

Aus dem **Institut für Didaktik und Ausbildungsforschung in der Medizin**

Institut der **Ludwig-Maximilians-Universität München**

Leitung: Prof. Dr. med. Martin R. Fischer

**Evidenzbasierte Medizin im Medizinstudium:
Bewerten von Studien – eine generische
Teilkompetenz mit Beeinflussung durch
fachspezifisches Wissen**

Dissertation

zum Erwerb des Doktorgrades der Medizin

an der Medizinischen Fakultät der

Ludwig-Maximilians-Universität zu München

vorgelegt von

Joana Melle

aus

Dachau

2018

Mit Genehmigung der Medizinischen Fakultät
der Universität München

Berichterstatter: Prof. Dr. med. Martin R. Fischer

Mitberichterstatter: Prof. Dr. Dr. Peter C. Scriba

Prof. Dr. Wolfgang Locher

Mitbetreuung durch die promovierten Mitarbeiter: Dr. Jan Kiesewetter

Dekan: Herr Prof. Dr. med. dent. Reinhard Hickel

Tag der mündlichen Prüfung: 15.03.2018

KURZZUSAMMENFASSUNG

Einleitung: Um evidenzbasierte Medizin (EbM) zu praktizieren, müssen Ärztinnen und Ärzte in der Lage sein, die aktuell beste Forschungsliteratur gewissenhaft, eindeutig und vernünftig für die individuelle Patientenentscheidung anzuwenden [2]. Dabei spielt die Teilkompetenz des Bewertens von Studien eine bedeutende Rolle. Denn wenn diese nicht korrekt beherrscht wird, kann es zu gravierenden Fehlentscheidungen zu Ungunsten des Patienten kommen [3]. Daher muss das Bewerten von Studien als ein Teil der EbM gut geschult werden. Das Ziel der hier vorgestellten Arbeit ist es, die Lehre der EbM im Medizinstudium zu verbessern. Um dies zu erreichen, muss die Vermittlung der Teilkompetenz „Bewerten von Studien“ tiefgründig verstanden werden. Es wurden dazu folgende zwei Fragestellungen bearbeitet: *Ist das Bewerten von Studien in der Medizin und Pädagogik eine fachspezifische oder generische Teilkompetenz? Wird die Teilkompetenz „Bewerten von Studien“ von Intelligenz und wissenschaftlichem Denken beeinflusst?*

Methode: In einem neu entwickelten Kompetenztest bearbeiteten Medizinstudierende (n= 161, 58% weiblich, 40% männlich) aus klinischen Semestern jeweils ein Szenario aus den Domänen Medizin und Pädagogik. Um eine vorgegebene Entscheidungssituation (z.B.: Wann ist der beste Zeitpunkt für eine Varizellenimpfung?) lösen zu können, mussten die Medizinstudierenden in beiden Szenarien jeweils vier Studien mit Hilfe des Bewertungssystems „BESTUEN“ (neustufige Ordinalskala mit Mittelpunkt) bewerten. Mit Hilfe von Experten wurde eine Musterlösung entwickelt. Die daraus berechnete Leistung im „Bewerten von Studien“ wurde zwischen den Domänen verglichen. Zusätzlich wurden das Vorwissen zu den Themen der Studien sowie Tests zur Intelligenz und zum wissenschaftlichen Denken abgeprüft.

Ergebnis: Die Leistungen beim Bewerten der Medizin-Studien korrelierten signifikant mit den Leistungen im Bewerten der Pädagogik-Studien ($r = ,42$; $p < ,01$). Außerdem gab es einen geringen, aber signifikanten Zusammenhang zwischen dem Bewerten von Medizin-Studien und dem Vorwissen in der Medizin ($r_{\text{Vorwissen Med}} = -,20$; $p < ,05$). Jedoch gab es keine Hinweise darauf, dass die Teilkompetenz des Bewertens von Studien von weiteren fachspezifischen Indikatoren wie höheres Fachsemester oder dem Besuch eines zum Thema passenden Seminars abhängt. Auch waren die Leistungen der Medizinstudierenden im Bewerten von Studien in der Domäne Medizin nicht besser als in der Domäne Pädagogik. Es konnte ein Zusammenhang zwischen dem Bewerten von Studien und der Fähigkeit wissenschaftlich zu Denken festgestellt werden. Zudem gab es marginale Hinweise auf einen Zusammenhang zwischen dem Bewerten von Studien und der Intelligenz.

Diskussion: Die Ergebnisse dieser Studie legen nahe, dass die Teilkompetenz des Bewertens von Studien generisch zu sein scheint und trotzdem von fachspezifischem Wissen sowie von der Fähigkeit, wissenschaftlich zu Denken beeinflusst wird. In dieser Studie konnte nur ein marginaler Einfluss von Intelligenz auf das Bewerten von Studien gezeigt werden.

INHALTSVERZEICHNIS

1.	Einleitung	6
1.1	Evidenzbasierte Medizin.....	6
1.2	Kompetenz des evidenzbasierten Arbeitens	10
1.3	Bewerten von Studien als Teilkompetenz evidenzbasierten Arbeitens .	11
1.4	Vergleich von zwei Domänen.....	15
1.5	Fachspezifisch oder generisch	16
1.6	Ziel der Arbeit.....	20
1.7	Fragestellungen und Hypothesen.....	20
2.	Methodenteil.....	25
2.1	Studiendesign.....	25
2.2	Teilnehmende.....	25
2.2.1	Voraussetzungen der Teilnahme.....	25
2.2.2	Demographische Daten der Teilnehmer	26
2.3	Kompetenztest: Aufbau und Entwicklung.....	28
2.3.1	Aufbau des Kompetenztests.....	28
2.3.2	Entwicklung der Szenarien.	29
2.3.3	Informationssuche.....	32
2.3.4	Bewertung von Studien (BESTUEN).	34
2.3.5	Argumentation.	38
2.4	Vortest.....	38
2.5	Studiendurchführung	40
2.6	Statistische Analysen	42
3.	Ergebnisse	44
3.1	Vorbereitende Analysen	44
3.1.1	BESTUEN.....	44
3.1.2	Vergleich der Varianzen.	46
3.1.3	Ausschluss eines Reihenfolgeneffekts.....	46
3.1.4	Vortest.....	47

3.2	Hypothesenüberprüfung.....	48
3.2.1	Fachspezifisch oder generisch.	48
3.2.2	Einfluss durch Intelligenz und wissenschaftliches Denken.	52
4.	Diskussion	53
4.1	Das Bewerten von Studien, eine generische Teilkompetenz mit fachspezifischem Einfluss	53
4.1.1	Fachspezifisch oder generisch.	53
4.1.2	Generisch oder übertragbar.....	55
4.1.3	Einfluss durch Intelligenz oder wissenschaftliches Denken.	56
4.1.4	Schlussfolgerung.....	57
4.2	Stärken und Schwächen des Kompetenztests	58
4.2.1	Studiendesign.....	58
4.2.2	Teilnehmende.....	59
4.2.3	Vorbereitende Analysen.....	59
4.2.4	Testgüte der BESTUEN.....	61
4.2.5	Inhaltliche Validität.	63
4.2.6	Weitere Forschung.	64
4.2.7	Zusammenfassung.....	66
4.3	Erkenntnisse für die Lehre der Evidenzbasierten Medizin.....	66
5.	Literaturverzeichnis.....	69
6.	Abbildungsverzeichnis	75
7.	Tabellenverzeichnis.....	76
8.	Anhang	77
8.1	Modulsystem an der Ludwig-Maximilians-Universität.....	77
8.2	Aufwärm szenario.....	78
8.3	Durchführungsleitfaden für die Testdurchläufe.....	79
8.4	Szenario Pädagogik.....	80
8.5	Hintergrundinformation Pädagogik	81
8.6	Szenario Medizin	82

8.7	Hintergrundinformation Medizin.....	83
8.8	Beispiel für eine Hauptstudien-Zusammenfassung	84
8.9	Suche von relevanten Studientiteln.....	85
8.10	Beispiel einer Kurzzusammenfassung die für die Informationssuche...87	
8.11	Leitfaden für die Suche der zehn Studien für die Informationssuche....88	
8.12	Beispiel zur Gewinnung der Musterlösung für die BESTUEN.....	90
9.	Eidesstattliche Versicherung	92

1 EINLEITUNG

1.1 EVIDENZBASIERTE MEDIZIN

Um eine optimale Patientinnen- und Patientenbetreuung zu gewährleisten, müssen Ärztinnen und Ärzte nicht nur eine große klinische Expertise mitbringen, sondern zusätzlich nach dem aktuellsten Forschungsstand behandeln können [2, 4-7]. Für die individuelle Patientenentscheidung sollte dabei die aktuell beste Forschungsliteratur gewissenhaft, eindeutig und vernünftig genutzt werden. Wenn dies berücksichtigt wird, wird von evidenzbasierter Medizin (EbM) gesprochen [2].

Mit der aktuell besten Forschungsliteratur ist praxisrelevante Forschungsliteratur gemeint, wie z.B. Studien über neue Medikamente oder Interventionen. Es geht dabei nicht um Literatur über Grundlagenforschung [5]. Schließlich kann nur in der Praxis Erprobtes sich unmittelbar auf die Patientinnen- und Patientenbehandlung auswirken. Evidenzbasierte Medizin möchte ein bestmögliches Behandlungsergebnis für Patientinnen und Patienten erreichen, so dass Forschungsliteratur mit möglichst hoher Evidenz bevorzugt wird. Der höchste Grad an Evidenz ist eine systematische Übersichtsarbeit, die ausschließlich randomisiert kontrollierte Studien einschließt, der niedrigste Grad an Evidenz ist eine einzelne Expertenmeinung [8]. Jedoch hilft die beste Forschungsliteratur nichts, wenn sie nicht von den Ärztinnen und Ärzten verstanden, bewertet und das resultierende Wissen für Patientinnen und Patienten eingesetzt werden kann.

Um EbM in Deutschland zu gewährleisten, erstellen wissenschaftliche medizinische Fachgesellschaften (von der Adipositas Gesellschaft bis zur Zytologie Gesellschaft) aus einzelnen Studien systematische Übersichtsarbeiten und daraus Leitlinien [9]. Als Hilfestellung dazu hat das Chochrane Institut Deutschland mit einer

Arbeitsgruppe das sogenannte GRADE-System entwickelt, mit dem systematische Übersichtsarbeiten und Leitlinien strukturiert und transparent erstellt werden können [10]. Die Aufbereitung soll den praktizierenden Ärztinnen und Ärzten ermöglichen, mit wenig zeitlichem Aufwand den aktuellen Stand der Forschung zu verfolgen und für Patientinnen und Patienten zu nutzen. So hat evidenzbasierte Medizin in Deutschland zwar bereits „(...)eine beeindruckende Wegstrecke zurückgelegt“ (S. 1640), dieser Weg ist jedoch noch nicht am Ende [11]. Denn trotz des immensen Aufwands und der großen Anzahl an systematischen Übersichtsarbeiten und Leitlinien, stellt uns der Fortschritt der EbM auch vor Schwierigkeiten. So wird es in der Medizin immer wieder individuelle Entscheidungen geben, welche mit keiner der bestehenden Leitlinien zu beantworten sind [12]. Auch werden zunehmend Ärztinnen und Ärzte benötigt, die fähig sind, diese Übersichten und Leitfäden zu erstellen. Zudem braucht eine wachsende evidenzbasierte Medizin Ärztinnen und Ärzte, welche Leitlinien nicht nur anwenden, sondern auch hinterfragen können [4]. Die Kompetenz, einzelne Studien kritisch zu bewerten und die so gewonnenen Erkenntnisse in den Behandlungsprozess mit einfließen zu lassen, ist also nach wie vor (und mehr als je) eine wünschenswerte Fähigkeit aller praktizierenden Ärztinnen und Ärzte. Da es aus zeitlichen und strukturellen Gründen oft schwer möglich ist, die Umsetzung von EbM im Berufsalltag zu erlernen, sollte die Lehre der EbM bereits im Studium eingeführt werden [13]. Denn um eine sinnvolle Anwendung der EbM zu gewährleisten, müssen junge Ärztinnen und Ärzte gut vorbereitet werden [14]. Bemerkenswert ist, dass EbM in der Lehre der Medizin noch nicht fest verankert und daher auch nicht in der aktuellen ärztlichen Approbationsordnung enthalten ist [15]. Der Wissenschaftsrat hat allerdings in einer Empfehlung zur Weiterentwicklung des Medizinstudiums vom 11.7.2014 die wissenschaftliche Kompetenz als Grundsatz mit aufgenommen [16].

Und der Fakultätentag hat am 4.6.2015 einen nationalen kompetenzbasierten Lernzielkatalog Medizin (NKLM) verabschiedet, in dem evidenzbasierte Medizin als Lernziel explizit erwähnt ist [17]. Die Lehre der EbM ist also bereits auf dem Weg dahin fest in die Lehre der Medizin verankert zu werden. Dies sollte durch fundierte Forschung bestmöglich unterstützt werden.

Nicht nur die Medizinpraxis selber, sondern auch die Lehre der Medizin wird zunehmend evidenzbasiert gestaltet. Die Didaktik der Lehre in der Medizin soll evidenzbasiert weiter entwickelt werden, um optimale Methoden zu finden, welche Medizinstudierende auf ihren zukünftigen Beruf vorbereiten. Hierfür hat sich der Begriff „Best Evidence Medical Education“ (BEME) etabliert [18].

Wie die Lehre der Medizin insgesamt, soll auch ein Curriculum zur Lehre von EbM im Medizinstudium evidenzbasiert entwickelt werden. Dafür ist eine geeignete Evaluation der Didaktik unumgebar. Für diese Evaluation gibt es bereits Fragebögen wie den Berlin-Questionnaire oder den Fresno-Test [19, 20]. Der validierte Berlin-Questionnaire beinhaltet 15 Multiple-Choice-Fragen über Wissen und Fähigkeiten in EbM. Der ebenfalls validierte Fresno-Test enthält vor allem offene Fragen zu EbM, welche durch ein standardisiertes „Grading“-System ausgewertet werden. Im Fragebogen wird anhand von klinischen Szenarios detailliert abgefragt, welche Schritte man durchlaufen muss, um evidenzbasiert zu arbeiten.

In der hier vorgestellten Studie wurde ein Kompetenztest neu entwickelt, in welchem die aktive Ausübung von evidenzbasiertem Arbeiten getestet werden soll. Dabei war es den Entwicklern des Kompetenztests² wichtig, nicht nur mit Fragebögen zu arbeiten. Vielmehr sollte ein Kompetenztest entwickelt werden, in welchem die einzelnen Teilschritte des evidenzbasierten Arbeitens in einem quasi-experimentellen Rahmen von den Studierenden selbst durchgeführt werden sollten. Dadurch konnte

sich die Testsituation einer realen Situation des evidenzbasierten Arbeitens noch mehr annähern. Der Kompetenztest wurde ursprünglich für die Durchführung mit sozialwissenschaftlichen Studierenden (Pädagogik und Lehramt) im Rahmen des KOMPARE-Projekts¹ entwickelt [1]. Dabei wurden Erkenntnisse der BEME in die Sozialwissenschaften übertragen. KOMPARE steht für *Kompetent argumentieren mit Evidenzen*, wobei hier nicht nur EbM in der Medizin bzw. BEME in der Lehre der Medizin gemeint ist, sondern evidenzbasiertes Arbeiten in verschiedenen Fachgebieten. In der hier vorgestellten Studie wurde das für sozialwissenschaftliche Studierende erarbeitete Konzept des KOMPARE-Projekts² übernommen, wieder in den Kontext der BEME gebracht und mit Medizinstudierenden durchgeführt (siehe Abbildung 1).

¹ Das KOMPARE-Projekt wurde vom Bundesministerium für Bildung und Forschung gefördert.

² Bei der Entwicklung des Kompetenztests der hier vorgestellten Studie im Rahmen des KOMPARE-Projekts waren folgende Personen beteiligt:

Jan Kiesewetter¹, Joana Melle¹, Eva Mörwald¹, Martin R. Fischer¹, Andreas Hetmanek², Christof Wecker², Frank Fischer², Kati Trempler³, Cornelia Gräsel³

1 Institut für Didaktik und Ausbildungsforschung in der Medizin am Klinikum der LMU München

2 Lehrstuhl für Empirische Pädagogik und Pädagogische Psychologie, LMU München

3 Lehrstuhl für Lehr-, Lern- und Unterrichtsforschung in der School of Education, Bergische Universität Wuppertal

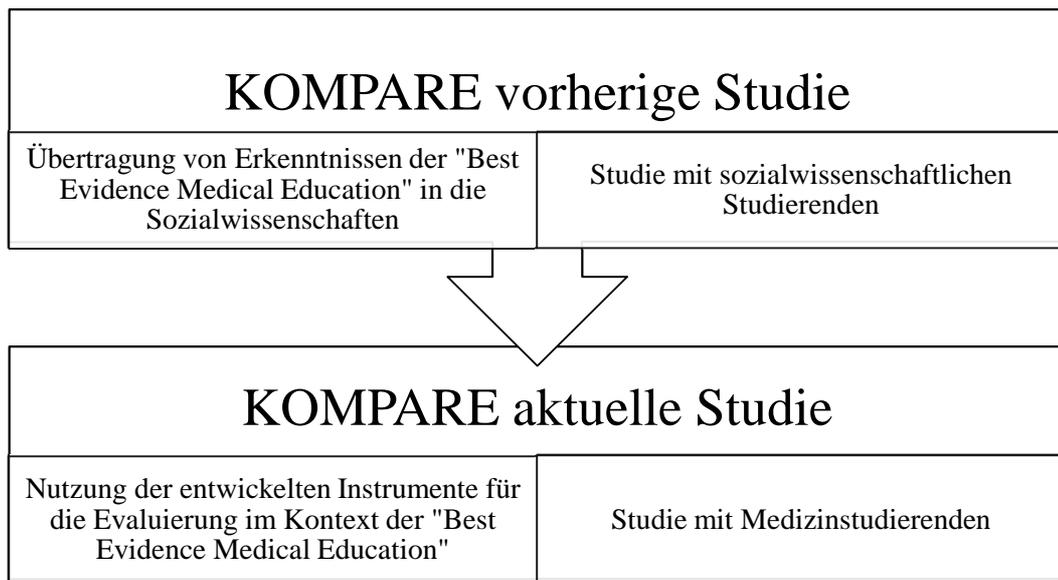


Abbildung 1: *Übersicht über das KOMPARE-Projekt.* (Autorin: Joana Melle)

Mit dem KOMPARE- Projekt insgesamt soll also die Kompetenz, evidenzbasiert zu arbeiten genauer analysiert werden. Um dies besser einordnen zu können, wird zunächst auf den Begriff Kompetenz näher eingegangen.

1.2 KOMPETENZ DES EVIDENZBASIERTEN ARBEITENS

Nach Hartig und Klieme ist Kompetenz definiert als erlernbare, bereichsspezifische „Leistungsdispositionen (...), die in einem gewissen Maß über ähnliche Situationen generalisierbar sind“ (S.129), und welche nicht mit Intelligenz gleich zu setzen seien [21]. In der Lehre der Medizin wird Kompetenz meist als eine Kombination aus Wissen, Fähigkeiten und anderen Komponenten wie Einstellung und Werten definiert [22]. Ähnliche Komponenten wurden zur Evaluation von EbM-Lehre verwendet: In der fünften internationalen Konferenz der Lehrer und Entwickler der evidenzbasierten Gesundheitsversorgung im Oktober 2009 auf Sizilien war man sich darin einig, dass zur Evaluation der Lehre von evidenzbasierter Medizin mehrere Komponenten untersucht werden müssen. Diese waren Wissen, Fähigkeiten, Einstellungen, Patientenoutcome, Verhalten, Selbsteffektivität und Reaktionen auf

die Lehre [23]. Ähnlich der Definition von Kompetenz in der Lehre der Medizin [22] sind also auch bei der Ausübung evidenzbasierter Medizin Wissen und Fähigkeiten entscheidende Komponenten. Es ist folglich naheliegend, dass die Ausführung evidenzbasierter Medizin als eine Kompetenz zu betrachten ist. Daraus kann wiederum geschlossen werden, dass auch evidenzbasiertes Arbeiten im Allgemeinen eine Kompetenz sein muss.

In der hier vorgestellten Studie wird die Kompetenz des evidenzbasierten Arbeitens mit dem Fokus auf den Teilaspekt der Fähigkeiten untersucht und analysiert, inwiefern diese generalisierbar und/oder bereichsspezifisch sind (siehe Hartig und Klieme) [21].

1.3 BEWERTEN VON STUDIEN ALS TEILKOMPETENZ EVIDENZBASIERTEN ARBEITENS

Zum Erlernen der Kompetenz des evidenzbasierten Arbeitens ist es hilfreich, die Kompetenz in einzelne Teile zu spalten und diese einzeln zu erlernen. Nach dem aktuellen Stand der Literatur wurden häufig fünf Stufen des evidenzbasierten Arbeitens beschrieben [24-26]. In Abbildung 2 sind fünf Stufen dargestellt, welche auf Literaturrecherche sowie Erfahrung der Delegierten der „Conference of Evidence-Based Health Care Teachers and Developers“ 2003 basieren [25]. M. Dawes et al. (2005) haben diese fünf Stufen niedergeschrieben:

Die fünf Stufen des evidenzbasierten Arbeitens

-
- 1) *Übersetzen von Unsicherheit in eine passende Fragestellung.*
 - 2) *Systematische Suche nach der besten Evidenz.*
 - 3) ***Kritisches Bewerten der Evidenzen bezüglich Validität, klinischer Relevanz und Anwendbarkeit.***
 - 4) *Anwendung der Ergebnisse in der Praxis.*
 - 5) *Evaluation.*

Abbildung 2: Die fünf Stufen des evidenzbasierten Arbeitens nach Dawes (2005). In der hier vorgestellten Arbeit liegt der Fokus auf dem kritischen Bewerten von Studien.

Im Folgenden wird von jeder Stufe als Teilkompetenz gesprochen, wobei der Begriff Teilkompetenz hier nur als Fähigkeit verstanden wird (siehe Kapitel 1.2).

Im Kompetenztest des dieser Arbeit zu Grunde liegenden KOMPARE-Projekts wurden im Besonderen drei Teilkompetenzen untersucht: Informationssuche, Bewerten von Studien und Argumentation. Während der Informationssuche sollten zur Fragestellung passende Studien gefunden werden. Sie entspricht der zweiten Stufe nach Dawes. Das Bewerten von Studien entspricht der dritten Stufe. Die Argumentation ist ein wesentlicher Bestandteil der vierten und fünften Stufe nach Dawes, denn nur durch Argumentation wird eine mögliche Anwendung begründet und evaluiert. Die tatsächliche Anwendung in der Praxis wurde nicht überprüft. Die erste Stufe nach Dawes wurde bewusst weggelassen, da dies mit dem Modell des Kompetenztests, in welchem eine Frage vorgegeben wird, nur schwer zu vereinbaren war [25]. Die vorliegende Arbeit konzentriert sich auf eine dieser drei Teilkompetenzen: Das Bewerten von Studien.

Warum steht das Bewerten von Studien in der hier vorgestellten Studie im Fokus? Evaluationsmethoden für die Teilkompetenz des Bewertens von Studien, in der Literatur oft „critical appraisal“ genannt, finden besondere Beachtung in der aktuellen

Forschung [27, 28]. So beschrieben Swennen et al. das Bewerten von Studien als eine besonders schwer zu erlernende Teilkompetenz in der EbM [6]. Nach Porzolt hingegen ist das Bewerten von Studien zwar nicht der schwierigste Teilschritt der evidenzbasierten Medizin, jedoch jener, welcher am längsten dauert und als Basis für die klinische Entscheidung von besonderer Wichtigkeit ist. Wenn die Teilkompetenz des Bewertens von Studien nicht korrekt beherrscht wird, kann es zu gravierenden Fehlentscheidungen zu Ungunsten des Patienten kommen [3]. Wie wichtig es ist, Studien richtig zu bewerten, wurde von Bowden et al. bekräftigt [29]. Auf Grund der 1973 von Meehl beschriebenen Diskrepanz zwischen statistischer Signifikanz und dem tatsächlichen klinischen Nutzen von Studien [30], sahen es Bowden et al. als essentiell an, eine Methode zu finden, mit welcher das Bewerten von Studien erleichtert werden kann. In ihrer Studie prüften sie die Anwendung der „Critically Appraised Topic“- Methode (CAT) von Straus, Richardson, Glasziou, & Haynes zur Bewertung von Studien anhand eines Beispiels aus der Neuropsychologie. Die CAT-Methode stellt drei Hauptfragen, welche man beim Lesen einer Studie beantworten sollte (Bowden nach Straus Richardson, Glasziou, & Haynes, 2011): 1) Sind die Ergebnisse der Studie valide? 2) Sind diese validen Ergebnisse wichtig? 3) Können die validen, wichtigen Ergebnisse auf die Behandlung des Patienten angewendet werden? Diese übergreifenden Fragen wurden durch weitere Unterfragen operationalisiert [31]. Bowden et al. kamen schließlich zu dem Schluss, dass die CAT-Methode, zusätzlich zu klinischer Expertise der Ärztinnen und Ärzte, sehr hilfreich ist zum Bewerten von Studien [29]. Ein systematisches Konzept scheint folglich wichtig zu sein für das erfolgreiche Bewerten von Studien.

Doch nicht nur Straus et al. entwickelten Methoden zur Erleichterung des Bewertens von Studien. Mit dem Ziel der Verbesserung von evidenzbasierter Lehre

in der Medizin (BEME), entwickelten Harden et al. ein weiteres Konzept als Hilfestellung beim Bewerten von Studien, welches aus sechs Dimensionen (QUESTS) besteht. QUESTS ist ein Akronym und steht für Qualität (quality), Nützlichkeit (utility), Reichweite (extent), Stärke (strength), Ziel (target), und Kontext (setting) [18]. Mit Hilfe dieser sechs Dimensionen sollte beim Lesen einer Studie beurteilt werden, wie gut bzw. relevant die Studie ist (Qualität), inwiefern die in der Studie angewandte Methode auf andere Situationen übertragen werden kann (Nützlichkeit), wie groß die Evidenz ist, also z.B. wie viele Studien in einem Review ausgewertet wurden (Reichweite), wie stark die Studie ist, in Bezug auf die statistische Auswertung (Stärke), ob mit der Studie das gemessen wurde, was gemessen werden sollte (Ziel) und inwiefern das Setting der Studie auf den Kontext der eigenen Fragestellung passt (Kontext).

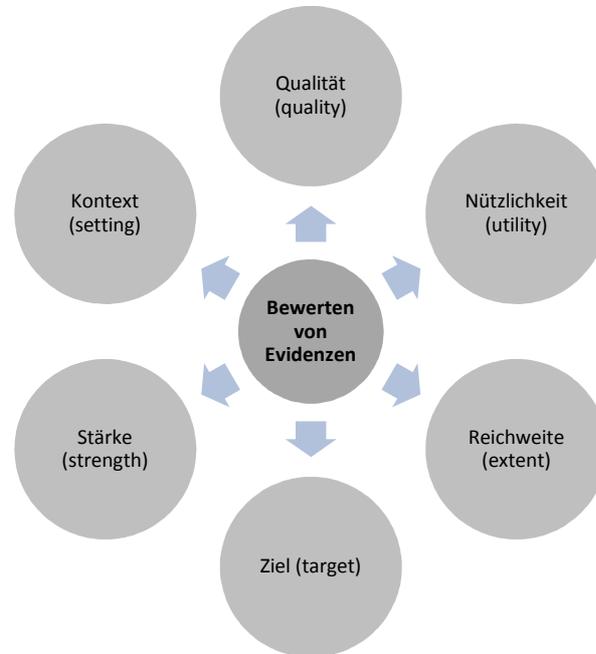


Abbildung 3: *QUESTS, sechs Dimensionen zur Evaluierung von Studien in der Medizindidaktik nach RM Harden (2000).*

Die QUESTS-Dimensionen wurden von den Entwicklern des Kompetenztests des KOMARE-Projekts als ein handhabbares Instrument in der Medizindidaktik

eingestuft, dass sich auch für eine Übertragung in die Sozialwissenschaften eignet. Deswegen wurden die QUESTS als Basis für eine neue Testskala mit den Bewertungsitems BESTUEN ausgewählt [1]. In die Entwicklung der Bewertungsitems BESTUEN gingen zusätzlich zu den beschriebenen sechs Dimensionen nach Harden et al. die UTOS- Kriterien nach Cronbach und Shapiro und die vier Aspekte der Validierung von Cook und Campbell mit ein [18, 32, 33]. Mit den BESTUEN wurde also ein neues systematisches Konzept entwickelt, welches das Bewerten von Studien in den Sozialwissenschaften sowie in der Medizin erleichtern soll und mit dem eine Evaluation der Teilkompetenz „Bewerten von Studien“ möglich ist. BESTUEN steht für „BEwerten von STUdiEN“. Diese Bewertungsitems sind der zentrale Bestandteil der hier vorgestellten Studie.

1.4 VERGLEICH VON ZWEI DOMÄNEN

In der hier vorgestellten Studie bewerteten Medizinstudierende Studien aus der Medizin und aus der Pädagogik. Mit den gleichen Testinstrumenten (BESTUEN) konnten somit die Teilkompetenz des Bewertens von Studien in zwei verschiedenen Fachgebieten (Domänen) getestet werden. Dadurch ist ein direkter Vergleich der beiden Domänen möglich.

Mit dem Begriff Domäne ist laut Duden ein Spezialgebiet gemeint, in dem sich jemand besonders gut auskennt [34]. Mandl, Gruber und Renkl definierten eine Domäne als abgegrenzten Inhalt mit bestimmten Aufgaben und Handlungen, spezifischem Wissen sowie „charakteristischen Problemlösestrategien“ (S.6) [35]. In der hier vorgestellten Studie wird die Definition von Mandl et al. zu Grunde gelegt.

Mit Hilfe des direkten Vergleichs der Leistungen in der eigenen und fremden Domäne kann schließlich untersucht werden, ob die Teilkompetenz des Bewertens von Studien fachspezifisch oder generisch ist.

1.5 FACHSPEZIFISCH ODER GENERISCH

Eine Kompetenz kann „fachspezifisch“ oder „generisch“ sein. Eine fachspezifische Kompetenz enthält Wissen, Fähigkeiten, Einstellung und Werte, die auf ein Fachgebiet ausgerichtet sind [22, 36]. Die Hochschuldidaktische Arbeitsstelle der TU Darmstadt definierte eine fachspezifische Kompetenz als ein Zusammenspiel aus „Fachkenntnisse(n) und -methoden, sowie deren Anwendung zur Bewältigung fachspezifischer Aufgaben“ (S.3) [37]. Häufig werden die Begriffe „fachspezifische Kompetenz“ und „Fachkompetenz“ synonym verwendet. In dieser Studie wird hingegen differenziert: Innerhalb eines Studiengangs bzw. einer Domäne können viele fachspezifische Kompetenzen erworben werden, deren Summe eine hohe Fachkompetenz ergibt.

Die These, dass die Teilkompetenz des Bewertens von Studien fachspezifisch, also eine speziell auf die Domäne Medizin ausgerichtete Teilkompetenz ist, scheint auf den ersten Blick unwahrscheinlich. Schließlich werden in vielen wissenschaftlichen Fächern Studien bewertet.

Nach Alexander ist domänenspezifisches Wissen erforderlich, um bestimmte Strategien anwenden zu können. Dabei definierte sie eine Strategie als eine Form des prozeduralen Wissens, welches unter anderem benötigt wird, um eine Aufgabe auszuführen. Wenn man die Teilkompetenz des Bewertens von Medizin-Studien als eine Aufgabe in der eigenen Domäne bezeichnet, könnte es also sein, dass domänenspezifisches Wissen aus dem Medizinstudium Strategien fördert, um diese Aufgabe, also Medizin-Studien zu bewerten, auszuführen [38]. Zudem kann domänenspezifisches Wissen helfen, einen Text aus derselben Domäne besser zu verstehen, denn laut Marti und Ulmi werden Informationen eines Textes mit „bereits vorhandenem Sprach-, Welt- und Fachwissen verbunden“(S.3) (Marti & Ulmi zitiert

nach Steffens, 2002) [39]. In der vorhergehenden Studie des KOMPARE-Projektes gab es bereits Hinweise darauf, dass pädagogisches Vorwissen in positivem Zusammenhang mit dem Bewerten von Pädagogik-Studien steht und somit die Teilkompetenz „Bewerten von Pädagogik-Studien“ von der Domäne abhängt [1]. Unter Betrachtung oben genannter Argumente kann vermutet werden, dass die Teilkompetenz des Bewertens von Studien möglicherweise von der Domäne beeinflusst wird, auch wenn es unwahrscheinlich erscheint, dass es sich um eine rein fachspezifische Kompetenz handelt.

Generische Kompetenzen sind hingegen unabhängig von der Domäne. Nach Strijbos sind die zehn häufigsten in der Literatur (zwischen 2000 und 2003) beschriebenen generischen Kompetenzen, die man am Ende eines Bachelorstudiengangs beherrschen sollte: Denkfähigkeit, Informationsprozessierung, Problemlösen, Kreativität, kritische Reflektion, lebenslanges Lernen, soziale Verantwortung, Teamwork, Kommunikation und Führung [40]. Nach dem Vorbild von Young & Chapman, fasste Strijbos die Kompetenzen Denkfähigkeit, Informationsprozessierung, Problemlösen und Kreativität in einem „cluster of conceptual skills“ (S. 25) zusammen: Studierende sollen am Ende des Bachelorstudiengangs in der Lage sein, zielgerichtete, evidenzbasierte und korrekte Ergebnisse zu liefern [40, 41]. Dies impliziert, dass Evidenzen für den Entscheidungsprozess genutzt werden sollen und dass diese dazu zunächst adäquat bewertet werden müssen. Die Teilkompetenz des Bewertens von Studien, als ein Teil des evidenzbasierten Arbeitens, könnte also generischer Natur sein. Ähnliches beschrieb Clanchy: Demnach gehören Methoden des Nachforschens zu den generischen Fähigkeiten, welche Studierende aller Fachrichtungen am Ende ihres Studiums beherrschen sollten [42]. Daraus kann man schlussfolgern, dass die

Methoden des Nachforschens nicht domänenspezifisch sind, sondern in jeglicher Fachrichtung gleichermaßen vorhanden sein sollten. Zu den Methoden des Nachforschens gehört auch die Teilkompetenz des Bewertens von Studien. Leutner et al. erklärten eine fächerübergreifende Kompetenz damit, dass sie „nicht eindeutig einem einzelnen Fach zuzuordnen“ (S. 12) ist [43]. Wenn das Bewerten von Studien zu den Fähigkeiten gehört, die alle Studierenden am Ende Ihres Studiums beherrschen sollen, dann ist sie nicht eindeutig einem Fach zuzuordnen und ist folglich eine fächerübergreifende, also generische Kompetenz. Auf Grund der genannten Argumente ist es durchaus möglich, dass die Teilkompetenz des Bewertens von Studien generisch ist.

In einem Review von Larkin wurden verschiedene starke Evidenzen dazu gefunden, dass domänenspezifisches Wissen übertragbar ist. Demnach können neue Aufgaben in einer bisher nicht bekannten Domäne schneller gelernt werden, wenn bereits Erfahrungen aus einer anderen Domäne vorhanden sind [44]. Nach Billing ist es für die Übertragbarkeit von Fähigkeiten wichtig, metakognitive Strategien zu erlernen. Das Erlernen von Prinzipien und Konzepten erleichtern dies. Metakognitive Strategien sind Strategien zum Lösen eines Problems, die selbst geplant, beobachtet und evaluiert werden können [45]. Wenn man die Bewertungssitem „BESTUEN“ (siehe Kapitel 1.3 und 2.3.4) als ein Konzept betrachtet, mit welchem das Erlernen metakognitiver Strategien zum Bewerten von Studien erleichtert wird, sollten die BESTUEN die Übertragbarkeit der Teilkompetenz des Bewertens von Studien von einer in eine andere Domäne fördern. Übertragbarkeit ist wiederum, wenn man Larkin berücksichtigt, nicht gleichzusetzen mit einer generischen Fähigkeit [44]. Das bedeutet: Selbst wenn das Bewerten von Studien eine fachspezifische Teilkompetenz wäre, könnte es sein, dass diese mit Hilfe der BESTUEN leichter übertragbar ist und

so eine generische Teilkompetenz vortäuscht. Zum besseren Verständnis findet sich in Abbildung vier eine Übersicht der verschiedenen Formen von Kompetenzen:

Übersicht über generische, fachspezifische und übertragbare Kompetenzen:
Übertragbar ist nicht gleich generisch!

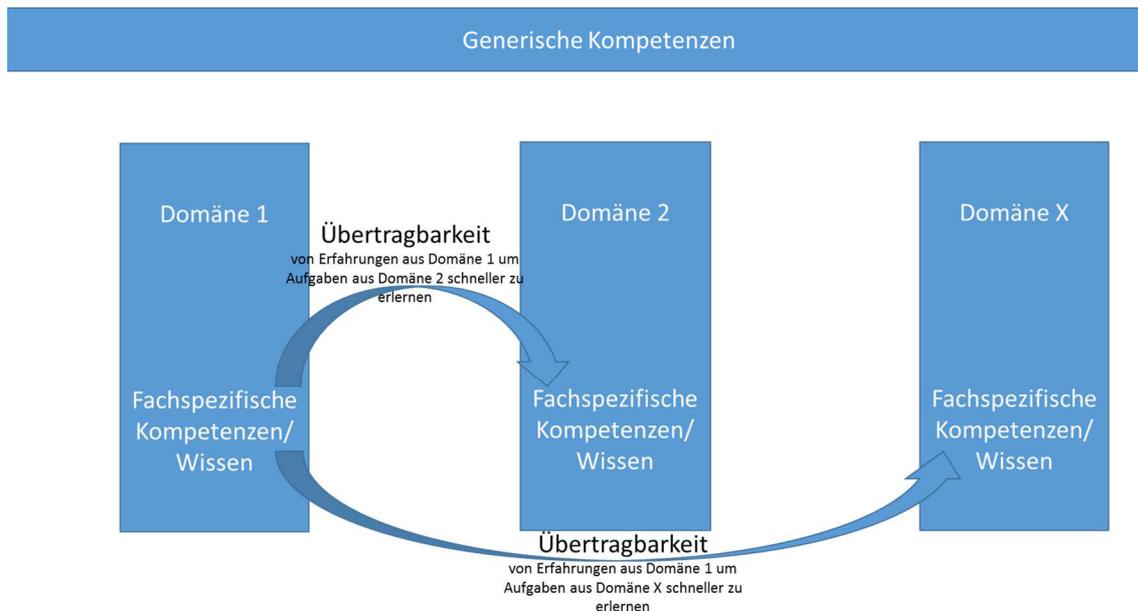


Abbildung 4: **Übersicht über generische, fachspezifische und übertragbare Kompetenzen.** (Autorin: Joana Melle)

Wie oben bereits erwähnt, wird Kompetenz nach Hartig und Klieme als eine erlernbare bereichsspezifische Leistungsdisposition definiert, „die in einem gewissen Maß über ähnliche Situationen generalisierbar“ (S.129) ist [21]. Wenn diese Definition einer Kompetenz auch für die Teilkompetenz des Bewertens von Studien zutrifft, dann wäre sie fachspezifisch (für die Domäne Medizin spezifisch) und gleichzeitig übertragbar.

Zusammenfassend scheint es plausibel zu sein, dass die Teilkompetenz des Bewertens von Studien generisch ist, da alle Studierenden nach Ende Ihres Studiums diese Teilkompetenz beherrschen sollten. Eine Beeinflussung durch die Domäne ist jedoch genauso möglich, denn Fachwissen hilft den Studierenden dabei den Inhalt einer Studie besser einzuordnen. In dieser Studie soll zunächst grundlegend untersucht

werden, ob die Teilkompetenz des Bewertens von Studien fachspezifisch oder generisch ist. Erst nach Beantwortung dieser Frage kann diskutiert werden, ob diese Teilkompetenz von der Domäne beeinflusst wird.

1.6 ZIEL DER ARBEIT

Ziel dieser Arbeit ist es, die Lehre von evidenzbasierter Medizin in das Medizinstudium zu integrieren. Als Voraussetzung dafür muss geeignete Forschung stattfinden. Die einzelnen Teilkompetenzen der evidenzbasierten Medizin sollten tiefgründig verstanden werden, damit sie optimal gelehrt werden können. In der hier vorgestellten Arbeit wird insbesondere die Teilkompetenz „Bewerten von Studien“ untersucht. Dazu wird zunächst grundlegend überprüft, ob das Bewerten von Studien fachspezifisch oder generisch ist. Schließlich werden verschiedene Einflüsse auf die Teilkompetenz „Bewerten von Studien“ ermittelt und bewertet. Die Erkenntnisse hieraus sollen schließlich als ein weiterer Baustein zur Entwicklung der Lehre der evidenzbasierten Medizin im Medizinstudium genutzt werden.

1.7 FRAGESTELLUNGEN UND HYPOTHESEN

Um schließlich die Einflüsse auf die Teilkompetenz „Bewerten von Studien“ zu ermitteln, werden in der hier vorgestellten Studie zwei Fragen untersucht.

- I. *Ist das Bewerten von Studien in der Medizin und Pädagogik fachspezifisch oder generisch?*
- II. *Wird die Teilkompetenz „Bewerten von Studien“ von Intelligenz und wissenschaftlichem Denken beeinflusst?*

I.

Wenn die Teilkompetenz des Bewertens von Studien fachspezifisch wäre, dann müsste sich eine hohe Fachkompetenz der Medizinstudierenden auf die Leistung im Bewerten der Medizinstudien positiv auswirken (siehe Kapitel 1.4) und die Bewertung der Studien der eignen Domäne (Medizin) müsste besser gelingen, als die Bewertung der domänenfremden Studien (Pädagogik). Das Besuchen eines Seminars zum Thema der zu bewertenden Studien, hier die Impfmedizin (siehe Kapitel 2.3), müsste sich positiv auf das Bewerten von Studien zu diesem Thema auswirken. Außerdem müsste den Medizinstudierenden aus höheren Semestern die Bewertung von Medizin-Studien besser gelingen, als denen aus niedrigeren Semestern. Denn Medizinstudierende aus höheren Semestern haben aufgrund der längeren Studiendauer in der Regel eine höhere Fachkompetenz als Studierende aus niedrigeren Semestern. Schließlich sollten diejenigen, welche ein besseres Fachwissen zum Thema der zu bewertenden Studien haben, auch besser bewerten können. Wenn jedoch die Teilkompetenz des Bewertens von Studien generisch wäre, dann sollten sich die Leistungen in der eigenen und in der fremden Domäne ähneln.

Wenn das Bewerten von Studien mit Hilfe der BESTUEN eine *fachspezifische* Teilkompetenz ist, dann müssten daher folgende vier Hypothesen zutreffen:

H1 Die Leistung im Bewerten von Studien bei Medizinstudierenden ist in der eigenen Domäne „Medizin“ besser, als in der fachfremden Domäne „Pädagogik“.

H2 Die Leistung im Bewerten von Studien der Impfmedizin nach Besuchen eines Impfseminars ist besser, als ohne Besuch des Impfseminars.

H3 Die Leistung im Bewerten von Studien aus der Domäne „Medizin“ ist bei Medizinstudierenden aus höheren Semestern besser, als bei Medizinstudierenden aus niedrigeren Semestern.

H4 Die Leistung im Bewerten von Studien ist umso besser, je besser das spezifische Vorwissen zum Thema der zu bewertenden Studien ist.

Wenn jedoch das Bewerten von Studien eine *generische* Teilkompetenz ist, dann müssen die oben genannten Hypothesen abgelehnt werden. Außerdem muss folgende Hypothese zutreffen:

H5 Je besser die Leistungen im Bewerten von Medizin-Studien sind, umso besser sind auch die Leistungen im Bewerten von Pädagogik-Studien.

Die Hypothesen zur Hauptfragestellung (I.) sind in Abbildung 5 zusammengefasst:

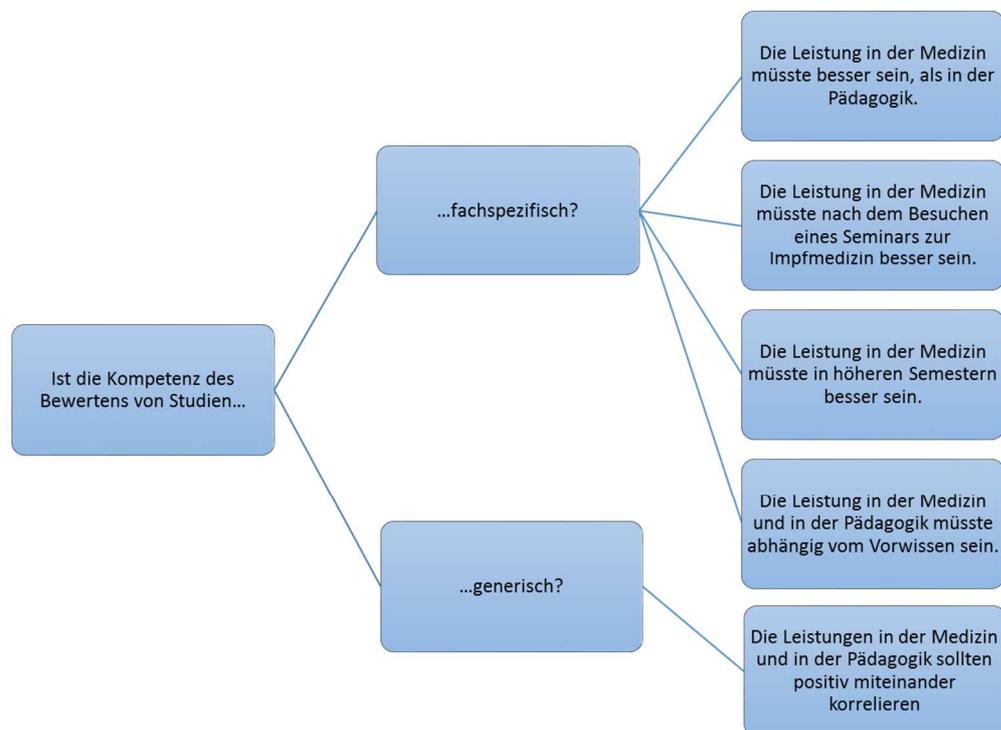


Abbildung 5: *Fünf Hypothesen zur Bearbeitung der Hauptfragestellung* (Autorin: Joana Melle).

II.

Um die Teilkompetenz des Bewertens von Studien tiefgründig zu verstehen, wird zusätzlich zur Hauptfragestellung - ob das Bewerten von Studien fachspezifisch oder generisch ist – eine zweite Fragestellung formuliert: Hängt die Teilkompetenz des Bewertens von Studien von *Intelligenz* und der *Fähigkeit wissenschaftlich zu Denken* ab?

Nach Gruber und Stamouli wird Intelligenz folgendermaßen definiert: „Intelligenz ist die Fähigkeit eines Menschen zur Anpassung an neuartige Bedingungen und zur Lösung neuer Probleme auf der Grundlage vorangehender Erfahrungen im gesellschaftlichen Kontext“ (S.29) [46]. Beim Bewerten von Studien sollten die Studierenden eine bisher unbekannte Studie mit einer neuen Methode, den Bewertungsitems „BESTUEN“, bewerten. Möglicherweise könnte also höhere Intelligenz dazu beitragen, diese Aufgabe besser zu meistern. Zudem wurde bereits oben erwähnt, dass nach Hartig und Klieme Kompetenz und Intelligenz positiv miteinander korrelieren, also möglicherweise auch Intelligenz mit der Teilkompetenz des Bewertens von Studien [21]. In der vorherigen Studie des KOMPARE-Projekts konnte bereits ein positiver Zusammenhang zwischen der Teilkompetenz des Bewertens von Studien und einzelnen Aufgaben des Intelligenz-Struktur-Tests festgestellt werden [1]. Daher ergibt sich die Frage, ob die Intelligenz auch in dieser Studie einen Einfluss auf die Teilkompetenz des Bewertens von Studien hat. Die Hypothese lautet:

H6 Je höher die Intelligenz der Probanden ist, umso besser wird die Teilkompetenz des Bewertens von Studien beherrscht.

Nicht nur die Intelligenz beeinflusst möglicherweise die Teilkompetenz des Bewertens von Studien. Es ist auch denkbar, dass die Fähigkeit wissenschaftlich zu Denken diese beeinflusst: Um die Fähigkeit des wissenschaftlichen Denkens zu testen, wurde von Lawson et al. ein Test entworfen in dem eine Situation analysiert, daraus eine Schlussfolgerung gezogen und/oder ein Problem gelöst werden sollte. Lawson et al. konnten dabei einen positiven Zusammenhang zwischen der Schwierigkeit dieser Aufgaben und dem Lösen von neuen Transferproblemen nachweisen [47]. Die Medizinstudierenden in der hier vorgestellten Studie sollten mit den neuen Bewertungsitens Studien analysieren und anschließend bewerten. Möglicherweise wird die Fähigkeit, dieses neue „Problem“ zu lösen, auch von der Fähigkeit beeinflusst, wissenschaftlich zu Denken. In der vorherigen Studie des KOMPARE-Projekts konnte bereits ein positiver Zusammenhang zwischen der Teilkompetenz des Bewertens von Studien und dem wissenschaftlichen Denken festgestellt werden [1]. Wenn mit Forschungserfahrung die Fähigkeit des wissenschaftlichen Denkens gefördert würde, dann zeigte sich in einer Studie von Zwolsman ein gegenteiliges Bild. Darin stand Forschungserfahrung nämlich weder in einem Zusammenhang mit der Fähigkeit, evidenzbasierte Medizin zu praktizieren, noch mit dem Wissen über EbM [48]. Zu diesen EbM-Fähigkeiten gehört auch das Bewerten von Studien. Daher stellt sich die Frage, ob auch in dieser Studie die Teilkompetenz des Bewertens von Studien mit der Fähigkeit wissenschaftlich zu Denken zusammenhängt. In dieser Arbeit wird also zusätzlich folgende Hypothese untersucht:

H7 Je besser die Fähigkeit der Probanden ist, wissenschaftlich zu Denken, umso besser wird die Teilkompetenz des Bewertens von Studien beherrscht.

2 METHODENTEIL

Im Folgenden wird zunächst auf das verwendete Studiendesign eingegangen, sowie auf die Teilnehmer der Studie. Anschließend werden der Aufbau und die Entwicklung des gesamten Kompetenztests genauer beschrieben. Daraufhin werden der Vortest und der Ablauf der Studiendurchführung dargestellt. Abschließend werden die zur Überprüfung der Hypothesen verwendeten statistischen Analysen kurz beschrieben.

2.1 STUDIENDESIGN

Die Studie hat ein quasi-experimentelles Querschnittsdesign mit zwei Domänen. Als Erhebungsmethode wurde ein Kompetenztest durchgeführt. Eine Gruppe von Medizinstudierenden sollte zu einem Zeitpunkt mit den Bewertungsitems BESTUEN Studien bewerten. Dabei wurde die Domäne der zu bewertenden Studien variiert. Es wird nun untersucht, inwiefern die unabhängigen Variablen „*Domäne, Intelligenz und wissenschaftliches Denken*“ die abhängige Variable „*Bewerten von Studien*“ beeinflussen. Ergänzend zum quasi-experimentellen Teil der Studie wird ein Zusammenhang zwischen dem „*Bewerten von Studien*“ in zwei verschiedenen Domänen untersucht.

2.2 TEILNEHMENDE

2.2.1 Voraussetzungen der Teilnahme

Die Studie wurde mit Medizinstudierenden der Ludwigs-Maximilians-Universität (LMU) in München durchgeführt. Die Teilnahme an der Studie war freiwillig. Es gab eine Aufwandsentschädigung von 40,00 € für die vollständige Bearbeitung des Vortests und des Kompetenztests (siehe folgende Kapitel 2.4, 2.3 und 2.5).

Um von einer Domäne „Medizin“ ausgehen zu können, wurden nur Medizinstudierende mit klinischen Grundlagen, also ab dem Modul 23 (6. Semester) zur Teilnahme an der Studie eingeladen. Erst ab diesem Semester konnte davon ausgegangen werden, dass die Medizinstudierenden Basiswissen und Basiskompetenzen aus ihrem Fachgebiet mitbringen. In Anhang 1 ist eine Übersicht über das Modulsystem des Studiengangs Humanmedizin der LMU zu betrachten.

2.2.2 Demographische Daten der Teilnehmenden

In die Auswertung der Studie gingen 161 Medizinstudierende ein. Davon waren 40 % männlich und 58% weiblich, 2% hatten keine Angabe gemacht. Das durchschnittliche Alter war 24,9 Jahre. Deutsch war die Muttersprache von 85 Prozent der Teilnehmenden. 53 Teilnehmende waren im sechsten oder siebten Fachsemester (Modul 23), 43 Teilnehmende im achten oder neunten Semester (Modul 4 oder 5), 34 Teilnehmende im Modul 6 oder im Freisemester und 25 Teilnehmende im Praktischen Jahr (siehe Tabelle 1). Drei Medizinstudierende hatten keine Angabe zu ihrem aktuellen Modul und zum Alter gemacht und drei weitere hatten gar keine Angabe zu ihren demographischen Daten gemacht. Insgesamt 22 Medizinstudierende hatten bereits ein zweistündiges Seminar zum Thema Impfmedizin besucht, welches in Modul 4 stattfindet.

Vier Studierende, die noch vor dem sechsten Fachsemester standen und trotzdem an der Studie teilgenommen haben wurden ausgeschlossen. Das heißt von ursprünglich 165 teilnehmenden Medizinstudierenden gingen 161 in die Auswertung ein.

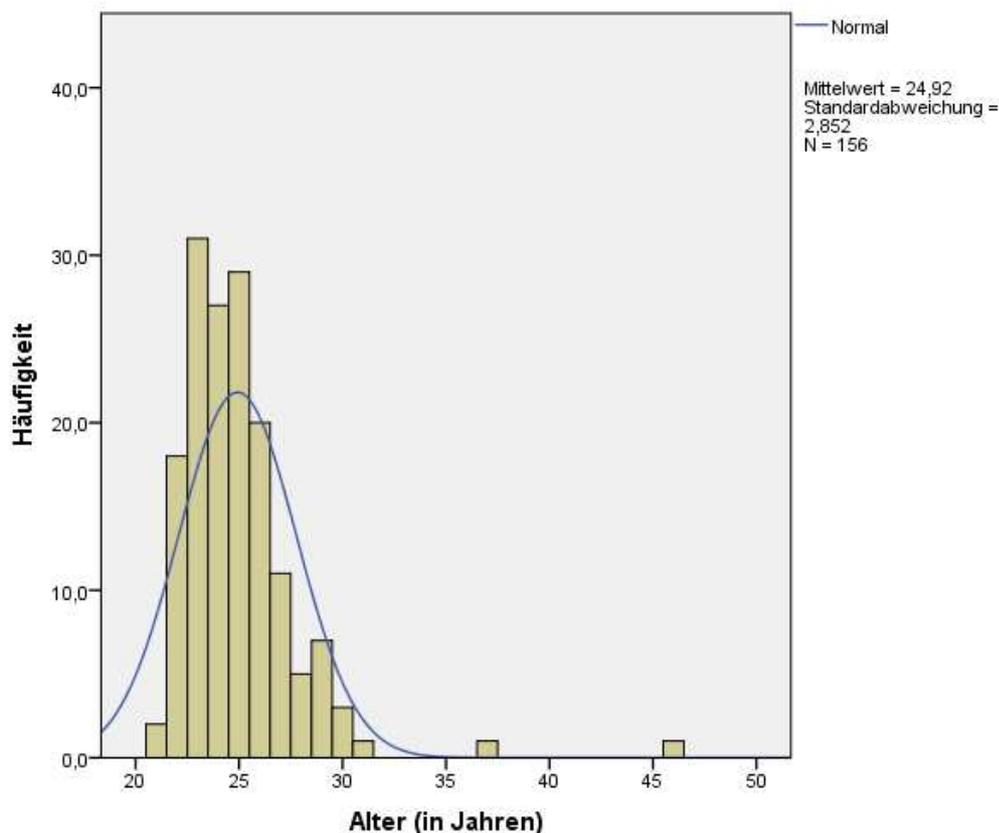


Abbildung 6: Altersverteilung der teilnehmenden Medizinstudierenden: zwei Medizinstudierende haben fälschlicherweise beim Alter „0“ angegeben. Diese beiden wurden in der Durchschnittsberechnung des Alters nicht berücksichtigt. Weitere drei haben gar keine Angabe zu den demografischen Daten gemacht.

	Häufigkeit	Prozent
Modul 23	53	32,9
Modul 4	21	13,0
Modul 5	22	13,7
Modul 6	19	11,8
Praktisches Jahr	25	15,5
Freisemester	15	9,3

Tabelle 1: Zugehörigkeit der Medizinstudierenden zu den einzelnen Semestern im Medizinstudium an der LMU. Die sechs Medizinstudierenden, welche keine Angaben zu den Modulen gemacht haben, sind nicht berücksichtigt.

2.3 KOMPETENZTEST: AUFBAU UND ENTWICKLUNG

Das Fundament des KOMPARE-Projekts sowie der hier vorgestellten Studie ist der Kompetenztest. In diesem wurden sowohl die Teilkompetenzen Informationssuche, Bewerten von Studien und Argumentation, als auch einige Kontrollvariablen untersucht. Um dem Leser einen vollständigen Eindruck von dem Test zu machen, den die Medizinstudierenden bearbeitet haben, soll der Kompetenztest und sein Entwicklungsprozess im Folgenden komplett berichtet werden - auch wenn nicht alles relevant für die Fragstellung ist.

2.3.1 *Aufbau des Kompetenztests*

Der Kompetenztest bestand aus zwei Szenarien – aus einem Pädagogik- und einem Medizin-Szenario. Mit dem Lesen der Szenarien wurden die teilnehmenden Medizinstudierenden jeweils in eine fiktive Situation gebracht, in welcher sie evidenzbasiert eine Entscheidung zwischen zwei Möglichkeiten fällen sollten. Im ersten Schritt sollten die Medizinstudierenden eine Bauchentscheidung ohne die Zuhilfenahme zusätzlicher Informationen fällen. In einem zweiten Schritt sollten sie passende Studien suchen und schließlich vier vorgegebene Studien lesen und bewerten. Danach sollten sich die Medizinstudierenden noch einmal informiert zwischen den beiden Optionen entscheiden und schließlich argumentieren, warum sie zu ihrer Entscheidung gekommen sind. Zudem wurden die Motivation und die Sicherheit der Medizinstudierenden zu den einzelnen Aufgaben und bei den Entscheidungen mittels Fragebogen ermittelt (siehe Abbildung 7).

Szenario

Wissensfragen
Bauchentscheidung
Informationssuche
Motivation bei der Testbearbeitung
Bewertung von Studien
Entscheidung
Argumentation

Abbildung 7: **Reihenfolge der zu bearbeitenden Aufgaben eines Szenarios**. In dieser Arbeit lag der Fokus auf dem Bewerten von Studien.

Wie bereits erläutert wurde, liegt der Fokus in dieser Arbeit auf nur einem Teil des Kompetenztests: dem Bewerten von Studien.

2.3.2 *Entwicklung der Szenarien*

Der Kompetenztest ist eine Weiterentwicklung des teilvalidierten Kompetenztests der bereits bestehenden Studie des KOMARE-Projekts (siehe auch Kapitel 1.1) [1]. Diese beinhaltete zwei Pädagogik-Szenarien und wurde mit sozialwissenschaftlichen Studierenden durchgeführt. In der hier vorgestellten Studie wurde nur eines der beiden Pädagogik-Szenarien aus dem vorhergehenden Projekt verwendet. Statt des zweiten Pädagogik-Szenarios kamen ein Medizin-Szenario sowie ein Aufwärm szenario (Anhang 2) neu hinzu. Die Bearbeitung von jeweils einem Szenario aus der Domäne Medizin und einem Szenario aus der Domäne Pädagogik durch dieselben Medizinstudierenden ermöglicht somit einen direkten Vergleich im Hinblick auf deren Teilkompetenz des Bewertens von Studien (siehe auch Kapitel 1.4). Das Aufwärm szenario deckte inhaltlich beide Domänen ab.

Um die Teilkompetenzen in den beiden Domänen miteinander vergleichen zu können, musste das Szenario aus der Medizin (Anhang 6) dem Szenario aus der Pädagogik (Anhang 4) möglichst ähnlich sein. Es sollte also ein Parallelszenario entwickelt werden, welches die gleiche Struktur hatte, möglichst gleiche Merkmale

der im Szenario vorkommenden Personen vorwies und in welchem eine möglichst gleiche Art der Entscheidung zu fällen war.

Inhaltlich wurde die Impfmedizin als Thema für das Medizin-Szenario gewählt. Dem lagen folgende Überlegungen zu Grunde: In der Impfmedizin werden Entscheidungen für eine Gruppe von Menschen getroffen und nicht nur für einen einzelnen Patienten. Eine gruppenspezifische Fragestellung ist demnach möglich. Dies entspricht dem Pädagogik-Szenario, in dem es um eine Entscheidung für eine Lehrmethode für eine Gruppe ging und nicht für eine einzelne Schülerin oder einen einzelnen Schüler. Zudem gibt es in der Impfmedizin häufig nur zwei Entscheidungsmöglichkeiten: „Impfen“ oder „nicht Impfen“. Dies entspricht ebenfalls dem Pädagogik-Szenario, in welchem die Entscheidung zwischen zwei Optionen getroffen werden musste.

Der Kompetenztest wurde mit dem Ziel entwickelt, ihn sowohl mit Medizinstudierenden in der hier vorgestellten Studie, als auch mit Studierenden der Sozialwissenschaft in einer weiteren hier nicht berichteten Studie durchzuführen (siehe Kapitel 4.2.6). Es musste also ein Thema gefunden werden, bei dem eine intuitive Entscheidung für beide Studierendengruppen möglich ist. Im Medizin-Szenario wurde das Thema der Varizelleninfektion gewählt. Dem lag die Überlegung zu Grunde, dass die Varizelleninfektion eine bekannte Kinderkrankheit ist, welche den meisten Studierenden jeglicher Fachrichtung aus der eigenen Erfahrung bereits bekannt sein sollte. Von eigenen Erfahrungen konnte auch im Pädagogik-Szenario ausgegangen werden. Jeder Studierende hat als Schüler/in bereits Kontakt zu verschiedenen Lehrmethoden gehabt. Somit konnte davon ausgegangen werden, dass beide Studierendengruppen in beiden Szenarien eine Bauchentscheidung auf Basis von eigenen Erfahrungen treffen konnten.

Innerhalb des Themas der Varizelleninfektion stand der richtige Zeitpunkt der Impfung zur Diskussion. Dies ist eine spannende und realistische Fragestellung, da dazu in Europa eine breite Studienlage sowie unterschiedliche Empfehlungen der Impfkommisionen existieren. In Deutschland wird empfohlen, die Varizellenimpfung im zweiten Lebensjahr durchzuführen [49]. In der Schweiz wird sie für das elfte Lebensjahr empfohlen, wenn zuvor noch keine Varizelleninfektion durchgemacht wurde [50]. Im Medizin-Szenario mussten sich die Medizinstudierenden in folgender Entscheidungssituation auf eine der beiden Möglichkeiten festlegen und diese anschließend begründen:

„Sollen Kinder flächendeckend mit zwei Jahren geimpft werden“

oder

„Sollen Kinder mit elf Jahren nur bei negativer oder unbekannter Varizellenanamnese geimpft werden?“

Damit wurde eine Entscheidung zwischen zwei Zeitpunkten generiert. Es ging also nicht - wie sonst in der Impfmedizin häufig der Fall - um das „Ob“, sondern über das „Wann“. Um das Medizin-Szenario mit dem Pädagogik-Szenario vergleichen zu können, wäre die Entscheidung „impfen oder nicht impfen“ daher nicht sinnvoll gewesen. Im Pädagogik-Szenario ging es nämlich auch nicht um das „Ob“ der Durchführung der Lerneinheit, sondern nur um das „Wie“.

Nachdem eine dem Pädagogik-Szenario ähnliche Fragestellung für das Medizinszenario gefunden wurde, wurden vier Original-Studien gesucht, welche die differierenden Meinungen der beiden Impfkommisionen widerspiegeln. Diese vier Original-Studien sollten möglichst valide sein und für beide Entscheidungsmöglichkeiten entsprechende Argumente beinhalten. Es sollte ein möglichst ähnliches Muster der Studienauswahl wie im Pädagogik-Szenario sein, um

die Vergleichbarkeit der beiden Domänen zu ermöglichen. Wie bei den Studien im Pädagogik-Szenario, wurden auch im Medizin-Szenario systematische Übersichtsarbeiten (Reviews und Metaanalysen), Fallstudien, Expertenmeinungen und Kostenanalysen bei der Datenbank-Suche ausgeschlossen. Diese wurden für die Testsituation bewusst weggelassen, da sie eine andere Struktur aufweisen und auch anders zu bewerten sind als einzelne Studien. Die Original-Studien wurden mit dem gleichen Schema, wie Studien im Pädagogik-Szenario, ins Deutsche übersetzt und zusammengefasst. In Anhang 8 ist ein Beispiel einer Zusammenfassung einer dieser Studien aus dem Medizinszenario zu lesen. Die vier Zusammenfassungen der Original-Studien (Hauptstudien) wurden schließlich verwendet, um die Teilkompetenz des Bewertens von Studien zu testen.

Um zu gewährleisten, dass die Studierenden überhaupt in der Lage waren, den Kompetenztest durchzuführen, sollte allen an der Studie teilnehmenden Studierenden ein möglichst einheitliches Basiswissen zur Verfügung gestellt werden. Dazu wurden, zusätzlich zu den beiden Szenarien, Hintergrundinformationen zur Varizelleninfektion (siehe Anhang 7) sowie zum Thema Gruppenpuzzle (Anhang 5) bereitgestellt.

Das Medizin-Szenario ist in Anhang 6 und das Pädagogik-Szenario ist in Anhang 4 zu lesen.

2.3.3 *Informationssuche*

Im Test zur Messung der Teilkompetenz „Informationssuche“, sollten die Medizinstudierenden anhand von Titeln und Kurzzusammenfassungen nach relevanten Studien suchen:

Zunächst sollte getestet werden, ob die Medizinstudierenden in der Lage waren bei der Datenbanksuche herauszufiltern, welche Studien für die Beantwortung der

Fragestellung sinnvoll waren. Es wurden in beiden Szenarien zehn Studientitel in folgendem Format vorgelegt, wobei die Titel alle in deutscher Sprache vorlagen:

„ Agarwal, N., Mittal, A., Kuldeep, C., Gupta, L. K., Khare, A. K. & Mehta, S. (2013). *Die chemische Wiederaufbau-Therapie von Hautnarben mittels 100%-iger Trichloressigsäure in der Behandlung von Windpockennarben im Gesichtsbereich: Eine Pilotstudie. J Cutan Aesthet Surg* 6(3). S. 144-147.“

Die Studierenden sollten auf einer vierstufigen Skala angeben, ob sie für die Entscheidung im Szenario auf Grundlage dieser Informationen mehr über die Studie wissen möchten. Die Antwortmöglichkeiten waren: *auf keinen Fall, eher nicht, eher schon, auf jeden Fall* (siehe auch Anhang 9). Anschließend hatten die Medizinstudierenden die Möglichkeit von allen zehn Studien eine Kurzzusammenfassung zu lesen (siehe Anhang 10). Sie sollten daraufhin entscheiden, ob sie auf der Grundlage dieser Informationen davon ausgehen können, dass die Studien für die Entscheidung eine wichtige Rolle spielen. Dies konnten die Teilnehmer auf einer 5-stufigen Skala (mit den Endpunkten „stimme gar nicht zu“; „stimme absolut zu“) angeben.

Zur Erstellung dieses Tests und seiner Musterlösung wurden auf Grundlage der vier hoch relevanten Hauptstudien (siehe Kapitel 2.3.2) je zwei mittel relevante, zwei wenig relevante und zwei irrelevante Studien gesucht. Dazu wurde ein Leitfaden für die Suche der zehn Studien in der Datenbank Pubmed erstellt, der in Anhang 11 nachzulesen ist. Die Kurzzusammenfassungen (Anhang 10) wurden nach einheitlichem Schema erstellt, ins Deutsche übersetzt und waren noch etwas kürzer gehalten als die vier Hauptstudien (Anhang 8). Es gab also zwei Versionen der vier, für das Bewerten von Studien relevanten, Original-Studien: Als Kurzzusammenfassungen im Teil „Informationssuche“ und als ausführlichere Zusammenfassungen (Hauptstudien) im Teil „Bewerten von Studien“. Die

„Informationssuche“ ging nicht in die Auswertung der hier vorgestellten Studie mit ein.

2.3.4 *Bewertung von Studien (BESTUEN)*

In Kapitel 1.3 wurde bereits begonnen die Entwicklung der Bewertungssitem BESTUEN darzustellen. Ebenfalls wurde dort bereits ausführlich erläutert, dass den BESTUEN, die QUESTS-Dimensionen zu Grunde liegen, die von Harden zur Evaluierung von Evidenzen in der medizinischen Ausbildung entwickelt wurden [18]. Auf deren Basis, unter Berücksichtigung der UTOS- Kriterien sowie der vier Aspekte der Validierung entwarfen die Entwickler des Kompetenztest des KOMPARE-Projekts zehn Items in deutscher Sprache, welche die Bewertung von Studien ermöglichten [32, 33]. Die Tabelle 2 veranschaulicht die Entwicklung der BESTUEN.

Die Entwicklung der BESTUEN begann in der vorherigen Studie, des KOMPARE-Projekts und wurde in der hier vorgestellten Studie fortgeführt. Nach Hinzufügen eines Szenarios aus der Medizin in der hier vorgestellten Studie wurden die zehn Items der BESTUEN nochmals angepasst, da nicht alle Formulierungen sowohl für Studien aus der Medizin, als auch aus der Pädagogik anwendbar waren. In Tabelle 3 ist das Ergebnis dieses Entwicklungsprozesses zu sehen. Jeweils nach dem Lesen der Zusammenfassungen der vier Hauptstudien (siehe Anhang 8) sollten die Medizinstudierenden mit Hilfe der BESTUEN die Studien bewerten.

QUESTS (Harden et al., 2000)	UTOS (Cronbach, 1982)	Art der Validität (Cook & Campbell, 1979)	Item der BESTUEN
Quality: Wie relevant ist die Studie für die Fragestellung?		Interne Validität	(2) Ob die Fragestellung zu meiner Entscheidung passt. (10) Ob die entscheidungsrelevanten Ergebnisse eindeutig auf die Maßnahme zurückzuführen sind.
Utility: Bis zu welchem Grad kann die Methode auf die Fragestellung übertragen oder adaptiert werden ohne sie zu modifizieren?	Treatment	Externe Validität	(3) Ob die Maßnahme untersucht wurde, um die es in meiner Entscheidung geht.
Extent: Wie groß ist die Evidenz? Anzahl der Studien in einer Metaanalyse oder einem Review			
Strength: Wie stark ist die Studie?		Statistische Validität	(9) Ob die entscheidungsrelevanten Ergebnisse statistisch belastbar sind.
Target: Was ist das Ziel? Was wurde gemessen? Wie valide ist die Evidenz?	Observations	Konstrukt Validität	(5) Ob die Erfolgskriterien mit den Erfolgskriterien bei meiner Entscheidung ausreichend übereinstimmen. (6) Ob die Erfolgskriterien mit geeigneten Verfahren erfasst wurden.
Setting: Inwiefern entspricht der Kontext oder das Setting dem der Fragestellung?	Setting	Externe Validität	(8) Ob die Rahmenbedingungen mit den Rahmenbedingungen bei meiner Entscheidung übereinstimmen.
	Units	Externe Validität	(7) Ob die Merkmale der Personen mit denen der Zielgruppe bei meiner Entscheidung übereinstimmen.
			(11) Unter Berücksichtigung der zuvor bewerteten Aspekte spielt die Studie für meine Entscheidung eine wichtige Rolle.

Tabelle 2: Entwicklung der BESTUEN [51].

1. *Die Studie spielt für meine Entscheidung eine wichtige Rolle.*
Aus dieser Kurzfassung geht hervor, dass....
2. *die Fragestellung zu meiner Entscheidung passt.*
3. *die Maßnahme untersucht wurde, um die es in meiner Entscheidung geht. (nur im Pädagogik-Szenario)*
4. *die Maßnahme leicht umsetzbar ist.*
5. *die Erfolgskriterien mit den Erfolgskriterien bei meiner Entscheidung ausreichend übereinstimmen.*
6. *die Erfolgskriterien mit geeigneten Verfahren erfasst wurden.*
7. *die Merkmale der Personen mit denen der Zielgruppe bei meiner Entscheidung übereinstimmen.*
8. *die Rahmenbedingungen mit den Rahmenbedingungen bei meiner Entscheidung übereinstimmen. (nur im Pädagogik-Szenario)*
9. *die entscheidungsrelevanten Ergebnisse statistisch belastbar sind.*
10. *die entscheidungsrelevanten Ergebnisse eindeutig auf die Maßnahme zurückzuführen sind.*
11. *Unter Berücksichtigung der zuvor bewerteten Aspekte spielt die Studie für meine Entscheidung eine wichtige Rolle.*

Tabelle 3: Die Bewertungssitems BESTUEN, mit Hilfe derer die Studierenden die vier Hauptstudien bewerten sollten³.

Bei jedem Item bearbeiteten die Medizinstudierenden eine neunstufige Ordinalskala mit Mittelpunkt (1 stimme überhaupt nicht zu; 9 stimme absolut zu). Es wurde eine ungerade Anzahl gewählt, da eine neutrale Einschätzung der Kriterien bei der Bewertung von Studien sinnvoll sein kann. Eine Studie kann zum Beispiel teilweise zu meiner Entscheidung passen, teilweise aber auch nicht, also in der

³ Bei der Entwicklung BESTUEN im Rahmen des KOMPARE-Projekts waren folgende Personen beteiligt:

Jan Kiesewetter¹, Joana Melle¹, Eva Mörwald¹, Martin R. Fischer¹, Andreas Hetmanek², Christof Wecker², Frank Fischer², Kati Trempler³, Cornelia Gräsel³

¹ Institut für Didaktik und Ausbildungsforschung in der Medizin am Klinikum der LMU München

² Lehrstuhl für Empirische Pädagogik und Pädagogische Psychologie, LMU München

³ Lehrstuhl für Lehr,- Lern- und Unterrichtsforschung in der School of Education, Bergische Universität Wuppertal

Wertung neutral sein. Da die Items drei und acht für die Medizin-Studien schlecht anwendbar waren, wurden im Medizin-Szenario diese beiden Items weggelassen (siehe Abbildung 8).

Aus dieser Kurzfassung geht hervor, dass ...

	überhaupt nicht zu	Ich stimme						völlig zu
die Fragestellung zu meiner Entscheidung passt.	<input type="radio"/>							
die Maßnahme untersucht wurde, um die es in meiner Entscheidung geht.	<input type="radio"/>							
die Erfolgskriterien mit denen bei meiner Entscheidung ausreichend übereinstimmen.	<input type="radio"/>							
die Erfolgskriterien mit geeigneten Verfahren erfasst wurden.	<input type="radio"/>							
die Merkmale der Personen mit denen der Zielgruppe bei meiner Entscheidung übereinstimmen.	<input type="radio"/>							
die Entscheidungsrelevanten Ergebnisse statistisch belastbar sind.	<input type="radio"/>							
die entscheidungsrelevanten Ergebnisse eindeutig auf die Maßnahme zurückzuführen sind.	<input type="radio"/>							
Unter Berücksichtigung der zuvor bewerteten Aspekte spielt die Studie für meine aktuelle Entscheidung eine wichtige Rolle.	<input type="radio"/>							

Abbildung 8: Die BESTUEN zum Bewerten von Medizinstudien mit nur acht Items und neunstufiger Ordinalskala 3.

Zur Auswertung des Tests zur Bewertung von Studien (BESTUEN), wurde eine Musterlösung entwickelt. Zwei Experten⁴ der Impfmedizin, sowie die Entwickler dieses Testes sammelten und gewichteten Argumente für die Bewertung der einzelnen Items. Dabei wurde darauf Rücksicht genommen, dass die vier Hauptstudien im Vergleich einheitlich bewertet wurden. Aus der Zusammenschau dieser Argumente entstand jeweils die Musterlösung für die einzelnen Items. Ein Beispiel für die Entstehung der Musterlösung ist in Anhang 12 zu betrachten. Die Musterlösung für das Pädagogik-Szenario wurde nach ähnlichem Schema von den Entwicklern des vorhergehenden KOMPARE-Projekts entwickelt und ohne Änderung übernommen[1].

⁴ Prof. Dr. med. Jörg Schelling, Komm. Direktor des Instituts für Allgemeinmedizin am Klinikum der Universität München und Dr. med. Hans-Jürgen Schrörs, Institut für medizinische Information, Berlin

2.3.5 Argumentation

In diesem Teil des Kompetenztests wurde getestet, inwiefern die Medizinstudierenden in der Lage sind, für ihre Entscheidung zu argumentieren.

Die Medizinstudierenden sollten in schriftlicher Form frei argumentieren, warum sie sich für eine Option entschieden haben. Dazu hatten sie nochmals die Möglichkeit, alle vier Hauptstudien auf dem Bildschirm zu lesen und daraus zu zitieren. Zusätzlich wurden den Medizinstudierenden die Hintergrundinformationen und das Szenario in gedruckter Form vorgelegt. Die Argumentation ging nicht in die Auswertung der hier vorliegenden Arbeit mit ein.

2.4 VORTEST

Um die Ergebnisse des Kompetenztests besser einordnen zu können, wurden in einem Online-Vortest, vor Bearbeitung des eigentlichen Kompetenztests, einige Kontrollvariablen abgefragt. Dieser Vortest wurde weitgehend von der vorhergehenden Studie des KOMPARE-Projekts übernommen. Er wurde um wenige Tests ergänzt und, um die maximale Testzeit in einem annehmbaren Rahmen zu halten, etwas gekürzt [1]. Da nicht alle der im Vortest geprüften Kontrollvariablen in der hier präsentierten Studie ausgewertet wurden, werden zunächst nur die relevanten Variablen erläutert. Die restlichen, nicht relevanten Variablen werden im Anschluss lediglich erwähnt.

Figurale und verbale Analogien: Um die Intelligenz der Medizinstudierenden einschätzen zu können, wurde ein Teil aus dem Intelligenz-Struktur-Test verwendet, welcher vor allem das schlussfolgernde Denken erfassen soll: Mit diesem Test von Liepmann sollten die Medizinstudierenden Beziehungen zwischen Wörtern erkennen und weitere Wörter entsprechend dieser Beziehung richtig zuordnen [52]. Des Weiteren wurde ein Test verwendet, in dem visuelle Muster erkannt und schließlich

vervollständigt werden sollten. Dieser Test wurde ursprünglich von Raven & Court entwickelt, die hier eingesetzte validierte Kurzfassung dieses Tests ist jedoch von Bors [53, 54].

Wissenschaftliches Denken: Beim wissenschaftlichen Denken sollten die Medizinstudierenden durch logisches Schlussfolgern eine Situation analysieren und schließlich eine Vorhersage machen oder ein Problem lösen [47]. Dabei sollte nach Lesen einer Problemsituation zuerst mittels Multiple-Choice-Verfahren die Lösung gefunden werden (*Identifizieren von Zusammenhängen*) und anschließend erklärt werden, warum dies die richtige Lösung sein muss (*Erklären der richtigen Zusammenhänge*). Die Auswahl der Items wurde nach Lippman vorgenommen [55].

Vorwissen in Pädagogik und Medizin: Im Wissenstest für Pädagogik und Medizin sollten die Medizinstudierenden im Medizin-Szenario fünf und im Pädagogik-Szenario sechs Fragen mit Multiple-Choice-Verfahren zu den spezifischen Themen „Impfmedizin“ und „Gruppenarbeit“ beantworten. Dies waren auch die Themen der zu bearbeitenden Studien. Der Schwierigkeitsgrad der Fragen war dabei ansteigend, um mit wenigen Fragen ein möglichst breites Spektrum an Wissen abzufragen. Die Vorwissenstests wurden für das KOMPARE-Projekt neu entwickelt.

Erhobene, aber an dieser Stelle nicht berichtete, Daten wurden mit weiteren Tests erhoben. Dazu gehörten Tests zur Evaluierung der Überzeugung der Medizinstudierenden zu evidenzbasierter Medizin/Bildungswissenschaft [56-58] und ein Test zur Evaluierung der Selbsteinschätzung von Interessen, Kenntnissen und Wissen zu den Themen Impfmedizin und Gruppenarbeit. Außerdem wurden zwei Tests zum Strategiewissen und zum Methodenwissen bearbeitet. Des Weiteren nicht berichtet werden die Ergebnisse des erhobenen Tests zum kritischen Denken [59], sowie jeweils drei Fragen in Multiple-Choice-Verfahren zu

Hintergrundinformationen, mit Hilfe derer überprüft wurde, ob die Medizinstudierenden die Hintergrundinformation gelesen und verstanden haben und den Kompetenztest nicht einfach „durchgeklickt“ haben. Diese Daten werden in anderen noch nicht veröffentlichten Arbeiten untersucht.

2.5 STUDIENDURCHFÜHRUNG

Im Zeitraum zwischen dem 20.11.2013 und 6.12.2013 wurden in einem Computerraum der Ludwig-Maximilians-Universität an sechs Tagen insgesamt 14 Testdurchläufe durchgeführt. Ein Testdurchlauf dauerte zwei Stunden und es nahmen jeweils vier bis 20 Medizinstudierende (MW = 11,6 Teilnehmer) teil. Während dieser Testdurchläufe bearbeiteten die Medizinstudierenden den Kompetenztest. Die Voraussetzung zur Teilnahme am Kompetenztest war eine vorherige vollständige Bearbeitung des Online-Vortests. Der Online-Vortest sollte selbstständig von Zuhause aus bearbeitet werden. Dabei gab es keine Zeitbeschränkung.

Die Medizinstudierenden wurden bei den Testdurchläufen nach dem Zufallsprinzip in zwei ungefähr gleich große Gruppen eingeteilt. Insgesamt nahmen 76 Medizinstudierende in Gruppe 1 und 89 Medizinstudierende in Gruppe 2 teil. Um sich an die BESTUEN zu gewöhnen, begannen beide Gruppen mit einem einheitlichen Aufwärmfall, welcher nicht in die Auswertung mit ein ging (Anhang 2). Anschließend bearbeitete Gruppe 1 zuerst das Medizin-Szenario und dann das Pädagogik-Szenario. Gruppe 2 bearbeitete beide Szenarios in umgekehrter Reihenfolge. Der Ablauf der gesamten Studie wird in Tabelle 4 veranschaulicht:

Online-Vortest

-
1. Wörter zuordnen
 2. Muster zuordnen
 3. Wissenschaftliches Denken
 4. Kritisches Denken
 5. Überzeugung der Medizinstudierenden zu evidenzbasierter Medizin/Bildungswissenschaft
 6. Selbsteinschätzung und Interessen zur Impfmedizin und zur Gruppenarbeit
 7. Verständnis zu Fachbegriffen aus Forschungsberichten (Methodenwissen)
 8. Selbsteinschätzung der Kenntnisse zur Varizellenimpfung und zum Gruppenpuzzle

Vortest

-
1. Strategiewissen
 2. Demographische Angaben
 3. Aufwärmfall zum Thema ADHS
(Einüben des Arbeitens mit den BESTUEN)

Kompetenztest

Gruppe 1	Gruppe 2
<ol style="list-style-type: none"> 1. Medizin-Szenario <ul style="list-style-type: none"> - Wissensfragen zur Medizin - Informationssuche - Bewertung von Studien - Entscheidung und Argumentation 	<ol style="list-style-type: none"> 1. Pädagogik-Szenario <ul style="list-style-type: none"> - Wissensfragen zur Pädagogik - Informationssuche - Bewertung von Studien - Entscheidung und Argumentation
<ol style="list-style-type: none"> 2. Pädagogik-Szenario <ul style="list-style-type: none"> - Wissensfragen zur Pädagogik - Informationssuche - Bewertung von Studien - Entscheidung und Argumentation 	<ol style="list-style-type: none"> 2. Medizin-Szenario <ul style="list-style-type: none"> - Wissensfragen zur Medizin - Informationssuche - Bewertung von Studien - Entscheidung und Argumentation

Tabelle 4: Schematischer Ablauf des Vortests und des Kompetenztests..

Um vergleichbare Bedingungen zwischen den einzelnen Testdurchläufen zu schaffen, wurde ein Durchführungsleitfaden erstellt (siehe Anhang 3). Damit alle Medizinstudierenden zur gleichen Zeit im gleichen Abschnitt des Kompetenztests arbeiteten, wurden sie dazu angehalten an bestimmten Stellen des Tests zu warten, bis alle diesen Stand der Bearbeitung erreicht haben. Insgesamt waren die

Medizinstudierenden im zweiten Szenario etwas schneller, als im ersten Szenario: Die Bearbeitungszeit für das Bewerten von vier Studien und der Argumentation war im Mittel beim ersten Szenario des Kompetenztests bei 26,14 Minuten (SD =1,83) und beim zweiten Szenario bei 22,25 Minuten (SD =3,33), unabhängig von der Gruppe. Unabhängig von der Domäne haben sich die Teilnehmer offenbar an die Items des Fragebogens gewöhnt und waren im zweiten Szenario schneller.

2.6 STATISTISCHE ANALYSEN

In den vorhergehenden Kapiteln wurden zum Verständnis des Testablaufs die Entwicklung und der Aufbau des kompletten Kompetenztests der hier vorgestellten Studie beschrieben. Im Folgenden wird auf die in Kapitel 1.7 genannten Fragstellungen eingegangen. Weitere Forschungsschwerpunkte, die sich aus dem Kompetenztest ableiten lassen, werden in anderen Arbeiten bearbeitet.

Um die Teilkompetenz des Bewertens von Studien analysieren zu können, wurde die *Leistung des Bewertens von Studien* mit Hilfe der Items zum Bewerten von Studien (BESTUEN) berechnet: Dazu wurden für jedes Item der Betrag der Differenz zwischen der Lösung der Medizinstudierenden und der in Kapitel 2.3.4 beschriebenen Musterlösung der Experten ermittelt. Der Mittelwert dieser Abweichungen über alle vier Studien wurde schließlich zur Auswertung verwendet. Es wurde somit die mittlere Abweichung von der Musterlösung berechnet: Je kleiner der Wert ist, umso besser ist die Leistung. Der Wert null besagt, dass die Lösung des Medizinstudierenden mit der Musterlösung übereinstimmt (siehe Abbildung 9).

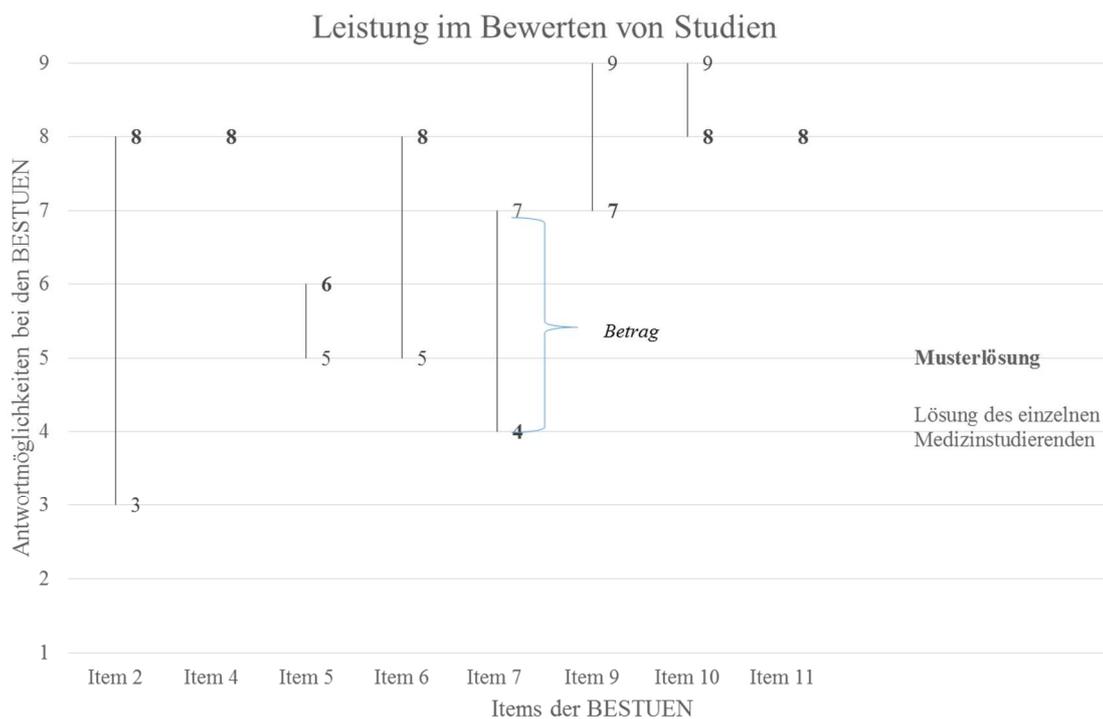


Abbildung 9: **Beispiel zur Berechnung der Leistung des Bewertens von Studien.** Es wurde der Betrag der Differenz von der Lösung der Medizinstudierenden und der Musterlösung berechnet. In diesem Beispiel wäre der Wert für die Leistung bei Item 7: $|7-4| = 3$, also eine Abweichung von 3 Antwortmöglichkeiten von der Musterlösung. Bei Item 4 und Item 11 stimmt die Lösung des Medizinstudierenden mit der Musterlösung überein, die Abweichung ist also null (Autorin: Joana Melle).

Um die in Kapitel 1.7 genannten Hypothesen zu überprüfen, wurden folgende Berechnungen durchgeführt: einfaktorielle Varianzanalysen, zweiseitige t-Tests bei unabhängigen Stichproben und Korrelationsanalysen nach Pearson.

Alle Berechnungen wurden mit dem Statistikprogramm SPSS Version 22 durchgeführt. Das alpha-Fehlerniveau wurde auf $\alpha \leq 0,05$ festgelegt. Nach Bonferroni wurde für die Hypothesen zwei und drei das alpha-Fehlerniveau auf $\alpha \leq 0,025$ korrigiert. Ein p-Wert von über 0,05, bzw. von über 0,025 bei den Hypothesen zwei und drei, wurden demnach als nicht signifikant (n.s.) ausgewiesen.

3 ERGEBNISSE

In den nächsten Unterkapiteln werden die Ergebnisse der statistischen Analysen dargestellt. Dabei wird zunächst auf die vorbereitenden Analysen eingegangen und schließlich werden die Hypothesen überprüft.

3.1 VORBEREITENDE ANALYSEN

Um eine Aussage über die Güte der für die Berechnung verwendeten Tests machen zu können, wurden vorbereitende Analysen durchgeführt, welche im Folgenden erläutert werden. Zunächst werden die Reliabilitätsanalysen der BESTUEN sowie zusätzliche Analysen aufgezeigt. Im Anschluss werden die Varianzen der verschiedenen Testergebnisse dargestellt und ein möglicher Reihenfolgeeffekt untersucht. Zuletzt werden die vorbereitenden Analysen der Vortests beschrieben.

3.1.1 BESTUEN

Wenn man die Messung der *Leistung im Bewerten von Studien* betrachtet, ergibt sich folgende Skalenreliabilität (siehe Tabelle 5):

	Item Anzahl	Minimum der Item- Mittelwerte	Maximum der Item- Mittelwerte	Interne Konsistenz der Skalen (Cronbachs α)	Stichprobe (gültige Fälle von n= 161)
Leistungen im Bewerten von Medizin-Studien	32	1,04	4,44	,57	148
Leistungen im Bewerten von Pädagogik-Studien	40	1,08	3,57	,63	137

Tabelle 5: Reliabilität der „BESTUEN“ nach Berechnung der Leistungen im Bewerten von Studien.

In der folgenden Tabelle 6 sind, zum besseren Verständnis der unterschiedlichen Schwierigkeiten und Dimensionen der einzelnen Items der BESTUEN, die

Mittelwerte, Standardabweichungen und Varianzen der einzelnen Items einer der vier Hauptstudien aus dem Medizinszenario dargestellt:

Deskriptive Statistiken der einzelnen Items einer Studie aus dem Medizinszenario						
	N	Minimum	Maximum	Mittelwert	Standard- abweichung	Varianz
...die Fragestellung zu meiner Entscheidung passt.	160	,00	6,00	1,73	1,47	2,15
...die Maßnahme untersucht wurde, um die es in meiner Entscheidung geht.	161	,00	6,00	3,06	1,94	3,75
...die Erfolgskriterien mit denen bei meiner Entscheidung ausreichend übereinstimmen.	161	,00	4,00	1,70	1,16	1,34
...die Erfolgskriterien mit geeigneten Verfahren erfasst wurden.	160	,00	6,00	1,34	1,14	1,31
...die Merkmale der Personen mit denen der Zielgruppe bei meiner Entscheidung übereinstimmen.	161	,00	5,00	3,08	1,50	2,25
...die Entscheidungsrelevanten Ergebnisse statistisch belastbar sind.	160	,00	7,00	4,34	1,85	3,42
...die entscheidungsrelevanten Ergebnisse eindeutig auf die Maßnahme zurückzuführen sind.	160	,00	6,00	1,50	1,32	1,75
Unter Berücksichtigung der zuvor bewerteten Aspekte spielt die Studie für meine aktuelle Entscheidung eine wichtige Rolle.	160	,00	5,00	1,64	1,20	1,50

Tabelle 6: Deskriptive Statistiken der Leistungen im Bewerten von Studien der einzelnen Items der ersten Studie aus dem Medizinszenario. Die Werte beziehen sich auf die Abweichung von der Musterlösung. Beispielsweise war die maximale Abweichung von der Musterlösung beim ersten Item 6. Im Durchschnitt haben die Medizinstudierenden beim ersten Item 1,73 Wertungseinheiten unterschiedlich zur Musterlösung bewertet.

3.1.2 Vergleich der Varianzen.

In der im Folgenden dargestellten Tabelle 7 werden die deskriptiven Daten der Vortests sowie der BESTUEN dargestellt. Im Gegensatz zu der oben gezeigten Tabelle 8, zeigt die folgende Tabelle nicht dichotomisierte Daten der Vortests auf. Die deskriptiven Statistiken der BESTUEN sind mit und ohne der Berechnung der Leistung (siehe Kapitel 2.6) zu betrachten. Durch diese Zusammenstellung können die Varianzen der einzelnen Tests verglichen werden.

Deskriptive Statistiken						
	N	Minimum	Maximum	Mittelwert	Standard- abweichung	Varianz
1 Verbale Analogien	158	2,00	9,00	5,88	1,57	2,45
2 Figurale Analogien	158	,00	12,00	7,82	2,36	5,58
3 Wissenschaftliches Denken (Identifikation der richtigen Zusammenhänge)	158	,00	4,00	2,21	1,08	1,16
4 Wissenschaftliches Denken (Bewerten der richtigen Zusammenhänge)	158	,00	5,00	3,30	1,24	1,54
5 Leistungen im Bewerten von Pädagogik-Studien (Differenz)	161	1,13	3,53	1,88	,35	,12
6 Leistungen im Bewerten von Medizin-Studien (Differenz)	161	1,28	3,41	2,08	,38	,14
7 Bewerten von Medizin- Studien (Rohwerte der BESTUEN)	161	2,84	9,00	6,58	,91	,83

Tabelle 7: Vergleich der Streuung um den Mittelwert in den Vortests (nicht dichotomisiert) und in den Leistungen beim Bewerten von Studien.

3.1.3 Ausschluss eines Reihenfolgeeffekts

Um auszuschließen, dass die beiden Gruppen, auf Grund der unterschiedlichen Bearbeitungsreihenfolge, unterschiedliche Leistungen im Bewerten von Studien erbracht haben, wurde ein Mittelwertvergleich durchgeführt.

Mittels t-Test bei unabhängigen Stichproben finden sich keine signifikanten Unterschiede zwischen den Mittelwerten der Leistungen im Bewerten von Studien in

Gruppe 1 (n= 75) und der Gruppe 2 (n= 86). Dies ist sowohl beim Vergleich der Leistungen im Bewerten von Medizin-Studien ($t(159)= 3,29$ n.s.), als auch beim Vergleich der Leistungen im Bewerten von Pädagogik-Studien ($t(159)= ,56$ n.s.) der Fall. Es fanden sich also keine Reihenfolgeeffekte in Bezug auf die Leistung im Bewerten von Studien zwischen Gruppe 1, welche mit dem Medizin-Szenario begonnen hatte und Gruppe 2, welche mit dem Pädagogik-Szenario begonnen hatte.

3.1.4 Vortest

In Tabelle 8 ist die ermittelte Reliabilität für die aus dem Vortest verwendeten Tests zur Ermittlung der *Intelligenz (verbale Analogien und figurale Analogien)*, zum *wissenschaftlichen Denken (Identifizieren und Erklären der richtigen Zusammenhänge)* und für die *Wissenstests in der Medizin und in der Pädagogik* zu sehen. Außerdem werden die jeweiligen deskriptiven Statistiken aufgezeigt:

	Item- Anzahl	Minimum der Item- Mittelwerte	Maximum der Item- Mittelwerte	Mittelwert der Item- Mittelwerte	Interne Konsistenz der Skalen (Cronbachs α)	Stichprobe (gültige Fälle von n=161)
Wissenschaftliches Denken (Identifizierung der richtigen Zusammenhänge)	5	,71	,82	,78	,50	119
Wissenschaftliches Denken (Erklärung der richtigen Zusammenhänge)	5	,05	,76	,39	,20	152
Verbale Analogien	10	,37	,99	,66	,42	135
Figurale Analogien	12	,31	,97	,66	,69	144
Vorwissen Medizin	5	,50	1,0	,73	,30	123
Vorwissen Pädagogik	6	,09	,83	,32	,25	161

Tabelle 8: Reliabilität und deskriptive Statistiken der in der Auswertung verwendeten Vortests. Alle Skalen sind dichotomisiert, das heißt die Medizinstudierendenantworten wurden in die zwei Gruppen richtig und falsch zusammengefasst. Die Item-Mittelwerte von Minimum und Maximum befinden sich deswegen zwischen den Werten 0 und 1 (1=richtig und 0=falsch).

Um zu überprüfen, ob der Besuch des Seminars „Impfmedizin“ in Modul 4 zu mehr Vorwissen „Impfmedizin“ führt, wurde eine bivariate Korrelationsanalyse nach Pearson durchgeführt. Danach korrelieren die Leistungen im Vorwissen der Medizin signifikant, mit einem geringen positiven Zusammenhang, mit dem Besuch des Seminars zur Impfmedizin ($r = ,21$; $p < ,01$; $n = 157$). Mit anderen Worten: Wer der Gruppe angehörte, welche das Seminar Impfmedizin besucht hatte (höherer Wert), schnitt auch gut im Test zum Vorwissen in der Medizin ab (höherer Wert bzw. nahe an 1).

Desweiteren konnte ein positiver Zusammenhang festgestellt werden zwischen den beiden Tests zur Messung der Intelligenz „Verbale Analogien“ und „Figurale Analogien“ ($r = ,37$; $p < ,01$; $n = 158$).

3.2 HYPOTHESENÜBERPRÜFUNG

Nachdem die vorbereitenden Analysen dargelegt wurden, wird nun die inferenzstatistische Überprüfung der Hypothesen dargestellt: Zunächst wurde geprüft, ob das Bewerten von Studien fachspezifisch oder generisch ist. Dazu wurden die Hypothesen H1 bis H5 überprüft. Anschließend wurden mögliche Einflussfaktoren des Bewertens von Studien analysiert (H6 und H7).

3.2.1 *Fachspezifisch oder generisch*

Um die Schwankungsbreite des Leistungsniveaus der Medizinstudierenden zu veranschaulichen, wurde die Verteilung der *Leistungen im Bewerten von Studien* in Abbildung 10 dargestellt:

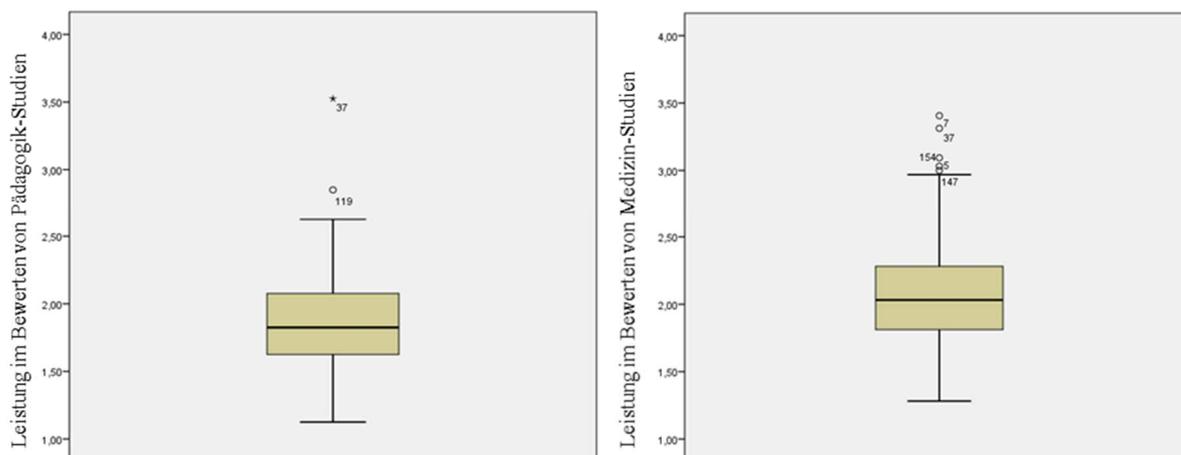


Abbildung 10: Verteilung der Leistungen von Medizinstudierenden im Bewerten von Pädagogik- und Medizin-Studien. ° und * liegen außerhalb der Leistungen von 95% der Medizinstudierenden (° mehr als 1,5 Interquartilsabstände; * mehr als drei Interquartilsabstände). Anmerkung: kleinere Werte bedeuten eine bessere Leistung, d.h. eine geringere Abweichung von der Musterlösung.

Es fällt auf, dass die Medizinstudierenden beim Bewerten von Pädagogik-Studien im Median ($Md = 1,83$) näher an der Musterlösung waren, als beim Bewerten von Medizin-Studien ($Md = 2,03$). Dies bestätigt sich auch in der einfaktoriellen Varianzanalyse:

H1 Die Abweichung von der Musterlösung beim Bewerten der Pädagogik-Studien war kleiner als beim Bewerten der Medizin-Studien ($MW_{\text{Päd}} = 1,88$; $p = ,00$; $MW_{\text{Med}} = 2,08$; $p = ,00$). Der Mittelwertunterschied wurde mittels einfaktorieller Varianzanalyse als signifikant ausgewiesen ($n = 161$; $F(58) = 1,81$; $p = 0,01$) und dies mit einem hohen Effekt ($\eta^2 = ,50$) nach Cohen (Cohen nach Kuckartz) [60].

➔ Die Leistung im Bewerten von Medizin-Studien ist nicht besser, als im Bewerten von Pädagogik-Studien, sondern schlechter.

H2 Es gab keinen signifikanten Unterschied zwischen den Medizinstudierenden, welche bereits das Seminar „Impfmedizin“ absolviert haben ($n = 22$; $MW = 2,23$; $SD = ,40$) und jenen, welche dies noch nicht absolviert haben ($n = 135$;

MW= 2,06; SD= ,37). Ein t-Test ergab, dass der Unterschied zwischen den beiden Gruppen bei der Bewertung der Medizin-Studien nicht signifikant war ($t(155)= 1,99$; n.s.). Die Gruppengröße von 22 hielt der Normalverteilungsüberprüfung nach Shapiro-Wilk stand.

➔ Die *Leistung im Bewerten von Medizin-Studien* ist nicht signifikant besser, wenn die Medizinstudierenden bereits ein zweistündiges *Seminar zum Thema Impfmedizin* absolviert haben.

H3 Mittels t-Test konnte kein signifikanter Mittelwertunterschied zwischen den Studierenden aus Modul 23 ($n= 53$; MW= 2,06; SD= ,42) und den Studierenden aus höheren Semestern ($n= 105$; MW= 2,10; SD= ,36) bei der Leistung beim Bewerten von Medizinstudien festgestellt werden ($t(156)= ,65$; n.s.).

➔ Im Mittel sind Medizinstudierende in *höheren Semestern* nicht besser im *Bewerten von Medizin-Studien*, als *Medizinstudierende des Moduls 23* (klinisches Basisjahr).

H4 Nach einer Korrelationsanalyse nach Pearson gab es einen geringen signifikanten Zusammenhang zwischen dem Bewerten von Medizin-Studien und dem Vorwissen in der Medizin ($r_{\text{Vorwissen Med}}= -,20$; $p < ,05$). Es konnte jedoch kein signifikanter Zusammenhang zwischen dem Pädagogik-Vorwissen und dem Bewerten von Pädagogik- Studien festgestellt werden ($r_{\text{Vorwissen Päd}}= -,01$; n.s.). Es gab auch keine signifikanten Zusammenhänge zwischen dem Vorwissen in der Medizin und dem Bewerten von Studien in der Pädagogik ($r_{\text{Vorwissen Med}}= -,12$; n.s.), sowie zwischen dem Vorwissen in der Pädagogik dem Bewerten von Studien in der Medizin ($r_{\text{Vorwissen Päd}}= -,03$ n.s.).

→ Die *Leistungen im Bewerten von Medizin-Studien* korrelieren gering mit dem *Vorwissen in der Medizin*.⁵

H5 Nach einer bivariaten Korrelationsanalyse nach Pearson korrelierten die Leistungen beim Bewerten der Medizin-Studien signifikant mit einem mittleren Zusammenhang ($r = ,42$; $p < ,01$) mit Leistungen im Bewerten der Pädagogik-Studien (siehe auch Abbildung 11). Wer bei der Bewertung der Pädagogik-Studien nah an den Musterlösungen war, war das auch bei der Bewertung der Medizin-Studien.

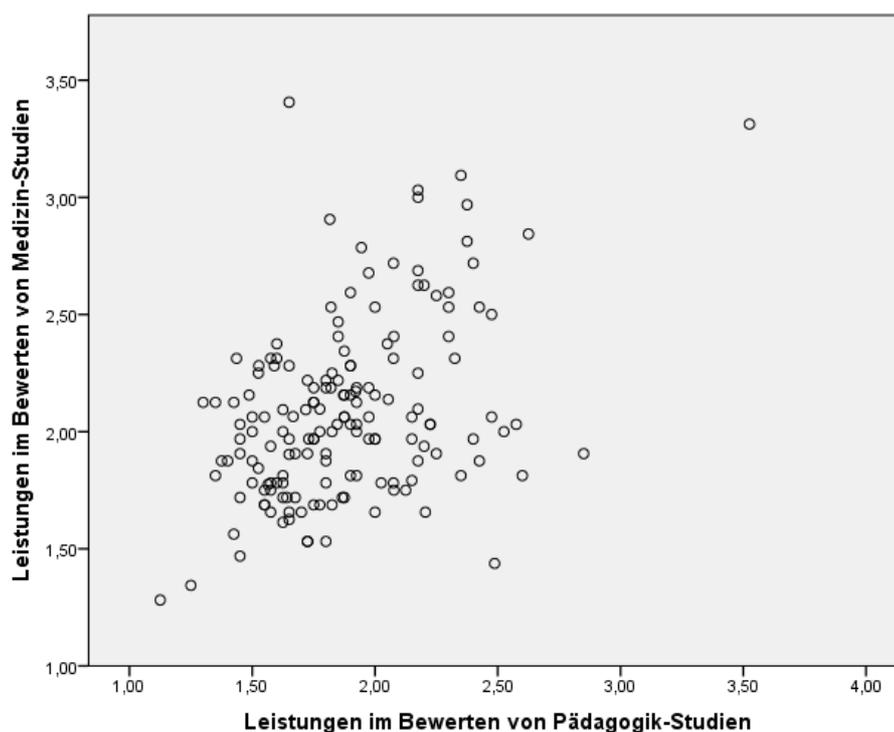


Abbildung 11: *Korrelation zwischen der Leistung im Bewerten von Medizin- und Pädagogik-Studien.*

→ Die *Leistungen im Bewerten von Medizin-Studien* korrelieren signifikant positiv mit den *Leistungen im Bewerten von Pädagogik-Studien*.

⁵Erinnerung: je kleiner der Wert der Leistung ist, desto besser ist diese. Darum ist ein primär negativer Zusammenhang als positiv zu werten.

3.2.2 Einfluss durch Intelligenz und wissenschaftliches Denken

Die letzten beiden Punkte stellen die Ergebnisse zur Überprüfung der möglichen Einflussfaktoren auf die Leistung des Bewertens von Studien dar: *Intelligenz* und *wissenschaftliches Denken*.

H6 Bei den Tests zur *Intelligenz* gab es keine signifikanten Zusammenhänge zwischen dem *Bewerten von Studien* und den *figuralen Analogien* ($r_{\text{Päd}} = -,14$; n.s.; $r_{\text{Med}} = -,13$; n.s.) und auch keinen signifikanten Zusammenhang zwischen dem *Bewerten von Medizin-Studien* und den *verbalen Analogien* ($r_{\text{Med}} = -,04$; n.s.). Jedoch konnte ein geringer signifikanter positiver Zusammenhang zwischen der *Leistung im Bewerten von Pädagogik-Studien* und den *verbalen Analogien* ($r = -,16$; $p = ,04$) festgestellt werden.

➔ Es gibt einen geringen positiven Zusammenhang zwischen dem Test der *verbalen Analogien* und dem *Bewerten von Pädagogik-Studien*.⁶

H7 Bei den Tests zum *wissenschaftlichen Denken* konnten positive Zusammenhänge zwischen dem *„Identifizieren der richtigen Zusammenhänge“* (Teilkompetenz des wissenschaftlichen Denkens) und dem *Bewerten von Pädagogik-Studien* ($r = -,17$; $p = ,03$) sowie dem *Bewerten von Medizin-Studien* ($r = -,21$; $p = ,01$) ermittelt werden. Außerdem gab es einen geringen positiven Zusammenhang zwischen dem *„Erklären der richtigen Zusammenhänge“* (Teilkompetenz des wissenschaftlichen Denkens) und dem *Bewerten von Medizinstudien* ($r = -,16$; $p = ,04$).

➔ Es gibt geringe positive Zusammenhänge zwischen dem *Bewerten von Studien* und den Tests zum *wissenschaftlichen Denken*.

⁶ Kommentar: je kleiner der Wert der Leistung ist, desto besser ist diese. Darum ist ein primär negativer Zusammenhang als positiv zu werten.

4 DISKUSSION

4.1 DAS BEWERTEN VON STUDIEN, EINE GENERISCHE TEILKOMPETENZ MIT FACHSPEZIFISCHEM EINFLUSS

Im Folgenden werden die Ergebnisse der beiden Fragestellungen noch einmal zusammengefasst und schließlich eine erste Schlussfolgerung daraus gezogen.

4.1.1 *Fachspezifisch oder generisch*

Die Hauptfragestellung in dieser Arbeit lautet: *Ist das Bewerten von Studien fachspezifisch oder generisch?* Um diese Hauptfragestellung zu beantworten, wurden mehrere Hypothesen überprüft.

Nach Auswertung der Ergebnisse kann man Folgendes resümieren: Die *Leistung der Medizinstudierenden im Bewerten von Medizin-Studien* ist nicht besser als beim *Bewerten von Pädagogik-Studien*, sondern signifikant schlechter. Nach Besuchen des zweistündigen *Seminars zur Impfmedizin* ist die Leistung im Bewerten von Medizin-Studien nicht besser als ohne den Besuch dieses Seminars. Es gibt keinen signifikanten Unterschied in der Leistung des Bewertens von Studien zwischen den *Medizinstudierenden in höheren Semestern* und den *Medizinstudierenden im klinischen Basisjahr (Modul 23)*. Jedoch gibt es einen geringen Zusammenhang zwischen den Leistungen im *Vorwissenstest Medizin* und den Leistungen im Bewerten von Medizin-Studien.

Ist das Bewerten von Studien nun eine *fachspezifische* Teilkompetenz? Wenn dies so wäre, würde man erwarten, dass *Medizinstudierende* mit höherer Fachkompetenz besser *Medizin-Studien* bewerten können. Fachkompetenz würde man im eigenen Fach erwarten. Die *Medizinstudierenden* hätten in ihrem Fach, der *Medizin*, also bessere Leistungen erbringen müssen. Außerdem würde man erwarten, dass die

Fachkompetenz mit der Studiendauer steigt. Medizinstudierende in höheren Semestern hätten demnach bessere Ergebnisse erzielen müssen. Vorkenntnisse aus dem Seminar zur Impfmedizin deuten ebenfalls auf eine höhere Fachkompetenz hin, so dass mit diesen Vorkenntnissen bessere Leistungen im Bewerten von Studien aus der Impfmedizin hätten erzielt werden müssen. Drei von vier untersuchten Indikatoren für erhöhte Fachkompetenz in der Medizin beeinflussen das Bewerten von Medizin-Studien jedoch nicht. Es kann also argumentiert werden, dass es sich bei der Kompetenz des Bewertens von Studien mit Hilfe der BESTUEN nicht um eine rein fachspezifische Teilkompetenz handelt. Der Zusammenhang zwischen Vorwissen und dem Bewerten von Studien gibt jedoch Hinweise darauf, dass die Teilkompetenz „Bewerten von Studien“ bei Medizinstudierenden trotzdem von Fachwissen aus der Domäne Medizin beeinflusst wird.

Als zweiter Teil der Hauptfragestellung wurde zu Beginn dieser Arbeit die Hypothese aufgestellt, dass das Bewerten von Studien eine *generische* Teilkompetenz ist. Nach Auswertung der Berechnungen ist festzuhalten, dass die *Leistungen im Bewerten von Studien* positiv miteinander korrelieren. Also haben diejenigen Medizinstudierenden, die im Bewerten von Medizin-Studien gute Leistungen erbrachten, auch in der fremden Domäne Pädagogik ähnlich gute Leistungen erzielt. Das Bewerten von Studien scheint damit unabhängig zu sein von der Domäne. Es ist also wahrscheinlich, dass die Teilkompetenz des Bewertens von Studien generisch ist.

Die Ergebnisse zur Hauptfragestellung sind in Abbildung 12 dargestellt:

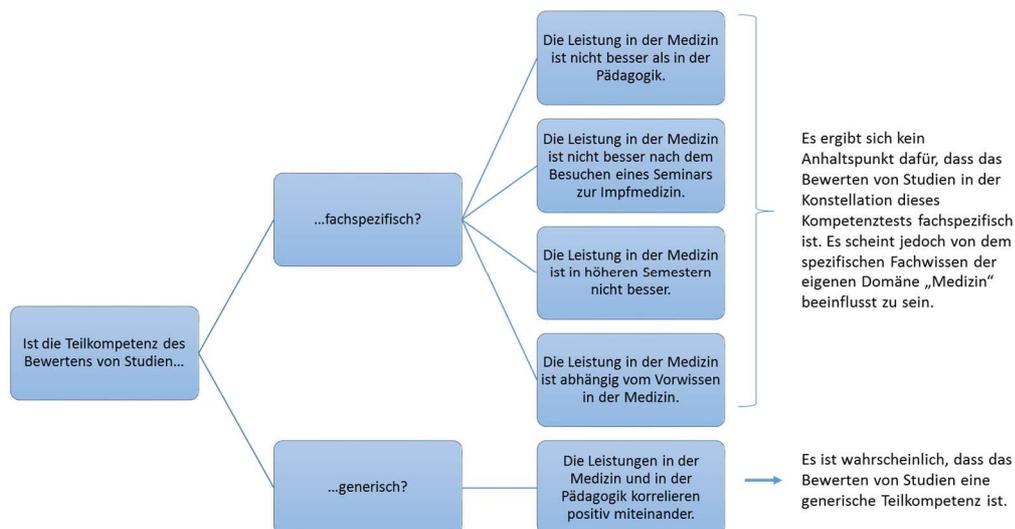


Abbildung 12 Zusammenstellung der Ergebnisse der Hypothesenüberprüfung zur Beantwortung der Hauptfragestellung (Autorin: Joana Melle).

4.1.2 Generisch oder übertragbar

Wird die generische Natur der Teilkompetenz des Bewertens von Studien nur durch Übertragbarkeit von fachspezifischen Komponenten vorgetäuscht? Übertragbarkeit bedeutet nach Larkin, dass neue Aufgaben in einer bisher nicht bekannten Domäne schneller gelernt werden, wenn bereits Erfahrungen aus einer anderen Domäne vorhanden sind [44]. Dies erscheint unwahrscheinlich, da die Medizinstudierenden auf Grund des Curriculums an der LMU bislang keine große Erfahrung im systematischen Bewerten von Studien mitgebracht haben, welche sie übertragen hätten können. Wenn Übertragbarkeit eine Rolle gespielt hätte, wäre es zu erwarten gewesen, dass die Medizinstudierenden in der eigenen Domäne besser gewesen wären und diese Fähigkeit dann in die fremde Domäne übertragen hätten, in der sie dann etwas schlechter gewesen wären. Warum die Medizinstudierenden jedoch im Bewerten von Studien in der fremden Domäne besser waren, als in der eigenen Domäne, wird weiter unten diskutiert.

4.1.3 Einfluss durch Intelligenz oder wissenschaftliches Denken

Zusätzlich zur Hauptfragestellung wurde folgende Frage gestellt: *Wird die Teilkompetenz „Bewerten von Studien“ von Intelligenz und wissenschaftlichem Denken beeinflusst?*

Bei der Untersuchung der möglichen Einflussgröße *Intelligenz*, gibt es keinerlei signifikante Zusammenhänge zwischen dem Bewerten von Medizin-Studien und den verwendeten Aufgaben und auch keinen Zusammenhang zwischen dem Bewerten von Pädagogik-Studien und dem Test „figurale Analogien“. Es korrelieren lediglich die Ergebnisse des Tests „verbale Analogien“ mit der Teilkompetenz des Bewertens von Pädagogik-Studien in geringem Ausmaß. Ein ähnliches Ergebnis konnte bereits in der vorhergehenden Studie des KOMPARE-Projekts gezeigt werden, wobei dort schwache Zusammenhänge zwischen beiden verwendeten Tests und dem Bewerten von Pädagogik-Studien festgestellt wurden [1]. Die Aussage von Hartig und Klieme (2006), welche von einer Korrelation zwischen Kompetenz und Intelligenz ausgingen, kann folglich durch die hier vorgestellte Studie nur sehr eingeschränkt bestätigt werden. Hierbei ist jedoch anzumerken, dass in der hier vorgestellten Studie nur ein Teil des Intelligenz-Strukturtest („verbale Analogien“) sowie ein weiterer Test zur Messung fluider Intelligenz („figurale Analogien“) verwendet wurden, welche zwar einen Einblick in die intellektuellen Fähigkeiten geben können, jedoch nicht alle Dimensionen von Intelligenz umfassen [52, 53]. Es ist an dieser Stelle daher lediglich eine tendenzielle Aussage über den Einfluss der Intelligenz möglich.

Die Einflussgröße *wissenschaftliches Denken* scheint mehr ins Gewicht zu fallen: Geringe positive Zusammenhänge zwischen den Tests zum wissenschaftlichen Denken und dem Bewerten von Medizin- und Pädagogik-Studien können festgestellt werden. Beim Betrachten der beiden Variablen „Identifizieren von wissenschaftlichen

Zusammenhängen“ und „Erklären von wissenschaftlichen Zusammenhängen“, ist aufgefallen, dass die Medizinstudierenden im Mittel recht gut identifizieren konnten, jedoch schlecht erklären (siehe Tabelle 8). Das Erklären der wissenschaftlichen Zusammenhänge ist also offensichtlich viel schwieriger, als das Identifizieren. Wer dennoch gut im Erklären von wissenschaftlichen Zusammenhängen war, also besonders gut im wissenschaftlichen Denken war, der konnte gut Medizin-Studien bewerten. Zwischen dem Erklären von Zusammenhängen und dem Bewerten von Studien in der fremden Domäne Pädagogik gibt es keinen Zusammenhang. Positive Zusammenhänge zwischen dem wissenschaftlichen Denken und der Teilkompetenz des Bewertens von Studien konnten ebenfalls in der vorhergehenden Studie des KOMPARE-Projekts festgestellt werden. Dabei haben Studierende der Sozialwissenschaften domäneneigene Studien bewertet [1]. Wer also gut wissenschaftlich denken und schlussfolgern kann, kann in der eigenen Domäne besser Studien bewerten. Die Domäne scheint also eine gewisse Rolle beim Bewerten von Studien zu spielen.

4.1.4 Schlussfolgerung

Zusammenfassend kann aus den Ergebnissen der hier vorgestellten Studie geschlossen werden, dass die Teilkompetenz des Bewertens von Studien unter den Voraussetzungen des Kompetenztests, zwar nicht fachspezifisch ist, aber von fachspezifischem Wissen beeinflusst wird. Es ist daher davon auszugehen, dass es sich um eine generische Teilkompetenz handelt, die von der Fähigkeit wissenschaftlich zu Denken in geringem Maße beeinflusst wird. Intelligenz hat nach Auswertung dieser Studie nur wenig bis keinen Einfluss auf die Teilkompetenz des Bewertens von Studien. Auf mögliche Limitationen des Kompetenztests wird im folgenden Kapitel eingegangen.

4.2 STÄRKEN UND SCHWÄCHEN DES KOMPETENZTESTS

Im Folgenden wird auf Stärken und Schwächen des Kompetenztests eingegangen. Dabei wird zunächst diskutiert inwiefern sich das Design der hier vorgestellten Studie auf die Ergebnisse auswirkt. Anschließend werden die Angaben der Teilnehmer und die Ergebnisse der vorbereitenden Analysen kritisch begutachtet sowie die inhaltliche Validität diskutiert. Außerdem werden Vorschläge gemacht wie die Forschung über das Bewerten von Studien zukünftig weiter fundiert werden könnte.

4.2.1 Studiendesign

In der hier vorgestellten Studie wurde ein Kompetenztest in quasiexperimentellem Design durchgeführt, welcher zwar einer realen Situation angenähert ist, einer solchen jedoch nicht entspricht. Mit der Ausschaltung der Fremdsprache Englisch, der weitgehenden Vereinfachung der Fachsprache sowie dem Vorliegen von Basis-Hintergrundwissen wurde also der reine Bewertungsvorgang, mit möglichst wenigen Störeffekten betrachtet. Schmälern diese vereinfachten Bedingungen den „Domäneneffekt“ und damit die Aussage dieser Studie? Dies kann nicht ausgeschlossen werden. Jedoch ist es spannend, dass trotz dieser vereinfachten Bedingungen, das Vorwissen in der Domäne Medizin einen geringen Einfluss auf das Bewerten von Medizin-Studien hatte. Der Einfluss des Vorwissens auf die Teilkompetenz des Bewertens von Studien wird also möglicherweise noch unterschätzt.

Beeinflussen diese Maßnahmen auch die Schlussfolgerung zur generischen Natur der Teilkompetenz des Bewertens von Studien? Dies ist unwahrscheinlich, da es sich nicht ausschließt, generisch zu sein und trotzdem von der Domäne beeinflusst zu werden (siehe Argumentation Kapitel 1.5). Die oben genannten Maßnahmen bei der Entwicklung des Kompetenztests waren also notwendig, um die generische Natur der

Teilkompetenz des Bewertens von Studien nicht durch schlechte Englischkenntnisse oder fachspezifischen Einfluss, wie zum Beispiel unbekannte Fachwörter, zu verdecken.

4.2.2 *Teilnehmende*

Nur 22 Teilnehmende haben angegeben, dass sie am Seminar Impfmedizin teilgenommen haben. Entsprechend des Studienfortschritts hätten aber mindestens 25 Teilnehmende das Impfseminar bereits besucht haben müssen. Denn jedenfalls die Studierenden im praktischen Jahr und die Absolventen des Moduls 5 haben nach dem Curriculum das Seminar bereits besucht, da es sich um eine Pflichtveranstaltung im Modul 5 handelt. Dies lässt die Vermutung zu, dass einige Teilnehmende sich nicht mehr an den Besuch erinnern konnten. Ein effektives Wissen aus diesem Seminar konnte daher auch nicht erwartet werden. Es ist also unwahrscheinlich, dass der Zusammenhang zwischen Besuchen des Seminars und dem Bewerten von Medizin-Studien, bei korrekterer Angabe der Teilnehmenden, höher gewesen wäre.

4.2.3 *Vorbereitende Analysen*

Vorwissenstests

Bei genauerer Betrachtung der Testgütekriterien, der im Kompetenztest verwendeten *Vorwissenstests*, fällt eine niedrige interne Konsistenz auf (siehe Tabelle 8): Die Wissenstests zum Vorwissen in der Medizin und in der Pädagogik beinhalten nur wenige Wissensfragen mit gleichbleibendem Thema, aber ansteigender Schwierigkeit. Wegen dieser ansteigenden Schwierigkeit der Items und vor allem auf Grund der niedrigen Item-Anzahl, ist ein hohes Cronbachs-Alpha, also eine hohe Korrelation zwischen den Items, nicht zu erwarten. Diese niedrig zu bewertende interne Validität wurde bewusst in Kauf genommen, um ein breites Spektrum an

Wissen durch wenige Fragen und damit wenig Zeitaufwand abzufragen. Auch gilt es bei der Betrachtung der Reliabilitätsanalysen zu bedenken, dass diese mit den bereits dichotomisierten Werten durchgeführt wurden. Das bedeutet, dass die Antworten der Medizinstudierenden in die zwei Gruppen richtig und falsch zusammengefasst wurden und sich die Item-Mittelwerte zwischen den Werten 0 und 1 (1=richtig und 0=falsch) befanden. Möglicherweise wurden die Cronbachs-Alpha-Werte auch dadurch beeinflusst.

Hinzu kommt eine generelle Kritik an Multiple-Choice-Aufgaben: Nach Schulze können Multiple Choice-Aufgaben nicht belegen, dass der Lernstoff auch tatsächlich beherrscht wird [61].

Insgesamt ist also vor allem die niedrige Anzahl an Items bei den Wissenstests, verbunden mit dem Format „Multiple-Choice“, als limitierend zu bewerten. Warum aus der Auswertung des Vorwissenstests „Medizin“ trotzdem sinnvolle Erkenntnisse gezogen werden können, wird in Kapitel 4.2.5 dargestellt.

Weitere Vortests

Ebenfalls niedrige Cronbachs-Alpha-Werte fallen bei dem Untertest zum wissenschaftlichen Denken „Erklären der richtigen Zusammenhänge“ und bei dem Untertest zur Intelligenz „Verbale Analogien“ auf (siehe Tabelle 8). Die beiden Tests zum wissenschaftlichen Denken „*Erklären der richtigen Zusammenhänge*“ und „*Identifizieren von wissenschaftlichen Zusammenhängen*“ haben jeweils nur fünf Items. Der Schwierigkeitsgrad beim „Erklären der richtigen Zusammenhänge“ war insgesamt höher und inhomogener als beim „Identifizieren von wissenschaftlichen Zusammenhängen“. Die schlechte interne Konsistenz könnte also durch die niedrige Item Anzahl in Kombination mit der Inhomogenität der Schwierigkeit erklärt werden. Warum der Test der „*verbalen Analogien*“ eine nur niedrige interne Konsistenz

aufweist ist schwierig zu sagen, vor allem weil die beiden Tests zur Messung der Intelligenz positiv miteinander korrelierten und der Test „*Figurale Analogien*“ eine recht gute interne Konsistenz aufweist. Auch hier wurden die Reliabilitätsanalysen mit den bereits dichotomisierten Werten durchgeführt. Dies hat die Cronbachs-Alpha-Werte möglicherweise verändert. Außerdem wurden die Vortests online von Zuhause aus bearbeitet und die Motivation während des Bearbeitens wurde weder abgeprüft, noch wurden die Medizinstudierenden beobachtet. Es ist daher lediglich spekulativ, dass die niedrige interne Konsistenz in den oben genannten Vortests durch schlechte Motivation der Medizinstudierenden beeinflusst wurde.

4.2.4 Testgüte der BESTUEN

Reliabilitätsanalyse

Auch die Bewertungssitems „BESTUEN“ haben in der Reliabilitätsanalyse nur mittlere Cronbachs-Alpha-Werte. Dies lässt sich am ehesten dadurch erklären, dass jedes einzelne Bewertungssitem unterschiedliche Dimensionen prüft. Die verschiedenen Dimensionen Qualität (quality), Nützlichkeit (utility), Reichweite (extent), Stärke (strength), Ziel (target), und Kontext (setting) nach Harden, welche in den BESTUEN berücksichtigt sind, stellen unterschiedlich schwierige Herausforderungen an die Medizinstudierenden [18]. Zudem wurden die Bewertungen von vier sehr unterschiedlichen Studien in die Berechnung der Reliabilität mit einbezogen. Nach Panayides sollte jeder Teil eines Tests unterschiedliche Dinge messen, auch wenn damit eine mittlere oder niedrige Homogenität der Items erreicht wird, denn nach Kline können Messungen von menschlichen Verhaltensweisen nicht zu konsistent sein [62, 63]. Auch nach Schmitt können Skalen mit niedrigen Cronbachs-Alpha-Werten in einigen Fällen durchaus

nutzbar sein [64]. Es ist also nachvollziehbar, dass die interne Konsistenz der Skala „BESTUEN“ nicht sehr groß sein kann und dies auch nicht unbedingt sein muss. Stattdessen wird durch Diversität der Items und der vier Studien eine hohe externe Validität erreicht.

In der vorhergehenden Studie des KOMPARE-Projekts, mit einer Probandengruppe aus den Sozialwissenschaften konnten in allen hier verwendeten Tests ähnliche (mittlere bis niedrige) Cronbachs-Alpha ermittelt werden, auch in den Tests zur Messung der Intelligenz und zum wissenschaftlichen Denken. Da nun bei gleichen Tests, mit verschiedenen Probandengruppen in zwei verschiedenen Studien, ähnliche Werte festgestellt werden können, ist die Skala der Bewertungitems BESTUEN, trotz der niedrigen Cronbachs-Alpha-Werte offenbar trotzdem reproduzierbar. Dies gilt es weiter zu untersuchen.

Vergleich der Varianzen

Zur ausführlicheren Analyse der BESTUEN sind weitere Punkte zu diskutieren: Beim Betrachten der Varianzen der Mittelwerte der Ergebnisse der verwendeten Skalen fällt auf, dass *Leistungen der Medizinstudierenden bei den Bewertungitems „BESTUEN“* weniger um den Mittelwert streuten, als die Leistungen in den Vortests. Dies ist eine logische Folge dessen, dass durch die Berechnung der Leistung (Differenz zwischen der Medizinstudierenden- und Musterlösung) die Spanne der Antworten kleiner und damit rein numerisch auch die Streuung um den Mittelwert wurde. So sieht man in Tabelle 7, dass die Varianzen der Rohwerte der BESTUEN deutlich größer waren, als bei den Leistungen des Bewertens von Studien. Zudem trug die Mittelung der acht, beziehungsweise zehn, verschiedenen Items aller vier Studien

zum Entstehen der niedrigen Varianzen bei. Bei Betrachtung der einzelnen Items waren die Varianzen nämlich deutlich größer (siehe Tabelle 6).

4.2.5 *Inhaltliche Validität*

Die *Leistungen im Bewerten von Studien* in der Medizin waren nicht besser, sondern sogar schlechter als in der Pädagogik. Woran könnte das liegen? Wurde überhaupt gemessen, was gemessen werden sollte? In der vorhergehenden Studie des KOMPARE-Projekts konnte eine „zufriedenstellende Übereinstimmung der Experten“ bei Bearbeitung der BESTUEN festgestellt werden, so dass die Autoren von einer inhaltlich „adäquaten“ Erfassung des Bewertens von Pädagogik-Studien ausgingen [1]. Die BESTUEN und die zugehörige Musterlösung in der Pädagogik stimmten also mit einem Konsens der befragten Experten ausreichend überein und sind damit inhaltlich validiert. Die Erstellung der Musterlösung im Medizin-Szenario der hier vorgestellten Studie erfolgte nach gleichem Konzept, ebenfalls unter Mitwirkung von Experten. Inhaltlich ist also das Bewerten von Studien mit den BESTUEN und deren Musterlösungen ausreichend gut abgebildet. Daher erscheint es zwar möglich, aber nicht vordergründig, dass die Erklärung für die unterschiedlichen Leistungen in einer Erstellung der beiden Musterlösungen, zu unterschiedlichen Zeitpunkten, von unterschiedlichen Experten besteht. Die Musterlösung für das Pädagogik-Szenario wurde nämlich bereits in der vorhergehenden Studie [1] und die Musterlösung für das Medizin-Szenario in der aktuellen Studie entwickelt. Eher scheint der Grund für die unterschiedlichen Leistungen in den beiden Domänen in der unterschiedlichen Schwierigkeit der Bewertung zu liegen: Die Studien der Domäne Medizin waren schwieriger zu bewerten als in der Domäne Pädagogik.

Bei der Frage nach der inhaltlichen Validität sollte auch diskutiert werden, inwiefern das Themen „Impfmedizin“ für Medizinstudierende tatsächlich die eigene

Domäne „Medizin“ widerspiegelt und ob das Thema „Gruppenpuzzle“ wirklich domänenfremd ist. Für die Medizinstudierenden war das Thema „Gruppenpuzzle“ mit Sicherheit domänenfremd, denn abgesehen von der eigenen Erfahrung als Schüler, ist das Thema Pädagogik nicht Teil des Medizinstudiums. Das Thema Impfmedizin gehört mit Sicherheit in die Domäne „Medizin“. Jedoch wurde dieses Thema im Medizinstudium an der LMU bisher nicht so sehr fokussiert wie andere Themen (z.B. „Herzinfarkt“) [65]. Die Medizinstudierenden waren also sehr wahrscheinlich keine Experten im Thema „Impfmedizin“. Es zeigte sich jedoch, dass die Medizinstudierenden im Vorwissen der (Impf-)Medizin im Mittel deutlich besser waren als im Vorwissen der Pädagogik (Tabelle 8). Außerdem fordert das Thema „Impfmedizin“ eine naturwissenschaftliche, „medizinische“ Denkweise heraus, die durchaus im Medizinstudium erlernt wird. Es kann also davon ausgegangen werden, dass das Thema Impfmedizin die eigene Domäne „Medizin“ widerspiegelt.

Zur Diskussion der inhaltlichen Validität gehört auch die Frage, inwiefern mit dem *Vorwissenstest* „Medizin“ tatsächlich das Konstrukt „fachspezifisches Vorwissen“ erfasst wurde, nachdem im Kapitel 4.2.3 bereits die geringe interne Validität dieses Vorwissenstests kritisch begutachtet wurde. Der geringe Zusammenhang zwischen dem Besuchen des Impfseminars und der Leistung im Vorwissenstest Medizin, welcher Inhalte zur Impfmedizin abfragte, lässt jedoch darauf schließen, dass der Vorwissenstest Medizin das Konstrukt „Wissen zu Impfmedizin“ auch tatsächlich testet.

4.2.6 *Weitere Forschung*

Um die BESTUEN mit den Musterlösungen weiter inhaltlich zu validieren, wäre eine größer angelegte Studie mit Experten aus der pädagogischen und medizinischen Wissenschaft notwendig. Außerdem wäre es interessant, eine weitere Studie im

gleichen Format wie die hier vorgestellte Studie durchzuführen. Jedoch sollte für die Domäne „Medizin“ ein Thema gewählt werden, das im Medizinstudium mehr fokussiert wird. Dazu wäre es sinnvoll, die Vorwissenstests zu erweitern, auch im Hinblick auf eine höhere interne Konsistenz. Möglicherweise tritt dann der Einfluss des fachspezifischen Vorwissens mehr in den Vordergrund. Um den Einfluss der Intelligenz auf das Bewerten von Studien weiter zu untersuchen, sollten alle Dimensionen von Intelligenz berücksichtigt werden und die Tests zur Intelligenz bei einer zukünftigen Studie erweitert werden.

Um die generische Natur des Bewertens von Studien in der tatsächlichen Praxis zu verstehen und zu überprüfen, ist weitere Forschung nötig: Man könnte diese Studie noch um ein Szenario in einer weiteren Domäne erweitern, um diese schließlich mit Probandengruppen aus verschiedenen Domänen durchzuführen. Um tiefergehend zu untersuchen, welchen Einfluss das Vorwissen hat, könnte man dies ohne Hintergrundinformation und mit einer direkten Übersetzung der englischen Forschungsliteratur ohne Vereinfachung und Zusammenfassung durchführen. Jedoch sollte dafür sichergestellt werden, dass die Probanden überhaupt ein entsprechendes Vorwissen haben. Da bereits gezeigt werden konnte, dass fachspezifisches Vorwissen einen Einfluss auf das Bewerten von Studien hat, würde komplett fehlendes Vorwissen möglicherweise zu einer Messung führen, welche das Konstrukt der Teilkompetenz „Bewerten von Studien“ nicht ausreichend erfassen kann.

Diesbezüglich wird auch die Auswertung einer weiteren Studie des KOMPARE-Projekts an der Bergischen Universität Wuppertal spannend. Diese wurde die in gleicher Form wie in München an der LMU mit Medizinstudierenden, nur dort mit Pädagogikstudierenden durchgeführt. Die Ergebnisse aus Wuppertal sind noch in Bearbeitung und werden an anderer Stelle ausgewertet. Wenn sich ähnliche

Ergebnisse in Wuppertal feststellen ließen, wie in der hier vorgestellten Studie, könnte dies weitere Indizien dafür liefern, dass die Teilkompetenz des Bewertens von Studien generisch ist und trotzdem von fachspezifischen Merkmalen abhängt.

4.2.7 Zusammenfassung

Zusammenfassend kann trotz einiger Limitationen dieser Studie argumentiert werden, dass die Teilkompetenz des Bewertens von Studien mit Hilfe der BESTUEN generisch ist, aber trotzdem von fachspezifischem Wissen beeinflusst wird. Die Fähigkeit wissenschaftlich zu Denken beeinflusst das Bewerten von Studien. Der Intelligenz kann nur ein sehr geringer Einfluss auf das Bewerten von Studien nachgewiesen werden. Weitere Forschung ist nötig (siehe 4.2.6).

4.3 ERKENNTNISSE FÜR DIE LEHRE DER EVIDENZBASIERTEN MEDIZIN

In Kapitel 1.1 wurde erläutert, warum evidenzbasierte Medizin unbedingt schon im Studium erlernt werden sollte. Dass die Kompetenz evidenzbasierte Medizin auszuführen grundsätzlich erlernbar ist, geht aus einer systematischen Übersichtsarbeit von Ilic & Maloney hervor. Darin wurden neun randomisiert kontrollierte Studien zu unterschiedlichen Methoden der Lehre der evidenzbasierten Medizin untersucht. Die Autoren zeigten in allen Studien nach EbM-Schulung eine Verbesserung in Kompetenz, Wissen, Einstellungen und Verhalten [66]. Doch kann evidenzbasierte Medizin so früh wie möglich gelehrt werden, also schon ab dem ersten Semester? In einem systematischen Review zeigen Ahmadi et al. auf, dass es bereits Strategien gibt, um Medizinstudierenden in der Vorklinik evidenzbasierte Medizin zu lehren [67]. Damit konnten sie bestätigen, dass große medizinische Fachkompetenz nicht zwingend notwendig ist, um das Erlernen von evidenzbasierter

Medizin zu beginnen, auch wenn der tatsächliche Nutzen für die evidenzbasierte Praxis noch nicht aufgezeigt werden konnte.

Wie können nun die Ergebnisse dieser Studie für das Erlernen der evidenzbasierten Medizin hilfreich sein? Eine generische Teilkompetenz des Bewertens von Studien kann für die Lehre in der evidenzbasierten Medizin bedeuten, dass in der Lehre in den Domänen variiert werden kann. Man könnte also durchaus in der Lehre des Bewertens von Studien Beispiele aus verschiedenen Fachgebieten verwenden. Gleichzeitig zeigen die Ergebnisse der hier vorgestellten Studie auf, dass es nicht sinnvoll erscheint, ohne Vorwissen Studien zu bewerten. Wenn also Studien aus verschiedenen Fachgebieten verwendet werden, dann ist ein gewisses Vorwissen trotzdem hilfreich für das Bewerten von Studien. Ein Beginn der Lehre im ersten Semester könnte also möglich sein, sobald Wissen zu den einzelnen vorklinischen Fächern vorhanden ist und die zu bewertenden Studien aus dem Themenbereich dieses vorklinischen Wissens stammen.

→ Prinzipiell können für die Lehre des Bewertens von Studien Übungsszenarien aus unterschiedlichen Domänen verwendet werden, jedoch sollte bereits ein gewisses Vorwissen zu dieser Domäne vorhanden sein.

Nach einem Erfahrungsbericht von Ohletz & Sellenthin ist das Lesen von englischer Fachsprache eine besondere Herausforderung für Medizinstudierende in der Vorklinik [68]. Das Übersetzen der Originalliteratur, wie es auch in der hier vorgestellten Studie gemacht wurde, scheint also wichtig zu sein um das Bewerten von Studien bereits in der Vorklinik einzuführen.

Die BESTUEN ermöglichen eine systematische Bewertung von Studien. In der hier vorgestellten Studie haben Medizinstudierende damit Studien bewertet, ohne

große Erfahrung mit dieser Methode zu haben. Dabei konnte eine schnelle Adaptation an dieses Bewertungssystem beobachtet werden (siehe Kapitel 2.5). Die Kombination aus einer vorgegebenen Bewertungssystematik und der Aufbereitung der Studien (Übersetzung ins Deutsche und wenig Fachsprache) waren hier vermutlich hilfreich.

→ In der Didaktik in der Medizin könnte das Bewerten von Studien mit Hilfe der BESTUEN, in Kombination mit einer aufbereiteten Fachliteratur und Übersetzung von englischer Fachsprache, ein erster Schritt in der Lehre des Bewertens von Studien sein.

Ziel dieser Arbeit war es, die Natur der Teilkompetenz „Bewerten von Studien“ besser zu verstehen: Es kann ein schwacher Einfluss der Fähigkeit des wissenschaftlichen Denkens auf die Teilkompetenz „Bewerten von Studien“ nachgewiesen werden. Intelligenz scheint nur marginalen Einfluss auf die Teilkompetenz des Bewertens von Studien zu nehmen. Der Einfluss von fachspezifischen Komponenten scheint zu existieren, wie stark diese Beeinflussung ist, muss jedoch noch weiter untersucht werden. Vor Allem aber werden in der hier vorgestellten Arbeit Indizien dafür geliefert, dass die Teilkompetenz des Bewertens von Studien mit Hilfe der BESTUEN generisch ist. Die hier gelieferten Erkenntnisse sollen schließlich als ein weiterer Baustein zur Entwicklung der Lehre der evidenzbasierten Medizin im Medizinstudium genutzt werden: Dies ist von besonderer Wichtigkeit, denn je besser die einzelnen Teilkompetenzen der evidenzbasierten Medizin verstanden werden, umso zielgerichteter kann sie gelehrt werden. Nur so besteht die Chance, dass aus dem Studium der Medizin junge Ärztinnen und Ärzte hervorgehen, welche evidenzbasiert handeln und somit Patientinnen und Patienten optimal und individuell versorgen können.

5 LITERATURVERZEICHNIS

1. Trempler K, H.A., Wecker C, Kiesewetter J, Wermelt M, Fischer F, Fischer M, Gräsel C *Nutzung von Evidenz im Bildungsbereich – Validierung eines Instruments zur Erfassung von Kompetenzen der Informationsauswahl und der Bewertung von Studien*. Z Pädagogik, 2015. **Beiheft 61**: p. 144-166.
2. Sackett, D.L., et al., *Evidence based medicine: what it is and what it isn't*. 1996. Clin Orthop Relat Res, 2007. **455**: p. 3-5.
3. Porzsolt, F. and H. Leonhardt-Huober, *Teaching Evidence-based Medicine*, in *Evidence-based Practice in Medicine and Health Care*, R. ter Meulen, et al., Editors. 2005, Springer Berlin Heidelberg. p. 35-43.
4. Guyatt, G.H., et al., *Practitioners of evidence based care - Not all clinicians need to appraise evidence from scratch but all need some skills*. British Medical Journal, 2000. **320**(7240): p. 954-955.
5. Raspe, H., *Konzept und Methoden der Evidenz-basierten Medizin: Besonderheiten, Stärken, Grenzen, Schwächen und Kritik*. Institut für Sozialmedizin Universitätsklinikum Schleswig-Holstein, Campus Lübeck. , 2007.
6. Swennen, M.H.J.M.D., et al., *Doctors' Perceptions and Use of Evidence-Based Medicine: A Systematic Review and Thematic Synthesis of Qualitative Studies*. Academic Medicine, 2013. **88**(9): p. 1384-1396.
7. Yost, J., D. Ciliska, and M. Dobbins, *Evaluating the impact of an intensive education workshop on evidence-informed decision making knowledge, skills, and behaviours: a mixed methods study*. BMC Medical Education, 2014. **14**: p. 9.
8. Harbour, R. and J. Miller, *A new system for grading recommendations in evidence based guidelines*. Bmj, 2001. **323**(7308): p. 334-6.
9. Bundesministerium für Gesundheit, *Evidenzbasierte Medizin*. 2016.[1.2.2016] <https://www.bundesgesundheitsministerium.de/service/begriffe-von-a-z/e/evidenzbasierte-medizin.html>
10. Langer, G., et al., *GRADE-Leitlinien: 1. Einführung – GRADE-Evidenzprofile und Summary-of-Findings-Tabellen*. Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen, 2012. **106**(5): p. 357-368.
11. Baethge, C., *Evidenzbasierte Medizin: In der Versorgung angekommen, aber noch nicht heimisch*. Dtsch Arztebl International, 2014. **111**(39): p. A-1636.

12. The Cochrane Collaboration, *Leitlinien*. 2016 [1.2.2016]; Available from: <http://www.cochrane.de/de/leitlinien>.
13. Gerhardus, A., C. Muth, and D. Luhmann, *Anpassung des „Curriculum Evidenzbasierte Medizin „für unterschiedliche Zielgruppen. Erfahrungen aus dem Aufbaustudiengang Public Health in Hannover und der Humanmedizinausbildung in Lubeck*. ZEITSCHRIFT FÜR ARZTLICHE FORTBILDUNG UND QUALITÄTSSICHERUNG, 2004. **98**(2): p. 155-160.
14. Taylor, D.K. and J. Buterakos, *Evidence-based medicine: not as simple as it seems*. Academic Medicine, 1998. **73**(12): p. 1221-2.
15. ÄAppO, *Approbationsordnung für Ärzte vom 27. Juni 2002 (BGBl. I S. 2405), die zuletzt durch Artikel 2 der Verordnung vom 2. August 2013 (BGBl. I S. 3005) geändert worden ist*. 2002.
16. Wissenschaftsrat, *Empfehlungen zur Weiterentwicklung des Medizinstudiums in Deutschland auf Grundlage einer Bestandsaufnahme der humanmedizinischen Modellstudiengänge*. 2014. **Drs. 4017-14**: p. 1-113.
17. Fischer, M.R., D. Bauer, and K. Mohn, *Finally finished! National Competence Based Catalogues of Learning Objectives for Undergraduate Medical Education (NKLM) and Dental Education (NKLZ) ready for trial*. Endlich fertig! Nationale Kompetenzbasierte Lernzielkataloge Medizin (NKLM) und Zahnmedizin (NKLZ) gehen in die Erprobung., 2015. **32**(3): p. 1-5.
18. Harden, R.M., et al., *Best Evidence Medical Education*. Adv Health Sci Educ Theory Pract, 2000. **5**(1): p. 71-90.
19. Fritsche, L., et al., *Do short courses in evidence based medicine improve knowledge and skills? Validation of Berlin questionnaire and before and after study of courses in evidence based medicine*. BMJ, 2002. **325**(7376): p. 1338-1341.
20. Ramos, K.D., S. Schafer, and S.M. Tracz, *Validation of the Fresno test of competence in evidence based medicine*. Bmj, 2003. **326**(7384): p. 319-21.
21. Hartig, J. and E. Klieme, *Kompetenz und Kompetenzdiagnostik*, in *Leistung und Leistungsdiagnostik*, K. Schweizer, Editor. 2006, Springer Berlin Heidelberg. p. 127-143.
22. Fernandez, N., et al., *Varying conceptions of competence: an analysis of how health sciences educators define competence*. Medical Education, 2012. **46**(4): p. 357-365.

23. Tilson, J., et al., *Sicily statement on classification and development of evidence-based practice learning assessment tools*. BMC Medical Education, 2011. **11**(1): p. 78.
24. Brand-Gruwel, S., I. Wopereis, and A. Walraven, *A descriptive model of information problem solving while using internet*. Computers & Education, 2009. **53**(4): p. 1207-1217.
25. Dawes, M., et al., *Sicily statement on evidence-based practice*. BMC Medical Education, 2005. **5**(1): p. 1.
26. Straus, S.E. and D.L. Sackett, *Using research findings in clinical practice*. Bmj, 1998. **317**(7154): p. 339-42.
27. Shaneyfelt, T., et al., *Instruments for evaluating education in evidence-based practice - A systematic review*. JAMA, 2006. **296**(9): p. 1116 - 1127.
28. Harris, J., et al., *Are journal clubs effective in supporting evidence-based decision making? A systematic review*. BEME Guide No. 16. Medical Teacher, 2011. **33**(1): p. 9-23.
29. Bowden, S.C., E.J. Harrison, and D.W. Loring, *Evaluating Research for Clinical Significance: Using Critically Appraised Topics to Enhance Evidence-based Neuropsychology*. The Clinical Neuropsychologist, 2013. **28**(4): p. 653-668.
30. Meehl, P.E., *What can the clinician do well? Psychodiagnosis selected papers*. 1973: p. 165 -173.
31. Clearinghouse, KT. *Diagnosis Critical Appraisal Worksheet*. 2000-2014 [2.1.2016]; Available from: <http://ktclearinghouse.ca/ceb/teaching/worksheets/diagnosis>.
32. Cronbach, L.J. and K. Shapiro, *Designing evaluations of educational and social programs*. 1982: Jossey-Bass.
33. Cook, T.D., D.T. Campbell, and A. Day, *Quasi-experimentation: Design & analysis issues for field settings*. Vol. 351. 1979: Houghton Mifflin Boston.
34. Bibliographisches Institut GmbH, *Domäne*. 2016 [2.1.2016]; Available from: <http://www.duden.de/rechtschreibung/Domaene>.
35. Mandl, H., H. Gruber, and A. Renkl, *Kontextualisierung von Expertise*. 1991.
36. Bibliographisches Institut GmbH, *fachspezifisch*. 2016 [1.2.2016]; Available from: <http://www.duden.de/rechtschreibung/fachspezifisch>.

37. Hollender, D.N., et al. *Formulierungshilfen für Modulhandbücher Handreichung zur Verstärkung der Kompetenzorientierung 2010* [cited 2016 12.07.2016]; Available from: <http://download.hrz.tudarmstadt.de/media/HDA/Handreichung.pdf>.
38. Alexander, P.A. and J.E. Judy, *The Interaction of Domain-Specific and Strategic Knowledge in Academic Performance*. *Review of Educational Research*, 1988. **58**(4): p. 375-404.
39. Marti, M. and M. Ulmi, *Lesend denken–Strategien im Umgang mit Fachtexten*. Kruse, Otto/Berger, Katja/Ulmi, Marianne (Hgg.): *Prozessorientierte Schreibdidaktik. Schreibtraining für Schule, Studium und Beruf*. Bern: Haupt Verlag, 2006: p. 175-194.
40. Strijbos, J., N. Engels, and K. Struyven, *Criteria and standards of generic competences at bachelor degree level: A review study*. *Educational Research Review*, 2015. **14**: p. 18-32.
41. Young, J., & Chapman, E, *Generic competency frameworks: A brief historical overview*. . *Education Research and Perspectives*, 2010. **37**.
42. Clanchy, J. and B. Ballard, *Generic skills in the context of higher education*. *Higher Education Research and Development*, 1995. **14**(2): p. 155-166.
43. Leutner, D., et al., *Problemlösefähigkeit als fächerübergreifende Kompetenz*, in *Problemlösekompetenz von Schülerinnen und Schülern*, E. Klieme, D. Leutner, and J. Wirth, Editors. 2005, VS Verlag für Sozialwissenschaften. p. 11-19.
44. Larkin, J.H., *What kind of knowledge transfers?*, in *Knowing, learning, and instruction: Essays in honor of Robert Glaser*. 1989, L Erlbaum Associates: Hillsdale. p. 283-305.
45. Billing, D., *Teaching for transfer of core/key skills in higher education: Cognitive skills*. *Higher Education*, 2007. **53**(4): p. 483-516.
46. Gruber, H. and E. Stamouli, *Intelligenz und Vorwissen*, in *Pädagogische Psychologie*, E. Wild and J. Möller, Editors. 2015, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 25-44.
47. Lawson, A.E., et al., *Development of Scientific Reasoning in College Biology: Do Two Levels of General Hypothesis-Testing Skills Exist?* *Journal of research in Science Teaching*, 2000. **37**(1): p. 81-101.
48. Zwolsman, S.E., et al., *How learning style affects evidence-based medicine: a survey study*. *Bmc Medical Education*, 2011. **11**: p. 8.

49. Ständige Impfkommission, *Epidemiologisches Bulletin Nr 34*, in *Epidemiologisches Bulletin*. 2014, Robert Koch Institut: Berlin. p. 307.
50. Eidgenössische Kommission für Impffragen (EKIF), Bundesamt für Gesundheit, *Schweizerischer Impfplan 2014. Richtlinien und Empfehlungen*. 2014, Bundesamt für Gesundheit: Bern.
51. Wermelt, M., et al., „Kompetent Argumentieren mit Evidenzen“ – *Entscheidungen aufgrund wissenschaftlicher Evidenzen – eine domänenspezifische oder übergreifende Fähigkeit?*, in *Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA)*. Hamburg, 25.-27.09.2014. 2014, Düsseldorf: German Medical Science GMS Publishing House; 2014. .
52. Liepmann D, B.A., Brocke B, Amthauer R, *I-S-T 2000 R. Intelligenz-Struktur-Test 200R* Vol. 2. 2007, Göttingen: Hogrefe.
53. Raven, J.C. and J.H. Court, *Raven's progressive matrices and vocabulary scales*. 1998: Oxford Psychologists Press.
54. Bors, D.A. and T.L. Stokes, *Raven's Advanced Progressive Matrices: Norms for first-year university students and the development of a short form*. Educational and Psychological Measurement, 1998. **58**(3): p. 382-398.
55. Lippman, J., *Improving and Predicting Novice Reasoning about the Evidentiary Connection between Studies and Theories*. 2011: University of Illinois at Chicago.
56. Jette, D.U., et al., *Evidence-based practice: beliefs, attitudes, knowledge, and behaviors of physical therapists*. Physical Therapy, 2003. **83**(9): p. 786-805.
57. Johnston, J.M., et al., *Clinical learning The development and validation of a knowledge, attitude and behaviour questionnaire to assess undergraduate evidence-based practice teaching and learning*. Medical Education, 2003. **37**(11): p. 992.
58. Young, J.M. and J.E. Ward, *Evidence-based medicine in general practice: beliefs and barriers among Australian GPs*. Journal of evaluation in clinical practice, 2001. **7**(2): p. 201-210.
59. Ennis R H, M.J., Tomko T N, *Cornell Critical Thinking Tests Level X and Level Z Manual*. Vol. 5. 2005: The Critical Thinking co.
60. Kuckartz, U., et al., *Statistik: eine verständliche Einführung*. 2013: Springer-Verlag.

61. Schulze J, D.S., Nürnberger F, Ochsendorf F, Schäfer V, Brandt C *Einfluss des Fragenformates in Multiple-choice-Prüfungen auf die Antwortwahrscheinlichkeit: eine Untersuchung am Beispiel mikrobiologischer Fragen*. GMS Z Med Ausbild, 2005. **22**(4): p. Doc218.
62. Kline, P., *Psychometrics and psychology*. 1979, London, United Kingdom: Academic Press.
63. Panayides, P., *Coefficient Alpha*. Europe's Journal of Psychology, 2013. **9**(4): p. 687-696.
64. Schmitt, N., *Uses and abuses of coefficient alpha*. Psychological Assessment, 1996. **8**(4): p. 350-353.
65. Beltermann, E., et al., *Lehre aktiv in der Impf- und Reisemedizin – Ein neues praxisnahes Format zur Förderung der Beratungskompetenz von Studierenden*. GMS Z Med Ausbild, 2015. **32**(3): p. doc 28.
66. Ilic, D. and S. Maloney, *Methods of teaching medical trainees evidence-based medicine: a systematic review*. Medical Education, 2014. **48**(2): p. 124-135.
67. Ahmadi, S.-F., H.R. Baradaran, and E. Ahmadi, *Effectiveness of teaching evidence-based medicine to undergraduate medical students: A BEME systematic review*. Medical Teacher, 2015. **37**(1): p. 21-30.
68. Ohletz A, S.C., *Integration der Evidence-Based Medicine in das Studium-Hexerei?* Med Ausbildung, 2000. **17**: p. 60-63.
69. Kraut, A.A., et al., *Comorbidities in ADHD children treated with methylphenidate: a database study*. BMC Psychiatry, 2013. **13**: p. 11.
70. Egeland, J., A.K. Aarli, and B.K. Saunes, *Few Effects of Far Transfer of Working Memory Training in ADHD: A Randomized Controlled Trial*. PLoS One, 2013. **8**(10): p. e75660.
71. Baxter, R., et al., *Long-term effectiveness of varicella vaccine: a 14-Year, prospective cohort study*. Pediatrics, 2013. **131**(5): p. e1389-96.
72. Agarwal, N., et al., *Chemical reconstruction of skin scars therapy using 100% trichloroacetic Acid in the treatment of atrophic facial post varicella scars: a pilot study*. J Cutan Aesthet Surg, 2013. **6**(3): p. 144-7.

6 ABBILDUNGSVERZEICHNIS

Abbildung 1: Übersicht über das KOMPARE-Projekt.....	10
Abbildung 2: Die fünf Stufen des evidenzbasierten Arbeitens nach Dawes (2005).	12
Abbildung 3: QUESTS, sechs Dimensionen zur Evaluierung von Studien in der Medizindidaktik nach RM Harden (2000).....	14
Abbildung 4: Übersicht über generische, fachspezifische und übertragbare Kompetenzen.	19
Abbildung 5: Fünf Hypothesen zur Bearbeitung der Hauptfragstellung.....	22
Abbildung 6: Altersverteilung der teilnehmenden Medizinstudierenden.....	27
Abbildung 7: Reihenfolge der zu bearbeitenden Aufgaben eines Szenarios.....	29
Abbildung 8: Die BESTUEN zum Bewerten von Medizinstudien.	37
Abbildung 9: Beispiel zur Berechnung der Leistung des Bewertens von Studien.....	43
Abbildung 10: Verteilung der Leistungen von Medizinstudierenden im Bewerten von Pädagogik- und Medizin-Studien.	49
Abbildung 11: Korrelation zwischen der Leistung im Bewerten von Medizin- und Pädagogik-Studien.	51
Abbildung 12 Zusammenstellung der Ergebnisse der Hypothesenüberprüfung zur Beantwortung der Hauptfragestellung.....	55

7 TABELLENVERZEICHNIS

Tabelle 1: Zugehörigkeit der Medizinstudierende zu den einzelnen Semestern im Medizinstudium an der LMU.	27
Tabelle 2: Entwicklung der BESTUEN.	35
Tabelle 3: Die Bewertungsitems BESTUEN, mit Hilfe derer die Studierenden die vier Hauptstudien bewerten sollten	36
Tabelle 4: Schematischer Ablauf des Vortests und des Kompetenztests.	41
Tabelle 5: Reliabilität der „BESTUEN“ nach Berechnung der Leistungen im Bewerten von Studien.	44
Tabelle 6: Deskriptive Statistiken der Leistungen im Bewerten von Studien der einzelnen Items der ersten Studie aus dem Medizinszenario.	45
Tabelle 7: Vergleich der Streuung um den Mittelwert in den Vortests (nicht dichotomisiert) und in den Leistungen beim Bewerten von Studien.....	46
Tabelle 8: Reliabilität und deskriptive Statistiken der in der Auswertung verwendeten Vortests.	47

8 ANHANG

8.1 MODULSYSTEM AN DER LUDWIG-MAXIMILIANS-UNIVERSITÄT

An der LMU werden im ersten klinischen Semester (fünftes Semester) nur die Grundlagen der klinischen Medizin gelehrt. Ab dem sechsten Fachsemester beginnt das interdisziplinäre Basisjahr (Modul 23), in dem ein Überblick über die Innere Medizin und die Chirurgie gegeben wird. Im Anschluss folgt das Modul 4, das Modul 5 oder das Modul 6 in variierender Reihenfolge. Das Modul 6 ist ein Pflichtsemester, das für wissenschaftliches Arbeiten freigehalten wird. Ein zusätzliches Freisemester wird oft für die Doktorarbeit oder aus privaten Gründen eingelegt.

1-4	Vorklinische Semester		
1. Staatsexamen (Physikum)			
5	Modul 1 : Grundlagen der klinischen Medizin		
6	Modul 23: klinisches Basisjahr		
7	Modul 23: klinisches Basisjahr		
8	Modul 4: Sensorium	Modul 5: Alter und Gehzeiten	Modul 6: wissenschaftliches Arbeiten
	oder	oder	
9	Modul 5: Alter und Gehzeiten	Modul 4: Sensorium	
	oder		
10	Modul 6: wissenschaftliches Arbeiten	Modul 5: Alter und Gehzeiten	Modul 4: Sensorium
	oder	oder	
2. Staatsexamen (schriftlicher Teil nach neuer ÄAppO von 2012)			
11	Praktisches Jahr		
12	Praktisches Jahr		
2. Staatsexamen (mündlicher Teil nach ÄAppO von 2012 bzw. schriftlicher und mündlicher Teil nach ÄAppO von 2002)			

Anhang 1: Grobes Schema des Modulsystems an der LMU. Die klinischen Fächer sind modularisiert. Zu Beginn des klinischen Abschnitts machen alle das Modul 1 (Grundlagen der klinischen Medizin), dann das Basisjahr Modul 23 und im Anschluss kann die Reihenfolge der Module variieren. Siehe auch: https://www.mecum-online.de/de/studium/studienabschnitt_2/index.html und <https://www.impp.de/internet/de/medizin/articles/gliederung-des-medizinstudiums.html>

8.2 AUFWÄRMSZENARIO

Aufwärm szenario

Bei Ihrem 12-jährigen Neffen wurde mildes ADHS diagnostiziert. Seine Mutter ist sich unsicher, ob ein kognitives Training oder eine medikamentöse Behandlung mit einem entsprechenden Medikament besser geeignet ist und fragt Sie deswegen um Rat. Sie sollen einen guten Ratschlag erteilen, mit welcher der zwei Behandlungsmethoden bei Ihrem Neffen die besten Erfolgsaussichten bei geringstem Risiko zu erwarten sind.

Option 1: Medikamentöse Behandlung

Ich würde zu einer medikamentösen Therapie raten. Dabei wird das Krankheitsbild ohne großes Risiko zum Positiven verändert.

Option 2: kognitives Training

Ich würde zu einer Behandlung mit einem kognitiven Training raten. Dabei wird das Krankheitsbild ohne großes Risiko zum Positiven verändert.

Anhang 2: Aufwärm szenario neu entwickelt für die hier präsentierte Studie. Für das Aufwärm szenario wurden zum Thema ADHS zusätzlich eine Medizin-Studie [69] und eine Pädagogik-Studie [70] nach dem gleichen Schema wie bei den vier Hauptstudien zusammengefasst.

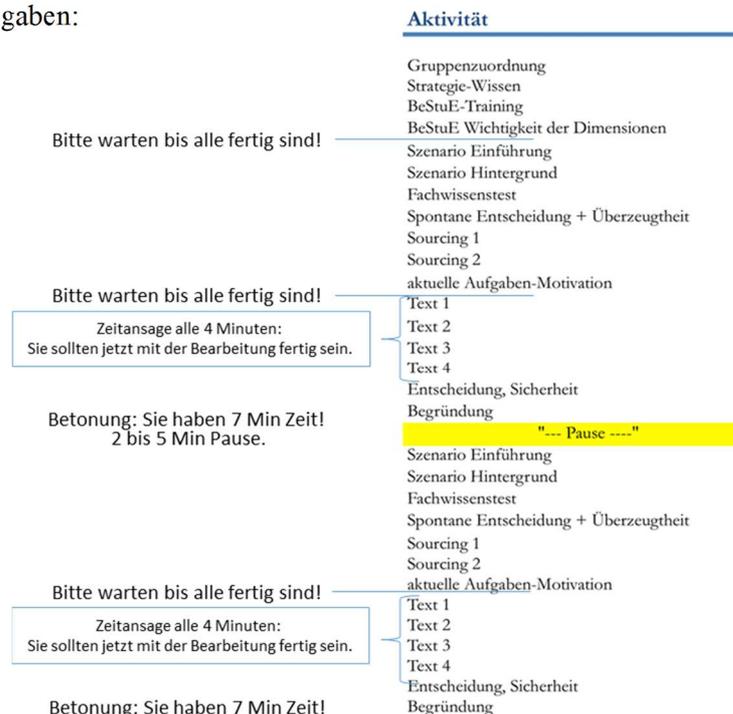
8.3 DURCHFÜHRUNGSLEITFADEN FÜR DIE TESTDURCHLÄUFE

Einführung

- Verteilen von Nummern-Karten; immer abwechselnd „1“ und „2“ [Bitte über die Sitzungen hinweg auf Ausbalancierung achten!!!]
- Teilnahme wird nur zugelassen bei vorher bearbeitetem Online-Vortest
- Auszahlung der 40 € am Ende des Testdurchlaufs bei vorher bearbeitetem Online-Vortest
- „Bitte beachten Sie, dass sie unbedingt bis zum Ende des Workshops an Ihren Plätzen bleiben und alles vollständig bearbeiten. Das ist Voraussetzung für die Auszahlung des Probandenhonorars. Es lohnt sich also nicht – Zeit-Einsparungsversuche zu unternehmen.“
- „Bitte versuchen Sie sich möglichst genau an meine **Zeitansagen** zu halten. Sollten Sie also voraus sein, dann können Sie sich ein bisschen mehr Zeit lassen, oder wenn Sie noch nicht an der entsprechenden Stelle sind, dann sollten Sie versuchen, ein bisschen schneller zu arbeiten.“

Arbeitsphase

- Verteilen: Szenario und Hintergrund-Info zweimal verteilen; jeweils nach der Bearbeitung des Wissenstests.
- Zeitangaben:



8.4 SZENARIO PÄDAGOGIK

Szenario Pädagogik

Sie sollen zusammen mit einer Kollegin ein *Präsentationstraining* zunächst planen und dann auch durchführen. Das Training richtet sich an *Schülerinnen und Schüler im Alter von ca. 15 Jahren mit ganz gemischtem Vorwissens- und Erfahrungsstand*. Die Gruppe von ca. 20 *Teilnehmenden* soll lernen, eine gute und überzeugende Präsentation mit Hilfe von PowerPoint-Folien zu halten.

Das Training ist Teil eines Förderprojektes „Business Start“, das aufwändig evaluiert wird. Daher ist gute und wissenschaftlich fundierte Vorbereitung des Trainings unbedingt notwendig, damit die Lernerfolge der Teilnehmer möglichst optimiert werden.

Sie haben von der Kollegin einen Vorschlag für die Strukturierung und Gestaltung des Trainings bekommen. Es gibt allerdings an einigen Stellen noch zwei Optionen, zwischen denen entschieden werden muss. Diese Entscheidungen wollen Sie gemeinsam mit der Kollegin treffen. Zur Vorbereitung auf dieses Treffen machen Sie sich schon einmal an die Recherche und suchen nach geeignetem Studienmaterial, auf dessen Grundlage Sie Ihre Entscheidung treffen und rechtfertigen können. Dabei legen Sie ein besonderes Augenmerk auf die Auswahl und Qualität der Untersuchungen.

Sie sind mit ihrer Planung jetzt an einer Stelle im Workshop, an der die Teilnehmenden sich mit einigen theoretischen Grundlagen guter Präsentationen beschäftigen und dabei möglichst viel zu allen Bereichen lernen und in der späteren praktischeren Phase des Workshops verfügbar haben sollen. Dabei sind fünf Themen abzudecken: 1.) Foliengestaltung, 2.) Eröffnung (inkl. Begrüßung), 3.) Wortwahl, 4.) Körpersprache und 5.) Stimmführung. Sie haben zu jedem der Themenbereiche einen kurzen informativen Einführungstext. Die Kollegin gibt ihnen zwei Möglichkeiten vor. Sie müssen sich nun entscheiden: *mit welcher der Optionen sie erwarten, dass die Teilnehmer mehr lernen und später besser darauf zugreifen können?*

<p>Option 1: Lehrervortrag Ich gebe eine Einführung in alle fünf Bereiche und teile die Texte als Hintergrundinformation zum Selbststudium aus. So stelle ich sicher, dass alle Teilnehmenden über alle Themen Bescheid wissen und das Wissen im weiteren Verlauf des Workshops anwenden können.</p>	<p>Option 2: Gruppenpuzzle Ich teile die Gruppe in Kleingruppen auf und arbeite mit der Methode „Gruppenpuzzle“. So stelle ich sicher, dass alle Teilnehmenden über alle Themen Bescheid wissen und das Wissen im weiteren Verlauf des Workshops anwenden können.</p>
---	--

8.5 HINTERGRUNDINFORMATION PÄDAGOGIK

Hintergrundinformation Pädagogik

Zur Durchführung eines „Gruppenpuzzles“ werden die Lernenden in gemischte *Stammgruppen* mit vier bis acht Mitgliedern eingeteilt. Eine *Stammgruppe* hat so viele Mitglieder, wie es Themenbereiche gibt. Jedes Mitglied einer Stammgruppe wählt eines der Themengebiete aus. Die Lernenden aus den verschiedenen Stammgruppen, die dasselbe Themengebiet gewählt haben, treffen sich in *Expertengruppen*, in denen sie den Stoff selbstständig bearbeiten und für die spätere Vermittlung in den Stammgruppen aufbereiten. Nach dieser Erarbeitungsphase kehren die Lernenden als Experten auf ihrem Teilgebiet in ihre Stammgruppen zurück, um dort ihr Wissen weiterzugeben. Die einzelnen Wissensteile der Experten werden, einem Puzzle gleich, in der Stammgruppe zu einem Ganzen zusammengesetzt. Diese Vermittlungsphase ist durch gegenseitiges Erklären und Fragen gekennzeichnet.

Insgesamt soll eine hohe Eigenaktivität beim Lernen erzeugt werden. Die Methode hat Vor- und Nachteile.

Anhang 5: Hintergrundinformationen für das Pädagogik- Szenario [1].

8.6 SZENARIO MEDIZIN

Szenario Medizin

Im Rahmen eines groß angelegten Kooperationsprojekts zwischen den Unikliniken Freiburg und Basel sollen Sie zusammen mit einem Schweizer Kollegen eine *Fortbildung für niedergelassene Kinderärzte zum Thema Windpockenimpfung* planen und durchführen. Die Fortbildung richtet sich an deutsche und Schweizer Kinderärzte aus dem Raum Freiburg und Basel.

Die teilnehmenden Kinderärzte praktizieren sowohl im ländlichen Raum als auch in Städten und betreuen pro Quartal bis zu 800 *Kinder und Jugendliche im Alter von 0 bis 16 Jahren*.

In diesem Workshop sollen die *Vor- und Nachteile der in der Schweiz und Deutschland verschiedenen Impfeempfehlungen beleuchtet* werden. Die teilnehmenden Kinderärzte sollen dabei für sich eine Impfstrategie, unabhängig von den länderspezifischen Empfehlungen der Impfkommisionen, entwickeln, die aber durch wissenschaftliche Untersuchungen gut abgesichert ist.

Diese wissenschaftliche Absicherung der Impfstrategie zur Windpockenimpfung wird von unabhängigen Experten überprüft. Damit die Teilnehmer zu guten Strategien kommen und eine gute wissenschaftliche Absicherung gewährleistet werden kann, ist eine gute Vorauswahl an Materialien für den Erfolg des Workshops unbedingt notwendig.

Sie machen sich direkt an die Recherche und suchen nach geeignetem Studienmaterial, das Sie den Teilnehmenden dann als Vorauswahl vorlegen können. Sie legen dabei ein besonderes Augenmerk auf die Auswahl und Qualität der Untersuchungen.

Nachdem Sie nun eine Zeit recherchiert haben, wird immer deutlicher, dass es zwei Optionen gibt (so ist dies auch in den unterschiedlichen Impfeempfehlungen festgehalten), zwischen denen sich Ihre Teilnehmer entscheiden werden müssen. Um wirklich gut informiert zu sein, machen Sie sich selbst auch daran eine Entscheidung zu diesem Punkt zu formulieren: halten Sie es mit der deutschen Impfeempfehlung und impfen mit 2 Jahren, oder halten Sie es mit der schweizer Impfeempfehlung und impfen mit 11 Jahren? Die entscheidende Frage ist: *Mit welcher der beiden Methoden ein größtmöglicher Schutz bei gleichzeitiger Minimierung der Risiken verbunden ist?*

Option 1: mit zwei Jahren	Option 2: mit 11 Jahren
<p>Ich würde alle Kinder mit ca. zwei Jahren impfen. Auf diese Weise kann ein größtmöglicher Schutz bei gleichzeitig minimierten Risiken gewährleistet werden.</p>	<p>Ich würde erst mit ca. elf Jahren impfen, wenn die Kinder noch keine Windpockenerkrankung durchgemacht haben bzw. nicht sicher ist, ob sie die Erkrankung schon hatten. Auf diese Weise kann ein größtmöglicher Schutz bei gleichzeitig minimierten Risiken gewährleistet werden.</p>

Anhang 6: Medizin-Szenario neu entwickelt für die hier vorgestellte Studie des KOMPARE-Projekts [51].

8.7 HINTERGRUNDINFORMATION MEDIZIN

Hintergrundinformation Medizin

Hintergrundinformation „Windpocken und Gürtelrose“:

Windpocken sind eine typische Kinderkrankheit, die durch das Varizella-Zoster-Virus ausgelöst wird. Die Erkrankung ist hoch ansteckend. Nachdem sie durchgemacht wurde, entwickelt der Körper eine lebenslange Immunität gegen das Virus.

Typische Symptome sind ein juckender Hautausschlag mit verkrustenden Bläschen und Fieber. Der Erkrankungsverlauf ist in der Regel gutartig. Es kann jedoch zu Komplikationen kommen. Diese treten umso häufiger auf, je älter der Patient bei der Erstinfektion ist.

Mögliche Komplikationen sind eine Entzündung der Lunge, der Leber oder des zentralen Nervensystems sowie zusätzliche bakterielle Infektionen des Hautausschlags. Es kann zu bleibenden Schäden kommen.

Nach der Erstinfektion, zieht sich das Virus in die zentralen Nervenzellen zurück und überlebt dort. Bei alten oder immun-geschwächten Personen kann er in Form der Gürtelrose, ohne erneute Infektion, wieder ausbrechen.

Die Gürtelrose zeigt sich als lokal begrenzter Hautausschlag. Vor allem bei Erwachsenen ist sie oft mit sehr starken Schmerzen verbunden. Nach Abheilung des Ausschlags kann über lange Zeit ein Schmerzsyndrom bestehen bleiben. Auch bei der Gürtelrose kann das zentrale Nervensystem betroffen sein.

Eine sichere und wirksame Schutzimpfung existiert.

Quelle:

http://www.rki.de/DE/Content/Infekt/EpidBull/Merkblaetter/Ratgeber_Varizellen.html;jsessionid=DB579C9EEFD9E5EC4014C5E6913F6BA1.2_cid381#doc2374554bodyText1
 nur noch bedingt an das RKI angelehnt...Zusätzliche Informationen aus Innere Medizin, Herold, 2009 und <http://de.wikipedia.org/wiki/Windpocken>

Anhang 7: Hintergrundinformation für das Medizin-Szenario[51].

8.8 BEISPIEL FÜR EINE HAUPTSTUDIEN-ZUSAMMENFASSUNG

„Langzeiteffektivität des Windpocken-Impfstoffs“; Roger Baxter, Paula Ray, Trung Tran, Steve Black & Henry Shinefield (410 Wörter)

In dieser Studie wurde untersucht, wie häufig und mit welchem Schweregrad Windpocken- und Gürtelroseerkrankungen bei geimpften Kindern in den USA auftraten. Damit sollte festgestellt werden, welche Auswirkungen die Windpockenimpfung auf die Epidemiologie der Windpocken- und Gürtelroseerkrankungen hat.

Methodik

Teilnehmer. Für die Studie wurden Daten von 7.585 Kindern im Alter von 12 bis 23 Monaten (51 % männlich und 49 % weiblich) ausgewertet. Alle Kinder wurden in den ersten fünf Monaten nach Zulassung der Windpockenimpfung geimpft. Das Patientenkollektiv entsprach der ethnischen und sozioökonomischen Zusammensetzung Nord-Kaliforniens. Die Eltern konnten frei über die Teilnahme ihrer Kinder an der Studie entscheiden.

Untersuchungsplan. Die erhobenen Daten zu Häufigkeit und Schweregrad von Windpocken- und Gürtelroseerkrankungen bei geimpften Kindern wurden mit Daten aus Studien zu Häufigkeit und Schweregrad in der Gesamtbevölkerung vor Zulassung der Windpocken-Impfung verglichen.

Ablauf der Untersuchung. Über einen Zeitraum von 14 Jahren wurden Patienten aus mehreren Einrichtungen eines großen Verbunds von Gesundheitszentren beobachtet und halbjährlich telefonisch befragt. Die Eltern wurden regelmäßig über Symptome und Erscheinungsbild der Erkrankungen informiert. Ihnen wurde außerdem eine gebührenfreie Telefonnummer mitgeteilt, bei der neue Windpocken oder Gürtelrose-Fälle gemeldet werden konnten.

Messung und Datenerfassung. *Häufigkeit der Windpockenerkrankung* und *Schweregrad der Windpockenerkrankung* wurden durch telefonische Befragung der Eltern ermittelt. Eine ärztliche Bestätigung wurde nicht benötigt, da Elternberichte in vorherigen Studien als verlässlich identifiziert werden konnten. Der *Schweregrad der Windpockenerkrankungen* wurde anhand der Ausprägung des Ausschlags definiert. Dabei galt ein Ausschlag mit weniger als 50 Pusteln als mild, ein Ausschlag mit 51 bis 300 Pusteln als mittelgradig und ein Ausschlag mit mehr als 300 Pusteln als schwergradig. Windpockenerkrankungen innerhalb der ersten sechs Wochen nach Impfung wurden nicht gewertet; dadurch wurden impfbezogene und bereits vor Impfung erworbene Windpockenerkrankungen ausgeschlossen. Das Auftreten einer *Gürtelrose* wurde durch eine ärztliche Bestätigung ohne apparativen Nachweis abgesichert. Es wurde nicht getestet, ob die *Gürtelrose* durch das geimpfte oder das natürliche Virus verursacht wurde.

Ergebnisse und Schlussfolgerungen

Der *Schweregrad der Windpockenerkrankungen*, die trotz Windpocken-Impfung auftraten, war in einem Großteil der Fälle mild (76 %), in einigen Fällen mittelschwer (24 %) und sehr selten schwergradig (2 %). Durch die Windpocken-Impfung kam es zu einer statistisch belastbaren Verminderung der *Häufigkeit der Windpockenerkrankungen* auf ein Zehntel der Häufigkeit vor Zulassung der Windpocken-Impfung. Im Vergleich zu Daten von Kindern vor Zulassung der Windpocken-Impfung, die eine natürliche Windpockeninfektion durchgemacht hatten, hatten geimpfte Kinder ein um 40% reduziertes Risiko, an *Gürtelrose* zu erkranken.

Die Autoren schließen daraus, dass der Windpocken-Impfstoff Windpocken effektiv verhindert und durch seinen Einsatz möglicherweise auch die Anzahl der Gürtelroseneuerkrankungen vermindert werden kann.

Anhang 8: Beispiel für eine Zusammenfassung einer der vier Hauptstudien. Der Test, der die Kompetenz des Bewertens von Studien misst (BESTUEN), bezieht sich auf diese Kurzzusammenfassungen [71].

8.9 SUCHE VON RELEVANTEN STUDIENTITELN

Fördern Schulnoten die Motivation? Eine quasi-experimentelle Studie zum Einfluss der Benotungserwartung auf selbst berichtete und verhaltensnah erhobene Motivationsqualitäten

Hänze, M., Berger, R. & Bianchy, K - Psychologie in Erziehung und Unterricht - 2009

- auf keinen Fall
 eher nicht
 eher schon
 auf jeden Fall

Prozesse und Effekte „Kooperativen Lernens“ im Sportunterricht

Bähr, I., Prohl, R. & Gröben, B. - Unterrichtswissenschaft - 2008

- auf keinen Fall
 eher nicht
 eher schon
 auf jeden Fall

Der Experteneffekt: Grenzen kooperativen Lernens in der Primarstufe?

Borsch, F., Gold, A., Kronenberger, J. & Souvignier, E. - Unterrichtswissenschaft - 2007

- auf keinen Fall
 eher nicht
 eher schon
 auf jeden Fall

Zur Vermittlung von systematischen Zusammenhängen im naturwissenschaftlichen Sachunterricht in der Grundschule: Ein Vergleich intruktionszentrierten und kooperativen Unterrichts gemäß Gruppenpuzzle

Doll, J., Rieck, K. & Fischer, M. - Unterrichtswissenschaft - 2007

- auf keinen Fall
 eher nicht
 eher schon
 auf jeden Fall

Anhang 9: Screenshot aus dem Original-Kompetenztest: Suche nach relevanten Studientiteln [1].

8.10 BEISPIEL EINER KURZZUSAMMENFASSUNG DIE FÜR DIE INFORMATIONSSUCHE

Zusammenfassung:

Hintergrund: Narben im Gesichtsbereich nach einer Windpockenerkrankung sind, wie Narben anderer Ursache, schwer zu behandeln. In der aktuellen Fachliteratur gibt es keine ausreichenden Empfehlungen für effektive Behandlungsmethoden für Windpockennarben. In dieser Studie wurde hoch konzentrierte Trichloressigsäure (TCA), die nachgewiesener Weise zum Umbau von Hautkollagenen führt, zur Behandlung von Windpockennarben eingesetzt. Ziel der Studie war es, die Effektivität der chemischen Wiederaufbau-Therapie von Windpockennarben festzustellen. Methoden: Insgesamt 26 Patienten mit Windpockennarben im Gesichtsbereich wurden mit lokaler Anwendung von 100%-iger TCA-Lösung behandelt, indem man sie mittels eines Zahnstochers auf die vernarbte Oberfläche drückte (CROSS-Technik). Insgesamt vier Sitzungen wurden in zwei wöchentlichen Intervallen durchgeführt und die Ergebnisse wurden drei Monate danach im Verlauf erhoben. Ergebnisse: Alle 13 Patienten, die die Studie abschlossen, zeigten deutliche klinische Verbesserungen, wobei 69% der Patienten das Ansprechen als außerordentliche Verbesserung bezeichneten und die restlichen 31% der Patienten über gute Verbesserungen berichteten. Bei keinem der Patienten traten bedeutende Komplikationen auf. Schlussfolgerungen: Die CROSS-Technik unter Verwendung von 100%-iger TCA ist eine sichere, günstige und effektive Therapie für die Behandlung von Windpockennarben.

Auf Grundlage dieser Informationen gehe ich davon aus, dass die Untersuchung eine wichtige Rolle für die Entscheidung spielt.

Dieser Aussage ...



Anhang 10: Beispiel einer Kurzzusammenfassung die für den Test zur Kompetenz der Informationssuche verwendet wurde. Sowohl für das Medizin- als auch für das Pädagogik-Szenario gab es zehn solcher Kurzzusammenfassungen, die Bewertet werden sollten. Hier eine Studie von Agarwal (2013). Darunter ist ein Screenshot aus dem Originalkompetenztest mit den Antwortmöglichkeiten zu sehen[51, 72].

8.11 LEITFADEN FÜR DIE SUCHE DER ZEHN STUDIEN FÜR DIE INFORMATIONSSUCHE

Leitfaden für die Suche von geeigneten Varizellen-Studien

Auf Grundlage der vier (hoch relevanten) Hauptstudien werden je zwei mittel relevante, zwei wenig relevante und zwei irrelevante Studie ausgesucht.

Datenbank: Pubmed

Mittel relevante Studien:

1. Eingabe der Keywords der Hauptstudien (meistens Angabe im Text) in Englisch in die „Advanced“- Suche unter Angabe: „all fields“ und „and“. Es werden keine „additional filters“ eingestellt. Unter „Display settings“ wird „relevance“ angeklickt.
2. Die entsprechende Hauptstudie muss in der Liste enthalten sein.
3. Die erste Studie der Liste wird ausgewählt, wenn sie in irgendeiner Weise relevant ist, nicht eine der Hauptstudien ist und auf die folgendes **nicht** zu trifft:
 - Reviews, Metaanalysen
 - Fallstudien, Expertenmeinungen
 - Kostenanalysen
 - Studien die gleich relevant sind, wie die vier Hauptstudien
3. Von den maximal vier möglichen Studien werden zwei ausgewählt, deren Abstract gut strukturiert ist.

Wenig relevante Studien:

1. Es werden die **zwei** Keywords in Englisch genommen, die in allen / den meisten der Hauptstudien vorkommen.
2. Eingabe dieser Keywords in die „Advanced“- Suche unter Angabe: „all fields“ und „and“. Es werden keine „additional filters“ eingestellt. Unter „Display settings“ wird „recently added“ angeklickt.
3. Die ersten beiden Studie der Liste werden ausgewählt, die nicht eine der Hauptstudien sind und auf die folgendes **nicht** zu trifft:
 - Reviews, Metaanalysen
 - Fallstudien, Expertenmeinungen
 - Kostenanalysen
 - Studien die gleich relevant sind wie die vier Hauptstudien oder die zwei mittelrelevanten Studien
 - Studien die absolut irrelevant sind

Nicht relevante Studien:

1. Es wird das Keyword in Englisch genommen, das in allen / den meisten der Hauptstudien vorkommt.
2. Eingabe dieses Keywords in die „Advanced“- Suche unter Angabe: „all fields“. Es werden keine „additional filters“ eingestellt. Unter „Display settings“ wird „recently added“ angeklickt.

3. Es werden die ersten zwei Artikel ausgewählt, auf die folgendes **nicht** zutrifft:

- Reviews, Metaanalysen
- Fallstudien, Expertenmeinungen
- Kostenanalysen

Anhang 11: Leitfaden für die Suche der zehn Studien, die im Kompetenztest für die Informationssuche verwendet wurden [51].). Es wurden systematische Übersichtsarbeiten (Reviews und Metaanalysen), Fallstudien, Expertenmeinungen und Kostenanalysen ausgeschlossen. Für die mittelrelevanten Studien wurden alle Keywords der Hauptstudien verwendet, die Ergebnisliste nach Relevanz sortiert und die in der Liste am höchsten stehenden Studien verwendet. Für die wenig relevanten Studien wurden nur noch mit den zwei am häufigsten in den Hauptstudien verwendeten Keywords gesucht und für die irrelevanten Studien mit einem Keyword. Die dabei entstandene Rangliste wurde bei der Auswertung als Lösung verwendet.

8.12 BEISPIEL ZUR GEWINNUNG DER MUSTERLÖSUNG FÜR DIE BESTUEN

<ul style="list-style-type: none"> - Kein Vergleich der Impfung zu den beiden Zeitpunkten, geht nicht um Auswirkungen des Impfzeitpunkts - Keine Untersuchung bei 11jährigen 	+	dass die Fragstellung zu meiner Entscheidung passt. 8	+ + +	<ul style="list-style-type: none"> - Effektivität und Auswirkung der Impfung wird untersucht - Untersucht die Impfung bei Kleinkindern - Auswirkungen ins Jugendalter werden auch untersucht
<ul style="list-style-type: none"> - aber nur bei 2 Jährigen 		dass die Maßnahme untersucht wurde, um die es in Ihrer Entscheidung geht. 8	+ +	<ul style="list-style-type: none"> - Maßnahme ist Impfung zw. 12 und 23 Monaten
<ul style="list-style-type: none"> - nicht so intensive Untersuchung der Risiken 		dass die Erfolgskriterien mit den Erfolgskriterien bei Ihrer Entscheidung übereinstimmen. 6	+ +	<ul style="list-style-type: none"> - Erfolgskriterium guter Schutz stimmt überein - Geht nur minimal auf Risikominimierung ein
<ul style="list-style-type: none"> - Pusteln nicht aussagekräftig für Schweregrad - Elternbefragung statt ärztlicher Untersuchung kritisch zu bewerten - Abzüge für Telefonumfrage 	+ + +	dass die Erfolgskriterien mit geeigneten Verfahren erfasst wurden. 8	+ + + + +	<ul style="list-style-type: none"> - ausreichende Zeit und großer Stichprobenumfang - Telefonumfrage mit immer denselben Eltern - regelmäßige Schulungen - Gürtelrose ärztlich bestätigt - Methoden werden validiert durch vorherige Studien
<ul style="list-style-type: none"> - Untersucht nicht die 11 jährigen - Nordkaliforniens Klima und ethnische Zusammensetzung entspricht nicht zwingend süddeutschen/schweizerischen Verhältnissen. 	+ + +	dass die Merkmale der Personen mit denen der Zielgruppe bei Ihrer Entscheidung übereinstimmen. 4	+ +	<ul style="list-style-type: none"> - 12 bis 23 Monate, damit die 2 jährigen abgedeckt
<ul style="list-style-type: none"> - Signifikanz steht bei Gürtelroseinzidenz nicht explizit drin. - Signifikanz bei Schweregrad der Erkrankung nicht angegeben - Sehr große Stichprobe, d.h. Signifikanz evtl. nicht sehr aussagekräftig 	+ +	dass die entscheidungsreleva nten Ergebnisse statistisch belastbar sind. 7	+ + +	<ul style="list-style-type: none"> - Große Stichprobe, - langer Untersuchungszeitraum - Bei einigen sehr relevanten Ergebnissen wird Signifikanz erwähnt
<ul style="list-style-type: none"> - Gürtelroseverminderung könnte auch natürliche Schwankung sein. 	+	dass die entscheidungsreleva nten Ergebnisse eindeutig auf die Maßnahme zurückzuführen	+ +	<ul style="list-style-type: none"> - Inzidenzverminderung in diesem Ausmaß kann nur aufgrund der Impfung sein, zu viel für natürliche Schwankung

		sind. Maßnahme = Impfung 8		
- Keine Aussage zum optimalen Impfzeitpunkt	+	Unter Berücksichtigung der zuvor bewerteten Aspekte spielt die Studie für Ihre Entscheidung eine wichtige Rolle. 8	+ + + +	- starke Studie, große Stichprobe, langer Untersuchungszeitraum - Gibt als einzige Studie Infos über die Impfung: z.B. Effektivität, Hohe(!) Durchbruchrate bei nur einer Impfdosis

Anhang 12: Beispiel einer Tabelle zur Gewinnung der Musterlösung für die BESTUEN. In der Mitte sind die Fragestellungen der Items aufgeführt und die Musterlösung. Auf der linken Seite sind Argumente gesammelt und gewichtet, die gegen diese Aussage sprechen. Auf der rechten Seite sind Argumente gesammelt und gewichtet, die für diese Aussage sprechen. Bei der Entstehung der Musterlösung wurde zusätzlich berücksichtigt, dass die vier Hauptstudien im Vergleich einheitlich bewertet wurden [51].

9 EIDESSTATTLICHE VERSICHERUNG

Melle, Joana Veronika

Ich erkläre hiermit an Eides statt,
dass ich die vorliegende Dissertation mit dem Thema

Evidenzbasierte Medizin im Medizinstudium: Bewerten von Studien – eine generische Teilkompetenz mit Beeinflussung durch fachspezifisches Wissen

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München, 11.4.2018

Joana Melle