1991).

Im Rückblick können Kuhlthau vor allem zwei wichtige Verdienste zugesprochen werden. Zum einen legte sie ein empirisch validiertes Modell der Informationssuche vor. Ihr Modell ist umfassender als die anderer Autoren – insbesondere sticht es durch die Einbeziehung von emotionalen Aspekten bei der Informationssuche hervor (Balceris, 2011; Homann, 2000). Zum zweiten machten ihre Arbeiten Bibliothekaren bewusst, dass traditionelle bibliothekarische Dienstleistungen, wie Schulungen oder persönliche Hilfe an der Auskunft, oft zu kurz greifen, weil sie letztendlich nur Schritt 5 behandeln – nämlich den eigentlichen Suchvorgang. Kuhlthau machte deutlich, dass der Rechercheprozess nicht erst mit der Eingabe von Suchbegriffen in entsprechende Abruf-Systeme beginnt, sondern mit der Spezifizierung des Themas (Isbell & Kammerlocher, 1998).

3.3.2 The Big Six Skills

Auch dieses Modell, entwickelt von Mike Eisenberg und Bob Berkowitz, beschreibt den Informationsprozess aus Sicht des Benutzers. Das Modell gliedert den Prozess in sechs aufeinanderfolgende Schritte, welche unterschiedliche Fähigkeiten erfordern (Eisenberg & Berkowitz, 2014; Lowe & Eisenberg, 2006). Auffällig ist, dass die drei Phasen "selection", "exploration" sowie "focus" des ISP-Modells in diesem Modell in zwei Phasen ("information seeking strategies" und "location and access") zusammengefasst sind. Ziel bei der Erstellung dieses Modells war es, die Lehrveranstaltungen für Informationskompetenz mithilfe des Modells besser gestalten und strukturieren zu können. Dies geschieht, indem für jede Phase zwei Handlungsanweisungen vorgegeben sind; entsprechend können diese Handlungen geübt, bzw. die dafür relevanten Fähigkeiten vermittelt werden. Auch bei diesem Modell ist

vorgesehen, dass der Prozess nicht immer in dieser linearen Reihenfolge verläuft; das Überspringen oder Wiederholen einzelner Schritte ist prinzipiell möglich. Die Autoren legen Wert darauf, dass dieses Modell auch die Fähigkeit kritischen Denkens impliziert (Lowe & Eisenberg, 2006, p. 66). Die sechs Phasen des Modells und die dazugehörigen Handlungsanweisungen sind wie folgt:

1. task definition
 - 1.1. define the problem
 - 1.2. identify the information needed;
2. information seeking strategies
 - 2.1. determine all possible sources
 - 2.2. select the best resource;
3. location and access
 - 3.1. locate sources
 - 3.2. find information within sources;
4. use of information
 - 4.1. engage (e.g. read, hear, view)
 - 4.2. extract relevant information;
5. synthesis
 - 5.1. organize the information from multiple sources
 - 5.2. present information;
6. evaluation
 - 6.1. judge the result
 - 6.2. judge the process.

3.3.3 Das dynamische Modell der Informationskompetenz (DYMIK)

Das dynamische Modell der Informationskompetenz wurde von Benno Homann auf Grundlage der beiden bereits skizzierten Modelle entwickelt (Homann, 2000). Das Modell sieht fünf Phasen des Informationsprozesses vor. Auch hier wird der Informationsprozess aus Sicht des Subjekts beschrieben und ein dynamischer Ablauf angenommen, d.h. es ist vorgesehen, dass einzelne Phasen wiederholt werden oder dahin zurückgekehrt wird. Wie bei Kuhlthau sind auch in diesem Modell emotionale Aspekte integriert. Um eine Grundlage für

die Vermittlung von Informationskompetenz zu schaffen, wurden die in jeder Phase erforderlichen Fähigkeiten definiert. So beginnt der Informationsprozess mit einem subjektiven Problem. Damit wird die erste Phase des Modells („Info-Bedarf“) initiiert, während der Art und Umfang der erforderlichen Informationen identifiziert werden. Hierfür sind vor allem kognitive und sozial-kommunikative Fähigkeiten erforderlich. In der zweiten Phase („Info-Quellen“) werden potentielle Informationsquellen ermittelt. Hierfür ist Wissen über Informationstypen und Informationseinrichtungen erforderlich. In der dritten Phase („Info-Zugang“) werden die benötigten Informationen abgerufen, wofür Kenntnisse über die dafür verwendeten technischen Systeme benötigt werden. Während der vierten Phase („Info-Erfassung“) wird die Information inhaltlich erschlossen, wofür Fähigkeiten in der Textanalyse unerlässlich sind. In der fünften und letzten Phase („Info-Bewertung“) wird überprüft, ob der in Phase 1 formulierte Informationsbedarf mit den gefundenen Informationen gedeckt werden kann. Hierfür sind vermutlich wieder kognitive und analytische Fähigkeiten erforderlich, dies geht jedoch nicht aus der Publikation (Homann, 2000) hervor. Da hier für (fast) jede Phase die erforderlichen Fähigkeiten und Kenntnisse spezifiziert sind, eignet sich dieses Modell besonders gut für die Konzeption von Schulungen zur Informationskompetenz (Homann, 2000). Die Einteilung in fünf Phasen ermöglicht zudem, die Vermittlung der erforderlichen Kenntnisse und Fähigkeiten auf mehrere Institutionen aufzuteilen. So fällt an der Universität Heidelberg die Lehre zu Phasen 2 und 3 (Quellen und Zugang) in den Zuständigkeitsbereich der Universitätsbibliothek, während die Lehre zu den übrigen Phasen bei anderen Institutionen liegt (Homann, 2000). Es fällt auf, dass Phase 5 („synthesis“) des Big Six-Modells komplett entfällt, bzw. in der Phase Info-Bewertung enthalten ist. Dies kann aber auch ein Zugeständnis an die Realität an deutschen Universitäten sein. Die Organisation und Verwendung von Informationen werden dort nicht unbedingt als Bestandteil von Informationskompetenz gesehen, sondern als Inhalte wissenschaftlichen Arbeitens (vgl. Homann, 2007).

3.4 Zusammenfassung

Beim Vergleich der Modelle zeigt sich, dass die Beschreibung von Informationskompetenz als Prozess am besten theoretisch untermauert ist (Balceris, 2011). Auch der Ansatz der SCONUL kann, trotz der Einbeziehung von Kompetenzstufen, als Aufzählung von Schritten (also als Prozess) verstanden werden. Der Informationsprozess wird in allen Modellen ähnlich konzeptualisiert; so enthalten alle Modelle das Erkennen des Informationsbedarfs, den Zugriff

auf Informationen und die Bewertung derselben (siehe auch Boon, Johnston, & Webber, 2007). Es wird zudem deutlich, dass das Big Six-Modell und das DYMIK für die Strukturierung von Lehrveranstaltungen konzipiert wurden, während das ISP-Modell den Informationsprozess aus wissenschaftlicher Sicht beschreibt.

Bei der Betrachtung der Modelle fällt auf, dass vorwiegend kognitive und emotionale Aspekte behandelt werden. Aus motivationaler Sicht könnte man den Ablauf eines Suchprozesses möglicherweise anhand des Rubikon-Modells (Achtziger & Gollwitzer, 2006, pp. 277–281) darstellen. Während der prädeziotionalen Phase wird das Informationsproblem definiert; während der präaktionalen Phase werden dann die zu verwendenden Strategien ermittelt, welche während der aktionalen Phase ausgeführt werden. Während der postaktionalen Phase erfolgt die Bewertung der Recherche.

Allerdings strukturieren diese Modelle den Informationsprozess, und die dafür erforderlichen Fähigkeiten, nur grob, sodass diese Modelle nur bedingt für die Konzeptualisierung von Lehrveranstaltungen oder die Konstruktion von Messinstrumenten geeignet sind. Hierfür müssten erforderliche Fähigkeiten oder Kompetenzen genauer spezifiziert werden, als das beim DYMIK oder Big Six-Ansatz der Fall ist. Eine genauere Definition und Gliederung der Fähigkeiten wird von Standards der Informationskompetenz geliefert, welche im Folgenden beschrieben werden.

4 Standards der Informationskompetenz

Durch die Standards wird der Informationsprozess nicht nur beschrieben, sondern soweit in Teilschritte bzw. Teilkompetenzen zerlegt, dass diese als Grundlage für Lehrveranstaltungen oder Messinstrumente genutzt werden können. Hierfür enthalten die Standards konkrete Verhaltensindikatoren (Balceris, 2011, p. 102). Im Folgenden werden die Standards der amerikanischen "Association of College and Research Libraries" (ACRL) und des deutschen Bibliotheksverbands beschrieben. Die Standards der ACRL sind von besonderer Relevanz, da hier eine auf das Fach Psychologie bezogene Version vorliegt. Die Standards des deutschen Bibliotheksverbands wurden aufgenommen, um die deutsche Sicht auf das Konzept abzubilden.

4.1 Information Literacy Standards for Higher Education

Die Standards wurden im Jahr 2000 festgelegt, um die Messung von Informationskompetenz auf eine Grundlage stellen zu können (ACRL, 2000). Es wurden insgesamt fünf Standards definiert, welche informationskompetentes Verhalten definieren. Um die Erfassung von Informationskompetenz zu ermöglichen, wurden für jeden Standard mehrere Verhaltensindikatoren und Verhaltensbeispiele definiert. Die Standards legen fest, dass ein Student als informationskompetent gilt, wenn er Folgendes beherrscht:

- Er bestimmt Art und Umfang der benötigten Informationen;
- er verschafft sich effizienten und effektiven Zugang zu den benötigten Informationen;
- er evaluiert Informationen und seine Quellen kritisch und integriert die ausgewählten Informationen in sein Wissen und Wertsystem;
- er nutzt Informationen effektiv – als Individuum oder Gruppenmitglied –, um ein bestimmtes Ziel zu erreichen;
- er versteht viele der ökonomischen, rechtlichen und sozialen Streitfragen, die mit der Nutzung von Informationen zusammenhängen und nutzt die Informationen in einer ethischen und legalen Weise.

(Übersetzung vgl. Homann, 2002)

Ein Problem dieses Verständnisses von Informationskompetenz ist, dass diese Fähigkeiten sehr breit aufgefasst werden; Informationskompetenz wird hier als kontextunabhängige Kompetenz dargestellt, die fast in jeder Lebenslage von Nutzen ist und lebenslanges Lernen

fördern soll. Damit entsteht ein Konzept, das in der Praxis nicht umsetzbar oder operationalisierbar ist. Letztendlich kann die Vermittlung solcher Inhalte weder von Bibliothekaren noch von den übrigen angesprochenen Gruppen (v.a. dem Lehrkörper an Universitäten) geleistet werden (Ingold, 2005, p. 99; Owusu-Ansah, 2003).

Im angloamerikanischen Raum existieren zum jetzigen Zeitpunkt mindestens zwei kommerziell erhältliche Multiple-Choice-Messinstrumente, welche auf diesen Standards basieren. Zum einen der "SAILS Test" (Project SAILS, 2000-2014), zum anderen der "Information Literacy Test" (ILT, Swain, Sundre, & Clarke, 2014). Dem Testmanual des ILT ist jedoch zu entnehmen, dass mit diesem Test nur Fähigkeiten bezogen auf Standards 1,2,3, und 5 erfasst werden. Fähigkeiten bezogen auf Standard 4 sind nach Ansicht der Autoren nicht durch Multiple-Choice-Items erfassbar und sollten stattdessen durch Begutachtung von studentischen Ergebnissen (z.B. Hausarbeiten) oder durch Bewertung des Prozesses untersucht werden (Swain et al., 2014, p. 4).

Recht bald erkannte man, dass Informationskompetenz nicht nur fachübergreifend verwendbare Fähigkeiten im Abruf von Informationen, sondern auch fachspezifisches Wissen, beispielsweise über Publikationsmechanismen im jeweiligen Fach, umfasst. Darüber hinaus sind für die Bewertung von Informationen, beispielsweise einem Zeitschriftenartikel, auch fachliches Wissen notwendig (Grafstein, 2002). Daher wurden die Standards der ACRL an die Besonderheiten einzelner Fächer angepasst. So erschienen die "Psychology Information Literacy Standards" mit dem Ziel, Informationskompetenzschulungen für Psychologiestudierende konzipieren und evaluieren zu können (ACRL, 2010). Bei der Überarbeitung wurde Standard 5 in die übrigen Standards integriert. Die vier psychologiespezifischen Standards lauten:

1. The information literate psychology student determines the nature and extent of the information needed;
2. the information literate psychology student accesses needed information effectively and efficiently;
3. the information literate psychology student evaluates information and its sources critically and incorporates selected information into her or his knowledge base;
4. the information literate psychology student, individually or as a member of a group, uses information effectively to accomplish a specific purpose.

Ein Verhaltensindikator für Standard 2 lautet beispielsweise: “constructs and implements effectively-designed search strategies”. Mögliche Ergebnisse sind: “uses appropriate psychological terminology for searching databases, recognizing the different effects of using keywords, synonyms, and controlled vocabulary from the database”. Die Indikatoren definieren die Anforderungen an Psychologiestudierende relativ präzise, so werden als Informationsquellen die Fachdatenbank PsycINFO, sowie Google Scholar und PubMed genannt. Daher eignet sich diese Definition sehr gut, um Messinstrumente für psychologiespezifische Informationskompetenz zu erstellen. Ein weiterer Vorteil gegenüber der allgemeinen Definition der ACRL ist, dass diese Definition von Informationskompetenz inhaltlich weniger breit ist und – zumindest im Wesentlichen – nur solche Aspekte beinhaltet, die realistischerweise gelehrt werden können. So sind beispielsweise in der psychologiespezifischen Definition mit dem Begriff Information nur solche Informationen gemeint, auf welche man über Bibliotheks- oder Internetressourcen o.ä. zugreifen kann. In den allgemeinen Standards der ACRL werden (ironischerweise) auch die Ergebnisse empirischer Experimente als Information bezeichnet; das Durchführen solcher Experimente gilt dementsprechend als Methode der Informationsbeschaffung. Dies kann kaum Inhalt von Informationskompetenzschulungen durch Bibliothekare sein.

An dieser Stelle soll erwähnt werden, dass die Informationskompetenzstandards der ACRL zurzeit (Stand Ende 2014) überarbeitet werden. An die Stelle der fünf Standards treten sogenannte “threshold concepts”, worunter Kernbegriffe zu verstehen sind, ohne deren Verständnis ein Zurechtfinden in modernen Informationsumwelten nicht möglich ist. Zu jedem dieser “threshold concepts” sollen Lernziele formuliert werden (Hapke, 2014, March 6).

4.2 Standards der Informationskompetenz für Studierende

Die vom Deutschen Bibliotheksverband (DBV, 2009) herausgegebenen Standards ähneln auf den ersten Blick den allgemeinen Standards der ACRL. Auch hier werden fünf Standards und dazugehörige Verhaltensindikatoren definiert. Standard 1 behandelt das Erkennen des Informationsbedarfs; Standard 2 den Zugang zu Informationen. Standard 3 definiert Fähigkeiten in der Bewertung von Informationen. Standard 4 beschreibt die Verarbeitung und Präsentation der Informationen, während Standard 5 die Verantwortung bei der Nutzung von Informationen hervorhebt. Bei genauerer Betrachtung der Verhaltensindikatoren zeigt sich

jedoch, dass hier ein engeres, für die Lehre und Messung realistischeres, Verständnis von Informationskompetenz vorliegt, das dem Verständnis der psychologiespezifischen ACRL-Standards ähnelt und teilweise noch enger konzipiert ist. So beinhalten beide Definitionen der ACRL unter Standard 3 die Fähigkeit, festzustellen, ob die neuen Informationen mit dem individuellen Wertesystem vereinbar sind und ggf. die Fähigkeit, Differenzen zu überbrücken. Auch dies fällt, zumindest im deutschsprachigen Raum, primär in den Bereich wissenschaftlichen Arbeitens und muss (oder sollte) daher nicht Bestandteil von Informationskompetenzschulungen sein.

4.3 Exkurs: “Information Behavior”

An dieser Stelle sollte daher der Begriff Informationsverhalten (“information behavior”) eingeführt werden. Damit werden sämtliche Handlungen bezeichnet, die Personen in Bezug auf Informationsquellen vornehmen. Hierunter fallen aktive und passive Informationssuche, also neben der Recherche in Informationssystemen auch Kommunikation mit anderen Personen und passive Rezeption von Fernsehprogrammen etc. Verwandte Begriffe sind “information-seeking behavior” und “information search behavior”. Information-seeking behavior bezeichnet das zielgerichtete Suchen nach Informationen aufgrund eines Informationsbedarfs. Information search behavior bezeichnet das Verhalten während der Interaktion mit Informationssystemen aller Art (Wilson, 2000). Wilson (1999) schlägt vor, eine geschachtelte Beziehung der Begriffe anzunehmen: Information-seeking behavior wäre dann ein Untergriff von Information behavior. Information search behavior wäre entsprechend ein Unterbegriff von Information-seeking behavior.

Unter Rückgriff auf diese Definitionen kann man also sagen, dass durch die Informationskompetenz-Standards das angemessene oder gewünschte Informationsverhalten definiert wird. Ginge es nur um Standards 2 und 3, könnte man auch sagen, dass so das gewünschte Information search behavior definiert wird; die Gesamtheit der Informationskompetenz-Standards beinhaltet allerdings auch Fähigkeiten bezüglich der Bewertung und Verwendung von Informationen. Angemessenes Informationsverhalten kann jedoch nicht mit Informationskompetenz gleichgesetzt werden, da Informationskompetenz darüber hinaus Hintergrundwissen enthält, das sich nicht unbedingt in Verhalten manifestiert. Angemessenes In-

formationsverhalten kann als ein Aspekt von Informationskompetenz angesehen werden (Timmers & Glas, 2010).

4.4 Zusammenfassung

Es lässt sich zusammenfassen, dass die Standards der Informationskompetenz für die Erstellung von Messinstrumenten oder die Gestaltung von Trainingsprogrammen wesentlich besser geeignet sind als die Modelle. Dies liegt in erster Linie an der Festlegung von Verhaltensindikatoren, welche das gewünschte Informationsverhalten und die dafür notwendigen Kenntnisse relativ genau beschreiben. Dabei stechen die psychologiespezifischen Standards der ACRL hervor, welche für Psychologiestudierende relevante Fähigkeiten und Wissensbestände recht genau definieren. Daher basieren die in dieser Arbeit vorgestellten Ansätze zur Förderung und Erfassung von Informationskompetenz in erster Linie auf diesen psychologiespezifischen Standards.

5 Förderung von Informationskompetenz

Wie oben erwähnt, sind Überlegungen zur Förderung von Informationskompetenz ein wichtiger Bestandteil der Informationskompetenz-Literatur. Der Großteil der Literatur behandelt Schulungen im akademischen Kontext, die üblicherweise von Universitätsbibliotheken angeboten werden. Als theoretische Grundlage für solche Schulungen können die oben beschriebenen Informationskompetenzstandards oder Modelle verwendet werden.

5.1 Inhalte

Eine wichtige Herausforderung bei der Gestaltung von Informationskompetenz-Lehre ist die Festlegung der Inhalte. Wie an den oben zusammengefassten Standards und Modellen deutlich wird, ist Informationskompetenz ein Konstrukt, das viele unterschiedliche Fähigkeiten beinhaltet. Diese Komponenten sind in unterschiedlichem Ausmaß durch Lehrveranstaltungen vermittelbar (Peter, 2015). Eine zentrale Komponente ist sicherlich eher technisches Wissen über in der jeweiligen Disziplin relevante Fachdatenbanken und deren Funktionen, welches sich relativ einfach definieren und lehren lässt. Weiterhin umfassen gängige Definitionen von Informationskompetenz aber auch die Fähigkeit zur kontextabhängigen Bewertung von Information, welche neben Fach- und Methodenwissen auch die Fähigkeit zum kritischen Denken beinhaltet (Peter, 2015). Dieses Fach- und Methodenwissen kann üblicherweise nicht im Rahmen von bibliothekarischen Schulungen vermittelt werden, sondern wird im Laufe des Studiums erworben. Die Lehre von kritischem Denken ist zwar prinzipiell möglich, jedoch aufwändig (van Gelder, 2005) – der Erfolg von Interventionen stark von deren Konzeption abhängig (Abrami et al., 2008). Auch Grassian und Kaplowitz (2001, pp. 111–115) ist diese Schwierigkeit bewusst. Sie argumentieren, dass Bibliothekare die Grundlagen kritischen Denkens lehren sollten, worauf im Rahmen anderer Lehrveranstaltungen aufgebaut werden kann. Unter den Grundlagen kritischen Denkens wird beispielsweise verstanden, dass man Studierende sensibilisiert, bei Informationen auf den Zweck der Veröffentlichung, den Hintergrund des Autors, eventuelle finanzielle Hintergründe der Veröffentlichung (z.B. Auftragsforschung) und das Alter der Veröffentlichung zu achten.

In der Praxis scheint dennoch vorwiegend das Erkennen eines Informationsbedarfs und das Recherchieren und Beschaffen von Informationen gelehrt zu werden. Dabei stehen meist die

Angebote der Bibliothek, wie zum Beispiel der Bibliothekskatalog, im Vordergrund (Ingold, 2005, p. 67). Teilweise enthalten solche Schulungen auch grundlegende IT-Fertigkeiten, wie zum Beispiel die Nutzung von Textverarbeitungsprogrammen (z.B. Andretta, 2005, pp. 68–69; Donnelly, 2000). Eine Umfrage unter Praktikern, in diesem Fall Bibliothekaren aus Kanada und Neuseeland, zeigt, dass vor allem das Auffinden von Informationen aus verschiedenen Quellen als Bestandteil von bibliothekarischen Schulungen angesehen wird. Die kritische Bewertung von Informationen ist nach Ansicht der befragten Bibliothekare ebenfalls ein Element von Informationskompetenz, allerdings sehen diese es nur teilweise als ihre Aufgabe an, diese Fähigkeit zu lehren (Julien, 2000).

Weitere Herausforderungen bei der Gestaltung der Inhalte sind der Einbezug von disziplinspezifischen und ortsspezifischen Inhalten (Krüger, 2008). Zu den nicht disziplinspezifischen Inhalten gehören beispielweise die (allgemeine) Funktionsweise von Fachdatenbanken, die Verwendung von Booleschen Operatoren oder die Verwendung von Google Scholar. Daneben gibt es aber auch disziplinspezifische Inhalte: Beispielsweise das Wissen über für das jeweilige Fach relevante Fachdatenbanken und deren Funktionen (Krüger, 2008). So unterscheiden sich Fachdatenbanken in ihrem Funktionsumfang und der Funktionsweise. Da vielfach sogenannte "hands-on-instruction", bei der bestimmte Vorgehensweisen nicht erst theoretisch, sondern von Anfang an mithilfe der verwendeten Recherchesysteme vermittelt werden (vgl. Bren, Hillemann, & Topp, 1998), angeboten wird, müssen in diesen Schulungen letztendlich die Datenbanken behandelt werden, die im späteren Studium auch genutzt werden. Auch bezüglich der Inhalte zum Erkennen eines Informationsbedarfs oder zum Bewerten von Informationen ist es vorteilhaft, wenn die Teilnehmer bei ihrer Fragestellung oder ihrem Fachgebiet „abgeholt“ werden können (Krüger, 2008). Zudem sollten ortsspezifische Inhalte behandelt werden. Hierbei handelt es sich nicht nur um Informationen über den lokal verwendeten Bibliothekskatalog oder die Systematik der Aufstellung der Bücher in der jeweiligen Universitätsbibliothek. Unterschiede gibt es auch beim Zugriff auf von der Universitätsbibliothek lizenzierte Fachdatenbanken oder der Nutzung sogenannter "Linkresolver" zum Zugriff auf Volltextartikel (siehe auch Peter, 2015).

5.2 Lehrveranstaltungen

Was die Form der Lehrveranstaltungen angeht, scheinen einstündige Schulungen zu dominieren – das gilt sowohl für den deutschsprachigen Raum (Homann, 2015), als auch für den angloamerikanischen Bereich (Mery, Newby, & Peng, 2012). Als Gründe hierfür werden vor allem mangelnde Ressourcen genannt, welche umfassenderen Lehrveranstaltungen entgegenstehen (Homann, 2003; Mery et al., 2012). Dies ist nicht unproblematisch, da sich in einer solchen einstündigen Schulung Informationskompetenz nicht umfassend lehren lässt. Meist lässt sich in dieser Zeit nur auf eine Informationsquelle (z.B. Fachdatenbank, Bibliothekskatalog) eingehen (Homann, 2015) und es ist fraglich, ob praxisrelevanter Lerngewinn erreicht werden kann (Grassian & Kaplowitz, 2001; Mery et al., 2012).

Ein weiteres Problem bei der Gestaltung entsprechender Lehrveranstaltungen ist die Integration in die Curricula der jeweiligen Studiengänge. So müssen die meisten im deutschsprachigen Raum angebotenen Schulungen auf freiwilliger Basis besucht werden (Homann, 2015). Im angloamerikanischen Bereich scheint zwar die Bereitschaft, Informationskompetenz-Schulungen in die Curricula einzubetten, größer zu sein, dennoch dominieren knappe, einstündige Schulungen, die jedoch Bestandteil eines fachlichen Seminars sind (Fister & Eland, 2008).

Grundsätzlich lassen sich Schulungen für Schlüsselqualifikationen, zu denen auch Informationskompetenz gezählt werden kann, in additive und integrative Ansätze einteilen (Heierle, 2006, siehe auch Peter, 2015). Bei additiven Ansätzen erfolgt die Vermittlung der Schlüsselqualifikationen unabhängig von der Vermittlung der Fachinhalte. Bei integrativen Ansätzen wird der Erwerb der Schlüsselqualifikationen mit dem Erwerb von Fachwissen kombiniert.

Fister und Eland (2008) beschäftigen sich umfassend mit den beiden Konzepten. Für das integrative Modell sprechen ihrer Meinung nach vor allem drei Gründe. Auf einer praktischen Ebene wird betont, dass es schwierig ist, einen separaten Kurs für Informationskompetenz im Curriculum zu verankern. Hierfür müssen zuerst die Lehrenden des jeweiligen Fachs und anschließend die Studierenden von der Notwendigkeit eines solchen Kurses überzeugt werden. Eine Kooperation mit Lehrenden ist dagegen wesentlich leichter zu initiieren. Auf konzeptueller Ebene wird betont, dass bei integrativen Ansätzen eine kontextbezogene Lehre von Informationskompetenz möglich ist. So ist es bei integrativen Ansätzen für die Studie-

renden viel leichter zu erkennen, welchen Nutzen die Informationskompetenz-Inhalte für ihr Studium haben können. Zudem kann somit fachspezifischen Anforderungen an Informationskompetenz-Lehre viel besser nachgekommen werden. Dabei identifizieren die Autoren mehrere Ansätze zur Integration. Im einfachsten Fall findet ein Einführungsseminar als Kooperationsveranstaltung zwischen Dozent und Bibliothekar statt. Dies gibt dem Bibliothekar die Möglichkeit, grundlegende Inhalte zur Informationskompetenz zu vermitteln. Andere Ansätze sehen Schulungen durch Bibliothekare im Laufe eines Seminars auf Nachfrage ("on demand") oder ein auf mehrere Semester angelegtes Schulungsprogramm vor. Dabei sind in mehrere aufeinander folgende Seminare jeweils Inhalte zur Informationskompetenz integriert. Dennoch, so wird argumentiert, ist bei integrativen Ansätzen eine konsistente und kohärente Informationskompetenz-Lehre nicht möglich. So behielten die Lehrenden des jeweiligen Fachs bei integrativen Ansätzen weiterhin die Kontrolle über den Lehrplan des Seminars, was einer umfassenden Informationskompetenz-Vermittlung entgegensteht. Eine umfassende Lehre von Informationskompetenz ist nach Ansicht der Autoren nur durch additive Lehrveranstaltungen möglich. Dabei wird betont, dass additive Lehrveranstaltungen Pflichtprogramm des Curriculums sein sollten. Beim Lesen fällt auf, dass hier die Situation in den USA beschrieben wird. In Deutschland sind Kooperationen zwischen Bibliothekaren und Lehrenden selten, stattdessen werden meist extracurriculare Informationskompetenz-Schulungen angeboten (vgl. Homann, 2015). Weisel (2015) sieht dafür vor allem die fehlende Wahrnehmung der Bedeutung von Informationskompetenz auf Seiten von Lehrenden und Studierenden als Ursache. Zudem bemängelt Weisel (2015, p. 64) eine fehlende „Anpassung der Standards und Konzepte an den Bedarf der Studierenden, Lehrenden und Forschenden sowie an die Geschäftsprozesse in Lehre und Forschung“. Dieses Defizit wurde von den Verantwortlichen erkannt; in den vergangenen fünf Jahren erschienen etliche Stellungnahmen und Empfehlungen, um diese Situation zu verbessern (z.B. HRK, 2012; für einen Überblick siehe Weisel, 2015). Mittlerweile existieren auch in Deutschland mehrere in das Curriculum integrierte Lehrveranstaltungen für Informationskompetenz. Beispielsweise wurde in Freiburg im Breisgau ein Blockseminar zu Informationsrecherchen entwickelt, das Teil eines universitätsweiten Programms zur Förderung beruflich relevanter Qualifikationen ist. Durch die Teilnahme können Kreditpunkte nach dem ECTS-System erworben werden (Ohloff, 2012).

5.3 Online-Lehre

Wie oben erwähnt scheinen mangelnde personelle Ressourcen ein wesentliches Problem bei der Lehre von Informationskompetenz zu sein (Julien, 2000; Mery et al., 2012). Als mögliche Lösung wird die Konzeption von Online-Tutorials vorgeschlagen, durch deren Einsatz es möglich ist, Informationskompetenz-Schulungen bei hohen Teilnehmerzahlen ressourcensparend anzubieten (Mery et al., 2012). Außerdem können mithilfe solcher Tutorials wesentlich mehr Informationen vermittelt werden, als in zeitlich begrenzten Schulungen. So liegt vielen der von Rauchmann (2002) betrachteten Online-Tutorials ein relativ breites Verständnis von Informationskompetenz zugrunde, welches über die klassischen Inhalte zur Informationssuche hinaus geht und beispielsweise auch die kritische Bewertung von Informationen beinhaltet. Weiterhin bieten solche Tutorials mehr Möglichkeiten zur Reflektion der Materialien; eine empirische Studie konnte zeigen, dass der Lerngewinn durch ein solches Online-Tutorial zur Informationskompetenz signifikant höher war als der Lerngewinn bei Teilnahme an einer einstündigen Schulung (Mery et al., 2012). Ein weiterer Vorteil von Online-Lehre ist sicherlich die zeitliche Flexibilität für die Teilnehmenden (Romero & Barberà, 2011). Allerdings gibt es etliche Studien, die über hohe Abbruchquoten bei Online-Lehre berichten (vgl. Levy, 2007). Eine mögliche Ursache hierfür sind mangelnde soziale Interaktionen bei Online-Lehre. Als Alternative zu reiner Online-Lehre hat sich Blended Learning, also die Kombination von Online- und Präsenzlehre (Garrison & Kanuka, 2004; Kerres, 2012) etabliert. Blended Learning kann die zeitliche Flexibilität für die Teilnehmenden und den geringen Ressourcenbedarf von Online-Lehre mit sozialen Interaktionen und entsprechend relativ niedrigen Abbruchquoten, welche man üblicherweise nur bei Präsenzlehre findet, kombinieren (López-Pérez, Pérez-López, & Rodríguez-Ariza, 2011).

Mittlerweile liegen meta-analytische Belege für die Wirksamkeit von Blended Learning vor. So zeigte eine Meta-Analyse von Bernard, Borokhovski, Schmid, Tamim und Abrami (2014), dass Blended Learning bezüglich des Lerngewinns effektiver war als reine Präsenzlehre. Eine andere Meta-Analyse verglich den Lernerfolg bei reiner Online-Lehre, Blended Learning und Präsenzlehre (Means, Toyama, Murphy, & Baki, 2013). Die Ergebnisse zeigen, dass Blended Learning effektiver war als reine Präsenzlehre oder reine Online-Lehre, welche sich nicht in ihrer Effektivität unterschieden. Dabei führen die Autoren den Vorteil von Blended Learning gegenüber reiner Online-Lehre auf die darin enthaltenen Möglichkeiten zur sozialen Interak-

tion zurück. Zwar sprechen bisher veröffentlichte Studien, die den Effekt von Blended Learning auf den Lerngewinn bei Informationskompetenz-Schulungen untersuchten und mit dem Lerngewinn bei Online- oder Präsenz-Schulungen verglichen (Anderson & May, 2010; Kraemer, Lombardo, & Lepkowski, 2007; Walton & Hepworth, 2013) nicht für die Überlegenheit von Blended Learning. Diese Studien beruhen jedoch auf nicht validierten Messinstrumenten und weisen teilweise weitere methodische Mängel (bspw. geringe Stichprobengrößen) auf (Peter, 2015).

5.4 Exkurs: Fachliche Recherchen

Ziel der meisten Bemühungen zur Förderung von Informationskompetenz bei Studierenden ist, diese zu befähigen, selbstständig effektiv nach Fachinformationen zu suchen. An dieser Stelle sollen daher die wichtigsten Informationssysteme bzw. Suchwerkzeuge im Fach Psychologie und deren Verwendung dargestellt werden. Damit sollen nicht nur relevante Fachbegriffe definiert werden, sondern auch ein besseres Verständnis der in Kapitel 12 dargestellten Rechercheaufgaben ermöglicht werden.

Die wichtigsten Informationssysteme oder Suchwerkzeuge für psychologische Inhalte sind Fachdatenbanken, wobei insbesondere PsycINFO und PSYINDEX von Relevanz sind (Lauber-Reymann, 2010, p. 338). Beide Fachdatenbanken beinhalten überwiegend Zeitschriftenartikel (z.B. ca. 80% von PsycINFO, vgl. American Psychological Association (APA), 2014), es sind aber auch andere Literaturgattungen (z.B. Monographien, Herausgeberwerke und Kapitel aus Herausgeberwerken) indexiert. PsycINFO beinhaltet Zeitschriftenartikel in mehreren Sprachen; bei anderen Literaturgattungen werden ausschließlich englischsprachige Publikationen aufgenommen. PSYINDEX beinhaltet ausschließlich Literatur von Autoren aus dem deutschsprachigen Raum, sofern diese in deutscher oder englischer Sprache veröffentlicht wurde (Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID), 2012). Auf beide Fachdatenbanken kann über Datenbankanbieter wie OvidSP oder EBSCO zugegriffen werden. Daneben besitzt auch Google Scholar Relevanz für fachliche Recherchen in Psychologie (ACRL, 2010). Auch der Bibliothekskatalog der lokalen Universitätsbibliotheken sollte an dieser Stelle erwähnt werden. Seit einiger Zeit steht zudem das Suchportal PubPsych für fachliche Recherchen im Fach Psychologie zur Verfügung; auf beide Angebote wird hier

jedoch nicht weiter eingegangen, da sie für das Verständnis der Rechercheaufgaben nicht notwendig sind.

Mithilfe von Fachdatenbanken kann relativ präzise recherchiert werden, da anhand eines kontrollierten Vokabulars (Thesaurus) gesucht werden kann. Bei komplexeren Suchanfragen können Schlagwörter (also Wörter aus dem kontrollierten Vokabular) mit Booleschen Operatoren (z.B. AND) verknüpft werden. Zudem bieten diese beiden Fachdatenbanken relativ spezifische Filterfunktionen, welche auf den verschiedenen Datenbankfeldern beruhen (Franke, Klein, & Schüller-Zwierlein, 2010). PsycINFO und PSYINDEX enthalten beispielsweise ein Datenbankfeld für die in der Publikation verwendete Methodik, welches es ermöglicht, gezielt nach Meta-Analysen, Review-Artikeln oder empirischen Studien zu suchen.

Google Scholar stellt seit ungefähr zehn Jahren für Psychologen eine Alternative zu den oben beschriebenen Fachdatenbanken dar. In der Literatur ist man sich relativ einig, dass Google Scholar ergänzend zu Fachdatenbanken verwendet werden sollte, jedoch keinesfalls Fachdatenbanken ersetzen kann (für eine Übersicht siehe Leichner, Peter, Mayer, & Krampen, 2014c). So wird das Fehlen spezifischer Filterfunktionen und eines kontrollierten Vokabulars beklagt (Giustini & Kamel Boulos, 2013). Zudem indexiert Google Scholar neben publizierter Fachliteratur auch sogenannte „graue Literatur“, also z.B. Berichte von Organisationen oder Instituten. Das kann durchaus als Vorteil angesehen werden, weil damit weitere relevante Literatur sichtbar gemacht wird, allerdings erfordert dies auch eine genauere Selektion und Qualitätskontrolle während oder nach der Recherche. Auch in Google Scholar können Boolesche Operatoren zur Verknüpfung von Suchbegriffen verwendet werden und durch Funktionen der erweiterten Suche (z.B. Autorensuche) ist eine weitere Präzisierung der Suche möglich (Google Inc., 2011).

Für die Recherche ist Google Scholar vor allem deswegen interessant, weil es damit zumindest in einigen Fällen möglich ist, im gesamten Artikel nach den Suchbegriffen zu suchen (Google Inc., 2011). Daher bietet sich eine Verwendung von Google Scholar insbesondere dann an, wenn die relevanten Thesaurusbegriffe oder Fachbegriffe nicht bekannt sind. So können mit Google Scholar relevante Publikationen möglicherweise auch dann gefunden werden, wenn die Fragestellung mit ähnlichen Begriffen beschrieben wird, die zwar im Text, aber nicht in Titel, Abstract oder Schlagwörtern enthalten sind. Prinzipiell können Google Scholar und Fachdatenbanken in einem solchen Fall auch kombiniert werden. Beispielsweise

könnte mit Google Scholar ein relevanter Artikel gefunden werden. Anhand dieses Artikels ist es möglich, die relevanten Fach- und Thesaurusbegriffe zu identifizieren und mit diesen anschließend eine Fachdatenbank zu durchsuchen. Aufgrund der benutzerfreundlichen Oberfläche, die auch eine Zeitersparnis bei der Recherche bewirkt, ist Google Scholar auch für einfache Suchanfragen geeignet (Giustini & Kamel Boulos, 2013).

5.5 Schlussfolgerung

Um Studierende auf fachliche Recherchen vorzubereiten, ist vor allem technisches Wissen über die verschiedenen Informationssysteme und deren Eigenschaften und Funktionsumfang notwendig. So sollten die Teilnehmer beispielsweise wissen, dass in PSYINDEX ausschließlich Literatur aus dem deutschsprachigen Raum enthalten ist, weshalb es für eine umfassende Recherche ungeeignet ist. Sollen aber gezielt deutschsprachige Publikationen gefunden werden, ist PSYINDEX das Mittel der Wahl. Auch der Funktionsumfang und die Verwendung der Fachdatenbanken (z.B. Verwendung der Filterfunktionen) sollten Teil von Schulungen sein. Da Google Scholar als sinnvolle Ergänzung zu Fachdatenbanken angesehen wird (ACRL, 2010), sollten dessen Stärken und Schwächen sowie die Verwendung ebenfalls Bestandteil von Schulungsbemühungen sein. Ergänzend sollte man die Studierenden sensibilisieren, insbesondere bei Recherchen mit Google Scholar, auf die Qualität und Glaubwürdigkeit der Publikationen zu achten.

6 Kritik

An dieser Stelle soll auch erwähnt werden, dass es Kritik am vorliegenden Konzept von Informationskompetenz und dessen Verbreitung und Umsetzung gibt.

Der erste Kritikpunkt bezieht sich auf die Terminologie. So wird kritisiert, dass überwiegend bibliothekarische Inhalte vermittelt werden, welche zuvor als "bibliographic instruction" vermarktet wurden (McCrack, 1991). Daher wird die Frage gestellt, ob die Umbenennung notwendig und sinnvoll war. Isaacson (2003, p. 42) schreibt: "Let's just say we want to teach people how to use libraries. This is not sexy. It doesn't billow the sails the way information literacy does. But why should we use windy phrases at all?".

Der zweite Kritikpunkt bezieht sich auf gängige Definitionen von Informationskompetenz. Informationskompetenz wird vielfach als kontextunabhängige Kompetenz im Umgang mit Informationen verstanden (Ingold, 2005, p. 99). Eine solche Definition schafft ein in der Praxis nicht umsetzbares Konstrukt, mit dem mehr versprochen wird, als von Bibliothekaren geboten werden kann (Ingold, 2005; Owusu-Ansah, 2003). Entsprechend kommen bei praktischen Umsetzungen verkürzte Definitionen zum Einsatz. Meist beschränken sich derartige Schulungen im akademischen Bereich auf für Studierende unmittelbar nutzbare fachliche Ressourcen, wie Bibliothekskataloge oder Fachdatenbanken. So zeigt die gemeinsame Schulungsstatistik von ungefähr 80 deutschen Bibliotheken (Krüger, 2014), dass die inhaltlichen Schwerpunkte der Schulungen im Jahr 2013 überwiegend bei der Bibliotheksbenutzung, einzelnen Katalogen/Datenbanken oder Suchstrategien lagen. Auch für die Jahre 2011 und 2012 zeigen sich ähnliche Schwerpunkte.

Ein weiterer Kritikpunkt ist die der Informationskompetenzbewegung zugrunde liegende Annahme, dass auf Seiten der Benutzer ein Defizit bestehe, welches durch Schulungen zu beheben sei. Foster (1993) beispielsweise argumentiert, dass hier von Bibliothekaren gezielt ein Übel heraufbeschworen werde, für welches ausschließlich Bibliothekare eine Lösung hätten (vgl. Ingold, 2005, p. 76). Kritiker sahen die Informationskompetenzbewegung, hinter der mehrheitlich Bibliothekare und deren Verbände standen, daher von Anfang an als Maßnahme, um den Status und die Wichtigkeit von Bibliotheken zu erhalten bzw. zu erhöhen (z.B.

McCrank, 1991). Eine Übersicht der kritischen Äußerungen zum Bereich Informationskompetenz findet man bei Ingold (2005, pp. 76–99).

7 Vergleich mit verwandten Konzepten

Eine enge inhaltliche Verbindung zeigt sich zwischen den Kompetenzfeldern Informationskompetenz und Medienkompetenz ("media literacy") (Gapski & Tekster, 2009). Gapski (2009) fasst bestehende Definitionen von Medienkompetenz so zusammen: „[die] Fähigkeit, Medien (kritisch/selbstbestimmt/kreativ/verantwortlich) (anwenden/verstehen/gestalten) zu können“. Auch ein aktueller Überblicksartikel von Potter (2010) kommt zu der Schlussfolgerung, dass ein wesentlicher Bestandteil vorliegender Definitionen von Medienkompetenz die Anwendung kritischen Denkens ist. Unterschiede gibt es allenfalls in der Art der Medien. So beziehen sich Medienkompetenz bei einigen Autoren nur auf ein Medium, z.B. Computer oder Fernsehen. Allerdings scheint sich in dieser Frage der Konsens durchzusetzen, dass zwischen den verschiedenen Massenmedien keine wesentlichen Unterschiede bestehen, die eine Einschränkung auf ein Medium notwendig machen.

Baacke (1998) definiert vier Dimensionen von Medienkompetenz: Medienkritik, -kunde, -nutzung und -gestaltung. Mit der Medienkritik ist die Fähigkeit gemeint, Medien und die durch sie transportierten Informationen kritisch zu bewerten. So sollte eine medienkompetente Person beispielsweise in der Lage sein, zu erkennen, dass sich viele Medien durch Werbung finanzieren und dass dies Einfluss auf Inhalte und Darstellung haben kann. Medienkunde bezeichnet das Wissen über verschiedene Informationssysteme einschließlich der Fähigkeit, diese auch einsetzen zu können. Mit Mediennutzung ist sowohl die rezeptive Nutzung, als auch die Erstellung und Verbreitung neuer Informationen gemeint. Eine medienkompetente Person ist zudem in der Lage, sich bei der Weiterentwicklung von Medien einzubringen (Mediengestaltung). Auch hier zeigt sich eine große inhaltliche Nähe der beiden Kompetenzfelder; Gapski (2012, p. 170) sieht beide daher auch als „begriffliches Geschwisterpaar“. Entsprechend sieht er bei beiden Begriffen das Problem der Eingrenzung. Beschränkt man Informationskompetenz auf fachliche Informationen und Medienkompetenz auf Computernutzung, so „verkürzt man die gesellschaftliche Reichweite und den pädagogischen Anspruch“. Beschreibt man beide Begriffe jedoch in der möglichen Breite und bezogen auf alle möglichen Kontexte und Nutzergruppen, so reflektiert dies die gesellschaftliche Bedeutung, es entsteht jedoch ein Konstrukt, das sich kaum noch operationalisieren lässt (Gapski, 2012, p. 170).

Bawden (2001) hingegen versteht die Felder "library literacy", "media literacy" und "computer literacy" als sogenannte "skill based literacies". Sein Verständnis von "media literacy" ist etwas enger als das oben erwähnte Medienkompetenzkonzept. So versteht Bawden (2001, p. 225) darunter die Fähigkeit, kritisches Denken auf Informationen aus Massenmedien anzuwenden. Die Elemente, die sich auf Informationsproduktion mithilfe von Medien beziehen, entfallen. "Library literacy" bezeichnet die Fähigkeit, Ressourcen der Bibliothek zu nutzen. Das Konzept "computer literacy" kann als die Fähigkeiten in der Computernutzung verstanden werden, die notwendig sind, um in der Informationsgesellschaft handeln zu können. Diese drei "skill based literacies" können als Bestandteile von Informationskompetenz angesehen werden, da sie erforderlich sind, um sich in Informationsumwelten zurechtzufinden (Bawden, 2001, p. 230).

Da dieser Arbeit ein relativ enges Verständnis von Informationskompetenz, bezogen auf das Auffinden und Nutzen von Fachinformationen und Bibliotheksressourcen, zugrunde liegt, ist eine Abgrenzung zu Medienkompetenz relativ klar und einfach vorzunehmen. So ist mit Medienkompetenz primär die Nutzung von Informationen aus Massenmedien gemeint, während sich Informationskompetenz auf die Nutzung von Fachinformationen und Bibliotheksressourcen bezieht.

8 Selbstwirksamkeit

Der Begriff Selbstwirksamkeit (engl. "self-efficacy") bezieht sich auf persönliche Überzeugungen, eine bestimmte Aufgabe oder eine Gruppe von Aufgaben erfolgreich bearbeiten oder lösen zu können (Bandura, 1977, 1982). Dabei geht es nicht um Routineaufgaben, sondern um mehr oder weniger komplexe Aufgaben, die Anstrengung und Ausdauer erfordern (Schwarzer & Jerusalem, 2002). Das Konstrukt ist in die sozial-kognitive Theorie von Bandura (u.a. Bandura, 2001) eingebettet, wonach subjektive Überzeugungen eine wichtige Rolle für kognitive, motivationale und emotionale Prozesse, sowie Handlungen spielen. Dabei geht es insbesondere um Konsequenzerwartungen und Selbstwirksamkeitserwartungen. Erstere beziehen sich auf die Annahme, dass bestimmte Handlungen bestimmte Konsequenzen haben. So wird ein Student beispielsweise annehmen, dass das Lernen in Lerngruppen eine bessere Note nach sich ziehen wird, als individuelles Lernen. Selbstwirksamkeitserwartungen (bzw. -überzeugungen) beziehen sich darauf, ob man sich in der Lage sieht, die relevanten und notwendigen Handlungen auszuführen (Schwarzer & Jerusalem, 2002). Man kann also davon ausgehen, dass auch bei der Recherche nach Fachinformationen Selbstwirksamkeitserwartungen im Spiel sind.

Grundsätzlich kann man zwischen spezifischen und allgemeinen Selbstwirksamkeitsüberzeugungen unterscheiden. Spezifische Überzeugungen beziehen sich auf bestimmte Handlungen in bestimmten Situationen, während allgemeine Selbstwirksamkeitsüberzeugungen nicht auf Situationen bezogen sind und mit Items wie „Schwierigkeiten sehe ich gelassen entgegen, weil ich immer auf meine Fähigkeiten vertrauen kann“ (Jerusalem & Schwarzer, 1999) erfasst werden. Forschungsergebnisse haben gezeigt, dass sich Verhalten anhand spezifischer Selbstwirksamkeitsüberzeugungen wesentlich besser vorhersagen lässt, als anhand allgemeiner Überzeugungen (Maddux, 1995; Schwarzer & Jerusalem, 2002).

Ein positiver Zusammenhang zwischen Selbstwirksamkeitserwartungen und Leistung, oder anderen positiven Ergebnissen, konnte in klinischen, akademischen und betrieblichen Settings gezeigt werden (Stajkovic & Luthans, 1998). So scheint die wahrgenommene Fähigkeit, mit Panikattacken umgehen oder diese kontrollieren zu können („Panik-Selbstwirksamkeit“, Übers. d. Verf.) den Erfolg kognitiver Verhaltenstherapie bei Panikstörungen zu vermitteln (Fentz, Arendt, O'Toole, Hoffart, & Hougaard, 2014). In akademischen Kontexten gibt es ver-

lässliche Hinweise, dass Selbstwirksamkeit mit Leistung in Verbindung steht, wie zwei Meta-Analysen zeigen (Richardson, Abraham, & Bond, 2012; Robbins et al., 2004). Auch in beruflichen Kontexten liegen meta-analytische Befunde für die positiven Effekte von Selbstwirksamkeit auf Leistung vor (Judge & Bono, 2001; Stajkovic & Luthans, 1998).

Dabei wird angenommen, dass Personen mit hohen Selbstwirksamkeitserwartungen mehr Anstrengung und Ausdauer bei der Bearbeitung der Aufgaben an den Tag legen (Maddux, 1995). Zudem wird davon ausgegangen, dass solche Personen sich auch höhere Ziele setzen, welche zu mehr Leistung anspornen (Locke & Latham, 2002). Allerdings kann gute Leistung umgekehrt auch eine Ursache für erhöhte Selbstwirksamkeit sein. Vergangene Erfolgserfahrungen sind die wichtigste Quelle von Selbstwirksamkeitserwartungen (Schwarzer & Jerusalem, 2002). Dieser Zusammenhang ist Bestandteil des "high-performance-cycle" (Latham, Locke, & Fassin, 2002, Locke & Latham, 1990a, 1990b), mit dem primär der Zusammenhang von Arbeitszufriedenheit und Leistung am Arbeitsplatz erklärt werden soll. Der Zyklus beginnt mit einer guten Leistung – bedingt durch hohe Selbstwirksamkeit und ambitionierte Ziele. Wird diese Leistung honoriert, führt dies zu Arbeitszufriedenheit und weiterem Engagement. Das Erreichen von Zielen kann auch zu einer Verbesserung der Selbstwirksamkeit führen, was den Effekt verstärkt. Dies entspricht den Ergebnissen einer Meta-Analyse von Sitzmann und Yeo (2013), die die Effekte von Selbstwirksamkeit im zeitlichen Verlauf auf Personenebene untersucht. Demnach ist Selbstwirksamkeit eher ein Ergebnis vergangener, als ein Prädiktor zukünftiger Leistung (Sitzmann & Yeo, 2013).

9 Kompetenzbegriff

In dieser Arbeit wurden die Begriffe Kompetenz und Fähigkeit teilweise synonym verwendet. Daher sollen an dieser Stelle Definitionen der Begriffe behandelt werden, um Unterschiede in den Definitionen aufzuzeigen.

Der Begriff Kompetenz wird insbesondere dann verwendet, wenn Leistungsfähigkeit in einem Anwendungskontext betrachtet wird (Hartig & Klieme, 2006). Dennoch besteht kein einvernehmlicher Konsens darüber, was unter einer Kompetenz zu verstehen ist und was Kompetenzen von Intelligenz unterscheidet. So scheinen, insbesondere im Kontext der ABO-Psychologie, eine Vielzahl von Definition zu existieren, deren gemeinsamer Nenner darin besteht, dass Kompetenz mit Leistung am Arbeitsplatz zusammenhängt (vgl. Shippmann et al., 2000). So schreibt Mirabile (1997): "a knowledge, skill, or ability associated with high performance on a job".

Für eine wissenschaftliche Beschäftigung ist jedoch eine klare Definition notwendig (Hartig & Klieme, 2006). Für den Bereich der Bildungsforschung und pädagogischen Psychologie hat man sich daher, zumindest im deutschsprachigen Raum, auf eine von Weinert (2001) empfohlene Definition geeinigt (vgl. Hartig & Klieme, 2006). Demnach sind Kompetenzen kontextspezifische kognitive Fähigkeiten und Fertigkeiten, die die Lösung bestimmter Probleme ermöglichen (siehe auch Weinert, 2014). An dieser Definition sind vor allem zwei Aspekte bemerkenswert. Zum einen werden für Leistung erforderliche affektive und motivationale Voraussetzungen nicht berücksichtigt, was die Messung erleichtert. Zum zweiten unterscheidet sich Kompetenz von Intelligenz durch die Trainierbarkeit und den Bezug auf spezifische Situationen. Beides wird bei dem Konstrukt Intelligenz typischerweise nicht angenommen; so gilt Intelligenz als nur in geringem Ausmaß trainierbar und als nicht an eine Situation gebunden (Hartig & Klieme, 2006). Damit folgt diese Definition einer Forderung von McClelland (1973), der dazu aufrief, im Bildungssystem anstelle von Intelligenz die für spezielle Tätigkeiten notwendigen Voraussetzungen zu erfassen (siehe auch Hartig & Klieme, 2006). Fähigkeiten und Fertigkeiten wären demnach Bestandteil von Kompetenzen, wenn sie in der durch die Kompetenz definierten Situation (bzw. Klasse von Situationen) von Nutzen sein können. Damit passt der Kompetenzbegriff in ein Modell von Mayer (2003), welches den Zusammenhang von Fähigkeit, Erfahrung, Kenntnis und Kompetenz zu erklären versucht.

Fähigkeit (“ability”), wird in diesem Modell als mehr oder weniger unveränderbar angenommen und mit fluider Intelligenz (vgl. Horn & Cattell, 1966) gleichgesetzt. Durch das Zusammenwirken von Fähigkeit und Erfahrung bzw. Übung (“experience”) entstehen Kenntnisse (“knowledge”). Eine Kompetenz (“competency”) wird hier als spezialisierte Form von Kenntnissen verstanden.

Eine Auslegungssache ist in diesem Zusammenhang jedoch das Ausmaß der Kontextspezifität. So ist eine gewisse Generalisierbarkeit über Situationen notwendig, um Schlüsselkompetenzen definieren zu können (Hartig & Klieme, 2006); das Problem kann hier aber nur angeschnitten werden. Weinert (2014) unterscheidet für den Bildungsbereich fachliche, überfachliche und Handlungskompetenzen. Fachliche Kompetenzen beziehen sich auf eine klar definierte Domäne, z.B. Mathematik, Physik. Beispiele für überfachliche Kompetenzen sind Teamfähigkeit oder Problemlösen. Mit Handlungskompetenzen ist gemeint, dass derjenige seine Kenntnisse und Fertigkeiten in Lebenssituationen anwenden kann.

9.1 Deklaratives und prozedurales Wissen

Man kann davon ausgehen, dass Kompetenzen sowohl deklarative als auch prozedurale Komponenten enthalten (Hartig & Klieme, 2006; Mayer, 2003). Diese Unterscheidung ist eine grundlegende Frage in vielen Konzepten und Theorien über Wissenserwerb, -speicherung, und -anwendung. Üblicherweise versteht man hierunter eine Unterscheidung zwischen Hintergrundwissen, beispielsweise über Konzepte, und Wissen über Prozeduren (Hiebert & Lefevre, 1986). So unterscheidet Piaget (1974/1978) zwischen konzeptuellem Wissen und erfolgreichem Handeln. Anderson (1983, pp. 215–241, 1996) unterscheidet zwischen deklarativem und prozeduralem Wissen. Diese Unterscheidung ist Teil von Anderson’s kognitiver Architektur “ACT” (später “ACT-R”, “adaptive control of thought-rational”) – einem formalisierten Modell, das mit dem Ziel entworfen wurde, Gedächtnisphänomene erklären zu können. Dieses Modell sieht eine Trennung zwischen deklarativen Inhalten, also Fakten, und Produktionsregeln (welche prozedurales Wissen enthalten) vor. Als Beispiel für deklaratives Wissen nennt Anderson (1996) den Fakt, dass die Addition der Zahlen drei und vier sieben ergibt. Eine entsprechende Produktionsregel beschreibt das Verfahren bei der Addition von Zahlen. Prozedurales und deklaratives Wissen sind in diesem Modell eng verbunden, da Produktionsregeln mit Begriffen beschrieben werden, deren Definitionen in de-

klarativen Wissensbeständen enthalten sind. Prozedurale Wissensbestände sind komplexe kognitive Fähigkeiten, wie z.B. mathematisches Problemlösen oder Computerprogrammierung (Anderson, 1983, p. 215).

Das Fach Mathematik eignet sich aufgrund des strukturierten und genau definierten Inhalts besonders gut für die Unterscheidung und Untersuchung von deklarativem (oder konzeptuellem) und prozeduralem Wissen (Hiebert & Lefevre, 1986, p. 1). Unter deklarativem Wissen wird hier Faktenwissen über Kernkonzepte und deren Zusammenhänge verstanden (Schneider & Stern, 2005). Mathematisches prozedurales Wissen besteht aus zwei Teilen. Zum einen handelt es sich um Symbole und entsprechendes Wissen über die syntaktischen Regeln, nach denen diese Symbole angeordnet und kombiniert werden dürfen. Zum anderen werden hierunter Regeln, Algorithmen und Prozeduren zum Lösen mathematischer Probleme verstanden. Prozedurales Wissen in Mathematik ist, zumindest teilweise, automatisiert und nicht immer verbalisierbar (Schneider & Stern, 2005).

Ähnlich kann man sich eine Unterscheidung zwischen deklarativem und prozeduralem Wissen im Bereich Informationskompetenz vorstellen. Unter die deklarativen Anteile fallen unter anderem Faktenwissen über verschiedene Recherchesysteme sowie deren Funktionsumfang und Benutzung, Wissen über Publikationsmechanismen und Bewertungskriterien für Informationen. Prozedurales Wissen beinhaltet in diesem Fall Recherchestrategien und Prozeduren bei der Verwendung von Suchwerkzeugen. So können etliche Suchvorgänge (z.B. Verwendung der Thesaurussuche, Verknüpfung von Suchbegriffen durch Boolesche Operatoren, Einschränkung der Suchergebnisse anhand von Filterfunktionen) als Prozeduren verstanden werden. Man kann sich vorstellen, dass Personen mit Erfahrung in der Recherche diese Vorgänge entsprechend relativ automatisiert vornehmen. Bei entsprechender Erfahrung sollte auch das Erkennen der in der jeweiligen Situation bestmöglichen Vorgehensweise einer relativ automatisch ablaufenden Prozedur folgen. Dennoch kann man davon ausgehen, dass auch Personen mit ausgeprägter Erfahrung in der Recherche stets bewusst ist, was sie tun, welcher Schritt folgt und welchen Zweck ihr Handeln verfolgt. Damit unterscheidet sich dieses Verständnis prozeduralen Wissens von einer Konzeption, wonach man hierunter primär (halb)automatisierte motorische Prozesse versteht, welche dem Ausführenden nur in begrenztem Ausmaß bewusst sind und sprachlich beschrieben werden können (z.B. Buchner & Brandt, 2005, pp. 498–499). Beispielsweise lernten Probanden in einer Reaktionszeitauf-

gabe, bei der, in Abhängigkeit des Stimulus, jeweils eine von vier Tasten gedrückt werden musste, eine Tastenfolge (Sequenz), ohne diese vollständig beschreiben zu können. Dass dennoch ein Lernprozess stattgefunden hatte, konnte anhand der Reaktionszeiten demonstriert werden (Willingham, Nissen, & Bullemer, 1989).

10 Erfassung von Informationskompetenz im fachlichen Kontext

Vor der Beschäftigung mit Messverfahren muss eine genaue Operationalisierung des Konstrukts vorliegen. Wie oben dargelegt, gibt es viele Definitionen von Informationskompetenz, die sich in der Breite der Definition stark unterscheiden; Informationskompetenz ist ein Konzept mit vielen Facetten (Andretta, 2005, p. 12). Die folgende Übersicht konzentriert sich auf Erfassungsinstrumente für fachliche Informationskompetenz, also die Fähigkeit, in einem akademischen Kontext Fachinformationen zu finden und zu bewerten.

Übersichtsartikel berichten eine Vielzahl von Ansätzen zur Erfassung von Informationskompetenz, die einzelne Bestandteile oder Aspekte erfassen (McCulley, 2009; Walsh, 2009). McCulley (2009) unterscheidet drei Typen von Erhebungsinstrumenten: Tests und Umfragen, informelle Assessments und sogenannte "Performance Assessments" (vgl. Chang et al., 2012). Unter die Bezeichnung informelle Assessments fallen beispielsweise Beobachtung während Schulungen oder die Befragung von Schulungsteilnehmern, um zu prüfen, ob diese den Lernstoff verstanden haben. Damit kann nicht nur der Lernerfolg überprüft, sondern die Teilnehmer auch zur Reflektion der Materialien angeregt werden (McCulley, 2009). Allerdings sind solche Verfahren wegen des informellen Charakters von begrenztem Wert (Chang et al., 2012), weshalb im Folgenden nicht weiter auf solche Ansätze eingegangen wird. Umfragen sind primär für die Erfassung der Einstellung gegenüber Informationssystemen etc. geeignet (McCulley, 2009), auch darauf wird hier nicht weiter eingegangen.

10.1 Standardisierte Tests

In diese Kategorie fallen Tests mit gebundenem Antwortformat ("fixed-choice"), also Tests, bei denen zwischen mehreren Antwortmöglichkeiten gewählt werden muss. Der wohl prominenteste Typ sind Multiple-Choice-Tests, bei denen mindestens zwei und meist maximal vier oder fünf (Tarrant, Ware, & Mohammed, 2009) Antwortoptionen vorgegeben sind. Andere Beispiele für Fixed-Choice-Tests sind Aufgaben, die die Zuordnung von Begriffen, oder Antworten auf kurze offene Fragen erfordern (McCulley, 2009).

Multiple-Choice-Tests sind die populärste Erhebungsmethode für Informationskompetenz (Walsh, 2009). Die Hauptvorteile dieser Tests sind die Ökonomie bei der Anwendung und Auswertung sowie die hohe Reliabilität, die mit solchen Tests erreicht werden kann. Ferner

ist es damit prinzipiell möglich, die Ergebnisse verschiedener Studien miteinander zu vergleichen (Chang et al., 2012; Oakleaf, 2008). Zudem handelt es sich um ein verbreitetes und anerkanntes Testformat, sodass die Akzeptanz relativ hoch ist (Oakleaf, 2008). Kritiker hingegen argumentieren, dass mit solchen Tests keine komplexeren Denkprozesse erfasst werden können (Oakleaf, 2008). Ein anderer Kritikpunkt ist, dass mit solchen Tests lediglich Leistung in einer künstlichen Situation gemessen wird (Shavelson, 2010) und ein positives Testergebnis nichts darüber aussagt, ob derjenige auch tatsächlich erfolgreich Rechercheaufgaben bearbeiten kann (Oakleaf, 2008).

Für die Erfassung von Informationskompetenz lässt sich festhalten, dass mit solchen Aufgaben vor allem deklarative Wissensbestände erfasst werden können. Auch Strategien (also prozedurale Komponenten einer Kompetenz) lassen sich hiermit teilweise erfassen. Komplexes prozedurales Wissen (z.B. eine aus vielen Teilschritten bestehende Recherchestrategie) kann hiermit nicht erfasst werden. Es wird sogar teilweise argumentiert, dass Multiple-Choice-Verfahren für die Erfassung von Informationskompetenz (oder verwandten Konzepten) überhaupt nicht geeignet sind (Goldhammer, Kröhne, Keßel, Senkbeil, & Ihme, 2014).

Neben vielen lokal entwickelten Tests, die oft bibliotheksspezifische Items beinhalten (z.B. Noe & Bishop, 2005; Ondrusek, Dent, Bonadie-Joseph, & Williams, 2005), existieren, wie bereits erwähnt, zwei veröffentlichte Testverfahren, welche nach Auskunft der Autoren bibliotheksübergreifend einsetzbar sein sollen (Project SAILS, 2000-2014; Swain et al., 2014). Beispielhaft soll hier daher der "Information Literacy Test" (ILT, Swain et al., 2014) beschrieben werden. Der Test basiert auf den Informationskompetenzstandards der ACRL (2000). Da Kompetenzen bezüglich des vierten Standards (Nutzung von Informationen) sich nach Ansicht der Autoren nicht im Multiple-Choice-Format abprüfen lassen, bezieht sich der Test lediglich auf die verbleibenden vier Standards. Anhand ihrer psychometrischen Eigenschaften wurden 60 Items aus einem größeren Itempool ausgewählt. Dieser Test wurde von einer Stichprobe bestehend aus $n = 1035$ College-Studierenden bearbeitet. Für den Gesamttest ergab sich eine interne Konsistenz von Cronbach's $\alpha = .91$. Zusätzlich wurden Subskalen für die vier Standards gebildet; die Werte für die interne Konsistenz liegen entsprechend niedriger (Range: $\alpha = .61$ bis $\alpha = .84$). Für die Validität liegen, neben Expertenratings, welche Inhaltsvalidität belegen, mehrere empirische Befunde aus weiteren Studien vor. So korrelierte die Leistung im ILT sowohl mit der Leistung in einem ähnlichen Test, als auch mit der Durch-

schnittsnote, was als konvergente Validität (vgl. Moosbrugger & Kelava, 2008) interpretiert werden kann. Zudem zeigte sich, dass fortgeschrittene Studierende höhere Werte erreichten als Studienanfänger (Swain et al., 2014).

10.2 Performance Assessments

Als Alternative zu standardisierten Tests wird Performance Assessment (Leistungsbewertung, Übers. d. Verf) vorgeschlagen (McCulley, 2009; Tung, 2010). Mit diesem Begriff werden Erhebungskonzepte bezeichnet, bei denen die Studierenden ihre Fähigkeiten anwenden müssen, um ein Produkt zu erstellen (McCulley, 2009; Tung, 2010). Ein klassisches Beispiel aus dem Bibliotheksbereich ist die Auswertung von Bibliographien, welche die Studierenden als Teil einer Hausarbeit erstellt haben (Knight, 2006). Dabei kann die Angemessenheit der gefundenen Literatur bewertet werden, wobei geprüft wird, ob diese für das vorgegebene Thema relevant ist (McCulley, 2009). Einen Spezialfall stellen Portfolio-Auswertungen dar, die mehrere Arbeitserzeugnisse der Probanden (bspw. ein Literaturverzeichnis, eine Dokumentation der Suchstrategien oder ein kurzes Essay über eigene Erfahrungen während der Informationssuche) enthalten (Sonley, Turner, Myer, & Cotton, 2007). Auch Rechercheaufgaben (z.B. Chen, 2009; Julien & Barker, 2009) fallen in diese Kategorie.

Um derartige Aufgaben standardisiert auswerten zu können, ist es unerlässlich, entsprechende Auswertungsschlüssel zu konstruieren (Moskal, 2000; Oakleaf, 2008, 2009). Dabei kann man Auswertungsschlüssel anhand von zwei Kriterien unterscheiden: Anhand des Produkts und anhand ihrer Spezifität. Es kann entweder das gesamte Produkt (holistischer Auswertungsschlüssel) oder einzelne Aspekte des Produkts (analytischer Auswertungsschlüssel) bewertet werden. Zudem kann es sich um einen spezifischen Auswertungsschlüssel handeln, der nur auf eine bestimmte Gruppe von Aufgaben anwendbar ist, oder um einen allgemeinen Auswertungsschlüssel, der auf eine größere Gruppe von Aufgaben angewendet werden kann (Moskal, 2000). Wie auch bei der Verhaltensbeobachtung gilt: Je präziser und konkreter die Beobachtungseinheiten definiert sind, desto reliabler ist die Auswertung möglich (Fisseni, 1997, p. 207).

Solche Verfahren bieten mehrere Vorteile gegenüber klassischen Fixed-Choice-Tests: Die Erfassung von Lernerfolg in einer authentischen Situation, den klaren Bezug zu Lernzielen und die Erfassung komplexerer Denkprozesse (Oakleaf, 2008). Außerdem wird argumentiert,

dass während der Bearbeitung solcher Aufgaben Lernprozesse stattfinden können (McCulley, 2009; Oakleaf, 2008). Die schwerwiegendsten Nachteile solcher Verfahren sind der hohe Aufwand bei der Konstruktion und – vor allem – der Auswertung sowie die mangelnde Standardisierung und geringe Reliabilität. Allerdings lassen sich viele Probleme bezüglich Reliabilität und Standardisierung auf schlecht/mangelhaft konstruierte Auswertungsschlüssel zurückführen (Oakleaf, 2009). So zeigte eine Übersichtsarbeit, dass die Reliabilität von Performance Assessments durch den Einsatz analytischer und spezifischer Auswertungsschlüssel entscheidend verbessert werden kann (Jonsson & Svingby, 2007).

Beispielhaft sollen hier drei Studien beschrieben werden, in denen dieser Ansatz verfolgt wurde. In der Studie von Dunn (2002) wurden Rechercheaufgaben im Rahmen einer umfangreichen Studie zur Erfassung des Standes der Informationskompetenz bei den Studierenden der California State University eingesetzt. Hier kamen zwei Methoden zum Einsatz. Um Informationsverhalten bei der Benutzung elektronischer Systeme zu erfassen, wurden den Studierenden Rechercheaufgaben vorgegeben und die gesamte Aktivität am PC wurde mithilfe von Screen-Capture-Software aufgezeichnet. Hiermit wurden sämtliche aufgerufenen Websites, einschließlich der verwendeten Suchbegriffe etc. registriert. Um das Informationsverhalten genauer zu verstehen, wurde eine kleine Gruppe Studierender von Untersuchern beobachtet, wobei Fragen zum Vorgehen gestellt wurden.

Einen ähnlichen Ansatz verfolgte eine Studie von Novotny und Cahoy (2006), in der über die Evaluation einer Informationskompetenzschulung für Studierende auf „undergraduate“-Niveau berichtet wird. Auch hier wurde die Rechercheaktivität mittels Screen-Capture-Software aufgezeichnet. Zudem wurden die Probanden aufgefordert, laut zu denken, was ebenfalls protokolliert wurde. Einen ökonomischeren Ansatz beschreibt Chen (2009), der über eine Untersuchung berichtet, die im Rahmen eines mehrwöchigen Informationskompetenzkurses durchgeführt wurde. Während des Kurses wurden Recherchen durchgeführt; die Teilnehmer wurden mithilfe von Fragebögen bezüglich der verwendeten Suchbegriffe und über ihre Zufriedenheit mit den Suchergebnissen befragt.

Mit solchen Verfahren kann überprüft werden, ob komplexe Recherchestrategien beherrscht werden und angewendet werden können. Unter dem Gesichtspunkt, dass Informationskompetenz vor allem als Prozess konzeptualisiert wird (vgl. Abschnitt 3, S. 8), erscheinen derartige Erfassungsverfahren wichtig. Bei der Bewertung von fertigen Produkten, beispielsweise

Bibliographien, kann jedoch im Nachhinein nicht mehr nachvollzogen werden, wie die Probanden diese erstellt haben. Daher ist es angemessener, ein Verfahren zu wählen, bei dem auch das Vorgehen dokumentiert und ausgewertet werden kann. Wird lediglich das Ergebnis der Recherche betrachtet, könnte es sich auch um einen „Zufallstreffer“ handeln, der nichts über die tatsächliche Informationskompetenz des Betreffenden aussagt (vgl. Leichner et al., 2014a).

10.3 Schlussfolgerung

Wie oben dargelegt, beinhaltet Informationskompetenz sowohl deklarative, als auch prozedurale Wissensanteile. Während sich deklarative und weniger komplexe prozedurale Anteile mit Multiple-Choice-Tests erfassen lassen, muss für die Erfassung komplexerer prozeduraler Anteile (z.B. eine aus vielen Schritten bestehende Recherchestrategie) auf Performance Assessments zurückgegriffen werden. Daher bietet sich in einem solchen Fall die Kombination mehrerer Messinstrumente im Sinne multimethodaler Erfassung an. Multimethodale Erfassung ist immer dann eine Option, wenn ein relevantes Phänomen mehrere Komponenten aufweist – wie in diesem Fall deklarative und prozedurale Anteile (vgl. Eid & Diener, 2006). Im Folgenden wird über die Entwicklung von zwei Erfassungsinstrumenten für Informationskompetenz, einen Multiple-Choice-Test und fachliche Rechercheaufgaben, berichtet. Damit soll fachliche Informationskompetenz multimethodal erfasst werden. Zudem wird eine Selbstwirksamkeitsskala entwickelt, mit der die wahrgenommene Fähigkeit in der Informationsrecherche erfasst werden kann.

11 Entwicklung von Multiple-Choice-Test und Selbstwirksamkeitsskala

Im Folgenden wird über die Entwicklung des Multiple-Choice-Tests und der Selbstwirksamkeitsskala berichtet. Ziel der Testentwicklung war es, einen Test für Informationskompetenz mit gebundenem Antwortformat zu erstellen, der für die ökonomische und reliable Erfassung von Informationskompetenz bei Psychologiestudierenden genutzt werden kann. Mit der Selbstwirksamkeitsskala sollten Zusammenhänge zwischen der selbsteingeschätzten Wirksamkeit und der tatsächlichen Leistung im Multiple-Choice-Test, bzw. den Rechercheaufgaben, untersucht werden. In beiden Fällen lagen keine einsetzbaren deutschsprachigen Instrumente vor.

11.1 Entwicklung des Multiple-Choice-Tests

Der Test basiert auf den auf das Fach Psychologie bezogenen Informationskompetenzstandards der ACRL (2010), wobei der Schwerpunkt, wie beim ILT (Swain et al., 2014), auf Standards 2 und 3, also dem Suchen und Bewerten von Informationen, lag. Fähigkeiten bei der Weiterverarbeitung von Informationen (Standard 4), werden von diesem Test nicht erfasst. Dafür gibt es mehrere Gründe: Zum einen sind Fähigkeiten bezogen auf Standard 4 schwerlich mit Multiple-Choice-Items erfassbar (vgl. Swain et al., 2014). Zum anderen wird die Weiterverarbeitung von Informationen an deutschen Universitäten primär im Rahmen von Schulungen zum wissenschaftlichen Arbeiten gelehrt und gehört nicht zu den traditionellen bibliothekarischen Inhalten (Homann, 2007). Auch rechtliche Aspekte bei der Nutzung von Informationen, sowie das Erkennen und Definieren eines Informationsbedarfs (Standard 1) werden nur begrenzt erfasst, da auch diese eher in den Bereich wissenschaftlichen Arbeitens fallen. Wie oben beschrieben, können mit derartigen Tests vor allem deklarative Aspekte von Informationskompetenz erfasst werden.

Die Testitems wurden anhand eigener Überlegungen und anhand von publizierten englischsprachigen Tests (insbesondere Noe & Bishop, 2005; Ondrusek et al., 2005) entwickelt. Dabei wurde darauf geachtet, psychologiespezifische Inhalte von Informationskompetenz zu erfassen. So erfordert die korrekte Beantwortung der Items die Kenntnis von psychologischen Fachdatenbanken und des Zitationsstils der APA. Ferner ist Wissen über den Funktionsum-

fang psychologischer Fachdatenbanken erforderlich. Es wurden 33 Multiple-Choice-Items mit drei Antwortoptionen erstellt. Da bei nur drei Antwortoptionen die Ratewahrscheinlichkeit nicht unerheblich ist, variierte die Anzahl korrekter Antwortalternativen von Item zu Item, bei einigen war keine einzige korrekt („pick any out of n“, vgl. Rost, 2004, pp. 63–64).

Zwei Items weichen von diesem Muster ab: Ein Multiple-Choice-Item verfügt lediglich über zwei Antwortoptionen, da hier kein weiterer geeigneter Distraktor gefunden werden konnte. Ein weiteres Item („Welche Kriterien können verwendet werden, um die Güte eines gefundenen Zeitschriftenartikels oder Buches einzuschätzen?“) erfordert eine kurze offene Antwort in Stichwortform, die anhand einer Auswertungsvorschrift ausgewertet werden kann. Die Hinzunahme dieses Items sollte die Testbearbeitung aus Sicht der Probanden abwechslungsreicher gestalten. Insgesamt wurden $k = 35$ Items entwickelt, davon waren anhand inhaltlicher Gesichtspunkte $k = 23$ Items Standard 2 und $k = 12$ Items Standard 3 der ACRL-Standards (ACRL, 2010) zugeordnet.

Für jedes Item wird maximal ein Punkt vergeben; für jede korrekt gewählte oder ausgelassene Antwortoption folglich 0.33 Teilpunkte. Bei dem Item mit nur zwei Antwortoptionen werden entsprechend 0.5 Teilpunkte pro Antwortoption vergeben. Für die korrekte Beantwortung des Items mit offenem Antwortformat wurden maximal 3 Punkte vergeben. Die bei diesem Item erreichten Punkte wurden anschließend durch 3 geteilt, sodass maximal ein Punkt erreicht werden konnte.

11.2 Entwicklung der Selbstwirksamkeitsskala

Basierend auf verschiedenen Informationskompetenz-Standards (ACRL, 2000; DBV, 2009), vorliegenden Online-Trainingsmaterialien (Krüger, 2008) und eigenen Überlegungen wurden $k = 10$ Items entwickelt. Jedes Item erforderte eine Stellungnahme zu einer Aussage wie z.B. „Wenn ich einen Vortrag vorbereite, kann ich zu Beginn gut einschätzen, wie viel Literatur ich ungefähr benötigen werde“ auf einer fünfstufigen Likert-Skala mit „keine Angabe“- bzw. „weiß-nicht“-Option. Diese Angaben wurden als fehlende Werte behandelt.

Dabei wurde darauf geachtet, alle relevanten Aspekte fachlicher Recherchen abzubilden. So wird die Selbstwirksamkeit beim Erkennen und Einschätzen eines Literaturbedarfs, bei der Recherche in verschiedenen Informationssystemen sowie bei der Beschaffung und Bewertung von Literatur erhoben. Um die Bearbeitungszeit gering zu halten und einen ökonomi-

schen Einsatz zu ermöglichen, wurde auf die Entwicklung weiterer Items verzichtet. Da viele Bereiche mit nur einem Item erfasst wurden, wurde zudem auf eine Zuordnung der Items zu einzelnen Bereichen von Informationskompetenz, wie bei der englischsprachigen Selbstwirksamkeitsskala von Kurbanoglu, Akkoyunlu und Umay (2006), verzichtet. Sämtliche Items sind im Anhang zu finden.

11.3 Vorstudie

Um beide Instrumente zu erproben und erste psychometrische Kennwerte zu erhalten, wurde eine Vorstudie an der Universität Trier durchgeführt. Bezüglich des Multiple-Choice-Tests sollte überprüft werden, ob sich Subskalen zu Standards 2 und 3 empirisch trennen lassen. Zudem wurden Gruppenunterschiede vorhergesagt: Je weiter der Fortschritt des Studiums, desto höhere Testwerte wurden erwartet. Bezüglich der Selbstwirksamkeitsskala wurde keine mehrdimensionale Struktur erwartet, da es sich nur um 10 Items handelt. Weil mit den Items Selbstwirksamkeit bezüglich unterschiedlicher Facetten von Informationsrecherchen erfasst wird, sollte die interne Konsistenz im mittleren Bereich liegen. Auch hier wurden Gruppenunterschiede, nämlich ein Anstieg der Selbstwirksamkeit im Verlauf des Studiums, erwartet. Dies wurde angenommen, da auch die Testwerte im Laufe des Studiums ansteigen sollten und zumindest eine gewisse Korrespondenz von Leistung und Selbsteinschätzung erwartet wurde. Entsprechend wurde auch eine signifikante Korrelation zwischen Multiple-Choice-Test und Selbstwirksamkeitsskala prädiziert.

11.3.1 Methode

An der Erprobungsstudie nahmen $N = 64$ Psychologiestudierende der Universität Trier teil. Dabei handelte es sich um $n = 22$ Studienanfänger aus dem 2. Semester des BSc-Studiengangs (Alter $M = 21.77$, 77% weiblichen Geschlechts) sowie um $n = 21$ fortgeschrittene Studierende aus dem BSc-Studiengang (ab 5. Semester) und aus den MSc- und Diplom-Studiengängen (Alter $M = 23.90$, 86% weiblichen Geschlechts). Zudem nahmen $n = 21$ Doktoranden an der Studie teil (Alter $M = 28.48$, 71% weiblichen Geschlechts). Die Probanden bearbeiteten die Instrumente in *paper-and-pencil*-Form. Die Datenerhebungen fanden als Einzelsitzungen in den Räumen des ZPID statt; die Probanden wurden für die Teilnahme entlohnt.

11.3.2 Ergebnisse

Multiple-Choice-Test

Bei explorativen Analysen zeigte sich, dass ein Item zu leicht war und von allen Teilnehmenden korrekt bearbeitet wurde. Dieses Item wurde von allen folgenden Analysen ausgeschlossen. Anschließend wurde eine Hauptkomponentenanalyse mit den verbleibenden 34 Items durchgeführt. Die Inspektion des Scree-Plots indizierte, dass die Daten am besten mit einer Ein-Faktoren-Lösung beschrieben werden können, obwohl weitere Faktoren mit Eigenwerten größer als eins extrahiert wurden (Verlauf der Eigenwerte > 1: 5.26, 2.57, 2.32, 2.06, 1.90, 1.61, 1.56, 1.50, 1.34, 1.27, 1.22, 1.13). Der erste Faktor konnte 15% der Varianz aufklären. Die Unterscheidung zwischen Standards 2 und 3 zeigte sich also nicht empirisch. Anschließend wurden alle 22 Items ausgewählt, die Faktorladungen auf dem ersten Faktor über $r_{it} = .20$ aufwiesen. Ein höherer Grenzwert (beispielsweise der üblich verwendete Grenzwert von $r_{it} = .30$) hätte zum Ausschluss sehr vieler Items geführt und den Umfang des Tests zu sehr eingeschränkt. Die Gesamtskala wurde durch Berechnen des Mittelwerts gebildet. Da eine Unterscheidung zwischen Standards 2 und 3 als inhaltlich bedeutsam angesehen wurde, wurden dennoch zwei Subskalen (Skala 1 und 2) durch Berechnen des Mittelwerts gebildet. Die Kennwerte für die Gesamtskala und die beiden Subskalen finden sich in Tabelle 1.

Tabelle 1

Deskriptive Statistiken und psychometrische Kennwerte für die drei Skalen des Multiple-Choice-Tests (Vorstudie)

| Skala | <i>k</i> | <i>M (SD)</i> | α | $r_{tt'}$ |
|--------------------|----------|---------------|----------|-----------|
| Gesamtskala | 22 | .64 (.18) | .82 | .78 |
| Skala 1 (Suchen) | 14 | .68 (.19) | .73 | .65 |
| Skala 2 (Bewerten) | 8 | .58 (.44) | .73 | .73 |

Anmerkung. *k* = Anzahl der Items; α = Cronbach's Alpha; $r_{tt'}$ = Split-Half-Reliabilität (Spearman-Brown).

Wie erwartet korrelierten die beiden Subskalen signifikant miteinander ($r = .60, p < .01$, einseitig). Bei der Analyse der Gruppendifferenzen zeigte sich, dass Doktoranden erwartungskonform höhere Werte erzielten als die fortgeschrittenen Studierenden, welche wiederum höhere Werte erzielten als die Studienanfänger; die Mittelwerte sind in Tabelle 2 wiedergegeben.

Tabelle 2

Punktwerte im Multiple-Choice-Test nach Studienfortschritt (Vorstudie)

| Skala | Studienanfänger <i>M(SD)</i> | Fortgeschrittene Studierende <i>M(SD)</i> | Doktoranden <i>M(SD)</i> |
|--------------------|---------------------------------|--|-----------------------------|
| Gesamtskala | .54 (.09) | .62 (.10) | .78 (.07) |
| Skala 1 (Suchen) | .59 (.11) | .68 (.09) | .79 (.06) |
| Skala 2 (Bewerten) | .47 (.13) | .53 (.13) | .74 (.11) |

Um die Annahmen genauer zu untersuchen, wurde für jede Skala eine Varianzanalyse berechnet. Auf jeder Skala zeigten sich hoch signifikante ($p < .01$) Unterschiede zwischen den Gruppen: $F(2,61) = 43.30$ für die Gesamtskala; $F(2,61) = 27.10$ für Skala 1 und $F(2,61) = 28.29$ für Skala 2. Geplante Kontraste ergaben, dass sich auf der Gesamtskala und Skala 1 alle Gruppen signifikant ($p < .01$, einseitig) unterschieden. Auf Skala 2 erreichten nur die Unterschiede zwischen Studienanfängern und Doktoranden sowie zwischen fortgeschrittenen Studierenden und Doktoranden das Signifikanzniveau ($p < .01$, einseitig). Der Unterschied zwischen Studienanfängern und fortgeschrittenen Studierenden zeigte zwar in die erwartete Richtung, erreichte jedoch das Signifikanzniveau nicht. Diese Ergebnisse wurden bereits in schriftlicher Form veröffentlicht (Leichner, Peter, Mayer, & Krampen, 2013a).

Selbstwirksamkeitsskala

Bei sechs Personen fehlten bei einem oder mehreren Items Werte. Da dies weniger als 10% der Daten betrifft, wurde auf komplexe Verfahren zur Behandlung fehlender Werte, wie z.B. multiple Imputation, verzichtet. Bei den Reliabilitäts- und Faktorenanalysen wurden diese fehlenden Werte mit paarweisem Fallausschluss behandelt.

Zu explorativen Zwecken wurde eine Hauptkomponentenanalyse durchgeführt. Der Scree-Plot wies eindeutig auf einen Faktor hin, der ca. 35% der Varianz aufklärt. Dennoch wurden vier Faktoren extrahiert. Da alle Items aus inhaltlichen Gründen beibehalten werden sollten, wurden keine Items entfernt. Die aus $k = 10$ Items bestehende Skala wies eine interne Konsistenz von $\alpha = .77$ und eine Split-Half-Reliabilität von $r_{tt'} = .77$ auf. Dabei erwies sich Item 5 („Wenn ich zu einem Thema recherchiere, kann ich relativ schnell entscheiden, ob eine Informationsquelle für mich interessant ist oder nicht“) als das leichteste Item mit einem Mittelwert von $M = 4.20$ auf einer fünfstufigen Skala. Item 10 („Verglichen mit anderen habe ich mehr Erfahrung bei der Recherche nach Fachinformationen“) hingegen war mit einem Mit-

telwert von $M = 2.60$ das schwierigste Item. Für die folgenden Analysen wurden die zehn Items gemittelt. Bei fehlenden Werten gingen entsprechend weniger Items in den Mittelwert ein. Deskriptive Statistiken für die drei Probandengruppen und Werte für die Gesamtstichprobe sind in Tabelle 3 abgebildet.

Tabelle 3

Mittelwerte auf der Selbstwirksamkeitsskala nach Studienfortschritt (Vorstudie)

| Gruppe | <i>n</i> | <i>M</i> | <i>SD</i> |
|------------------------------|----------|----------|-----------|
| Studienanfänger | 22 | 2.99 | .51 |
| Fortgeschrittene Studierende | 21 | 3.56 | .56 |
| Doktoranden | 21 | 3.80 | .47 |
| Gesamt | 64 | 3.44 | .61 |

Um die Gruppenunterschiede zu untersuchen, wurde eine einfaktorielle Varianzanalyse berechnet, die signifikante Unterschiede ($F[2,61] = 13.99, p < .01$) anzeigte. Um diese genauer zu untersuchen, wurden geplante Kontraste berechnet. Hierbei zeigten sich signifikante Differenzen zwischen allen Gruppen ($p < .01$, einseitig) – mit Ausnahme der Differenz zwischen fortgeschrittenen Studierenden und Doktoranden. Die Selbstwirksamkeitsskala korrelierte zudem signifikant ($p < .01$, einseitig) mit der Gesamtskala ($r = .54$) sowie den beiden Subskalen des Multiple-Choice-Tests (mit Skala 1 $r = .52$, mit Skala 2 $r = .44$).

11.4 Online-Studie

Die Online-Studie wurde durchgeführt, um auf einer größeren Stichprobe beruhende Kennwerte für die Instrumente, sowie weitere Tests und Fragebögen zu ermitteln. Da sich in der Zwischenzeit Änderungen am Bibliothekskatalog und an anderen Recherchesystemen ergeben hatten, wurde die in der Vorstudie entwickelte aus $k = 22$ Items bestehende Version des Multiple-Choice-Tests um einige Items ergänzt. Die hier eingesetzte Version besteht aus $k = 35$ Items (vgl. Leichner, 2015) und ist im Anhang zu finden. Auch in dieser Studie wurden Gruppenunterschiede im Multiple-Choice-Test und auf der Selbstwirksamkeitsskala erwartet. Zudem sollten Testwerte und Selbstwirksamkeitseinschätzung erneut positiv miteinander korrelieren.

11.4.1 Methode

Für diese Studie wurden Psychologiestudierende der Universität Trier aus allen Studiengängen (BSc-, MSc-, oder Diplomstudiengang) unabhängig von ihrem Fachsemester rekrutiert.

Diese bearbeiteten die Instrumente als elektronische Version an einem Ort ihrer Wahl über die Umfragesoftware *Unipark*, wofür sie bezahlt wurden. Insgesamt beendeten $N = 184$ Teilnehmende die Befragung. In einem ersten Schritt wurden $n = 9$ Fälle aus dem Datensatz entfernt, da die Probanden bereits an der Vorstudie teilgenommen hatten und daher nicht den Einschlusskriterien dieser Studie entsprachen. Infolge deskriptiver Analysen wurden $n = 2$ Fälle entfernt, da die Probanden die Befragung nach Bearbeitung der ersten Hälfte abgebrochen hatten und somit sehr viele fehlende Werte aufwiesen. Die folgenden Analysen beruhen daher auf einer Stichprobe von $N = 173$. Darunter waren $n = 35$ Studienanfänger (BSc 1. und 2. Semester), $n = 106$ fortgeschrittene Studierende (BSc ab 4. Semester) sowie $n = 32$ Studierende aus den MSc- und Diplomstudiengängen; von allen Probanden waren 147 (ca. 85%) weiblichen Geschlechts. Das mittlere Alter der Stichprobe betrug $M = 22.68$ Jahre ($SD = 2.69$).

11.4.2 Ergebnisse

Multiple-Choice-Test

Eine explorativ durchgeführte Hauptkomponentenanalyse ergab nicht eindeutig interpretierbare Ergebnisse. Es wurden 14 Faktoren mit Eigenwerten größer eins extrahiert; der Scree-Plot wies jedoch auf eine Lösung mit drei Faktoren hin. Da die Items anhand inhaltlicher Überlegungen zusammengestellt worden waren, wurden sämtliche 35 Items zu einer Skala kombiniert. Itemanalysen zeigten, dass die Itemschwierigkeiten stark streuten. Beim leichtesten Item wurden im Schnitt über 90% der erreichbaren Punkte erreicht; beim schwierigsten Item im Durchschnitt 10%. Dennoch gab es kein für alle Probanden unlösbares Item. Wie in der Vorstudie wurden wieder Subskalen für die Facetten „Suchen nach Informationen“ und „Bewerten von Informationen“ gebildet. Reliabilitätskennwerte für die Gesamtskala und die beiden Subskalen sind in Tabelle 4 angegeben.

Tabelle 4

Deskriptive Statistiken und psychometrische Kennwerte für die drei Skalen des Multiple-Choice-Tests (Online-Studie)

| Skala | <i>k</i> | <i>M</i> (<i>SD</i>) | α | $r_{tt'}$ |
|--------------------|----------|------------------------|----------|-----------|
| Gesamtskala | 35 | .62 (.06) | .49 | .45 |
| Skala 1 (Suchen) | 22 | .63 (.06) | .23 | .30 |
| Skala 2 (Bewerten) | 13 | .61 (.10) | .48 | .43 |

Anmerkung. *k* = Anzahl der Items; α = Cronbach's Alpha; $r_{tt'}$ = Split-Half-Reliabilität (Spearman-Brown).

Beide Subskalen korrelierten signifikant miteinander ($r = .29, p < .01$, einseitig). Zur Untersuchung von Gruppenunterschieden wurde die Stichprobe in die oben genannten Gruppen eingeteilt. Die Gruppengrößen und die entsprechenden Mittelwerte sind in Tabelle 5 dargestellt.

Tabelle 5

Punktwerte im Multiple-Choice-Test nach Studienfortschritt (Online-Studie)

| Skala | Gruppe 1: BSc 1. - 3. Sem. <i>M</i> (<i>SD</i>) | Gruppe 2: BSc ab 4. Sem. <i>M</i> (<i>SD</i>) | Gruppe 3: MSc-u. Dip- lom <i>M</i> (<i>SD</i>) |
|-------------|--|--|---|
| Gesamtskala | .60(.05) | .62(.06) | .65(.05) |
| Skala 1 | .59(.05) | .63(.06) | .66(.04) |
| Skala 2 | .61(.09) | .60(.09) | .64(.10) |

Erwartungskonform stieg der Testwert mit dem Studienfortschritt an. Separate einfaktorielle Varianzanalysen ergaben signifikante Differenzen zwischen den drei Gruppen auf der Gesamtskala ($F[2,170] = 8.63, p < .01$) sowie Skala 1 ($F[2,170] = 14.15, p < .01$). Auf Skala 2 zeigte die Varianzanalyse keine signifikante Differenz zwischen den Gruppen ($F[2,170] = 2.20, ns$). Um die Gruppenunterschiede genauer zu analysieren, wurden geplante Kontraste berechnet. Diese zeigten signifikante Unterschiede ($p < .05$, einseitig) auf der Gesamtskala und Skala 1 zwischen allen drei Gruppen. Auf Skala 2 zeigte sich nur zwischen Gruppen 2 und 3 eine signifikante Differenz.

Selbstwirksamkeitsskala

Fehlende Werte wurden auch hier wieder bei den Reliabilitäts- und Faktorenanalysen mit paarweisem Fallausschluss behandelt. Der Scree-Plot einer Hauptkomponentenanalyse ließ sich nicht zweifelsfrei interpretieren. Da keine mehrdimensionale Struktur erwartet worden war und es sich lediglich um $k = 10$ Items handelte, wurden diese wieder durch Bildung des

Mittelwerts zu einer Skala kombiniert, welche eine interne Konsistenz von $\alpha = .77$ und eine Split-Half-Reliabilität von $r_{tt'} = .76$ aufwies. Deskriptive Statistiken für die einzelnen Gruppen sind in Tabelle 6 wiedergegeben.

Tabelle 6

Mittelwerte der Selbstwirksamkeitsskala nach Studienfortschritt (Online-Studie)

| Gruppe | <i>n</i> | <i>M</i> | <i>SD</i> |
|------------------|----------|----------|-----------|
| 1 BSc 1.-3. Sem. | 35 | 2.90 | .72 |
| 2 BSc ab 4. Sem. | 106 | 2.95 | .67 |
| 3 MSc- u. Diplom | 32 | 3.22 | .59 |
| Gesamt | 173 | 2.99 | .67 |

Eine Varianzanalyse signalisierte marginal signifikante Unterschiede zwischen den Probandengruppen ($F[2,170] = 2.48, p < .10$). Geplante Kontraste zeigten signifikante ($p < .05$, einseitig) Unterschiede zwischen Gruppen 1 und 3 sowie zwischen Gruppen 2 und 3. Lediglich die Differenz zwischen Gruppen 1 und 2 verfehlte das Signifikanzniveau. Erwartungskonform korrelierte die Selbstwirksamkeitsskala mit der Gesamtskala des Multiple-Choice-Tests ($r = .16$), sowie mit Skala 1 ($r = .16$, beide $p < .05$, einseitig). Die Korrelation mit Skala 2 erreichte das Signifikanzniveau nicht ($r = .10, ns$).

11.5 Diskussion

Bezüglich des Multiple-Choice-Tests lässt sich zusammenfassen, dass hier in zwei Schritten ein verwendbarer Test zur Erfassung deklarativer und weniger komplexer prozeduraler Anteile von Informationskompetenz erstellt wurde. Die Vorstudie diente unter anderem dazu, erste Erfahrungen mit der Formulierung von Items zum Themengebiet Informationskompetenz zu sammeln. So erwies sich beispielsweise ein Item als zu leicht. Die Konstruktion der Items beruhte auf den Informationskompetenz-Standards der ACRL und auf eigenen Überlegungen. In Fachbegriffen ausgedrückt wurden hier also eine rationale und eine intuitive Konstruktionsmethode verwendet (Jankisz & Moosbrugger, 2008, pp. 36–37). Bei der rationalen Konstruktionsmethode wird auf vorliegende Theorien zurückgegriffen. Etliche Items, beispielsweise über die Verwendung von Fachdatenbanken, beruhen auf den psychologiespezifischen Informationskompetenzstandards der ACRL (2010). Jedoch beinhalten die Standards nicht alle relevant erscheinenden Facetten von fachlicher Informationskompetenz und keine deutschen landesspezifischen Inhalte. Daher erschien eine rationale Konstruktion – aus-

schließlich basierend auf diesen Standards – als nicht ausreichend, sodass weitere Items – basierend auf eigenen Erfahrungen und Überlegungen – hinzugefügt wurden. Dies wird auch als intuitive Konstruktionsmethode bezeichnet. Ein solches Verfahren kann angewendet werden, wenn das jeweilige Forschungsfeld im Entstehen begriffen ist, wie das bei der Erfassung von Informationskompetenz im akademischen Kontext der Fall ist (vgl. Pinto, 2015). Bei solchen Forschungsfeldern liegen gewöhnlich nicht hinreichend ausformulierte Theorien vor, die als (ausschließliche) Grundlage für die Konstruktion von Testverfahren nach der rationalen Methode verwendet werden können. Die Vermischung mehrerer Konstruktionsprinzipien bei der Testentwicklung ist ein übliches Vorgehen; so schreiben Jankisz und Moosbrugger (2008, p. 36), dass die Testkonstruktion selten ausschließlich einer Konstruktionsmethode folgt. Mit der Online-Studie konnten die Instrumente in einem etwas anderen Kontext getestet werden, wobei Daten anhand einer wesentlich größeren Strichprobe erhoben wurden.

Zur Bestimmung der Reliabilität wurden Split-Half-Reliabilität und interne Konsistenz berechnet. Dabei zeigten sich für die Gesamtskala in beiden Studien akzeptable Werte. Zur Berechnung der Split-Half-Reliabilität wird der Test in zwei möglichst äquivalenten Hälften geteilt, welche miteinander korreliert werden, wobei die Korrelation auf die Gesamtlänge des Tests hochgerechnet werden muss. Eine hohe Split-Half-Reliabilität bestätigt, dass mit beiden Testhälften das gleiche (oder zumindest etwas Ähnliches) erfasst wird. Zur Berechnung der internen Konsistenz wird der Test nicht in zwei Hälften geteilt, sondern in genau so viele Teile, wie er Items enthält. Beide Kennwerte steigen und fallen mit der Homogenität der Testhälften, bzw. Testitems.

Sicherlich liegen die Kennwerte in der Online-Studie im niedrigen Bereich (vgl. Fisseni, 1997, p. 124), wenn man die strengen Kriterien für Individualdiagnostik anlegt. Allerdings können Reliabilitätskennwerte durchaus niedriger sein, wenn der Test lediglich zur Gruppendiagnostik genutzt wird, da Gruppenmittelwerte auch bei stärker mit Messfehler behafteten individuellen Werten interpretiert werden können (Schermelleh-Engel & Werner, 2008, pp. 129–130). Auch hängt die realistischere zu erwartende Reliabilität auch vom Forschungsfeld ab (Schermelleh-Engel & Werner, 2008, p. 129). So ist Informationskompetenz, wie bereits mehrfach dargelegt, per se ein heterogenes Konzept (Peter, 2015) und beinhaltet nicht nur technisches Wissen über Informationssysteme, wie z.B. Fachdatenbanken, sondern auch

Hintergrundwissen zu Publikationstypen und zum Wissenschaftssystem im Fach Psychologie. Zudem scheint es, dass sich Studierende ihr Wissen zum Thema Informationskompetenz vor allem selbstständig nach dem Versuch-und-Irrtum-Prinzip aneignen (Klatt, Gavriilidis, Kleinsimlinghaus, & Feldmann, 2001). Entsprechend ist auf Seiten der Studierenden keine umfassende und kohärente Wissensbasis zu erwarten, was ebenfalls das Erreichen einer hohen internen Konsistenz erschweren dürfte (vgl. Schneider, Rittle-Johnson, & Star, 2011). Sicherlich wäre es möglich gewesen, den Test inhaltlich homogener zu gestalten. So hätte man beispielsweise ausschließlich nach Funktionen von Fachdatenbanken fragen können, allerdings hätte man dann einen Test mit fragwürdiger Validität entwickelt, da nur ein kleiner Ausschnitt von Informationskompetenz erfasst wird. So argumentieren Cattell und Kline (1977 as cited in Kline, 1998, p. 9), dass sich eine sehr hohe interne Konsistenz und Validität nahezu ausschließen. Inhaltlich heterogene Tests mit niedriger interner Konsistenz sind jedoch, insbesondere in anwendungsnahen Kontexten, nichts Ungewöhnliches und können durchaus Vorhersage-Validität besitzen (Schermelleh-Engel & Werner, 2008, p. 127). So zeigen sich im vorliegenden Fall signifikante Unterschiede zwischen im Studium mehr oder weniger fortgeschrittenen Studierendengruppen, was für die Validität des Tests spricht.

Dennoch sollte an dieser Stelle die Frage diskutiert werden, warum die interne Konsistenz (und entsprechend auch die Split-Half-Reliabilität) in der Online-Studie wesentlich geringer waren als in der Vorstudie. Durch Analysen konnte ausgeschlossen werden, dass die neu hinzugekommenen Items für die niedrige interne Konsistenz verantwortlich waren. Werden nur die auch in der Vorstudie eingesetzten Items ausgewertet, ist die interne Konsistenz nicht wesentlich höher als bei Einbezug aller Items. So bleiben zwei Unterschiede zwischen den beiden Studien als mögliche Ursachen: Die homogenere Stichprobe in der Online-Studie und die Erhebungsmodalität. Die Stichprobe der Vorstudie ist durch den Einbezug der Doktoranden wesentlich leistungsheterogener als die Stichprobe der Online-Studie, was die Varianz erhöht (vgl. Tabelle 1 und Tabelle 4). Varianz ist allerdings eine Voraussetzung für Korrelationen zwischen den Items – und damit für eine hohe interne Konsistenz. Um dieser Vermutung nachzugehen, wurde die interne Konsistenz anhand der Vorstudien Daten für jede Probandengruppe separat bestimmt. Hierbei zeigten sich interne Konsistenzwerte um $\alpha = .60$ für die beiden studentischen Stichproben und eine interne Konsistenz von $\alpha = .41$ für die leistungsstärkere und vermutlich auch leistungshomogenere Gruppe der Doktoranden.

Die Werte für die einzelnen Gruppen liegen also deutlich unter der für die Gesamtstichprobe bestimmten internen Konsistenz von $\alpha = .82$. Diese Befunde sprechen dafür, dass ein Teil der hohen internen Konsistenz, welche in der Vorstudie gefunden wurde, auf die Heterogenität der Stichprobe zurückzuführen ist. Die zweite Ursache, Erhebungsmodalität, kann insofern eine Erklärung sein, als dass die Probanden den Test in der Online-Studie unter unkontrollierten Bedingungen an einem Ort ihrer Wahl bearbeiteten. Möglicherweise waren sie dabei unkonzentrierter, was den Messfehler erhöht haben kann. Ein erhöhter Messfehler wiederum kann zu schlechteren psychometrischen Kennwerten führen (vgl. Leichner, Peter, Mayer, & Krampen, 2013b).

Die Ergebnisse der Hauptkomponentenanalysen beider Studien zeigten, dass sich die beiden von den Informationskompetenz-Standards postulierten Subskalen nicht empirisch trennen lassen. Dennoch verhalten sich beide Subskalen unterschiedlich. So waren die Gruppenunterschiede auf Skala 2 (Bewerten) schwächer ausgeprägt, als auf Skala 1 (Suchen). Da die Standards nicht empirisch entwickelt worden sind, ist es auch nur begrenzt erstaunlich, dass sich keine empirische Unterscheidung zwischen den beiden Standards zeigt. Eine Einteilung in verschiedene inhaltlich relativ homogene Standards ist sicherlich zum Zwecke der Lehre sinnvoll. So können separate Trainings für die einzelnen Standards angeboten werden oder die Lehre bezüglich einzelner Standards auf verschiedene Institutionen aufgeteilt werden (vgl. Homann, 2000). Dieser Befund wirft jedoch die Frage auf, ob eine Trennung der beiden Standards bei der Erfassung von Informationskompetenz sinnvoll ist. Da sich für die Gesamtskala eine akzeptable interne Konsistenz und eine zufriedenstellende Split-Half-Reliabilität zeigte, während sich insbesondere Skala 1 in der Online-Studie als intern sehr inkonsistent erwies, wird in Folge auf eine separate Berechnung der beiden Subskalen verzichtet.

Einschränkend soll an dieser Stelle darauf hingewiesen werden, dass der hier vorliegende Test bislang lediglich bei Psychologiestudierenden eingesetzt worden ist. Da der Test etliche auf das Fach Psychologie bezogene Items enthält, ist davon auszugehen, dass er allenfalls bei Studierenden benachbarter Fächer durchgeführt werden kann. Ob er bei berufstätigen Psychologen verwendet werden kann, erscheint ebenfalls fraglich, da diesen etliche Informationssysteme, wie z.B. Fachdatenbanken, oftmals nicht zur Verfügung stehen und diese Personengruppen daher nicht die Möglichkeit haben, hierin Erfahrungen zu sammeln. Für die

Zwecke der vorliegenden Untersuchung erscheint dies jedoch unproblematisch, da der Test primär zur Evaluation eines Informationskompetenz-Trainings für Psychologiestudierende eingesetzt werden soll. Zuletzt soll noch besprochen werden, dass das verwendete "pick any out of n"-Konzept mit dem Nachteil verbunden ist, dass möglicherweise auch für nicht bearbeitete Items Teilpunkte vergeben werden. Im Nachhinein ist nicht mehr zu unterscheiden, ob ein Item nicht bearbeitet wurde oder ob der Proband sämtliche Antwortoptionen für verkehrt hielt und daher keine gewählt hat. Sofern es richtig war, eine oder mehrere Antwortoptionen auszulassen, erhält der Proband entsprechend Teilpunkte, obwohl er das Item nicht bearbeitet hat.

Auch die Konstruktion der Selbstwirksamkeitsskala scheint gelungen. Wie erwartet ist die Skala homogen: Die interne Konsistenz lag in beiden Studien über $\alpha = .70$, was vielfach als Untergrenze für verlässliche Diagnostik gesehen wird (Fisseni, 1997; Schmitt, 1996). Zudem zeigten sich erwartungskonforme Gruppenunterschiede und Zusammenhänge mit den Skalen des Multiple-Choice-Tests. Über die Kausalrichtung des Zusammenhangs zwischen Selbstwirksamkeit und Testwerten kann hier jedoch keine Aussage getroffen werden. Beim Vergleich der Studien zeigt sich, dass die Zusammenhänge und Gruppenunterschiede in der Online-Studie wesentlich schwächer ausgefallen sind als in der Vorstudie. Auch hierfür können die homogene Stichprobe und die unkontrollierten Bedingungen in der Online-Studie als Gründe angeführt werden.

12 Entwicklung der Rechercheaufgaben

Ziel dieser Studie war die Konstruktion von Rechercheaufgaben, mit denen die prozeduralen Komponenten fachlicher Informationskompetenz erfasst werden können. Es wurden drei Typen von Aufgaben entwickelt, welche nach ihrer antizipierten Schwierigkeit geordnet werden können. Damit besteht die Möglichkeit, durch die Aufgaben mit unterschiedlicher Schwierigkeit zwischen Personen mit unterschiedlich ausgeprägten Recherchefähigkeiten zu unterscheiden. Um die Auswertung standardisiert zu gestalten, wurden Auswertungsschlüssel entwickelt.

Vor der Gestaltung der Aufgaben wurde eine umfangreiche Literatursuche durchgeführt, um möglicherweise bereits erprobte Aufgaben als Vorbild verwenden zu können. Rechercheaufgaben wurden bereits mehrfach eingesetzt. So konzipierte Kim (2009) drei Typen von Rechercheaufgaben und untersuchte, inwieweit Aufgabentypen die verwendeten Suchstrategien beeinflussen. Eine andere Studie untersuchte, wie gut die Teilnehmer ein Poster identifizieren konnten, das in einem bestimmten Poster-Wettbewerb der UNESCO erfolgreich gewesen war (Scott & O'Sullivan, 2005). In der Praxis bereits durchgeführte Rechercheaufgaben waren jedoch meist mit dem Ziel entwickelt worden, Informationsverhalten zu beschreiben und nicht zu bewerten. Zudem handelte es sich meist um Webrecherchen – und nicht um fachliche Recherchen. Da für Webrecherchen andere Suchwerkzeuge zum Einsatz kommen, können diese Aufgaben nicht für die vorliegenden Zwecke eingesetzt werden. Auch in solchen Studien, die eine Messung von Informationskompetenz im akademischen Kontext mittels Rechercheaufgaben versprechen, erfolgt oftmals lediglich eine Beschreibung des Informationsverhaltens (Dunn, 2002; Novotny & Cahoy, 2006). Indizien, mit denen die Schwierigkeit bestimmter Aufgaben(typen) ermittelt werden kann, sind in diesen Studien nicht genannt. Die Erstellung von Aufgaben unterschiedlicher Schwierigkeit ist jedoch Voraussetzung für die Messung von Informationskompetenz mittels Rechercheaufgaben.

Im ersten Schritt wurde eine Aufgabentaxonomie mit drei Aufgabentypen unterschiedlicher antizipierter Schwierigkeit entwickelt. Die antizipierte Schwierigkeit basierte auf den für die Aufgabenlösung erforderlichen Fähigkeiten. Fähigkeiten wurden hier als Wissen über ein Suchwerkzeug (z.B. Google Scholar oder eine Fachdatenbank), und die für die Benutzung notwendigen Kenntnisse, verstanden. An der Universität Trier muss auf die für das Fach Psy-

chologie relevanten Fachdatenbanken über den Datenbankanbieter OvidSP zugegriffen werden. Entsprechend ist die Beherrschung der Suchoberfläche von OvidSP für die Aufgabelösung mittels Fachdatenbanken notwendig. Die erforderlichen Fähigkeiten wurden mit Rückgriff auf die psychologiespezifischen Informationskompetenzstandards (ACRL, 2010) definiert. Die Aufgaben folgen dem Prinzip additiver Fähigkeiten, d.h. die für eine Aufgabe erforderlichen Fähigkeiten sind auch für die Bearbeitung der nachfolgenden Aufgabe vonnöten, welche jedoch weitere Fähigkeiten voraussetzen. So ist es möglich, drei Aufgabentypen mit aufsteigender Schwierigkeit zu entwickeln. Die Aufgabentaxonomie ist in Tabelle 7 dargestellt (vgl. Leichner et al., 2014a); bei Aufgabentypen 2 und 3 sind lediglich die Fähigkeiten angegeben, die zusätzlich zu den für den vorherigen Aufgabentyp erforderlichen notwendig sind.

Tabelle 7

Taxonomie der Rechercheaufgaben

| Aufgabentyp | Beschreibung und Beispielaufgabe | Empfohlenes Vorgehen und erforderliche Fähigkeiten |
|-------------------------------------|---|--|
| Typ 1 (niedrigste Schwierigkeit) | Zwei Publikationen zu einem Fachbegriff, die nach einem bestimmten Jahr publiziert worden sind, sollen gefunden werden. Suchbegriff ist vorgegeben. Beispiel: Finden Sie zwei nach 2005 publizierte wissenschaftliche Artikel, die sich mit falschen Erinnerungen beschäftigen. Bitte verwenden Sie den englischsprachigen Suchbegriff "false memories". | Eingabe des Suchbegriffs in Google Scholar und Einschränkung des Publikationsdatums über entsprechende Filterfunktion. Fähigkeiten: Entwickeln einer einfachen Suchstrategie; Verwendung des angemessenen Suchwerkzeugs; Verwendung einfacher Einschränkungen. |
| Typ 2 (mittlere Schwierigkeit) | Zwei Publikationen sollen gefunden werden, die sich mit einem Thema befassen, das durch zwei vorgegebene Fachbegriffe definiert ist. Die Publikationen müssen bestimmte Anforderungen erfüllen (Verwendung einer bestimmten Methodik [z.B. Meta-Analyse] und bestimmter Publikationszeitraum). Beispiel: Gibt es nach 2005 publizierte Längsschnittstudien ("longitudinal study"), die Risikofaktoren ("risk factors") für die generalisierte Angststörung ("Generalized Anxiety Disorder") untersuchen? Bitte geben Sie – wenn möglich - zwei entsprechende Publikationen an. | Idealerweise werden beide Suchbegriffe in die Suchmaske von OvidSP eingegeben und als Schlagwörter erkannt. Anschließend Verknüpfung der Begriffe mit AND und Verwendung der Einschränkungen für Methodik und Alter. Auch mit Google Scholar lösbar. Fähigkeiten: Verständnis und Nutzung der Schlagwortsuche; Verständnis und Nutzung von Booleschen Operatoren; Nutzung spezifischer Funktionen, z.B. Einschränkungen in OvidSP oder Wortgruppensuche in Google Scholar. |
| Typ 3 (hohe Schwierigkeit) | Suche nach Publikationen, welche sich mit einem Thema beschäftigen, das durch nichtwissenschaftliche Begriffe beschrieben wird. Beispiel: Suchen Sie zwei Publikationen, die sich mit der Frage beschäftigen, ob Altersunterschiede Einfluss auf das Therapieergebnis bei der Behandlung von Depressionen besitzen. | Identifikation von Suchbegriffen über Thesaurussuche oder durch mehrere Suchanfragen in Google Scholar. Dann Recherche mit Google Scholar oder Fachdatenbank. Fähigkeiten: Entwicklung einer mehrschrittigen Suchstrategie; Beherrschung fortgeschrittener Funktionen von Fachdatenbanken, z.B. Thesaurussuche, oder Beherrschung eines anderen Vorgehens zur Identifikation von Suchbegriffen. |

Für die Auswertung wurden zwei Auswertungsschlüssel entwickelt: (a) Für das Ergebnis der Recherche (Wurden Publikationen gefunden, die den Anforderungen entsprachen?) und (b) für das Vorgehen während der Recherche (Prozess) – gewissermaßen wird hiermit also das Informationsverhalten bewertet. Es erschien ratsam, sich nicht auf das Auswerten des Rechercheergebnisses zu beschränken, da aus einem guten Ergebnis der Recherche nicht zwangsläufig auf ein angemessenes Vorgehen geschlossen werden kann. Es besteht die Möglichkeit, dass Probanden aufgrund eines Zufallstreffers ein hohes Niveau von Informationskompetenz zugeschrieben wird. Bei Typ 2-Aufgaben ist es beispielsweise möglich, dass ein Proband beide Suchbegriffe in die Suchmaske von Google Scholar zusammen mit der Methodik, z.B. "meta-analysis", eingibt und eine oder möglicherweise zwei relevante Publikationen findet, obwohl ein angemessenes Informationsverhalten hier die Verwendung einer Fachdatenbank und entsprechender Filterfunktionen beinhalten würde. Da aus Zeitgründen nur eine Rechercheaufgabe von jedem Typ eingesetzt werden kann, erscheint diese Möglichkeit durchaus plausibel. Durch den Einsatz von Auswertungsschlüsseln kann wesentlich konsistenter und reliabler bewertet werden als ohne festgelegte Auswertung. Prinzipiell kann man Auswertungsschlüssel nach den Evaluationskriterien und nach ihrer Spezifität unterscheiden. Evaluationskriterien können sich auf das gesamte Produkt (holistischer Schlüssel) oder auf einzelne Facetten (analytischer Schlüssel) beziehen. Spezifische Auswertungsschlüssel können lediglich für einen bestimmten Aufgabentyp verwendet werden, während allgemein gehaltene Auswertungsschlüssel für die Auswertung einer Reihe von Aufgaben genutzt werden können. Ein Literaturüberblick hat gezeigt, dass die Reliabilität der Auswertung von offenen Aufgaben wesentlich verbessert werden kann, wenn analytische Auswertungsschlüssel eingesetzt werden, welche auf einzelne Aufgaben zugeschnitten sind (spezifischer Auswertungsschlüssel) (Jonsson & Svingby, 2007). Daher wurden hier für jeden Aufgabentyp spezifische und analytische Auswertungsschlüssel für Ergebnis und Prozess erstellt.

Der Auswertungsschlüssel für das Rechercheergebnis sah vor, dass für jedes in der Aufgabenbeschreibung genannte Kriterium (Thematik der Publikation, verwendete Methodik, Alter) Teilpunkte vergeben werden. Dem Auswertungsschlüssel für den Prozess zufolge werden Teilpunkte für eine angemessene Wahl des Informationssystems, für die Verwendung der angemessenen Funktionen des Informationssystems (Filter, Verknüpfung mittels Boolescher Operatoren) und für die Verwendung der Thesaurussuche (nur Typ 3-Aufgaben) ver-

geben. Jedes dieser Kriterien wurde separat bewertet. Bei einer Typ 2-Aufgabe konnte die maximale Punktzahl erreicht werden, wenn eine Fachdatenbank verwendet wurde, wobei Boolesche Operatoren zur Verknüpfung der beiden Suchbegriffe und die entsprechenden Filterfunktionen zur Einschränkung der Treffermenge genutzt wurden. Damit folgt dieser Auswertungsschlüssel den in den Informationskompetenzstandards der ACRL (2010) als informationskompetent definierten Vorgehensweisen. Beide Auswertungsschlüssel sind im Anhang zu finden.

Trotz aller Bemühungen, die Datenerhebungen durch Verwendung einer Aufgabentaxonomie und von Auswertungsschlüsseln zu standardisieren, muss bei diesen Aufgabentypen mit einem relativ hohen Messfehler gerechnet werden, der zum Teil auf durch nicht kontrollierbare Faktoren zurückgeführt werden kann. So ist den Studierenden prinzipiell freigestellt, welche Suchwerkzeuge sie für die Recherche verwenden, auch wenn zum Zeitpunkt der Studiendurchführung lediglich Google Scholar und die beiden Fachdatenbanken PsycINFO und PSYINDEX, über die via OvidSP zugegriffen werden kann, als geeignete Recherchewerkzeuge in Frage kamen. Dabei ist es durchaus möglich, dass ein eigentlich relativ informationskompetenter Proband an der unübersichtlichen Suchoberfläche von OvidSP scheitert. Umgekehrt ist es bei Verwendung von Google Scholar auch ohne weitgehende Kenntnisse über Informationssysteme möglich, geeignete Treffer zu erzielen, da die Suchoberfläche intuitiv zuverlässig zu bedienen ist. Sicherlich ist Google Scholar nicht für umfassende Recherchen geeignet (vgl. Abschnitt 5.4, S. 26); für die Bearbeitung einer Rechercheaufgabe, bei der nur zwei Publikationen gefunden werden müssen, kann sich dieses Suchwerkzeug jedoch als hilfreich erweisen. Die oben beschriebenen Szenarien würden nicht nur zu einem relativ hohen Messfehler, sondern auch zu einer Divergenz von Ergebnis- und Prozesspunktwerten führen, da für die Verwendung von Fachdatenbanken bei Aufgabentypen 2 und 3 mehr Prozesspunkte vergeben werden, als für die Verwendung von Google Scholar. Ein weiteres Problem ist die Abhängigkeit der einzelnen Schritte. Wer beispielsweise unpassende Suchbegriffe wählt, wird auch dann keine passenden Treffer erzielen, wenn er sämtliche relevanten Funktionen des verwendeten Suchwerkzeugs beherrscht. Sicherlich hätte man diesem Problem durch stärkere Standardisierung entgegenwirken können; beispielsweise hätte man die Instruktion geben können, ausschließlich Fachdatenbanken für die Recherche zu verwenden. Damit wäre allerdings in gewissem Maß eine künstliche Situation geschaffen worden, da für fachliche

Recherchen in der realen Welt eben nicht nur Fachdatenbanken zur Verfügung stehen. Da einer der Gründe für den Einsatz von Rechercheaufgaben die Forderung nach Leistungsmessung mithilfe von realitätsnahen Aufgaben (z.B. Shavelson, 2010) war, kam diese Option nicht in Frage. Eine weitere Möglichkeit wäre der Einsatz von Simulationsumgebungen gewesen, wie sie in anderen Studien zum Einsatz kamen (Goldhammer et al., 2014; Katz, 2007). Solche Aufgaben sind jedoch nicht nur extrem aufwendig und kostspielig in der Erstellung (Sireci & Zenisky, 2006), sondern erfordern zudem die Zerlegung des Rechercheprozesses in einzelne Schritte, für die jeweils eine Simulationsaufgabe erstellt würde. Auch dies würde den formulierten Anforderungen an eine realitätsnahe Erfassung nicht gerecht.

13 Multimethodale Trainingsevaluation

Im Folgenden wird über die Evaluation eines Blended Learning-Trainings für psychologiespezifische fachliche Informationskompetenz berichtet, das an der Universität Trier durchgeführt wurde. Zur Evaluation kamen die Taxonomie von Rechercheaufgaben und der Multiple-Choice-Test (MC-Test) zum Einsatz. Auch die Selbstwirksamkeitsskala wurde von den Probanden bearbeitet.

Das Training richtete sich an Studierende im BSc-Studiengang Psychologie an der Universität Trier. Die Trainingsinhalte orientierten sich an den Psychologie-Informationskompetenzstandards der ACRL (2010) und waren an diese Zielgruppe angepasst. Sie bestanden aus grundlegenden Informationen über Literaturrecherchen für Psychologiestudierende: Zunächst wurden die im Fach Psychologie wichtigsten Publikationstypen (Monographien, Sammelwerke, Zeitschriftenartikel) vorgestellt. Nach einer Erläuterung der Funktionsweise Boolescher Operatoren folgte eine Einführung in die für die Psychologie relevanten Fachdatenbanken (v.a. PsycINFO und PSYINDEX) und deren Funktionsumfang. Hier wurden insbesondere die Thesaurussuche und die erweiterten Einschränkungen beschrieben, mit denen beispielsweise nur solche Publikationen angezeigt werden können, bei denen es sich um Überblicksarbeiten oder Meta-Analysen handelt. Darüber hinaus wurde Google Scholar behandelt, da diese Ressource für Recherchen im Fach Psychologie als ergänzende Quelle genutzt werden kann (ACRL, 2010). Zudem wurden die Angebote der lokalen Universitätsbibliothek, wie der elektronische Bibliothekskatalog oder die Fernleihe, beschrieben. Ein weiterer Abschnitt der Trainingsmaterialien behandelte Kriterien zur Bewertung von Informationen, wie z.B. "Journal Impact Factor". Das Training deckte insgesamt Standards 1 bis 3 der ACRL-Definition ab. Standard 4, also die Weiterverarbeitung von Informationen, wurde nicht behandelt, da es sich hier um einen Inhalt handelt, der an deutschen Universitäten üblicherweise in Veranstaltungen zum wissenschaftlichen Arbeiten gelehrt wird (Homann, 2007). Das Training dauerte zwei Wochen. In dieser Zeit bearbeiteten die Teilnehmenden die Online-Materialien (erwartete Bearbeitungszeit ca. fünf Stunden) und nahmen an zwei Präsenzseminaren (je 90 Min.) teil. Es wurde ein Wartekontrollgruppendesign eingesetzt, bei dem die Kontrollgruppe zeitlich versetzt am Training teilnimmt, um die Effekte durch einen Kontrollgruppenvergleich absichern zu können. Gleichzeitig erlaubt dieses Design, sämtliche Teilnehmer für die Be-

rechnung eines Trainingseffekts heranzuziehen. Eine genauere Beschreibung der Trainingsinhalte und Trainingsgestaltung findet man bei Leichner, Peter, Mayer und Krampen (2014b).

13.1 Methode

Die Gesamtdauer der Evaluationsstudie betrug vier Wochen; in dieser Zeit fanden drei Datenerhebungen (t1 bis t3) statt. Die erste Erhebung (t1) fand zu Beginn der Studie statt. Anschließend nahm Gruppe 1 am zweiwöchigen Training teil, während Gruppe 2 als Wartekontrollgruppe fungierte. Es folgte Erhebung t2 und darauf die Trainingsteilnahme von Gruppe 2. Die dritte Erhebung (t3) fand statt, nachdem beide Gruppen das Training absolviert hatten.

Die Stichprobe bestand aus $N = 71$ Psychologiestudierenden, die im 2. oder 4. Fachsemester des BSc-Studiengangs studierten. Die Probanden wurden für die Teilnahme an den Datenerhebungen bezahlt. Im Laufe des Trainings brachen drei Teilnehmende das Training aus persönlichen Gründen ab, eine Person wurde aufgrund eines extremen Altersunterschieds zu den übrigen Probanden aus dem Datensatz entfernt, sodass für die Auswertung eine Gesamtstichprobe von $N = 67$ Probanden zur Verfügung stand. Im Mittel waren die Teilnehmenden $M = 21.67$ ($SD = 2.38$) Jahre alt. Zur Umsetzung des Wartekontrollgruppendesigns waren die Probanden randomisiert auf zwei Gruppen ($n = 37$ und $n = 30$) aufgeteilt worden. Um die Gruppengrößen für die Datenerhebungen und die Präsenzseminare hinreichend klein zu halten, wurden diese Gruppen noch einmal geteilt.

Die Aufgabentaxonomie wurde verwendet, um mehrere Aufgaben von jedem Typ zu erstellen. Um Testeffekte (die Verbesserung der Leistung durch die mehrfache Bearbeitung einer Aufgabe- oder eines Aufgabentyps, vgl. Hausknecht, Halpert, Di Paolo, & Moriarty Gerrard, 2007) soweit wie möglich zu reduzieren, bearbeiteten die Probanden zu jedem Messzeitpunkt jeweils eine andere Rechercheaufgabe vom selben Typ.

Sämtliche Datenerhebungen fanden in einem Computerraum der Universität Trier als Gruppenversuch statt. Alle Aufgaben und Fragebögen sowie Tests wurden mithilfe der Umfragesoftware *Unipark* präsentiert, die auch die Antworten registrierte. Dabei bearbeiteten die Probanden drei Rechercheaufgaben (eine von jedem Typ, aufsteigend gereiht nach ihrer Schwierigkeit), zwei Fragebögen zu epistemologischen Überzeugungen, den aus $k = 35$ Items bestehenden Multiple-Choice-Test sowie die Selbstwirksamkeitsskala. Für die Bearbeitung

der Rechercheaufgaben konnten die Probanden sämtliche an den Computern verfügbaren Ressourcen (Internetzugang sowie Zugang zu den von der Universität Trier lizenzierten Fachdatenbanken und Zugriff auf den elektronischen Bibliothekskatalog) nutzen. Nach der Bearbeitung jeder Rechercheaufgabe wurden einige Fragen zum Vorgehen (bspw. nach den verwendeten Informationssystemen und Suchbegriffen) gestellt, welche für die Auswertung des Vorgehens während der Recherche genutzt wurden. Die Bearbeitungszeit für jede Aufgabe war begrenzt, um die Erhebung so standardisiert wie möglich zu gestalten.

13.1.1 Hypothesen

1. Bezüglich der Rechercheaufgaben wurde erwartet, dass sich die Aufgaben in ihrer Schwierigkeit unterscheiden, was sich in den mittleren erreichten Punktwerten widerspiegeln sollte. Konkret wurde erwartet, dass Typ 1-Aufgaben am leichtesten sind, gefolgt von Typ-2- und Typ-3-Aufgaben.
2. Die zweite Hypothese bezieht sich auf den Zusammenhang der Punktwerte für Ergebnis und Prozess während der Recherche. Diesbezüglich wurde erwartet, dass die beiden Punktwerte signifikant, aber im moderaten Bereich (vgl. Bühner & Ziegler, 2012), miteinander korrelieren und dass die Korrelation $r = .50$ nicht überschreitet. Hier wurden mittlere Korrelationen erwartet, da es zum einen zu Zufallstreffern kommen kann, d.h. Probanden finden ohne adäquate Suchstrategie geeignete Publikationen. Zum anderen ist es auch möglich, dass trotz überwiegend informationskompetenten Verhaltens keine passenden Publikationen gefunden werden, da, wie oben beschrieben, die einzelnen Suchschritte voneinander abhängig sind und ein unpassender Schritt (z.B. Wahl eines ungeeigneten Suchbegriffs) das Ergebnis wesentlich verschlechtern kann.
3. Die dritte Hypothese bezieht sich auf den Zusammenhang von Rechercheaufgaben und Multiple-Choice-Test. Hier wurde ebenfalls eine moderate Korrelation angenommen, da mit den Instrumenten unterschiedliche Komponenten von Informationskompetenz (prozedurale und deklarative Aspekte) erfasst werden. Zudem behandelten einige Fragen des Multiple-Choice-Tests solche Aspekte von Informationskompetenz, welche für die Bearbeitung der Rechercheaufgaben nicht relevant waren (bspw. den Fernleihdienst der Universitätsbibliothek).

4. Weiterhin wurde damit gerechnet, dass sich die Trainingsteilnahme positiv auf die Punktwerte im Multiple-Choice-Test und in den Rechercheaufgaben auswirkt. Dabei sollte Gruppe 1 während t2 höhere Werte erreichen als Gruppe 2, da zu diesem Zeitpunkt nur Gruppe 1 am Training teilgenommen haben würde. Bei t3 sollte Gruppe 2 „aufgeholt“ haben und es sollten sich keine Unterschiede zwischen den Gruppen zeigen. Zu explorativen Zwecken sollte zudem eine Clusteranalyse durchgeführt werden, um zu prüfen, ob sich Gruppen mit unterschiedlichen Fähigkeitsprofilen identifizieren lassen. Es erscheint plausibel, dass einige Probanden gute Leistungen im MC-Test erreichen, jedoch bei der Bearbeitung der Rechercheaufgaben Probleme haben, während andere Probanden auf sämtlichen Variablen relativ hohe Punktwerte erreichen.
5. Die letzte Hypothese behandelt motivationale oder persönlichkeits-theoretische Komponenten, nämlich den Zusammenhang von Selbstwirksamkeit bei der Recherche und tatsächlicher Leistung. Hier wurden signifikante Korrelationen zwischen der Selbstwirksamkeitsskala und den Maßen für Informationskompetenz erwartet. Da hier messwiederholte Daten vorliegen, kann explorativ der Frage nach der Wirkrichtung der Selbstwirksamkeitseinschätzung auf die Leistungsmaße nachgegangen werden.

13.2 Ergebnisse

Die Rechercheaufgaben wurden von zwei voneinander unabhängigen Beurteilern ausgewertet. Zur Untersuchung der Beurteilerübereinstimmung wurden für jede Aufgabe Pearson-Korrelationen zwischen den von beiden Beurteilern vergebenen Punktwerten berechnet. Für die Ergebnisbewertungen lagen die Korrelationen zwischen $r = .62$ und $r = .87$. Für die Prozessbewertungen wurde die Beurteilerübereinstimmung für jede Facette (z.B. verwendetes Suchwerkzeug) bestimmt; die Korrelationen lagen zwischen $r = .72$ und $r = .92$. Damit lag die Beurteilerübereinstimmung oberhalb der in der Literatur genannten Untergrenze von $r = .60$ (Hartmann, 1977). Bei unterschiedlichen Bewertungen einigten sich die Beurteiler auf eine gemeinsame Bewertung. Anschließend wurden die für jede Aufgabe vergebenen Punktwerte addiert und so skaliert, dass sie im Bereich zwischen 0 und 1 liegen, also mit 100% multipliziert als Prozentsatz der maximal erreichbaren Punktwerte interpretiert werden können. Die Werte sind in Tabelle 8 dargestellt.

Tabelle 8

Ergebnis- und Prozesspunktwerte für die Rechercheaufgaben in der Trainingsevaluationsstudie

| Aufgabe und Messzeitpunkt | Ergebnispunktwerte $M(SD)$ | Prozesspunktwerte $M(SD)$ |
|---------------------------|----------------------------|---------------------------|
| Aufgabe 1 t1 | .77(.29) | .55(.20) |
| Aufgabe 2 t1 | .50(.31) | .46(.17) |
| Aufgabe 3 t1 | .32(.26) | .36(.17) |
| Aufgabe 1 t2 | .90(.21) | .76(.17) |
| Aufgabe 2 t2 | .67(.30) | .71(.21) |
| Aufgabe 3 t2 | .31(.29) | .63(.24) |
| Aufgabe 1 t3 | .91(.21) | .79(.11) |
| Aufgabe 2 t3 | .87(.19) | .72(.16) |
| Aufgabe 3 t3 | .61(.36) | .74(.18) |

Die Ergebnispunktwerte zu t1 korrelierten zwischen $r = .03$ und $r = .15$ (alle *ns*), die Prozesspunktwerte korrelierten zwischen $r = .27$ und $r = .43$ (alle mindestens $p < .05$, einseitig). Bei Messzeitpunkt 2 lagen die Korrelationen der Ergebnispunktwerte in einem ähnlichen Bereich: Zwischen $r = .04$ und $r = .13$ (alle *ns*); die Korrelationen der Prozesspunktwerte lagen etwas höher: Zwischen $r = .48$ und $r = .61$ (alle $p < .01$, einseitig). Bei Messzeitpunkt 3 lagen die Interkorrelationen der Ergebnispunktwerte im Bereich zwischen $r = .00$ und $r = .19$ (alle *ns*). Die Interkorrelationen der Prozesspunktwerte lagen zwischen $r = .13$ (*ns*) für die Korrelation zwischen Aufgaben 2 und 3, $r = .22$ für die Korrelation zwischen Aufgaben 1 und 2 sowie $r = .27$ für die Korrelation zwischen Aufgaben 1 und 3 (beide $p < .05$, einseitig).

Als nächstes wurden die in der Tabelle enthaltenen Ergebnis- und Prozesspunktwerte für jeden Messzeitpunkt gemittelt. Damit resultierte für jeden Messzeitpunkt ein Wert für das Ergebnis der Recherche und ein Wert für den Prozess. Damit soll der Messfehler reduziert werden, da die Ergebnisse nun nicht mehr von nur einer Aufgabe abhängig sind. Die Werte sind, nach Gruppenzugehörigkeit geteilt, in Abbildungen 1 und 2 dargestellt.

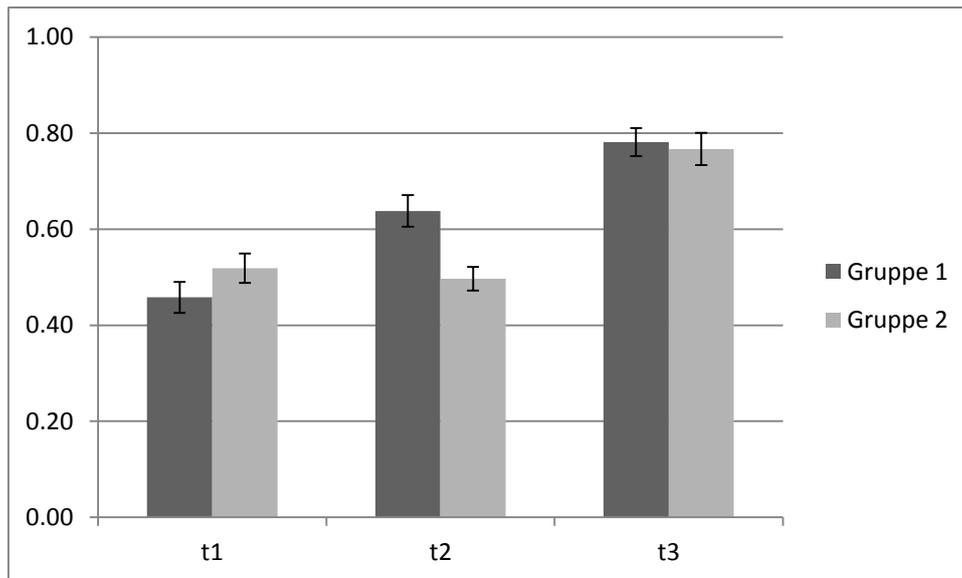


Abbildung 1. Ergebnispunktwerte zu den drei Messzeitpunkten. Fehlerbalken = SE.

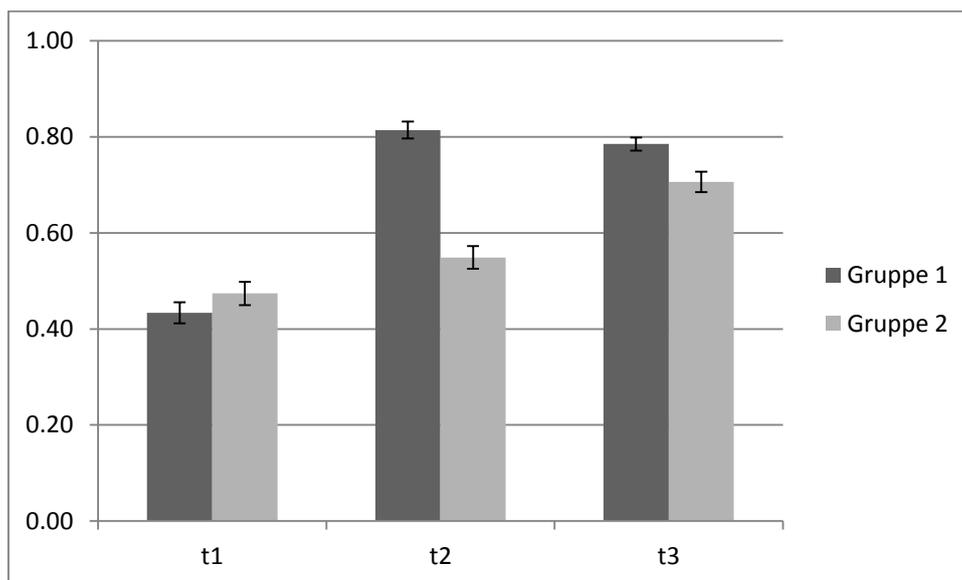


Abbildung 2. Prozesspunktwerte zu den drei Messzeitpunkten. Fehlerbalken = SE.

Der MC-Test wurde in derselben Weise ausgewertet wie in der Online-Studie. Aufgrund der Ergebnisse der Online-Studie wurde diesmal auf eine Berechnung von Subskalen verzichtet. Reliabilitätskennwerte für die Gesamtskala zu den drei Messzeitpunkten sind in Tabelle 9 angegeben. Die Punktwerte sind, nach Gruppenzugehörigkeit geordnet, in Abbildung 3 dargestellt. Da im Rahmen dieser Studie Daten zu drei Zeitpunkten erhoben worden waren, konnte – zumindest mit einer Teilstichprobe – die Retest-Reliabilität der Instrumente bestimmt werden. Hierfür bot es sich an, die zu t1 und t2 erhobenen Daten von Gruppe 2 zu

verwenden, da diese Gruppe in der Zwischenzeit als Wartekontrollgruppe fungiert hatte; der Zeitraum zwischen beiden Datenerhebungen betrug 2 Wochen. Für die Ergebnispunktwerte ergab sich eine Korrelation von $r = .53$; für die Prozesspunktwerte $r = .63$ und für den MC-Test $r = .71$. Im Folgenden werden die zuvor definierten Hypothesen überprüft.

Tabelle 9

Reliabilitätskennwerte für den MC-Test in der Trainingsevaluationsstudie

| Kennwert | Messzeitpunkt t1 | Messzeitpunkt t2 | Messzeitpunkt t3 |
|-----------------------------|------------------|------------------|------------------|
| Cronbach's α | .63 | .80 | .55 |
| Split-Half (Spearman-Brown) | .57 | .84 | .61 |

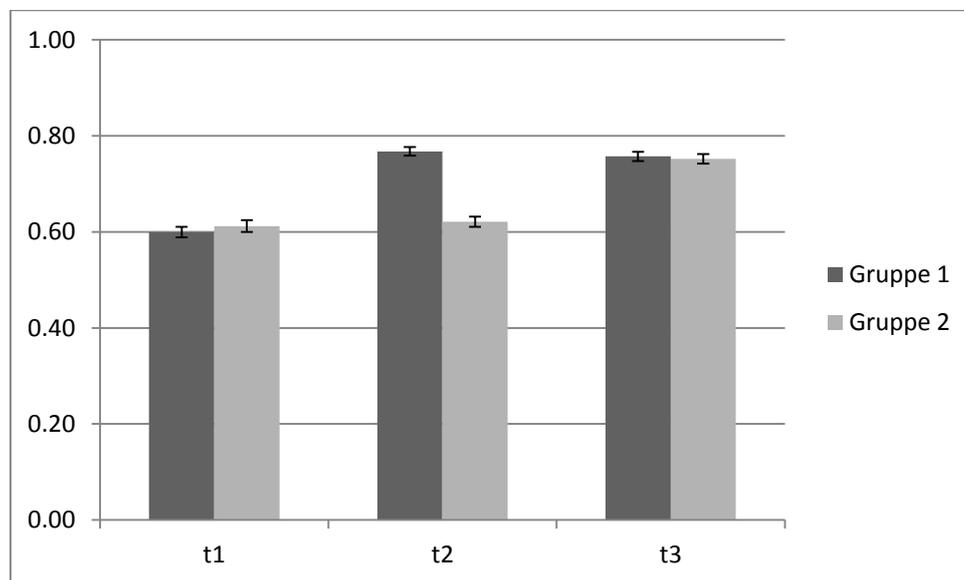


Abbildung 3. Punktwerte im MC-Test zu den drei Messzeitpunkten. Fehlerbalken = SE.

13.2.1 Hypothese 1

Zur Prüfung der ersten Hypothese wurde untersucht, ob sich die antizipierten Aufgabenschwierigkeiten in den durchschnittlich erreichten Punktwerten (vgl. Tabelle 8, S. 67) widerspiegeln. Hierzu wurden zu jedem Messzeitpunkt Varianzanalysen mit Messwiederholung für Ergebnis- und Prozesspunktwerte berechnet. Für t1 zeigten sich bezüglich der Ergebnispunktwerte ($F[2,132] = 45.66, p < .01$) und der Prozesspunktwerte ($F[2,132] = 29.65, p < .01$) signifikante Unterschiede zwischen den Aufgaben. In beiden Fällen zeigte sich ein signifikanter linearer Trend. Die Daten von t2 zeigten ebenfalls signifikante Unterschiede bei den Ergebnispunktwerten ($F[2,132] = 87.51, p < .01$) und Prozesspunktwerten ($F[2,132] = 14.80, p < .01$); auch hier zeigte sich ein signifikanter linearer Trend. Das gleiche Bild ergab sich bei

t3: Signifikante Unterschiede zwischen den Aufgaben auf den Ergebnispunktwerten ($F[2,132] = 29.26, p < .01$) und den Prozesspunktwerten ($F[2,132] = 5.02, p < .05$); der lineare Trend erreichte Signifikanz.

13.2.2 Hypothese 2

Zur Untersuchung der zweiten Hypothese wurden Korrelationen zwischen den Ergebnis- und Prozesspunktwerten für jeden Aufgabentyp und jeden Messzeitpunkt berechnet. Die Ergebnisse sind in Tabelle 10 dargestellt.

Tabelle 10

Korrelationen von Ergebnis- und Prozesspunktwerten nach Aufgabentyp und Messzeitpunkt

| Aufgabentyp | Messzeitpunkt t1 | Messzeitpunkt t2 | Messzeitpunkt t3 |
|---------------|------------------|------------------|------------------|
| Aufgabentyp 1 | .28* | -.32** | .21* |
| Aufgabentyp 2 | .09 | .39* | .38* |
| Aufgabentyp 3 | .08 | .34** | .28* |

Anmerkung. ** $p < .01$; * $p < .05$, beide einseitig.

Wurden die für jeden Messzeitpunkt separat gemittelten Ergebnis- und Prozesspunktwerte verwendet, ergab sich für t1 eine Korrelation von $r = .22$ ($p < .05$); für t2 eine Korrelation von $r = .41$ ($p < .01$) und für t3 eine Korrelation von $r = .32$ ($p < .01$, alle p einseitig).

13.2.3 Hypothese 3

Zur Prüfung der dritten Hypothese wurden Korrelationen zwischen den aggregierten Ergebnis- und Prozesspunktwerten und den im MC-Test erreichten Punktwerten berechnet. Die Korrelationskoeffizienten sind in Tabelle 11 zusammengefasst.

Tabelle 11

Korrelationen zwischen dem MC-Test und den Punktwerten der Rechercheaufgaben nach Messzeitpunkt

| Korrelation | Messzeitpunkt t1 | Messzeitpunkt t2 | Messzeitpunkt t3 |
|-----------------------------------|------------------|------------------|------------------|
| Cor (Ergebnispunktwerte, MC-Test) | .30** | .49** | .00 |
| Cor (Prozesspunktwerte, MC-Test) | .49** | .70** | .15 |

Anmerkung. ** $p < .01$; * $p < .05$, beide einseitig.

13.2.4 Hypothese 4

Zuerst musste sichergestellt werden, dass sich beide Gruppen vor Beginn des Trainings (t_1) nicht unterschieden. Erwartungskonform zeigte sich weder bei den Ergebnispunktwerten ($t[65] = 1.34, ns$), noch bei den Prozesspunktwerten ($t[65] = 1.23, ns$) oder dem MC-Test ($t[65] = .78, ns$) eine signifikante Differenz der Gruppen. Den Erwartungen entsprechend wurden signifikante Differenzen bei t_2 gefunden (für die Ergebnispunktwerte $t[65] = 3.32, p < .01$; für die Prozesspunktwerte $t[65] = 9.21, p < .01$ und $t[65] = 10.47, p < .01$ für den MC-Test). In allen Fällen erreichte Gruppe 1 höhere Werte als Gruppe 2. Für den Messzeitpunkt 3 zeigten sich erwartungsgemäß keine Differenzen in den Ergebnispunktwerten ($t[65] = .33, ns$) und im MC-Test ($t[65] = .37, ns$). Bei den Prozesspunktwerten zeigte sich jedoch wieder ein Unterschied ($t[65] = 3.21, p < .01$, alle p einseitig), wobei Gruppe 1 höhere Werte erreichte als Gruppe 2.

Als nächstes wurden für die drei messwiederholt erfassten Variablen separate Varianzanalysen mit Messwiederholung berechnet, wobei die Gruppenzugehörigkeit als Zwischensubjektfaktor einging. Für alle drei Variablen zeigte sich eine signifikante Interaktion von Messwiederholungs- und Zwischensubjektfaktor (für die Ergebnispunktwerte $F[2,130] = 5.45, p < .01$; für die Prozesspunktwerte $F[2,130] = 37.38, p < .01$ und für den MC-Test $F[2,130] = 73.13, p < .01$). Anschließend wurden separate Analysen für beide Gruppen berechnet.

Bei der Analyse der Ergebnispunktwerte zeigte sich für beide Gruppen ein signifikanter Effekt des Messwiederholungsfaktors (für Gruppe 1 $F[2,72] = 27.17, p < 0.01$; für Gruppe 2 $F[2,58] = 25.63, p < 0.01$). Mit SPSS berechnete wiederholte Kontraste zeigten bei Gruppe 1 signifikante ($p < .01$, einseitig) Differenzen zwischen t_1 und t_2 sowie zwischen t_2 und t_3 ; bei Gruppe 2 jedoch nur zwischen t_2 und t_3 .

Die Analyse der Prozesspunktwerte ergab ebenfalls signifikante Effekte des Messwiederholungsfaktors für beide Gruppen (für Gruppe 1 $F[2,72] = 130.55, p < .01$; für Gruppe 2 $F[2,58] = 54.31, p < .01$). Anhand wiederholter Kontraste konnte festgestellt werden, dass sich die Werte von Gruppe 1 nur zwischen t_1 und t_2 signifikant ($p < .01$, einseitig) unterschieden. Die Werte von Gruppe 2 unterschieden sich sowohl zwischen t_1 und t_2 als auch zwischen t_2 und t_3 signifikant.

Bezüglich des MC-Tests zeigte sich für beide Gruppen ebenfalls ein signifikanter Effekt des Messwiederholungsfaktors ($F[2,72] = 183.12$, $p < .01$ für Gruppe 1 und $F[2,58] = 113.45$, $p < .01$ für Gruppe 2). Wiederholte Kontraste zeigten für Gruppe 1 signifikante ($p < .05$, einseitig) Differenzen zwischen t1 und t2 sowie zwischen t2 und t3. Bei Gruppe 2 erreichte nur die Differenz zwischen t2 und t3 das Signifikanzniveau. Die Ergebnisse der Trainingsevaluation mittels Rechercheaufgaben wurden bereits schriftlich veröffentlicht (Leichner et al., 2014a).

Clusteranalyse

Wie die oben berichteten Ergebnisse zeigen, korrelierten die drei Variablen, mit denen Informationskompetenz erfasst wurde, moderat miteinander. Zudem zeigte sich ein recht ähnliches Verlaufsmuster über die drei Messzeitpunkte. Daher wurde anschließend mittels clusteranalytischer Methoden untersucht, ob sich Probandengruppen mit unterschiedlichen Fähigkeitsprofilen zeigen. Clusteranalytische Verfahren werden eingesetzt, um Personen (oder andere Objekte) anhand bestimmter Variablen zu gruppieren. Die Mitglieder einer Gruppe sollen sich bezüglich dieser Variablen möglichst ähnlich sein – Mitglieder verschiedener Gruppen entsprechend möglichst unähnlich (Bühl, 2008, p. 545). Hier wurden die beiden Punktwerte der Rechercheaufgaben und die Punktwerte im MC-Test verwendet. Es sollte also überprüft werden, ob sich Gruppen von Probanden identifizieren lassen, die sich in Bezug auf diese drei Variablen ähnlich sind. Dabei wurde auf die Daten von t2 zurückgegriffen, da zu diesem Zeitpunkt stärker ausgeprägte Unterschiede zwischen den Probanden – und damit mehr Varianz – vorlag.

Grundsätzlich kann man solche Verfahren in hierarchische und nicht-hierarchische einteilen (Bortz, 1999, p. 555). Bei den üblicherweise verwendeten agglomerativ-hierarchischen Verfahren (z.B. Analyse nach Ward) werden in jedem Schritt zwei nach bestimmten Fusionskriterien ausgewählte Cluster zusammengefasst, bis sich alle Fälle in einem Cluster befinden. Dabei steigt mit jedem Fusionierungsschritt die Distanz der beiden fusionierten Cluster. Solche Verfahren können benutzt werden, um die optimale Clusterzahl zu bestimmen: Steigt die Distanz nach Fusion zweier Cluster sprunghaft an, kann dies als Indiz gewertet werden, den Fusionierungsprozess an dieser Stelle abzubrechen und die beiden Cluster nicht zu aggregieren (Bühl, 2008, p. 552). Neben den hier beschriebenen agglomerativ-hierarchischen

Verfahren existieren auch divisive Verfahren, die mit einem Gesamtcluster beginnen, welches in Teilcluster aufgeteilt wird (Bortz, 1999, p. 554); darauf wird hier nicht eingegangen. Bei nicht-hierarchischen Verfahren (z.B. k-means-Methode) hingegen muss die Clusterzahl vorab festgelegt werden. Zudem muss eine Startgruppierung vorliegen, d.h. jeder Fall muss bereits einem Cluster zugeordnet sein. Der Algorithmus versucht in der Folge, durch schrittweises Verschieben der Fälle, bzw. Objekte, die Clusterlösung zu optimieren. In der Literatur findet man die Empfehlung, zunächst anhand einer hierarchischen Analyse nach Ward die optimale Clusterzahl zu bestimmen. Diese Lösung kann dann mittels der nicht-hierarchischen k-means-Methode optimiert werden, indem die Clusterzentren der durch die Ward-Analyse ermittelten Cluster als Startpartition der k-means-Methode vorgegeben werden (Bortz, 1999, pp. 555–556).

Vor Beginn der Clusteranalyse empfiehlt es sich, die beteiligten Variablen zu standardisieren, da unterschiedliche Skalierungen das Ergebnis der Analyse verzerren können (Bühl, 2008, p. 549). Daher wurde eine z-Standardisierung der Ergebnis- und Prozesspunktwerte der Rechercheaufgaben sowie der im MC-Test erreichten Punktwerte von Messzeitpunkt 2 vorgenommen. Anschließend wurde eine Ward-Analyse durchgeführt. Die Ward-Analyse fusioniert diejenigen Cluster, durch deren Fusion die Fehlerquadratsumme (Summe der quadrierten Abweichungen vom Clustermittelwert) am wenigsten ansteigt. Die Inspektion der Fehlerquadratsumme zeigte, dass diese bei dem Schritt von drei zu zwei Clustern sprunghaft anstieg, entsprechend wurde davon ausgegangen, dass hier drei Cluster gebildet werden sollten. Diese Clusterzahl und die Gruppenmittelwerte wurden als Startwerte der folgenden k-means-Analyse verwendet. Die k-means-Analyse verschiebt Fälle immer dann in ein anderes Cluster, wenn die Distanz zum Clusterzentrum des anderen Clusters geringer ist, als die Distanz zum eigenen Clusterzentrum. Auch hier wird die quadrierte euklidische Distanz als Distanzmaß verwendet. Danach werden die Clustermittelpunkte neu berechnet und es wird wieder verschoben, bis kein Objekt existiert, das einem anderen Cluster näher ist. Daher kann dieses Verfahren verwendet werden, um eine Clusterlösung zu optimieren. Die Clustergrößen und Clusterzentren sind in Tabelle 12 angegeben.

Tabelle 12

Clusterzentren und Größen der Cluster

| Cluster | n | MC-Test | | Ergebnispunktwerte | | Prozesspunktwerte | |
|-----------|----|---------|-----|--------------------|-----|-------------------|-----|
| | | M | SD | M | SD | M | SD |
| Cluster 1 | 28 | -.97 | .57 | -.46 | .63 | -.95 | .65 |
| Cluster 2 | 24 | .96 | .40 | 1.00 | .70 | .79 | .59 |
| Cluster 3 | 15 | .27 | .56 | -.74 | .64 | .50 | .44 |

Es ist auffällig, dass die Varianz innerhalb der Cluster deutlich geringer ist als 1.0 (also die Varianz der z-standardisierten Variablen in dieser Stichprobe). Dies spricht dafür, dass es gelungen ist, ähnliche Fälle zu gruppieren. Offenbar befinden sich in Cluster 1 solche Probanden, die relativ niedrige Werte auf allen drei Variablen erreichen. Cluster 2 beinhaltet einheitlich leistungsstarke Probanden. Cluster 3 hingegen scheint Probanden zu beinhalten, die Schwierigkeiten bei der Verwendung der Suchwerkzeuge haben. Sie erreichen überdurchschnittliche Punkte im MC-Test und bei den Prozesspunktwerten, finden aber keine geeigneten Publikationen, wie die unterdurchschnittlichen Ergebnispunktwerte zeigen.

Um zu prüfen, ob zwischen den Clustern signifikante Differenzen bezüglich der drei Variablen vorliegen, wurde für jede Variable eine einfaktorielle Varianzanalyse berechnet. Erreichte der jeweilige F-Wert das Signifikanzniveau, wurden Post-hoc-Tests mit Bonferroni-Korrektur berechnet, um einzelne Gruppenunterschiede zu untersuchen; hierfür wurde das Signifikanzniveau auf $p < .05$ festgelegt. Für den MC-Test lieferte die Varianzanalyse signifikante Differenzen ($F[2,64] = 93.53, p < .01$), Post-hoc-Tests mit Bonferroni-Korrektur zeigten signifikante Unterschiede in jedem Gruppenvergleich. Für die Ergebnispunktwerte ergaben sich ebenfalls signifikante Gruppenunterschiede ($F[2,64] = 44.20, p < .01$), Post-hoc-Tests mit Bonferroni-Korrektur ergaben, dass sich lediglich Cluster 1 und 3 nicht signifikant voneinander unterschieden. Bezüglich der Prozesspunktwerte wurden ebenfalls signifikante Gruppenunterschiede ermittelt ($F[2,64] = 63.81, p < .01$). Bonferroni-korrigierte Post-Hoc-Tests zeigten, dass lediglich zwischen Cluster 2 und 3 keine signifikanten Unterschiede vorlagen.

Im nächsten Schritt wurde untersucht, wie sich die beiden Gruppen der vorliegenden Studie auf die Cluster aufteilten. Dabei zeigte sich, wie in Tabelle 13 dargestellt, dass Probanden aus Gruppe 1 überwiegend in Clustern 2 und 3 zu finden waren, während die meisten Probanden aus Gruppe 2 dem Cluster 1 zugeordnet wurden ($X^2[2, N = 67] = 45.08, p < .01$).

Tabelle 13

Clusterzugehörigkeit nach Gruppe

| Gruppe | n | Cluster | | |
|----------|----|---------|----|----|
| | | 1 | 2 | 3 |
| Gruppe 1 | 37 | 2 | 23 | 12 |
| Gruppe 2 | 30 | 26 | 1 | 3 |

13.2.5 Hypothese 5

In diesem Abschnitt wird Hypothese 5 behandelt, welche sich auf Wechselwirkungen zwischen der Selbstwirksamkeitsskala und der Leistung in MC-Test und Rechercheaufgaben bezieht. Die Selbstwirksamkeitsskala wurde wieder durch Berechnen des Mittelwerts der $k = 10$ Items gebildet; die Werte sind, nach Gruppenzugehörigkeit und Messzeitpunkt geordnet, in Abbildung 4 wiedergegeben. Die zu den drei Messzeitpunkten ermittelten Werte für die interne Konsistenz und Split-Half-Reliabilität (Spearman-Brown) der Skala sind in Tabelle 14 angegeben. Fehlende Werte wurden hierbei wieder mit paarweisem Fallausschluss behandelt.

Tabelle 14

Reliabilitätskennwerte für die Selbstwirksamkeitsskala (Trainingsevaluationsstudie)

| Kennwert | Messzeitpunkt 1 | Messzeitpunkt 2 | Messzeitpunkt 3 |
|---------------------------------------|-----------------|-----------------|-----------------|
| Interne Konsistenz (α) | .84 | .78 | .68 |
| Split-Half-Reliabilität ($r_{tt'}$) | .84 | .77 | .68 |

Eine Varianzanalyse mit Messwiederholung mit Messwiederholungs- und Gruppenfaktor ergab eine signifikante Interaktion beider Faktoren ($F[2,130] = 18.28, p < .01$). Anschließend wurden separate Varianzanalysen mit Messwiederholung für beide Gruppen berechnet. Es zeigte sich für beide Gruppen ein Effekt des Messwiederholungsfaktors (für Gruppe 1 $F[2,72] = 48.93, p < .01$, für Gruppe 2 $F[2,58] = 39.47, p < .01$). Wiederholte Kontraste ergaben, dass die Werte von Gruppe 1 zwischen t1 und t2 signifikant ($p < .05$, einseitig) voneinander abwichen. Die Werte von Gruppe 2 unterschieden sich zwischen t1 und t2 sowie zwischen t2 und t3 signifikant.

Im Folgenden wird der Frage nachgegangen, ob sich Zusammenhänge zwischen der Selbstwirksamkeitsskala und den Leistungsmaßen für Informationskompetenz zeigen. Zu Messzeitpunkt 1 korrelierte die Selbstwirksamkeitsskala signifikant mit dem MC-Test ($r = .46$,

$p < .01$, einseitig), die Korrelationen mit den Rechercheaufgaben erreichten jedoch nur teilweise das Signifikanzniveau ($r = .12$, ns für die Ergebnispunktwerte und $r = .45$, $p < .01$, einseitig für die Prozesspunktwerte). Zu t_2 erreichten alle drei Korrelationskoeffizienten das Signifikanzniveau ($r = .68$ für den MC-Test, $r = .33$ für die Ergebnispunktwerte und $r = .55$ für die Prozesspunktwerte, alle $p < .01$, einseitig). Bei t_3 erreichte jedoch keiner der Zusammenhänge das Signifikanzniveau ($r = .09$, $r = .12$ und $r = -.01$, alle ns , einseitig).

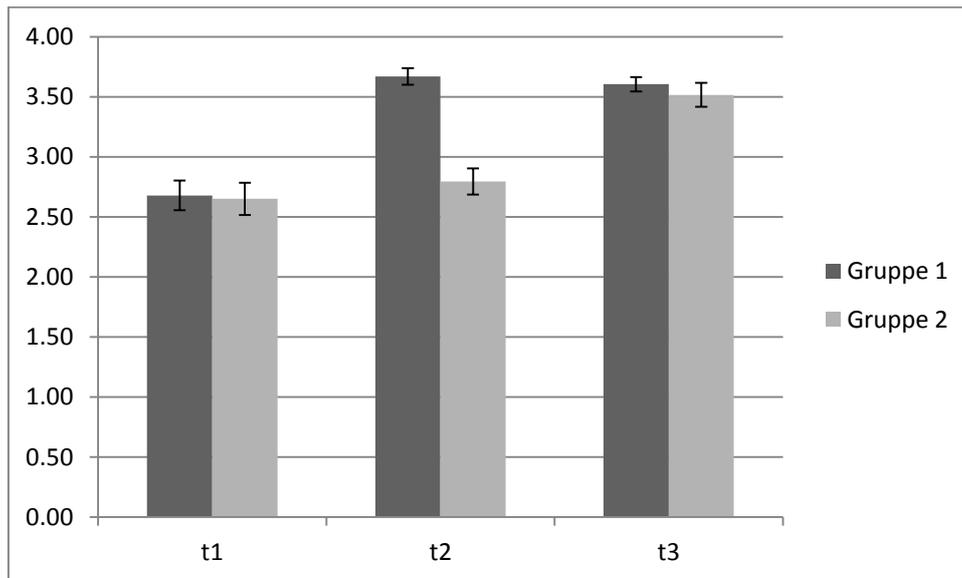


Abbildung 4. Werte auf der Selbstwirksamkeitsskala. Fehlerbalken = SE.

Anschließend wurde die Richtung des Zusammenhangs von Selbstwirksamkeit und Leistung untersucht. Es sollte also die Frage geklärt werden, ob Selbstwirksamkeit in dieser Studie primär ein Prädiktor für Leistung oder ob Leistung umgekehrt ein Prädiktor für Selbstwirksamkeit ist. Dies ist möglich, da hier ein Design mit mehreren Messzeitpunkten vorliegt. Hierfür wurde ausschließlich auf den MC-Test als Leistungsindikator zurückgegriffen, da die Rechercheaufgaben stärker mit Messfehler behaftet sind. Zudem lässt sich diese Fragestellung nur dann verlässlich untersuchen, wenn keine Intervention zwischen Erfassung der Selbstwirksamkeit und Erfassung des Leistungsindikators stattgefunden hat. Im Folgenden werden daher für beide Probandengruppen separate Analysen berichtet.

Für Gruppe 1 wurden daher die Zusammenhänge anhand der Daten von t_2 und t_3 untersucht, da diese Gruppe zu diesem Zeitpunkt bereits am Training teilgenommen hatte. Dabei zeigte sich, dass die Selbstwirksamkeit zu t_2 nicht signifikant mit der Testleistung zu t_3 ($r = .07$, ns) korrelierte. Die Testleistung zu t_2 korrelierte mit der Selbstwirksamkeit zu t_3

ebenfalls nicht signifikant ($r = .20$, ns). Die Ergebnisse sprechen gegen einen Zusammenhang von Selbstwirksamkeit und Testleistung über zwei Messzeitpunkte hinweg.

Dieses Ergebnis sollte mithilfe multipler Regression, in die sowohl Selbstwirksamkeit, als auch Testleistung zum vorherigen Messzeitpunkt eingehen, abgesichert werden. Prinzipiell wäre es möglich, dass durch den Einbezug der jeweils anderen Variablen ein signifikanter Zusammenhang entsteht, indem die andere Variable irrelevante Varianzen unterdrückt – ein sogenannter Suppressionseffekt (Bortz, 1999, p. 444). Mithilfe einer multiplen Regressionsanalyse wurde zunächst untersucht, ob die Leistung im MC-Test zu t3 durch die Leistung im Test und die Selbstwirksamkeitsskala, beide erhoben zu t2, vorhergesagt werden kann. Für die Testleistung zu t2 ergab sich ein standardisiertes β von $.81$ ($p < .01$), während sich die zu t2 erhobene Selbstwirksamkeit als nicht signifikanter Prädiktor erwies ($\beta = -.09$, ns). Die Varianzaufklärung durch das Modell betrug $R^2 = .64$. Als nächstes wurde untersucht, ob die zu t2 erhobene Testleistung die zu t3 erhobene Selbstwirksamkeit vorhersagt, wenn auch die zu t2 erhobene Selbstwirksamkeit als Prädiktorvariable in das Modell eingesetzt wird. Hier erwies sich lediglich die zu t2 erhobene Selbstwirksamkeit als signifikanter Prädiktor ($\beta = .59$, $p < .01$); der Einfluss der Testleistung zu t2 hatte keinen signifikanten Einfluss ($\beta = .09$, ns); die Varianzaufklärung durch das Modell betrug $R^2 = .38$.

Für Gruppe 2 wurden die Zusammenhänge anhand der Daten von t1 und t2 untersucht, da diese Gruppe erst danach am Training teilgenommen hatte. Hier zeigte sich eine marginal signifikante Korrelation zwischen Selbstwirksamkeit zu t1 und Testleistung zu t2 ($r = .28$, $p < .10$, einseitig). Zwischen der Testleistung zu t1 und der Selbstwirksamkeit zu t2 zeigte sich eine signifikante Korrelation ($r = .51$, $p < .01$).

In diesem Fall sollte mithilfe der Regressionsanalyse überprüft werden, ob die Zusammenhänge auch dann Signifikanz erreichen, wenn der Einfluss von Selbstwirksamkeit, bzw. Testleistung, zu t1 einbezogen wird. Wurde die Testleistung zu t2 durch Selbstwirksamkeit und Testleistung zu t1 vorhergesagt, hatte lediglich die Testleistung zu t1 ($\beta = .72$, $p < .01$) signifikanten Einfluss. Der Einfluss der zu t1 erhobenen Selbstwirksamkeit erreichte dann nicht mehr das Signifikanzniveau ($\beta = -.01$, ns); die gesamte Varianzaufklärung betrug $R^2 = .51$. Wurde die Selbstwirksamkeit zu t2 durch Selbstwirksamkeit und Testleistung zu t1 vorhergesagt, ergaben sich für Testleistung ($\beta = .21$, $p < .10$) und Selbstwirksamkeit von t1 ($\beta = .70$,

$p < .01$, alle einseitig) zumindest marginal signifikante Koeffizienten; die Varianzaufklärung betrug $R^2 = .66$.

13.3 Diskussion

Die Ergebnisse bezüglich der Rechercheaufgaben sprechen dafür, dass ein verlässliches und valides Messinstrument für prozedurale Anteile von Informationskompetenz entwickelt wurde. Mit den Aufgaben kann sowohl das Vorgehen (Prozesspunktwerte), als auch das Ergebnis der Recherche bewertet werden.

Bei der Analyse der Beurteilerübereinstimmung zeigte sich, dass die Korrelationen zwischen den Bewertungen der beiden Beurteiler in allen Fällen über dem Wert von $r = .60$ lagen, der in der Literatur (Hartmann, 1977) als unterer Grenzwert genannt wird. Dabei sollte betont werden, dass die meisten Korrelationen über $r = .70$ lagen. Die Retest-Reliabilität erreichte ebenfalls einen akzeptablen Bereich (vgl. March, Sullivan, & Parker, 1999), was angesichts des relativ hohen Messfehlers und der oben beschriebenen Probleme bei der Durchführung von Rechercheaufgaben fast schon erstaunlich ist. Dabei ist auffällig, dass beide Reliabilitätskennwerte für die Prozesspunktwerte höher lagen, als für die Ergebnispunktwerte. Da die Entscheidung, ob eine gefundene Publikation ein bestimmtes Thema behandelt, wesentlich weniger eindeutig zu treffen ist, als die Entscheidung, ob ein bestimmtes Suchwerkzeug benutzt wurde, ist dieser Befund gut erklärbar. Zusammen genommen sprechen diese Befunde dafür, dass sich die Entwicklung spezifischer und analytischer Auswertungsschlüssel (Jonsson & Svingby, 2007; Moskal, 2000) ausgezahlt hat.

Was den Multiple-Choice-Test angeht, hatte sich gezeigt, dass die interne Konsistenz stark von der Heterogenität der Stichprobe abhängig war. In dieser Studie erreichte der Test zu Messzeitpunkt 1 eine interne Konsistenz von über $\alpha = .60$. Angesichts der relativ homogenen Stichprobe, die ausschließlich aus Studierenden des BSc-Studiengangs bestand, ist dies durchaus erstaunlich. Unter Berücksichtigung der bereits dargelegten Heterogenität des Konzepts Informationskompetenz und der Tatsache, dass bei den meisten Studierenden keine kohärente Wissensbasis zum Thema Informationskompetenz vorliegt, sind, wie in Abschnitt 11.5 (S. 52) dargelegt, schon allein durch die Natur des Testgegenstandes bedingt keine hohen Werte für die interne Konsistenz zu erwarten. Die Retest-Reliabilität liegt mit

$r = .71$ oberhalb der in der Literatur genannten Untergrenze von $r = .70$ (Fisseni, 1997). Offenbar sind die Mittelwerte zeitlich relativ stabil, was für die Verlässlichkeit des Tests spricht.

Da der Test im Laufe der Studie dreimal bearbeitet wurde, kann auch untersucht werden, wie sich die interne Konsistenz im Laufe der Studie verändert. Interessanterweise lag sie zu Messzeitpunkt 2 wesentlich höher als bei Messzeitpunkt 1, sank jedoch bei Messzeitpunkt 3 leicht unter das Niveau des ersten Messzeitpunkts. Der außergewöhnlich hohe Kennwert zu Messzeitpunkt 2 kann durch die Varianz, welche durch den unterschiedlichen Trainingsstand der Probanden bedingt ist, erklärt werden. Zu diesem Zeitpunkt hatte Gruppe 1 bereits am Training teilgenommen, Gruppe 2 jedoch noch nicht. Wird die interne Konsistenz für beide Gruppen separat berechnet, ergeben sich Werte von $\alpha = .57$ für Gruppe 1 und $\alpha = .54$ für Gruppe 2. Interessanter ist jedoch die niedrige interne Konsistenz zu Messzeitpunkt 3. Zuvor war postuliert worden, dass bei den Studierenden keine kohärente Wissensbasis zum Thema Informationskompetenz vorliegt, was einer hohen internen Konsistenz entgegensteht. Nach Trainingsteilnahme sollte bei den Probanden eine relativ kohärente Wissensbasis vorliegen – und die interne Konsistenz zumindest das Niveau von Messzeitpunkt 1 erreichen. Dass dies nicht der Fall ist, kann daran liegen, dass zwar eine kohärente Wissensbasis vorliegt, die Stichprobe nun aber wieder homogener geworden ist, was zu einer geringen Varianz – und damit niedrigen Korrelationen zwischen den Items führt. Eine andere, in diesem Fall möglicherweise etwas gewagte, Vermutung wäre, hier eine Ausdifferenzierung der Fähigkeiten anzunehmen, auch wenn sich diese Hypothese primär auf Intelligenztests bezieht. Der Differenzierungshypothese zufolge werden die Korrelationen der einzelnen Subtests eines Intelligenztests mit zunehmendem Fähigkeitsniveau geringer (zusammenfassend Deary et al., 1996), was mehrfach bestätigt wurde (z.B. Jensen, 2003). Prinzipiell könnte man annehmen, dass ein solcher Effekt auch bei Wissenstests auftritt, wobei im vorliegenden Fall keine Subtests existieren, sodass auf die interne Konsistenz zurückgegriffen werden muss.

13.3.1 Hypothesen 1-3

Die erste Hypothese konnte uneingeschränkt bestätigt werden. Typ 1-Aufgaben fielen den Probanden am leichtesten, gefolgt von Typ 2- und Typ 3-Aufgaben. Dieses Muster zeigte sich zu allen Messzeitpunkten, war bei t3 jedoch schwächer ausgeprägt, was auf ein allgemein höheres Leistungsniveau und in einigen Fällen auch auf Deckeneffekte zurückzuführen ist.

Dies bestätigt die zugrundeliegenden Annahmen über die für die Bearbeitung der einzelnen Aufgaben erforderlichen Fähigkeiten, auf deren Basis die Aufgabenschwierigkeiten postuliert worden waren.

Hypothese 2 bezog sich auf Interkorrelationen zwischen Ergebnis- und Prozesspunktwerten der Rechercheaufgaben. Hier wurden mittlere positive Korrelationen (maximal $r = .50$) erwartet. Bei Messzeitpunkt 1 zeigte sich lediglich für die Aufgabe vom ersten Typ eine erwartungskonforme Korrelation zwischen Ergebnis- und Prozesspunktwerten. Bei Aufgabentypen 2 und 3 ergaben sich lediglich nicht-signifikante Korrelationen der beiden Punktwerte. Dies kann darauf zurückgeführt werden, dass in der Stichprobe zu diesem Zeitpunkt relativ wenige Kenntnisse über fortgeschrittene Recherchemöglichkeiten vorlagen, sodass die Probanden durch Aufgabentypen 2 und 3 vermutlich überfordert waren, was einer verlässlichen Messung durch Performance Assessment entgegensteht. Bei t2 zeigte sich bei der Aufgabe vom ersten Typ unerwartet eine negative Korrelation von Ergebnis- und Prozesspunktwerten. Dieses Ergebnis kann nicht zufriedenstellend erklärt werden. Möglicherweise liegt hier ein außergewöhnlich großer Messfehler vor. Die Ergebnisse sind nicht auf den unterschiedlichen Trainingsstand der beiden Gruppen zurückzuführen: Wird die Korrelation für beide Gruppen separat berechnet, findet man keine relevanten Unterschiede. Bei Messzeitpunkt 3 zeigten sich erwartungskonform mittlere Korrelationen zwischen den Variablen. Bei Verwendung der gemittelten Werte für Ergebnis- und Prozesspunktwerte ergaben sich zu allen Messzeitpunkten erwartungsgemäß mittlere Zusammenhänge, was dafür spricht, dass der Messfehler durch die Mittelung reduziert werden konnte. Anhand dieser Daten kann die Hypothese uneingeschränkt bestätigt werden.

Bezüglich Hypothese 3 zeigten sich zu t1 und t2 erwartungskonforme Zusammenhänge zwischen den Rechercheaufgaben und dem MC-Test, welche sich bei t3 nicht replizieren ließen. Dabei fallen die unerwartet hohen Zusammenhänge zu t2 auf, welche jedoch durch die erhöhte Varianz (bedingt durch den unterschiedlichen Trainingsstand der beiden Gruppen) erklärt werden können. Schwerer fällt die Erklärung der nichtsignifikanten Zusammenhänge bei t3. Die Varianz ist zu diesem Zeitpunkt geringer, was als Erklärung dienen kann. Jedoch erscheint fraglich, ob der vollständige Effekt hierauf zurückzuführen ist. Alternativ oder ergänzend kann die Ausdifferenzierungshypothese als Erklärung angeführt werden, wonach Interkorrelationen von Subtests bei höherem Leistungsniveau geringer werden. Eine Erklä-

rung für die Ausdifferenzierungshypothese ist die Annahme, dass Personen unterschiedliche Lernerfahrungen machen, welche zwar das allgemeine Leistungsniveau heben, jedoch zu unterschiedlichen Fähigkeitsprofilen führen, welche die niedrigen Interkorrelationen der Subtests bedingen (Ferguson, 1954; zusammenfassend Harris, Vernon, & Jang, 2005). Für die Bearbeitung der Rechercheaufgaben sind teilweise andere Fähigkeiten erforderlich als für die Bearbeitung des MC-Tests, welche sich durch das Training auch in unterschiedliche Richtungen entwickeln können. Vor diesem Hintergrund erscheint diese Erklärung durchaus plausibel.

13.3.2 Hypothese 4

Die vierte Hypothese besagt, dass sich die Trainingsteilnahme positiv auf die Punktwerte in den Rechercheaufgaben und im MC-Test auswirkt. Diese Hypothese kann auch überwiegend bestätigt werden: Die Probanden erreichten im Durchschnitt nach der Trainingsteilnahme auf allen Variablen höhere Werte als vor dem Training, was nicht nur als Indiz für die Wirksamkeit des Trainings, sondern auch als Indiz für die Validität der Rechercheaufgaben gewertet werden kann. Wie beschrieben wurde hier ein Wartekontrollgruppendesign eingesetzt, sodass die beiden Gruppen zeitversetzt am Training teilnahmen. Bei genauerer Betrachtung zeigte sich jedoch, dass die Punktwerte teilweise auch dann anstiegen, wenn die jeweilige Gruppe gar nicht an der Intervention teilgenommen hatte (bspw. der Anstieg der Ergebnis-punktwerte von Gruppe 1 zwischen t2 und t3). Hierfür können im Nachhinein zwei Erklärungen gefunden werden. Möglicherweise handelt es sich um Testeffekte, also einen Anstieg der Punktwerte, der alleine durch die mehrfache Bearbeitung derselben Aufgaben (MC-Test) oder Aufgaben desselben Typs (Rechercheaufgaben) bedingt ist (vgl. Hausknecht et al., 2007). Eine andere Erklärung ist, dass die Probanden durch die Datenerhebung oder das Training angeregt wurden, sich selbstständig mit den Inhalten auseinander zu setzen, was ebenfalls die Leistung beim nächsten Messzeitpunkt erhöhen sollte. Auch die Differenz zwischen den Gruppen bei den Ergebnis-punktwerten zu t3 sollte an dieser Stelle angesprochen werden. Angesichts der relativ geringen Gruppengrößen ist diese möglicherweise als Messfehler zu erklären.

Die Ergebnisse der Clusteranalyse unterstreichen, dass mit den verwendeten Messinstrumenten unterschiedliche Fähigkeiten (oder zumindest unterschiedliche Schwerpunkte von

Fähigkeiten) erfasst wurden. Es ließen sich drei Gruppen von Probanden identifizieren, die über unterschiedliche Fähigkeitsprofile verfügen. In Cluster 1 befinden sich solche Probanden, die relativ niedrige Werte auf allen drei Variablen erreichen. Cluster 2 wurden einheitlich leistungsstarke Probanden zugeordnet. Cluster 3 hingegen scheint Probanden zu beinhalten, die Schwierigkeiten bei der Verwendung der Suchwerkzeuge haben. Sie erreichen überdurchschnittliche Punkte im MC-Test und bei den Prozesspunktwerten der Rechercheaufgaben, erreichen aber nur unterdurchschnittliche Ergebnispunktwerte in den Rechercheaufgaben. Die Tatsache, dass sich Probanden aus der nicht-trainierten Gruppe 2 überwiegend in Cluster 1 (niedrige Fähigkeiten) befinden, während sich die meisten Probanden aus Gruppe 1 in Clustern 2 und 3 befinden, bestätigt die Interpretation der Clusterlösung. Möglicherweise lassen sich diese Einteilungen nutzen, um verschiedene Trainingsangebote für die Gruppen zu konzipieren. So benötigen die Probanden aus Cluster 1 ein umfassendes Training, während Probanden in Cluster 3 möglicherweise nur in der Verwendung der Suchwerkzeuge geschult werden müssten.

Besonders interessant ist die Existenz von Cluster 3, also von Probanden, die zwar über konzeptuelles Wissen über Literaturrecherchen verfügen, jedoch nur bedingt in der Lage sind, selbst effektiv Recherchen durchzuführen. Dies unterstreicht die in der Literatur mehrfach geäußerte Feststellung, dass konzeptuelles Wissen nicht zwangsläufig mit der Fähigkeit der Wissensanwendung verbunden ist (Miller, 1990; Oakleaf, 2008; Shavelson, 2010).

13.3.3 Hypothese 5

Die in den vorherigen Studien gesammelten Belege für die Reliabilität und Validität der Selbstwirksamkeitsskala konnten bestätigt werden. Vor Beginn des Trainings erreichte die Skala akzeptable bis gute psychometrische Kennwerte und korrelierte erwartungsgemäß mit den Leistungsmaßen für Informationskompetenz. Interessant ist, dass die interne Konsistenz zu Messzeitpunkt 3 sank und zu diesem Messzeitpunkt auch keine signifikanten Korrelationen mit den Leistungsmaßen für Informationskompetenz zu beobachten waren. Dies ist umso erstaunlicher, da die Punktwerte auf allen Instrumenten zu t3 höher lagen – sich also in dieselbe Richtung wie die Leistungsmaße bewegt hatten. Die verminderte interne Konsistenz lässt sich möglicherweise darauf zurückführen, dass die Probanden nach Trainingsende nicht nur mehr Kenntnisse in diesem Bereich hatten, sondern auch ihre Stärken und Schwächen

besser einschätzen konnten. So gesehen wundert es nicht, dass die einzelnen Facetten von Informationskompetenz, welche mit den $k = 10$ Items erfasst werden, nach Trainingsende weniger stark miteinander korrelierten. Die niedrigen Zusammenhänge mit den Leistungsmaßen weisen darauf hin, dass die Probanden ihre Leistungsfähigkeit mit steigendem Leistungsniveau schlechter einschätzen konnten. Dies ist ein erstaunlicher Befund, da die meisten Studien über schlechte Selbsteinschätzungsfähigkeiten bei leistungsschwachen Studierenden berichten, während leistungsstarke Studierende ihre Leistung wesentlich besser einschätzen können (Karnilowicz, 2012; Kennedy, Lawton, & Plumlee, 2002). Allerdings war die Varianz in der Stichprobe zu t_3 wesentlich geringer, sodass dieser Befund möglicherweise auch auf statistischer Ebene zu erklären ist.

Explorativ wurde versucht, die Wirkrichtung des Zusammenhangs von Selbstwirksamkeit und Leistung zu ermitteln. Allerdings konnte nicht verlässlich gezeigt werden, dass es in dieser Studie überhaupt Zusammenhänge zwischen Selbstwirksamkeit und Leistung über Messzeitpunkte hinweg gibt. Anhand der Daten von Gruppe 2 zeigte sich, dass sich die Selbstwirksamkeit bei t_2 durch die Testleistung zu t_1 vorhersagen lässt, auch wenn die zu t_1 erhobene Selbstwirksamkeit in die Prädiktion eingeht – allerdings erreichte der Zusammenhang nur marginale Signifikanz. Prinzipiell sprechen diese Befunde für die Annahme, dass Selbstwirksamkeit als das Resultat vergangener Leistung anzusehen ist – und nicht als Prädiktor für zukünftige Leistung. Allerdings ließ sich dieser Befund nicht anhand der Daten von Gruppe 1 bestätigen. Sicherlich kann die geringe Stichprobengröße (und entsprechend niedrige Teststärke) als Begründung angeführt werden. Allerdings sollte auch erwähnt werden, dass Gruppe 1 bereits am Training teilgenommen hatte, als die hierfür relevanten Daten erhoben wurden – Gruppe 2 jedoch nicht. Weiterhin sollte erwähnt werden, dass die Probanden keine direkte Rückmeldung über ihre Leistung erhielten. In der Literatur finden sich Hinweise, dass das Vorhandensein von Leistungsrückmeldungen Einfluss auf den Zusammenhang von Selbstwirksamkeit und Leistung hat (Schmidt & DeShon, 2010). Zudem hatten die hier erbrachten Leistungen keinerlei Einfluss auf das zukünftige Leben der Probanden; die Daten waren ausschließlich zu Forschungszwecken erhoben worden und konnten daher nicht in die Benotung eingehen. Möglicherweise müssen die Leistungen eine gewisse Relevanz für die Probanden haben, damit sich Effekte von Selbstwirksamkeit zeigen; bei den meisten veröffentlichten Studien war dies sicherlich der Fall.

14 Gesamtdiskussion

Es kann festgehalten werden, dass hier zwei Messinstrumente für Informationskompetenz entwickelt wurden, mit denen die hierfür notwendigen Fähigkeiten mit unterschiedlichen Schwerpunkten erfasst werden können. Während mit dem Test eher deklarative oder konzeptuelle Schwerpunkte gesetzt werden, erfassen die Rechercheaufgaben prozedurale Aspekte. Wie weiter oben dargelegt, wird in dieser Arbeit unter deklarativem Wissen Faktenwissen über verschiedene Informationssysteme und deren Funktionen sowie deren Stärken und Schwächen verstanden. Konkrete Vorgehensweisen, Recherchestrategien und Abläufe hingegen fallen unter die Rubrik prozedurales Wissen. Deklaratives und prozedurales Wissen sind eng miteinander verbunden, da Vorgehensweisen mit Begriffen beschrieben werden müssen, die mithilfe deklarativer Begriffe definiert werden. Zudem ist deklaratives Hintergrundwissen für die Bearbeitung der Rechercheaufgaben erforderlich. Die Unterscheidung zwischen deklarativem (oder konzeptuellem) und prozeduralem Wissen ist ein zentrales Element psychologischer Theorien über Wissen und Wissenserwerb (Hiebert & Lefevre, 1986) und Grundlage etlicher Forschungsarbeiten im Bereich der Lernforschung. Dabei stammen auffällig viele Forschungsarbeiten zu diesem Thema aus dem Bereich der Mathematik, was möglicherweise auf den klar strukturierten und definierten Inhalt dieser Domäne zurückzuführen ist (Hiebert & Lefevre, 1986, p. 1).

Eine grundlegende Frage derartiger Forschungsarbeiten ist die nach der angemessenen Erfassung der beiden Wissensarten. Für den Bereich Mathematik hat es sich offenbar eingebürgert, strukturell ähnliche Aufgaben für die Erfassung prozeduralen Wissens und strukturell andersartige (sogenannte Transferaufgaben) für die Erfassung konzeptuellen Wissens zu verwenden (Schneider et al., 2011; Schneider & Stern, 2005). Dahinter steht die Annahme, dass die für die Lösung neuartiger Aufgaben erforderlichen Prozeduren erst anhand von vorliegenden konzeptuellen Wissensstrukturen generiert werden müssen, während die Prozeduren für die Lösung strukturell ähnlicher Aufgaben bereits vorliegen (Rittle-Johnson, Siegler, & Alibali, 2001). Teilweise werden auch Wissensfragen zur Erfassung konzeptuellen Wissens verwendet (Rittle-Johnson & Star, 2007). Die Verwendung von strukturell gleichartigen und neuartigen Aufgaben ist im vorliegenden Fall aus zwei Gründen nicht anwendbar. Erstens ist die Domäne fachlicher Informationskompetenz nicht so eng strukturiert, dass es

möglich wäre, Aufgaben zu konstruieren, für deren Lösung lediglich auf erprobte Prozeduren zurückgegriffen werden muss. Zweitens setzt dieses Verfahren voraus, dass ein einheitlicher Wissensstand auf Seiten der Probanden vorliegt – nur dann wäre es prinzipiell möglich, strukturell bekannte und strukturell neuartige Aufgaben zu konstruieren. Stattdessen wurden in dieser Arbeit Wissensfragen zur Erfassung konzeptuellen Wissens verwendet und Rechercheaufgaben für die Erfassung prozeduralen Wissens konstruiert. Wie die Korrelationen der beiden Messinstrumente zeigen, welche zu t1 im mittleren Bereich liegen, ist es gelungen, mit diesen Instrumenten unterschiedliche Wissensbestände zu erfassen.

Konstruktvalidität kann angenommen werden, wenn es möglich ist, anhand empirischer Ergebnisse die Zuverlässigkeit der Interpretation von Testergebnissen zu belegen (Hartig, Frey, & Jude, 2008). Dabei werden anhand eines theoretisch formulierten Konstrukts (oder mehrerer Konstrukte) Vorhersagen über empirisch beobachtbare Ergebnisse gemacht, welche überprüft werden können. Für die Konstruktvalidität der Verfahren können etliche Befunde vorgelegt werden. Insbesondere die Unterschiede zwischen den Probandengruppen auf den Testwerten des MC-Tests und die mittleren Zusammenhänge zwischen MC-Test und Rechercheaufgaben sind hier zu nennen. Die Ergebnisse der Clusteranalyse, welche zeigen, dass sich Probandengruppen mit unterschiedlichen Fähigkeitsprofilen identifizieren lassen, können als Beleg für die Konstruktvalidität der einzelnen Erhebungsverfahren und auch für das gesamte Messkonzept gewertet werden. Allerdings handelt es sich bei fachlicher Informationskompetenz nicht um ein klar definiertes Konstrukt wie beispielsweise Intelligenz. Prinzipiell kann also die Frage gestellt werden, ob hier überhaupt von Konstruktvalidität gesprochen werden kann.

Bezüglich der Reliabilität sind in diesem Kontext vor allem Retest-Reliabilität sowie interne Konsistenz von Belang. Als Retest-Reliabilität bezeichnet man die Korrelationen zwischen zwei Durchführungen des Testverfahrens, zwischen denen ein gewisser Zeitraum vergangen sein sollte. Dies ermöglicht Aussagen über die zeitliche Stabilität der Testwerte, durch unsystematische Veränderungen der wahren Werte kann dieser Reliabilitätskoeffizient jedoch verzerrt werden. Die interne Konsistenz erlaubt primär Aussagen über die Homogenität der Items (Schermelleh-Engel & Werner, 2008). Wie bereits dargelegt, weist der Test, insbesondere in der Online-Studie, eine interne Konsistenz auf, welche nicht den strengen Anforderungen für Individualdiagnostik entspricht (vgl. Schmitt, 1996). Allerdings handelt es sich bei

der Domäne Informationskompetenz um ein wesentlich inhaltsheterogeneres Feld als üblicherweise erfasste psychologische Konstrukte, wie zum Beispiel Selbstwirksamkeit, Intelligenz oder Persönlichkeitsdimensionen. Zudem sind Inhalte zur Informationskompetenz meist nicht im Curriculum verankert, sodass sich die Studierenden ihr Wissen nach dem Versuch-und-Irrtum-Prinzip aneignen (Klatt et al., 2001). Dementsprechend ist auf Seiten der Studierenden keine umfassende und kohärente Wissensbasis zu erwarten, was Voraussetzung für eine hohe interne Konsistenz des Tests ist. Sicherlich wäre es unter diesen Umständen auch möglich gewesen, einen Test mit homogeneren Items zu konstruieren, welcher dann aber nur einen Teilausschnitt des Konzepts Informationskompetenz abgebildet hätte und somit von eingeschränkter Validität gewesen wäre. Angesichts der Heterogenität der Domäne Informationskompetenz und der nicht systematischen Vermittlung von Informationskompetenzinhalten ist in diesem Kontext die Retest-Reliabilität als der wichtigere Reliabilitätskennwert anzusehen, welcher bei dem MC-Test im akzeptablen bis guten Bereich lag.

Die Rechercheaufgaben sind aus strenger psychometrischer Sicht durchaus als problematisch zu bewerten. Allerdings ist hierbei zu beachten, dass bei derartig offen gestalteten Rechercheaufgaben mit erhöhtem Messfehler gerechnet werden muss. Dieser entsteht zum einen dadurch, dass es sich um komplexe Aufgaben handelt, bei denen die einzelnen Schritte in starkem Ausmaß voneinander abhängig sind. Zum anderen sind die Probanden in der Wahl ihrer Vorgehensweise vollkommen frei, was ebenfalls zu einem erhöhten Messfehler beiträgt. Diesem Problem könnte man durch die Verwendung von Teilaufgaben entgegenwirken. Die komplette Rechercheaufgabe würde dann in mehrere Teilschritte zerlegt (Identifizieren von Suchbegriffen, Wahl des Suchwerkzeugs, Suche, Verwendung der Filterfunktionen), die separat bearbeitet werden. Das Problem der Abhängigkeit der einzelnen Schritte von den vorhergehenden Schritten könnte man lösen, indem man das Ergebnis des vorgehenden Schrittes in der Aufgabenstellung der folgenden Aufgabe nennt. Jedoch erscheint es fraglich, ob man das Ziel, Recherchefähigkeiten in einer natürlichen Umgebung zu erfassen, mit so gestalteten Aufgaben hätte erreichen können. Auch die Verwendung von Simulationsumgebungen hätte eine Zerlegung der Aufgaben in stärker standardisierte Teilschritte erfordert, und wäre damit den Anforderungen an Erfassung in einer natürlichen Umgebung nicht gerecht geworden. Zudem sollte die Bewertung der Reliabilität stets unter Berücksich-

tigung des zu erfassenden Merkmals und der Einsatzbedingungen erfolgen (Schermelleh-Engel & Werner, 2008), was die teilweise problematischen Werte relativiert.

Ein besonders wichtiger Beleg für die Konstruktvalidität der Erfassungsinstrumente und des Messkonzepts wird durch die Ergebnisse der Clusteranalyse geliefert. Hier konnten drei Gruppen von Studierenden identifiziert werden, welche unterschiedliche Fähigkeitsprofile oder Schwerpunkte in ihren Fähigkeiten aufweisen. Interessant ist, dass sich neben pauschal leistungsstarken und pauschal leistungsschwachen Probanden auch eine Gruppe zeigte, die offenbar Schwierigkeiten hatte, ihr Hintergrundwissen anzuwenden. Diese Gruppe erreichte im MC-Test und bei den Prozesspunktwerten überdurchschnittliche Werte, erlangte allerdings unterdurchschnittliche Werte, was die Qualität der gefundenen Literatur oder die Passung der gefundenen Literatur zur Aufgabenstellung anging. Damit wird das Messkonzept gestützt, wonach sich Informationskompetenz in deklarative und prozedurale Wissenskomponenten trennen lässt, wobei die prozedurale Komponente mit zwei Kennwerten erfasst wird.

Da die Selbstwirksamkeitsskala teilweise im moderaten Bereich mit den Messinstrumenten korreliert hatte, könnte prinzipiell darüber nachgedacht werden, den aufwendigen Einsatz von Messinstrumenten durch Selbsteinschätzungen zu ersetzen. Sofern man den Studierenden zusichert, dass die Daten nur in aggregierter Form – und nicht personenbezogen – ausgewertet werden, wäre es vielleicht sogar möglich, zu vermeiden, dass die Studierenden sich bewusst besser darstellen (“faking good”) und damit die Messung verzerren. Allerdings gibt es empirische Hinweise, dass Studierende ihre Fähigkeiten im Bereich Informationskompetenz überschätzen (Heinze, 2008). Zudem zeigten sich in der vorliegenden Studie bei Messzeitpunkt 3 keine signifikanten Zusammenhänge zwischen der Selbstwirksamkeitsskala und den Messinstrumenten. Folglich erscheint die Verwendung von Selbsteinschätzungsmaßen zur Erfassung von Informationskompetenz fragwürdig.

Als Einschränkungen der vorliegenden Arbeit soll hier vor allem die recht kleine Stichprobe genannt werden, anhand der die Rechercheaufgaben erprobt wurden. Zudem liegt kein empirisches Außenkriterium für die vorgestellten Messverfahren vor; insofern fehlt der Nachweis der Kriteriumsvalidität anhand eines externen Außenkriteriums. Die Korrelationen zwischen den Rechercheaufgaben und dem MC-Test können zwar als ein solcher Nachweis interpretiert werden, jedoch wurden beide Instrumente von derselben Forschergruppe entwi-

ckelt, was den Wert dieses Befundes schmälern kann. Hierfür wäre es jedoch notwendig, dass ein Außenkriterium (beispielsweise in Form eines weiteren Verfahrens einer anderen Forschergruppe) im deutschsprachigen Raum zur Verfügung steht.

Zusammenfassend lässt sich sagen, dass die Erfassung von Informationskompetenz durch MC-Aufgaben alleine den Anforderungen an eine umfassende Erhebung der relevanten Fähigkeiten und Wissensbestände nicht gerecht wird. Die Verwendung von Rechercheaufgaben ist jedoch mit hohem Zeitaufwand bei Durchführung und Auswertung verbunden, was dem Einsatz in der Praxis entgegensteht. Hier wäre es Aufgabe der Forschung, alternative Erhebungsverfahren zu entwickeln, mit denen prozedurale Komponenten von Informationskompetenz in einer ökonomischeren Weise erfasst werden können. Eine Möglichkeit wäre beispielsweise die Entwicklung eines "situational judgement test", bei dem Probanden hypothetische Situationen und mehrere Verhaltensmöglichkeiten als Antworten auf jede Situation vorgelegt werden. Die Probanden müssen sich dann für die ihrer Meinung nach angemessenste Verhaltensmöglichkeit entscheiden (Lievens, Peeters, & Schollaert, 2008). Um solche Verfahren realitätsnäher zu gestalten, können Videos oder andere Multimediaformate anstelle von schriftlichen Beschreibungen der Situation und der Antwortmöglichkeiten verwendet werden. Im konkreten Fall würde durch die Verwendung von Screenshots vermutlich bereits ein erheblicher Zuwachs an Realitätsnähe erreicht.

15 Literatur

- Abrami, P. C., Bernard, R. M., Borokhovski, E., Wade, A., Surkes, M. A., Tamim, R., & Zhang, D. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of Educational Research, 78*(4), 1102–1134. doi:10.3102/0034654308326084
- Achtziger, A., & Gollwitzer, P. M. (2006). Motivation und Volition im Handlungsverlauf. In J. Heckhausen & H. Heckhausen (Eds.), *Motivation und Handeln* (3rd ed., pp. 277–302). Heidelberg, Germany: Springer.
- Amelang, M., Bartussek, D., Stemmler, G., & Hagemann, D. (2006). *Differentielle Psychologie und Persönlichkeitsforschung* (6th ed.). Stuttgart, Germany: Kohlhammer.
- American Psychological Association (APA). (2014). *PsycINFO – quick facts*. Retrieved from <http://www.apa.org/pubs/databases/psycinfo/index.aspx?tab=2>
- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist, 51*(4), 355–365. doi:10.1037/0003-066X.51.4.355
- Anderson, K., & May, F. A. (2010). Does the method of instruction matter? An experimental examination of information literacy instruction in the online, blended, and face-to-face classrooms. *The Journal of Academic Librarianship, 36*(6), 495–500. doi:10.1016/j.acalib.2010.08.005
- Andretta, S. (2005). *Information literacy: A practitioner's guide*. Oxford, United Kingdom: Chandos.
- Association of College and Research Libraries (ACRL). (1989). *Presidential committee on information literacy: Final report*. Retrieved from <http://www.ala.org/acrl/publications/whitepapers/presidential>
- Association of College and Research Libraries (ACRL). (2000). *Information literacy competency standards for higher education*. Retrieved from <http://www.ala.org/acrl/standards/informationliteracycompetency>
- Association of College and Research Libraries (ACRL). (2010). *Psychology information literacy standards*. Retrieved from http://www.ala.org/acrl/standards/psych_info_lit
- Baacke, D. (1998). *Zum Konzept und zur Operationalisierung von Medienkompetenz*. Retrieved from http://www.produktive-medienarbeit.de/ressourcen/bibliothek/fachartikel/baacke_operationalisierung.shtml
- Balceris, M. (2011). *Medien- und Informationskompetenz. Modellierung und Messung von Informationskompetenz bei Schülern* (Doctoral thesis, Universität Paderborn, Paderborn, Germany). Retrieved from urn:nbn:de:hbz:466:2-8199
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215. doi:10.1037/0033-295X.84.2.191
- Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 37*(2), 122–147. doi:10.1037/0003-066X.37.2.122
- Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology, 52*(1), 1–26. doi:10.1146/annurev.psych.52.1.1
- Bawden, D. (2001). Information and digital literacies: A review of concepts. *Journal of Documentation, 57*(2), 218–259. doi:10.1108/EUM0000000007083
- Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982). ASK for information retrieval: Part I. Background and theory. *Journal of Documentation, 38*(2), 61–71. doi:10.1108/eb026722

- Bernard, R. M., Borokhovski, E., Schmid, R. F., Tamim, R. M., & Abrami, P. C. (2014). A meta-analysis of blended learning and technology use in higher education: From the general to the applied. *Journal of Computing in Higher Education*, 26(1), 87–122. doi:10.1007/s12528-013-9077-3
- Boon, S., Johnston, B., & Webber, S. (2007). A phenomenographic study of English faculty's conceptions of information literacy. *Journal of Documentation*, 63(2), 204–228. doi:10.1108/00220410710737187
- Bortz, J. (1999). *Statistik für Sozialwissenschaftler* (5th ed.). Berlin, Germany: Springer.
- Bowles-Terry, M. (2012). Library instruction and academic success: A mixed-methods assessment of a library instruction program. *Evidence Based Library and Information Practice*, 7(1), 82–95. Retrieved from <http://ejournals.library.ualberta.ca/index.php/EBLIP/article/view/12373>
- Bren, B., Hillemann, B., & Topp, V. (1998). Effectiveness of hands-on instruction of electronic resources. *Research Strategies*, 16(1), 41–51. doi:10.1016/S0734-3310(98)90005-2
- Bruce, C. S. (n.d.). *Seven faces of information literacy in higher education*. Retrieved from <http://www.christinebruce.com.au/informed-learning/seven-faces-of-information-literacy-in-higher-education/>
- Bruce, C. S. (1997). *The seven faces of information literacy*. Adelaide, Australia: Auslib Press.
- Buchner, A., & Brandt, M. (2005). Gedächtniskonzeptionen und Wissensrepräsentationen. In J. Müseler & W. Prinz (Eds.), *Allgemeine Psychologie* (1st ed., pp. 493–543). Heidelberg, Germany: Spektrum.
- Bühl, A. (2008). *SPSS 16: Einführung in die moderne Datenanalyse* (11th ed.). München, Germany: Pearson.
- Bühner, M., & Ziegler, M. (2012). *Statistik für Psychologen und Sozialwissenschaftler* (3rd ed.). München, Germany: Pearson.
- Chang, Y.-k., Zhang, X., Mokhtar, I. A., Foo, S., Majid, S., Luyt, B., & Theng, Y.-l. (2012). Assessing students' information literacy skills in two secondary schools in Singapore. *Journal of Information Literacy*, 6(2). doi:10.11645/6.2.1694
- Chen, H.-L. (2009). An analysis of undergraduate students' search behaviors in an information literacy class. *Journal of Web Librarianship*, 3(4), 333–347. doi:10.1080/19322900903328807
- Deary, I. J., Egan, V., Gibson, G. J., Austin, E. J., Brand, C. R., & Kellaghan, T. (1996). Intelligence and the differentiation hypothesis. *Intelligence*, 23(2), 105–132. doi:10.1016/S0160-2896(96)90008-2
- Deutscher Bibliotheksverband (DBV). (2009). *Standards der Informationskompetenz für Studierende*. Retrieved from http://www.bibliotheksverband.de/fileadmin/user_upload/Kommissionen/Kom_Dienstleistung/Publikationen/Standards_Infokompetenz_03.07.2009_endg.pdf
- Donnelly, K. (2000). Reflections on what happens when librarians become teachers. *Computers in Libraries*, 20(3), 46–49.
- Dunn, K. (2002). Assessing information literacy skills in the California State University: A progress report. *The Journal of Academic Librarianship*, 28(1-2), 26–35. doi:10.1016/S0099-1333(01)00281-6
- Eid, M., & Diener, E. (2006). The need for multimethod measurement in Psychology. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in Psychology* (1st ed., pp. 3–8). Washington, DC: American Psychological Association.
- Eisenberg, M. B. & Berkowitz, R. E. (2014). *Big6 skills overview*. Retrieved from <http://big6.com/pages/about/big6-skills-overview.php>
- Fentz, H. N., Arendt, M., O'Toole, M. S., Hoffart, A., & Hougaard, E. (2014). The mediational role of panic self-efficacy in cognitive behavioral therapy for panic disorder: A systematic review and meta-analysis. *Behaviour Research and Therapy*, 60, 23–33. doi:10.1016/j.brat.2014.06.003

- Ferguson, G. A. (1954). On learning and human ability. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 8(2), 95–112. doi:10.1037/h0083598
- Fisseni, H.-J. (1997). *Lehrbuch der psychologischen Diagnostik* (2nd ed.). Göttingen, Germany: Hogrefe.
- Fister, B., & Eland, T. (2008). Curriculum issues in information literacy instruction. In C. N. Cox & E. B. Lindsay (Eds.), *Information literacy instruction handbook* (pp. 94–112). Chicago, IL: Association of College and Research Libraries.
- Foster, S. (1993). Information literacy: Some misgivings. *American Libraries*, 24(4), 344–346.
- Franke, F., Klein, A., & Schüller-Zwierlein, A. (2010). *Schlüsselkompetenzen: Literatur recherchieren in Bibliotheken und Internet*. Stuttgart, Germany: Metzler.
- Franke, F. & Schüller-Zwierlein, A. (2008). *Wie informationskompetent sind die bayerischen Studierenden im Jahr 2007?* Retrieved from http://www.bsb-muenchen.de/fileadmin/imageswww/pdf-dateien/bibliotheksforum/2008-1/BFB_0108_12-Franke.pdf
- Gapski, H. (2012). Informations- und Medienkompetenz aus Sicht der Kommunikations- und Medienwissenschaft. In W. Sühl-Strohmer (Ed.), *Handbuch Informationskompetenz* (pp. 167–175). Berlin, Germany: de Gruyter Saur.
- Gapski, H. & Tekster, T. (2009). *Informationskompetenz in Deutschland: Überblick zum Stand der Fachdiskussion und Zusammenstellung von Literaturangaben, Projekten und Materialien zu einzelnen Zielgruppen*. Retrieved from http://www.bui.haw-hamburg.de/pers/ursula.schulz/teaching_library_2/informationskompetenz_in_deutschland.pdf
- Garrison, D. R., & Kanuka, H. (2004). Blended learning: Uncovering its transformative potential in higher education. *The Internet and Higher Education*, 7(2), 95–105. doi:10.1016/j.iheduc.2004.02.001
- Giustini, D., & Kamel Boulos, M. N. (2013). Google Scholar is not enough to be used alone for systematic reviews. *Online Journal of Public Health Informatics*, 5(2). doi:10.5210/ojphi.v5i2.4623
- Goldhammer, F., Kröhne, U., Keßel, Y., Senkbeil, M., & Ihme, J. M. (2014). Diagnostik von ICT-Literacy. *Diagnostica*, 60(1), 10–21. doi:10.1026/0012-1924/a000113
- Google Inc. (2011). *Google Scholar-Hilfe*. Retrieved from <http://scholar.google.de/intl/de/scholar/help.html>
- Grafstein, A. (2002). A discipline-based approach to information literacy. *The Journal of Academic Librarianship*, 28(4), 197–204. doi:10.1016/S0099-1333(02)00283-5
- Grassian, E. S., & Kaplowitz, J. R. (2001). *Information literacy instruction: Theory and practice*. New York, NY: Neal-Schuman.
- Hancock, V. E. (1993). *Information literacy for lifelong learning. ERIC digest*. Retrieved from <http://files.eric.ed.gov/fulltext/ED358870.pdf>
- Hapke, T. (2014, March 6). *"Threshold concepts" statt Standards zur Informationskompetenz* [Web log post]. Retrieved from <http://blog.hapke.de/information-literacy/threshold-concepts-statt-standards-zur-informationskompetenz/>
- Harris, J. A., Vernon, P. A., & Jang, K. L. (2005). Testing the differentiation of personality by intelligence hypothesis. *Personality and Individual Differences*, 38(2), 277–286. doi:10.1016/j.paid.2004.04.007
- Hartig, J., Frey, A., & Jude, N. (2008). Validität. In H. Moosbrugger & A. Kelava (Eds.), *Testtheorie und Fragebogenkonstruktion* (pp. 135–163). Berlin, Germany: Springer.
- Hartig, J., & Klieme, E. (2006). Kompetenz und Kompetenzdiagnostik. In K. Schweizer (Ed.), *Leistung und Leistungsdiagnostik* (pp. 127–143). Berlin, Germany: Springer.
- Hartmann, D. P. (1977). Considerations in the choice of interobserver reliability estimates. *Journal of Applied Behavior Analysis*, 10, 103–116. doi:10.1901/jaba.1977.10-103

- Hausknecht, J. P., Halpert, J. A., Di Paolo, N. T., & Moriarty Gerrard, M. O. (2007). Retesting in selection: A meta-analysis of coaching and practice effects for tests of cognitive ability. *Journal of Applied Psychology, 92*(2), 373–385. doi:10.1037/0021-9010.92.2.373
- Heierle, L. (2006). *Schlüsselqualifikationen in der universitären Lehre am Beispiel des Fachs Geographie. Theorie, empirische Untersuchung und konzeptionelle Überlegungen* (Doctoral thesis, Universität Basel, Basel, Switzerland). Retrieved from <http://edoc.unibas.ch/410/>
- Heinze, N. (2008). *Bedarfsanalyse für das Projekt i-literacy: Empirische Untersuchung der Informationskompetenz der Studierenden der Universität Augsburg*. Retrieved from http://opus.bibliothek.uni-augsburg.de/opus4/files/685/imb_Arbeitsbericht_19.pdf
- Heinze, N., Fink, J., & Wolf, S. (2009). *Informationskompetenz und wissenschaftliches Arbeiten: Studienergebnisse und Empfehlungen zur wissenschaftlichen Recherche im Hochschulstudium*. Retrieved from http://opus.bibliothek.uni-augsburg.de/opus4/files/1269/imb_Arbeitsbericht_21.pdf
- Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in Mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge. The case of Mathematics* (1st ed., pp. 1–27). Hillsdale, NJ: Lawrence Erlbaum.
- Hochschulrektorenkonferenz (HRK). (2012). *Hochschule im digitalen Zeitalter: Informationskompetenz neu begreifen – Prozesse anders steuern*. Retrieved from http://www.hrk.de/uploads/media/Entschliessung_Informationskompetenz_20112012_01.pdf
- Homann, B. (2000). Dynamisches Modell der Informationskompetenz (DYMIK) – Didaktisch-methodische Grundlage der Vermittlung von Methodenkompetenzen an der UB Heidelberg. *Theke, 86–93*. Retrieved from <http://journals.ub.uni-heidelberg.de/index.php/Theke/article/view/107/86>
- Homann, B. (2001, August). *Schwierigkeiten und neue Ansätze der Benutzerschulung in Deutschland*. Paper presented at the 67th IFLA Council and General Conference, Boston, MA. Retrieved from <http://archive.ifla.org/IV/ifla67/papers/072-126g.pdf>
- Homann, B. (2002). Standards der Informationskompetenz. *Bibliotheksdienst, 36*(5), 625–633.
- Homann, B. (2003). German libraries at the starting line for the new task of teaching information literacy. *Library Review, 52*(7), 310–318. doi:10.1108/00242530310487407
- Homann, B. (2007). Standards und Modelle der Informationskompetenz – Kooperationsgrundlage für bibliothekarische Schulungsaktivitäten. In U. Krauß-Leichert (Ed.), *Teaching Library - eine Kernaufgabe für Bibliotheken*. (pp. 81–99). Frankfurt, Germany: Lang.
- Homann, B. (2015). Bibliothekarische Aktivitäten zur Vermittlung von Informationskompetenz. In A.-K. Mayer (Ed.), *Informationskompetenz im Hochschulkontext – Interdisziplinäre Forschungsperspektiven* (pp. 165–178). Lengerich, Germany: Pabst.
- Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology, 57*(5), 253–270. doi:10.1037/h0023816
- Ingold, M. (2005). Das bibliothekarische Konzept der Informationskompetenz. *Berliner Handreichungen zur Bibliothekswissenschaft, 128*. Retrieved from <http://www.ib.hu-berlin.de/~kumlau/handreichungen/h128/>
- Isaacson, D. (2003). Let's talk libraries, not "Information Literacy". *Library Journal, 128*(19), 42.
- Isbell, D., & Kammerlocher, L. (1998). Implementing Kuhlthau: A new model for library and reference instruction. *Reference Services Review, 26*(3/4), 33–44. doi:10.1108/00907329810307722
- Jankisz, E., & Moosbrugger, H. (2008). Planung und Entwicklung von psychologischen Tests und Fragebogen. In H. Moosbrugger & A. Kelava (Eds.), *Testtheorie und Fragebogenkonstruktion* (pp. 27–72). Berlin, Germany: Springer.
- Jensen, A. R. (2003). Regularities in Spearman's law of diminishing returns. *Intelligence, 31*(2), 95–105. doi:10.1016/S0160-2896(01)00094-0

- Jerusalem, M. & Schwarzer, R. (1999). *Allgemeine Selbstwirksamkeitserwartung (SWE)*. Retrieved from <http://userpage.fu-berlin.de/health/germscal.htm>
- Johnston, B., & Webber, S. (2003). Information literacy in higher education: A review and case study. *Studies in Higher Education, 28*(3), 335–352. doi:10.1080/03075070309295
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review, 2*(2), 130–144. doi:10.1016/j.edurev.2007.05.002
- Judge, T. A., & Bono, J. E. (2001). Relationship of core self-evaluations traits – self-esteem, generalized self-efficacy, locus of control, and emotional stability – with job satisfaction and job performance: A meta-analysis. *Journal of Applied Psychology, 86*(1), 80–92. doi:10.1037/0021-9010.86.1.80
- Julien, H. (2000). Information literacy instruction in Canadian academic libraries: Longitudinal trends and international comparisons. *College & Research Libraries, 61*(6), 510–523. doi:10.5860/crl.61.6.510
- Julien, H., & Barker, S. (2009). How high-school students find and evaluate scientific information: A basis for information literacy skills development. *Library & Information Science Research, 31*(1), 12–17. doi:10.1016/j.lisr.2008.10.008
- Karnilowicz, W. (2012). A comparison of self-assessment and tutor assessment of undergraduate Psychology students. *Social Behavior and Personality: An international journal, 40*(4), 591–604. doi:10.2224/sbp.2012.40.4.591
- Katz, I. R. (2007). Testing information literacy in digital environments: ETS's iSkills assessment. *Information Technology and Libraries, 26*(3), 3–12. doi:10.6017/ital.v26i3.3271
- Kelly, G. A. (1986). *Die Psychologie der persönlichen Konstrukte* (E. Danzinger-Tholen, Trans.). Paderborn, Germany: Junfermann. (Original work published 1955)
- Kennedy, E. J., Lawton, L., & Plumlee, E. L. (2002). Blissful ignorance: The problem of unrecognized incompetence and academic performance. *Journal of Marketing Education, 24*(3), 243–252. doi:10.1177/0273475302238047
- Kerres, M. (2012). *Mediendidaktik: Konzeption und Entwicklung mediengestützter Lernangebote* (3rd ed.). München, Germany: Oldenbourg.
- Kim, J. (2009). Describing and predicting information-seeking behavior on the web. *Journal of the American Society for Information Science and Technology, 60*(4), 679–693. doi:10.1002/asi.21035
- Klatt, R., Gavriilidis, K., Kleinsimlinghaus, K., & Feldmann, M. (2001). *Nutzung elektronischer wissenschaftlicher Information in der Hochschulausbildung: Barrieren und Potenziale der innovativen Mediennutzung im Lernalltag der Hochschulen*. Retrieved from <http://opus.bsz-bw.de/hdms/volltexte/2004/334/pdf/NutzungwissInfo.pdf>
- Kline, P. (1998). *The handbook of psychological testing*. London, United Kingdom: Routledge.
- Knight, L. A. (2006). Using rubrics to assess information literacy. *Reference Services Review, 34*(1), 43–55. doi:10.1108/00907320610640752
- Kommission der europäischen Gemeinschaften. (2001). *Einen europäischen Raum des lebenslangen Lernens schaffen*. Retrieved from http://www.bildungspolitik.de/uploads/dokumente_informelles_lernen/2001_11_kom_III.pdf
- Kraemer, E. W., Lombardo, S. V., & Lepkowski, F. J. (2007). The librarian, the machine, or a little of both: A comparative study of three information literacy pedagogies at Oakland University. *College & Research Libraries, 68*(4), 330–342. doi:10.5860/crl.68.4.330
- Krüger, N. (2008). LOTSE: Ein ganzheitlicher Ansatz zur Online-Vermittlung von Informationskompetenz. In E. Pipp (Ed.), *Informationskonzepte für die Zukunft: ODOK '07* (pp. 71–81). Graz, Austria: Neugebauer.

- Krüger, N. (2014). *Gemeinsame Schulungsstatistik*. Retrieved from <http://www.informationskompetenz.de/veranstaltungsstatistik/>
- Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science and Technology*, 42(5), 361–371.
- Kuhlthau, C. C. (1993). *Seeking meaning: A process approach to library and information services. Information management, policy, and services*. Norwood, NJ: Ablex.
- Kurbanoglu, S. S., Akkoyunlu, B., & Umay, A. (2006). Developing the information literacy self-efficacy scale. *Journal of Documentation*, 62(6), 730–743. doi:10.1108/00220410610714949
- Latham, G. P., Locke, E. A., & Fassin, N. E. (2002). The high performance cycle: Standing the test of time. In S. Sonnentag (Ed.), *Psychological management of individual performance* (pp. 201–228). Chichester, United Kingdom: Wiley.
- Lauber-Reymann, M. (2010). *Informationsressourcen: Ein Handbuch für Bibliothekare und Informationsspezialisten*. Berlin, Germany: de Gruyter.
- Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID). (2012). *Ausgewählte wissenschaftliche psychologische Informationsquellen und ihre Webzugänge*. Retrieved from http://www.zpid.de/pub/info/lehre_webzugang.pdf
- Leichner, N. (2015). Multimethodale Erfassung von Informationskompetenz. In A.-K. Mayer (Ed.), *Informationskompetenz im Hochschulkontext – Interdisziplinäre Forschungsperspektiven* (pp. 83–102). Lengerich, Germany: Pabst.
- Leichner, N., Peter, J., Mayer, A.-K., & Krampen, G. (2013a). Assessing information literacy among German psychology students. *Reference Services Review*, 41(4), 660–674. doi:10.1108/RSR-11-2012-0076
- Leichner, N., Peter, J., Mayer, A.-K., & Krampen, G. (2013b). Erfassung von Wissen über Informationsrecherchen: Konzeptuelle Überlegungen und empirische Befunde. *B.I.T. Online*, 16(4), 298–306.
- Leichner, N., Peter, J., Mayer, A.-K., & Krampen, G. (2014a). Assessing information literacy programmes using information search tasks. *Journal of Information Literacy*, 8(1). doi:10.11645/8.1.1870
- Leichner, N., Peter, J., Mayer, A.-K., & Krampen, G. (2014b). Fostering information literacy in German Psychology students using a Blended Learning approach. *Proceedings of the 6th International Conference on Computer Supported Education*, 353–359. doi:10.5220/0004795103530359
- Leichner, N., Peter, J., Mayer, A.-K., & Krampen, G. (2014c). Google Scholar versus Fachdatenbanken: Wie erfolgreich nutzen Studierende verschiedene Suchwerkzeuge? In M. Ockenfeld (Ed.), *Informationsqualität und Wissensgenerierung: 3. DGI-Konferenz, 66. Jahrestagung der DGI; Proceedings* (pp. 37–46). Frankfurt am Main, Germany: DGI.
- Levy, Y. (2007). Comparing dropouts and persistence in e-learning courses. *Computers & Education*, 48(2), 185–204. doi:10.1016/j.compedu.2004.12.004
- Lievens, F., Peeters, H., & Schollaert, E. (2008). Situational judgment tests: A review of recent research. *Personnel Review*, 37(4), 426–441. doi:10.1108/00483480810877598
- Locke, E. A., & Latham, G. P. (1990a). Work motivation and satisfaction: Light at the end of the tunnel. *Psychological Science*, 1(4), 240–246. doi:10.1111/j.1467-9280.1990.tb00207.x
- Locke, E. A., & Latham, G. P. (1990b). Work motivation. The high performance cycle. In U. Kleinbeck, H.-H. Quast, H. Thierry, & H. Häcker (Eds.), *Work motivation* (pp. 3–26). Hillsdale, NJ: Lawrence Erlbaum.
- Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, 57(9), 705–717. doi:10.1037/0003-066X.57.9.705

- López-Pérez, M. V., Pérez-López, M. C., & Rodríguez-Ariza, L. (2011). Blended learning in higher education: Students' perceptions and their relation to outcomes. *Computers & Education, 56*(3), 818–826. doi:10.1016/j.compedu.2010.10.023
- Lowe, C. A., & Eisenberg, M. B. (2006). Big6 skills for information literacy. In K. E. Fisher, S. Erdelez, & L. McKechnie (Eds.), *Theories of information behavior* (2nd ed.). Medford, NJ: Information Today.
- Maddux, J. E. (1995). Self-efficacy theory. An introduction. In J. E. Maddux (Ed.), *Self-efficacy, adaptation, and adjustment. Theory, research, and application* (pp. 3–33). New York, NY: Plenum Press.
- March, J. S., Sullivan, K., & Parker, J. (1999). Test-retest reliability of the multidimensional anxiety scale for children. *Journal of Anxiety Disorders, 13*(4), 349–358. doi:10.1016/S0887-6185(99)00009-2
- Mayer, R. E. (2003). What causes individual differences in cognitive performance? In R. J. Sternberg & E. L. Grigorenko (Eds.), *The psychology of abilities, competencies, and expertise* (pp. 263–273). Cambridge, United Kingdom: Cambridge University Press.
- McClelland, D. C. (1973). Testing for competence rather than for "intelligence". *American Psychologist, 28*(1), 1–14. doi:10.1037/h0034092
- McCrack, L. J. (1991). Information literacy: A bogus bandwagon? *Library Journal, 116*(8), 38–42.
- McCulley, C. (2009). *Mixing and matching: Assessing information literacy*. Retrieved from http://digitalcommons.linfield.edu/cgi/viewcontent.cgi?article=1001&context=librariesfac_pubs
- Means, B., Toyama, Y., Murphy, R., & Baki, M. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature. *Teachers College Record, 115*(3), 1–47.
- Mery, Y., Newby, J., & Peng, K. (2012). Why one-shot information literacy sessions are not the future of instruction: A case for online credit courses. *College & Research Libraries, 73*(4), 366–377. doi:10.5860/crl-271
- Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine, 65*(9), S63–S67.
- Mirabile, R. J. (1997). Everything you wanted to know about competency modeling. *Training & Development, 51*(8), 73–77.
- Moosbrugger, H., & Kelava, A. (2008). Qualitätsanforderungen an einen psychologischen Test (Testgütekriterien). In H. Moosbrugger & A. Kelava (Eds.), *Testtheorie und Fragebogenkonstruktion* (pp. 7–26). Berlin, Germany: Springer.
- Moskal, B. M. (2000). Scoring rubrics: What, when and how? *Practical Assessment, Research & Evaluation, 7*(3). Retrieved from <http://PAREonline.net/getvn.asp?v=7&n=3>
- Noe, N. W., & Bishop, B. A. (2005). Assessing Auburn University Library's tiger information literacy tutorial (TILT). *Reference Services Review, 33*(2), 173–187. doi:10.1108/00907320510597372
- Novotny, E., & Cahoy, E. S. (2006). If we teach, do they learn? The impact of instruction on online catalog search strategies. *portal: Libraries and the Academy, 6*(2), 155–167. doi:10.1353/pla.2006.0027
- Oakleaf, M. (2008). Dangers and opportunities: A conceptual map of information literacy assessment approaches. *portal: Libraries and the Academy, 8*(3), 233–253. doi: 10.1353/pla.0.0011
- Oakleaf, M. (2009). Using rubrics to assess information literacy: An examination of methodology and interrater reliability. *Journal of the American Society for Information Science and Technology, 60*(5), 969–983. doi:10.1002/asi.21030
- Ohloff, R. (2012). Förderung von Informationskompetenz als Schlüsselqualifikation: Modell Freiburg im Breisgau. In W. Sühl-Strohmenger (Ed.), *Handbuch Informationskompetenz* (pp. 250–260). Berlin, Germany: de Gruyter Saur.

- Ondrusek, A., Dent, V. F., Bonadie-Joseph, I., & Williams, C. (2005). A longitudinal study of the development and evaluation of an information literacy test. *Reference Services Review*, 33(4), 388–417. doi:10.1108/00907320510631544
- Owusu-Ansah, E. K. (2003). Information literacy and the academic library: A critical look at a concept and the controversies surrounding it. *The Journal of Academic Librarianship*, 29(4), 219–230. doi:10.1016/S0099-1333(03)00040-5
- Peter, J. (2015). Lehre von Informationskompetenz im Fach Psychologie: Konzeptionelle und strategische Überlegungen sowie empirische Befunde. In A.-K. Mayer (Ed.), *Informationskompetenz im Hochschulkontext – Interdisziplinäre Forschungsperspektiven* (pp. 79–194). Lengerich, Germany: Pabst.
- Piaget, J. (1978). *Success and understanding* (A. J. Pomerans, Trans.). Cambridge, MA: Harvard University Press. (Original work published 1974)
- Pinto, M. (2015). Viewing and exploring the subject area of information literacy assessment in higher education (2000–2011). *Scientometrics*, 102(1), 227–245. doi:10.1007/s11192-014-1440-2
- Potter, W. J. (2010). The state of media literacy. *Journal of Broadcasting & Electronic Media*, 54(4), 675–696. doi:10.1080/08838151.2011.521462
- Project SAILS (Standardized Assessment of Information Literacy Skills). (2000-2014). SAILS Test [Measurement instrument]. Retrieved from <https://www.projectsails.org/>
- Rauchmann, S. (2002). *Die Vermittlung von Informationskompetenz in Online-Tutorials: Eine vergleichende Bewertung der US-amerikanischen und deutschen Konzepte* (Diplomarbeit, Fachhochschule Potsdam, Potsdam, Germany). Retrieved from <http://fiz1.fh-potsdam.de/volltext/diplome/05100.pdf>
- Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, 138(2), 353–387. doi:10.1037/a0026838
- Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in Mathematics: An iterative process. *Journal of Educational Psychology*, 93(2), 346–362. doi:10.1037//0022-0663.93.2.346
- Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology*, 99(3), 561–574. doi:10.1037/0022-0663.99.3.561
- Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, 130(2), 261–288. doi:10.1037/0033-2909.130.2.261
- Romero, M., & Barberà, E. (2011). Quality of e-learners' time and learning performance beyond quantitative time-on-task. *The International Review of Research in Open and Distance Learning*, 12(5), 125–137. Retrieved from <http://www.irrodl.org/index.php/irrodl/article/view/999>
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion* (2nd ed.). Bern, Switzerland: Huber.
- Schermelleh-Engel, K., & Werner, C. (2008). Methoden der Reliabilitätsbestimmung. In H. Moosbrugger & A. Kelava (Eds.), *Testtheorie und Fragebogenkonstruktion* (pp. 113–133). Berlin, Germany: Springer.
- Schmidt, A. M., & DeShon, R. P. (2010). The moderating effects of performance ambiguity on the relationship between self-efficacy and performance. *Journal of Applied Psychology*, 95(3), 572–581. doi:10.1037/a0018289
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8(4), 350–353. doi:10.1037/1040-3590.8.4.350

- Schneider, M., Rittle-Johnson, B., & Star, J. R. (2011). Relations among conceptual knowledge, procedural knowledge, and procedural flexibility in two samples differing in prior knowledge. *Developmental psychology*, 47(6), 1525–1538. doi:10.1037/a0024997
- Schneider, M., & Stern, E. (2005). Conceptual and procedural knowledge of a mathematics problem: Their measurement and their causal interrelations. In B. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th annual conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum. Retrieved from <http://www.psych.unito.it/csc/cogsci05/frame/talk/f610-schneider.pdf>
- Schwarzer, R., & Jerusalem, M. (2002). Das Konzept der Selbstwirksamkeit. In M. Jerusalem & D. Hopf (Eds.), *Zeitschrift für Pädagogik: Beiheft 44. Selbstwirksamkeit und Motivationsprozesse in Bildungsinstitutionen* (pp. 28–53). Weinheim, Germany: Beltz.
- Scott, T. J., & O'Sullivan, M. K. (2005). Analyzing student search strategies: Making a case for integrating information literacy skills into the curriculum. *Teacher Librarian*, 33(1), 21–25.
- Shavelson, R. J. (2010). On the measurement of competency. *Empirical research in vocational education and training*, 2(1), 41–63.
- Shippmann, J. S., Ash, R. A., Battista, M., Carr, L., Eyde, L. D., Hesketh, B., . . . Sanchez, J. I. (2000). The practice of competency modeling. *Personnel Psychology*, 53(3), 703–740. doi:10.1111/j.1744-6570.2000.tb00220.x
- Sireci, S. G., & Zenisky, A. L. (2006). Innovative item formats in computer-based testing: In pursuit of improved construct representation. In S. M. Downing & T. S. Haladyna (Eds.), *Handbook of test development* (pp. 329–348). Mahwah, NJ: Lawrence Erlbaum.
- Sitzmann, T., & Yeo, G. (2013). A meta-analytic investigation of the within-person self-efficacy domain: Is self-efficacy a product of past performance or a driver of future performance? *Personnel Psychology*, 66(3), 531–568. doi:10.1111/peps.12035
- Society of College, National and University Libraries (SCONUL). (1999). *Information skills in higher education: A SCONUL position paper*. Retrieved from http://392274175.webhosting.wanadoo.nl/informationSkillsUK_SCONUL.pdf
- Society of College, National and University Libraries (SCONUL). (2011). *The SCONUL Seven Pillars of information literacy*. Retrieved from <http://www.sconul.ac.uk/sites/default/files/documents/coremodel.pdf>
- Sonley, V., Turner, D., Myer, S., & Cotton, Y. (2007). Information literacy assessment by portfolio: A case study. *Reference Services Review*, 35(1), 41–70. doi:10.1108/00907320710729355
- Soria, K. M., Fransen, J., & Nackerud, S. (2014). Stacks, serials, search engines, and students' success: First-year undergraduate students' library use, academic achievement, and retention. *The Journal of Academic Librarianship*, 40(1), 84–91. doi:10.1016/j.acalib.2013.12.002
- Stajkovic, A. D., & Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin*, 124(2), 240–261. doi:10.1037/0033-2909.124.2.240
- Swain, M., Sundre, D. L., & Clarke, K. (2014). *The Information Literacy Test (ILT): Test manual*. Retrieved from http://www.madisonassessment.com/uploads/ILT%20Test%20Manual%20May%202014%20pdf_3.pdf
- Tarrant, M., Ware, J., & Mohammed, A. M. (2009). An assessment of functioning and non-functioning distractors in multiple-choice questions: A descriptive analysis. *BMC Medical Education*, 9(1), 40–48. doi:10.1186/1472-6920-9-40
- Taylor, R. S. (1968). Question-negotiation and information seeking in libraries. *College & Research Libraries*, 29(3), 178–194. doi:10.5860/crl_29_03_178
- Thompson, S. (2003). *Information literacy meeting of experts: Prague, The Czech Republic September 20-23, 2003*. Retrieved from

http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/themes/info_lit_meeting_prague_2003.pdf

- Timmers, C. F., & Glas, C. A. W. (2010). Developing scales for information-seeking behaviour. *Journal of Documentation*, 66(1), 46–69. doi:10.1108/00220411011016362
- Tung, R. (2010). *Including performance assessments in accountability systems: A review of scale up efforts*. Retrieved from <http://files.eric.ed.gov/fulltext/ED509787.pdf>
- van Gelder, T. (2005). Teaching critical thinking: Some lessons from cognitive science. *College Teaching*, 53(1), 41–46. doi: 10.3200/CTCH.53.1.41-48
- Walsh, A. (2009). Information literacy assessment: Where do we start? *Journal of Librarianship and Information Science*, 41(1), 19–28. doi:10.1177/0961000608099896
- Walton, G., & Hepworth, M. (2013). Using assignment data to analyse a blended information literacy intervention: A quantitative approach. *Journal of Librarianship and Information Science*, 45(1), 53–63. doi:10.1177/0961000611434999
- Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45–65). Seattle, WA: Hogrefe & Huber.
- Weinert, F. E. (2014). Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Ed.), *Leistungsmessungen in Schulen* (3rd ed., pp. 17–32). Weinheim, Germany: Beltz.
- Weisel, L. (2015). Informationskompetenz im Hochschulkontext – Initiativen und Programme. In A.-K. Mayer (Ed.), *Informationskompetenz im Hochschulkontext – Interdisziplinäre Forschungsperspektiven* (pp. 63–82). Lengerich, Germany: Pabst.
- Willingham, D. B., Nissen, M. J., & Bullemer, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(6), 1047–1060. doi:10.1037/0278-7393.15.6.1047
- Wilson, T. D. (1999). Models in information behaviour research. *Journal of Documentation*, 55(3), 249–270. doi:10.1108/EUM0000000007145
- Wilson, T. D. (2000). Human information behavior. *Informing science*, 3(2), 49–56. Retrieved from <http://www.informingscience.org/Journals/InformingSciJ/Articles>

16 Anhang

16.1 Multiple-Choice-Test

Die richtigen Antwortoptionen sind fett markiert.

1. 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. Welche Aussagen über die Recherche in Google Scholar und in Fachdatenbanken sind korrekt?
 - Google Scholar verwendet Volltextsuche, d. h. es wird überprüft, ob der Suchbegriff in einem Dokument enthalten ist.**
 - Mithilfe von Fachdatenbanken können Metadaten (Abstracts, Schlagwörter, Titel etc.) durchsucht werden.**
 - Google Scholar bietet Volltextzugriff zu allen Zeitschriftenartikeln.

3. Bei der Recherche zu einem Thema sind Sie auf ein Buch gestoßen, das sieben Jahre alt ist. In Fachdatenbanken sind Sie außerdem auf aktuellere Zeitschriftenartikel gestoßen. Welche Gründe gibt es, dennoch zuerst das Buch zu lesen?
 - Das Buch vermittelt wahrscheinlich einen besseren Überblick über das Thema und verschiedene Forschungsrichtungen.**
 - Bücher werden erst dann geschrieben, wenn Wissen gesichert ist.
 - Zeitschriftenartikel behandeln oft einen Teilaspekt des gesamten Themas, der für mich möglicherweise nicht relevant ist.**

4. Sie suchen Informationen zu einem bestimmten Thema, beispielsweise für einen Vortrag. Welches sind angemessene Vorgehensweisen, um Suchbegriffe zu identifizieren?
 - Das Thema in einigen Sätzen zusammenfassen. Die zentralen Begriffe der Zusammenfassung können als Suchbegriffe verwendet werden.**
 - Nach Synonymen für die zentralen Begriffe suchen, beispielsweise durch Brainstorming oder die Verwendung eines Lexikons.**
 - Einen Zeitschriftenartikel zu dem Thema lesen und verwandte Konzepte oder Ideen notieren.**

5. Welche Antwortmöglichkeiten sind korrekt?

Der SFX-Button in der Ergebnisansicht einer Fachdatenbank kann verwendet werden, um...

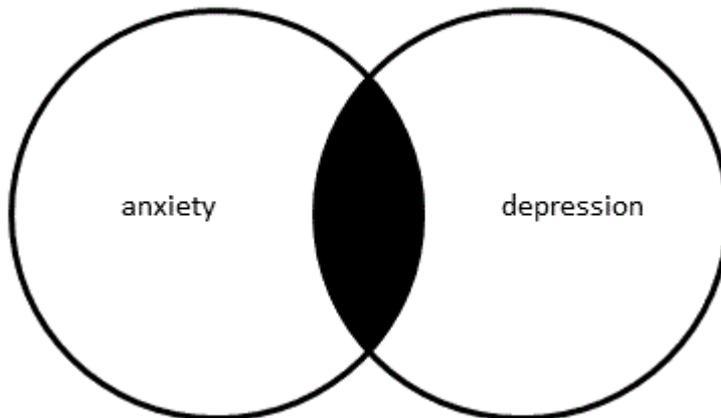
 - herauszufinden, ob eine bestimmte Zeitschrift in der (Universitäts-)Bibliothek verfügbar ist.**

- **herauszufinden, ob der fragliche Zeitschriftenartikel für Mitglieder dieser Einrichtung online verfügbar ist.**
 - **herauszufinden, ob ein bestimmtes Buch in der (Universitäts-)Bibliothek vorhanden ist.**
6. Welche Probleme treten beim Peer-Review von Zeitschriftenartikeln auf?
- Keine, deshalb ist Peer-Review der Königsweg.
 - **Reviewer können dazu tendieren, solche Zeitschriftenartikel, die ihren Befunden widersprechen, abzulehnen. Neue abweichende Befunde werden benachteiligt.**
 - **Es werden oft nicht ausreichend qualifizierte Reviewer gefunden.**
7. Welche Gründe gibt es, die Reputation des Autors/der Autorin nicht als Kriterium zur Bewertung von Publikationen zu verwenden?
- **Qualifizierte jüngere Autor/innen werden benachteiligt, weil sie noch keine Reputation erwerben konnten.**
 - **Auch unbekannte Autor/innen können gute Zeitschriftenartikel veröffentlichen, die dann übersehen werden.**
 - **Das Peer-Review-Verfahren garantiert (im Idealfall) eine gleichbleibende Qualität unabhängig vom Autor/der Autorin.**
8. Welche Aussagen über die Reihenfolge der Autor/innen einer Publikation sind korrekt?
- **In einigen Fällen wird die Reihenfolge nach Zufall festgelegt, damit keine/r der Autor/innen benachteiligt ist.**
 - **Oft stehen die Autor/innen, die viel zu der Studie beigetragen haben, an den vorderen Stellen.**
 - **Die etablierten Autor/innen stehen häufig an den hinteren Stellen.**
9. Welche Antwortmöglichkeit trifft zu? Die Reputation des Verlags ist vor allem wichtig bei der Beurteilung von...
- Zeitschriften, weil die Reputation wichtiger ist als der Journal Impact Factor.
 - **Monographien und Herausgeberwerken, weil in diesen Fällen kein Impact Factor berechnet wird.**
10. Wie kann die Reputation des Verlags festgestellt werden?
- **Andere (z. B. Kolleg/innen) fragen, ob sie Auskunft geben können.**
 - Informationen auf Wikipedia heranziehen.
 - **Eine/n wissenschaftliche/n Mitarbeiter/in oder eine/n Mitarbeiter/in der (Universitäts-)Bibliothek fragen.**
11. Welche Antwortmöglichkeit trifft zu? Der Journal Impact Factor (JIF) einer Zeitschrift gibt an...
- **wie oft Autor/innen Artikel aus der fraglichen Zeitschrift in einem bestimmten Zeitraum zitiert haben.**
 - wie viele (Universitäts-)Bibliotheken die Zeitschrift abonniert haben.
 - welche Relevanz der Zeitschrift von einem Expertengremium zugesprochen wurde.

12. Der Zugriff auf den JIF steht nicht in jeder (Universitäts-)Bibliothek (kostenfrei) zur Verfügung. Welche Optionen haben Studenten?
- Bei der Uni nach Kostenerstattung fragen.
 - **Eigenfactor.org ist eine frei zugängliche Alternative.**
 - **Die JIF-Werte sind auf den Homepages der Zeitschriften zu finden.**
13. Was unterscheidet Fachdatenbanken beispielsweise von einer Internetsuchmaschine wie Google Scholar?
- **Fachdatenbanken ermöglichen eine Thesaurussuche.**
 - In Fachdatenbanken ist es möglich, Boolesche Operatoren zur Präzision einer Suche zu verwenden.
 - **Die Reihenfolge der Treffer in der Ergebnisliste einer Fachdatenbank ist unabhängig davon, wie häufig die Artikel innerhalb der Datenbank angeklickt wurden.**
14. Eine Publikation wurde 50 Mal zitiert. Was kann daraus eindeutig geschlossen werden?
- Die Publikation wurde von der Fachwelt positiv aufgenommen.
 - Der Autor genießt eine hohe Reputation.
 - Der Inhalt der Publikation kann als zutreffend angesehen werden.
15. Mit welcher der drei genannten Möglichkeiten gelangt man am schnellsten zu einer Liste von Artikeln, die thematisch zu einem vorliegenden Artikel passen?
- **„Zitiert-Durch“-Suchfunktion in Google Scholar oder Fachdatenbanken.**
 - Durchsuchen der Literaturliste des vorliegenden Artikels.
 - Publikationsliste auf der Homepage des Erstautors.
16. DOIs ...
- sind bibliometrische Indizes, die zur Einschätzung des wissenschaftlichen Einflusses eines Autors genutzt werden.
 - **sind Bezeichner, die zur Identifikation von Publikationen genutzt werden.**
 - sind in ihrer Verwendung in der sozialwissenschaftlichen Fachwelt sehr umstritten.
17. Trunkierung eines Begriffs bedeutet unter anderem ...
- **dass bei einer Suche nach dem Wortstamm des Begriffs mit verschiedenen Endungen gesucht wird.**
 - dass bei mehrdeutigen Begriffen die Suche auf eine bestimmte Bedeutung beschränkt wird.
 - dass die Suche nach einem bestimmten Begriff durch verwandte Begriffe aus dem Thesaurus erweitert wird.
18. Welche der folgenden Aussagen treffen auf Google zu?
- **Wenn in einer Suchanfrage zwei durch ein Leerzeichen getrennte, blanke Begriffe nebeneinander stehen, werden die Begriffe automatisch durch AND miteinander verknüpft.**
 - **Genaue Wortgruppen können durch Anführungszeichen definiert werden.**

- **Durch die Ergänzung filetype:pdf kann die Suche auf PDF-Dokumente eingeschränkt werden.**

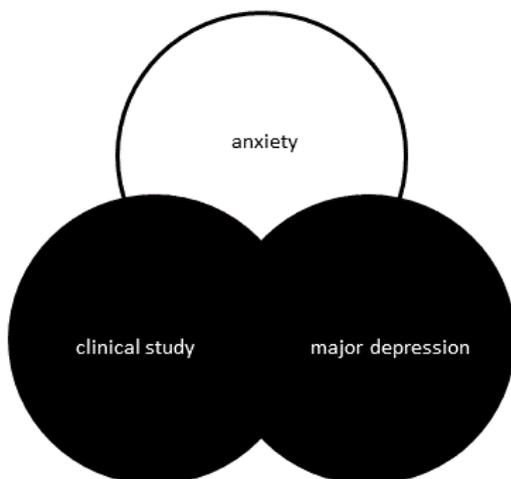
19. Sei ein Kreis eine Menge von Artikeln in einer wissenschaftlichen Fachdatenbank zu einem bestimmten Begriff oder einer Begriffskombination. Die dunkle Fläche sei die Menge der Treffer zu einer Suchanfrage, die hellen Flächen hingegen Artikel, die nicht in der Trefferliste erscheinen. Welche Suchanfrage(n) entspricht/entsprechen der Grafik?



Suchanfragen:

- **anxiety AND depression**
- anxiety OR depression
- anxiety NOT depression

20. Sei ein Kreis eine Menge von Artikeln in einer wissenschaftlichen Fachdatenbank zu einem bestimmten Begriff oder einer Begriffskombination. Die dunkle Fläche sei die Menge der Treffer zu einer Suchanfrage, die hellen Flächen hingegen Artikel, die nicht in der Trefferliste erscheinen. Welche Suchanfrage(n) entspricht/entsprechen der Grafik?



Suchanfragen:

- "clinical study" AND "major depression" NOT anxiety
- **"clinical study" OR "major depression"**
- "clinical study" OR "major depression" NOT anxiety

21. Warum ergibt die Suche nach *Personality AND false memories* in Fachdatenbanken weniger Treffer als die Suche nach *Personality OR false memories*?

- **Bei der Verknüpfung der Suchbegriffe mit AND wird nur nach solchen Einträgen gesucht, die mit beiden Suchbegriffen verknüpft sind. Bei der Verknüpfung mit OR werden auch Einträge gefunden, die nur mit einem der beiden Begriffe verknüpft sind.**
- Das Wort OR ist mit mehr Einträgen verknüpft als das Wort AND.
- OR impliziert, dass weitere Fachdatenbanken einbezogen werden, die sonst nicht verwendet würden.

22. Der h-Index ...

- ist eine Nummer zur eindeutigen Identifikation wissenschaftlicher Zeitschriftenbeiträge.
- ist eine Nummer, die zur eindeutigen Identifikation eines Autors genutzt wird, da Autoren häufig unter verschiedenen Schreibweisen gelistet werden.
- **ist ein bibliometrisches Maß, das auf Zitationen der Publikationen eines Autors basiert.**

23. Sie finden die folgende Literaturangabe:

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–402). New York: Academic Press.

Um welche Art von Publikation handelt es sich?

- Artikel in einer Zeitschrift.
- Monographie.
- **Beitrag in einem Herausgeberwerk.**

24. Sie finden die folgende Literaturangabe:

Tulving, E., & Watkins, M. J. (1974). On negative transfer: Effects of testing one list on the recall of another. Journal of Verbal Learning & Verbal Behavior, 13(2), 181–193.

Welche der folgenden Methoden ist die effizienteste, sich Zugang zu der Publikation zu verschaffen?

- Im elektronischen Bibliothekskatalog nach der Zeitschrift zu suchen.
- In der elektronischen Zeitschriften-Bibliothek nach der Zeitschrift zu suchen.
- **In einer Aufsatzdatenbank nach dem Zeitschriftenartikel zu suchen.**

25. Welche der Antwortmöglichkeiten ist korrekt?

Die Thesaurussuche wird üblicherweise verwendet, um ...

- Teilbereiche der Psychologie zu identifizieren.
- **die Suche zu präzisieren.**
- **weitere verwandte Suchbegriffe zu finden.**

26. Welche Kriterien können verwendet werden, um die Güte eines gefundenen Zeitschriftenartikels oder Buches einzuschätzen? *Bitte aufschreiben*

*Für folgende Angaben (oder ähnliche) wurde je 1 Punkt vergeben: Peer-Review, Reputation des Autors/Verlags, JIF/eigenfactor.org, Alter der Publikation. **Max. 3 Punkte dürfen vergeben werden!***

27. Welches sind wichtige Fachdatenbanken für die Psychologie?

- **PsycINFO**
- **PSYINDEX**
- PsycBASE

28. Welche der folgenden Aussagen sind korrekt?

- In wissenschaftlichen Datenbanken wie PSYINDEX sind ausschließlich Publikationen enthalten, die ein Peer-Review Verfahren durchlaufen haben.
- Google Scholar durchsucht neben wissenschaftlichen Datenbanken wie PSYINDEX und PsycINFO auch nahezu das gesamte World Wide Web.
- In PsycINFO sind neben deutschsprachiger Literatur auch deutschsprachige Tests dokumentiert.

29. Wofür kann der elektronische Bibliothekskatalog verwendet werden?

- **Um herauszufinden, welche Bücher in der (Universitäts-)Bibliothek vorhanden sind.**
- **Um herauszufinden, welche Zeitschriften von der (Universitäts-)Bibliothek abonniert wurden.**
- **Um auf Volltexte von Artikeln zuzugreifen.**

30. Welche der folgenden Aussagen sind korrekt?

- Reviews werden in erster Linie für eine nicht-wissenschaftliche Leserschaft verfasst.
- **Um ein Review zu verfassen, muss der Autor keine eigenen Datenerhebungen an Probanden vorgenommen haben.**
- **Reviews fassen eine Menge ausgewählter Publikationen zu einem bestimmten Themengebiet zusammen.**

31. Welche Aussagen über den elektronischen Bibliothekskatalog sind korrekt?

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

32. Welche der folgenden Funktionen, die in PSYINDEX (via OVID) verfügbar sind, sind in Google Scholar nicht vorhanden?

- Suche nach Publikationen, die in einer bestimmten Zeitschrift veröffentlicht wurden.
- **Suche anhand von Schlagwörtern.**

- **Schrittweiser Aufbau von Suchen.**
33. Welche der folgenden Aussagen sind korrekt?
- **Ergänzend zum Verzeichnis „Testverfahren“ bietet das ZPID eine Liste der Rezensionen an, die zu den jeweiligen Tests verfasst wurden.**
 - Im Verzeichnis „Testverfahren“ des ZPID sind ausschließlich hochwertige Tests aufgeführt, die psychometrische Mindestanforderungen erfüllen.
 - **Die Website www.testzentrale.de ist ein Angebot des Hogrefe-Verlags.**
34. Welche der folgenden Aussagen sind korrekt?
- Meta-Analysen berücksichtigen ausschließlich experimentelle Untersuchungen.
 - **In einer Meta-Analyse werden statistische Koeffizienten aggregiert.**
 - In Meta-Analysen werden Hypothesen aus verschiedenen Studien diskutiert, um Rückschlüsse auf eine zugrunde liegende Theorie zu ziehen.
35. Eine Feldsuche mit OVID ermöglicht, ...
- **spezifisch nach bestimmten Publikationstypen zu suchen (z. B. Buch).**
 - spezifisch nach Publikationen zu suchen, die bestimmte bibliometrische Mindestanforderungen erfüllen (z. B. Autoren-Einfluss).
 - **spezifisch nach solchen Publikationen zu suchen, in denen eine bestimmte Methodik angewandt wurde (z. B. Interview).**

16.2 Selbstwirksamkeitsskala

Zur Beantwortung der Items wurde eine fünfstufige Likert-Skala [1 = „trifft auf mich nicht zu“; 5 = „trifft auf mich zu“] mit „weiß-nicht“-Option vorgegeben.

1. Wenn ich einen Vortrag vorbereite, kann ich zu Beginn gut einschätzen, wie viel Literatur ich ungefähr benötigen werde.
2. Wenn ich während der Recherche zu einem Thema zu der Einschätzung gelangt bin, dass ich einen ausreichenden Überblick über die Literaturlage gewonnen habe, liege ich meist richtig.
3. Wenn ich zu einem Thema recherchiere, weiß ich genau, in welcher Reihenfolge ich die verschiedenen verfügbaren Informationsquellen verwende.
4. Während der Recherche zu einem Thema gehe ich zielgerichtet vor.
5. Wenn ich zu einem Thema recherchiere, kann ich relativ schnell entscheiden, ob eine Informationsquelle für mich interessant ist oder nicht.
6. Wenn ich weiß, dass es in einem Zeitschriftenartikel oder einem Buch relevante Informationen gibt, weiß ich gut, wie ich mir die Quellen beschaffen kann.
7. Wenn ich an meine bisherigen Recherchen (z.B. für Vorträge oder Berichte) zurückdenke, fallen mir wenige Optionen zur Verbesserung ein.
8. Ich weiß, wie ich Fachdatenbanken nutzen kann, um relevante Literatur zu finden.
9. Wenn ich Internetquellen verwende, weiß ich, worauf ich achten muss, um nicht falsche Informationen zu erhalten.
10. Verglichen mit anderen habe ich mehr Erfahrung bei der Recherche nach Fachinformationen.

16.3 Rechercheaufgaben

16.3.1 Messzeitpunkt 1

1. Finden Sie zwei nach 2005 publizierte wissenschaftliche Publikationen, die sich mit falschen Erinnerungen beschäftigen. Bitte verwenden Sie den englischsprachigen Suchbegriff "false memory".
2. Gibt es nach 2005 publizierte Längsschnittstudien (longitudinal study), die Risikofaktoren ("risk factors") für die generalisierte Angststörung ("Generalized Anxiety Disorder") untersuchen? Bitte geben Sie – wenn möglich – zwei entsprechende Publikationen an.
3. Suchen Sie zwei Publikationen, die sich mit der Frage beschäftigen, ob Altersunterschiede Einfluss auf das Therapieergebnis bei der Behandlung von Depressionen besitzen.

16.3.2 Messzeitpunkt 2

1. Finden Sie zwei nach 2005 publizierte wissenschaftliche Publikationen zum Thema Kurzzeitgedächtnis. Verwenden Sie den englischsprachigen Suchbegriff "Short Term Memory".
2. Gibt es nach 2005 publizierte Meta-Analysen, die Risikofaktoren ("risk factors") für die Entstehung einer posttraumatischen Belastungsstörung ("Posttraumatic Stress Disorder") untersuchen? Bitte geben Sie – wenn möglich – zwei entsprechende Publikationen an.
3. Suchen Sie zwei Studien, die sich damit beschäftigen, ob Unterschiede in der Prävalenz psychischer Störungen zwischen ethnischen Gruppen bestehen.

16.3.3 Messzeitpunkt 3

1. Finden Sie zwei nach 2003 (nicht einschließlich 2003) publizierte wissenschaftliche Publikationen, die sich mit Panikattacken beschäftigen. Verwenden Sie den englischsprachigen Suchbegriff "Panic Attack".
2. Gibt es nach (nicht einschließlich) 2003 publizierte Meta-Analysen, die die Wirksamkeit der sogenannten kognitiven Verhaltenstherapie ("cognitive behavior therapy") bei der Behandlung von Depression ("Major Depression") untersuchen? Bitte geben Sie – wenn möglich – zwei entsprechende Publikationen an.
3. Suchen Sie zwei Studien, die sich damit beschäftigen, welche spezifischen Arten der Schmerzbehandlung es für Personen mit Persönlichkeitsstörungen gibt.

16.4 Fragen zur Prozessbewertung der Rechercheaufgaben

1. Welche Informationssystem (Suchmaschine, bzw. Fachdatenbank) haben Sie verwendet? (Bitte nur das zuletzt verwendete Informationssystem angeben).
2. Welche Suchbegriffe haben Sie verwendet?
3. Wie haben sie die Suchbegriffe identifiziert? (nur bei Typ 3-Aufgaben gefragt)
4. Haben Sie spezifische Funktionen des Informationssystems (z.B. Filterfunktion) genutzt? Wenn ja, welche?
5. Haben Sie weitere Informationssysteme benutzt? Wenn ja, welche?
6. Trugen die weiteren Informationssysteme zum Rechercheergebnis bei?

16.5 Auswertungsschlüssel zu den Rechercheaufgaben

16.5.1 Bewertung des Ergebnisses

Hinweis: Jede gefundene Publikation wird separat bewertet.

Typ 1-Aufgaben

- Thematische Relevanz: Der angegebene Suchbegriff (bzw. ein Synonym) ist im Titel und/oder Abstract und/oder Schlagworten enthalten: 0.5 Punkte, sonst 0 Punkte (es erfolgt keine Inspektion des Volltexts der Publikation, da der erforderliche Aufwand hierfür zu groß wäre).
- Zeitliche Passung: Die Publikation wurde in dem lt. Aufgabenstellung erforderlichen Zeitfenster (z.B. nach 2005) veröffentlicht: 0.5 Punkte, sonst 0 Punkte.

Pro Treffer kann somit max. 1 Punkt erreicht werden; für die gesamte Aufgabe sind daher max. 2 Punkte erreichbar.

Typ 2-Aufgaben

- Thematische Relevanz: Die angegebenen Suchbegriffe (bzw. Synonyme) sind im Titel und/oder Abstract und/oder Schlagworten enthalten: 0.5 Punkte pro Suchbegriff (d.h. bei zwei Suchbegriffen max. 1 Punkt für thematische Relevanz).
- Falls Titel und/oder Abstract zeigen, dass dieser Artikel etwas anderes als das Gesuchte behandelt, der Artikel allerdings in einer Fachdatenbank mit beiden Schlagworten verknüpft ist: 0.5 Punkte. In diesem Fall werden dennoch Teilpunkte vergeben, weil man Schlagwörter als Indikator des Inhalts ansehen kann.
- Zeitliche Passung: Die Publikation wurde in dem lt. Aufgabenstellung erforderlichen Zeitfenster (z.B. nach 2005) veröffentlicht: 0.5 Punkte Punkte, sonst 0 Punkte.
- Methodische Passung: Die Publikation entspricht dem lt. Aufgabenstellung erforderlichen Typus (z.B. Meta-Analyse): 0.5 Punkte.

Pro Treffer können somit max. 2 Punkte erreicht werden; für die gesamte Aufgabe sind max. 4 Punkte erreichbar.

Typ 3-Aufgaben

- Thematische Relevanz: Die relevanten Suchbegriffe (siehe Liste) oder Synonyme sind in Titel, Abstract oder Schlagwortliste enthalten: 2 Punkte.
- Ist nur einer der relevanten Suchbegriffe in Titel, Abstract oder Schlagwortliste enthalten: 0.5 Punkte.

Pro Treffer können somit max. 2 Punkte erreicht werden; für die gesamte Aufgabe sind max. 4 Punkte erreichbar.

Liste der relevanten Suchbegriffe für Typ 3-Aufgaben

Messzeitpunkt 1: age differences; major depression; treatment outcomes oder psychotherapeutic outcomes oder treatment effectiveness evaluation.

Messzeitpunkt 2: racial and ethnic differences; epidemiology; mental disorders.

Messzeitpunkt 3: personality disorders; pain management.

16.5.2 Bewertung des Vorgehens (Prozesspunktwerte)

Typ 1-Aufgaben

- Verwendete Suchmaschine: Google Scholar, MS Academic Search oder Fachdatenbanken: 1 Punkt, sonst 0 Punkte.
- Suchbegriffe wurden aus der Aufgabenstellung übernommen: 1 Punkt, sonst 0 Punkte.
- Nutzung komplexer Funktionen: Filterfunktion für Publikationsdatum verwendet: 1 Punkt, sonst 0 Punkte.
- Weitere verwendete Informationssysteme: Google Scholar, MS Academic Search oder Fachdatenbanken: je 1 Punkt.

Maximal wurden 4 Prozesspunkte für diese Aufgabe vergeben.

Typ 2-Aufgaben

- Verwendete Suchmaschine: Fachdatenbanken: 1 Punkt, Google Scholar oder MS Academic Search: 0.5 Punkte, sonst 0 Punkte.
- Suchbegriffe wurden aus der Aufgabenstellung übernommen: 1 Punkt, sonst 0 Punkte.
- Nutzung komplexer Funktionen: Verwendung Boolescher Operatoren, Filterfunktion für Publikationsdatum, Filterfunktion für Methodik oder jede weitere komplexe Funktion, deren Verwendung hier sinnvoll erschien: je 1 Punkt.
- Weitere verwendete Informationssysteme: Google Scholar, MS Academic Search oder Fachdatenbanken: je 0.5 Punkte.

Maximal wurden 5 Prozesspunkte für diese Aufgabe vergeben.

Typ 3-Aufgaben

- Verwendete Suchmaschine: Fachdatenbanken: 1 Punkt, Google Scholar oder MS Academic Search: 0.5 Punkte, sonst 0 Punkte.
- Suchbegriffe: 1 Punkt für jeden Begriff aus der Liste, 0.5 Punkte für jeden ähnlichen Begriff, sonst 0 Punkte.
- Identifikation der Suchbegriffe: Aus vorliegender Literatur entnommen² oder durch Thesaurussuche ermittelt: 1 Punkt, sonst 0 Punkte.
- Nutzung komplexer Funktionen: Nicht ausgewertet.
- Weitere verwendete Informationssysteme: Google Scholar, MS Academic Search oder Fachdatenbanken: je 0.5 Punkte.

Maximal wurden 5 Prozesspunkte für diese Aufgabe vergeben.

Die Antworten auf die Frage, ob die weiteren Informationssysteme zum Rechercheergebnis beitrugen, wurden bei allen Aufgabentypen nur deskriptiv ausgewertet und gingen nicht in die Bewertung des Vorgehens ein.

² Hierfür muss der Proband zunächst eine relevante Publikation finden, was mithilfe von Google Scholar u.U. auch ohne Kenntnis der genauen wissenschaftlichen Fachtermini möglich ist. Aus dieser Publikation können dann die exakten Suchbegriffe entnommen werden.