

Aus der Klinik für Pädiatrie mit Schwerpunkt Endokrinologie und  
Diabetologie  
des Otto Heubner Centrums für Kinder- und Jugendmedizin  
der Medizinischen Fakultät Charité – Universitätsmedizin Berlin

DISSERTATION

Skript Konkordanz Test: Eine neue Methode zur Messung der  
Entwicklung der Kompetenz des klinischen Denkens in der  
Kinderheilkunde bei Studierenden der Humanmedizin

zur Erlangung des akademischen Grades  
Doctor medicinae (Dr. med.)

vorgelegt der Medizinischen Fakultät  
Charité – Universitätsmedizin Berlin

von

Kai Sostmann  
aus Hamm/Westfalen

Datum der Promotion  
10. März 2017

---

# Gliederung

|  |    |
|--|----|
| Gliederung.....  | 2  |
| Abstrakt.....  | 6  |
| Einleitung.....  | 6  |
| Methodik.....  | 6  |
| Ergebnisse.....  | 6  |
| Schlußfolgerung.....   | 7  |
| Abstract - English.....  | 7  |
| Introduction.....  | 7  |
| Methods.....   | 7  |
| Results.....   | 8  |
| Conclusion.....  | 8  |
| Abkürzungsverzeichnis.....                                     | 9  |
| Glossar.....   | 9  |
| Einleitung.....  | 11 |
| Problemstellung.....   | 12 |
| Zielsetzung der Arbeit.....                                    | 13 |
| Stand der internationalen Forschung.....                       | 14 |
| Definition des Kompetenzbegriffes in der Humanmedizin.....     | 15 |
| Professionelle ärztliche Kompetenzen.....                      | 16 |
| Methoden zur Beschreibung klinischen Denkens.....              | 17 |
| Informationen im medizinischen Alltag.....                     | 17 |
| Hypothetikededuktive Methode - Prozessorientiertes Wissen..... | 18 |
| Strukturtheorien – Wissensverkapselung.....                    | 19 |
| Skripttheorie.....   | 20 |
| Mustererkennung.....   | 21 |
| Bedeutung von medizinischem Grundlagenwissen.....              | 22 |
| Bedeutung von klinischem Wissen.....                           | 23 |
| Semantische Verknüpfung von Wissen.....                        | 24 |
| Methoden zur Messung des klinischen Wissenserwerbs.....        | 25 |
| Multiple-Choice-Fragen (MCQ).....                              | 26 |

---

|  |    |
|--|----|
| Extended-matching questions (EMQ) .....                            | 27 |
| Objective structured clinical examination (OSCE) .....             | 27 |
| Key-Feature-Fragen (KF).....                                       | 27 |
| Skript-Konkordanz-Test.....  | 28 |
| Beispiele für den SKT an der Charité in Berlin .....               | 30 |
| Methodik .....   | 31 |
| Material und Methoden .....  | 31 |
| Methodisches Vorgehen.....   | 31 |
| Forschungsleitende Thesen .....                                    | 31 |
| Fragestellungen .....  | 32 |
| Studiendesign .....  | 33 |
| Beschreibung der Auswahl der Stichprobe .....                      | 33 |
| Studierende.....   | 33 |
| Experten_innen.....  | 33 |
| Multiple Choice-Klausur-Durchführung.....                          | 34 |
| Skript Konkordanz Test-Durchführung .....                          | 34 |
| Blueprint Entwicklung Kinder- und Jugendmedizin.....               | 35 |
| Lernzielerstellung für den Unterricht am Krankenbett (UaK) .....   | 35 |
| Lernzielerstellung Blockpraktikum.....                             | 35 |
| SKT-Item-Konstruktion .....  | 35 |
| Fragenerstellung MC-Klausur.....                                   | 36 |
| Fragenerstellung SKT .....   | 36 |
| Item-Anzahl - Reliabilität.....                                    | 38 |
| Referenzpanel.....   | 39 |
| Antwort-Auswertung.....  | 40 |
| Durchführung der Prüfung .....                                     | 40 |
| Methodik der Testanalyse .....                                     | 41 |
| Definition der Antwortschlüssel.....                               | 41 |
| Berechnung der Probengröße .....                                   | 41 |
| Statistische Analyse - Methoden .....                              | 42 |
| Ergebnisse .....   | 43 |
| Ergebnisse des Skript-Konkordanz-Tests Analyse – Studierende ..... | 43 |
| Demographische Daten und Ergebnisse.....                           | 43 |

---

|  |    |
|--|----|
| Reliabilität der SKT- Gesamtergebnisse der Studierenden.....         | 44 |
| Reliabilitätsanalyse nach Wochen: Ergebnisse des SKT .....           | 44 |
| Item-Reliabilitätsanalyse des SKT .....                              | 45 |
| Operationalisierung über die Trennschärfe .....                      | 46 |
| Ergebnisse der MC-Klausur .....                                      | 47 |
| Reliabilität der MC-Klausur.....                                     | 48 |
| Vergleich der MC-Klausurergebnisse mit dem SK-Test .....             | 49 |
| Vergleich der Ergebnisse getrennt nach Geschlecht .....              | 50 |
| Einfluss der Dozierenden auf das Gruppenergebnis .....               | 50 |
| Ausbildungsstatus und Geschlechterverteilung .....                   | 50 |
| Ergebnisse des Skript Konkordanz Testes – Dozierenden.....           | 50 |
| Berufserfahrung Dozierenden in Bezug auf die Gruppenergebnisse ..... | 51 |
| Zusammenfassung.....   | 53 |
| Diskussion .....   | 54 |
| Validitätskriterien der MC-Klausur .....                             | 55 |
| Validitätskriterien des Skript-Konkordanz-Tests.....                 | 56 |
| Validität.....   | 56 |
| Inhaltsvalidität.....  | 56 |
| Konstruktvalidität.....  | 57 |
| Kriteriumsvalidität – Trennschärfe der Items .....                   | 58 |
| Kriteriumsvalidität – Antworten der Dozierenden.....                 | 58 |
| Durchführungsqualität.....   | 59 |
| Beziehungen der Variablen .....                                      | 60 |
| Operationalisierung der Trennschärfe .....                           | 61 |
| Zusammenfassung.....   | 61 |
| Limitationen .....   | 63 |
| Schlussfolgerungen .....   | 63 |
| Ausblick .....   | 64 |
| Schriftenverzeichnis .....   | 66 |
| Abbildungen .....  | 73 |
| Tabellen.....  | 77 |
| Lernziele/Leitfäden .....  | 81 |
| Publikationen Kai Sostmann 1999 - 2015.....                          | 86 |

---

|                    |    |
|--------------------|----|
| 1999.....          | 86 |
| 2008.....          | 86 |
| 2010.....          | 86 |
| 2013.....          | 86 |
| 2014.....          | 87 |
| Buch-Beiträge..... | 87 |
| Danksagung.....    | 88 |
| Lebenslauf.....    | 89 |

---

## **Abstrakt**

### **Einleitung**

Die Prüfungen im universitären Umfeld bestehen in der studentischen Ausbildung zum großen Teil aus Multiple-Choice-Fragen. Dieses Format überprüft medizinisches Faktenwissen. Medizinische Handlungspraxis basiert neben dem theoretischen medizinischen Fachwissen auf unterschiedlichen Kompetenzen. Eine davon ist das klinisch-differentialdiagnostische Denken. Das Ziel dieser Studie ist die Überprüfung der Eignung des Skript Konkordanz Tests (SKT) als neues Prüfungsinstrument für die Messung der Kompetenz des klinisch-differentialdiagnostischen Denkens, im Fach der Kinderheilkunde bei Studierenden der Humanmedizin. Es soll überprüft werden inwiefern eine Beziehung zwischen dem Grad der Ausprägung klinischen Experten\_innen-Wissens der Dozierenden und dem Abschneiden der Studierendengruppen dieser Dozierenden im SKT besteht.

### **Methodik**

In einem Workshop wurde ein Fragenkatalog mit 150 Fragen aus den elf Fachbereichen des OHCs für Studierende des Blockpraktikums im sechsten klinischen Semesters (Regelstudiengang) entwickelt. Im Wintersemester 2008/2009 wurde eine prospektive Kohortenstudie durchgeführt. Alle 50 Dozierenden haben die Fragen freiwillig beantwortet. Die Ergebnisse des SKT wurden mit den Ergebnissen der MC-Klausur (Sommersemester 2008) verglichen. Die Ergebnisse der Dozierenden wurden korreliert mit den Ergebnissen der von Ihnen betreuten Gruppen.

### **Ergebnisse**

287 Studierende nahmen an der semesterübergreifenden auf Multiple Choice Fragen basierenden Abschlussklausur des fünften klinischen Semesters im Sommersemester 2008 teil. Von diesen nahmen 232 am Skript-Konkordanz Test des folgenden sechsten klinischen Semesters teil (Wintersemester 2008/2009). 64% von ihnen waren weiblich, 36% männlich. Alle 51 Dozenten, die im Blockpraktikum des Wintersemesters 2008/2009 unterrichteten nahmen freiwillig teil. Die SK-Testergebnisse der Studierenden lagen zwischen 49 und 95 Prozent der erreichbaren Punkte, mit einem Mittelwert von 80,4 Prozent. In der MC- Klausur erreichten die Studierenden Werte zwischen 12 und 38 von 39 erreichbaren (Mittelwert 30 Punkte) Punkten.

---

Die Dozierenden erreichten Ergebnisse zwischen 63 und 96 Prozent im SKT (Mittelwert 82,5 Prozent). Der SKT erreichte bei den Studierenden einen optimierten Reliabilitätskoeffizienten, ausgedrückt durch den Faktor Cronbach Alpha von 0,72. (MC-Klausur 0,76). Die Korrelationskoeffizienten ( $r=0,11$   $r=0,04$ ) der MC-Klausur und dem SKT waren statistisch nicht signifikant. Der Vergleich der Mittelwerte der Dozierenden und Studierenden ergab einen statistisch signifikanten Unterschied, insbesondere unter Berücksichtigung des Einflusses der Dozierenden auf das Ergebnis ihrer Gruppe ( $r=0,55$ ,  $p<0,001$ ).

### **Schlußfolgerung**

Der SK-Test erfüllt durch seine hohen Reliabilitäts- und Validitätswerte die Anforderungen, die an eine notengebenden Prüfung gerichtet werden. Der SKT unterscheidet zudem signifikant zwischen Studierenden und Dozierenden. Die schwache Korrelation zwischen den Ergebnissen der MC-Klausur und dem SKT ist ein Hinweis auf die unterschiedlichen Wissensdomänen sein, die von den beiden Formaten gemessen werden.

## **Abstract - English**

### **Introduction**

Examinations in undergraduate medical education consist to a large extent of Multiple Choice Questions. Clinical reasoning represents one core competency addressed by undergraduate medical education. Aim of this study is to approve the appropriateness of a tool to assess clinical reasoning competence, the Script Concordance Test (SCT). It aims to find validity for the test as a tool to assess clinical reasoning competency by a large group of medical students related to the grade of competence of clinical experts from the same department.

### **Methods**

A workshop was held to define a pool of 150 SCT-questions was developed, based on the learning objectives in eleven pediatric subspecialties. A prospective cohort study was conducted with all students of the sixth clinical semester during winter semester 2008/2009. Scores of the SCT were compared to student's MCQ-scores. Scores of the teachers' group were compared to those of their students to analyze their influence on the development of clinical reasoning competence during the course.

---

## Results

287 students took part in the obligatory MC-assessment in summer semester 2008 of whom 232 took part in the SCT in winter semester 2008/2009 at the end of the “Blockpraktikum Pädiatrie” course (64 percent female, 36 percent male). All 51 teachers of the course took part in SCT test. The SCT results of all students were correlated to their MC-test results (summer semester 2008) and to the results of their groups’ teachers results. The scores on the SCT of the students ranged from 49 to 95 percent (mean 80,4 percent). Teachers scores ranged from 63 to 96 percent (mean 82,5 percent). The optimized reliability of the SCT scores, using Cronbach’s alpha, was 0.72, of the MC test 0,76. The student’s correlations between the scores of the SC-Test and the MC-Test results were with  $r= 0,11$  and  $r=0,04$  statistically non-significant. Correlation between students’ and teachers’ mean scores was significant ( $r=0,55$ ,  $p<0,001$ )

## Conclusion

The SK-Test results could significantly differentiate between students and residents. The SC-Test could not show an increase in performance between residents and senior house officers. The poor correlation between the performance on the SCT and the performance on the MC-Test are due to the fact that the test evaluated a different domain. SCT has proven its ability to fulfill the needs of a high stakes examination tool and to assess students’ clinical reasoning capabilities as summative test tool.



---

## Abkürzungsverzeichnis

|       |  |
|-------|--|
| ÄAppO | Approbationsordnung für Ärzte                        |
| ACGME | Accreditation Council for Graduate Medical Education |
| BP    | Blockpraktikum                                       |
| MCQ   | Multiple Choice Questions                            |
| OHC   | Otto-Heubner-Centrum für Kinder- und Jugendmedizin   |
| OSCE  | Objective Structured Clinical Examination            |
| SKT   | Skript Konkordanz Test                               |
| UaK   | Unterricht am Krankenbett                            |

## Glossar

**Aggregat-Scoring-Methode:** Methode um die Variabilität von mehreren Experten\_innenantworten durch eine Likertskala abzubilden ohne einen Konsensus herbeizuführen.

**Cronbach's-Alpha:** Misst die interne Testkonsistenz, daher die Reliabilität mit der die verwendeten Testitems die festgelegten Kriterien messen.

**Cueing:** Rein optisches Erkennen der richtigen Antwort unter eine Reihe von Falschantworten in Multiple-Choice-Fragen auftretend.

**Deckeneffekt ( Ceiling-Effect ):** entweder stoßen im Rahmen vorgegebener Zeitgrenzen leistungsstarke Versuchspersonen bei der Bearbeitung von Testaufgaben mit ihrer Testleistung an eine Grenze ("Decke") oder der Maximalwert wird geringer Schwierigkeit der Aufgaben von relativ vielen Versuchspersonen erreicht wird.

**High-stakes Examen:** Prüfung auf notengebenden Niveau, daher starke Testgütekriterien, staatsexamensäquivalent.

**Item:** Grundlegender Testbestandteil, in diesem Fall einer Frage und der zu ihr gehörigen Antwort.

**Kompetenz:** Anhand definierter Kriterien messbare, theoretische und praktische Fertigkeiten und Kenntnisse.

**Likertskala:** Aus 5,7 oder 11 Merkmalsausprägungen bestehende Skala, mittels derer TestTeilnehmenden Aussagen in einem Fragenitem zustimmen oder ablehnen können.

---

**Kritisches klinisches Denken:** Denkprozess, dem mehrere Methoden zugrunde liegen können, um Patienten\_innen -Informationen zu erfassen, analysieren und bewerten, um klinische Diagnosen abzuleiten.

**Murrhardter Kreis:** Ein „Arbeitskreis Mediziner Ausbildung“ bestehend aus medizinischen Hochschullehrern, initiiert von der Robert Bosch Stiftung (1982-1995).

**Outcome:** definierte theoretisch oder praktisch prüfbare Ebene einer oder mehrerer Kompetenzen, die am Ende eines Ausbildungsabschnittes erreicht werden soll.

**Unterricht am Krankenbett (UaK):** Durch die ÄAppO vorgeschriebenes praktisches Unterrichtsformat. Mit Patienten\_innen 3 Studierende pro Patient.

**Reliabilität:** repräsentiert die Verallgemeinerbarkeit und Übertragbarkeit der Testergebnisse

**Validität:** gibt den Umfang an, in dem der Test den Grad der Kompetenz misst, zu deren Messung er eingesetzt wird.

**Volition:** „...stellt die zweite Phase des Motivationsprozesses dar, wenn aus dem Wunsch der motivationalen Phase der entschlossene Wille geworden ist, der zu einer zielgerichteten Realisierung der Absichten beiträgt.“ (Heckhausen, H. (1977). Motivation: Kognitionspsychologische Aufspaltung eines summarischen Konstrukts. Psychologische Rundschau, 28, 175-189.)

---

## Einleitung

Die universitäre medizinische Ausbildung stellt die Grundlage für die Ausübung des ärztlichen Berufes dar. Auszubildenden Medizinerinnen und Mediziner wird im Laufe ihres Studiums ein umfangreiches kognitiv-theoretisches und praktisches Wissensrepertoire angeboten, dessen Erwerb sie zur Ausübung ärztlicher Tätigkeiten befähigen soll. Traditionelle humanmedizinische Curricula beinhalten in dem ersten, vorklinischen Teil des Studiums vor allem die Fächer, die die theoretischen und wissenschaftlichen Grundlagen für den zweiten, klinischer orientierten Studienabschnitt vermitteln sollen. Klinische Bezüge werden durch interdisziplinäre Vorlesungen und Seminarveranstaltungen aufgebaut. Diese Curricula stützen sich im großen Umfang auf Großgruppenformate vor allem aber auf Vorlesungen als Hauptunterrichtsformat zur Vermittlung theoretischen und systematischen Wissens [1]. Der klinische Abschnitt des Studiums der Humanmedizin dient der stufenweisen Vermittlung klinischen Wissens in fachorientierten Kursen: der Lehre der inneren Erkrankungen muss der Unterricht der Pathologie der internistischen Erkrankungen vorausgegangen sein. Die kontroverse Diskussion um die Vor- und Nachteile dieses klassischen Ausbildungsweges und die internationale Etablierung problemorientierter Curricula mündeten in die Forderungen des Murrhardter Kreises und der Hochschulrektorenkonferenz nach einer problemorientierten und stärker Patienten\_innen orientierten Umstrukturierung der medizinischen Curricula, in denen stärker praxisrelevantes Wissen vermittelt wird [2]. Diese Forderungen führten zur Einführung der Modellstudiengangsklausel 1999 und hatten inhaltlich starken Einfluss auf die Reform der ärztlichen Approbationsordnung 2002. Diese verlangt von medizinischen Curricula die Vermittlung theoretischer und praktischer, differentialdiagnostischer und therapeutischer Fähigkeiten: *„Die Ausbildung zum Arzt wird auf wissenschaftlicher Grundlage und praxis- und Patienten\_innen bezogen durchgeführt. Sie soll die für das ärztliche Handeln erforderlichen allgemeinen Kenntnisse, Fähigkeiten und Fertigkeiten in Diagnostik, Therapie, ... und praktische Erfahrungen im Umgang mit Patienten\_innen, einschließlich der fächerübergreifenden Betrachtungsweise von Krankheiten und der Fähigkeit, die Behandlung zu koordinieren.“* [3]. Im Rahmen dieser Forderungen wurden praktische und theoretische Ausbildungsziele eingeführt, die neue Unterrichtsformate erforderten, die Patienten\_innen stärker in den Mittelpunkt der praktischen Ausbildung setzen sollten. Es wurden Formate, wie der Unterricht am Krankenbett (UaK) von der ÄAppO gefordert und deswegen in den neuen Curricula mit einem großen

---

Stundenumfang eingeführt. Die Lernziele des UaKs beinhalten die Vermittlung praktischer ärztlicher Fertigkeiten und differentialdiagnostischen Denkens im Kleingruppenunterricht mit maximal drei Studierenden am Krankenbett (sechs in der Vorstellung klinischer Fälle) [3]. Die Durchführung eines praxisorientierten Unterrichts benötigt die Einführung von Bewertungskriterien, für die erworbenen praktischen Kompetenzen bei den Studierenden, die über das reine Abfragen von theoretischem Wissen hinausgehen [4].

Eine Möglichkeit der Definition von einheitlichen Bewertungskriterien für erfolgreiches ärztliches Handeln, liegt in der Definition der für das Erreichen dieses Zieles notwendigen praktischen und theoretischen ärztlichen Kompetenzen. Diese Ausbildungsziele (Outcomes) legen in verschiedenen Dimensionen die Tiefe der zu erreichenden Kompetenzniveaus fest. Eine dieser Kompetenzen ist die der medizinischen Entscheidungsfindung. Kritisches klinisches Denken (Clinical Reasoning) stellt eine Kernkomponente dieser Kompetenz dar [5]. Den Theorien der Mustererkennung, die dieser Arbeit als theoretische Grundlage dienen werden, liegt die Idee zugrunde, dass klinische Experten\_innen den ihnen bekannten Erkrankungen mit Hilfe ihres kritischen klinischen Denkens typische Symptome und klinische Merkmale zuordnen [6]. Diese Zusammenstellung ergibt ein bestimmtes Muster, das in den Wissensstrukturen abgelegt wird und zu Vergleichszwecken im klinischen Denken heran gezogen wird [7, 8]. Auf dieser Annahme baut die Skripttheorie auf, die die Grundlage für die Konzeption des Skript- Konkordantestes darstellt, der der Untersuchungsgegenstand dieser Arbeit sein wird.

## **Problemstellung**

Mit der Einführung des Kompetenzbegriffes in den reformierten medizinischen Curricula im angloamerikanischen Raum [9], vollzog sich ein Wandel der Ausbildungssystematik. Im deutschen Sprachraum folgte 1999 der Einführung der Modellstudiengangsklausel die Gründung des Reformstudiengangs Medizin an der Charité als erstem Curriculum mit einer Kompetenzorientierung der Ausbildungsziele [10].

In der Entwicklung eines kompetenzbasierten Curriculums wird der Schwerpunkt auf den Kompetenzerwerb der Studierenden gelegt. Für diese werden Outcomes, also zu erreichende Basiskompetenzen, definiert, die in unterschiedliche Niveaustufen gegliedert sind. Diese sollte jeder Studierende im Laufe seines Studiums erworben haben, um als Arzt oder Ärztin tätig zu sein. Ausgehend von diesen Outcomes werden Lernziele für die einzelnen Unterrichtsveranstaltungen festgelegt, die zur Erreichung der jeweiligen Kompetenzniveaus

---

notwendig sind. Für die Lernziele werden die notwendigen Lerninhalte durch die Fachvertreter festgelegt [11].

Eine dieser Kompetenzen medizinisch tätiger Ärzte\_innen ist das differentialdiagnostische Denken, das klinischen Entscheidungen zugrunde liegt. Dieser Begriff (engl: Clinical Reasoning) wird in dieser Arbeit eingesetzt, um die professionelle Kompetenz des kritischen differentialdiagnostischen Denkens in alle medizinischen Bereichen für das ärztliche Personal zu beschreiben, sowohl im ambulanten als auch im stationären Bereich. Für die Messung dieser Kompetenz im Bereich der ärztlichen Fort- und Weiterbildung wurde von den Kanadiern R. Gagnon, B. Charlin und B. Carrière eine neue Testmethode entwickelt und validiert, der Script Concordance Test (deutsch: Skript Konkordanz Test, SKT) [12]. Im studentischen Unterricht des Regel- und Reformstudiengangs der Charité wurden bis zu diesem Zeitpunkt Multiple-Choice-Prüfungen im großen Umfang zur Überprüfung theoretischen Wissens eingesetzt. OSCEs wurden als summatives Prüfungsformat zur Überprüfung anwendungsorientierten Handlungswissens verwendet. Ein Format, das speziell die Kompetenz des differentialdiagnostischen klinischen Denkens überprüft war bisher jedoch nicht im Einsatz. Aus diesem Grund sollte die Eignung des Skript Konkordanz Tests untersucht werden, den Erwerb dieser Kompetenz in der studentischen Ausbildung zu messen.

## **Zielsetzung der Arbeit**

Zielsetzung dieser Arbeit ist es, die Verwendungsmöglichkeiten einer bisher im Bereich der ärztlichen Fort- und Weiterbildung im Einsatz befindlichen Testmethode, für den studentischen, klinischen Unterricht der Kinderheilkunde zu validieren. Der Skript Konkordanz Test wurde als Prüfmethode entwickelt, um die Kompetenz des klinischen Denkens auf unterschiedlichen Experten\_innen-Niveaustufen bei Ärzten\_innen zu messen [12-16]. Es werden die kognitions-psychologischen Theorien dargestellt, die die Grundlage für die Entwicklung der Methodik darstellen, mit der der Skript Konkordanz Test (SKT) klinisches Denken bei den Studierenden und Kliniker\_innen auf dem Gebiet der Kinderheilkunde überprüft. Es wird versucht die Fragestellung zu beantworten, inwiefern ein Einfluss des theoretischen Wissens auf das klinische Denken nachgewiesen werden kann und wie sich der Einfluss der Dozierenden im Unterricht auf die Testergebnisse ihrer Studierenden auswirkt.

---

Die Erläuterung der Definition des Kompetenzbegriffes als Grundlage der Kompetenzdiagnostik stellt den Ausgangspunkt der Überlegungen dar. Ausgehend von der Darstellung des Begriffes der professionellen Kompetenz in der Medizin, werden die Theorien, die die Entwicklung der Kompetenz des klinischen Denkens bei den Experten\_innen beschreiben, dargestellt. Es werden die zugrundeliegenden Theorien der Kognitionspsychologie erläutert, die die mentalen Prozesse beschreiben, die bei der Entwicklung klinischer Denk- und Entscheidungsprozesse bei Anfängern\_innen und Experten\_innen zugrunde gelegt werden. In Bezug auf diese Theorien werden die unterschiedlichen Prüfungsmethoden und die von ihnen überprüften Kompetenzen aufgezeigt. Der Fokus der Theorien liegt auf der Skripttheorie und dem Skript Konkordanz Test [17, 18].

## **Stand der internationalen Forschung**

Der Schwerpunkt der vorliegenden Untersuchung wird auf der Messung der Kompetenz des klinischen, differentialdiagnostischen Denkens konzentriert sein. Dem Prozess, der der Entwicklung dieser Kompetenz zugrunde liegt, wird eine integrierende Rolle bei der Nutzung und Vernetzung, der in den anderen Kompetenzen enthaltenen Fertigkeiten und Fähigkeiten zugeschrieben. Daher wird dieser Darstellung eine Herleitung des Kompetenzbegriffs vorangestellt [19-20].

Die Kompetenz des klinischen Denkens beruht auf der Interpretation und Integration der Informationen, die mit den Fertigkeiten der Anamneseerhebung, körperlicher und klinischer Untersuchungsmethoden erhoben werden. Mit Hilfe der Zusammenfassung der Daten wird dieser in die Lage versetzt aus Symptomen und Befunden differentialdiagnostische Hypothesen zu bilden, die die Veranlassung weiterer diagnostischer und therapeutischer Schritte nach sich ziehen [21-24]. Die Theorie, die dem Prozess der Hypothesenbildung zugrunde liegt, wird in dem Kapitel „Methoden zur Beschreibung klinischen Denkens“ beschrieben.

Im zweiten Teil der Arbeit werden die grundlegenden Erklärungsansätze für die Beschreibung der Entwicklung von Experten\_innen-Wissen dargestellt. Historisch bezieht sich der erste Erklärungsansatz auf die Erforschung der Lösung von klinischen Problemstellungen [25-26].

Dieser Forschungszweig wurde stark durch die Forschungen auf dem Gebiet der Entscheidungsfindung von Experten\_innen im Umfeld klinisch unpräziser Informationsgrundlagen geprägt. Die Grundlage für das Verständnis dieser Theorien ist die Verarbeitung unsicherer oder ungenauer medizinisch Informationen im klinischen Alltag

---

(Uncertainty in clinical reasoning). Die Erklärung dieser Theorie wird deswegen der Darstellung der kognitionspsychologischen Theorien vorangestellt. Im Anschluss werden die aufeinander aufbauenden Theorien der medizinischen Entscheidungsfindung dargestellt. Es werden die Modelle der medizinischen Wissensrepräsentation dargestellt und wie deren Entwicklung von medizinischen Anfängern bis zu klinisch erfahrenen Experten\_innen deren Entscheidungsfindung beeinflusst.

Der Kompetenzbegriff wird in unterschiedlichen Zusammenhängen, mit den zu dem jeweiligen wissenschaftlichen Kontext passenden Definitionen eingesetzt. Kompetenzen können sich auf praktische und kognitive Fertigkeiten beziehen. Allgemein werden Kompetenzen definiert, als die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, zur Lösung umschriebener Probleme. Schön definiert professionelle Kompetenz als die Fähigkeit, die ihren Träger in die Lage versetzt, nicht eindeutige Problemstellungen auf Basis unsicherer Entscheidungsgrundlagen aufgrund begrenzter Informationen zu einer Problemlösungsstrategie zu integrieren [27]. Teile dieser Kompetenzen beinhalten, die für die Verwirklichung der Problemlösungsstrategien in variablen Situationen notwendigen motivationalen, volitionalen und sozialen Befähigungen und Fertigkeiten [28].

## **Definition des Kompetenzbegriffes in der Humanmedizin**

Im Bereich der Humanmedizin wurden die wichtigsten Definitionen in den siebziger Jahren über die Definition der professionellen Kompetenzen begründet [29]. Die fachgebietsspezifische Definition in der Humanmedizin entspringt den unterschiedlichen Bedürfnissen, die die verschiedenen internationalen Gesundheitssysteme für das ärztliche Handeln erfordern. Für die Gesundheitssysteme in Nordamerika wurde in den siebziger Jahren der Begriff der professionellen ärztlichen Kompetenz eingeführt [30]. Dieser wurde im weiteren Verlauf durch das CANMEDS-Konzept mit einem kompletten Curriculum unterlegt [31]. Diese Definitionen werden in dieser Arbeit als theoretische Grundlage verwendet, weil sie im Verlauf der Entwicklungsgeschichte moderner Curricula Eingang gefunden haben. Auch die Nomenklatur in den wichtigsten europäischen Kompetenzkatalogen werden diesen entsprechend vertreten [32,33]. Ein stärker auf die medizinische Praxis bezogene Eingrenzung des in den genannten Katalogen sehr weit gefassten Kompetenzbegriffes stellt die ganzheitliche Aufgaben-Definition („whole-task“) von Schuwirth und van der Vleuten dar. Diese beschreiben professionelle Kompetenz als die

---

Befähigung zur Lösung komplexer professioneller Aufgaben, durch Anwendung, der für deren Bewältigung notwendigen kognitiven, psychomotorischen und affektiven Fähigkeiten, [34].

Professionelle Kompetenz in der Humanmedizin wird von Epstein und Hundert mit dem „gewohnheitsmäßigen und vernunftgemäßen Einsatz von Kommunikations-, Wissens- und technischen Fertigkeiten, klinischem Denken, Gefühlen, Werten und Reflexionen in der alltäglichen Praxis zum Wohle der Individuen und der Gemeinschaft“ definiert [35]. Im Rahmen der gleichen Publikation unterteilen die Autoren\_innen professionelle medizinische Kompetenz in mehrere Dimensionen, die in einem Abhängigkeitsverhältnis voneinander stehen (s. Abbildung 2), auf die im nachfolgenden Kapitel näher eingegangen werden soll.

Als Folge dieser Verfeinerung der Definition messbarer ärztlicher Kompetenzen, wurden die Empfehlungen der amerikanischen Behörde für die Akkreditierung von medizinischen Absolventen\_innen (AGCME) verändert. Der Fokus der Prüfungsmethoden wurde von dem ausschließlich wissensbasierten Prüfen auf die Überprüfung der Kompetenzen kritischen klinischen Denkens erweitert [36].

## **Professionelle ärztliche Kompetenzen**

Epstein et al teilen die professionellen ärztlichen Kompetenzen in sieben Ebenen ein [35] (s. Abbildung 2). Die Ebene der technischen Kompetenz enthält unter anderem die praktischen Fertigkeiten für die Durchführung von körperlichen Untersuchungstechniken und chirurgischen Prozeduren. Der Bewusstseinsbene werden der selbstkritische Umgang mit dem eigenen Denken, den eigenen Emotionen und Arbeitstechniken, Aufmerksamkeit und kritische Neugier, Fehlermanagement und die Berücksichtigung emotionaler und kognitiver Schwankungen zugeschrieben. Auf der kontextuellen Ebene finden sich der Umgang mit dem klinischen Kontext und das Zeitmanagement wieder. Die Dimension der Beziehungen vereinigt die Kommunikationsfertigkeiten, die Fähigkeit Konflikte zu lösen miteinander (s. Abbildung 1) Im Mittelpunkt dieser Arbeit und der weiteren Beschreibungen steht das kritische klinische Denken, das Epstein et al auf der Ebene der integrativen Kompetenzen verorten. In dieser kognitiven Ebene sind die Fertigkeiten zu selbstständigem, problemorientierten, selbstkritischen Denken und lebenslangem Lernen, um das theoretische und praktische Basiswissen und das informelle Wissen zu akquirieren und zu integrieren. Ebenso werden ihr grundlegende Kommunikationsfertigkeiten zugerechnet.



---

Diese Ebene wird deshalb eine große Bedeutung zugemessen, weil sie die kritische Beurteilungsfähigkeit mit der Anwendung klinisch-differentialdiagnostischen Denkens, der interdisziplinären Vernetzung von Grundlagenwissen mit klinischem Wissen und die Fähigkeit mit unbestimmten Situationen umzugehen vereinigt.

## **Methoden zur Beschreibung klinischen Denkens**

Die Beschreibung klinischen Denkens setzt eine Analyse der Wissensstrukturen voraus, die dem Erwerb dieser ärztlichen Kompetenz zugrunde liegen. Es ist unter den meisten medizinischen Fachgesellschaften im angloamerikanischen Sprachraum unstrittig, dass der Erwerb dieser Kompetenz eine der wesentlichen Grundlagen für erfolgreiches ärztliches Handeln im klinisch-wissenschaftlichen Kontext darstellt und somit Bestandteil der Lehr- und Prüfungsformate und der Forschung auf dem Gebiet der medizinischen Ausbildung sein muss [31]. Für die Definition der unterschiedlichen Arten der Abbildung klinischen Denkens wurden psychologische Erklärungsmodelle entwickelt. Die Theorien bauen sowohl in der Wissenschaftshistorie als auch in der zugrunde liegenden Theorie aufeinander auf und werden im Folgenden dargestellt. All diesen Theorien liegen als Kontext unpräzise medizinische Informationen zugrunde, die einen Unsicherheitsfaktor als Entscheidungsgrundlage darstellen. Sie beeinflussen maßgeblich den Weg der Entscheidungsfindung von Kliniker\_innen. Aufgrund dessen wird eine kurze Beschreibung dieses bedeutsamen Phänomens der Darstellung der verschiedenen theoretischen Modelle voran gestellt.

## **Informationen im medizinischen Alltag**

Die Praxis des differentialdiagnostischen Denkens ähnelt hinsichtlich der Wissensbasis, die den Ärzten\_innen bei dem ersten Zusammentreffen mit einem neuen Patienten\_innen zur Verfügung stehen dem Verlauf nach einer umgekehrten Pyramide [37]. Im medizinischen Alltag arbeiten Ärztinnen und Ärzte bei ihrer Entscheidungsfindung ständig im Rahmen von informell unsicheren Bedingungen. Gründe für diese fehlende Präzision bei der Entstehung von Entscheidungen für Experten\_innen sind laut Kahnemann und Gryzmala-Busse zusammengefasst die folgenden: die Wahrscheinlichkeit, dass ein Ereignis überhaupt eintritt, die wissenschaftlich zum Teil nicht leicht begründbaren Überzeugungen der Experten\_innen, die Uneindeutigkeit der zugrundeliegenden Informationen, eine Ausnahme von der Regel oder dass die Regel an sich zu kompliziert ist und deswegen vereinfacht in die Alltagsarbeit

---

implementiert werden muss und die unpräzise klassifizierte Quantifizierung von Messwerten oder die Ungenauigkeit von Messmethoden zur Erfassung von Messwerten [38,39].

In der Humanmedizin geht das der Skripttheorie zugrunde liegende theoretische Konzept, von der Tatsache aus, dass jeder Patienten\_innen kontakt medizinisch relevante Informationen über das vorliegende Krankheitsbild bei den behandelnden Ärztinnen und Ärzte hinterlässt [40]. Es erfolgt eine Kategorisierung durch Einteilung der vorgefundenen Symptome, anamnestische Angaben, klinische Befunde und Therapieinformationen. Diesen werden zu jedem einzelnen Patienten\_innen bestimmte Attribute zugeordnet, die in einem für jede Diagnose typisch zeitlich-räumlichen Abfolge auftreten können. Die Kombination der Dimensionen bildet für jede Diagnose und für jedes Krankheitsbild einen spezifischen Sinnzusammenhang, der ähnliche Befunde weiterer Patienten\_innen zu einem Netzwerk miteinander verbindet. Skripte befinden sich demnach in einem ständigen Umbauprozess und fortwährender Erweiterung. Bestehendes Wissen wird, mit den neuen Informationen, nach den zuvor genannten Dimensionen in Zusammenhang gesetzt. Klinische Expertise hängt in diesem Sinne von der Anzahl der Krankheitsbilder ab, die von dem Arzt, der Ärztin im Laufe seiner Laufbahn bis dahin untersucht und behandelt wurden.

## **Hypothetikededuktive Methode - Prozessorientiertes Wissen**

Die Theorie des hypothetiko-deduktiven Denkens beschreibt den Ablauf der mentalen Prozesse, die dem medizinisch-klinischen Denken bei Ärzten und Ärztinnen aus kognitionspsychologischer Sicht zugrunde gelegt werden [21, 41]. Sie geht von der Annahme aus, dass bei Ärzten\_innen vom Erkennen der ersten Symptome an, die Generierung von Hypothesen über die zugrundeliegenden Differentialdiagnosen beginnt [40]. Es wird ein Prozess initiiert, der mit der Sammlung medizinischer Informationen über den Patienten\_innen anfängt. Auf dieser Basis werden die bisher erstellten Theorien um neue Hypothesen ergänzt oder reduziert. Aufgrund dessen werden weitere diagnostische und therapeutische Maßnahmen geplant und durchgeführt, die wiederum eine Ergänzung der Hypothesenbildung nach sich ziehen. Am Ende des Prozesses wählt die Kliniker\_innen\_innen die wahrscheinlichste Diagnose, die für die vorliegenden Befunde und Behandlungsergebnisse spricht (s. Abbildung 2).

Die psychologischen Hintergründe, die für die Auswahl der spezifischen Inhalte und die Organisation der in Frage kommenden Differentialdiagnosen dieser Theorie zufolge verantwortlich sind, werden selbst allerdings nur unzureichend erklärt [24]. Die

---

Unterscheidung zwischen erfahrenen und unerfahrenen klinischen Experten\_innen zeigt, dass dieser Erklärungsansatz alleine nicht ausreicht, um den Werdegang einer klinischen Entscheidung quantitativ und qualitativ nachzuvollziehen. Beide Gruppen nutzen die Methode in ihren differentialdiagnostischen Denkprozessen. Die erfahreneren Experten\_innen nutzen sie vor allem bei differentialdiagnostisch schwierig einzuordnenden oder untypischen Befunden, also bei sehr starken Abweichungen von ihren bereits bestehenden Konzepten, während Anfänger diese gleich zu Beginn ihres differentialdiagnostischen Denkens nutzen [42].

## **Strukturtheorien – Wissensverkapselung**

Im Gegensatz zu den prozessorientierten Theorien, wie der hypothetiko-deduktiven, gehen die Strukturtheorien von der These aus, dass kausale Wissensstrukturen als Grundlage für diagnostische Wissensprozesse dienen [40]. Ausgangspunkt dieser Theorien ist die These, dass der Zuwachs des Verständnisses kausaler Zusammenhänge in einem Spezialgebiet die Entwicklung klinischer Expertise repräsentiert. Im Gegensatz zu der Annahme des reinen Wissensgewinnes steht die These, dass die stetige Veränderung und Neustrukturierung des Wissens durch zunehmende Erfahrung die Entwicklung klinischen Denkens bestimmt. Der Entwicklungsprozess, der die Entstehung der Wissensstrukturen steuert, wird von den Vertretern dieser Theorie wie folgt beschrieben: während die Neulinge in einem Fachgebiet für die Erklärung und Herleitung von klinischen Symptomen kausale Wirkungsketten abrufen und auf Übereinstimmung mit den vorliegenden Symptomen überprüfen, wenden Kliniker\_innen mit zunehmender Erfahrung das vergleichende Erkennen typischer klinischer Muster für die Diagnoseerstellung an [43]. Die Kenntnisse in den grundlegenden Wissenschaften der Pathophysiologie beziehungsweise in den medizinischen Grundlagenwissenschaften (Anatomie, Biochemie, Physiologie) werden für das entsprechende Krankheitsbild zu einem Wissensblock zusammengefasst oder besser eingekapselt [20]. Der Weg, mit dem diese mentalen Strukturen erreicht werden, kann als mehrstufiger Lernprozess beschrieben werden. Auf Grundlage der Wissensverkapselung, entstehen kontextspezifische Wissensstrukturen, die unabhängig von kausalen Zusammenhängen Informationen in einen vergleichenden Sinnzusammenhang setzen. Auf der Basis von Wahrscheinlichkeiten setzen Experten\_innen mit zunehmendem Erfahrungshintergrund die klinischen Informationen über ihre Patienten\_innen in einen vergleichenden Bezugsrahmen. Der Bezugsrahmen wird durch Krankheitskripte (Illness

---

Scripts) repräsentiert [17]. Diese sammeln und repräsentieren die wichtigsten und charakteristischen Informationen, für die von dem Arzt behandelten Patienten\_innen und deren Krankheitsbilder, die durch den Experten\_innen bereits zuvor behandelt wurden. Medizinische Experten\_innen bilden mit zunehmenden Erfahrungsgrad ein semantisches Netzwerk zwischen den einzelnen Skripten aus. Die Skripttheorie postuliert ein vergleichendes Vorgehen bei der Neuvorstellung von Patienten\_innen, indem Übereinstimmungen zwischen den vorhandenen Symptome und den bereits vorhandenen, miteinander vernetzten Skripten gesucht werden, bis eine Verdachtsdiagnose gestellt werden kann [19].

## **Skripttheorie**

Die Skripttheorie ist ein theoretischer Ansatz, der die mentale Organisation klinischen Wissens bei Ärzten\_innen beschreibt [17]. Sie basiert auf dem „Schema“-Begriff, der erstmals von Minsky eingeführt wurde. Ein Schema beschreibt Gedächtnisstrukturen, die Informationsverarbeitung von neuen Informationen durch die Verknüpfung mit bereits bestehenden Wissensstrukturen ermöglichen.

Schemata repräsentieren dieser Theorie zufolge prototypische kognitive Abbildungen von Objektinstanzen, die durch bestimmte Parameter und ihre Beziehungen zueinander definiert sind [7]. Eine besondere Form der Schemata stellen Skripte dar. Die Idee der Ausbildung von Skripten, in diesem Zusammenhang, wurde von Schank und Abelson begründet. Skripte werden dort als Situationen im sozialen Kontext beschrieben, die mit häufig wiederkehrenden ähnlichen Verhaltensweisen gekoppelt sind [44]. Es wird postuliert, dass Skripte untereinander ein semantisches Netzwerk ausbilden. Dies liegt der Verknüpfung aller relevanten Informationen, die für den kognitiven und praktischen Umgang mit medizinischen Situationen notwendig sind, zugrunde.

Der Ansatz der Skript-Prozessierung, den Feltovich und Barrows verfolgen, basiert auf folgendem Konzept: die bei jedem klinischen Kontakt entstehenden Skripte durchlaufen einen auf Kausalketten basierenden Plausibilisierungsprozess, an dessen Ende keineswegs eine völlige Übereinstimmung mit Merkmalen vorbestehender Skripte erreicht werden muss. Für die Diagnosestellung werden auslösende Faktoren, pathogenetische Fehlfunktionen und daraus resultierende Konsequenzen mit bestehenden Skripten semantisch assoziiert. Am Ende des Prozesses muss der Arzt entscheiden, inwiefern die fehlende Übereinstimmung mit

---

bestehenden Skripten trotzdem eine Diagnoseentscheidung für ein bestimmtes Skript rechtfertigt [45].

Schmidt et al formulierten basierend auf dieser Theorie ein Konzept für die Beschreibung des differentialdiagnostischen Wertes von Skripten für die Behandelnden. Er entsteht durch die Tatsache, dass jedes Skript, das sich der oder die jeweilige Untersucher\_in erstellt hat, durch seine Attribute und den diesen zugeordneten Werten die jeweilige Instanz eines Krankheitsprozesses zu einem bestimmten Zeitpunkt charakterisiert. Stimmen bei den untersuchten Patienten\_innen die den Attributen zugewiesenen Werte nicht mit den in den bisherigen Skripten vorliegenden Werten überein, werden weitere, semantisch verknüpfte Skripte heran gezogen, bis eine maximale Übereinstimmung erreicht ist [19,20]. Die einzelnen Werte sind nicht exklusiv für die jeweilige Erkrankung, sondern können bei verschiedenen Erkrankungen vorkommen. Das entscheidungsbildende Skript, das ausschlaggebend für die Diagnosestellung ist, entsteht durch die Verknüpfung mit den Symptomen und anamnestischen Angaben, die der Patient zu dem jeweiligen Zeitpunkt aufweist oder aufgewiesen hat. Diese werden im diagnostischen Prozess im Rahmen der Hypothesenbildung hypothetiko-deduktiv mit ähnlichen Skripten und dem individuellen medizinischen Grundlagenwissen des Untersuchers abgeglichen, bis eine größtmögliche Übereinstimmung mit einem bestehenden Skript hergestellt ist. Dieser Prozess vollzieht sich unbewusst während des gesamten Patienten\_innen-Kontakts und leitet die Anamneseerhebung und Diagnostik. Wichtig für diesen Prozess ist eine enge Interaktion der Skripte miteinander. Nelson zufolge erfolgt die Speicherung durch die Interpretation sozialer Beziehungen innerhalb bestimmter Kontexte mit den dazugehörigen Informationen über Personen, Objekte, Aktionen und deren Beziehungen zueinander [46]. Die Ausbildung klinischer Expertise entsteht durch die ständige Erweiterung der in Skript-Datenbasis für jede einzelne Erkrankung. Je mehr individuelle Skripte vorhanden sind, auf die die Ärzte\_innen mit zunehmender Erfahrung zugreifen können, desto mehr und schneller wird die Wahrscheinlichkeit für zutreffende Verdachtsdiagnosen zunehmen.

## **Mustererkennung**

Eine frühe Form der Expertise-Forschung hat sich mit der Mustererkennung auseinander gesetzt. Die ersten Ansatzpunkte für die Darstellung von Mustern in der Entwicklung von Expertise auf einem bestimmten Gebiet, haben sich auf dem Gebiet der Schachforschung entwickelt. Auf dem Gebiet der Schachexpertise konnte belegt werden, dass diese sich durch

---

das Wiedererkennen typischer Stellungsmuster auszeichnet. Es konnte belegt werden, dass das Wiedererkennen von bestimmten Stellungen und Positionen ein größerer Stellenwert beizumessen ist, als der vorausschauenden Planung [47]. Der Unterschied für den Erfolg zwischen erfahrenen und unerfahrenen Schach-Experten\_innen, zeichnet sich durch das Auswahlverhalten der erfahrenen Experten\_innen und die Rückgriffsmöglichkeiten auf eine größere Anzahl von erinnerten Positionen aus [48]. Schachgroßmeister zeichnen sich neben ihren kombinatorischen Kompetenzen durch die Fähigkeit aus, eine spezifischere Auswahl unter den ihnen umfangreicher als Laien erinnerlichen Stellungen zu treffen und somit schneller zum Ziel zu kommen. Diese Forschungsergebnisse hatten entscheidenden Einfluss auf die kognitionspsychologischen Aspekte der Experten\_innen-Forschung in anderen Fachrichtungen, insbesondere der Humanmedizin [49,50].

Die Theorie der rein wissensbasierten Entscheidungsfindung wurde durch die Darstellung des „Intermediate Effektes“ durch Schmidt et al widerlegt. Er konnte belegen, dass klinische Experten\_innen, die sich in einem „Zwischenstadium“ des klinischen Experten\_innentums befanden, bei der Memorierung klinischen Detailwissens der ihnen zu Beginn der Studien vorgelegten Patienten\_innen -Fälle deutlich besser abschneiden, als Anfänger und weiter fortgeschrittene Experten\_innen. Voraussetzung ist, dass man ihnen ausreichend Zeit für die Reaktivierung des Wissens lässt [51].

## **Bedeutung von medizinischem Grundlagenwissen**

Grundlage für den Erwerb und die Nutzung klinischen Experten\_innenwissens ist der Rückgriff auf medizinisches Grundlagenwissen. Es gibt zwei verschiedene Ansätze, die Bedeutung von Grundlagenwissen bei der Ausübung klinischer Expertise einzuschätzen. Patel und Groen konnten in Studien zeigen, dass der Rückgriff auf medizinisches Grundlagenwissen von Studierenden verschiedener Ausbildungsstufen stärker für die Ableitung von Schlussfolgerungen genutzt wird, als von erfahrenen Kliniker\_innen [43]. Die Studienergebnisse wurden hinsichtlich der curricularen Einflüsse auf den späteren Einsatz von medizinischem Grundlagenwissen, das während des Studiums erlernt wurde, durch Schmidt und Patel später selbst in Zweifel gezogen [51, 52].

Traditionelle Curricula unterstützen durch die zeitliche strenge Trennung der vorklinisch-wissenschaftlichen, grundlagenwissenschaftlichen von den klinischen Inhalten, die spätere mentale Trennung von biomedizinischem Grundlagenwissen und klinischem Experten\_innen-

---

Wissen [52]. Änderungen in der Versuchsanordnung legen einen Einfluss der zugrundeliegenden biomedizinischen Daten auf das kontextuelle Lernen nahe. Studierende, die ihren Fokus während der Fallbearbeitung auf das Erlernen biomedizinischer Daten legen, wiesen in der Studie von Woods einen stärkeren Langzeitgedächtnis von grundlagenmedizinischen Fakten auf, als die Studierenden, die nur auf der Basis von Wahrscheinlichkeiten gelernt haben [53].

Henk Schmidt konnte mit seiner Forschungsgruppe mit Hilfe von Think-aloud-Protokollen und Post-hoc-Analysen die Ergebnisse unabhängig von Patel bestätigen. Klinische Experten\_innen gelingt es, mehr Schlussfolgerungen aus den angebotenen Daten abzuleiten, zeichnen sich durch effizientere Lösungswege für klinische Problemstellungen aus und greifen bei schwierigen Problemstellungen, für die sich aus den vorhandenen Schemata keine Lösungen ermitteln lassen, vermehrt auf biomedizinische Erklärungsansätze zurück [51]. Hinsichtlich des Einflusses problemorientierter, kontextbasierter Lernmodelle auf die Integration und den Einfluss biomedizinischen Wissens auf die Diagnosefindung besteht eine kontroverse Diskussion, trotz einer vermehrten Integration biomedizinischen Wissens durch die PBL-Studierenden, die keine eindeutigen Vorteile bei der Diagnosestellung hierdurch belegen konnten [50].

Weitere Studien haben die Bedeutung und den Umfang des Erwerbs von medizinischem Grundlagenwissen bei der Entwicklung vom/von der klinischen Anfänger/in zum/zur Experten\_innen bestätigt. Sie konnten die Thesen widerlegen, dass der Anteil des medizinischen und wissenschaftlichen Grundlagenwissens bei zunehmender klinischer Expertise weniger in Anspruch genommen wird [42]. Schmidt et al und später Verkoeijen konnten anhand der Theorie des Intermediate-Effektes belegen, dass die Entwicklung klinischer Expertise mehrstufig verläuft. Sie basiert ihren Modellen zufolge auf einem parallelen Zuwachs von biomedizinischem und klinischem Wissen [54]. Der Prozess der Wissensverkapselung (s. Abbildung 4) führt zu einer autonomen, semantischen Vernetzung der klinischen Erfahrungen mit den dazugehörigen theoretisch-wissenschaftlichen Hintergrundinformationen.

## **Bedeutung von klinischem Wissen**

Klinisches Wissen kann kurz gefasst unterteilt werden in das klinisch-theoretische, grundlagen-medizinische Wissen, das sich auf dem theoretischen Grundlagenwissen aufbaut

---

und dem praktischen, Patienten\_innen bezogenen Wissen, das auf den klinischen Fällen beruht, die von dem/der Medizinerin behandelt wurden [55].

## **Semantische Verknüpfung von Wissen**

Die Prozesse, die zur Skriptbildung führen und Skripte semantisch miteinander verknüpfen, stellen die Grundlage für den mentalen Entwicklungsprozess, des Aufbaus, der Ergänzung und des Umbaus der Skripte bei Kliniker\_innen dar. Es stehen verschiedene Theorien zur Begründung dieser Aktivitäten zur Verfügung, deren Bezug zur Skriptentstehung im Folgenden erläutert werden soll.

Elstein zufolge ist die psychologische Grundlage für die Entwicklung von Problemlösungsstrategien die Hypothesenbildung. Sehr großen Datenmengen, die mental in klinischen Situationen bewältigt werden müssen, stehen begrenzten kognitiven Kapazitäten gegenüber. Experten\_innen unterteilen in klinischen Situationen die vorhandenen Datenmengen in kleinere klinische Analysepakete, denen sie klinische Hypothesen zuweisen. Anschließend wird im Rahmen einer klinischen Datensammlung versucht, die Informationen abzugleichen, um die Thesen zu bestätigen oder zu widerlegen [5].

Ausgehend von der Theorie Elsteins haben Grant und Marsden versucht, den Prozess der aktiven Informationsverarbeitung bei Kliniker\_innen darzustellen. Demzufolge eröffnen bestimmte klinische Schlüsselinformationen den Zugang zu bereits bestehenden Erinnerungsstrukturen. Die Erinnerungsstrukturen und das Vorwissen des Kliniker\_innen s geben vor, wie während der Interpretation der klinischen Symptome, der Prozess der Bestätigung und Widerlegung bestimmter Krankheitseigenschaften bis zur endgültigen Diagnose verläuft [56].

Patel und Groen verwenden in ihren Analysen einen strukturalistischen Ansatz, der von einer propositionalen Vorgehensweise ausgeht. Die von Experten\_innen verwendeten semantischen Strukturen werden durch medizinische Begrifflichkeiten repräsentiert [43]. Deren kausale Abhängigkeiten voneinander werden durch grammatikalische Kausalketten beschrieben, die von den Experten\_innen verwendet werden, um die klinischen Zusammenhänge darzustellen. Da die Begrifflichkeiten mit dieser Messmethode in einem zeitlich linearen Zusammenhang auftreten, stellten Patel und Groen die Theorie des Vorwärtsdenkens („Forward Reasoning“) auf. Diese Methodik wird von erfahrenen Experten\_innen während des klinisch-diagnostischen Denkens verwendet. Unerfahrene Mediziner\_innen oder Studierende verwenden hingegen zeitintensiveres analytisch-retrogrades Ableiten von Schlussfolgerungen



---

(„Backward Reasoning“) als Methode zur Erstellung von Diagnosen. Die Theorie wurde von verschiedenen Seiten angezweifelt [57].

Bordage zufolge liegt der Wissensbildung bei klinisch Experten\_innen ein semantisches Netzwerk aus miteinander verknüpften Wissensblöcken dem erfolgreichen klinischen Denken zugrunde. Der Ausbildung eines derartigen Wissensnetzes liegen die Erfahrungen zugrunde, die jeweiligen klinische Experten\_innen in Form von klinischen Fällen gesammelt haben. Jeder klinische Fall weist charakteristische Merkmale seiner Instanzen auf. Diese Informationen werden im Langzeitgedächtnis als Wissensblöcke gespeichert. Deren Mehrwert für das klinische Denken steigt durch den Grad der Vernetzung, den die in ihnen enthaltenen Einzelinformationen enthalten. Die Vernetzung beruht auf abstrakten Gedächtnisstrukturen, die die konkreten Instanzen repräsentieren. Diese werden während des klinischen Denkprozesses vergleichend zu den real wahrgenommenen Symptomen und Befunden der Patienten\_innen in Beziehung gesetzt [40].

## **Methoden zur Messung des klinischen Wissenserwerbs**

Traditionell wurden in der Humanmedizin im deutschen Sprachraum zur Prüfung des theoretischen Wissens der Studierenden mehrheitlich Multiple-Choice-Fragen und mündliche Prüfungen eingesetzt. Bedingt durch den inhaltlichen Wandel medizinischer Curricula, hin zu einer stärkeren Überprüfung theoretischer und praktischer Lernziele und der damit verbundenen Kompetenzen, wurden praktische Prüfungsformate entwickelt, die den klinischen Kontext und die Überprüfung praktischer Fertigkeiten mit in die Prüfung einbeziehen. Zu diesen zählen in Bezug auf die Überprüfung klinischer Kompetenzen, die Objective-Structured-Clinical-Examination (OSCE) und Mini-Clinical Examination-Exercise (Mini-CEX) [4, 58]. Schuwirth und van der Vleuten empfehlen für die Messung spezifischer Kompetenzen die Auswahl eines geeigneten Fragentyps unter Berücksichtigung der folgenden Kriterien [23, 34] :

1. Reliabilität
2. Validität
3. Pädagogisches Wirkungsspektrum
4. Ressourcenintensität
5. Akzeptanz
6. Umsetzbarkeit

---

Prüfungsformate, die diese Kriterien in ausreichender Güte aufweisen sollen, setzen einen hohen Grad der Standardisierung, der in den Prüfungen enthaltenen Informationen in Form von Checklisten voraus. Von diesen Formaten wird nur ein Teil der mit dem klinischen Denken in Zusammenhang stehenden Kompetenzen überprüft.

Klinisches Denken und Entscheiden wird auf der Basis eines unvollständigen und immer wieder wechselnden Informationspools durchgeführt. Diese Informationen können in bestimmten frühen Stadien der klinischen Informationssammlung und auf dieser basierenden Entscheidungsfindung, im Gegensatz zu den sehr präzise umschriebenen Inhalten der MC- und OSCE-Fragenformate und deren fest definierten Antwortvorgaben, als unsicher beschrieben werden. Die Antworten erfahrener Kliniker\_innen auf die die vage umschriebenen Probleme können in Praxis und ohne Konsensusprozess stark voneinander variieren. In der vorliegenden Studie soll die Kompetenz des klinischen Denkens bei den Studierenden, ausgedrückt durch die Fähigkeit differentialdiagnostische und therapeutische Entscheidungen zu treffen, gemessen werden. Für die Messung dieser professionellen Entscheidungs- und Handlungskompetenz wurde der Skript-Konkordanztest als neues Prüfungsinstrument seit 1999 neu entwickelt [59]. Im weiteren Verlauf dieser Arbeit wird der deutsche Begriff Skript-Konkordanz-Test (SKT) synonym für den Script Concordance Test verwendet. Im Folgenden wird in Abgrenzung zu den etablierten Prüfungsformaten der SKT beschrieben.

## **Multiple-Choice-Fragen (MCQ)**

MCQs stellen das Hauptprüfungsformat in der medizinischen Aus-, Fort- und Weiterbildung im deutschsprachigen Raum dar. Es gibt zwei Antworttypen: der am häufigsten eingesetzte der einfachen Richtig-Falsch-Antworten mit einer Mehrfach-Auswahl von drei bis fünf verschiedenen Falschantworten und einer Richtig-Antwort und den seltener eingesetzten True/False-Fragen, die aus einem Stamm mit einem richtigen und einem falschen Anteil besteht. Der Vorteil besteht in der einfachen Auswertbarkeit der Fragen, der mit wenig Prüfungszeit im Vergleich zu allen anderen Methoden erreichbaren hohen Reliabilität und bei sorgfältiger Planung ebenfalls hohen Validität [60]. Nachteil ist der hohe Zeitaufwand, der für die Erstellung der Fragen aufgewendet werden muss [61]. Inhaltlich kann mit diesem Test sowohl klinisch-orientiertes Handlungswissen, als auch theoretisches Grundlagenwissen geprüft werden [62].

---

## **Extended-matching questions (EMQ)**

EM-Fragen bieten zehn bis zwanzig Diagnoseoptionen an, die die zu Prüfenden anschließend Fallvignetten zuordnen müssen, die sie in sequenzieller Abfolge angeboten bekommen. Die Diagnosen können auf eine oder aber auch auf alle Vignetten zutreffen. Ziel ist die Überprüfung von Entscheidungs- und Problemlösungskompetenz. Vorteil des Formats ist die Minimierung des Cueing-Effektes, also der Fähigkeit des Prüflings, die richtige Antwort zu erkennen, statt aus dem eigenen Wissen heraus zu erkennen [63]. Nachteil ist die Notwendigkeit, das Format aufgrund der Mehrfachauswahlen computerbasiert prüfen zu müssen, was den Aufwand für die Erstellung und Durchführung der Prüfung erhöht.

## **Objective structured clinical examination (OSCE)**

OSCEs gehören zu den nicht schriftlichen Prüfungsformaten. Ziel des Formates ist die Überprüfung von praktischem, kommunikativen und handlungsorientierten Wissen und Fertigkeiten. Im Rahmen von sieben bis fünfzehn nacheinander zu absolvierenden und fünf bis fünfzehn Minuten dauernden Prüfungsstationen, werden die Dozent mit einer typischen klinischen Situation konfrontiert, in der sie ihre praktischen Fertigkeiten auf dem Gebiet der Anamneseerhebung, Beratung und Diagnostik unter Beweis stellen müssen. Die Prüfungskriterien sind vorher auf einer Checkliste festgehalten worden, an die Prüfer\_innen sich halten müssen. Nachteile des Formats sind die mit der Durchführung verbundenen hohen Kosten. Um eine zu MC vergleichbare Reliabilität zu erreichen, muss ein Drittel mehr Prüfungszeit veranschlagt werden, einschließlich der mit dem Format verbundenen Bereitstellung von infrastrukturellen Rahmenbedingungen in Form eines Trainingszentrums für das Trainieren und die Überprüfung der praktischen Fertigkeiten [60].

## **Key-Feature-Fragen (KF)**

Der Key-Feature-Ansatz beruht auf Fallvignetten, die eine kurze und präzise Beschreibung aller Symptome und für das Verständnis des Falles wichtigen Daten enthalten. Die dargestellten Key Features sichern den inhaltlichen Bezug zu dem Fall, weil sie die entscheidenden Informationen für die Lösung des Falles enthalten. Die folgende Beschreibung erfordert weitere Handlungsschritte von den Geprüften [64]. Nach der Beantwortung einer Frage werden neue Informationen über den weiteren Verlauf des Falles angeboten, um eine sich thematisch anschließende Frage, die sich auf ein neues Key-Feature bezieht, zu stellen. Vorteile des Formates sind die einfache Erstellung von Fragen durch

---

Kliniker\_innen und die Überprüfung von klinischem Handlungswissen. Die Auswahl der Antworten kann per Long-Menu-Fragen erfolgen, so dass eine computerbasierte Prüfung ein vorteilhaftes Prüfungssystem darstellt [65].

## **Skript-Konkordanz-Test**

Die Kompetenz für die Lösung von mit unpräzisen Informationen ausgestatteten Problemstellungen, wird als professionelle Problemlösungskompetenz umschrieben [66]. Diese Kompetenz lässt sich mit den herkömmlichen Testformaten (MCQ, OSCE, MEQ) nicht überprüfen [58, 66]. Die klassischen Prüfungsmethoden MCQ oder OSCE erreichen ihr hohes Maß Validität und Reliabilität durch präzise beschriebene Problemstellungen und bezogen auf diese, die Überprüfung detailliert beschriebener Outcomes in Bezug auf die für die Lernziele des Tests zugrunde gelegten Kompetenzen. Diese Problemstellungen erfordern von den Kliniker\_innen eine variable Lösungsstrategie, die sie in die Lage versetzt, im Kontext der unsicheren Entscheidungsfindung ein strukturiertes Vorgehen zu entwickeln, das zu einer hohen Lösungswahrscheinlichkeit für die vorliegende Aufgabe führt. Ausgangspunkt für diese ist die genaue Definition des Problems. Weitere Merkmale unscharf umschriebener, klinischer Problemstellungen, sind die unterschiedlichen Lösungsstrategien, die von Experten\_innen angewendet werden können, um Lösung des Problems zu kommen. Die verschiedenen Lösungsansätze ziehen nicht unbedingt eine einzige richtige Antwort nach sich [59].

Die Prozesse, die zu der richtigen Lösung von klinisch unscharf beschriebenen Problemstellungen beitragen, werden mit diesen Methoden nicht abgebildet, noch überprüft. Für die Messung der kognitiven Kompetenzen, die für die Entwicklung von differentialdiagnostischen und therapeutischen Strategien und für die Abbildung von Einstellungen und Haltungen wesentlich sind, sind die vorab genannten Testformate nicht geeignet [23, 68].

Der Skript-Konkordanz-Test wurde von Charlin und Kollegen als Ergänzung zu den bereits bestehenden Prüfungsverfahren entwickelt, um ein Prüfungsformat einsetzen zu können, das die Kompetenz zur Lösung typischer, klinischer, unpräzise definierter Problemstellungen, hypothetisch also das damit verbundene kritische klinische Denken zu prüfen in der Lage ist [59]. Die Differenzierung in präzise und unpräzise definierte (Ill-defined) Problemstellungen geht auf Schön zurück und wurde für die humanmedizinische Forschung durch Fox und Kollegen erweitert. Im klinischen Alltag müssen sich die Kliniker\_innen mit unpräzise

---

umschriebenen Problemstellungen in verschiedenen Lösungsstadien auseinandersetzen, die sich durch die Kombination von untypischen Symptomen und bruchstückhaften Informationen auszeichnen, die vervollständigt werden müssen, um zu einer Verdachtsdiagnose zu gelangen oder eine Therapieentscheidung treffen zu können [27, 37].

Eine SKT-Frage kann Diagnose-, Untersuchungs- oder Therapieentscheidungen überprüfen. Eine Skript Konkordanz Test-Frage, hier als Test-Item bezeichnet, besteht aus einer wenige Sätze umfassenden Beschreibung einer oder eines Patienten\_innen /in (Fallvignette). Diese enthält den ersten Teil der klinischen Informationen des beschriebenen Falles. Der Inhalt der Beschreibung beschränkt sich auf einen Datensatz, der Ärzten\_innen in typischen klinischen Alltagssituationen vor der Erhebung der ersten Verdachtsdiagnosen zur Verfügung steht. Er kann sich auf die verschiedenen, typischen Stadien klinischer Situationen beziehen. Anhand dieser Kriterien folgt der Fallvignette der Teil des Items, der neue Informationen zu dieser anbietet. Diese können entweder körperliche Befunde, klinische Zeichen, Laborbefunde oder diagnostische Befunde sein. Die Bewertung wird auf einer Fünf-Punkt-Likert-Skala abgegeben.

Die Reliabilität und Validität des Skript Konkordanz Testes wurde in mehreren Studien mit etablierten Prüfungsverfahren vergleichbar hoch gemessen. Die Reliabilitätskoeffizienten erreichten in den ersten beiden vergleichenden Studien in der Frauenheilkunde/Geburtshilfe und der Radiologie zwischen PJ-Studierenden und AssistenzÄrzten\_innen (n=103) Cronbach Alpha-Werte zwischen 0,79 und 0,81 in einer dritten Studie wurden mit PJ-Studierenden der Allgemeinmedizin (n=24) Werte zwischen 0,45 und 0,34 erreicht werden [59, 69]. Weitere Studien evaluierten die Methode bei Weiterbildungs-Assistenten in der Urologie und in der pädiatrischen Notfallmedizin und konnten die vielversprechenden Ergebnisse der ersten Studien hinsichtlich deren Reliabilität (Cronbach's Alpha 0,77) und Validität bestätigen [16, 70)].

## Beispiele für den SKT an der Charité in Berlin

### Illustration von drei verschiedenen Fragetypen

#### 1. Differentialdiagnostischer Typ

Ein 5jähriges Mädchen wird mit Gewichtsverlust um 3kg in den letzten 2 Wochen und zunehmender Müdigkeit vorgestellt.

| <i>Wenn Sie an folgende Diagnose denken:</i> | <i>Und Sie erfahren:</i>            | <i>wird Ihre Verdachtsdiagnose:</i> |
|--|-------------------------------------|-------------------------------------|
| <b>Diabetes mellitus</b>                     | <b>der Urinstix ist unauffällig</b> | <b>-2   -1   0   +1   +2</b>        |

Legende:

-2: ausgeschlossen oder fast ausgeschlossen

-1: weniger wahrscheinlich

0: hat keinen Einfluss auf Ihre Verdachtsdiagnose

+1: wahrscheinlicher

+2: fast oder vollständig gesichert

#### 2. Untersuchungstyp

Ein 5jähriges Mädchen wird mit Gewichtsverlust um 3kg in den letzten 2 Wochen und zunehmender Müdigkeit vorgestellt.

| <i>Wenn Sie an folgende Untersuchung denken:</i> | <i>Und Sie erfahren:</i>  | <i>bedeutet das für Ihre Untersuchung:</i> |
|--|---|--|
| <b>manuell differenziertes Blutbild</b>          | <b>sie hat in den letzten Nächten wiederholt eingenässt, obwohl das lange nicht mehr vorkam</b> | <b>-2   -1   0   +1   +2</b>               |

Legende:

-2: nicht indiziert

-1: weniger sinnvoll

0: hat keinen Einfluss auf Ihre Entscheidung

+1: sinnvoller

+2: notwendig

### 3. Therapietyp

Ein 5jähriger Junge mit einer schweren Diarrhoe und einer nachgewiesenen Salmonelleninfektion wird zur stationären Behandlung eingewiesen.

| <i>Wenn Sie an folgende Therapie denken:</i> | <i>Und Sie erfahren:</i>                                     | <i>wird Ihre Verdachtsdiagnose:</i> |
|--|--|-------------------------------------|
| <b>Antibiotikatherapie</b>                   | <b>Im Blutbild zeigt sich eine Leukozytose von 22.000/nl</b> | <b>-2   -1   0   +1   +2</b>        |

Legende:

-2: kontraindiziert

-1: weniger sinnvoll

0: hat keinen Einfluss auf Ihre Therapie

+1: sinnvoller

+2: notwendig

## Methodik

### Material und Methoden

#### Methodisches Vorgehen

Die Methodik der Testanalyse der SKT-Items beruht auf den Methoden, die von Norcini et al eingeführt und von Charlin et al validiert wurden [14, 18]. Die Testfragen wurden nach den elf am Otto-Heubner Centrum (OHC) vertretenen Fachbereichen der Kinder- und Jugendmedizin fachlich gegliedert und proportional aufgeteilt.

#### Forschungsleitende Thesen

Die für die Entwicklung des differentialdiagnostischen Denkens führende Kompetenz des kritischen klinischen Denkens, sowie die für dessen Entwicklung zugrundeliegenden Hypothesen, wurden in den vorangehenden Kapiteln ausführlich dargestellt. Die Hypothese legt einen Entwicklungsprozess des klinischen Denkens zugrunde. Basierend auf dem wissenschaftlichen Grundlagenwissen, geleitet durch handlungsorientiertes Wissen finden Kliniker\_innen problemorientiert fallspezifische Lösungen für ihre Patienten\_innen. Die für

---

die Festlegung der passenden diagnostischen und therapeutischen Schritte Differentialdiagnosen, beruhen auf klinischen Informationen über den Patienten\_innen . Die Unsicherheit für den Kliniker\_innen besteht in der untypischen und unscharfen Zusammenstellung der zu einem bestimmten Zeitpunkt gegebenen Informationen. Die Ärzte\_innen müssen in diesem vorerst ungenauen informellen Umfeld ihre Hypothesen aus ihren bisherigen Erfahrungen mit Patienten\_innen mit ähnlichen Symptomen ableiten. Medizinstudierende konnten diese Erfahrungen noch nicht sammeln oder in einem sehr geringen Umfang sammeln und könnten theoretisch die fehlenden klinischen Erfahrung durch die Aktivierung von klinischen Grundlagenwissen ausgleichen, dies postulieren Theorien die hypothetiko-deduktiven Denkmodelle analysiert haben.

Aus diesen Theorien leiten sich die folgenden Thesen ab:

1. Der Skript-Konkordanz-Test ist ein Instrument um das kritische klinische Denken von Studierenden nach Abschluss des Unterrichts in der Kinderheilkunde reliabel und valide zu messen.
2. Der SKT ist in der Lage den Unterschied des klinischen Denkens zwischen Studierenden und Experten\_innen zu messen.
3. Die Dozierenden des Blockpraktikums haben durch ihren Unterricht einen Einfluss auf den Erfolg der Studierenden im SKT.
4. Die theoretischen Grundkenntnisse der Kinderheilkunde, repräsentiert durch die Multiple Choice Klausur, beeinflussen das Ergebnis der Studierenden im SKT.
5. Geschlechtsunterschiede haben keinen Einfluss auf das Ergebnis im SKT.

### **Fragestellungen**

1. Ist der SKT ein valides Testinstrument für den studentischen Unterricht in der klassischen Medizin am Beispiel der Kinderheilkunde?
2. Inwieweit unterscheidet sich die Kompetenz der Studierenden im kritischen klinischen Denken in der klassischen Medizin hier der Kinderheilkunde von der Kompetenz der Experten\_innen (Ärzte\_innen/ Fachärzte\_innen/ Professoren\_innen)?
3. Haben unterschiedliche Kompetenzniveaus klinischer Experten\_innen einen Einfluss auf die Entwicklung des kritischen klinischen Denkens bei den Studierenden?



- 
4. Welchen Einfluss haben die grundlagentheoretischen Kenntnisse der Studierenden in der klassischen Medizin speziell am Beispiel der Kinderheilkunde auf den Erfolg des klinischen Denkens?

### **Studiendesign**

Es wurde eine prospektive Kohortenstudie durchgeführt. Alle Studierenden des sechsten klinischen Semesters des Regelstudiengangs wurden verpflichtet, am Ende des Blockpraktikums Kinderheilkunde an einem Skript-Konkordanz-Test als notengebenden Test teilzunehmen. Einverständnis für die Klausureinsicht wurde von dem Assessmentbereich der Charité-Universitätsmedizin eingeholt. Alle Studierenden dieses Semesters wurden schriftlich um ihr Einverständnis gebeten, der Studienleitung Einsicht und anonyme Auswertung ihrer MC-Klausurergebnisse aus dem Vorsemester (Gynäkologie, Humangenetik, Pädiatrie) in Bezug zu ihren SKT Ergebnissen zu geben.

### **Beschreibung der Auswahl der Stichprobe**

#### **Studierende**

Im Wintersemester 2008/2009 nahmen an der Studie alle Studierenden, die für die Pflichtveranstaltung „Blockpraktikum Kinderheilkunde“ im sechsten klinischen Semester des Regelstudiengangs Medizin der Charité-Universitätsmedizin Berlin zugelassen waren, verpflichtend teil. Es wurden alle SKTs zur Auswertung heran gezogen, deren Studierende eine Multiple-Choice-Klausur (MC-Klausur) in den voran gegangenen Semestern absolviert hatten. Die MC-Klausur wurde am Ende des fünften klinischen Semesters im Rahmen des Kurses „Unterricht am Krankenbett“ in der Pädiatrie durch die Studierenden als notengebende Klausur absolviert.

Insgesamt konnten 283 gültige Klausuren des fünften klinischen Semesters ausgewertet werden. Von diesen Studierenden legten im folgenden Semester 232 den Skript-Konkordanz-Test ab. Die Unterschiede in den Kohorten ergaben sich durch Studierende, die nach dem fünften klinischen Semester nicht weiter studierten (Erasmus-Studierende, Aussetzer aus unterschiedlichen Gründen).

#### **Experten\_innen**

Für das Referenzpanel wurden 51 Experten\_innen der Kinderkliniken des Otto-Heubner-Centrums gewählt, die sich in verschiedenen Ausbildungsstufen befanden

---

(Assistenzärzte\_innen in Weiterbildung, Fachärzte\_innen, Oberärzte\_innen und Professoren\_innen). Es wurde Wert auf eine proportionale Verteilung der Experten\_innen entsprechend ihrem tatsächlichen zahlenmäßigen Anteil in den verschiedenen Ausbildungsstufen gelegt. Die Experten\_innen konnten in die folgenden Ausbildungsstufen unterteilt werden:

| <b>Ausbildungsstatus</b> | <b>Häufigkeit</b> | <b>Prozent</b> |
|--------------------------|-------------------|----------------|
| Assistenzarzt 3.-5. Jahr | 13                | 25,0           |
| Facharzt (ab 6. Jahr)    | 7                 | 13,5           |
| Oberarzt (ab 8. Jahr)    | 28                | 53,8           |
| Professor                | 3                 | 5,8            |

Tabelle Ausbildungsstatus der Experten\_innen

Einschlusskriterium für die Teilnahme an der Veranstaltung des Blockpraktikums der Kinderheilkunde war die fortdauernde Zugehörigkeit zum Lehrkörper des OHCs während des Semesters. Ausschlusskriterien konnten nicht definiert werden, da der Test das Denken der Studierenden überprüfen sollte. Aufgrund dessen sollten alle Ärzte\_innen die Fragen entsprechend ihrem Status beantworten können. Den Experten\_innen wurden in einer Schulung das Konzept und die Durchführungsprinzipien des Skript-Konkordanz-Tests und der Bezug zu den Lernzielen erläutert.

### **Multiple Choice-Klausur-Durchführung**

Die MC-Klausur wird am Ende des fünften klinischen Semesters als semesterübergreifende Klausur zusammen mit den Fächern Frauenheilkunde/Geburtshilfe (40 Fragen) und Humangenetik (20 Fragen), durchgeführt. Für die Kinderheilkunde müssen die Studierenden 40 Multiple Choice-Fragen mit Einfach-Auswahl (eine richtige Lösung) aus drei bis fünf Antwortmöglichkeiten wählen. Die Fragen beziehen sich auf den gesamten Lehrstoff, der in den Vorlesungen und dem Unterricht am Krankenbett vermittelt wurde. Die Studierenden wurden beaufsichtigt und befanden sich gemeinsam in einem Hörsaal. Sie haben insgesamt eine Stunde Zeit für die Beantwortung aller Fragen der Kinder- und Jugendmedizin.

### **Skript Konkordanz Test-Durchführung**

Die Studierenden hatten während der Praktikumswoche ausreichend Zeit, sich mit ihren Dozierenden auf die Durchführung der Prüfung und den praktischen Umgang mit dem für sie

---

neuen Konzept des SKT auseinanderzusetzen. In dieser Zeit wird das Beantworten der Fragen mit den Dozierenden des Blockpraktikums eingeübt, da es sich um den ersten Kontakt der Studierenden mit diesem Testformat handelt. Alle Dozierenden haben vor dem ersten Einsatz im Blockpraktikum eine Schulung für die Anwendung des SKTs durchlaufen.

Am Ende der Blockpraktikumswoche musste der SKT innerhalb einer Stunde von jedem Studierenden absolviert werden. Die Studierenden befanden sich während der Durchführung unter Aufsicht der für sie zuständigen Lehrkraft in der aus sechs Personen bestehenden Gruppe, mit der sie das Praktikum durchlaufen hatten. Jeder Studierende musste den Bogen allein ausfüllen, es gab keine fachlichen Hinweise durch die Dozierenden der Gruppe.

## **Blueprint Entwicklung Kinder- und Jugendmedizin**

### **Lernzielerstellung für den Unterricht am Krankenbett (UaK)**

Die Erstellung der Lernziele der Kinderheilkunde wurde in einem Konsensus-Verfahren durch die koordinierenden Oberärzte\_innen und Direktoren\_innen der elf Fachbereiche durchgeführt. Jeder Unterrichtseinheit (2 Unterrichtsstunden à 45 Minuten) wurden drei theoretische Lernziele zugeordnet, die mit einem vorab definierten Vokabular ausgearbeitet wurden. Nach der Erstellung wurden diese durch ein Review-Komitee bestehend aus den Semesterbeauftragten der jeweiligen Fachsemester überprüft und gegebenenfalls revidiert.

### **Lernzielerstellung Blockpraktikum**

Für das Blockpraktikum der Kinderheilkunde im sechsten Semester standen insgesamt fünfzehn Unterrichtsstunden zur Verfügung. Die Lernziele wurden diesen Anforderungen angepasst und an die Anforderungen des Blockpraktikums und des Gegenstandskatalogs, das inhaltlich aus der praktischen Teilnahme der Studierenden an der ärztlichen Stationsarbeit besteht, angepasst. Die Fragenitems des SK-Tests bezogen sich auf diese und die Lernziele des UaKs.

### **SKT-Item-Konstruktion**

Das Ziel der Item-Konstruktion ist es differentialdiagnostische, diagnostische und therapeutische Hypothesenbildung bei Studierenden in der Kinderheilkunde zu überprüfen. Als Item-Elemente werden, basierend auf dem Key-Feature Ansatz, Fragen eingesetzt, die einen sinnvollen Zusammenhang im Sinne einer Kontextualisierung der beschriebenen

---

klinischen Problem herstellen. Zu diesem Zweck wird ein Lösungsweg vorgegeben, dessen Sinnhaftigkeit in Bezug gesetzt wird zu einer Zusatzinformation. Die mögliche Lösung wird durch die Antwortenden auf einer Likertskala festgestellt. Konsekutive Fragen in einem Fragen-Item, also mehrere Fragen zu der gleichen Fallvignette, unterscheiden sich durch unterschiedliche Hypothesen und Zusatzinformationen und bauen nicht aufeinander auf. Die Folgefrage in einem Item ist nicht von den Informationen der vorhergehenden Frage abhängig. Items sind unabhängig voneinander.

Der Cueing-Effekt bezeichnet die Beantwortung einer MC-Frage durch die zu Prüfenden durch Erkennen der richtigen Antwort anhand der Antwortzusammenstellung, anstatt durch spontane Beantwortung aus dem eigenen Wissen heraus [71]. Zum Ausschluss eines Cueing-Effektes wurde auf eine gleichmäßige Verteilung der Antwortmöglichkeiten auf der gesamten Likertskala aller Fragenitems Wert gelegt [60].

### **Fragenerstellung MC-Klausur**

Die Erstellung der Klausurfragen der Multiple Choice Klausuren wird durch den Assessmentbereich der Charité organisiert. Die inhaltliche Grundlage stellt der beschriebene Blueprint für die studentische Ausbildung in der Kinder- und Jugendmedizin dar. Die Neuerstellung der Klausurfragen erfolgt unter Berücksichtigung eines für die Autoren\_innen bereit gestellten Kataloges von Qualitätskriterien (s. Anhang: „Übersicht: Qualitätskriterien für die SKT-Fragenentwicklung“). Anschließend werden die Fragen in den Reviewprozess übergeben. In den Reviewkomitees befindet sich ein interdisziplinäres Team von Experten\_innen, die die Fragen auf die Einhaltung der vorgegebenen Qualitätskriterien überprüfen. Anschließend werden die Fragen entweder zur Überarbeitung an die Autoren\_innen zurück gegeben oder erneut in den Reviewprozess eingebracht und nach abschließender Begutachtung in die Klausuren übernommen. Der Anteil an Neufragen in den Klausuren beträgt mindestens 30 Prozent an der Gesamtfragenmenge. Bei der Zusammenstellung der MC-Klausur wird der gleiche Verteilungsschlüssel (s. Tabelle Blueprint Kinder- und Jugendmedizin) eingesetzt, der auch bei der Zusammenstellung der SK-Fragen Berücksichtigung fand.

### **Fragenerstellung SKT**

Die Zusammenstellung der Prüfungsfragen basiert auf dem Blueprint des Lernzielkataloges für den studentischen Unterricht in der Kinderheilkunde (Uak und Blockpraktikum). Für die einzelnen Klausuren wurde von 2006 bis 2008 ein Fragenpool von insgesamt 150 Fragen von

vier erfahrenen Ärzten\_innen des OHCs erstellt, die zuvor eine mehrtägige Schulung zur Test- und Fragenerstellung mit den Entwicklern des SKTs durchlaufen hatten. In jeder Woche wurde eine neue Auswahl von Fragen nach dem oben genannten Schema zusammengestellt. Vor ihrem Einsatz wurden die Fragen einem Review durch das Experten\_innen-Panel, bestehend aus den Kurs-Dozierenden des aktuellen Semesters unterzogen.

Die Anteile der Fragen am Gesamtkatalog der Fragen, als auch in der Zusammenstellung der einzelnen Klausuren, wurden entsprechend den Unterrichtsanteilen der einzelnen Fachgebiete gewichtet (s. Tabelle Unterrichtsanteile Kinderheilkunde). Die erste Erprobung neu erstellter Fragen erfolgte durch die Dozierenden des Blockpraktikums und durch ärztliche Mitarbeiter\_innen des OHCs (Assistenzärzte\_innen, Fachärzte\_innen, Professoren\_innen). Die Fragen wurden im Hinblick auf inhaltliche Qualität, fachliche Relevanz und Fragenaufbau evaluiert (s. Übersicht Qualitätsanforderungen).

Die Fragen bezogen sich inhaltlich auf die gemeinsamen kognitiven Lernziele des fünften und sechsten klinischen Semesters des OHCs der Charité. Die Lernziele wurden zuvor in einem Konsensus Prozess durch die Lehrverantwortlichen der einzelnen Fachbereiche erstellt. Die Erstellung der Testaufgaben orientierte sich an den Vorgaben, die von den Entwicklern des SKT für die Entwicklung von Testfragen vorgeschlagen wurden [72]. Den Experten\_innen lag ein Checkliste vor, die die Qualitätskriterien für eine gute SKT-Frage nach den Empfehlungen von Charlin enthält (s. Tabelle Checkliste Qualitätskriterien). Die Fragen sollten typische und häufige klinische oder ambulante Fälle beschreiben. Die Prüfungsfragen rekrutierten sich aus den elf Themenbereichen des pädiatrischen Unterrichts am OHC und wurden auf die kognitiven Lernziele der Fachbereiche abgestimmt. Die Fragenanzahl wurde auf die Lehranteile der einzelnen Fachbereiche am Gesamtunterricht der Kinderheilkunde abgestimmt (s. Tabelle Lehranteile). Der Schwierigkeitsgrad der Fragen bezog sich inhaltlich weitgehend auf allgemeinpädiatrisches klinisches Basiswissen. Im nächsten Schritt wurden die wichtigsten klinischen Informationen, Fragen und Laboruntersuchungen definiert. Anhand dieser Daten wurden die Test-Items erstellt.

#### Blueprint Studentischer Unterricht Kinderheilkunde

| <b>Fachgebiet</b>  | <b>Lehranteil</b> | <b>Fragenanzahl</b> |
|--|-------------------|---------------------|
| 1. Allgemeine Pädiatrie (Infektiologie, Stoffwechselmedizin) | 20%               | 23,6 %              |
| 2. Pädiatrische Hämato-/Onkologie                            | 10%               | 14,6 %              |

|  |     |        |
|--|-----|--------|
| 3. Pädiatrische Endokrinologie         | 5%  | 11,4 % |
| 4. Pädiatrische Nephrologie            | 5%  | 3,3 %  |
| 5. Pädiatrische Neurologie             | 6%  | 7,3 %  |
| 6. Pädiatrische Pulmologie/Immunologie | 10% | 11,4 % |
| 7. Pädiatrische Gastroenterologie      | 3%  | 6,5 %  |
| 8. Kinderchirurgie                     | 9%  | 5,7 %  |
| 9. Kinder- u. Jugendpsychiatrie        | 7%  | 0,8 %  |
| 10. Kinderkardiologie                  | 5%  | 7,3 %  |
| 11. Neonatologie                       | 20% | 8,1 %  |

Tabelle: der Unterrichtsanteile Kinderheilkunde Stand Wintersemester 2008/2009

### Item-Anzahl - Reliabilität

Die Reliabilität einer Prüfungsmethode misst die Reproduzierbarkeit der Prüfungsergebnisse [68]. Die Reliabilität eines Testverfahrens wird durch die folgenden Faktoren beeinflusst:

1. Kontextspezifität
2. Anzahl der Testitems
3. Prüfungs-Teilnehmenden (Prüfer\_innen, Patienten\_innen )
4. Standardisierung

Van der Vleuten betont, dass in Bezug auf die oben genannten Kategorien für eine umfassende Überprüfung von Kompetenzen klinischen Denkens, die ausreichende Anzahl der Testitems und deren Kontextspezifität, als auch die inhaltliche Themenbreite der Fragen ein hervor gehobene Rolle spielen [34].

Der SK-Test wird als summative Prüfung eingesetzt und stellt die Grundlage für die Vergabe des Leistungsnachweises und die Benotung der Studierenden in dem Blockpraktikum dar.

Die Vorgabe der Prüfungskommission der Kinderheilkunde war, dass die Fragen eines Testblocks durch die Studierenden innerhalb einer Stunde zu beantworten sein sollten. Generalisierbarkeitsstudien ergaben eine Mindestanzahl von 60 Fragen, um ein Alpha von 0,80 zu erreichen [13]. Unter Berücksichtigung der vorhergehenden Studien durch Charlin et al und einer angestrebten Testzeit von 60 Minuten wurde die Mindestanzahl auf 45 Fragen pro SKT durch die Lehrkoordinatoren\_innen des OHCs festgelegt [13,73]. Damit wurde theoretisch ein Cronbach-alpha-Koeffizient von 0,70 angestrebt.

---

## Referenzpanel

Das Antwortverhalten medizinischer Experten\_innen in typischen klinischen Situationen weist eine große Variationsbreite auf. Wie in der Skripttheorie beschrieben, unterscheiden sich die Wege der Informationssammlung und deren Verarbeitung bei Kliniker\_innen verschiedener Erfahrungsstufen signifikant.

Für die Bewertung der abgegebenen Antworten muss ein Experten\_innen-Panel zusammengestellt werden, das auf die inhaltlichen Bedürfnisse der zu testenden Wissensbereiche und Kompetenzen bei den Test-Teilnehmenden abgestimmt ist. Damit wird eine ausreichende Variationsbreite im Antwortverhalten gewährleistet [73]. Es wurden mehrheitlich Assistenzärzte\_innen und in abnehmender Zahl Fachärzte\_innen, sowie einige Oberärzte\_innen und Professoren\_innen in das Experten\_innen-Panel aufgenommen und um Beantwortung der Fragen gebeten. Da Experten\_innen-Antworten in diesem Testumfeld in der Regel eine bestimmte Variationsbreite aufweisen, ist deren Anzahl in dem Referenzpanel ein wichtiges Kriterium für die Antwortbreite der einzelnen Fragen. [74]. Brailovsky und Charlin konnten zeigen, dass diese These auch für den Skript-Konkordanztest zutrifft [12]. Die Punktzahl, die für jedes Item vergeben wird, resultiert aus Konkordanz, also der Übereinstimmung, der studentischen Antwort mit den Antworten der Experten\_innen. Als Antwortskala wurde für den SKT eine Likert-Skala gewählt. Bewertet wird die Anzahl der Antworten, die von den Experten\_innen pro Punkt auf der Likertskala vergeben wurden. Die Experten\_innen-Antworten für die Einzelpunkte auf der Likertskala wurden summiert. Punkte werden im Skriptkonkordanztest für Konkordanz, also die Übereinstimmung der Antworten der Studierenden mit den Antworten der Experten\_innen vergeben. Der Wert auf der Skala für ein Fragenitem, auf dem die absolute Mehrheit der Experten\_innen ihre Stimme abgegeben haben, wird für Studierende, die diesen in der Abstimmung gewählt haben, mit einem Punkt bewertet. Studierende erhalten für die Wahl von Skalenwerten, für die sich die verbliebene Minderheit der Experten entschieden hat anteilig Punkte. Diese werden anteilig der Anzahl der Experten\_innen vergeben, die sich für den jeweiligen Punkt entschieden haben. 35 von 50 Experten haben beispielsweise die +2 gewählt. Also bekommen Studierende, die die +2 wählen einen ganzen Punkt. Die verbliebenen 15 Experten\_innen haben die +1 gewählt. Jeder Studierenden, der sich für diesen Wert entscheidet an den werden  $15/50$ , also  $3/10$  Punkte vergeben. Antworten, die mit von keinem Experten\_innen ausgewählt wurden, werden mit der Punktzahl null bewertet.

---

In Bezug auf die Auswertung der Antworten und Ausschluss nicht valider Antworten wurde folgendes Verfahren fest gelegt: Antworten, die eine zu geringe Varianz in der Antwortbreite aufwiesen, daher alle Experten\_innen haben die gleiche Antwort gewählt, Antworten, die aus Sicht der Experten\_innen inhaltlich nicht angemessen waren, zu schwierig und vor allem nicht eindeutig formuliert waren, wurden aus dem Fragenpool entfernt oder einer Bearbeitung übergeben.

### **Antwort-Auswertung**

Die auf der Likert-Skala abgegebenen Antworten der Experten\_innen und die der Studierenden können durch unterschiedliche Methoden ausgewertet werden. Vergleichende Untersuchungen von Charlin zu dem Antwortverhalten von Experten\_innen innerhalb eines Panels haben gezeigt, dass das Antwortverhalten der Kliniker\_innen in den verschiedenen Methoden insofern voneinander abweicht, als dass die Varianz zwischen den Antworten durch einen Konsensus ersetzt wird, sobald diese gebeten werden, eine gemeinsame „Richtige“ Antwort zu ermitteln [16]. Für die Darstellung der klinischen Variationsbreite im Antwortverhalten der Experten\_innen wurde aufgrund dessen die Aggregat-Scoring-Methode als Antwortmessmethode untersucht. Die Antworten werden auf Likert-Skalen abgegeben. Die Methode gewichtet die Antworten der Experten\_innen. Die Gewichtung entsteht durch die Summe der Antworten der Experten\_innen, die sich für die gleiche Antwort entschieden haben. Anschließend wird die Anzahl der gewählten Lösungsmöglichkeiten durch die Anzahl der Experten\_innen geteilt werden, die diese gewählt haben [18]. Die Adaptation dieser Methode an den Skript Konkordanz Test hat sich als sicher erwiesen, um ein reliables und valides Ergebnis zu erzielen.

Die Anzahl von 15 bis 20 Panel-Mitglieder stellt eine stabile Größe für das Erreichen eines validen Testergebnisses dar. Hinsichtlich der Korrelation der Einzelergebnisse mit dem Gesamtergebnis, der Reliabilität des Testergebnisses und der Variabilität der Standardabweichungen zwischen den Ergebnissen können somit Werte erreicht werden, die für notengebende Prüfungen erforderlich sind [73].

### **Durchführung der Prüfung**

Die Studierenden hatten während der Unterrichtswoche Zugriff auf den Gesamtkatalog der 150 Prüfungsfragen und wurden angehalten, die zugrundeliegenden Problemstellungen mit ihren Dozierenden während des Unterrichts zu thematisieren. Es wurden keine Lösungsvorschläge veröffentlicht. Missverständlich formulierte oder nicht dem



---

Ausbildungsstand der Studierenden angemessene Fragen wurden aus dem Fragenkatalog entfernt oder umformuliert.

### **Methodik der Testanalyse**

Die von den Studierenden am Ende jeder Woche ausgefüllten Testatbögen wurden durch das Lehrsekretariat des OHCs ausgewertet. Die Antworten wurden in eine zu diesem Zweck präparierte Exceltabelle eingegeben, die mit den Ergebnissen des Experten\_innen-Panels synchronisiert wurde. Somit konnte aus dem Antwortverhalten der Experten\_innen die Konkordanz der Antworten der Studierenden und die entsprechenden Punkt-, und Teilpunktzahlen für die Studierenden ermittelt werden. (Tabelle Antwortauswertung Dozierenden)

### **Definition der Antwortschlüssel**

Die Antworten der Experten\_innen des Panels wurden in einer Excel-Tabelle erfasst. Anhand der Anzahl der Experten\_innen, wurde der Antwortschlüssel für jede Lösungsmöglichkeit nach dem folgenden Schema errechnet:

Ein Punkt wurde vergeben, wenn die Mehrheit der Experten\_innen sich für den Punkt entschieden hatte. Für Antworten, für die sich nur ein Teil der Experten\_innen entschieden haben, wurden anteilig Punkte vergeben (Gesamtzahl Experten\_innen geteilt durch Anzahl der Experten\_innen, die sich für den Punkt entschieden haben).

### **Berechnung der Probengröße**

Die zentrale Forschungsfrage der Studie war, die Aussagekraft des SKTs als valides Testinstrument im studentischen Unterricht der Kinderheilkunde zu beweisen, belegt durch die Differenzierung zwischen den verschiedenen Stufen der Expertise der Studierenden und Dozierenden. Ausgehend von einem Alpha-Fehler von 0,05 und einer Power von 0,90 wurde eine erforderliche Probengröße von mindestens 71 Testteilnehmenden  $n$  als ausreichend berechnet, um mindestens eine Korrelation von 0,4 zwischen der MC-Klausur und dem SKT zu messen. Damit konnte eine Semesterkohorte ( $n=300$ ) unter Berücksichtigung der voraussehbaren Ausschlussfaktoren und Ausfälle von Teilnehmenden, wie Nichtteilnahme an einem der beiden Tests, als ausreichend betrachtet werden. Die Berechnung erfolgte unter Zuhilfenahme der Software Gpower (V. 3.1.3).

---

## Statistische Analyse - Methoden

Im folgenden Teil werden bezogen auf die einzelnen Forschungsfragen die jeweiligen Ergebnisse dargestellt. Die Daten wurden mit der Tabellenkalkulation Microsoft Office Excel (2003) gesammelt und mittels der Statistiksoftware SPSS ausgewertet (Version 16 - 18). Die Ergebnisse wurden als statistisch signifikant angesehen, wenn die p-Werte kleiner 0,05 waren.

### **Frage1: Ist der SKT ein valides Testinstrument für den Einsatz im studentischen Unterricht in der klassischen Medizin am Beispiel der Kinderheilkunde?**

Die Feststellung und Beschreibung der internen Konsistenz eines Tests steht im Mittelpunkt des Nachweises der Evidenz der durch den Test erhobenen Daten. Nach Downing setzt sich die Konstruktvalidität eines Testinstruments aus verschiedenen Typen von Evidenzen zusammen. An erster Stelle wurde die Reliabilität untersucht. Nach Schuwirth repräsentiert die Reliabilität die Verallgemeinerbarkeit und Übertragbarkeit der Testergebnisse, da die Testfragen nur einen Ausschnitt aller möglicherweise zu stellenden Fragen repräsentieren [60].

Der Cronbach alpha Koeffizient wurde als Messgröße für die 45 Fragen der Einzeltest verwendet. Der erste Cronbach alpha Wert wurde für die Mediane aller SKT-Fragen berechnet. Negativ korrelierte Items wurden durch Messung der Interitem-Korrelationen ausgeschlossen. Die Interitem-Korrelationen wurden nochmals für jede einzelne Woche ermittelt.

### **Frage 2: Inwieweit unterscheidet sich die Kompetenz der Studierenden des Blockpraktikums (6. Klinisches Semester) im kritischen klinischen Denken in der klassischen Medizin am Beispiel der Kinderheilkunde von der Kompetenz der Experten\_innen (Ärzte/ Fachärzte/ Professoren\_innen)?**

Die Cronbach's Alpha Koeffizienten wurden für jede Woche gesondert berechnet. Gruppen mit negativen Korrelationen wurden aus der Berechnung ausgeschlossen. Mit den verbleibenden Daten wurde die Berechnung optimiert. Für die Ergebnisse der Studierenden und der Dozierenden wurde der Pearson-Korrelationskoeffizient gebildet. Gemessen werden konnten die Unterschiede zwischen den Gruppen.

### **Frage 3: Gibt es eine Korrelation zwischen der Leistungsfähigkeit der Dozierenden im klinischen Unterricht am Krankenbett und dem den Ergebnissen der Studierenden im SKT?**

Die Gruppe der Dozierenden wurde nach den Jahren der Berufserfahrung aufgeteilt. Es konnten vier Gruppen identifiziert werden. Dozierenden (Assistenzärzte) mit drei bis fünf

---

Jahren Berufserfahrung, Dozierenden mit Facharztniveau (mindestens sechs Jahre Berufserfahrung) und Dozierenden mit Oberarztniveau (länger als vier Jahre Facharzt).

Die Mittelwerte der Gruppen jedes einzelnen Dozierenden wurden Z- standardisiert und mittels des Pearson-Korrelationskoeffizienten miteinander in Beziehung gesetzt.

#### **Frage 4: Besteht eine Korrelation der MC-Klausuren-Ergebnisse zu den Ergebnissen der Studierenden im SKT?**

Nach Downing gehört zum Beweis der internen Validität eines Tests der Beleg verschiedener Strukturprinzipien. Zur Absicherung dieser kann zusätzlich der Vergleich durch Korrelation mit ähnlichen Variablen zur Bestätigung der Validitätsthese heran gezogen werden oder um diese abzulehnen. Die Ergebnisse der MC-Klausuren und der SK-Tests jedes einzelnen Studierenden wurden miteinander korreliert.

## **Ergebnisse**

Der Bereich ist aufgegliedert in die demographischen Daten der Dozierenden und Studierenden und anschließend daran der Korrelationskoeffizienten zwischen den beiden Testformaten und den Dozierenden in Bezug auf die Gruppen.

### **Ergebnisse des Skript-Konkordanz-Tests Analyse – Studierende**

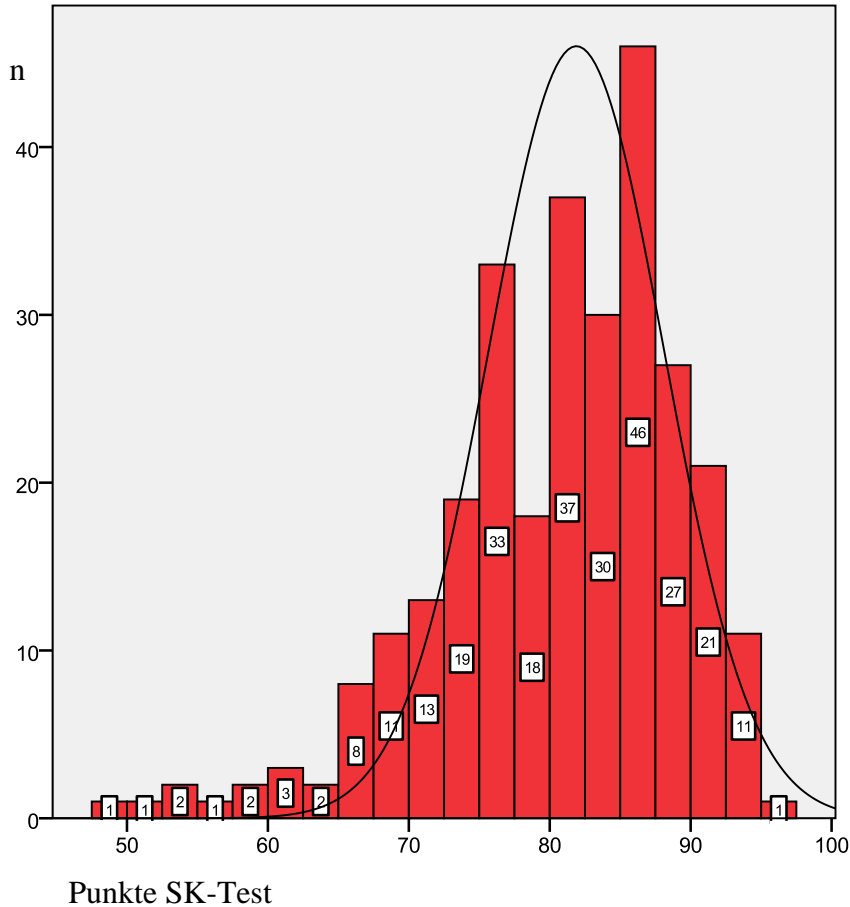
#### **Demographische Daten und Ergebnisse**

287 Studierenden des fünften klinischen Semesters im Regelstudiengang der Charité – Universitätsmedizin Berlin hatten die fächerübergreifende Abschlussklausur Humangenetik, Gynäkologie, Pädiatrie absolviert. In der vorliegenden Studie werden nur die Daten des Pädiatrieteils der Klausur mit einbezogen. Es konnten 232 Teilnehmenden mit in die Wertung einbezogen werden, die an beiden Testformaten teilgenommen hatten. 64 Prozent der Studierenden in beiden Formaten waren weiblich, 36 Prozent männlichen Geschlechts. Das Alter der Studierenden lag im Durchschnitt bei 27 Jahren (SD = 3,8, Spanne = 24 – 46,5 Jahre).

Die Auswertung des SKT ergab einen Mittelwert für alle Studierenden von 80,45 von maximal 100 Prozent der erreichbaren Punkte (Standardabweichung 8,4). Der Kolmogorov-Smirnov-Test zeigt eine Normalverteilung der Ergebnisse der Studierenden bei einer zweiseitigen Signifikanz von  $p=0,007$ .

|     | N   | Mittelwert | Standardabweichung | Minimum | Maximum |
|-----|-----|------------|--------------------|---------|---------|
| SKT | 287 | 80,44      | 8,415              | 49      | 95      |

Tabelle Gesamtergebnis SK-Test im Wintersemester 2008/2009



### Reliabilität der SKT- Gesamtergebnisse der Studierenden

Fragen mit negativer Korrelation wurden aus der Wertung ausgeschlossen. Für den SK-Test errechnet sich ein optimierter Cronbach-Alpha Wert von 0,79. Die einzelnen Fragen wiesen einen stabilen Einfluss auf die Gesamtreliabilität der Testergebnisse auf.

### Reliabilitätsanalyse nach Wochen: Ergebnisse des SKT

Es wurden die Cronbach Alpha-Werte für jede einzelne Testwoche wie folgt ermittelt: Wochen mit einer Varianz von Null oder kleiner wurden aus der Auswertung entfernt.

| Woche    | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   | 12    | 13   | 14   |
|----------|------|------|------|------|------|------|------|------|------|------|------|-------|------|------|
| $\alpha$ | 0,78 | 0,64 | 0,28 | 0,49 | 0,72 | 0,26 | 0,68 | 0,69 | 0,08 | 0,77 | 0,71 | -0,58 | 0,75 | 0,34 |

Tabelle Intertitemkorrelation SKT

---

Die Wochen mit niedrigen Varianzwerten weisen auf eine Kombination von Items hin, die eine geringe Interpretationsbreite aufweisen. Eine andere Möglichkeit ist eine Zunahme der Konkordanz durch eine verstärkte Abstimmung der Gruppenantworten der Studierenden und oder des Dozierenden mit der Gruppe.

### Item-Reliabilitätsanalyse des SKT

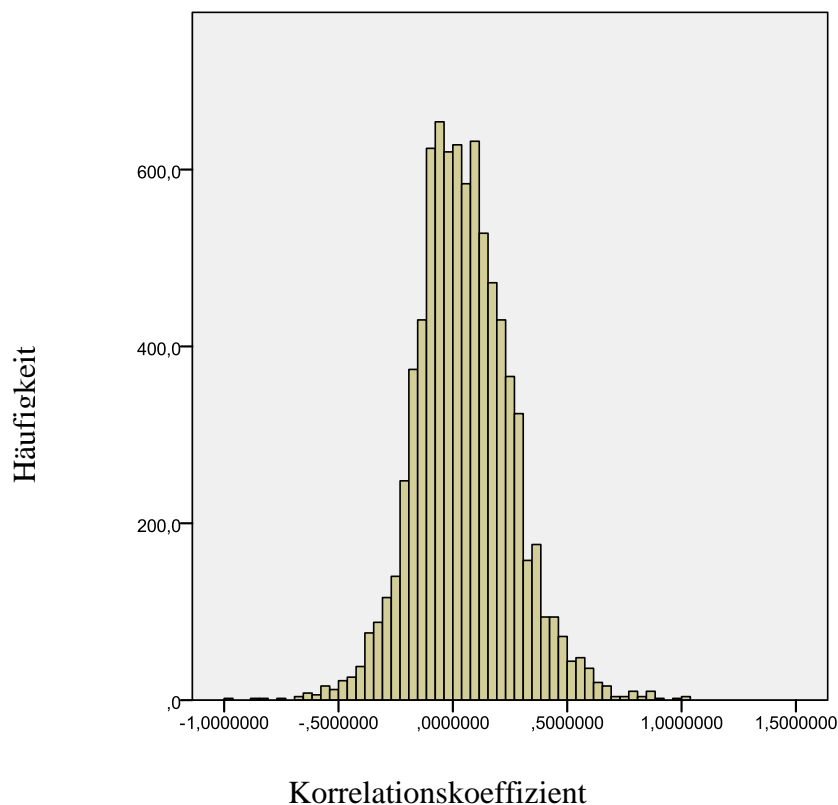
Der Cronbach's Alpha Koeffizient wurde nach der folgenden Formel berechnet:

$$\alpha = \frac{N * r}{+(N-1)*r}$$

Anschließend wurde die mittlere Inter-Itemkorrelation für SKT-Items aus der Datenstichprobe geschätzt. Als Grundlage diente die Matrix der erzielten Itempunkte pro Testteilnehmenden.

Die Korrelationen werden unter paarweisem Ausschluss der Missings ermittelt. Kommt eine Fragenkombination nicht vor, wurde diese aus der Analyse ausgeschlossen (Missing).

Es ergeben sich 8272 Paarungen von Items innerhalb des gesamten Datensatzes, so dass die mittlere Interitemkorrelation aus 8272 Werten geschätzt werden kann. Daraus ergibt sich ein Mittelwert von 0,5 (Standardabweichung 0,21) für die folgende Verteilung:



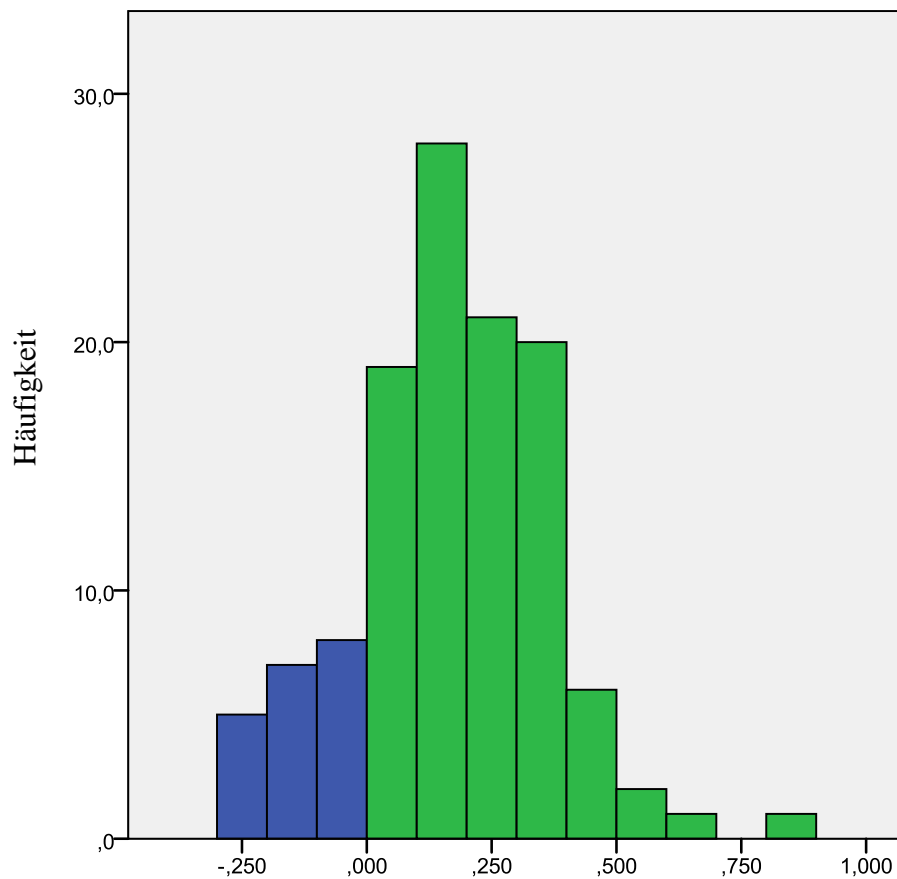
Die mittlere Interitemkorrelation weist einen Wert von 0.05 auf, wenn man einen Test aus 45 derartigen Items zusammenstellen würde, ließe sich ein Alpha von:

$$45 \cdot 0.0497 / (1 + 44 \cdot 0.0497) = 0.70 \text{ erwarten}$$

### Operationalisierung über die Trennschärfe

Die Trennschärfe der einzelnen Items wurde durch die Korrelation des Itempunktwertes mit dem Gesamtpunkt看wert des restlichen Tests (Gesamtwert minus Itemwert) ermittelt. Da die Items meist in verschiedenen Sets (Wochen) und von verschiedenen Personen vorliegen, werden sowohl Itempunktwert als auch Gesamtesttestpunkt看wert innerhalb der Sets (Wochen) vor der Korrelationsbildung z-standardisiert.

Tabelle Trennschärfe Z-Transformiert



Legende:

- = ≤ 0
- = > 0

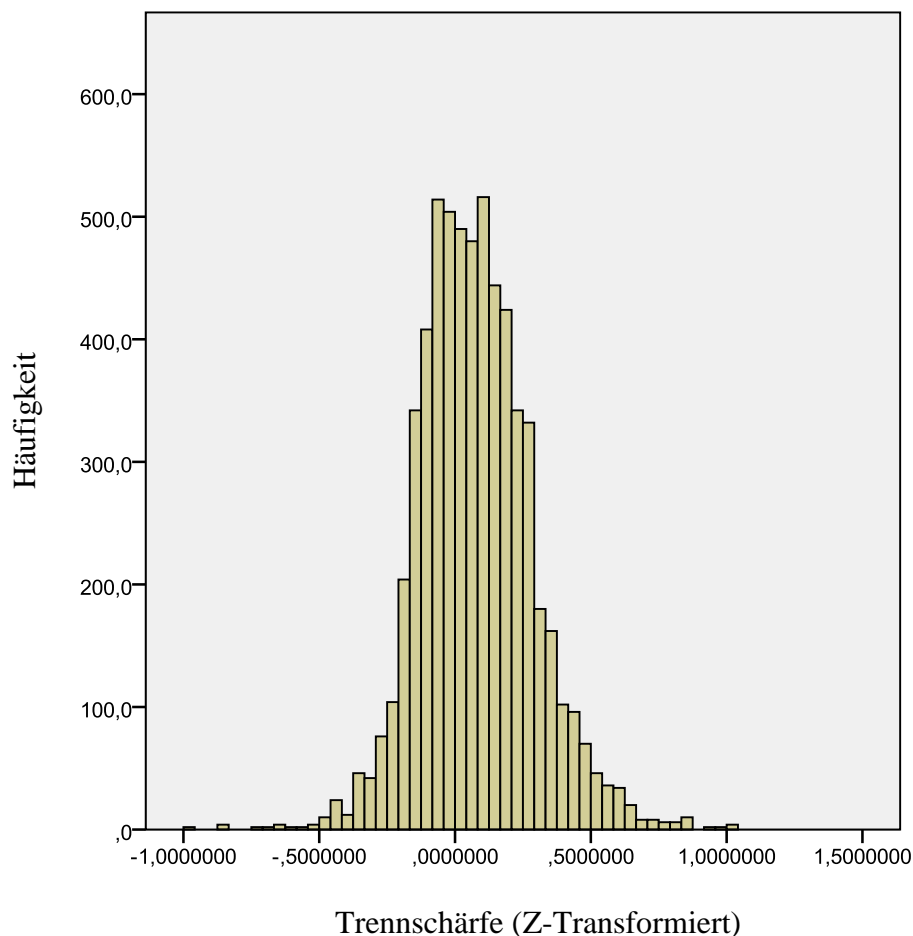
Trennschärfe (Z-Transformiert)

---

Unter Ausschluss der Items mit negativer Korrelation in die mittlere Interitemkorrelation ergab sich eine höhere mittlere Interitemkorrelation von 0.076. Mit den optimierten Items lässt sich ein Alpha von

$$45 \cdot 0.076 / (1 + 44 \cdot 0.076) = 0.79 \text{ errechnen.}$$

Tabelle Optimierte Interitemkorrelation



## Ergebnisse der MC-Klausur

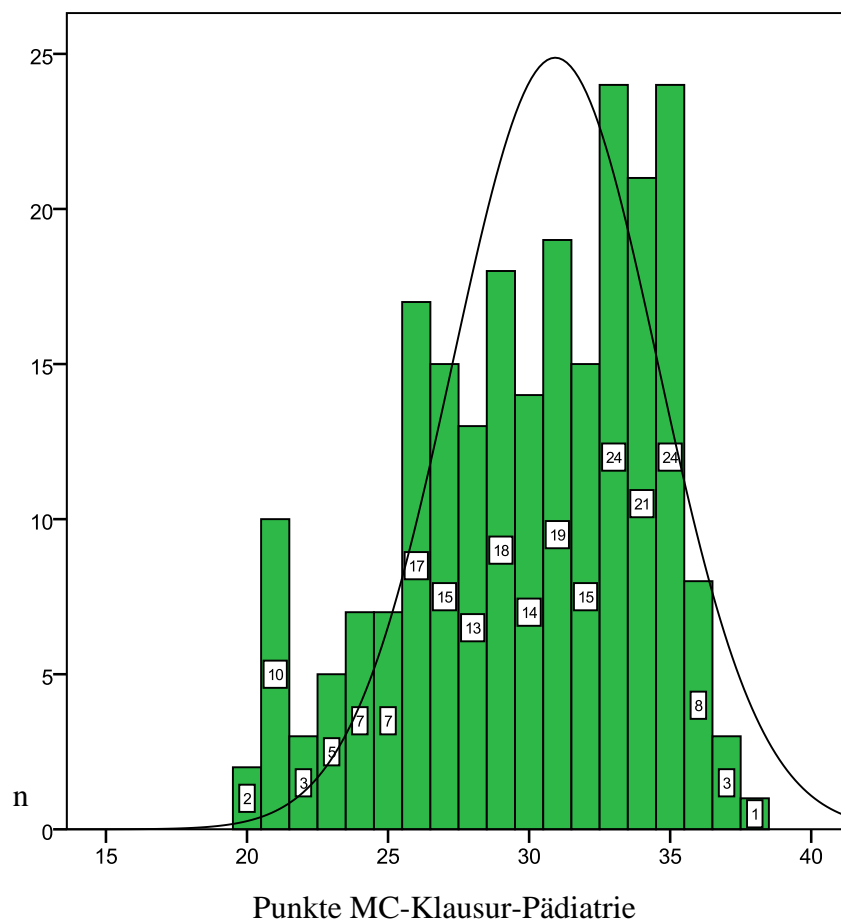
64 Prozent der teilnehmenden Studierenden an der MC-Klausur im Sommersemester 2008 waren weiblich, 36 Prozent männlich. Die MC-Klausur bestand aus 40 Fragen-Items, von denen eines aufgrund eines Einspruches der Studierenden aus der Wertung genommen werden musste. Von den 287 Teilnehmenden des fünften klinischen Semesters wurden die 232 Studierenden mit in die Wertung einbezogen, die auch das Blockpraktikum im Folgesemester mit dem Skript-Konkordanz-Test abgeschlossen haben.

## Reliabilität der MC-Klausur

Die Auswertung der Klausur ergab einen Mittelwert von 29,7 von maximal 39 Punkten für alle Studierenden (Standardabweichung 4,7). Für diese Werte lässt sich ein Cronbach's-Alpha-Wert von 0,77 errechnen. Die einzelnen Fragen wiesen einen stabilen Einfluss auf die Gesamtreliabilität der Testergebnisse auf. Der Kolmogorov-Smirnov-Test zeigt eine Normalverteilung der Ergebnisse der Studierenden und erweist sich mit einem Wert von 0,007 als signifikant. Die Ergebnisse der MC-Klausur basieren somit auf einer Normalverteilung der Punkte.

Tabelle Ergebnisse MC-Klausur Sommersemester 2008

|                          | N   | Mittelwert | Standardabweichung | Minimum | Maximum |
|--------------------------|-----|------------|--------------------|---------|---------|
| Punkte MC-Klausur SoSe08 | 232 | 29,7       | 4,6                | 12      | 38      |





## Vergleich der MC-Klausurergebnisse mit dem SK-Test

Insgesamt nahmen 232 Studierende an beiden Klausuren teil. Der Korrelationskoeffizient nach Pearson weist einseitig keinen signifikanten Zusammenhang zwischen der Höhe der Ergebnisse der Studierenden in der Multiple Choice Klausur und deren Ergebnissen in dem Skript Konkordanz Test ein Semester später auf. Die lineare Regressionsanalyse liegt mit einem  $r=0,11$  im statistisch nicht signifikanten Bereich.

Tabelle Lineare Regression MC-Klausur – SKT

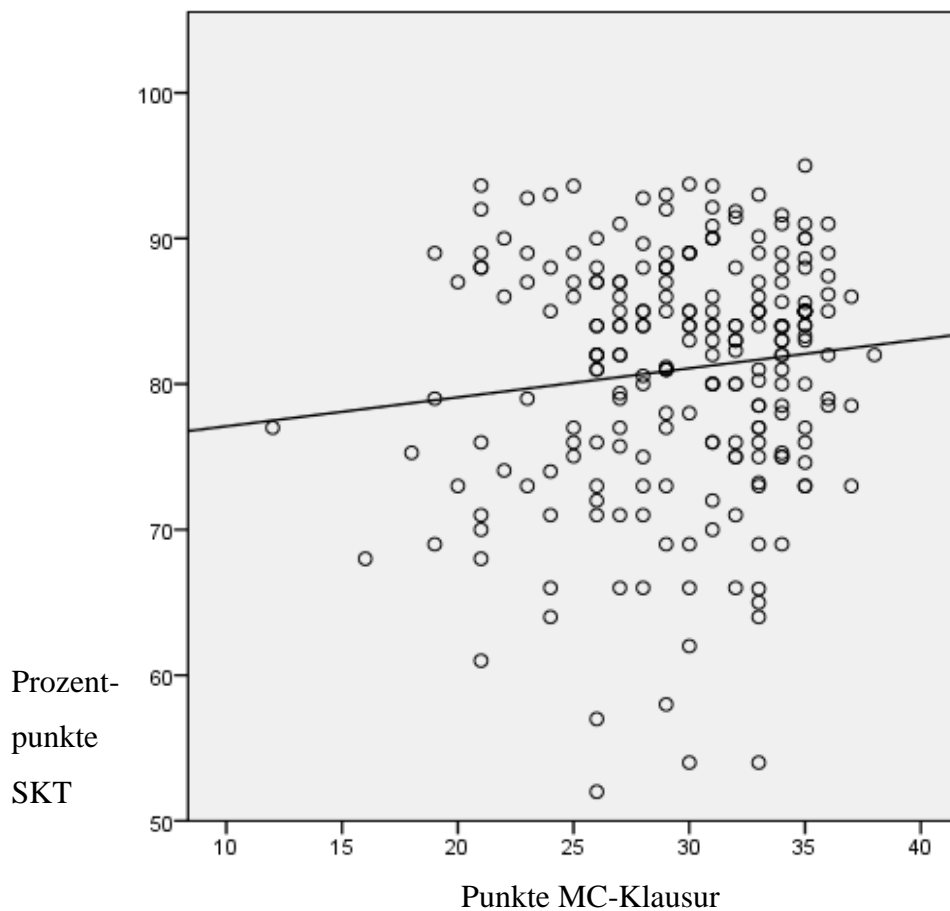


Tabelle Korrelationen SKT-Test – MC-Klausur nach Pearson

|                          | Punkte MC-Klausur SoSe2008 | Punkte SKT WiSe 2008/09 |
|--------------------------|----------------------------|-------------------------|
| Korrelation nach Pearson | 1                          | $r = 0.11$              |
| Signifikanz (1-seitig)   |                            | $p = 0.04$              |

---

## Vergleich der Ergebnisse getrennt nach Geschlecht

Um den Grad der Testgüte im Vergleich mit anderen Testformaten darstellen zu können, wurden die Testergebnisse beider Testformate nach Geschlechtern getrennt bewertet. Der Median der männlichen Studierenden lag in der MC-Klausur bei 30,0 Punkten und im SK-Test bei 79,4 Prozent. Die weiblichen Studierenden erreichten einen Median von 31 Punkten in der MC-Klausur und 83,0 Prozent im SK-Test.

## Einfluss der Dozierenden auf das Gruppenergebnis

### Ausbildungsstatus und Geschlechterverteilung

Es nahmen insgesamt 50 Dozierende aus den neun Fachrichtungen des Otto Heubner Centrums für Kinder- und Jugendmedizin an dem Test teil. Von den Dozierenden wurden alle 150 Fragen beantwortet. Die Dozierenden waren gleichzeitig Seminargruppenleiter im Blockpraktikum.

Die Dozierenden des Praktikums rekrutierten sich mehrheitlich aus Oberärzte\_innen der Kliniken. Ein Viertel der Dozierenden wird durch Assistenzärzte im 3. bis 5. Weiterbildungsjahr repräsentiert, 13,5 Prozent der Dozierenden werden von Fach-Ärzten\_innen ohne Oberarztstatus repräsentiert und nur 5,8 Prozent der Dozierenden werden von Professoren\_innen repräsentiert. Weiblich waren 46 Prozent der Dozierenden, männlich 54 Prozent.

Tabelle Dozierenden Ausbildungs-Status

| Ausbildungsstatus        | Häufigkeit | Prozent |
|--------------------------|------------|---------|
| Assistenzarzt 3.-5. Jahr | 12         | 25,0    |
| Facharzt (ab 6. Jahr)    | 7          | 13,5    |
| Oberarzt (ab 8. Jahr)    | 28         | 53,8    |
| Professor                | 3          | 5,8     |

## Ergebnisse des Skript Konkordanz Testes – Dozierenden

Die Dozierenden erreichten im SK-Test einen Mittelwert von 82,35 Prozent. In Bezug auf die geschlechtsgetrennte Verteilung der Ergebnisse erreichten die weiblichen Teilnehmenden einen Mittelwert von 84,63 Prozent (SD 5,8). Die männlichen Teilnehmenden erreichen einen Mittelwert von 80,00 Prozent (SD 8,0). Die Werte unterscheiden sich bei einem

angenommenen  $p < 0,05$  signifikant voneinander ( $r = 0,027$ ). In Bezug auf die Erfahrung der Dozierenden und die Geschlechterzugehörigkeit zeigt sich keine signifikante Zunahme der erreichten Punktzahlen mit Zunahme des Erfahrungsgrades der Dozierenden innerhalb der Gruppe der Dozierenden (s. Tabelle Ergebnisse der Dozierenden).

Tabelle SKT Ergebnisse der Dozierenden nach Geschlecht

| <b>Gruppe</b>   | <b>N</b> | <b>Mittelwert</b> | <b>Median</b> | <b>Std.-Ab.</b> | <b>Min.</b> | <b>Max.</b> |
|-----------------|----------|-------------------|---------------|-----------------|-------------|-------------|
| <b>Weiblich</b> | 23       | 84,6              | 84,0          | 5,8             | 73          | 95          |
| <b>Männlich</b> | 27       | 80,0              | 79,0          | 8,8             | 63          | 96          |
| <b>Gesamt</b>   | 50       | 82,1              | 82,5          | 7,4             | 63          | 96          |

Zwischen der Gruppe der Assistenzärzte und Oberärzte\_innen zeigte sich keine signifikante Zunahme der Ergebnisse. Die Gruppe der Professoren\_innen nahm nur in geringer Zahl ( $n=3$ ) an den Tests teil, fiel jedoch in ihren Ergebnissen gegenüber den Oberärzte\_innen ab.

Tabelle SKT-Ergebnisse der Dozierenden

| <b>Gruppe</b>            | <b>N</b> | <b>Mittelwert</b> | <b>Bereich</b> | <b>Std.-Ab.</b> |
|--------------------------|----------|-------------------|----------------|-----------------|
| Assistenzarzt 3.-5. Jahr | 12       | 81,15             | 67-95          | 7,01            |
| Facharzt (ab 6. Jahr)    | 7        | 79,43             | 72-86          | 5,09            |
| Oberarzt (ab 8. Jahr)    | 28       | 84,48             | 67-96          | 7,03            |
| Professor                | 3        | 71,33             | 63-78          | 7,63            |

Der Test unterscheidet somit zwischen den Experten\_innen-Stufen der Studierenden und Ärzten\_innen in verschiedenen Weiterbildungsstadien, jedoch nicht signifikant zwischen den verschiedenen Ausbildungsstadien der Ärzte untereinander.

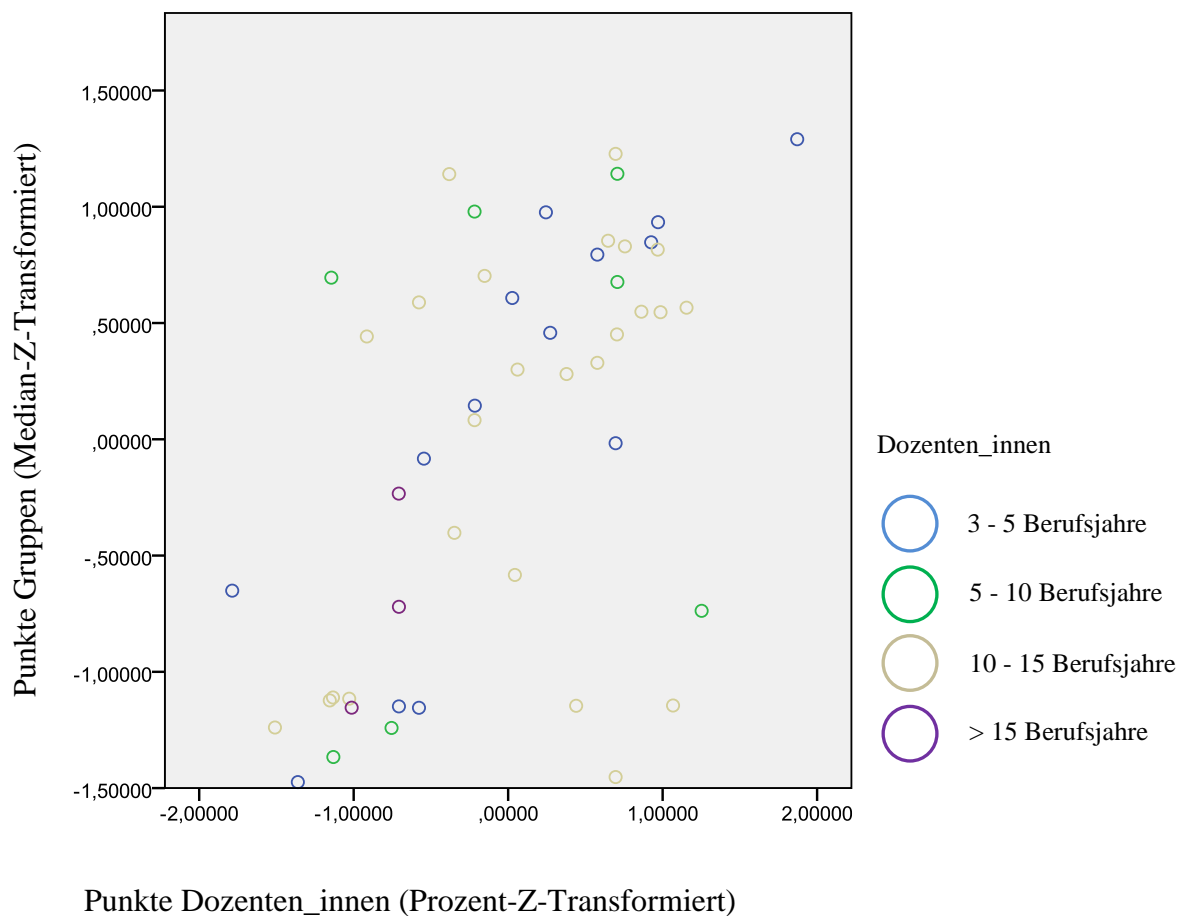
### **Berufserfahrung Dozierenden in Bezug auf die Gruppenergebnisse**

Es wurden für diese Untersuchung die Ergebnisse der Dozierenden (in Prozent) in Bezug gesetzt zum Mittelwert der Ergebnisse (in Punkten) der von ihnen betreuten Gruppen. Es

wurden die Ergebnisse von 47 Dozierenden mit in die Auswertung einbezogen, von drei Dozierenden lagen keine Gruppenergebnisse vor. Der Korrelationskoeffizient nach Pearson weist einen Wert von  $r = 0,48$  auf, der bei einem angenommenen Signifikanzniveau von  $p < 0,05$  mit  $0,007$  zweiseitig signifikant ist.

Um eine Verzerrung durch den möglicherweise unterschiedlichen Schwierigkeitsgrade der in den jeweiligen Wochen eingesetzten Tests auf das Gesamtergebnis auszuschließen, wurden die Gruppenmittelwerte und die Ergebnisse der Studierenden nach Z-transformiert. Die über alle Wochen ermittelten Z-Werte wurden miteinander korreliert und konnten die Korrelation mit einem  $r = 0,55$  ( $p < 0,001$ ) bestätigen. Damit wird die in vorhergehenden Untersuchungen des SKT im Bereich der postgraduierten Ausbildung unterstellte These unterstützt, dass unabhängig vom Schwierigkeitsgrad der Fragenitems das Abschneiden der Dozierenden im Test im Zusammenhang steht zu dem Niveau der Ergebnisse der Studierenden der Gruppe des Dozierenden (24, 76).

Tabelle Korrelation Punkteverteilung Dozierenden – Gruppen SKT

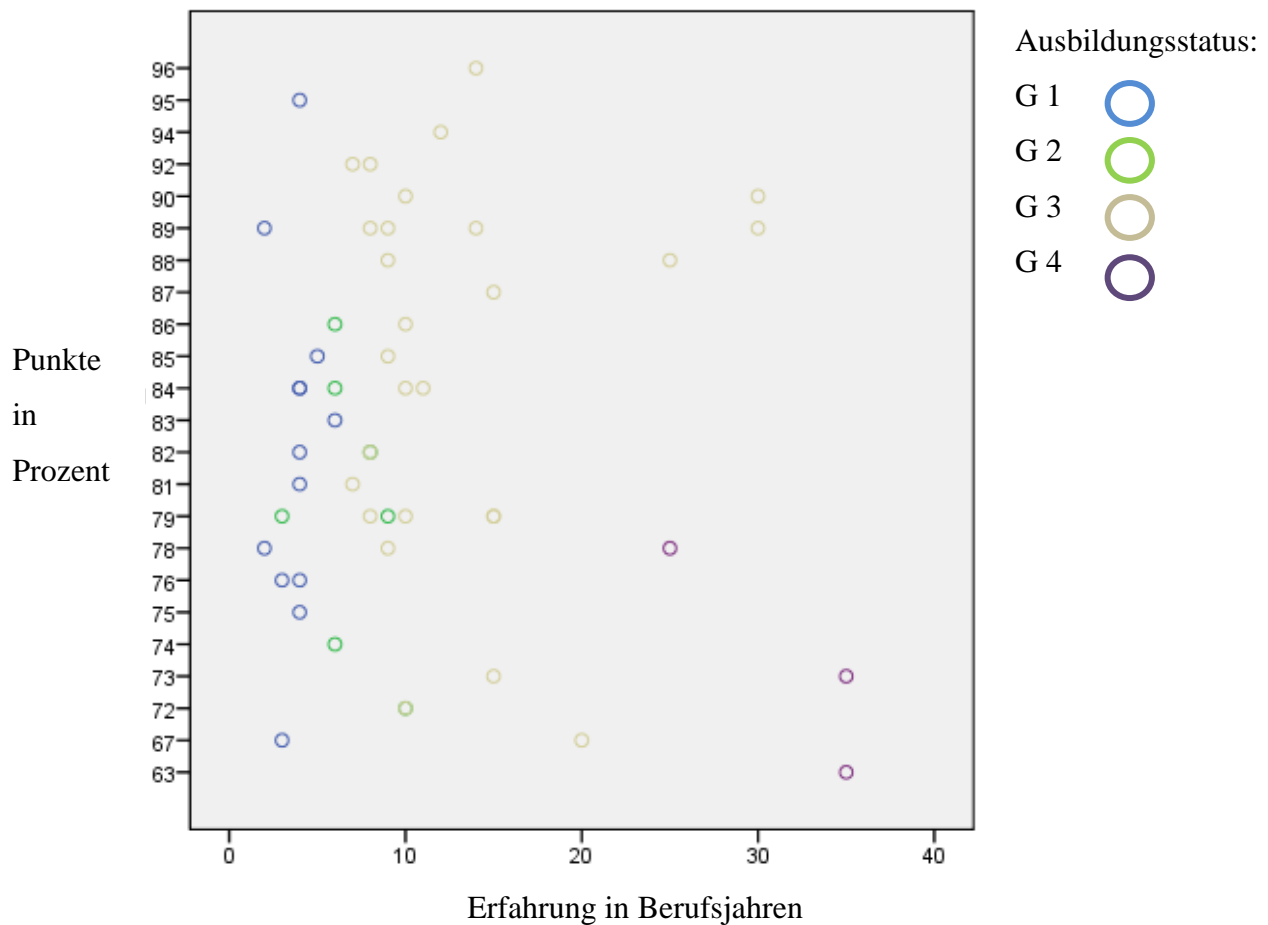


---

## Zusammenfassung

Die MC-Klausur enthielt 39 von 40 Fragen, die in die Auswertung gelangten. Die 232 in die Wertung gelangten Studierenden erreichten in der MC-Klausur im Mittel 29,6 von 40 möglichen Punkten (Maximal: 38 Punkte, Minimal 12 Punkte, Standardabweichung 4,7). Die Reliabilität der Klausur erreichte ein Cronbach's Alpha von 0,76. Die Studierenden erreichten im SK-Test im Mittel 80,4 von 100 erreichbaren Punkten (Maximal: 95 Punkte, Minimal 49 Punkte, Standardabweichung 8,4). Der SK-Test in Berlin enthielt nach der Optimierung 123 von 150 verwendbaren Items. Die Reliabilität des SK-Testes erreichte optimiert einen Alpha-Wert von 0,79. Die interne Validität der SK-Items wies eine signifikanten mittleren Korrelationskoeffizienten von  $r = 0,49$  auf. Die Korrelation nach Pearson zwischen den Ergebnissen der MC-Klausur und den Ergebnissen des SK-Testes ergaben einen Wert von  $p = 0,11$  und war somit statistisch nicht signifikant (Signifikanz 0,09,  $p < 0,05$ ). Die SKT-Ergebnisse der Dozierenden korrelieren statistisch signifikant mit den SKT-Ergebnissen der Studierenden ihrer Seminargruppe ( $r = 0,48$ , Signifikanzniveau 0,007). Dozierenden mit einem besseren Ergebnis als ihre Mit-Dozierenden, bilden auch Gruppen mit besseren Ergebnissen aus.

Die Unterschiede in der Berufserfahrung der Dozierenden weisen keine Korrelation zu deren SKT-Ergebnissen auf. Der Test unterscheidet in diesem Fall nicht signifikant zwischen erfahrenen und unerfahrenen Dozierenden.



Männliche und weibliche Dozierenden \_innen und männliche und weibliche Studierende unterscheiden sich in ihren SKT-Ergebnissen jeweils statistisch signifikant ( $p=0,027$ ) voneinander. Männliche Dozierende erreichten 79% (Median 80,0%) gegenüber 84,6% bei den weiblichen Dozierenden. Dieses Ergebnis korrespondiert mit Ergebnissen aus anderen Studien, die sich mit der Geschlechterdifferenzierung in schriftlichen Prüfungen befasst haben (Mc Donough 2000]. In den bisher durchgeführten Studien mit dem SKT wurde dieser Faktor nicht erwähnt oder stand im Mittelpunkt der Untersuchungen.

## Diskussion

Ausgehend von der Diskussion der Testgüte-Kriterien des SK-Testes bei pädiatrischen Studierenden, werden die Korrelationen der MC-Klausurergebnisse in Bezug auf die SKT-Ergebnisse und der Bezug der Ergebnisse dargestellt. Die Limitationen der Studie werden in Bezug auf die beschriebenen Testgütekriterien betrachtet und die daraus resultierenden Implikationen und Schlussfolgerungen diskutiert.

---

Der erreichte Cronbach alpha Wert von 0,79 in dieser Studie liegt knapp unter den Anforderungen (0,80), die in Bezug auf diesen Koeffizient als Wert an eine notengebende Prüfung (High stakes Examen) gestellt werden [77].

Laut Downing steht die Validität im Mittelpunkt der Beurteilung der Qualität eines Testinstruments. Eine Reihe von Faktoren können die Validität der Testgütekriterien in den Testergebnissen einschränken [78]. Zusammengefasst wurden diese Faktoren von Messick ursprünglich unter den Kategorien „Construct under representation“ (CU) und „Construct-irrelevant variance“ (CIV) [75]. Unter CU werden die folgenden Kriterien subsumiert:

1. Ausreichende Anzahl von Items um die überprüfte Domäne zu repräsentieren.
2. Bias oder nicht repräsentativer Ausschnitt an Items aus dem Fachbereich.
3. Umfassende Beispiele aus dem Fachbereich.
4. Niedrige Ergebnis Reliabilität.

Messick fasst unter CIV die folgenden Faktoren zusammen:

1. Fehlerhafte Itemformate.
2. Bias zwischen den Items.
3. Fehlerhafte oder missverständliche Formulierung der Items.
4. Items diskriminieren nicht oder sind zu schwer oder zu leicht.
5. Betrug bei der Beantwortung der Items
6. Unangemessene Bestehensgrenzen
7. Schlecht trainierte Dozierenden

Die genannten Faktoren werden in den folgenden Abschnitten diskutiert.

## **Validitätskriterien der MC-Klausur**

Die MC-Klausur entspricht den Validitätskriterien, die gemäß den Kriterien der ACGME für notengebende MC-basierte Prüfungen empfohlen werden [36]. Bei der Erstellung der Fragen wurden die Kriterien für die Erstellung nach Palmer beachtet, die in die Charité-eigenen Kriterien für die Erstellung und Pflege von MC-Fragen eingeflossen sind [62]. Der Kolmogorov-Smirnov-Test zeigt eine Normalverteilung der Ergebnisse der Studierenden bei einer zweiseitigen Signifikanz von  $p = 0,007$ . Daraus lässt sich ableiten, dass die Zusammenstellung der Fragen-Items erfolgreich zwischen den unterschiedlichen Kenntnissen der Studierenden unterscheiden kann.

---

## **Validitätskriterien des Skript-Konkordanz-Tests**

### **Validität**

Die Validität einer Testmethode gibt die Zuverlässigkeit an, mit der diese das misst, was diese zu messen vorgibt. Lienert et al unterscheiden zwischen Inhaltsvalidität, Konstruktvalidität, Itemvalidität und Kriteriumsvalidität [79]. Zusätzlich werden die von Downing et al in den „Standards for Educational and Psychological Testing“ empfohlenen Kriterien in der Diskussion berücksichtigt [80].

### **Inhaltsvalidität**

Die Inhaltsvalidität eines Tests wird als wichtiges Element für die Bewertung einer Testmethode erachtet [80]. Der Skript Konkordanz Test ist hinsichtlich der Inhaltsvalidität in der studentischen Ausbildung in diesem Umfang als notengebendes Prüfungsformat zum Zeitpunkt der Studiendurchführung noch nicht mit einer großen Kohorte überprüft worden. Den Ergebnissen aus anderen Studien zufolge, kann für das SKT-Konstrukt ein hohes Maß an Inhaltsvalidität zugrunde gelegt werden [16].

Die Strukturierung der Inhalte erfolgte anhand eines von pädiatrischen Fachexperten\_innen definierten Blueprints für die Erstellung von Prüfungsfragen für den studentischen Unterricht in der Pädiatrie. Der Test wurde entsprechend der Gewichtung der elf pädiatrischen Fachgebiete zusammengestellt. Es wurde sichergestellt, dass die von den Testitems überprüften Inhalte den Lernzielen des Blockpraktikums und des Unterrichts am Krankenbett entsprechen (s. Übersicht Lernziele Blockpraktikum). Die Lernziele und die Fragenitems selber beziehen sich auf wichtiges und grundlegendes pädiatrisches Wissen, die typischen Krankheitsbilder, Diagnosen und wichtige therapeutische Ziele, die nach einer pädiatrischen Ausbildung, ungeachtet der später gewählten Fachrichtung, von jedem Arzt am Ende seiner Ausbildung beherrscht werden sollten.

Die Gewichtung der Fachgebiete erfolgte anhand ihres Anteils am Gesamtunterricht der Charité, der sich anhand der Zahl der vorgehaltenen Betten in dem Bereich berechnet. Bei der praktischen Umsetzung des SK-Tests im Blockpraktikum der Kinderheilkunde wurde der Umfang der Fragen weitgehend an den Lehranteilen der einzelnen Fachbereiche ausgerichtet. Ein Übergewicht entstand zugunsten der pädiatrischen Endokrinologie und pädiatrischen Gastroenterologie, die signifikant mehr Fragen zugewiesen bekamen, als die Neonatologie, Kinderchirurgie und Kinder- und Jugendpsychiatrie. Dies rechtfertigt sich durch den Anteil der Lernziele der genannten Fachbereiche an denen des Blockpraktikums der



---

Kinderheilkunde. Der Unterrichtsschwerpunkt der Kinder- und Jugendpsychiatrie liegt im fünften klinischen Semester (UaK). Trotzdem wird hier bei der Entwicklung zukünftiger SK-Fragen ein Schwerpunkt in den untergewichteten Fachbereichen gesetzt werden müssen, um dieses Ungleichgewicht zu vermindern.

Der Fokus dieser Studie wurde bei der Auswahl der Fragen inhaltlich auf das differentialdiagnostische Denken und die diagnostische Entscheidungsfindung gelegt. Fragen zu therapeutischen Entscheidungen spielen in den Lernzielen der Kinderheilkunde des OHCs eine sekundäre Rolle. Dies entspricht den Anforderungen, die in vorhergehenden Studien von Fournier et al bezüglich der Zusammenstellung von Testfragen im studentischen Bereich aufgestellt wurden, um eine inhaltliche Reliabilität der Fragen zu erreichen, die den Ansprüchen einer notengebenden Prüfung zu entsprechen. Es konnte sichergestellt werden, dass mit den Fragen des SK-Tests prozedurales kognitives Wissen überprüft wurde. Die Abfrage von präzise definiertem, theoretischem Faktenwissen bleibt der MC-Klausur vorbehalten [14].

Die Fragen des SK-Tests wurden von einem Team aus erfahrenen Pädiatern des OHCs erstellt. Bei der Fragenerstellung wurde das Qualitätssicherungsraster, von Charlin eingesetzt, eingesetzt (s. Anhang Qualitätskriterien). Anschließend wurden die Fragen einer Experten\_innen-Bewertung durch zwölf Mitglieder des Dozierenden-Teams des Blockpraktikums unterworfen und in einer Pilotstudie über ein Semester getestet (Sommersemester 2008). Mit den so gesammelten Rückmeldungen wurden mit dem Ziel der Verbesserung der Reliabilität des Tests zweideutige und schlecht diskriminierende Fragen aus dem Fragenpool ausgeschlossen. Dies betraf insgesamt 13 von 150 Fragen.

### **Konstruktvalidität**

Die Konstruktvalidierung bildet nach Messick die theoretische Grundlage für die praktische Integration der kriterienbezogenen und inhaltlichen Validitätsaspekte [75]. Nach Downing bezieht sich die interne Struktur als Teilaspekt der Validität auf die statistischen Daten des Tests. Die interne Struktur beschreibt das psychometrische Modell der Erstellung der Testergebnisse. Ferner sind hierunter die Daten der Reproduzierbarkeit und Übertragbarkeit der Testergebnisse abzubilden.

Die Reliabilität der vorhergehenden Studien zum SKT wurde auf der Basis der eingesetzten Fragen ermittelt [81]. Dort konnten Cronbach-alpha-Werte zwischen 0,77 (60 Items), 0,79 (80 Items), und 0,9 (70 Items) erreicht werden. Die einzelnen Fragen sind in diesen

---

Teststrukturen jeweils zu mehreren einem Fall zugeordnet. Da in dieser Studie die Fälle aus ein bis maximal drei Fragen bestanden, war nicht zu erwarten, dass die Berechnung der Alpha-Werte für die Fälle gegenüber denen für die einzelnen Fragen einen statistisch signifikanten Unterschied erbringen würde. Diese Annahme bezieht sich auf Erfahrungen aus voran gegangenen Studien mit diesem Instrument, in denen aber die Fälle mit bis zu acht Fragen ausgestattet wurden, was zu einer geringfügig größeren statistischen Signifikanz für die einzelnen Fälle geführt hat [59, 82]. Für den erstmals in der studentischen Ausbildung eingesetzte Berliner Pädiatrie SKT wurde mit 45 Items ein Cronbach's-Alpha-Koeffizient von 0,79 erreicht. Dieser Wert wird als ausreichend betrachtet für eine universitäre Examensprüfung. Die MC-Klausur weist bei 39 berechneten Items einen Cronbach's-Alpha-Koeffizienten von 0,72 auf. Der Vergleich der MC-Klausurergebnisse mit denen des SKT-Testes der Studierenden zeigt in der linearen Regressionsanalyse keine Signifikanz auf ( $r=0,11$ ). Dies kann als Hinweis auf die unterschiedlichen Kompetenzdomänen interpretiert werden, die durch die beiden Tests gemessen werden.

#### **Kriteriumsvalidität – Trennschärfe der Items**

Als Bewertungsschlüssel wurde die Aggregat-Scoring Methode verwendet [18]. Die Besonderheit der Methode besteht in der Tatsache, dass mit dieser Methode mehr als eine richtige Antwort als Lösungsschlüssel angenommen werden. Aufgrund dessen wurden die Gesamtergebnisse der Studierenden in absoluten Prozenträngen ermittelt. Anschließend wurden nicht positiv trennscharfe Items aus der Untersuchung ausgeschlossen. Insgesamt verblieben aber 118 der 138 zugelassenen Testitems in der Itemanalyse. Unter Ausschluss der 20 nicht trennscharfen Items ergibt sich eine höhere mittlere Interitemkorrelation von 0.076. Hinsichtlich seiner Konstruktvalidität erzielt Skript Konkordanz Test stabile Werte innerhalb der verschiedenen Ranggruppen. Beide der verglichenen Testformate weisen in dieser Hinsicht die für High-stakes Examen vorausgesetzten psychometrischen Rahmendaten in Bezug auf Validität und Reliabilität auf.

#### **Kriteriumsvalidität – Antworten der Dozierenden**

Die verschiedenen Ebenen der internen Validität des Testinstrumentes wurden in dieser Studie dokumentiert, um eine für eine notengebende Prüfung ausreichende Sicherheit der Kriterien zu belegen.

Die mit den verwendeten 40 Fragen erreichte hohe Reliabilität des Tests rechtfertigt die Fortsetzung des Einsatzes als abschließende notengebende Prüfung mit starkem klinischen

---

Bezug und als Ergänzung zu der MC-Prüfung, die auf die Überprüfung von klinischem und klinisch-theoretischem Wissen ausgelegt ist.

Der Test unterscheidet statistisch signifikant zwischen den klinischen Erfahrungsniveaus der Studierenden und der Dozierenden. Klinische Experten\_innen unterscheiden sich in den Testwerten entsprechend ihren klinischen Berufsjahren nicht signifikant voneinander. Dies spricht gegen die Konstruktvalidität des Tests. Zugleich widerspricht dieses Ergebnis den Ergebnissen, die von Charlin und Carrière mit dem SKT auf dem Gebiet der postgradualen Fortbildung erzielt wurden [12,15,59,69]. Ein möglicher Erklärungsansatz für diesen Unterschied ist der Schwierigkeitsgrad der verwendeten Fragenitems. Dieser repräsentiert ein Niveau, das dem Wissens- und Kompetenzstand in klinischem Denken der seit kurzem mit den klinischen Problemen in der Kinderheilkunde in Kontakt stehenden Studierenden der Humanmedizin entspricht. Die Abnahme der Varianz in den Wochen [5,7,8,12] weist auf einen einseitigen Deckeneffekt hin. Dieser könnte durch den für Fachexperten\_innen geringen Schwierigkeitsgrad des SKT-Pädiatries begründet werden, durch den eine große Anzahl der Ärzte in die Lage versetzt wird, in ihren Ergebnissen sich dem maximal erreichbaren Testergebnis anzunähern.

Der Einfluss der Dozierenden auf das Abschneiden der Studierenden im SKT lässt sich wie folgt beschreiben: Die durchschnittlich erreichten Noten der Studierenden korrelieren signifikant mit denen der Experten\_innen der jeweiligen Gruppe. Gute Dozierenden unterrichten tendenziell Gruppen, die ein besseres SKT Ergebnis vorweisen können (Korrelation  $N=47$   $r=0.478$ ,  $p<0,001$ , nach Z-Standardisierung  $r=0,545$ ,  $p<0,001$ ). Durch die fehlende Möglichkeit, eine entsprechende Korrelation zu den Ergebnissen der MC-Klausur zu ermitteln, kann dieser Zusammenhang nicht mit einem externen Kriterium validiert werden. Auf dem diagnostischen Niveau zeigt sich, dass die einzelnen Dozierenden ihre Seminargruppe nicht wesentlich in Richtung ihres eigenen fachlichen Schwerpunktes beeinflussen (s. Übersicht: „Einfluss Fächerzugehörigkeit“). In Bezug auf die einzelnen Fachgebiete zeigt sich, dass die Zugehörigkeit der Dozierenden zu einem Fach keinen Einfluss auf das Abschneiden ihrer Studierenden in dem SKT allgemein oder in den zu dem Fachgebiet gehörenden Fragen im Speziellen hat.

### **Durchführungsqualität**

Die Durchführungsqualität als Validitätsquelle für Tests wird von Downing entsprechend den Leitlinien der American Educational Research Association, den Standards for Educational

---

and Psychological Testing, mit den Kriterien umschrieben, die sicher stellen, dass der Beantwortungsprozess des Testes zu allen Zeitpunkten, einschließlich seiner Vorbereitung durch die durchführenden Institution und der Dozierenden und der anschließenden Datenverarbeitung mit großer Genauigkeit eingehalten wird [83, 84].

Im Rahmen des SKTs im studentischen Unterricht der Kinderheilkunde wurde bereits in der dem eigentlichen Test vorangehenden Pilotphase eine während der Testphase wiederholte Schulung der an dem Test teilnehmenden Dozierenden des Blockpraktikums durchgeführt. Den Studierenden standen vorab Testfragen zur Verfügung, die sie mit ihren Seminar Dozierenden während der Testwoche bearbeiten sollten, ein Teil von diesen gehörte zum Katalog der Testfragen. Die Zusammenstellung der Fragen und die Erstellung der Neufragen und die Auswertung der Fragen in Form von Excel-Tabellen wurden durch eine mit der Betreuung des Testes beauftragte Assistenzärztin der Kinderklinik beaufsichtigt. Die Durchführung des Testes erfolgte unter Aufsicht der Lehrkräfte der jeweiligen Woche. Die Notengebung erfolgte nach einem Modell, das den Prozenträngen der Studierenden zufolge fest in zehn Prozent Schritten Noten von fünf bis eins vergab. Die Bestehensgrenze lag bei fünfzig Prozent. Die Aussagekraft der Ränge in Bezug auf die untersuchte Kompetenz des klinischen Denkens ist mit dieser Methode nicht validiert worden. In Bezug auf die spätere Entwicklung der Studierenden als klinische Experten\_innen gibt es erste Hinweise auf die Aussagekraft eines guten Abschneidens im SKT als Messinstrument für ein im späteren Berufsverlauf gut organisiertes klinisches Experten\_innen-Wissen [12].

Obwohl verschiedene Scoring-Methoden für den SKT diskutiert wurden, konnte kein eindeutiger Vorteil im Hinblick auf die Wahl der Antwortskalenitems der Likertskala (eins bis fünf oder eins bis drei) fest gestellt werden, so dass basierend auf diesen Vorstudien von der Annahme ausgegangen werden konnte, mit der vorliegende Skala stabile Test-Ergebnisse ermitteln zu können.

### **Beziehungen der Variablen**

Die Beziehungen der einzelnen untersuchten Variablen zueinander werden durch Korrelationen dargestellt. Für die Darstellung des Zusammenhangs zwischen den gemessenen Variablen sind die bestätigenden Messdaten ebenso von Bedeutung, wie die Daten, die keinen Zusammenhang belegen. Im Rahmen dieser Arbeit wurde die Art des Zusammenhangs zwischen dem SKT als neuen Testformat und der MC-Klausur als etabliertem Testformat untersucht. Die MC-Klausur dient als Maßstab der Leistungsbeurteilung der Studierenden.

---

Negative Cronbach alpha Koeffizient in bestimmten Wochen zeigten an, dass dieser Teil der Items den Qualitätsanforderungen für den Test nicht entsprachen, die Item-Kombinationen zu leicht waren oder einzelne Items eine zu geringe Trennschärfe auswiesen. Diese Items wurden aus der Bewertung ausgeschlossen.

### **Operationalisierung der Trennschärfe**

Die Trennschärfe misst die Korrelation des Itempunktwertes mit dem Gesamtpunktwert des übrigen Tests. Da die Items meist in verschiedenen Kombinationen (pro Woche) und von verschiedenen Personen vorliegen, werden sowohl der Itempunktwert als auch Gesamtesttestpunktwert innerhalb der jeweiligen Wochenkombination vor der Korrelationsbildung z-standardisiert, um einen Bezug der Items der Testpersonen in Bezug auf die jeweilige Woche herzustellen.

Die Korrelation zwischen den SCT-Ergebnissen und den Ergebnissen der MC-Klausur waren schwach ausgeprägt ( $r=0,11$ ;  $p<0,09$ ).

### **Zusammenfassung**

Die Güte der internen Struktur des SK-Testes in der pädiatrischen Ausbildung repräsentiert durch die an den internationalen psychometrischen Standards für Validitätskriterien erhobenen Messdaten, erreicht die für den Einsatz im Rahmen eines High-stakes-Examen notwendigen Werte. Die gefundenen Ergebnisse stimmen mit den Hypothesen überein. Der SKT stellt einen Test dar, der klinisches, differentialdiagnostisches Denken überprüfen soll. Die MC-Fragen in der Form, in der sie in der Klausur verwendet wurden, überprüfen theoretisches Faktenwissen und bieten eine Auswahl von Lösungsvarianten mit einem hohen Detailgrad an. Die hohe Reliabilität beider Testformate betont die Messgenauigkeit beider Formate in ihrer jeweiligen Domäne. Es wurden keine Subdomänen beider Tests im Speziellen miteinander verglichen, da keine ausreichende Item-Anzahl in den jeweiligen Domänen für einen reliablen Vergleich erreicht werden konnte, weitere untersuchte Variablen belegen deren Abhängigkeit untereinander. Die Ergebnisse zeigen einen Unterschied in den Leistungen abhängig von verschiedenen Trainingsgraden der Teilnehmenden und des Experten\_innen-Panels. Diese Unterschiede repräsentieren den hohen Grad der Validität des Testkonstruktes, allerdings mit der Einschränkung, dass das Format in dieser Form der Fragenformulierung nicht ausreichend zwischen den beteiligten Fachärzten\_innen mit deutlich weniger Berufserfahrung zu unterscheiden in der Lage ist. Diese Ergebnisse widersprechen den Ergebnissen aus den Ausgangsstudien mit diesem Format, die aber das

---

Testformat inhaltlich auf Zielgruppen aus der beruflichen Fort- und Weiterbildung ausgerichtet haben. Da der Test in der vorliegenden Form in der Lage ist zwischen den Ergebnissen der Studierenden und Ärzten\_innen in verschiedenen Ausbildungsstufen zu diskriminieren, sollten kommende Studiendesigns versuchen, die Kompetenzen kritischen klinischen Denkens bei den Studierenden in verschiedenen Ausbildungsstufen fächerübergreifend zu messen.

Die Variablen, die eine Korrelation zwischen der Dozierenden-Qualität und den Ergebnissen ihrer Gruppen herstellen, entsprachen den Hypothesen, mit einer hohen Korrelation ( $p < 0,05$ ,  $t = 0,007$   $r = 0,48$ ). Wie in den vorhergehenden Abschnitten in den Themenbereichen Validität beschrieben, ist die Stärke des Tests, präzise zwischen Teilnehmenden in verschiedener Ausbildungsstufen zu unterscheiden, ein Merkmal, das das Konstrukt als geeignet für den Einsatz in notengebenden Prüfungen ausweist. Ein weiterer Schritt zur Bestätigung dieser Differenzierungseigenschaften in diesem Testformat muss die konsekutive Befragung von Studierenden verschiedener Ausbildungsstufen in dem gleichen Umfang sein, in dem dies in dieser Testvariante durchgeführt wurde. Im Rahmen dessen muss in Frage gestellt werden, inwiefern die Ergebnisse der Studien, die den SK-Test mit deutlich kleineren Teilnehmenden-Zahlen von Studierenden und Ärzten\_innen in verschiedenen Ausbildungsstufen eingesetzt haben, valide sind. Es muss berücksichtigt werden, dass der Entwicklungsansatz für die Fragen in diesen Testvarianten vornehmlich auf Ärzte\_innen oder Fachärzte\_innen bezogen war, so dass die überprüften Kompetenzen anderen Stufen der Expertise und Erfahrung der Teilnehmenden unterlagen [70,82]. Diese SK-Varianten konnten zudem keine Unterscheidung zwischen Ärzten\_innen verschiedener Fachrichtungen und gleich langer Ausbildungszeit herstellen, was aber auch in der geringen Power der Formate begründet liegen kann. Der Schwierigkeitsgrad der eingesetzten Fragen, dieser für die berufliche Kompetenzevaluation bestimmten Formate, entwickelt sich in diesen SK-Varianten in die entgegengesetzte Richtung in Bezug auf den Expertisegrad der Teilnehmenden. Ein weiterer Faktor, der Bestandteil weiterer Untersuchungen sein muss, ist der Unterscheidungsgrad in Bezug auf die Berufswünsche der Studierenden und deren spätere berufliche Performance [12]. Ein weiterer Gegenstand künftiger Untersuchungen könnte der Grad der Erfahrung im Umgang mit dem SKT selbst sein und mit der Tatsache begründet werden, dass der Test in den vorhergehenden Umgebungen bis auf wenige Ausnahmen in reinen Forschungsszenarien eingesetzt wurde.

---

## Limitationen

Die Studie weist in Bezug auf den Messgegenstand Beschränkungen auf. Die Definition des Messgegenstands selbst, des kritischen klinischen Denkens auf Basis der Skript-Theorie kann nicht sicher feststellen, welche Teilaspekte des klinischen Denkens in der Messung abgebildet werden. Es gibt bisher keine Studien, die die externe Validität des Tests in Bezug die gemessenen Kriterien anhand eines externen Messkriteriums sicher validieren, dies wurde auch in dieser Studie nicht durchgeführt.

Eine vergleichende Messung der einzelnen Subdomänen beider Tests wurde nicht durchgeführt. Die fehlende Diskriminierung des Formats zwischen der Ergebnissen der ärztlichen Teilnehmenden lässt zumindest die Möglichkeit für Zweifel an der Validität des Formats, auch wenn der SKT-Pädiatrie nicht für Fachärzte konzipiert wurde.

## Schlussfolgerungen

Zahlreiche Studien haben die Einsatzmöglichkeiten des Skript-Konkordanz Tests als formatives Testformat in der ärztlichen Fort- und Weiterbildung bewiesen. In der studentischen Ausbildung wurden diese Konzepte bisher nicht in diesem Umfang erprobt. Diese Studie belegt erstmals für den Rahmen der studentischen Ausbildung mit einer großen Kohorte von Teilnehmenden n, dass der Einsatz des SKT im studentischen Unterricht in der Form eines summativen Testformates den Ansprüchen eines High-stakes Examens genügen kann. Die Reliabilität des Skript-Konkordanz-Tests weist mit 45 verarbeiteten Items einen Alpha-Koeffizienten von 0,79 bei einer Testzeit von einer Stunde auf. Damit kann der Test für eine notengebende medizinische Prüfung sicher eingesetzt werden. Er würde sogar nur knapp die Qualitätskriterien für ein High stakes Examen verfehlen. Es konnten Hinweise gewonnen werden, dass das Kompetenzniveau der Dozierenden einen Einfluss auf den Erfolg der Blockpraktikumsgruppe hat und maßgeblich zu einem besseren Abschneiden der Studierenden im SKT beiträgt. Es kann daher empfohlen werden in Blockpraktika eher erfahrene Dozierende einzusetzen. Der fehlende Zusammenhang zwischen den Ergebnissen der Studierenden im SKT und der MC-Klausur weist auf die unterschiedlichen Kompetenzen hin, die mit den beiden Formaten bei den Studierenden überprüft werden. Als Hinweis auf die fehlende klinische, also berufspraktische Relevanz MC-basierter Prüfungsformate kann deren moderater prädiktiver Aussagewert im Hinblick auf die spätere berufliche Performance der Studierenden gewertet werden (85). Daher bleibt in zukünftigen Studien zu überprüfen, inwiefern der SKT als ein Parameter genutzt werden kann, der möglicherweise stärker als

---

MC-Fragen Schlussfolgerungen auf die spätere berufliche Performance der Studierenden schließen lässt.

## **Ausblick**

Die Kompetenz des klinisch-differentialdiagnostischen Denkens steht im Mittelpunkt ärztlichen medizinischen Denkens und Handelns. Unbestritten ist die Vertiefung dieser Kompetenz während der fachärztlichen Weiterbildung. Die Grundlagen für den Kompetenzerwerb müssen während des Medizinstudiums gelegt werden. Klinisch orientierte Unterrichtsformate wie der von der ÄAppO geforderte Unterricht am Krankenbett müssen sich mit praktischen Anteilen der Unterrichtseinheit an klinischen Problemstellungen orientieren. Ergänzend dazu müssen Prüfungsformate eingesetzt werden, die sich auf die Überprüfung der Kompetenzen, die für das Erlernen medizinischen Denkens notwendig sind konzentrieren. In der ärztlichen Ausbildung und der Fort- und Weiterbildung haben sich MC-Fragen als leicht standardisierbare, einfach zu verteilende, auswertbare und kostengünstige Testvariante etabliert. In der medizinischen Ausbildung werden neben den auf MC-Fragen basierenden Testformaten (einschließlich Key-Feature) und den zur Überprüfung praktischer Kompetenzen eingesetzten OSCEs, für die Messung der Kompetenz des differentialdiagnostischen Denkens mündlichen Prüfungen an vielen Fakultäten eingesetzt. Der große Vorteil des Skript-Konkordanz-Tests ist der hochgradige Praxisbezug zusammen mit den für Studierenden wie Ärzte reproduzierbaren hohen Niveaus der Testgütekriterien. Deshalb ist der SKT für den Einsatz im letzten Studienabschnitt überaus sinnvoll. Die dargestellten Ergebnisse legen die Schlüsse nahe, die mit den Hypothesen dieser Arbeit postuliert wurden:

Der Skript Konkordanz Test für die klassische Medizin in der Kinderheilkunde überprüft das Konstrukt klinischen Denken in der vorliegenden Definition im Umfeld unsicherer klinischer Informationen, indem er klinische Denkstrukturen und die daraus resultierenden differentialdiagnostischen und therapeutischen Schritte von Studierenden mit den Ergebnissen eines Experten\_innen-Panels mit Hilfe der Aggregat-Scoring Methode vergleicht. Der Test weist auch in der studentischen Ausbildung stabile Qualitätsmerkmale (Reliabilität und Validität) auf, die die Anforderungen für ein High-stakes Examen mit notengebenden Rahmenbedingungen erfüllen. Die Interitemkorrelationen und die Trennschärfe der Testitems erzielen reliable Ergebnisse bei den Studierenden, die eine Notengebung aufgrund der Ergebnisse erlauben.



---

Mit den Resultaten dieser Studie bietet sich die Möglichkeit, den SKT studienbegleitend mit auf MC-Fragen basierenden Testformaten wie dem Progress Test Medizin oder dem klinisch orientierten Key-Feature-Format zu kombinieren. Niedrige Korrelationen zwischen dem MC-Test Ergebnissen der Studierenden und deren SKT-Ergebnissen lassen den Schluss zu, dass der Test zwischen den Kenntnissen, die mit dem MC-Format getestet werden und klinisch orientierten Denkstrukturen, die mit dem SKT überprüft werden, unterscheiden kann. Der SKT unterscheidet zwischen Anfängern und Ärzten\_innen. Hier müssen weitergehende Studien ansetzen, die genauer untersuchen müssen, inwiefern die Hypothesen in Bezug auf die durch den SKT getesteten Kompetenzen differentialdiagnostisches Denken belegt werden können.

Die SKT-Ergebnisse des studentischen Testformates differenzieren nicht ausreichend zwischen den verschiedenen beruflichen Experten\_innen-Stufen. Das Auftreten dieses Effektes wurde bereits von Charlin beschrieben und weist auf eine nicht ausreichende Variabilität der Fragen in Bezug auf die verschiedenen Experten\_innen-Niveaus hin. Dieser Test jedoch intendiert keine Messung der Kompetenzen der Experten\_innen, sondern der Schwierigkeitsgrad der Fragen wurde für die Studierenden angepasst [86]. Für die ärztliche Weiterbildung muss eine Testversion eingesetzt werden, die schwierigere Fragen auf Experten\_innen-Niveau enthält, um in der Lage zu sein, verschiedene Expertise Stufen im ärztlichen Umfeld zuverlässig voneinander zu unterscheiden.

Ein wesentlicher Faktor für den Einsatz des SK-Tests im studentischen Unterricht ist die Ausrichtung der Lernziele auf das differentialdiagnostische Denken und medizinische Entscheidungsfindung. Dem sollte eine differenzierte Analyse und Festlegung der Anteile an kognitiven Wissen und handlungsorientiertem Wissen in den Lernzielen voran gehen. Eine Veränderung der curricularen Intervention kann aus den vorliegenden Ergebnissen aufgrund der Unterschiedlichkeit der fachlichen Ansprüche an die späteren Absolventen\_innen nicht abgeleitet werden. Die Ergebnisse dieser belegen die sichere Verwendungsmöglichkeit des Skript-Konkordanz-Testes als Methode zur Überprüfung klinischer Denkprozesse bei Studierenden der Humanmedizin ergänzend zu weiteren Testmethoden. Außerdem eröffnet der SKT die Möglichkeit, die für die Staatsexamensprüfungen vorhandenen Testformate zu erweitern. Der Skript Konkordanz Test füllt auf diesem Gebiet eine wichtige Lücke, in dem er dem klinisch-medizinischen Denken dem ihm gebührenden Platz einräumt.

---

## Schriftenverzeichnis

- 1 Kaufman D, Mann K: Comparing achievement on the Medical Council of Canada Qualifying Examination Part I of students in conventional and problem-based learning curricula. *Academic Medicine* 1998, 74:853-858.
- 2 Murrhardter Kreis: Das Arztbild der Zukunft : Analysen künftiger Anforderungen an den Arzt ; Konsequenzen für die Ausbildung und Wege zu ihrer Reform / Arbeitskreis Mediziner Ausbildung der Robert Bosch Stiftung - Murrhardter Kreis. Gerlingen: Bleicher 1995
- 3 Approbationsordnung für Ärzte vom 27. Juni 2002 (BGBl. I S. 2405), zuletzt geändert durch Artikel 5 des Gesetzes vom 2. Dezember 2007 (BGBl. I S. 2686), Zuletzt geändert durch Art. 5 G v. 2.12.2007 I 2686;  
[Http://www.bundesrecht.juris.de/\\_appro\\_2002/BJNR240500002.html](http://www.bundesrecht.juris.de/_appro_2002/BJNR240500002.html)
- 4 Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. *Br Med J* 1975;1(5955):447-51.
- 5 Elstein AS, Schwartz A.; Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. *BMJ*. 2002 Mar 23;324(7339):729-32. Review. No abstract available. Erratum in: *BMJ*. 2006 Nov 4;333(7575):944.
- 6 Feltovich PJ, Barrows HS. Issues of generality in medical problem solving. In: Schmidt HG, De Volder ML, eds. *Tutorials in Problem-Based Learning*. Assen / Maastricht: Van Gorcum 1984;128–142.
- 7 Schallert DL. The significance of knowledge: a synthesis of research related to schema theory. In: Otto W, White S (eds). *Reading Expository Prose*. New York: Academic, 1982:13–48.
- 8 Papa FJ, Shores JH, Meyer S.; Effects of pattern matching, pattern discrimination, and experience in the development of diagnostic expertise. *Acad Med*. 1990 Sep;65(9 Suppl):S21-2.
- 9 Frank, JR. (Ed). 2005. *The CanMEDS 2005 physician competency framework. Better standards. Better physicians. Better care*. Ottawa: The Royal College of Physicians and Surgeons of Canada.
- 10 Burger W. The Berlin reformed medical curriculum at the Charité. Experiences with the first cohort; *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*. 2006 Apr;49(4):337-43. German.
- 11 Harden R.; M. AMEE Guide No. 14: Outcome-based education: Part 1-An introduction to outcome-based education, *Medical Teacher*, Volume 21, Number 1, 1 January 1999 , pp. 7-14(8)

- 
- 12 Brailovsky C, Charlin B, Beausoleil S, et al, Measurement of clinical reflective capacity early in training as a predictor of clinical reasoning performance at the end of residency: an experimental study on the script concordance test. *Med Educ.* 2001 May;35(5):430-6
  - 13 Charlin B, *Scripts and Medical Diagnostic Knowledge: Theory and Applications for Clinical Reasoning Instruction and Research*, *Acad. Med.* 2000;75:182–190.
  - 14 Charlin B., *Research Basic to Medical Education: The Script Concordance Test: A tool to assess the reflective clinician*, *Teaching and learning in medicine*, 2000, 12:4, 189-195
  - 15 Sibert L, Charlin B, Corcos J, et al. Stability of clinical reasoning assessment results with the Script Concordance Test across two different linguistic, cultural and learning environments. *Med Teach.* 2002;24:522-527
  - 16 Carrière B, Gagnon R, Charlin B, et al: Assessing clinical reasoning in pediatric emergency medicine: validity evidence for a Script Concordance Test. *Ann Emerg Med.* 2009 May;53(5):647-52. Epub 2008 Aug 22.
  - 17 Feltovich PJ, Johnson PE, Moller JH, et al: the role and development of medical knowledge in diagnosis expertise. In: Clancey WJ, Shortliffe EH (eds). *Readings in Medical Artificial Intelligence: The First Decade*. Reading, MA: Addison–Wesley, 1984: 275–319.
  - 18 Norcini et al. The Use of Aggregate Scoring for a Recertifying Examination *Eval Health Prof.*1990; 13: 241-251
  - 19 Schmidt HG, Norman GR, Boshuizen HP. A cognitive perspective on medical expertise: theory and implication (published correction appears in *Acad Med.* 1992;67:287]. *Acad Med.* 1990;65:611-621.
  - 20 Schmidt HG, Rikers RM. How expertise develops in medicine: knowledge encapsulation and illness script formation. *Med Educ.* 2007 Dec;41(12):1133-9.
  - 21 Elstein AS, Shulman LS, Sprafka SA. *Medical Problem Solving: an Analysis of Clinical Reasoning*. Cambridge, MA: Harvard University Press 1978.
  - 22 Schmidt HG, Machiels-Bongaerts M, Hermans H, ten Cate TJ, Venekamp R, Boshuizen HPA. The development of diagnostic competence: comparison of a problem-based, and integrated, and a conventional medical curriculum. *Acad Med* 1996;71 (6):658–64.
  - 23 van der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ* 1996;1 (1):41–67.
  - 24 Coderre S, Mandin H, Harasym PH, Fick GH. Diagnostic reasoning strategies and diagnostic success. *Med Educ.* 2003 Aug;37(8):695-703.
  - 25 Bordage G, Zacks R.; The structure of medical knowledge in the memories of medical students and general practitioners: categories and prototypes. *Med Educ.* 1984 Nov;18(6): 406-16.

- 
- 26 Barrows HS, Tamblyn, R. Problem-Based Learning: An Approach to Medical Education. New York: Springer, 1980.
- 27 Schön DA. The Reflective Practitioner. New York, NY: Basic Books; 1983.
- 28 Weinert F. E. (2001) Vergleichende Leistungsmessung in Schulen - eine umstrittene Selbstverständlichkeit in F. E. Weinert (Hrsg.), Leistungsmessung in Schulen. Weinheim und Basel: Beltz-Verlag; 45-58
- 29 McGaghie, William C., World Health Organization: Competency-base Curriculum Development in Medical Education: An Introduction, World Health Organization, 1978 Original University of Michigan
- 30 Hubbard J., Measuring medical education National Board of Medical Examiners; Edition: 2, illustrated Lea & Febiger, 1978
- 31 Frank JR, Danoff D. The CanMEDS initiative: implementing an outcomes-based framework of physician competencies. Med Teach. 2007 Sep;29(7):642-7.
- 32 Madeley R., Tomorrow's doctors. J Public Health Med. 1994 Dec;16(4):379-80.
- 33 Simpson JG, Furnace J, Crosby J, et al: The Scottish doctor-learning outcomes for the medical undergraduate in Scotland: a foundation for competent and reflective practitioners. Med Teach. 2002 Mar;24(2):136-43.
- 34 van der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. Med Educ. 2005 Mar;39(3):309-17.
- 35 Epstein RM, Hundert EM. Defining and assessing professional competence. JAMA 2002;287:226-35.
- 36 ACGME Outcome Project. Accreditation Council for Graduate Medical Education Web site. Available at: <http://www.acgme.org>. 2000. Accessed October 1, 2001.
- 37 Fox R. Medical Uncertainty Revisited. In: Albrecht G, Fitzpatrick, R., Scrimshaw, S., editor. Handbook of Social Studies in Health and Medicine. London: Sage Publications; 2000
- 38 Kahneman D, Slovic P, Tversky A, eds. Judgment under uncertainty: heuristics and biases. New York: Cambridge University Press, 1982.
- 39 Grzymala-Busse JW, Managing Uncertainty in Expert Systems., Kluwer Academic Publishers, Dordrecht, Boston, London (1991).
- 40 Bordage G, Lemieux M. Semantic structures and diagnostic thinking of experts and novices. Acad Med. 1991;66:S70-S72.
- 41 Marcdante KW, Simpson D., How pediatric educators know what to teach: the use of teaching scripts., Pediatrics. 1999 Jul; 104(1 Pt 2):148-50.

- 
- 41 Neufeld VR, Norman GR, Barrows HS, Feightner JW. Clinical problem-solving by medical students: a longitudinal and cross-sectional analysis. *J. of Med Educ* 1981;15:315–22.
- 42 Boshuizen HPA, Schmidt HG. Biomedical knowledge and clinical expertise. *Cogn Sci* 1992;16:153–84.
- 43 Patel VL, Groen GJ. Knowledge based solution strategies in medical reasoning. *Cogn Sci.* 1986;10:91–116.
- 44 Schank RC., Abelson RP. (1977). *Scripts, Plans, Goals and Understanding: an Inquiry into Human Knowledge Structures* (Chap. 1-3), L. Erlbaum, Hillsdale, NJ
- 45 Meterissian S, Zabolotny B, Gagnon R, Charlin B. 2007. Is the Scriptconcordance test a valid instrument for assessment of intra-operative decision-making skills? *American J Surg* 4193:248–251.
- 46 Nelson K. Event knowledge and cognitive development. In: Nelson K (ed). *Event Knowledge: Structure and Function in Development*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1986.
- 47 Gobet, F. & Simon, H. A. (1996). The roles of recognition processes and lookahead search in time-constrained expert problem solving: Evidence from grandmaster level chess. *Psychological Science*, 7, 52-55.
- 48 De Groot A.D. (1965). *Thought and choice in chess* (2<sup>nd</sup> ed. 1978). The Hagues: Mouton
- 49 Charness, N. (1992). The impact of chess research on cognitive science. *Psychological Research*, 54, 4-9.
- 50 Colliver JA, Effectiveness of problem-based learning curricula: research and theory. *Acad Med* 2000;75:259-66.
- 51 Schmidt HG, Boshuizen HP. On the origin of intermediate effects in clinical case recall. *Mem Cognit.* 1993 May;21(3):338-51.
- 52 Patel VL, Groen GJ, Norman GR. Effects of conventional and problem-based medical curricula on problem solving. *Acad Med.* 1991 Jul;66(7):380-9.
- 53 Woods NN, Brooks LR, Norman GR. It all make sense: biomedical knowledge, causal connections and memory in the novice diagnostician. *Adv Health Sci Educ Theory Pract.* 2007 Nov;12(4):405-15. Epub 2007 Feb 22.
- 54 Verkoeijen PP, Rikers RM, Schmidt HG, et al, Case representation by medical experts, intermediates and novices for laboratory data presented with or without a clinical context. *Med Educ.* 2004 Jun;38(6):617-27.

- 
- 55 de Bruin AB, Schmidt HG, Rikers RM. The role of basic science knowledge and clinical knowledge in diagnostic reasoning: a structural equation modeling approach. *Acad Medicine* 2005 Aug;80(8):765-73.
- 56 Grant J, Marsden P. The structure of memorized knowledge in students and clinicians: an explanation for diagnostic expertise. *Med Educ*. 1987 Mar;21(2):92-8
- 57 Eva KW. What every teacher needs to know about clinical reasoning. *Med Educ* 2005;39:98–106.
- 58 Norcini JJ, Blank LL, Arnold GK, Kimball HR. The mini-CEX (clinical evaluation exercise): a preliminary investigation. *Ann Intern Med*. 1995;123: 795-9.
- 59 Charlin B, Brailovsky C, Leduc C, Blouin D. The Diagnosis Script Questionnaire: A New Tool to Assess a Specific Dimension of Clinical Competence. *Adv Health Sci Educ* 1998;3(1):51-58.
- 60 Schuwirth LW, Different written assessment methods: what can be said about their strengths and weaknesses? *Med Educ*. 2004 Sep;38(9):974-9. Wright Expertise and Decision Support George Wright, Fergus Bolger
- 61 Downing SM. Reliability: on the reproducibility of assessment data. *Med Educ* 2004;38:1006-1012.
- 62 Palmer E, Devitt P: Constructing multiple choice questions as a method for learning. *Ann Acad Med Singap* 2006, 35:604-08.
- 63 Schuwirth LW, van der Vleuten CPM, Donkers HHLM. A closer look at cueing effects in multiple-choice questions. *Med Educ* 1996;30:44–9.
- 64 Hatala R, Norman GR Adapting the Key Features Examination for a clinical clerkship. *Med Educ*. 2002 Feb;36(2):160-5.
- 65 Fischer MR, Kopp V, Holzer M, Ruderich F, Junger J.; A modified electronic key feature examination for undergraduate medical students: validation threats and opportunities. *Med Teach*. 2005 Aug;27(5):450-5.
- 66 Wright G., Bolger F., Expertise and decision support, Plenum Press New York 1992
- 67 Downing SM. Threats to the validity of locally developed multiple-choice tests in medical education: construct-irrelevant variance and construct underrepresentation. *Adv Health Sci Educ Theory Pract*. 2002;7(3):235-41.
- 68 Downing SM., Assessment of Knowledge with written Test Forms; *International Handbook of Research in Medical Education*; Norman et al Springer, 2002, pp.658-664
- 69 Charlin B, Desaulniers M, Gagnon R, Blouin D, van der Vleuten C. Comparison of an aggregate scoring method with a consensus scoring method in a measure of clinical reasoning capacity. *Teach Learn Med*. 2002 Summer;14(3):150-6.

- 
- 70 Sibert L, Charlin B, Gagnon R, Corcos J, Khalaf A, Grise P. (Evaluation of clinical reasoning in urology: contribution of the Script Concordance Test]. *Prog Urol* 2001;11(6):1213-9.
- 71 Norman GR, Smith EKM, Powles AC, et al. Factors underlying performance on written tests of knowledge. *Med Educ* 1987;21:297–304.
- 72 Fournier JP, Demeester A, Charlin B. Script concordance tests: guidelines for construction. *BMC Med Inform Decis Mak.* 2008 May 6;8:18.
- 73 Gagnon R, Charlin B, Coletti M ; Assessment in the context of uncertainty: how many members are needed on the panel of reference of a script concordance test? *Med Educ.* 2005 Mar;39(3):284-91.
- 74 Ekwo EE, Loening-Baucke V.; Clinical problem solving and clinical knowledge. *Med Educ.* 1979 Jul;13(4):251-6.
- 75 Messick S., Validity. In: *Educational Measurement*, 3rd edn. Ed: Linn RL. New York: American Council on Education and Macmillan 1989:13–104.
- 76 Heemskerk L, Norman G, Chou S, et al The effect of question format and task difficulty on reasoning strategies and diagnostic performance in Internal Medicine residents. *Adv Health Sci Educ Theory Pract.* 2008 Nov;13(4):453-62.
- 77 Downing SM: *Assessment of Knowledge with written test forms*; in *International Handbook of research in medical education*; pp. 647- 672, GR Norman et al; Dordrecht 2002
- 78 Downing SM, Haladyna TM. Validity threats: overcoming interference with proposed interpretations of assessment data. *Med Educ.* 2004 Mar;38(3):327-33.
- 79 Lienert & Raatz (1994), *Testaufbau und Testanalyse*. 5. völlig neu bearbeitete und erweiterte Auflage; S.283-285. Weinheim: Beltz – PVU.
- 80 Downing SM. Validity: on meaningful interpretation of assessment data. *Med Educ* 2003;37(9):830-7.
- 81 Lambert C, Gagnon R, Nguyen D, Charlin B. The script concordance test in radiation oncology: validation study of a new tool to assess clinical reasoning. *Radiat Oncol.* 2009 Feb 9;4:7.
- 82 Brazeau-Lamontagne L, Charlin B, Gagnon R, Samson L, Van Der Vleuten C. Measurement of perception and interpretation skills during radiology training: utility of the script concordance approach. *Med Teach* 2004;26(4):326-32.
- 83 Downing SM. Validity: on meaningful interpretation of assessment data.; *Med Educ.* 2003 Sep;37(9):830-7.

---

84 American Educational Research Association, American Psychological Association, National Council on Measurement in Education. Standards for Educational and Psychological Testing. Washington, DC: American Educational Research Association 1999.

85 Hamdy H, Prasad K, Anderson MB, Scherpbier A, Williams R, Zwierstra R, Cuddihy H. BEME systematic review: predictive values of measurements obtained in medical schools and future performance in medical practice. *Med Teach*. 2006 Mar;28(2):103-16.

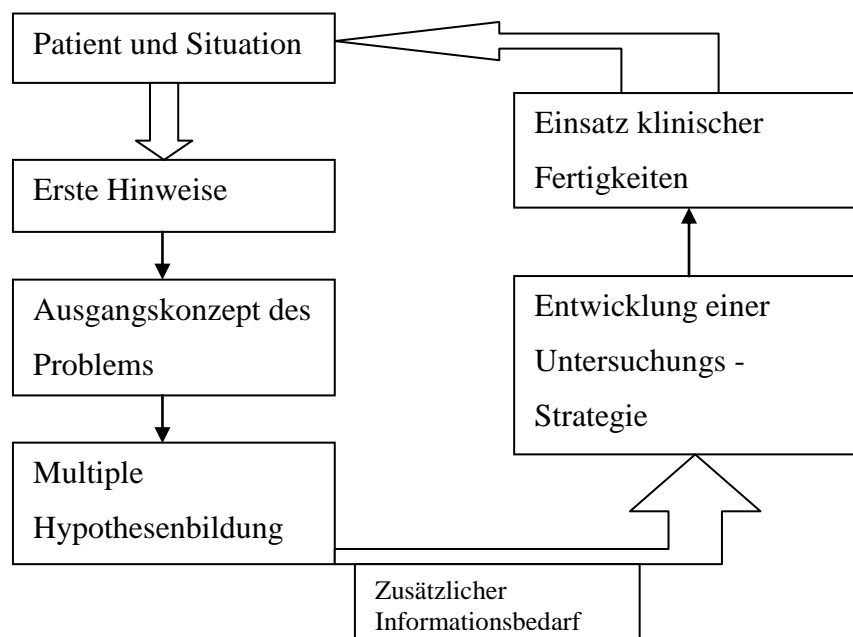
86 Charlin B, Gagnon R, Pelletier J, et al, Assessment of clinical reasoning in the context of uncertainty: the effect of variability within the reference panel. Coletti M, Abi-Rizk G, Nasr C, Sauvé E, van der Vleuten C. *Med Educ*. 2006 Sep;40(9):848-54.



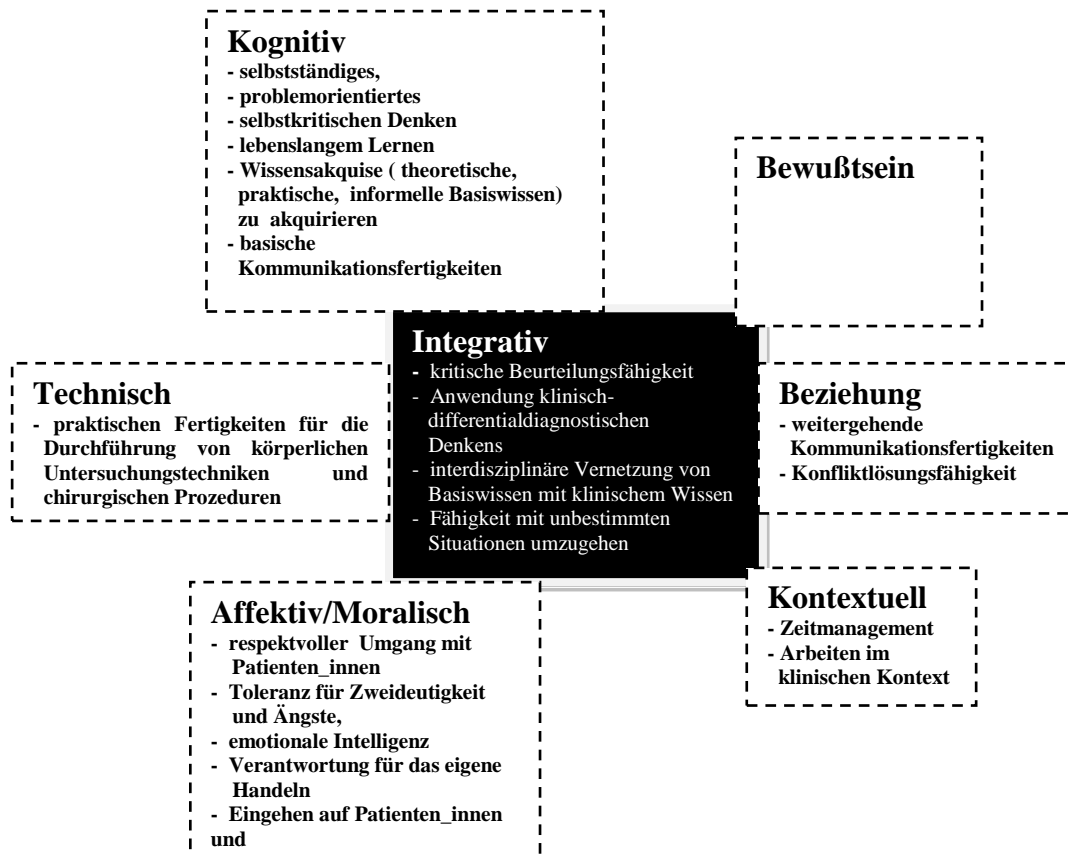
---

## Abbildungen

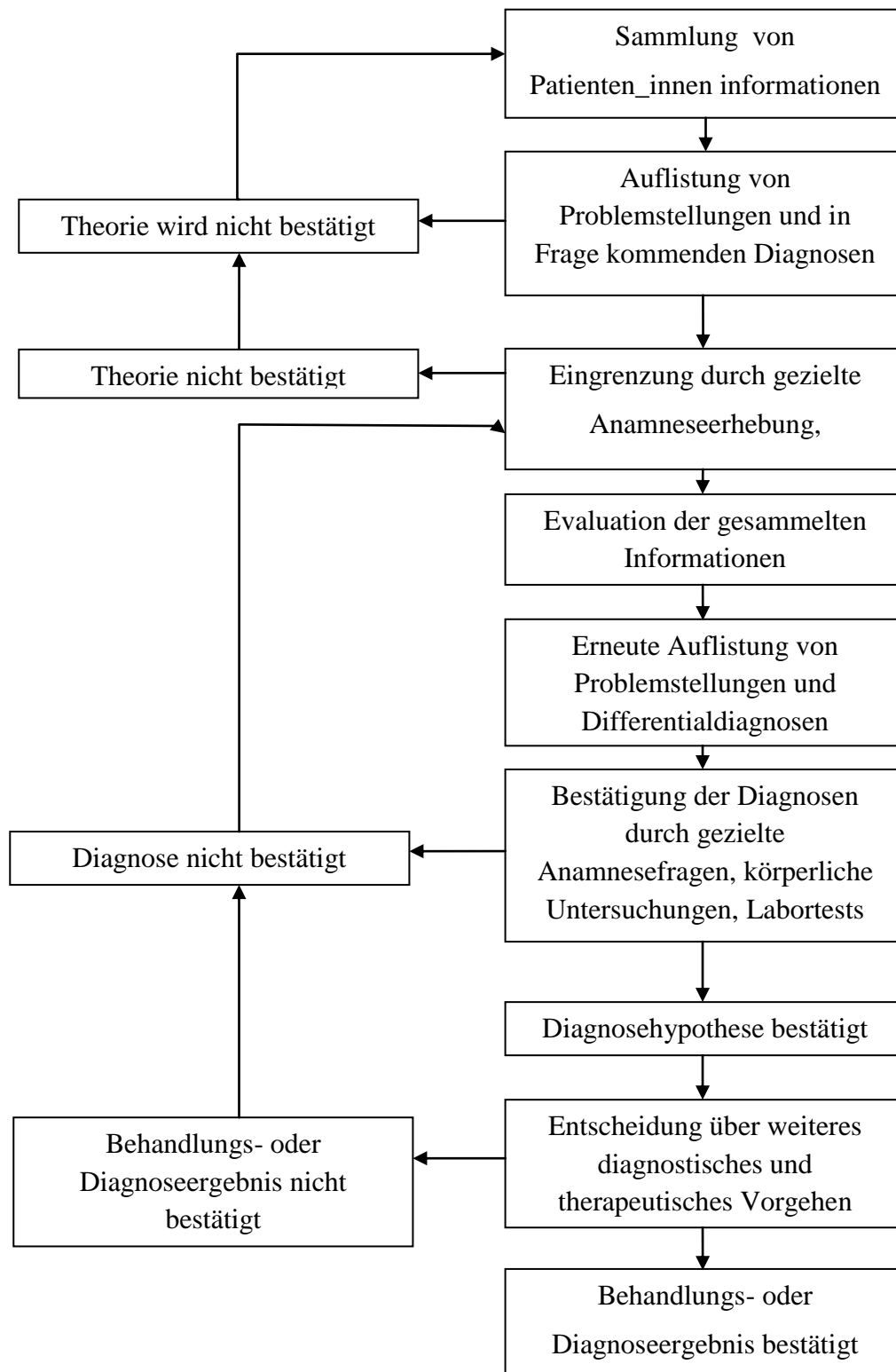
**Abbildung 1:** Ablaufdiagramm des klinischen Denkprozesses (adaptiert nach H. Barrows „Problem-based Learning“)



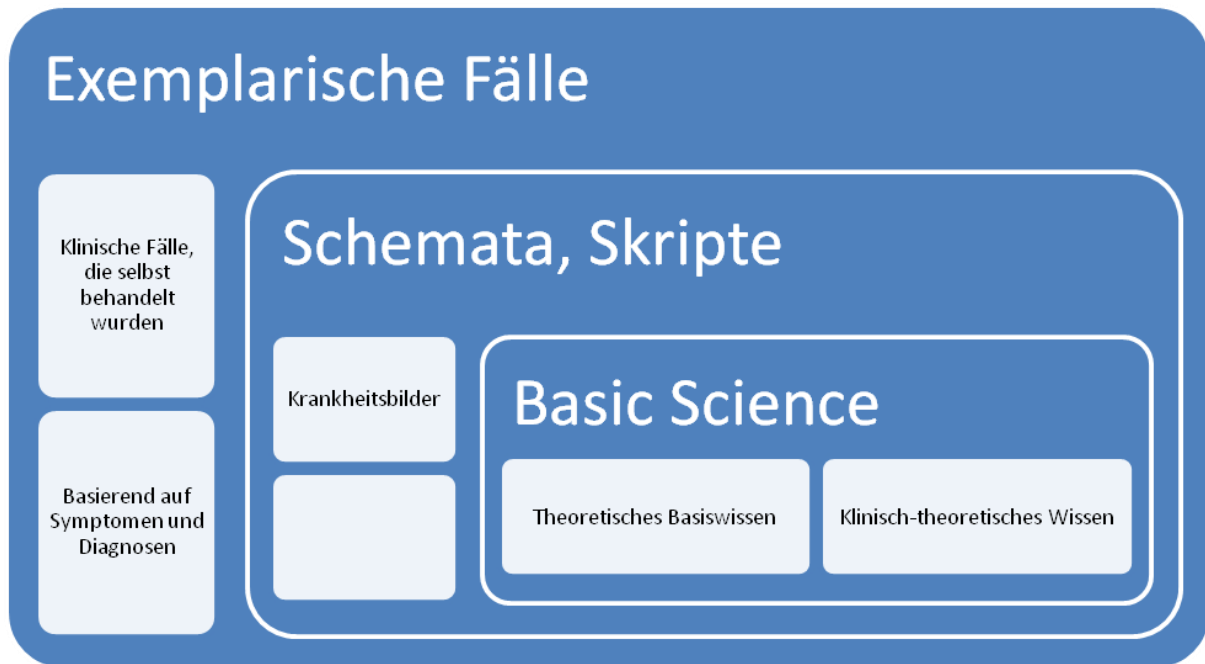
**Abbildung 2 : Ebenen der professionellen Kompetenz**



**Abbildung 3:** Hypothetiko-deduktives Modell der Diagnoseentwicklung: modifiziert nach *Clinical Method*, Robin C Fraser 1987. (Nach Epstein et al JAMA 2002)



**Abbildung 4 :** Formen mentaler Wissenrepräsentation



(nach Schmidt et al 1990)

## Tabellen

**Tabelle 1:** Entwicklungsphasen professioneller Kompetenz bei Lehrenden

|  |   |
|--|---|
| I Frühe Phase                          | 0. Bedenken hinsichtlich der eigenen Fertigkeiten   |
| II Zwischenphase (Kompetenzerwerb)     | 1. Bedenken hinsichtlich der Erwartungshaltungen und Akzeptanz im professionellen Umfeld.<br>2. Bedenken hinsichtlich der eigenen Zulänglichkeit und Fähigkeiten der Seminargruppenleitung.<br>3. Bedenken hinsichtlich der zwischenmenschlichen Umgangs mit den Studierenden.                  |
| III Späte Phase (Professionalisierung) | 4. Bedenken hinsichtlich der Aneignung des gelehrten Stoffes durch die Studierenden.<br>5. Bedenken hinsichtlich der Bedarfsgerechtigkeit des gelehrten Stoffes.<br>6. Bedenken hinsichtlich des eigenen Beitrags das Verhalten der Studierenden auf zwischenmenschlicher Ebene zu beeinflussen |

Adaptiert nach M. Eraut (Fuller 1970) Developing Professional Competence and Knowledge

**Tabelle 2:** Vergleich verschiedener Testmethoden hinsichtlich der Reliabilität

| Fragentyp                           | Beschreibung | Reliabilität nach Testzeit |      |      |
|-------------------------------------|--------------|----------------------------|------|------|
|                                     |              | nach 1,2,4 Stunden         |      |      |
| Multiple Choice Fragen              |              | 0.62                       | 0.76 | 0.93 |
| Mündliche Prüfung                   |              | 0.5                        | 0.69 | 0.82 |
| OSCE                                |              | 0.54                       | 0.69 | 0.82 |
| Skript Konkordanz Test              |              | (0,72)                     |      |      |
| (Ergänzt nach Van der Vleuten 2005) |              |                            |      |      |

**Tabelle 3 :** Ausschnitt aus der Excel-Antworttabelle der Dozierenden

| Numm<br>er des<br>Dozen<br>t/in | Note                | Frage    | Frage    | Frage    | Frage    | Frage    | Frage    | Frage    | Frage    | Frage    | Frage     |
|---------------------------------|---------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|
|                                 |                     | <i>1</i> | <i>2</i> | <i>3</i> | <i>4</i> | <i>5</i> | <i>6</i> | <i>7</i> | <i>8</i> | <i>9</i> | <i>10</i> |
|                                 | <i>Prozen<br/>t</i> |          |          |          |          |          |          |          |          |          |           |
| Dozen<br>t/in 1                 | 83,85               | 1,00     | 1,00     | 1,00     | 1,00     | 1,00     | 0,41     | 1,00     | 1,00     | 1,00     | 1,00      |
| Dozen<br>t/in 2                 | 93,89               | 1,00     | 1,00     | 1,00     | 1,00     | 1,00     | 1,00     | 0,16     | 1,00     | 1,00     | 1,00      |
| Dozen<br>t/in 3                 | 79,26               | 0,27     | 1,00     | 1,00     | 1,00     | 0,07     | 1,00     | 1,00     | 1,00     | 1,00     | 1,00      |
| Dozen<br>t/in 4                 | 90,34               | 1,00     | 0,19     | 1,00     | 1,00     | 1,00     | 1,00     | 1,00     | 1,00     | 1,00     | 1,00      |
| Dozen<br>t/in 5                 | 71,79               | 1,00     | 1,00     | 1,00     | 1,00     | 1,00     | 0,41     | 1,00     | 1,00     | 0,20     | 1,00      |
| Dozen<br>t/in 6                 | 92,28               | 1,00     | 1,00     | 1,00     | 1,00     | 1,00     | 1,00     | 1,00     | 1,00     | 1,00     | 1,00      |
| Dozen<br>t/in 7                 | 72,07               | 0,87     | 1,00     | 1,00     | 1,00     | 1,00     | 0,41     | 0,04     | 1,00     | 1,00     | 1,00      |
| Dozen<br>t/in 8                 | 78,82               | 1,00     | 1,00     | 1,00     | 1,00     | 1,00     | 0,41     | 1,00     | 1,00     | 0,20     | 1,00      |
| Dozen<br>t/in 9                 | 78,43               | 0,87     | 0,19     | 1,00     | 1,00     | 1,00     | 0,41     | 0,04     | 1,00     | 1,00     | 1,00      |

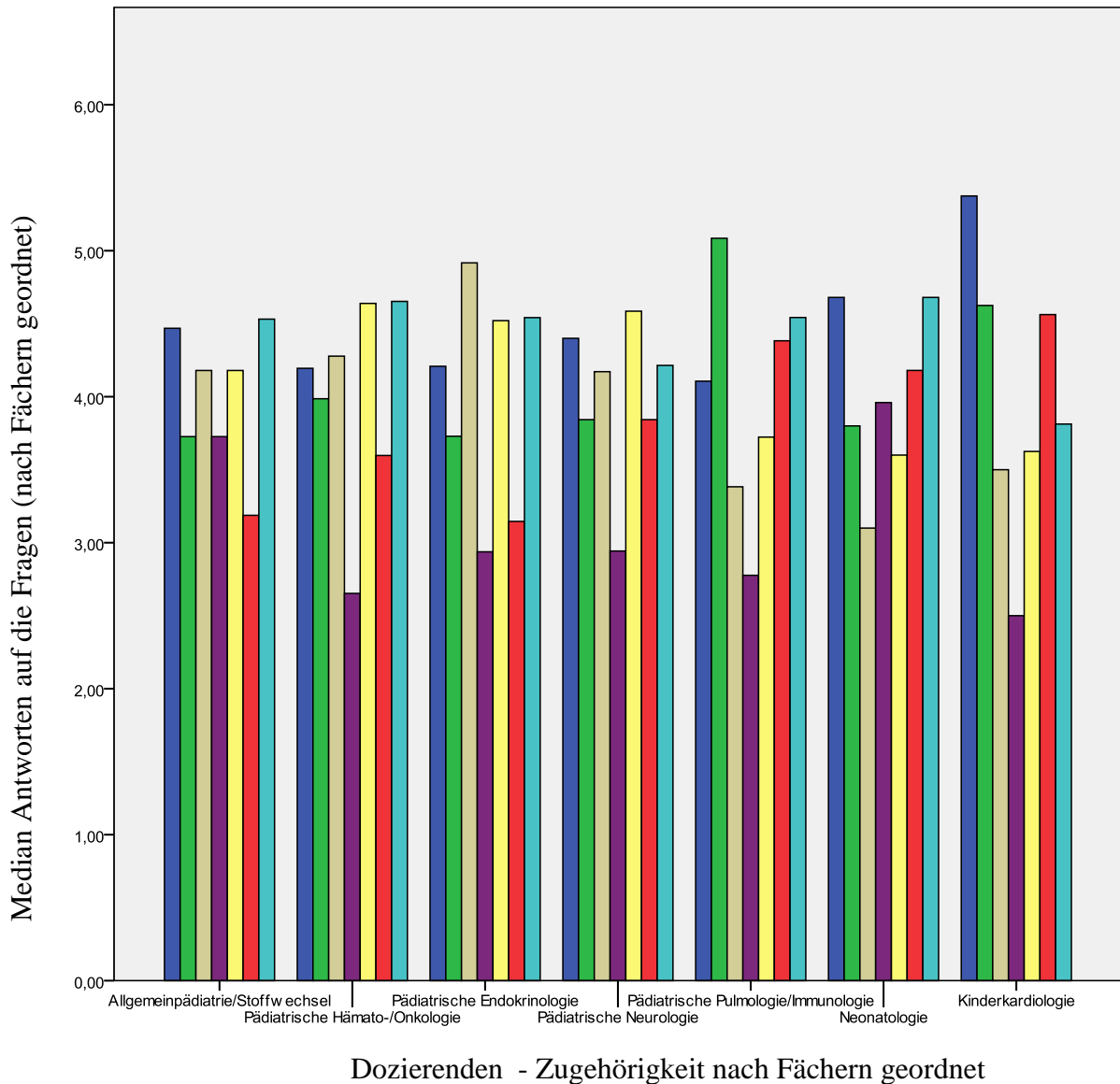
**Tabelle 4:** Ausschnitt Exceltabelle Korrekturschlüssel Referenzpanel

| Korrekturschlüssel ausgehend von den Mitglieder des Referenzpanels |                               |      |      |      |      |      |
|--|-------------------------------|------|------|------|------|------|
|  |                               | -2   | -1   | 0    | 1    | 2    |
| <b>Frage 1</b>   | <b>Korrekturschlüssel</b>     | 0,00 | 1,00 | 0,87 | 0,27 | 0,00 |
|  | <b>Anzahl Antworten Panel</b> | 0    | 15   | 13   | 4    | 0    |
| <b>Frage 2</b>   | <b>Korrekturschlüssel</b>     | 0,00 | 0,00 | 0,00 | 0,19 | 1,00 |
|  | <b>Anzahl Antworten Panel</b> | 0    | 0    | 0    | 5    | 27   |
| <b>Frage 3</b>   | <b>Korrekturschlüssel</b>     | 0,00 | 0,03 | 0,03 | 1,00 | 0,00 |
|  | <b>Anzahl Antworten Panel</b> | 0    | 1    | 1    | 29   | 0    |
| <b>Frage 4</b>   | <b>Korrekturschlüssel</b>     | 0,00 | 0,00 | 0,00 | 1,00 | 0,10 |
|  | <b>Anzahl Antworten Panel</b> | 0    | 0    | 0    | 29   | 3    |
| <b>Frage 5</b>   | <b>Korrekturschlüssel</b>     | 0,00 | 0,07 | 1,00 | 0,00 | 0,00 |
|  | <b>Anzahl Antworten Panel</b> | 0    | 2    | 30   | 0    | 0    |
| <b>Frage 6</b>   | <b>Korrekturschlüssel</b>     | 0,00 | 0,41 | 1,00 | 0,41 | 0,06 |
|  | <b>Anzahl Antworten Panel</b> | 0    | 7    | 17   | 7    | 1    |
| <b>Frage 7</b>   | <b>Korrekturschlüssel</b>     | 0,04 | 0,04 | 0,04 | 0,16 | 1,00 |
|  | <b>Anzahl Antworten Panel</b> | 1    | 1    | 1    | 4    | 25   |
| <b>Frage 8</b>   | <b>Korrekturschlüssel</b>     | 0,00 | 0,00 | 0,11 | 1,00 | 0,07 |
|  | <b>Anzahl Antworten Panel</b> | 0    | 0    | 3    | 27   | 2    |
| <b>Frage 9</b>   | <b>Korrekturschlüssel</b>     | 0,04 | 0,04 | 0,00 | 1,00 | 0,20 |
|  | <b>Anzahl Antworten Panel</b> | 1    | 1    | 0    | 25   | 5    |
| <b>Frage 10</b>  | <b>Korrekturschlüssel</b>     | 0,00 | 0,00 | 0,00 | 1,00 | 0,03 |

**Tabelle 5: MC-Klausur: Item-Konsistenz-Statistik**

|              | Skalenmittelwert, wenn<br>Item weggelassen | Skalenvarianz, wenn<br>Item weggelassen | Korrigierte Item-Skala-<br>Korrelation | Cronbachs Alpha, wenn<br>Item weggelassen |
|--------------|--|---|--|---|
| richtig.1866 | 29,01                                      | 20,580                                  | ,276                                   | ,763                                      |
| richtig.1871 | 28,73                                      | 21,090                                  | ,375                                   | ,761                                      |
| richtig.1872 | 28,97                                      | 20,172                                  | ,391                                   | ,757                                      |
| richtig.1881 | 28,92                                      | 20,759                                  | ,264                                   | ,763                                      |
| richtig.1892 | 28,93                                      | 21,103                                  | ,172                                   | ,768                                      |
| richtig.1897 | 29,15                                      | 20,117                                  | ,363                                   | ,758                                      |
| richtig.1906 | 28,82                                      | 20,816                                  | ,318                                   | ,761                                      |
| richtig.1909 | 28,69                                      | 21,685                                  | ,178                                   | ,767                                      |
| richtig.1916 | 28,95                                      | 20,418                                  | ,335                                   | ,760                                      |
| richtig.1927 | 28,82                                      | 21,049                                  | ,246                                   | ,764                                      |
| richtig.1930 | 28,68                                      | 21,820                                  | ,133                                   | ,768                                      |
| richtig.1932 | 28,74                                      | 21,234                                  | ,290                                   | ,763                                      |
| richtig.1938 | 28,87                                      | 21,204                                  | ,169                                   | ,768                                      |
| richtig.1957 | 28,69                                      | 21,737                                  | ,145                                   | ,768                                      |
| richtig.1959 | 28,91                                      | 20,987                                  | ,211                                   | ,766                                      |
| richtig.1960 | 29,12                                      | 19,492                                  | ,512                                   | ,749                                      |
| richtig.4770 | 28,94                                      | 20,754                                  | ,254                                   | ,764                                      |
| richtig.4782 | 29,59                                      | 22,131                                  | -,081                                  | ,774                                      |
| richtig.4783 | 29,00                                      | 20,303                                  | ,345                                   | ,759                                      |
| richtig.5104 | 29,55                                      | 22,335                                  | -,145                                  | ,778                                      |
| richtig.5105 | 28,83                                      | 21,041                                  | ,238                                   | ,764                                      |
| richtig.5106 | 28,95                                      | 21,036                                  | ,183                                   | ,767                                      |
| richtig.5107 | 28,67                                      | 21,979                                  | ,013                                   | ,769                                      |
| richtig.5114 | 29,00                                      | 19,900                                  | ,442                                   | ,754                                      |
| richtig.5115 | 29,02                                      | 20,229                                  | ,356                                   | ,758                                      |
| richtig.5120 | 28,69                                      | 21,948                                  | ,022                                   | ,770                                      |
| richtig.5122 | 28,94                                      | 21,160                                  | ,156                                   | ,769                                      |
| richtig.5149 | 28,69                                      | 21,780                                  | ,118                                   | ,768                                      |
| richtig.5152 | 29,13                                      | 20,844                                  | ,198                                   | ,767                                      |
| richtig.5153 | 28,98                                      | 20,424                                  | ,321                                   | ,760                                      |
| richtig.5156 | 28,91                                      | 19,949                                  | ,480                                   | ,753                                      |
| richtig.5197 | 28,71                                      | 21,611                                  | ,183                                   | ,767                                      |
| richtig.5204 | 28,75                                      | 21,113                                  | ,304                                   | ,763                                      |
| richtig.5205 | 28,76                                      | 21,240                                  | ,249                                   | ,764                                      |
| richtig.5215 | 28,84                                      | 21,480                                  | ,107                                   | ,770                                      |
| richtig.5218 | 28,84                                      | 20,882                                  | ,281                                   | ,763                                      |
| richtig.5221 | 28,92                                      | 21,374                                  | ,108                                   | ,771                                      |
| richtig.5230 | 28,83                                      | 20,126                                  | ,522                                   | ,752                                      |
| richtig.5231 | 28,69                                      | 21,581                                  | ,244                                   | ,766                                      |

**Tabelle 6:** Einfluss Fächerzugehörigkeit auf Beantwortung der Fachfragen



Legende:

- Fach
- Allgemeinpädiatrie/Stoffw echsel
  - Pädiatrische Hämato-/Onkologie
  - Pädiatrische Endokrinologie
  - Pädiatrische Neurologie
  - Pädiatrische Pulmologie/Immunologie
  - Neonatologie
  - Kinderkardiologie



---

## Lernziele/Leitfäden

### Übersicht: Lernziele des Blockpraktikums Kinderheilkunde

Die kognitiven Lernziele des voran gegangenen Semesters (UaK) stellen die theoretische Grundlage des Blockpraktikums dar.

### Anwendungsbezogene Lernziele

1. Körperliche Untersuchung bei Neugeborenen (U1 mit Reifezeichen, U2 an Untersuchungspuppe), Säuglingen, Kleinkindern, Schulkindern und Adoleszenten durchführen können.
2. Grundlagen der Anamneseerhebung bei Neugeborenen, Säuglingen, Kleinkindern, Schulkindern und Adoleszenten kennen und anwenden können.
3. Eine Epikrise für häufige pädiatrische Krankheitsbilder erstellen können.
4. In der Lage sein, eine vorgegebene Gewichts- und Längenentwicklung in Perzentilenkurven einzutragen und zu interpretieren (BMI/Zielgrößen).
5. Pubertätsstadien nach Tanner beurteilen und anwenden können.
6. Klinische Zeichen einer Meningitis / Enzephalitis in verschiedenen Altersgruppen beurteilen können.
7. Untersuchungstechniken bei Verdacht auf Appendizitis anwenden können.
8. Zeichen der klinischen Dehydratation erkennen können und den täglichen Flüssigkeitsbedarf für verschiedene Altersgruppen berechnen können.
9. Besonderheiten der Ohrinspektion bei Kindern verschiedener Altersgruppen kennen und anwenden können.
10. Klinische Zeichen einer Pneumonie bei einem 3jährigen Kind anhand der körperlichen Symptomatik, der Röntgenbefunde einschätzen und die notwendige Therapie leitlinienbasiert und evidenzbasiert planen können.
11. Die Diagnostik und anschließende Akutbehandlung eines Fieberkrampfs durchführen können.
12. Eine Beratung zur Anfallsprophylaxe von Kind mit Fieberkrampf durchführen können.
13. In der Lage sein, Masern, Röteln, Exanthema subitum, Windpocken und Scharlach differentialdiagnostisch anhand von Hauterscheinungen und Laborwerten und klinischen Erscheinungsbild des Patienten\_innen zu unterscheiden.

---

14. Ein Aufklärungsgespräch vor einer invasiven Maßnahme (Blutentnahme, Liquorpunktion) durchführen können.

---

## Übersicht: Qualitätskriterien für die SKT-Fragenentwicklung

(Workshop Charlin et al Berlin 2006, nach Downing 2004)

### Technische Kriterien

#### Vignetten

- Beschreibt eine schwierige/unklare Situation (es liegen keine ausreichenden oder widersprüchliche Daten).
- Die dargestellte Situation ist dem Ausbildungsgrad der Probanden angemessen.
- Die Vignette ist für das Verständnis der Frage notwendig. Sie versetzt den Prüfling in den spezifischen klinischen Kontext.

#### Optionen (Diagnostik, Untersuchung, Therapie)

- Steht in Bezug zur klinischen Situation, aus Sicht des Experten\_innen.
- Die gleiche Option kommt nicht in zwei aufeinanderfolgenden Fragen vor.

#### Zusatzinformationen (2. Spalte)

- Überprüft die Stärke des Zusammenhangs zwischen der Option und der neuen Information im Rahmen des klinischen Kontexts der Vignette
- Die Fragen sollten so konstruiert sein, dass die Antwortverteilung möglichst gleichmäßig auf der folgenden Skala erscheint:

-2, -1, 0, +1, +2

- Die Items sind so konstruiert, dass die Fragen dem klinischen Denken der Disziplin angenähert sind.

#### Likert Skala - Anker

- Stehen im engen Bezug zum realen klinischen Denken (z.B. Eine einzelne Information hilft kaum weiter, um eine Verdachtsdiagnose zu bestätigen oder zu verwerfen)
- Müssen im gleichen Form im gesamten Test erscheinen (negative Items rechts, Null in der Mitte, und positive Items links)
- die Bedeutung des „0“ muss eindeutig ausgewiesen werden.

---

## Eidesstattliche Versicherung

„Ich, Kai Sostmann versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: : „Skript Konkordanz Test: Eine neue Methode zur Messung der Entwicklung der Kompetenz des klinischen Denkens in der Kinderheilkunde bei Studierenden der Humanmedizin“ selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe.

Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen anderer Autoren beruhen, sind als solche in korrekter Zitierung (siehe „Uniform Requirements for Manuscripts (URM)“ des ICMJE -[www.icmje.org](http://www.icmje.org)) kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen) entsprechen den URM (s.o) und werden von mir verantwortet.

Meine Anteile an etwaigen Publikationen zu dieser Dissertation entsprechen denen, die in der untenstehenden gemeinsamen Erklärung mit dem/der Betreuer/in, angegeben sind. Sämtliche Publikationen, die aus dieser Dissertation hervorgegangen sind und bei denen ich Autor bin, entsprechen den URM (s.o) und werden von mir verantwortet.

Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§156,161 des Strafgesetzbuches) sind mir bekannt und bewusst.“

Datum

Unterschrift

---

### **Anteilerklärung an erfolgten Publikationen**

Kai Sostmann hatte folgenden Anteil an den folgenden Publikationen:

Publikation 1 Poster: Höffe J; Gaedicke G; Knigge M; Krude H; Müller S; Sostmann K (2007): Entwicklung eines Skript-Konkordanz-Testes für die pädiatrische Ausbildung. In: Jahrestagung der Gesellschaft für Medizinische Ausbildung - GMA. Hannover, 16.-18.11.2007. Düsseldorf: German Medical Science GMS Publishing House; 2007. Doc07gma56 40.

Beitrag: Fragenentwicklung, Testatdurchführung, Evaluation

Publikation 2: Höffe J; Sostmann K; Guthmann F; Knigge M; Gaedicke G (2009): New methods in assessing clinical reasoning: the script concordance test. Poster presentation: Research in Medical Education - Chances and Challenges 2009. Heidelberg, 20.-22.05.2009. Düsseldorf: German Medical Science GMS Publishing House; 2009. Doc09rmeE9

Beitrag: Mitdurchführung Fragenworkshop, Fragenentwicklung, Administration Fragebogen, Auswertung

Publikation 3 Vortrag: Sostmann K; Höffe J; Gaedicke G (2009): Skript Konkordanz Test in der Kinderheilkunde: Vergleich des klinischen Denkens bei Studierenden und Ärzten; Jahrestagung der Gesellschaft für Medizinische Ausbildung - GMA. Freiburg im Breisgau, 08.-10.10.2009. Düsseldorf: German Medical Science GMS Publishing House; 2009. Doc09gmaT5P085; Beitrag: Workshopdurchführung, Fragenerstellung, Evaluation, Administration., statistische Auswertung,

Unterschrift des Doktoranden/der Doktorandin

---

## Publikationen

### Kai Sostmann 1999 - 2015

#### 1999

Sommerburg, K Sostmann, T Grune, J. H. H. Ehrich; Oxidative stress in hemodialysis patients treated with a dialysis membrane which has alpha-tocopherol bonded to its surface. BioFactors 01/1999; 10(2-3):121-4

#### 2008

Huwendiek S, Muntau AC, Maier EM, Tönshoff B, Sostmann K (2008): E-Learning in der medizinischen Ausbildung: Ein Leitfaden zum sinnvollen Einsatz. Monatschr Kinderheilkd, 156: 458-463, 2008.

H.M. Bosse, G. Gaedicke, M. Gross, J. Forster, G.F. Hoffmann, M. Krüger, A.C. Muntau, D. Reinhardt, U.A. Schatz, K. Sostmann und A. Superti-Furga; 4 innovative pädiatrische Curricula Umsetzung der Approbationsordnung 2002; Monatsschr Kinderheilkd 2008 · 156:436–445

Sostmann K, Hoeffe J, Mueller S, Krude H, Gaedicke G; New Methods and Trends in Paediatric Education. Undergraduate Training; Pediatriki 2008; 71: 1-4

#### 2009

Sostmann K, Gaedicke K., Problem-based collaborative learning with virtual patients in pediatrics: a practical example; GMS Medizinische Informatik, Biometrie und Epidemiologie 2009, Vol. 5(1), ISSN 1860-9171

Sostmann K , Müller S, Höffe J, Gross M, Gaedicke G; Problembasiertes kollaboratives Lernen mit virtuellen Patienten in der Kinderheilkunde: ein Beispiel aus der Ausbildungspraxis ; GMS Medizinische Informatik, Biometrie und Epidemiologie. 01/2009;

#### 2010

Sostmann K, Tolks D, Fischer M, Buron S; Serious Games for Health: Spielend lernen und heilen mit Computerspielen? GMS Medizinische Informatik, Biometrie und Epidemiologie. 01/2010; 6(2).

Kaschny M, Buron S, von Zadow U, Sostmann K, „Medical education on an interactive surface“, ACM International Conference on Interactive Tabletops and Surfaces, S. 267–268, 2010.

#### 2013

von Zadow U, Buron S, Sostmann K, Dachsel R, The SimMed experience: medical education on interactive tabletops; Proceedings of the 2013 ACM international conference on Interactive tabletops and surfaces, 297-300, ACM

---

von Zadow U, Buron S, Harms T, Behringer F, Sostmann K, Dachselt R; Simmed: combining simulation and interactive tabletops for medical education, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2014, pp. 1469-1478, ACM

#### **2014**

Hoff E, Haberstroh N, Sostmann K, Perka C, Putzier M, Schmidmaier G, E-Learning in der Orthopädie und Unfallchirurgie: Eine vergleichende Pilotstudie zur Akzeptanz und zum Wissenszuwachs unter Nutzern und Nichtnutzern (Originalien), Der Orthopäde 43 (7), 674-680

Back DA, Haberstroh N, Sostmann K, Schmidmaier G, Putzier M, Perka C, High Efficacy and Students' Satisfaction After Voluntary vs Mandatory Use of an e-Learning Program in Traumatology and Orthopedics—A Follow-up Study; Journal of surgical education 71 (3), 353-359

Back DA, Haberstroh N, Antolic A, Sostmann K, Schmidmaier G, Hoff E; Blended learning approach improves teaching in a problem-based learning environment in orthopedics—a pilot study BMC medical education 14 (1), 17

#### **Buch-Beiträge**

Manfred Gross, Tina Fix, Markus Kötzle, Kai Sostmann; E-Learning in der Berliner Hochschulmedizin - Strategie der Fakultät und Umsetzung ; eUniversity - Update Bologna, Eudaction Quality forum edited by Reinhard Keil, 01/2007: pages 319-334; Waxmann Verlag.

Schnabel, Kai P. / Ahlers, Olaf / Dashti, Hiwa / Georg, Waltraud / Schwantes, Ulrich; Ärztliche Fertigkeiten, Kai Sostmann, Gerhard Gaedicke: Kapitel Kinderheilkundliche Untersuchungen; Anamnese, Untersuchung, ausgewählte Anwendungsgebiete; ISBN 978-3-8047-2446-4; 2010

Sostmann, K, Henning, J. Ehlers J.; Human- und Tiermedizin - Technologieeinsatz im Gesundheitswesen; Sandra Schön, Martin Ebner; In Lehrbuch für Lernen und Lehren mit Technologien (Februar 2011)

Mitherausgeber GMS-ZMA Sonderausgabe "Einsatz von Social Media und Web2.0 in der (tier/zahn)medizinischen Aus- und Fortbildung"; 2013

---

## **Danksagung**

Mein Dank für die Unterstützung während der Fertigstellung dieser Arbeit gilt vor allem meinen geduldigen Kindern. Meiner Mutter Brigitte und in liebevollem Gedenken meinem Vater Jörg Sostmann. Beide haben mich allzeit in meinen Bemühungen vorbehaltlos unterstützt.

Mein großer Dank gilt auf fachlicher Seite meinem Doktorvater Prof. Dr.med. Gerhard Gaedicke, Frau Dr. med. Julia Höffe, Waltraud Georg und Dr. med. Kai Schnabel.



---

## **Lebenslauf**

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

Berlin, den 11. Januar 2017