

Aus dem Institut für Didaktik und Ausbildungsforschung in der Medizin  
der Ludwig-Maximilians-Universität München

Vorstand: Prof. Dr. med. Martin R. Fischer, MME (Bern)

# **Die Auswirkung des Falldarbietungsformats und des Vorwissens auf diagnostische Kompetenzen bei Medizinstudierenden**

Dissertation

zum Erwerb des Doktorgrades der Medizin

an der medizinischen Fakultät der

Ludwig-Maximilians-Universität zu München

vorgelegt von

Valentina Manuela Jung

aus

Schwäbisch Gmünd

Jahr

2021

---

Mit Genehmigung der Medizinischen Fakultät  
der Universität München

Berichterstatter: Prof. Dr. med. Martin R. Fischer, MME (Bern)

Mitberichterstatter: Prof. Dr. Hans-Walter Pfister

Prof. Dr. Frank Fischer

Mitbetreuung durch den  
promovierten Mitarbeiter: PD Dr. Jan Kiesewetter, Dipl.-Psych

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

Tag der mündlichen Prüfung: 22.07.2021

# Hinweise

---

## **VERÖFFENTLICHUNGEN** Teile dieser Arbeit wurden veröffentlicht

**BMC Medical Education, Article number 73 (2020)**, veröffentlicht 14.03.2020:

Learning clinical reasoning: how virtual patient case format and prior knowledge interact. *Jan Kiesewetter, Michael Sailer, Valentina M. Jung, Regina Schönberger, Elisabeth Bauer, Jan M. Zottmann, Inga Hege, Hanna Zimmermann, Frank Fischer & Martin R. Fischer*

**GMA Jahrestagung Wien 2018:**

Der Effekt von Vorwissen und Falldarbietungsformat auf die Diagnoserichtigkeit und den prozeduralen Wissenserwerb Medizinstudierender. *Valentina Jung, Regina Schönberger, Michael Sailer, Elisabeth A. Bauer, Frank Fischer, Martin R. Fischer, Jan Kiesewetter*

**AMEE annual conference Basel 2018:**

Evaluation of virtual case simulations considering the extraneous cognitive load and emotions of medical students. *Regina Schönberger, Valentina Jung, Michael Sailer, Elisabeth A. Bauer, Frank Fischer, Martin R. Fischer, Jan Kiesewetter*

## **GENDER**

Nach Art. 3 Abs. 2 des Grundgesetzes sind Frauen und Männer gleichberechtigt. Alle Personen- und Funktionsbezeichnungen gelten daher für Frauen und Männer in gleicher Weise. Zum Zweck der besseren Lesbarkeit wird auf die geschlechtsspezifische Schreibweise verzichtet. In dieser Arbeit wird standardmäßig immer die männliche Form verwendet. Wenn zum Beispiel von Assistenzärzten gesprochen wird, sind auch Assistenzärztinnen gemeint.

## **FORMATIERUNG**

Englische Begriffe für die es keine eindeutige oder gebräuchliche deutsche Übersetzung gibt, bzw. bei denen der englische Ausdruck im Deutschen gebräuchlich ist, werden im Text in *Kursivschrift* dargestellt. Die *Kursivschrift* wird aber auch für andere Elemente, z.B. Überschriften oder Fragestellungen benutzt.

## **EINSICHT ONLINEFÄLLE UND ITEMS**

Die in der Studie verwendeten Onlinefälle und Items können über das Institut für Didaktik und Ausbildungsforschung der Medizin der Ludwig-Maximilians-Universität München oder über Valentina Jung eingesehen werden.

# Inhaltsverzeichnis

---

<b>Inhaltsverzeichnis .....</b>	<b>4</b>
<b>1 Einleitung.....</b>	<b>7</b>
<b>1.1 AKTUELLER FORSCHUNGSSTAND.....</b>	<b>7</b>
1.1.1 <i>Diagnostische Kompetenzen .....</i>	<i>7</i>
1.1.2 <i>Clinical Reasoning .....</i>	<i>7</i>
1.1.3 <i>Diagnoserichtigkeit und Diagnoseeffizienz.....</i>	<i>8</i>
1.1.4 <i>Deskriptives und prozedurales Wissen .....</i>	<i>8</i>
1.1.5 <i>Förderung diagnostischer Kompetenzen .....</i>	<i>9</i>
1.1.6 <i>Förderung diagnostischer Kompetenzen mit virtuellen Patienten in Onlinefällen ..</i>	<i>10</i>
1.1.7 <i>Falldarbietungsformate .....</i>	<i>12</i>
<b>1.2 NOTWENDIGKEIT VORWISSENSABHÄNGIGER FALLDARBIETUNG .....</b>	<b>12</b>
<b>2 Zielsetzung .....</b>	<b>14</b>
<b>2.1 ZIELE DER STUDIE.....</b>	<b>14</b>
<b>2.2 FRAGESTELLUNGEN .....</b>	<b>14</b>
2.2.1 <i>Wirken sich verschiedene Falldarbietungsformate für Studierende mit unterschiedlichen Vorwissensstufen positiv auf deren Diagnoserichtigkeit aus?.....</i>	<i>14</i>
2.2.2 <i>Kann im serial cue-Format eine positive Auswirkung des effizienten Diagnostizierens auf die Diagnoserichtigkeit beobachtet werden? .....</i>	<i>15</i>
2.2.3 <i>Wirken sich verschiedene Falldarbietungsformate für Studierende mit unterschiedlichen Vorwissensstufen positiv auf deren Wissenszuwachs aus?.....</i>	<i>16</i>
<b>3 Material und Methoden .....</b>	<b>17</b>
<b>3.1 STUDIENDESIGN .....</b>	<b>17</b>
<b>3.2 STUDIENTEILNEHMER.....</b>	<b>18</b>
3.2.1 <i>Voraussetzungen.....</i>	<i>18</i>
3.2.2 <i>Demographie .....</i>	<i>19</i>
<b>3.3 VORWISSENSSTUFE .....</b>	<b>19</b>
<b>3.4 PLATTFORM.....</b>	<b>19</b>
<b>3.5 DIAGNOSERICHTIGKEIT .....</b>	<b>20</b>
3.5.1 <i>Themen .....</i>	<i>20</i>
3.5.2 <i>Fallentwicklung.....</i>	<i>20</i>
3.5.3 <i>Diagnoseeffizienz und Diagnoseineffizienz.....</i>	<i>23</i>
3.5.4 <i>Review der Onlinefälle .....</i>	<i>23</i>
3.5.5 <i>Bildmaterial .....</i>	<i>24</i>
3.5.6 <i>Auswahl.....</i>	<i>24</i>

<b>3.6 MESSINSTRUMENTE</b> .....	25
3.6.1 <i>Test des deskriptiven Wissens</i> .....	25
3.6.2 <i>Test des prozeduralen Wissens</i> .....	27
<b>3.7 PILOTIERUNG</b> .....	30
<b>3.8 STUDIENDURCHFÜHRUNG</b> .....	31
3.8.1 <i>Ethische Unbedenklichkeit</i> .....	31
3.8.2 <i>Lernumgebung und Räumlichkeiten</i> .....	31
3.8.3 <i>Studienablauf</i> .....	31
3.8.4 <i>Technische Ausstattung</i> .....	32
<b>3.9 AUFARBEITUNG UND ANALYSE DER STATISTISCHEN DATEN</b> .....	32
3.9.1 <i>Aufarbeitung der statistischen Daten</i> .....	32
3.9.2 <i>Analyse der statistischen Daten</i> .....	36
3.9.3 <i>Hypothesentestung</i> .....	36
<b>4 Ergebnisse</b> .....	<b>37</b>
<b>4.1 VORGEORDNETE ANALYSEN</b> .....	37
4.1.1 <i>Einteilung der Vorwissensstufe</i> .....	37
4.1.2 <i>Test auf Normalverteilung der Residuen</i> .....	37
4.1.3 <i>Adjustierung der Signifikanzniveaus</i> .....	39
4.1.4 <i>Stichprobeneigenschaften</i> .....	39
4.1.5 <i>Einfluss demographischer Variablen</i> .....	40
4.1.6 <i>Deskriptive Statistik der abhängigen Variablen</i> .....	42
4.1.7 <i>Korrelative Zusammenhänge nach Pearson</i> .....	45
<b>4.2 HYPOTHESENTESTUNG</b> .....	47
4.2.1 <i>Auswirkung des Falldarbietungsformats und der Vorwissensstufe auf die Diagnoserichtigkeit</i> .....	47
4.2.2 <i>Auswirkung der Anzahl der diagnostischen Schritte im serial cue-Format auf die Diagnoserichtigkeit</i> .....	48
4.2.3 <i>Auswirkung des Falldarbietungsformats und der Vorwissensstufe auf den Wissenszuwachs</i> .....	50
<b>5 Diskussion</b> .....	<b>52</b>
<b>5.1 ZUSAMMENFASSUNG DER HAUPTERGEBNISSE</b> .....	52
<b>5.2 DISKUSSION DER ERGEBNISSE</b> .....	54
5.2.1 <i>Diskussion der Auswirkung des Falldarbietungsformats und der Vorwissensstufe auf die Diagnoserichtigkeit</i> .....	54
5.2.2 <i>Diskussion der Auswirkung der diagnostischen Schritte im serial cue-Format auf die Diagnoserichtigkeit</i> .....	60
5.2.3 <i>Diskussion der Auswirkung des Fallformats und der Vorwissensstufe auf den Wissenszuwachs</i> .....	64

<b>5.3 DISKUSSION DER METHODEN</b> .....	66
5.3.1 Studiendesign.....	66
5.3.2 Studienteilnehmer.....	68
5.3.3 Vorwissensstufe.....	69
5.3.4 Plattform.....	70
5.3.5 Diagnoserichtigkeit.....	72
5.3.6 Messinstrumente.....	74
5.3.7 Pilotierung.....	75
5.3.8 Studiendurchführung.....	77
<b>6 Kurzzusammenfassung</b> .....	<b>78</b>
<b>7 Literaturverzeichnis</b> .....	<b>80</b>
<b>8 Abbildungsverzeichnis</b> .....	<b>84</b>
<b>9 Tabellenverzeichnis</b> .....	<b>84</b>
<b>10 Anhang</b> .....	<b>85</b>
10.1 Anhang 1: Erläuterung des Modulsystems des klinischen Abschnitts.....	85
10.2 Anhang 2: Tabelle der Onlinefälle mit zur Verfügung stehender Diagnostik und Literaturverweis.....	87
10.3 Anhang 3: Auswertung der statistischen Daten.....	90
10.3.1 Test auf Normalverteilung der Residuen.....	90
10.3.2 Median des Vorwissens.....	92
10.3.3 MANOVA der Stichprobenverteilung.....	92
10.3.4 ANOVA KF prä Gruppen.....	95
10.3.5 ANOVA der demographischen Variablen.....	95
10.3.6 Korrelationen nach Pearson.....	96
10.3.7 Diagnoseeffizienz und Diagnoseineffizienz.....	100
10.3.8 Hypothesentestung H1 und H2.....	101
10.3.9 Hypothesentestung H3 und H4.....	106
10.3.10 Hypothesentestung H5 und H6.....	107
<b>11 Eidesstaatliche Versicherung</b> .....	<b>116</b>
<b>12 Danksagung</b> .....	<b>117</b>
<b>13 Lebenslauf</b> .....	<b>118</b>

# 1 Einleitung

---

## 1.1 AKTUELLER FORSCHUNGSSTAND

### 1.1.1 Diagnostische Kompetenzen

Diagnostische Kompetenzen sollten als unabdingbare Voraussetzungen für die Ausübung der Arztprofession während des Medizinstudiums vermittelt werden (Nationaler Kompetenz-basierter Lernzielkatalog Medizin (NKLM) 2015).

Mit diagnostischen Kompetenzen sind Fähigkeiten gemeint, die es einem ermöglichen wichtige Informationen zu sammeln, Daten zu analysieren und richtige Diagnosen zu stellen (Hasegawa, Ogasawara, & Katz, 2007). Es handelt sich also um kontextspezifische, kognitive Leistungen (Koeppen, Hartig, Klieme, & Leutner, 2008). Ein wichtiger Teil der diagnostischen Kompetenzen ist das *clinical reasoning* (Modi, Gupta, & Singh, 2015).

### 1.1.2 Clinical Reasoning

*Clinical reasoning* ist ein Kernelement diagnostischer Kompetenzen (Modi et al., 2015). Es verbessert diagnostische Leistungen, ermöglicht individuelle Behandlungen und hat einen positiven Einfluss auf den Behandlungserfolg (Fleming, 1991). Der Begriff *clinical reasoning* wurde bisher nicht eindeutig in der Literatur definiert (Simmons, 2010). Anderson (2006) definierte *clinical reasoning* als die Fähigkeit, verschiedenes Wissen anzuwenden, Evidenzen abzuwägen, kritisch zu denken und diagnostische Prozesse zu reflektieren. Grob übersetzt ins Deutsche bedeutet *clinical reasoning* klinische Schlussfolgerung. Gemeint ist damit die Fähigkeit des klinischen Denkens und Argumentierens. Es handelt sich dabei aber nicht nötiger Weise um die Fähigkeit, seine Entscheidungen explizit begründen zu können (Mattingly, 1991). Von

der Fähigkeit des *clinical reasoning* hängt es ab richtige Diagnosen in adäquater Zeit zu stellen (Linn, Khaw, Kildea, & Tonkin, 2012). Das bedeutet, dass *clinical reasoning* als diagnostische Kompetenz Einfluss nimmt auf andere diagnostische Kompetenzen, wie die Diagnoserichtigkeit und die Diagnoseeffizienz.

### *1.1.3 Diagnoserichtigkeit und Diagnoseeffizienz*

Die Diagnoserichtigkeit ist eine wichtige diagnostische Kompetenz (Hasegawa et al., 2007). Der Begriff selbst ist eine Übersetzung des englischen Begriffs *diagnostic accuracy*. Er bezeichnet die Fähigkeit, richtige Diagnosen stellen zu können. Voraussetzungen dafür sind kognitive Prozesse, wie zum Beispiel die systematische und reflexive Verarbeitung und die Kategorisierung von Informationen (Durning, Artino Jr, Schuwirth, & van der Vleuten, 2013; Helmke, 2010; Schmidt & Boshuizen, 1993). Die Diagnoserichtigkeit hängt damit auch vom effizienten Umgang mit Informationen ab (Elstein & Schwarz, 2002).

Diagnoseeffizienz bezeichnet die Fähigkeit, richtige Diagnosen in angemessener Zeit und unter adäquatem Einsatz diagnostischer Untersuchungen zu stellen. Diagnoseeffizienz ist für den klinischen Alltag sehr relevant (Braun et al., 2017), da Budgets für diagnostische Untersuchungen Grenzen haben und Ärzte meist unter Zeitdruck stehen. Sowohl richtiges als auch effizientes Diagnostizieren erfordern die Integration von individuellen Erfahrungen und theoretischem Wissen (Abs, 2007; Patel, Evans, & Groen, 1989). Diagnostizieren hängt somit vom Wissen ab.

### *1.1.4 Deskriptives und prozedurales Wissen*

Wissen, in der Medizin biologisches, physiologisches und klinisches Wissen über Symptome und Krankheitsbilder, ist eine Voraussetzung für erfolgreiches Diagnostizieren und spielt eine wichtige Rolle beim Erwerb von diagnostischen Kompetenzen (Elstein, Shulman, & Sprafka, 1978). Es kann dabei deskriptives Wissen



von prozeduralem Wissen unterschieden werden (Schmidt & Moust, 2000). Deskriptives Wissen entspricht Faktenwissen (Kopp, Möltner, & Fischer, 2006), zum Beispiel Symptome einer Krankheit oder Diagnosekriterien. Prozedurales Wissen entspricht dem Wissen, wie man etwas praktisch umsetzt (Kopp et al., 2006; Schmidmaier et al., 2013). Deskriptives und prozedurales Wissen entsprechen also den Konzepten „Wissen, dass“ und „Wissen, wie“ (Ryle, 1949). Mit zunehmender Menge und Erfahrung organisiert sich medizinisches Wissen in sogenannte *illness scripts* (Van De Wiel, Boshuizen, & Schmidt, 2000). Dies sind klinische Arbeitsschemata, die Symptom- und Krankheitskonzepte mit Wissen über Diagnosen, deren Abklärung und relevante Differentialdiagnosen verknüpfen (Custers, Boshuizen, & Schmidt, 1998; Patel et al., 1989; Schmidt, Norman, & Boshuizen, 1990; Van De Wiel et al., 2000). Wissen ist also eine wichtige Voraussetzung zur Erlernung und Förderung diagnostischer Kompetenzen.

### *1.1.5 Förderung diagnostischer Kompetenzen*

Die Förderung von diagnostischen Kompetenzen in der medizinischen Ausbildungsforschung konzentriert sich oft auf die Diagnoserichtigkeit (Schmidt & Mamede, 2015). Kiesewetter, Ebersbach, Görlitz, Holzer, Fischer und Schmidmaier (2013) haben gezeigt, dass Integrationsprozesse und Repräsentation von klinischen Informationen die Diagnoserichtigkeit verbessern können. Reflektieren allein verbessert dabei aber nicht unbedingt die Richtigkeit einer Diagnose (Schmidt et al., 2014). Auch das *clinical reasoning*, klinisches Denken und Schlussfolgern (siehe 1.1.2), kann erlernt werden (Linn et al., 2012). Das Problemlösen und die Entscheidungsfindung sind dabei Grundvoraussetzungen (Elstein & Schwarz, 2002). Um Prozesse des *clinical reasoning* zu fördern, braucht es zudem geeignete Bewertungsverfahren (Ilgen et al., 2012) und gezielte Anleitung (Mutnick & Barone,

2014). Weiter benötigt es geeignete Aufgaben, die mit einem Problem beginnen und die Entscheidungsfindung von Studierenden trainieren (Dhaliwal & Ilgen, 2016). Virtuelle Patienten und digitale Medien erfüllen die oben genannten Punkte und sind somit geeignet das *clinical reasoning* zu fördern.

### *1.1.6 Förderung diagnostischer Kompetenzen mit virtuellen Patienten in Onlinefällen*

Virtuelle Patienten werden derzeit schon vielfältig in der Lehre und der Forschung angewandt (Braun et al., 2017; Mamede, Schmidt, Rikers, Penaforte, & Coelho-Filho, 2007; Pinnock, Spence, Chung, & Booth, 2012). Sie werden meist in Onlinefällen präsentiert. Die Förderung diagnostischer Kompetenzen durch analoge wie digitale Medien weist dabei sowohl Stärken als auch Schwächen auf (Mamede et al., 2007). Die Auswirkung digitaler Medien auf das Lernen hängt dabei stark von deren Einsatz und Nutzung ab (Stegmann, 2016). Je kreativer der Umgang und je strukturierter die interaktive Anwendung, desto größer ist der positive Effekt auf den Wissenserwerb (Stegmann, 2016). Studierende wünschen sich generell einen Wechsel vom passiven Lehransätzen hin zu aktiven, multisensorischen und experimentellen Lehrmethoden (Deshpande & Huang, 2011; Prensky, 2001), wie virtuelle Patienten in Onlinefällen. Studien haben zudem gezeigt, dass die Anwendung von virtuellen Patienten im Hochschulstudium Vorteile bringt (Berman, Durning, Fischer, Huwendiek, & Triola, 2016; Fischer et al., 2008). Um diagnostische Kompetenzen, wie das *clinical reasoning* und die Diagnoserichtigkeit, bei Studierenden, die auf einer niedrigeren Lernebene nur Informationen interpretieren, zu höheren Lernebenen, wie dem Problemlösen, zu befördern, braucht es geeignete Medien und Lernmöglichkeiten (Gagne, Briggs, & Wagner, 1992). Diese sollten eigenständiges Entscheiden fördern (Gagne et al., 1992). Die Entwicklung von *clinical reasoning* zum Beispiel benötigt Situationen, in

denen Studierende klinische Probleme kritisch beurteilen und Interventionen planen und durchführen können (Fleming, 1991), was bei Onlinefällen zutrifft. Boeker, Ansel, Vach und Frankenschmidt (2013) fanden heraus, dass Studierende, die Lerninhalte in Onlinefällen bearbeiteten, einen signifikant höheren Wissenserwerb aufwiesen als Studierende, die dasselbe Material in Skriptform bearbeitet hatten. Zudem gaben die Studierenden an, mehr Spaß beim Lernen gehabt zu haben und sich sicherer im Umgang mit den Lerninhalten gefühlt zu haben (Boeker, Ansel, Vach, & Frankenschmidt, 2013). Die Definition von virtuellen Patienten scheint derzeit noch zu variieren (Ellaway, Poulton, Fors, McGee, & Albright, 2008). Cook und Triola (2009) definieren virtuelle Patienten als spezifische Simulationen, die klinische Szenarios aus dem echten Leben darstellen. Lernende schlüpfen dabei in eine Rolle und führen diagnostische Schritte durch (Association of American Medical Colleges, 2007). Sie treffen in dieser Rolle diagnostische und therapeutische Entscheidungen für ihre Patienten (Association of American Medical Colleges, 2007). Lernprozesse des *clinical reasoning*, während Famulaturen zum Beispiel, sind schwer zu beurteilen, da Studierende meist nur eine beobachtende Rolle ausüben (Kogan, Holmboe, & Hauer, 2009). Virtuelle Patienten sind dahingehend besser geeignet klinische Fähigkeiten und diagnostische Kompetenzen zu vermitteln und zu prüfen (Lang, Kogan, Berman, & Torre, 2013), da Studierende auch aktiv agieren können. Diagnostizieren, Patientenmanagement und klinische Entscheidungsfindung können dort trainiert werden, ohne echten Patienten zu schaden oder sie Risiken auszusetzen (Lang et al., 2013). Auch aus ethischen Gründen ist es daher erforderlich, diagnostische Situationen zu simulieren (Kiesewetter, Kiessling, & Fischer, 2013; Ziv, Wolpe, Small, & Glick, 2003). Um diese virtuellen Patienten effektiv einsetzen zu können, braucht es Untersuchungen über deren optimale Gestaltung und Anwendung (Cook & Triola,

2009). Zum Beispiel können sich Onlinefälle in ihrem Falldarbietungsformat unterscheiden.

### 1.1.7 Falldarbietungsformate

Zur Förderung diagnostischer Kompetenzen bieten sich verschiedene Falldarbietungsformate an, die interaktive und konstruktive Lernaktivitäten ermöglichen (Berman et al., 2016; Stegmann, 2016). Grundsätzlich können die zwei Falldarbietungsformate *serial cue* und *whole case* voneinander unterschieden werden (Schmidt & Mamede, 2015). Beim *whole case*-Format handelt es sich um ein holistisches Falldarbietungsformat, das alle Informationen von Anfang an bereitstellt. Das *serial cue*-Format ist ein sequenzielles Falldarbietungsformat, das Informationen schrittweise offenlegt. Studierende können im *serial cue*-Format auswählen, welche Informationen sie als nächstes bekommen möchten, während Studierende im *whole case*-Format ohne Wahlmöglichkeiten alle Informationen, das heißt auch irrelevante, in vorgegebener Reihenfolge erhalten (Schmidt & Mamede, 2015). Diese beiden Falldarbietungsformate *serial cue* und *whole case* können als instruktionale Gestaltung und Unterstützung gesehen werden, welche Lernende bei anspruchsvollen Problemstellungen benötigen (Kirschner, Sweller, & Clark, 2006). Die Wirkung dieser Gestaltung und Unterstützung scheint von der Phase des Kompetenzerwerbs, in dem sich der Lernende befindet, und seinem Vorwissen abzuhängen (Kalyuga, 2007).

## 1.2 NOTWENDIGKEIT VORWISSENSABHÄNGIGER

### FALLDARBIETUNG

Es ist unklar, wie sich verschiedene Falldarbietungsformate auf den Wissenszuwachs auswirken (Schmidt & Mamede, 2015). Ein Wissenszuwachs kann allein durch die optimale Darbietung eines Problems erfolgen (Avrahami et al., 1997). Ein an die

Vorwissenstufe angepasstes Falldarbietungsformat könnte als optimale Problemdarbietung fungieren und so den Wissenszuwachs fördern. Doch ist unbekannt welches Falldarbietungsformat für welche Vorwissenstufe bei Medizinstudierenden am besten geeignet wäre, um den Wissenszuwachs und diagnostische Kompetenzen zu fördern. Schmidt und Mamede (2015) vermuten, dass sich das *serial cue*-Format besser für fortgeschrittene Medizinstudierende mit höherem Vorwissen und vorhandenen *illness scripts* (siehe 1.1.3) eigne. Studierende mit höherem Vorwissen könnten dadurch zielgerichteter diagnostizieren und würden ihr Arbeitsgedächtnis dabei weniger belasten (Nendaz, Raetzo, Junod, & Vu, 2000). Für Studierende mit niedrigerem Vorwissen könnte das *serial cue*-Format gerade deshalb ungeeignet sein, da sie noch über keine solche *illness scripts* verfügen (Nendaz et al., 2000) und die Problemstellung sie überfordert (Cook & Triola, 2009). Für diese Studierende könnte sich das *whole case*-Format als vorteilhaft erweisen, da es das Arbeitsgedächtnis weniger belastet (Van Merriënboer & Sweller, 2010). Studierende mit niedrigerem Vorwissen benötigen zudem strukturierte Lösungsbeispiele (Cook & Triola, 2009) und als ein solches könnte sich das *whole case*-Format erweisen. Daher erscheinen wissensangepasste Falldarbietungsformate für Lernende notwendig, um diagnostische Kompetenzen optimal zu fördern und Wissenszuwachs zu erzeugen.

## 2 Zielsetzung

---

### 2.1 ZIELE DER STUDIE

Die vorliegende Studie soll ermitteln, wie sich *e-learning* in Onlinefällen mit virtuellen Patienten optimal nutzen lässt, um diagnostische Kompetenzen wie die Diagnoserichtigkeit und den Wissenszuwachs zu fördern. Speziell ist ein Ziel der Studie, ein geeignetes Falldarbietungsformat für Studierende mit unterschiedlichen Vorwissensstufen zu finden.

Die Ergebnisse dieser Studie könnten Auswirkungen darauf haben, wie künftig *e-learning* mit virtuellen Patienten bei Studierenden individuell und in verschiedenen Phasen des Studiums eingesetzt werden sollte, um die diagnostische Kompetenz Diagnoserichtigkeit zu fördern und einen möglichst großen Wissenszuwachs bei der Bearbeitung der Onlinefälle zu erzeugen.

### 2.2 FRAGESTELLUNGEN

*2.2.1 Wirken sich verschiedene Falldarbietungsformate für Studierende mit unterschiedlichen Vorwissensstufen positiv auf deren Diagnoserichtigkeit aus?*

Schmidt und Mamede (2015) weisen darauf hin, dass das *serial cue*-Format für fortgeschrittene Medizinstudierende geeigneter sein könnte, da bei diesen klinische Arbeitsschemata vorhanden wären, sogenannte *illness scripts* (siehe 1.1.4). Dazu ergänzend vermuten Nendaz et al. (2000), dass das *serial cue*-Format weniger für Studierende mit niedrigem Vorwissen geeignet wäre. Das *whole case*-Format könnte für diese Studierenden geeigneter sein, da es das Arbeitsgedächtnis weniger belastet

(Van Merriënboer & Sweller, 2010). Studierende mit niedrigem Vorwissen benötigten zudem Lösungsbeispiele (Cook & Triola, 2009). Als ein solches könnte das *whole case*-Format fungieren. Daraus ergeben sich folgende Hypothesen:

**H1 Studierende mit höherem Vorwissen erreichen im *serial cue*-Format eine höhere *Diagnoserichtigkeit* als im *whole case*-Format.**

**H2 Studierende mit niedrigerem Vorwissen erreichen im *whole case*-Format eine höhere *Diagnoserichtigkeit* als im *serial cue*-Format.**

*2.2.2 Kann im serial cue-Format eine positive Auswirkung des effizienten Diagnostizierens auf die Diagnoserichtigkeit beobachtet werden?*

Elstein und Schwarz (2002) zeigen, dass Diagnoserichtigkeit weniger vom strategischen Vorgehen und mehr vom effizienten Umgang mit Informationen abhängt. Daher lauten die Hypothesen wie folgt:

**H3 Studierende, die effiziente Diagnostik im *serial cue*-Format durchführen, erreichen eine höhere *Diagnoserichtigkeit* als Studierende die keine effiziente Diagnostik durchführen.**

**H4 Studierende, die keine ineffiziente Diagnostik im *serial cue*-Format durchführen, erreichen eine höhere *Diagnoserichtigkeit* als Studierende, die ineffiziente Diagnostik durchführen.**

### 2.2.3 Wirken sich verschiedene Falldarbietungsformate für Studierende mit unterschiedlichen Vorwissensstufen positiv auf deren Wissenszuwachs aus?

Wie sich *serial cue*- und *whole case*-Formate auf den Wissenszuwachs auswirken ist bisher unklar (Schmidt & Mamede, 2015). Die optimale Darbietung von Problemen kann einen großen Wissenszuwachs erzeugen (Avrahami et al., 1997; R. M. Hatala, Brooks, & Norman, 2003). Das für eine Vorwissensstufe am besten geeignete Falldarbietungsformat könnte als eine optimale Problemdarbietung fungieren und daher Wissenszuwachs fördern. Angelehnt an die Hypothesen H1 und H2 werden folgende Hypothesen formuliert:

**H5 Studierende mit *höherem Vorwissen* haben im *serial cue*-Format einen größeren Wissenszuwachs als im *whole case*-Format.**

**H6 Studierende mit *niedrigerem Vorwissen* haben im *whole case*-Format einen größeren Wissenszuwachs als im *serial cue*-Format.**



# 3 Material und Methoden

---

## 3.1 STUDIENDESIGN

Die Studie wurde in einem 2 x 2 faktoriellen Design aufgebaut, in dem das Falldarbietungsformat und die Vorwissensstufe als Faktoren variiert wurden. Für eine bildliche Darstellung siehe Abbildung 1. Betreffend des Falldarbietungsformats handelte es sich dabei um eine experimentelle Studie. Betreffend der Vorwissensstufe handelte es sich um eine quasi-experimentelle Studie, da die Vorwissensstufe von den Probanden abhing und nicht randomisiert zugeteilt werden konnte.

Die unabhängigen Variablen waren das Falldarbietungsformat, *serial cue*- beziehungsweise *whole case*-Format und die Vorwissensstufen, niedriges beziehungsweise hohes Vorwissen. Aus dem Design resultierten so vier experimentelle Gruppen, *serial cue*-Format und niedriges Vorwissen, *serial cue*-Format und hohes Vorwissen, *whole case*-Format und niedriges Vorwissen und *whole case*-Format und hohes Vorwissen, siehe auch Abbildung 1.

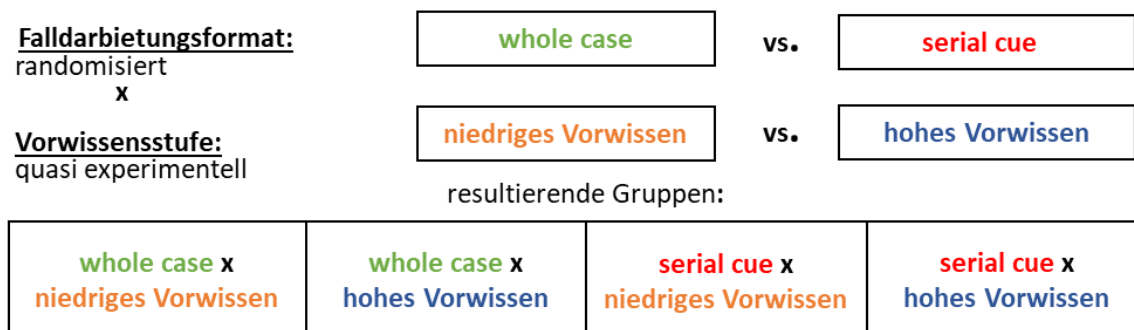


Abbildung 1: 2 x 2 faktorielles Studiendesign und die vier resultierenden experimentellen Gruppen

Die Diagnoserichtigkeit und das prozedurale Wissen wurden in der Studie als abhängige Variablen bestimmt. Es wurde so untersucht, wie die Faktoren Falldarbietungsformat und Vorwissensstufe sich einzeln und in Wechselwirkung auf die Diagnoserichtigkeit auswirken und ob dabei ein unterschiedlicher Lerneffekt auf das prozedurale Wissen entsteht.

Die Diagnoserichtigkeit wurde mittels acht zu lösenden Onlinefällen, in denen virtuelle Patienten diagnostiziert werden mussten, erhoben. Die Vorwissensstufe wurde durch einen deskriptiven Wissenstest bestehend aus Multiple-Choice-Fragen in einem Prätest bestimmt. Zusätzlich wurde deskriptives Wissen in Form von Multiple-Choice-Fragen in einem Abschlusstest bestimmt. Das prozedurale Wissen wurde in Form von Key-Feature Problems sowohl in einem Prätest als auch in einem Posttest abgefragt und verglichen, um einen möglichen Wissenszuwachs zu untersuchen. Der genaue Studienablauf wird in 3.8.3 erklärt, siehe dazu auch Abbildung 5.

## **3.2 STUDIENTEILNEHMER**

### *3.2.1 Voraussetzungen*

Als Probanden dienten Studierende des klinischen Abschnitts der Humanmedizin der Ludwig-Maximilians-Universität München. Es wurden Studierende aus den Semestern Modul 23, Modul 4, Modul 5, Modul 6 und dem praktischen Jahr eingeladen teilzunehmen. Studierende des Modul 23 mussten die Bedingung erfüllen, Modul 23 Teil 1 L6 schon abgeschlossen zu haben, damit ein ausreichendes Basiswissen zu den Themen Fieber und Rückenschmerzen vorausgesetzt werden konnte. Diese Themen waren die Inhalte der Onlinefälle und der Wissensfragen. Eine Erklärung des

Modulsystems des klinischen Abschnitts des Studiengangs Humanmedizin an der Ludwig-Maximilians-Universität findet sich im Anhang, Kapitel 10.1 Anhang 1.

### *3.2.2 Demographie*

Das aktuelle Semester bzw. Modul, das Geschlecht, das Alter, die Muttersprache, die Abiturnote und die Physikumsnote wurden bei der Datenerhebung erfragt.

## **3.3 VORWISSENSSTUFE**

Die Vorwissensstufe wurde durch das Ergebnis der Teilnehmer in einem deskriptiven Wissenstest im Prätest festgelegt. Der Median der erreichten Punktzahl wurde bestimmt und nach diesem die zwei Vorwissensstufen, niedriges und hohes Vorwissen, unterschieden. Siehe dazu auch Kapitel 5.2.3.

## **3.4 PLATTFORM**

Als Plattform zur Erstellung der Onlinefälle und der Messinstrumente, sowie zur Datenerhebung der Pilotierung und der eigentlichen Studie wurde das *e-learning* Portal CASUS® (Version 2017, Instruct AG, München) verwendet. Das von CASUS® entwickelte System ermöglichte es, Onlinekurse anzulegen und diese mit virtuellen Patientenfällen zu füllen. Ein Fall konnte dabei inhaltlich individuell mit Texten, Bildern und Aufgaben gestaltet werden. Hierfür standen sogenannte Karten zur Verfügung, deren Bearbeitungsreihenfolge festgelegt werden konnte. Die Kurse konnten anschließend online bearbeitet werden und sämtliche Daten, die dadurch entstanden, konnten direkt in andere Datensysteme, zum Beispiel in Excel oder SPSS, exportiert werden.

## 3.5 DIAGNOSERICHTIGKEIT

Die Diagnoserichtigkeit wurde als Variable aus der Anzahl der richtig gelösten Onlinefälle aufsummiert. Sie wurde binär codiert (richtig = 1 Punkt, falsch = 0 Punkte). Es wurden keine halben Punkte vergeben.

### 3.5.1 Themen

Inhaltlich behandelten die Onlinefälle die Leitsymptome Fieber oder Rückenschmerzen. Diese Themen wurden aufgrund ihrer hohen praktischen Relevanz und ihren differentialdiagnostischen Möglichkeiten mit den daraus folgenden Konsequenzen für die diagnostische Abklärung ausgewählt. Es sind fachübergreifend relevante Symptome, deren Basisdiagnostik jeder Arzt beherrschen sollte.

### 3.5.2 Fallentwicklung

Alle Onlinefälle wurden von fachkundigen Ärzten überprüft, siehe 3.5.4. Es sollte ein möglichst heterogenes Bild aus verschiedensten Ätiologien und Diagnosen entstehen. Es wurden zwölf Onlinefälle, sechs zum Thema Fieber und sechs zum Thema Rückenschmerzen entwickelt. Als Quellen zu den jeweiligen Krankheitsbildern dienten Gerd Herold und Mitarbeiter Innere Medizin und die duale Reihe Orthopädie und Unfallchirurgie. Für genaue Quellenangaben pro Fall siehe Anhang 10.2. Den Studierenden wurde in den Onlinefällen die Rolle eines Assistenzarztes in einer internistischen Notaufnahme zugedacht, zu dem ein Patient mit Beschwerden kommt. Die Studierenden sollten dann in dieser Rolle den Patienten diagnostizieren. Der Aufbau eines Onlinefalles orientierte sich dabei an der Studie „*Representation scaffolds improve diagnostic efficiency in medical students*“ von Braun et al. (2017). In unserer Studie bestanden die Onlinefälle aus der Anamnese, der körperlichen Untersuchung, einem Standardlabor, jeweiligen Bildgebungen und zusätzlichen

diagnostischen Untersuchungen. Das Standardlabor umfasste Differentialblutbild, CRP, Elektrolytwerte, Blutglucosespiegel, Kreatinin, GOT, GPT, yGT, aP, Billirubin, LDH, Quick, INR und PTT. Für eine genaue Aufzählung der jeweils möglichen diagnostischen Schritte und eine genaue Literaturangabe pro Onlinefall siehe Tabelle der Onlinefälle mit zur Verfügung stehender Diagnostik im Anhang 10.2. Wichtig ist, dass die Onlinefälle mehr diagnostische Schritte enthielten als nötig, das heißt, es waren auch irrelevante diagnostische Untersuchungen gegeben.

Inhaltlich entsprachen die Onlinefälle im *whole case*-Format eins zu eins denen im *serial cue*-Format. Ein *serial cue*-Fall begann stets mit der Anamnese und der anschließenden Auswahlleiste für die nächsten diagnostischen Untersuchungen. Ein *whole case*-Fall begann mit der Anamnese, der körperlichen Untersuchung und dem Standard Labor, anschließend folgte die jeweilige weitere Diagnostik in zufällig gewählter Reihenfolge.

FrauHoffmann Jung Valentina Gehe zu Hilfe

Gehe zu: Top

**Anamnese**

Es kommt die **36-jährige Frau Hoffmann** zu Ihnen. Seit fast einer Woche bestehe ein fiebriger grippaler Infekt, ohne Anschein der Besserung. Zusätzlich bestünden Abgeschlagenheit, Appetitverlust und Übelkeit. Ebenso Durchfall. Keiner in der näheren Umgebung zeige die gleichen Symptome. Sie gibt an, keine Gelenkschmerzen zu haben. Sie habe keine Veränderung der Stuhl- und Urinfarbe bemerkt. Es seien keine Vorerkrankungen bekannt. Sie sei vor 20 Jahren nach einem Handgelenksbruch operiert worden. Medikamente nehme sie keine ein, bis auf homöopathische Globuli und homöopathische Tropfen bei Bedarf. Es sei eine Penicillin- und Nickelallergie bekannt. Sie sei Nichtraucherin und es bestünde gelegentlicher Alkoholkonsum. Eine Schwangerschaft sei ausgeschlossen. Sie sei vor einem Monat von einer Rundreise aus Costa Rica zurückgekehrt. Keine der empfohlenen Reiseimpfungen seien davor durchgeführt worden.

Bitte wählen Sie Ihren nächsten diagnostischen Schritt aus.

**HINWEIS:** Sie brauchen nicht jede hier aufgelistete Diagnostik, um die richtige Diagnose stellen zu können. Die Reihenfolge in der Navigation ist willkürlich gewählt.

Körperliche Untersuchung Standard Labor Blutkulturen Dicker Tropfen und Blutausschrieb EKG HIV-Test Hepatitis-Serologie

Echokardiographie Röntgen Thorax Diagnose eingeben

Experte Vorschau schließen

Abbildung 2: Darstellung eines Onlinefalles im *serial cue*-Format in Casus®

FrauHoffmann Jung Valentina Gehe zu Hilfe

Gehe zu: Top

Es kommt die **36-jährige Frau Hoffmann** zu Ihnen. Seit fast einer Woche bestehe ein fiebriger grippaler Infekt, ohne Anschein der Besserung. Zusätzlich bestünden Abgeschlagenheit, Appetitverlust und Übelkeit. Ebenso Durchfall. Keiner in der näheren Umgebung zeige die gleichen Symptome. Sie gibt an, keine Gelenkschmerzen zu haben. Sie habe keine Veränderung der Stuhl- und Urinfarbe bemerkt. Es seien keine Vorerkrankungen bekannt. Sie sei vor 20 Jahren nach einem Handgelenksbruch operiert worden. Medikamente nehme sie keine ein, bis auf homöopathische Globuli und homöopathische Tropfen bei Bedarf. Es sei eine Penicillin- und Nickelallergie bekannt. Sie sei Nichtraucherin und es bestünde gelegentlicher Alkoholkonsum. Eine Schwangerschaft sei ausgeschlossen. Sie sei vor einem Monat von einer Rundreise aus Costa Rica zurückgekehrt. Keine der empfohlenen Reiseimpfungen seien davor durchgeführt worden.

Vitalparameter: Körpertemperatur 38,4°C, Herzfrequenz 93 Schläge/min, Blutdruck 115/75mmHg.


Allgemeinzustand: Patientin wach und orientiert, freundlich und zugewandt, Normalgewicht mit 65kg bei 1,70m. BMI von 22,5. Skleren nicht verfärbt, Schleimhäute rosig. Haut gebräunt. Weitere Inspektion unauffällig.

Respiratorisches System: Beidseits sonore Klopfschall über der Lunge. Lungengrenzen atemverschieblich. Atemgeräusch beidseits vesikulär. Keine Zyanosezeichen.

Kardiovaskuläres System: Herztöne regelrecht auskultierbar. Keine Herzgeräusche vorhanden. Keine Strömungsgeräusche. Keine gestauten Jugularvenen.

Gastrointestinales System: Lebhaftes Darmgeräusche über allen vier Quadranten. Keine Flankendämpfung. Bauchdecke weich und ohne Abwehrspannung. Schmerzen bei Palpation im rechten Oberbauch. Keine Resistenzen und kein Milzrand palpabel. Leber perkutorisch in MCL 14 cm. Kein Nierenklopfschmerz und kein Klopfschmerz an der LWS.

Lymphknoten: Cervikal, axillar und inguinal keine vergrößerten Lymphknoten tastbar.



Experte Vorschau schließen

Abbildung 3: Darstellung eines Onlinefalles im *whole case*-Format in Casus®

### *3.5.3 Diagnoseeffizienz und Diagnoseineffizienz*

Die Anzahl der ausgewählten diagnostischen Schritte im *serial cue*-Format wurde in zwei separaten, binär codierten Variablen verwendet, Diagnoseeffizienz und Diagnoseineffizienz.

Die Voraussetzung für die Diagnoseeffizienz war, dass nur die für die Diagnosefindung nötigen diagnostischen Untersuchungen in einem Onlinefall angewandt worden waren. Wurden mehr oder weniger Untersuchungen angewandt, lag keine Diagnoseeffizienz vor. Die Diagnoseeffizienz wurde dann als Summe der Onlinefälle berechnet (nur die nötigen diagnostischen Schritte im Onlinefall angewandt = 1 Punkt, mehr oder weniger diagnostische Schritte angewandt = 0 Punkte). Ein Assistenzarzt der Inneren Medizin an der Medizinischen Klinik und Poliklinik IV Klinikum der Universität München hatte festgelegt, welche diagnostischen Schritte für die Diagnosefindung nötig waren. Der Median der Punktzahl der Diagnoseeffizienz wurde bestimmt und nach diesem in zwei Gruppen unterschieden, hohe und niedrige Diagnoseeffizienz.

Diagnoseineffizienz war gegeben, wenn alle zur Verfügung stehenden diagnostischen Schritte in einem Onlinefall ausgewählt worden waren. Die Diagnoseineffizienz wurde als Summe aller angewandten diagnostischen Schritte pro Onlinefall berechnet (alle diagnostischen Schritte angewandt = 1 Punkt, nicht alle diagnostische Schritte angewandt = 0 Punkte). Der Median der Punktzahl wurde bestimmt und nach diesem in hohe und niedrige Diagnoseineffizienz unterteilt.

### *3.5.4 Review der Onlinefälle*

Eine Fachärztin für Allgemeinmedizin am Institut für Didaktik und Ausbildungsforschung in der Medizin am Klinikum der Universität München und ein

Assistenzarzt der Inneren Medizin an der Medizinischen Klinik und Poliklinik IV Klinikum der Universität München haben die Onlinefälle auf deren Inhalt und Form überprüft.

### *3.5.5 Bildmaterial*

Die verwendeten Bildmaterialien wurden realen Patientenfällen entnommen und anonymisiert. Die EKGs wurden von einem leitenden Oberarzt der Medizinischen Klinik und Poliklinik IV Klinikum der Universität München zur Verfügung gestellt und den Onlinefällen entsprechend zugeordnet. Das radiologische Bildmaterial wurde von einer Assistenzärztin der Klinik und Poliklinik für Radiologie der Universität München passend zum jeweiligen virtuellen Patienten ausgewählt.

### *3.5.6 Auswahl*

In der Studie sollten letztendlich acht Onlinefälle, vier Onlinefälle pro Symptom, verwendet werden. Aus den zwölf zur Verfügung stehenden Onlinefällen wurden die acht Besten ausgewählt. So wurde der Fall eines Bandscheibenprolapses als zu leicht aussortiert, da er von allen Pilotteilnehmern richtig gelöst worden war. Der Fall einer Malaria tertiana Infektion war von keinem richtig gelöst worden und wurde folglich als zu schwer aussortiert. Den übrigen zehn Onlinefällen wurden Nummern von eins bis zehn zugewiesen, welche dann auf kleine zwei mal zwei Zentimeter große, weiße Papierstücke geschrieben wurden. Nach zweimaligem Falten wurden nacheinander acht blind gezogen. So wurden aus zehn übrigen Onlinefällen acht für die Studiendurchführung ausgelost. Die Onlinefälle wurden also nach den Ergebnissen der Pilotierung ausgewählt und dann ausgelost.

Die Ermittlung der Reihenfolge der Fälle erfolgte ebenso zufällig. Alle übrigen acht Onlinefälle bekamen Nummern von eins bis acht zugeordnet und es wurde auf die



gleiche Art gelöst. Der erste gezogene Fall war Fall Nummer eins, der nächste Fall Nummer zwei und so folgend. Die Reihenfolge der in der Studie verwendeten Onlinefälle ist in Tabelle 1 dargestellt.

Fallnummer	Diagnose
1	Akute Hepatitis-A-Infektion
2	Hyperthyreosis factitia
3	Osteoporotische Wirbelkörperfraktur
4	Spondylitis ankylosans
5	Bakterielle Endokarditis
6	Lumbago
7	Spinalkanalstenose
8	HIV-Infektion

Tabelle 1: Reihenfolge der Onlinefälle in der Studiendurchführung

## 3.6 MESSINSTRUMENTE

Die Messinstrumente wurden von einer Fachärztin für Allgemeinmedizin am Institut für Didaktik und Ausbildungsforschung in der Medizin am Klinikum der Universität München überprüft.

### 3.6.1 Test des deskriptiven Wissens

Deskriptives Wissen zu den Themen Fieber und Rückenschmerzen wurde durch Multiple-Choice Fragen erhoben. Für die Studie wurde ein Single-Choice Verfahren verwendet, bei dem eine richtige Antwort aus fünf möglichen Antworten auszuwählen war. Pro richtige Antwort gab es einen Punkt. Insgesamt wurden 20 Multiple-Choice-Fragen in der Studie abgefragt, zehn pro Thema. Als Quellen fungierten abermals Gerd Herold und Mitarbeiter Innere Medizin und die duale Reihe Orthopädie und Unfallchirurgie. Brendels Anleitung zur Überprüfung von Multiple-Choice-Fragen wurde zur Qualitätskontrolle angewandt (Brendel, 2017). Die 20 Fragen wurden in

jeweils zehn Fragen für den Prä- und Posttest aufgeteilt. Um möglichst gleiche Bedingungen im Prä- und Posttest zu erreichen, wurden die Multiple-Choice-Fragen nach den Ergebnissen der Pilotierung gepaart, siehe auch Kapitel 3.7. Die Schwierigkeit der Fragen im Prätest entsprach so ungefähr der im Posttest.

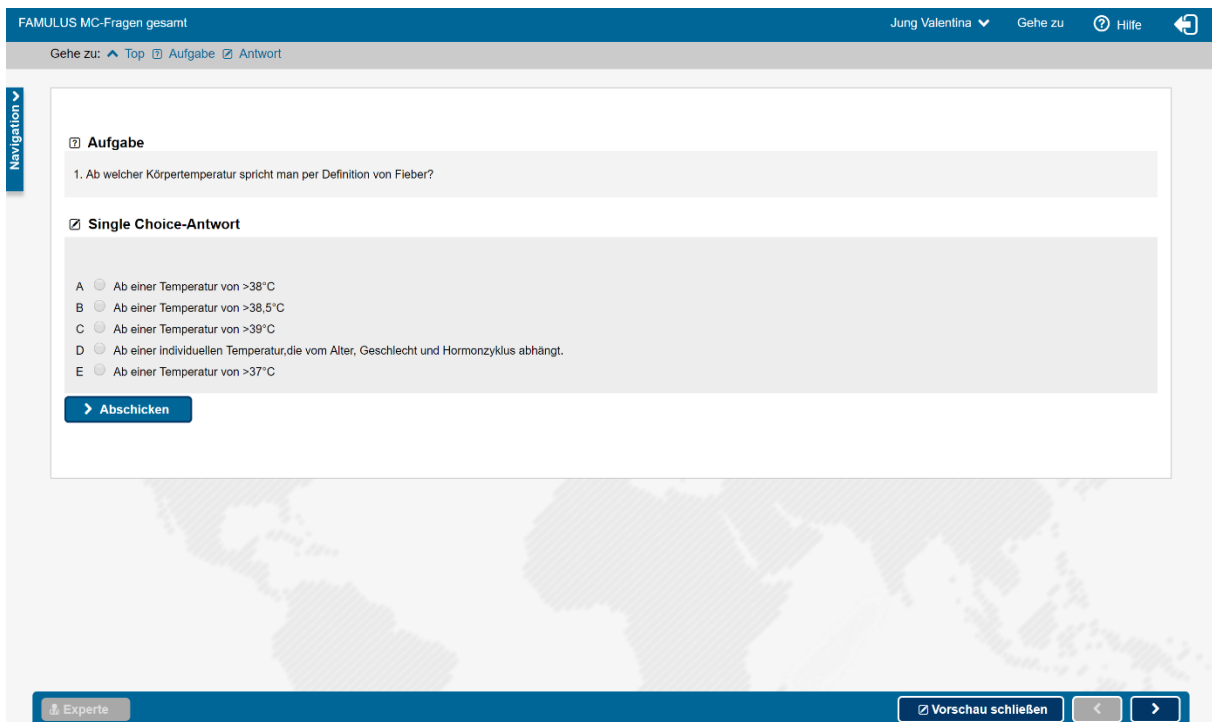


Abbildung 4: Darstellung einer Multiple-Choice-Frage in Casus®

Prätest	Erfolgsquote	Posttest	Erfolgsquote
1	0,13	10	0,14
2	0,63	19	0,71
3	0,29	4	0,26
5	1	12	0,86
6	0	20	0
7	0,4	9	0,14
8	0,72	17	0,72
11	0,29	16	0,29
13	0,86	18	0,86
14	0,57	15	0,57

Tabelle 2: Paarung der Fragen im Prä- und Posttest nach Schwierigkeit (die Schwierigkeit entspricht der Erfolgsquote) und Nummern der Fragen nach der Reihenfolge der Erstellung

### 3.6.2 Test des prozeduralen Wissens

Prozedurales Wissen wurde anhand von Key-Feature Problems erfragt. „Key Feature werden [...] als kritische Entscheidungen definiert, die getroffen werden müssen, um ein klinisches Problem zu lösen.“ (Kopp et al. 2006, S.2) . Ein Key-Feature Problem konzentriert sich dabei auf einen fehleranfälligen Schritt im Umgang mit einem Problem in einer zu lösenden Aufgabe (Page, Bordage, & Allen, 1995). Ein Key-Feature Problem besteht aus einer kurzen Patientenvignette und maximal drei Fragen dazu (Kopp et al., 2006). Der Inhalt thematisierte wieder die Leitsymptome Fieber und Rückenschmerzen. Die Entwicklung erfolgte nach der achtschrittigen Anleitung zur Entwicklung eines Key-Feature Problems von Kopp et al. (2006). Der Kontext wurde als Freitagnachmittag in einer Notaufnahme festgelegt, um somit die Notwendigkeit einer Entscheidung des Bearbeiters zu unterstreichen. Je nach Fallvignette handelte es sich um eine klinische Bagatelle oder eine klinisch ernste bzw. lebensbedrohliche Situation. Als Key-Features sollten die wichtigsten Differentialdiagnosen genannt,

weitere Untersuchungen ausgewählt und das weitere Vorgehen geplant werden. Die klinische Situation, die Rolle des Bearbeiters, der Name, das Geschlecht und das Alter des Patienten, dessen Symptome und seit wann diese bestehen, sowie Zusatzinformationen zum Patienten wurden beschrieben. Die Key-Feature Fragen wurden wie folgt festgelegt:

- An welche wichtigen Differentialdiagnosen denken Sie?
- Welche gezielten Untersuchungen führen Sie durch, bzw. leiten Sie weiter ein?
- Wie gehen Sie weiter mit dem Patienten/ der Patientin vor?

Es wurde das Short Menu als Antwortformat gewählt. Als Anmerkung zur Umsetzung sollte hier erwähnt werden, dass in CASUS® das Long Menu von der Funktion dem eines Short Menus im Sinne von Kopp et al. (2006) entspricht. Die Maximalpunktzahl pro Frage wurde auf 1,0 festgelegt und pro Antwortmöglichkeit wurden Teilpunkte vergeben. Pro Key-Feature Problem waren so bei drei Fragen 3,0 Punkte maximal erreichbar und bei acht Key-Feature Problems insgesamt 24,0 Punkte erreichbar. Review und Korrektur der Fallvignetten und Antwortmöglichkeiten geschah durch eine Fachärztin für Allgemeinmedizin am Institut für Didaktik und Ausbildungsforschung in der Medizin am Klinikum der Universität München. Die Key-Feature Problems wurden wie von Kopp et al. (2006) empfohlen pilotiert, siehe Kapitel 3.7. Alle acht Key-Feature Problems wurden als ein zusammenhängender Test pilotiert. Ähnlich den Multiple-Choice-Fragen wurde auch hier eine Paarung, resultierend aus den Ergebnissen der Pilotierung vorgenommen. Es wurden so vier Key-Feature Problems dem Prätest zugeordnet und vier dem Posttest. Die Schwierigkeit des Prätests entsprach so ungefähr der des Posttests.

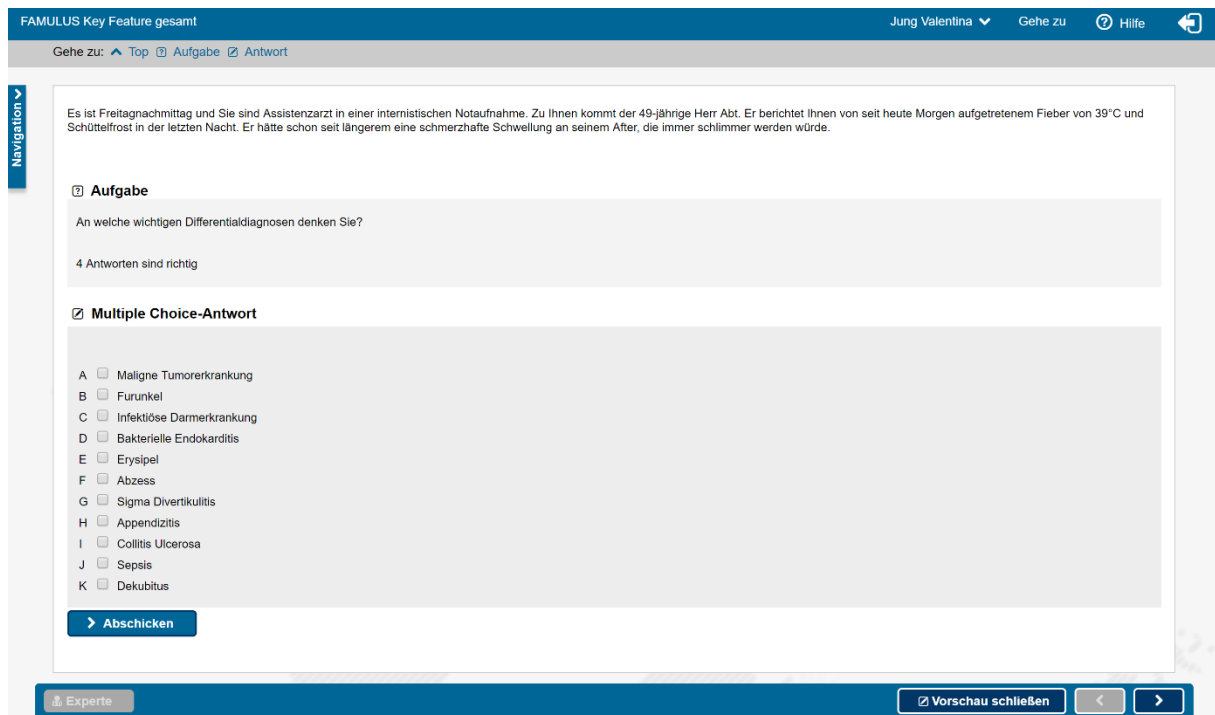


Abbildung 5: Darstellung eines Key-Feature Problems in Casus®

Prätest	Erfolgsquote	Posttest	Erfolgsquote
<b>Key-Feature Problem 1</b> (Abszess mit Sepsisverdacht)	MW: 0,68	<b>Key-Feature Problem 5</b> (Bakterielle Meningitis)	MW: 0,63
<b>Key-Feature Problem 2</b> (Grippaler Infekt)	MW: 0,79	<b>Key-Feature Problem 4</b> (Lumbago)	MW: 0,79
<b>Key-Feature Problem 3</b> (Cauda-Syndrom)	MW: 0,76	<b>Key-Feature Problem 6</b> (Gastroenteritis)	MW: 0,72
<b>Key-Feature Problem 7</b> (Spinalkanalstenose)	MW: 0,71	<b>Key-Feature Problem 8</b> (Herpes Zoster)	MW: 0,67

Tabelle 3: Zuordnung der Key-Feature Problems in Prä- und Posttest nach Erfolgsquote und Nummern nach der Reihenfolge der Erstellung

### 3.7 PILOTIERUNG

Die Pilotierung der Onlinefälle, des deskriptiven Wissenstests und des prozeduralen Wissenstests erfolgte online durch acht freiwillige Teilnehmer im Zeitraum vom 14.08.2017 bis zum 01.09.2017. Daran nahmen sechs Medizinstudierende der Ludwig-Maximilians-Universität und zwei Ärztinnen im ersten Jahr der Assistenzarztausbildung teil. Die Aufgabenreihenfolge des Pilotkurses entsprach der folgenden: Erst sollten die Multiple-Choice-Fragen, dann die Key-Feature Problems und zum Schluss die Onlinefälle bearbeitet werden. Sie entsprach somit nicht der Aufgabenreihenfolge der späteren Studie, siehe 3.8.3, da die Key-Feature Problems und die Multiple-Choice-Fragen nach den Ergebnissen der Pilotierung in einen Prä- und Posttest aufgeteilt wurden, siehe auch 3.6.1 und 3.6.2. Die Teilnehmer bekamen einen zufälligen Zugang zu den Onlinefällen im *whole case*- oder im *serial cue*-Format. Zwei Studierende beendeten die Pilotierung nicht vollständig, daher gehen in die Auswertung nur insgesamt sechs Teilnehmer mit ein. Von diesen sechs Teilnehmern bearbeiteten die eine Hälfte die Onlinefälle im *serial cue*-Format und die Andere im *whole case*-Format. Für eine vollständige Bearbeitung der Onlinefälle gab es ein Entgelt von 30,00 €. Ausgewertet wurde die Pilotierung mit IBM SPSS Statistics (Version 24,0, SPSS Inc., Chicago, Illinois). Durch die ermittelten Ergebnisse der Pilotteilnehmer konnte dem deklarativen Wissenstest eine Schwierigkeit basierend auf der Erfolgsquote zugeordnet werden. Sie wurde berechnet aus Anzahl richtiger Antworten subtrahiert durch Anzahl der Antworten gesamt. Die Erfolgsquote des prozeduralen Wissenstests wurde durch den Mittelwert der jeweils drei Fragen festgelegt. Im Mittel brauchten die Teilnehmer 3 Stunden, um die Onlinefälle vollständig zu bearbeiten. Teilnehmer der Pilotierung wurden als Probanden für die eigentliche Studiendurchführung ausgeschlossen.

## 3.8 STUDIENDURCHFÜHRUNG

### 3.8.1 Ethische Unbedenklichkeit

Die Ethikkommission der medizinischen Fakultät der Ludwig-Maximilians-Universität hatte der Studiendurchführung zugestimmt (Code 17-249).

### 3.8.2 Lernumgebung und Räumlichkeiten

Die Datenerhebung erfolgte vom 25.10.2017 bis zum 15.11.2017. Als Erhebungsort fungierten der Seminarraum 4 und 5 der Klinik für Allgemeine, Unfall- und Wiederherstellungschirurgie, Nußbaumstraße 20, 80336 München. Es erfolgten 23 Erhebungen mit minimal einem Teilnehmer und maximal 15 Teilnehmern.

### 3.8.3 Studienablauf

Es fand eine Gruppenerhebung am Einzelnen statt. Das bedeutete man erhob gleichzeitig mehrere Einzelpersonen. Vor Eintreffen der Studienteilnehmer wurden die Computer hochgefahren und die Einwahlseite [famulus.casus.net](http://famulus.casus.net) aufgerufen. Als Internetbrowser wurde nach Empfehlung der Instruct AG Mozilla Firefox verwendet.

Nach einer kurzen Begrüßung und Einführung begann die eigentliche Datenerhebung. Die Studierenden wählten sich mit einem anonymen Pin, welchen sie vorher zufällig gezogen hatten, unter der Einwahlseite ein. Darauf erfolgte eine automatische Weiterleitung zum jeweiligen Kursformat mit Patientenfällen im *serial cue*- oder *whole case*- Format. Die Teilnehmer absolvierten der Reihenfolge nach einen Prätest, acht Onlinefälle und einen Posttest. Der genaue Ablauf der Studie wird in Abbildung 5 bildlich dargestellt. Der Prä- und Posttest war in beiden Kursformaten gleich gestaltet. Die Probanden hatten drei Stunden Zeit, den Kurs zu bearbeiten. Eine 15-minütige Pause fand nach eineinhalb Stunden Bearbeitungszeit statt. Die Reihenfolge der Patientenfälle war, wie schon in 3.5.6 erwähnt, zufällig gewählt.

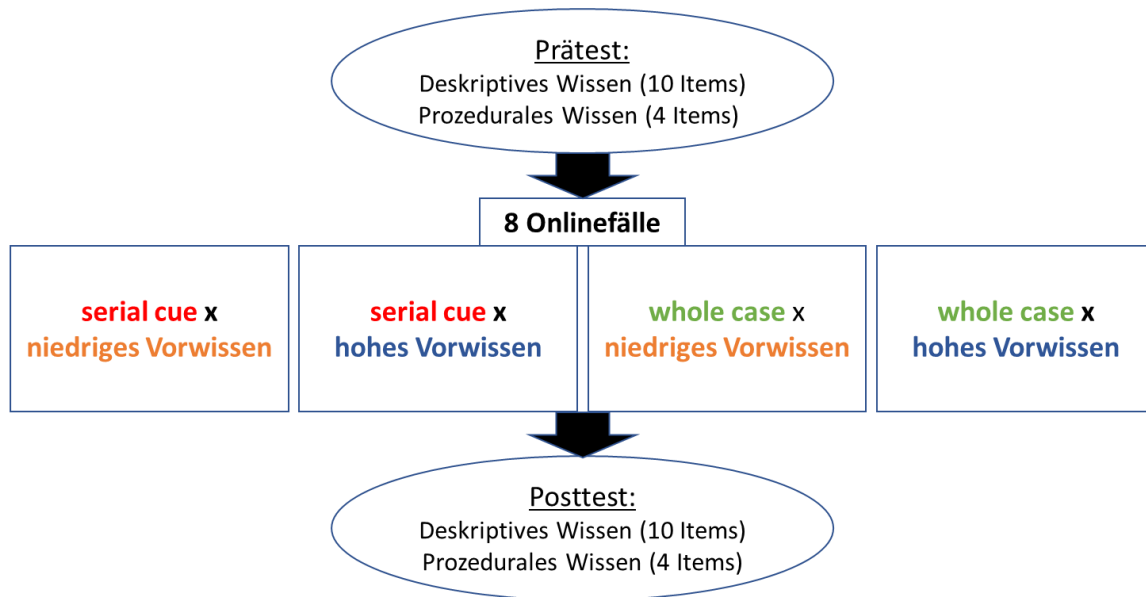


Abbildung 6: Bildliche Darstellung des Studienablaufs

### 3.8.4 Technische Ausstattung

Bei den verwendeten Computern handelte es sich um Intel® Core™ i5-2400 CPU 3.10 GHz Prozessoren mit 8,00 GB Arbeitsspeicher und dem Betriebssystem Microsoft Windows, Edition Windows 7. Bei den zugehörigen Bildschirmen handelte es sich um LG Flatron E2210 22 Zoll 16:10 Widescreen Bildschirme.

## 3.9 AUFARBEITUNG UND ANALYSE DER STATISTISCHEN DATEN

### 3.9.1 Aufarbeitung der statistischen Daten

Die ermittelten Daten wurden am letzten Erhebungstag direkt von CASUS® als Excel Tabelle des jeweiligen Kurses exportiert und zu einem Datensatz in die Software IBM



SPSS Statistics (Version 24.0, SPSS Inc. Chicago, Illinois) importiert, welche für alle statistischen Berechnungen genutzt wurde.

Die Skalenniveaus entsprachen minimal der Nominalskala (Geschlecht, Modul, Muttersprache, Fallformat, Vorwissensstufe), im Mittel der Intervallskala (Abiturnote, Physikumsnote, berechnete Mittelwerte) und maximal der Verhältnisskala (Alter, Diagnoserichtigkeit, Multiple-Choice-Fragen Prätest, Multiple-Choice-Fragen Posttest, Key-Feature Problems Prätest, Key-Feature Problems Posttest, Diagnoseeffizienz, Diagnoseineffizienz, Bearbeitungszeit). Eine Zusammenfassung der Variablendefinitionen und Berechnungsvorschriften findet sich in Tabelle 4. Eine gesonderte Skalenniveauprüfung ist nicht erfolgt. Listenweiser Fallausschluss wurde angewandt, um fehlende Werte aus der Analyse zu entfernen. In der Datentabelle wurden diese mit 999 bezeichnet. Aufgrund limitierter technischer Möglichkeiten wurde bei 186 Diagnoseeingaben nachträglich sichergestellt, dass die von den Studierenden eingegebene Diagnose einer kodierten Diagnose entsprach und ggf. Rechtschreibfehler ausgeräumt. In 49 Fällen wurden mehrere Diagnosen angegeben. Hier wurde ein strenger Auswertungsmodus gewählt, sodass diese Antworten alle als falsch gewertet wurden, selbst wenn sich die richtige Diagnose darunter befand.

Einzelne Antwortmöglichkeiten des Tests des prozeduralen Wissens wurden nachkorrigiert (KF1\_pre\_2 -.33, KF2\_pre\_2 -.16, KF3\_pre3 x.5 und KF2\_pre\_3 als zu schwer aus der Wertung genommen; KF7\_post\_1 als zu leicht aus der Wertung genommen).

Item	Skala	Definition	minimal	maximal
Geschlecht	Nominal	0 = männlich 1 = weiblich	-	-
Modul	Nominal	0 = 23 1 = 4 2 = 5 3 = 6 4 = PJ	-	-
Muttersprache	Nominal	0 = deutsch 1 = nicht deutsch	-	-
Fallformat	Nominal	1 = serial cue 2 = whole case	-	-
Vorwissensstufe	Nominal	0 = niedriges Vorwissen 1 = hohes Vorwissen	-	-
Abiturnote	Intervall	1,0 (sehr gut) bis 4,0 (ausreichend)	1,0	4,0
Physikumsnote	Intervall	1,0 (sehr gut) bis 4,0 (ausreichend)	1,0	4,0
Alter	Verhältnis	Anzahl Lebensjahre	0	-
Diagnoserichtigkeit (8 Onlinefälle)	Verhältnis	Summe richtig gestellter Diagnosen	0	8
Multiple-Choice-Fragen Prätest (10 Items)	Verhältnis	Summe richtig gelöster Aufgaben	0	10
Multiple-Choice-Fragen Posttest (10 Items)	Verhältnis	Summe richtig gelöster Aufgaben	0	10

Item	Skala	Definition	minimal	maximal
Key-Feature Problems Prätest  (4 Items)	Verhältnis	Summe richtig gelöster Aufgaben	0	12
Key-Feature Problems Posttest  (4 Items)	Verhältnis	Summe richtig gelöster Aufgaben	0	12
Bearbeitungszeit	Verhältnis	Summe Minuten	0	180
Diagnoseeffizienz im <i>serial cue</i>  (8 Onlinefälle)	Verhältnis	0 = mehr als die nötigen angewandt  1 = die nötigen angewandt	0	8
Diagnoseineffizienz im <i>serial cue</i>  (8 Onlinefälle)	Verhältnis	0 = nicht alle angewandt  1 = alle angewandt	0	8

Tabelle 4: Variablendefinition und Berechnungsvorschriften, Tabelle zweigeteilt über S.34 und 35

### *3.9.2 Analyse der statistischen Daten*

Es erfolgten vorgeordnete Analysen der Daten, siehe dazu 4.1. Der Median der Ergebnisse des deklarativen Wissenstests wurde bestimmt und wie in Kapitel 3.3 erwähnt zur Einteilung der Vorwissensstufen verwendet. Eine Testung auf Normalverteilung der Residuen der gemessenen Variablen wurde durchgeführt. Das Signifikanzniveau wurde bei mehrfacher Testung adjustiert. Die Stichprobeneigenschaften wurden deskriptiv dargestellt und eine mögliche Konfundierung mit einer univariaten Varianzanalyse berechnet. Die Antworten der demographischen Fragen wurden mittels einer deskriptiven Statistik dargestellt und der Mittelwert des Alters wurde berechnet. Eine deskriptive Statistik für die Hauptskalen wurde erstellt. Der Median der Diagnoseeffizienz und Diagnoseineffizienz wurde berechnet und nach diesem in jeweils zwei Gruppen unterschieden. Mögliche korrelative Zusammenhänge und Einflüsse der Kontrollvariablen wurden nach Pearson berechnet.

### *3.9.3 Hypothesentestung*

Zur Überprüfung von H1 und H2 wurde eine univariante Varianzanalyse mit der abhängigen Variable Diagnoserichtigkeit und dem Falldarbietungsformat und der Vorwissensstufe berechnet.

H3 und H4 wurden durch eine univariante Varianzanalyse der abhängigen Variable Diagnoserichtigkeit mit der Diagnoseeffizienz und der Diagnoseineffizienz überprüft.

Eine einstufige messwiederholte Varianzanalyse der Mittelwerte des prozeduralen Wissens des Prä- und Posttests wurde mit dem Falldarbietungsformat und der Vorwissensstufe berechnet, um H5 und H6 zu überprüfen.

# 4 Ergebnisse

---

## 4.1 VORGEORDNETE ANALYSEN

### 4.1.1 Einteilung der Vorwissensstufe

Der Median der Ergebnisse des deskriptiven Wissenstests im Prätest ergab den Punktwert 4.00. Als niedriges Vorwissen galten Ergebnisse unter 4.00 (59 Teilnehmer, 41.5%) und als hohes Vorwissen galten Ergebnisse ab 4.00 (83 Teilnehmer, 58.5%).

### 4.1.2 Test auf Normalverteilung der Residuen

Ein Kolmogorov-Smirnov-Test wurde für die Residuen der abhängigen Variablen Diagnoserichtigkeit und Prozedurales Wissen im Prä- und Posttest berechnet, um zu überprüfen, ob die Daten normalverteilt waren. Die Residuen der Ergebnisse des prozeduralen Wissens im Prätest (Kolmogorov-Smirnov  $p = .20$ ) und im Posttest (Kolmogorov-Smirnov  $p = .20$ ) waren normalverteilt. Die Residuen der Diagnoserichtigkeit wiesen keine Normalverteilung auf (Kolmogorov-Smirnov  $p=.004$ ). Betrachtete man jedoch die Residuen der Diagnoserichtigkeit im Q-Q-Diagramm waren keine großen Abweichungen zwischen den Verteilungen aufgetreten, siehe Abbildung 6. Das bedeutet die Verteilungen waren annähernd linear. In einem ergänzend durchgeführten Boxplot fielen dazu passend keine extremen Ausreißer auf, die über zwei Standardabweichungen lagen, siehe Abbildung 7. Die Residuen der Diagnoserichtigkeit konnte daher annähernd als normalverteilt betrachtet werden und in parametrischen Tests verarbeitet werden. Ein Mitarbeiter des Instituts für medizinische Informationsverarbeitung, Biometrie und Epidemiologie an der Ludwig-Maximilians-Universität München hatte die Residuenverteilung zusätzlich begutachtet und keine Einwände gegen parametrische Testungen erhoben.

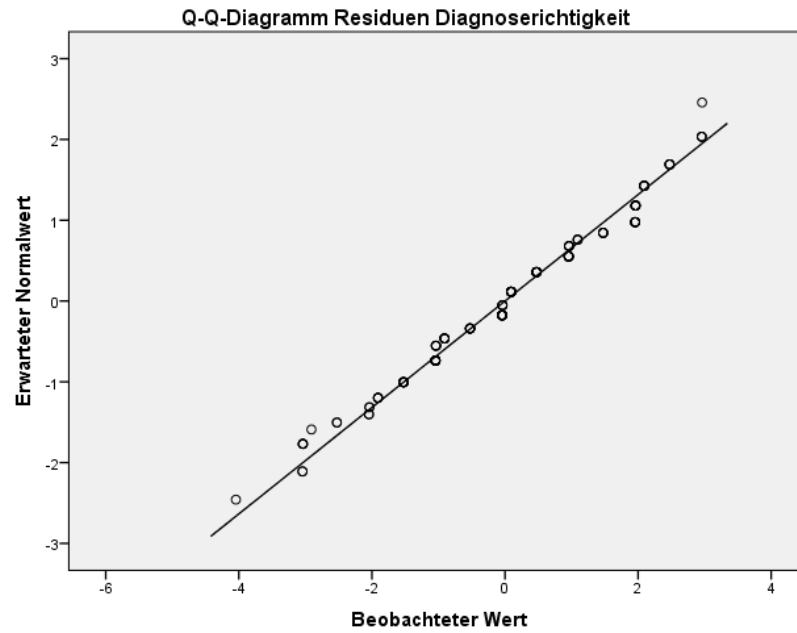


Abbildung 7: Q-Q-Diagramm der Residuen der Diagnoserichtigkeit, Verteilungen annähernd linear

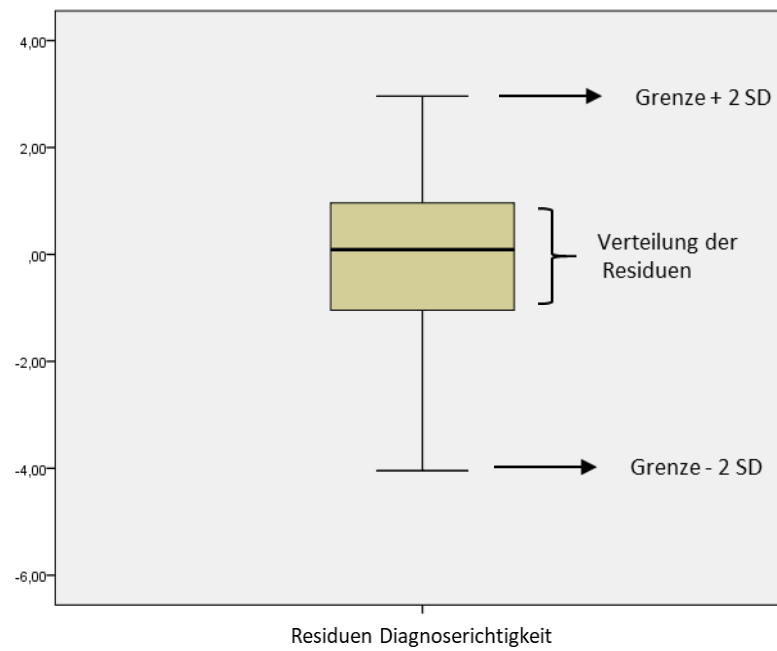


Abbildung 8: Boxplots der Residuen der Diagnoserichtigkeit, keine Werte über zwei Standardabweichungen, SD = Standardabweichung

#### 4.1.3 Adjustierung der Signifikanzniveaus

Das Alpha-Fehlerniveau ( $\alpha$ ) war konventionsgemäß auf 5% festgelegt. Bei 2-maliger Testung der Stichproben bezogen auf die Diagnoserichtigkeit wurde die Bonferroni-Korrektur zur Adjustierung des p-Werts angewandt. P-Werte von kleiner gleich .025 wurden daher bei der Testung von H1, H2, H5 und H6 als signifikant erachtet.

#### 4.1.4 Stichprobeneigenschaften

Es nahmen 142 Personen teil, davon liegen 135 (95 %) komplette Datensätze vor. Das Geschlechterverhältnis entsprach weiblich zu männlich gleich 2.6 zu 1 (N weiblich = 102 (72%), N männlich = 40 (28%)). Minimal waren die Teilnehmer 21 Jahre alt und maximal 35 Jahre alt. Der Mittelwert des Alters ergab 24.4 Jahre (SD = 2.90). 121 Teilnehmer (85%) gaben Deutsch als ihre Muttersprache an, 18 Teilnehmer (13%) gaben nicht Deutsch als ihre Muttersprache an und 3 Teilnehmer (2%) gaben an zweisprachig zu sein. Die Abiturnote lag im Durchschnitt bei 1.6 (SD = .56) und die Physikums Note bei 2.6 (SD = .85). 71 Teilnehmer (50%) bearbeiteten die Patientenfälle im *whole case*-Format und 71 Teilnehmer im *serial cue*-Format.

Eine multivariante Varianzanalyse ergab keine signifikanten Unterschiede der vier experimentellen Gruppen, betreffend das Alter, das Geschlecht, die Muttersprache, die Abiturnote, die Physikumsnote, das Modul und den Mittelwert der Bearbeitungszeit ( $F(21,399) = .74, p = .79, n.s.$ ). Das heißt hinsichtlich dieser demographischen Variablen war die Zusammensetzung der Gruppen ungefähr gleich. Tabelle 7 zeigt die Stichprobeneigenschaften der vier experimentellen Gruppen. Hinsichtlich des prozeduralen Wissens im Prätest traten in einer univariaten Varianzanalyse keine signifikanten Unterschiede in den Gruppen auf ( $F(3,138) = .76, p = .52, n.s.$ ). Das bedeutet im Prätest hatten alle vier experimentellen Gruppen ungefähr gleiche prozedurale Wissenswerte erreicht.

#### *4.1.5 Einfluss demographischer Variablen*

Um eine mögliche Konfundierung der Haupteffekte im Vorfeld auszuschließen, wurden Effekte der demographisch erhobenen Variablen (Geschlecht, Alter, Muttersprache, Abiturnote, Physikumsnote, Semester/ bzw. Modul) auf die abhängige Variable Diagnoserichtigkeit mit einer univariaten Varianzanalyse berechnet. Die Muttersprache wirkte als einzige Variable einen signifikanten Effekt aus ( $F(2,105) = 4.32$ ,  $p = .02$ ,  $\eta^2 = .08$ ). Eine Sichtung der Daten zeigte, dass Probanden mit der Muttersprache Deutsch eine signifikant höhere Diagnoserichtigkeit erreichten als nicht Muttersprachler. Auf die Ergebnisse dieser Studie hatte dies jedoch keinen Effekt, da sich wie oben in 4.1.4 erwähnt die Zusammensetzung der Studiengruppen hinsichtlich der demographischen Variablen, einschließlich der Muttersprache, nicht signifikant unterschied, siehe 4.1.4 und Tabelle 5.



	<b>serial cue x</b> <b>hohes Vorwissen</b>	<b>serial cue x</b> <b>niedriges vorwissen</b>	<b>whole case x</b> <b>hohes Vorwissen</b>	<b>whole case x</b> <b>niedriges vorwissen</b>
<b>N Teilnehmer</b> (Prozent)	38 (26,7 %)	33 (23,2 %)	45 (31,7 %)	26 (18,3 %)
<b>Geschlecht</b>				
N männlich	13	8	13	6
N weiblich	25	25	32	20
<b>MW Alter</b>	24.53	25.12	24.33	23.46
SD	3.27	3.49	2.24	2.34
<b>Muttersprache</b>				
N deutsch	32	28	41	23
N nicht-deutsch	6	5	5 *	5 **
<b>MW Abiturnote</b>	1.56	1.68	1.56	1.50
SD	0.61	0.60	0.59	0.37
<b>MW Physikumsnote</b>	2.50	2.70	2.62	2.69
SD	0.85	0.88	0.83	0.87
<b>Modul 23</b>	13	12	13	12
4	10	8	7	6
5	5	4	7	0
6	6	7	10	7
PJ	4	2	8	1
<b>MW BZ</b> <b>der Onlinefälle gesamt</b>	46.11 min	44.86 min	44.15 min	45.09 min
SD	13.14	11.20	12.62	12.15
<b>MW BZ</b> <b>einzelner Onlinefall</b>	5.76 min	5.61 min	5.52 min	5.64 min
SD	1.64	1.40	1.57	1.52

*Tabelle 5: Stichprobeneigenschaften der vier experimentellen Gruppen*

N = Anzahl, MW =Mittelwert, SD = Standardabweichung, BZ = Bearbeitungszeit, \* ein Teilnehmer zweisprachig, \*\* zwei Teilnehmer zweisprachig, **keine signifikanten Unterschiede in den Gruppen**

#### 4.1.6 Deskriptive Statistik der abhängigen Variablen

Der Mittelwert der Diagnoserichtigkeit aller Teilnehmer gesamt entsprach 4.46, siehe Tabelle 5. Die minimal erreichte Diagnoserichtigkeit betrug 1.0 von der minimal möglichen Punktzahl von 0. Das bedeutete kein Teilnehmer hatte alle Onlinefälle falsch gelöst und mindestens ein Onlinefall war von jedem Teilnehmer richtig gelöst worden. Die Mittelwerte der Diagnoserichtigkeit der vier experimentellen Gruppen unterschieden sich mit 4.53, 3.91, 5.05, und 4.04 von dem Mittelwert aller Teilnehmer gesamt, wobei die Gruppe *whole case x hohes Vorwissen* mit 5.04 den höchsten Mittelwert erreichte, siehe Tabelle 6.

	MW	SD	95% Konfidenzintervall	Minimal erreicht	Maximal erreicht
<b>Diagnoserichtigkeit gesamt</b>	4.46	1.59	Untergrenze 4.12 Obergrenze 4.64	1.0	8.0

Tabelle 6: Diagnoserichtigkeit aller Teilnehmer gesamt, MW = Mittelwert, SD = Standardabweichung

	<b>serial cue x hohes Vorwissen</b>	<b>serial cue x niedriges Vorwissen</b>	<b>whole case x hohes Vorwissen</b>	<b>whole case x niedriges Vorwissen</b>
<b>MW Diagnoserichtigkeit</b>	4.53	3.91	5.04	4.04
<b>SD</b>	1.37	1.28	1.61	1.89

Tabelle 7: Diagnoserichtigkeit der vier experimentellen Gruppen gesamt, MW = Mittelwert, SD = Standardabweichung

Betrachtete man die Onlinefälle einzeln, so zeigte sich ein erheblicher Unterschied in der Anzahl der richtig gestellten Diagnosen pro Fall, siehe auch Tabelle 7. Die Fälle akute Hepatitis-A-Infektion und HIV-Infektion wurden am häufigsten richtig gelöst. 116 Teilnehmer (81.7 %) hatten die richtige Diagnose gestellt. Der Fall Lumbago wurde am seltensten richtig gelöst. Lediglich 39 Teilnehmer (27,5 %) hatten die richtige Diagnose gestellt.

	N	N Richtige Diagnose	N Falsche Diagnose	MW	SD
<b>Bakterielle Endokarditis</b>	142	56	86	.39	.49
<b>Akute Hepatitis-A-Infektion</b>	142	116	25	.82	.38
<b>HIV-Infektion</b>	142	116	26	.82	.39
<b>Hyperthyreosis factitia</b>	142	99	63	.30	.46
<b>Lumbago</b>	141*	39	102	.28	.45
<b>Osteoporotische Wirbelkörperfraktur</b>	142	84	58	.59	.50
<b>Spinalkanalstenose</b>	141*	74	67	.52	.50
<b>Spondylitis ankylosans</b>	142	105	37	.74	.44

Tabelle 8: Lösung der einzelnen Onlinefälle, N = Anzahl, MW = Mittelwert, SD = Standard Abweichung, \*ein Wert fehlend

Bezüglich der Diagnoseeffizienz zeigte sich, dass lediglich 3 von 71 Teilnehmer (4.2%) effizient im Sinne der Diagnoseeffizienz gearbeitet hatten, siehe auch Tabelle 8. Der Median der Diagnoseeffizienz ergab 0.0. Teilnehmer mit einem Ergebnis von 0 wurden der Gruppe niedrige Diagnoseeffizienz zugeordnet, während Ergebnisse über 0 zur Gruppe hohen Diagnoseeffizienz gezählt wurden. Der Mittelwert der Diagnoserichtigkeit der Gruppe hohe Diagnoseeffizienz war mit 4.5 leicht höher als der Mittelwert der Gruppe niedrige Diagnoseeffizienz mit 4.0, siehe Tabelle 9.

Es zeigte sich auch, dass nur 7 von 71 Teilnehmern (9,9%) die Onlinefälle ineffizient im Sinne der Diagnoseineffizienz bearbeitet hatten, siehe Tabelle 8. Der berechnete Median ergab 2.0. 0 bis einschließlich 2.0 Punkte wurden als niedrige Diagnoseineffizienz betrachtet. Ergebnisse über 2.0 wurden als hohe Diagnoseineffizienz betrachtet. Es zeigte sich, dass der Mittelwert der Diagnoserichtigkeit der Teilnehmer mit niedriger Diagnoseineffizienz mit 4.4 höher ausfiel, als der Mittelwert der Teilnehmer mit hoher Diagnoseineffizienz mit 4.0, siehe auch Tabelle 9.

	N	Minimum	Maximum	MW	SD	Median
<b>Diagnoseeffizienz</b>	71	0	3	.61	.90	0
<b>Diagnoseineffizienz</b>	71	0	7	3.15	2.84	2

Tabelle 9: Diagnoseeffizienz und Diagnoseineffizienz, N = Anzahl, MW = Mittelwert, SD = Standard Abweichung

	hohe Diagnoseeffizienz	niedrige Diagnoseeffizienz	hohe Diagnoseineffizienz	niedrige Diagnoseineffizienz
<b>N</b>	28	43	33	38
<b>%</b>	30.4	60.6	46.5	53.5
<b>MW DR</b>	4.5	4.0	4.0	4.4
<b>SD</b>	1.3	1.3	1.4	1.3

Tabelle 10: Ergebnisse der Diagnoserichtigkeit des Median-Splits der Diagnoseeffizienz und der Diagnoseineffizienz, N = Anzahl, MW = Mittelwert, DR = Diagnoserichtigkeit, SD = Standardabweichung

Betrachtete man die Ergebnisse des prozeduralen Wissens, ergaben sich Mittelwerte des Prätests von .60 (*serial cue* x hohes Vorwissen), .60 (*serial cue* x niedriges Vorwissen), .62 (*whole case* x hohes Vorwissen) und .61 (*whole case* x niedriges Vorwissen), die sich wie in 4.1.4 erwähnt, nicht signifikant voneinander unterschieden. Die Mittelwerte des Posttests wichen von den Ergebnissen des Prätests ab. Am stärksten ausgeprägt war dies in der Gruppe *whole case* x hohes Vorwissen mit .62, siehe Tabelle 10.

	serial cue x hohes Vorwissen	serial cue x niedriges Vorwissen	whole case x hohes Vorwissen	whole case x niedriges Vorwissen
<b>MW Prätest</b>	.60	.60	.62	.61
<b>SD</b>	.08	.07	.08	.06
<b>MW Posttest</b>	.64	.61	.67	.59
<b>SD</b>	.08	.11	.11	.08

Tabelle 11: Ergebnisse des prozeduralen Wissenstestes, MW = Mittelwert, SD = Standardabweichung

#### 4.1.7 Korrelative Zusammenhänge nach Pearson

Die Orientierungshilfe von Cohen (Cohen, 1988), welche  $r = .10$  einen kleinen,  $r = .30$  einen moderaten und  $r = .50$  einen starken Zusammenhang zuordnet, wurde zur Interpretation verwendet. Es wurden nur Korrelationskoeffizienten berichtet, die auf dem 1%-Niveau 2-seitig signifikant sind.

Ein stark positiver Zusammenhang ( $r = .68$ ) ergab sich zwischen Alter und Abiturnote (je älter, desto schlechter die Abiturnote und umgekehrt) und ein moderater zwischen Alter und Modul ( $r = .38$ ) (je älter, desto höheres Modul und umgekehrt).

Zwischen der Abiturnote und der Physikumsnote gab es einen kleinen positiven Zusammenhang ( $r = .22$ ) (je schlechter die Abiturnote, desto schlechter die Physikumsnote und umgekehrt). Die Physikumsnote ergab mit der Diagnoserichtigkeit einen kleinen negativen Zusammenhang ( $r = -.29$ ) (je schlechter die Physikumsnote, desto kleiner die Diagnoserichtigkeit und umgekehrt) und einen moderate negativen ( $r = -.35$ ) mit dem Mittelwert des prozeduralen Wissens im Posttest (je schlechter die Physikumsnote, desto schlechter das KF Post Ergebnis und umgekehrt).

Die Diagnoserichtigkeit wies einen kleinen positiven Zusammenhang ( $r = .30$ ) mit dem Mittelwert des prozeduralen Wissens im Prätest (je höher die Diagnoserichtigkeit,

desto höher das KF Prä Ergebnis und umgekehrt) und einen positiv moderaten ( $r = .33$ ) mit dem Mittelwert des prozeduralen Wissens im Posttest auf (je höher die Diagnoserichtigkeit, desto höher das KF Post Ergebnis und umgekehrt). Ebenso wies sie einen kleinen positiven Zusammenhang ( $r = .27$ ) mit dem Mittelwert des deskriptiven Wissenstests im Prätest und einen positiv moderaten Zusammenhang ( $r = .36$ ) mit dem Mittelwert des Posttests auf (je höher die Diagnoserichtigkeit, desto höher die MC Prä und Post Ergebnisse und umgekehrt).

Die Diagnoseineffizienz zeigte einen stark negativen Zusammenhang mit der Diagnoseeffizienz ( $r = -.54$ ) (je höher die Diagnoseineffizienz, desto niedriger die Diagnoseeffizienz und umgekehrt). Es zeigten sich auch moderat positive Zusammenhänge der Diagnoseineffizienz mit der Muttersprache ( $r = .32$ ) (war die Muttersprache nicht Deutsch, desto höher war die Diagnoseineffizienz und umgekehrt) und der Bearbeitungszeit ( $r = .38$ ) (je höher die Diagnoseineffizienz, desto höher die Bearbeitungszeit und umgekehrt).

Das Modul wies einen moderat positiven Zusammenhang ( $r = .37$ ) mit dem Mittelwert des prozeduralen Wissenstests im Posttest auf (je höher das Modul, desto höher das KF Post Ergebnis und umgekehrt).

Mittelwerte des prozeduralen Wissenstests im Prätest zeigten kleine positive Zusammenhänge ( $r = .29$ ) zu Mittelwerten im prozeduralen Wissenstest im Posttest (je höher das KF Prä Ergebnis, desto höher das KF Post Ergebnis und umgekehrt).

Mittelwerte des deskriptiven Wissenstests im Prätest hatten einen kleinen positiven Zusammenhang ( $r = .26$ ) zu Mittelwerten des prozeduralen Wissenstests im Posttest (je höher das MC Post Ergebnis, desto höher das KF Post Ergebnis und umgekehrt).

## 4.2 HYPOTHESENTESTUNG

Die Ergebnisse der Hypothesentestung werden nach den Forschungsfragen gegliedert dargestellt. Nach den Richtlinien für Effektstärken von Cohen (1988) wird das Effektstärkemaßes folgendermaßen bewertet:

$\eta^2 > .01$  kleiner Effekt,  $\eta^2 > .06$  mittlerer Effekt,  $\eta^2 > .14$  großer Effekt.

### 4.2.1 Auswirkung des Falldarbietungsformats und der Vorwissensstufe auf die Diagnoserichtigkeit

Eine univariante Varianzanalyse wurde mit der abhängigen Variable Diagnoserichtigkeit der vier experimentellen Gruppen und dem Falldarbietungsformat und der Vorwissensstufe durchgeführt. Der Effekt der Vorwissensstufe mit Falldarbietungsformat war nicht signifikant ( $F(1,138) = .55$ ,  $p = .46$ , n.s.). Es zeigten sich also keine signifikanten Unterschiede der Diagnoserichtigkeit, ob nun ein Studierender mit niedrigem Vorwissen einen Onlinefall im *serial cue*-Format oder im *whole case*-Format bearbeitet hatte und dasselbe galt für Studierende mit hohem Vorwissen. H1 und H2 konnten nicht bestätigt werden.

Das Falldarbietungsformat allein betrachtet hatte auch keinen signifikanten Interaktionseffekt auf die Diagnoserichtigkeit ( $F(1,138) = 1.52$ ,  $p = .22$ , n.s.). Die Ergebnisse der Diagnoserichtigkeit der Falldarbietungsformate unabhängig von der Vorwissensstufe unterschieden sich also nicht signifikant. Das bedeutete, ob nun ein Teilnehmer die Onlinefälle im *serial cue*- oder *whole case*-Format bearbeitet hatte, beeinflusste die Diagnoserichtigkeit nicht signifikant.

Die Vorwissensstufe allein jedoch hatte einen signifikanten Effekt mittlerer Stärke ( $F(1,138) = 9.54$ ,  $p = .002$ ,  $\eta^2 = .07$ ) auf die Diagnoserichtigkeit. Das bedeutete, dass sich die Diagnoserichtigkeit in den Gruppen hohes und niedriges Vorwissen signifikant

unterschied. Betrachtete man die Messwerte, siehe auch Tabelle 6, dann erreichten Studierende mit hohem Vorwissen eine höhere Diagnoserichtigkeit als Studierende mit niedrigem Vorwissen. Sie diagnostizierten demnach signifikant mehr Fälle richtig als Studierende mit niedrigem Vorwissen.

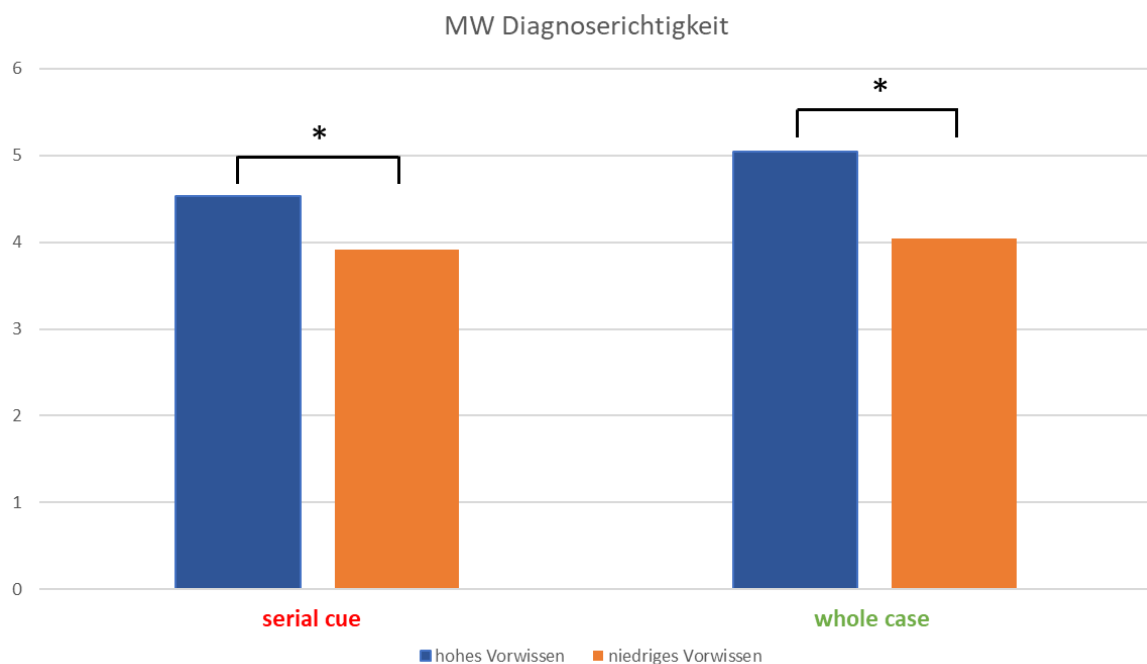


Abbildung 9: Diagramm der Diagnoserichtigkeit der vier experimentellen Gruppen, \* signifikanter Unterschied gekennzeichnet

#### 4.2.2 Auswirkung der Anzahl der diagnostischen Schritte im serial cue-Format auf die Diagnoserichtigkeit

Eine univariante Varianzanalyse ergab, dass die Diagnoseeffizienz im *serial cue*-Format keinen signifikanten Effekt auf die Diagnoserichtigkeit ( $F(1,67) = .46, p = .50, n.s.$ ) hatte. Studierende mit hoher Diagnoseeffizienz erreichten also keine höhere Diagnoserichtigkeit, diagnostizierten also nicht signifikant mehr Onlinefälle richtig als Studierende mit niedriger Diagnoseeffizienz. Die Diagnoseineffizienz hatte auch keinen signifikanten Effekt auf die Diagnoserichtigkeit ( $F(1,67) = 2.8, p = .09, n.s.$ ). Studierende mit niedriger Diagnoseineffizienz erreichten auch keine höhere



Diagnoserichtigkeit, diagnostizierten also auch nicht signifikant mehr Onlinefälle richtig als Studierende mit hoher Diagnoseineffizienz. Jedoch war bei Betrachtung der Ergebnisse und  $p = .09$  eine Tendenz erkennbar, dass Probanden mit niedriger Diagnoseineffizienz eine höhere Diagnoserichtigkeit erzielten, siehe auch Tabelle 9 und Abbildung 8, jedoch war dies nicht signifikant. H3 und H4 konnten nicht bewiesen werden.

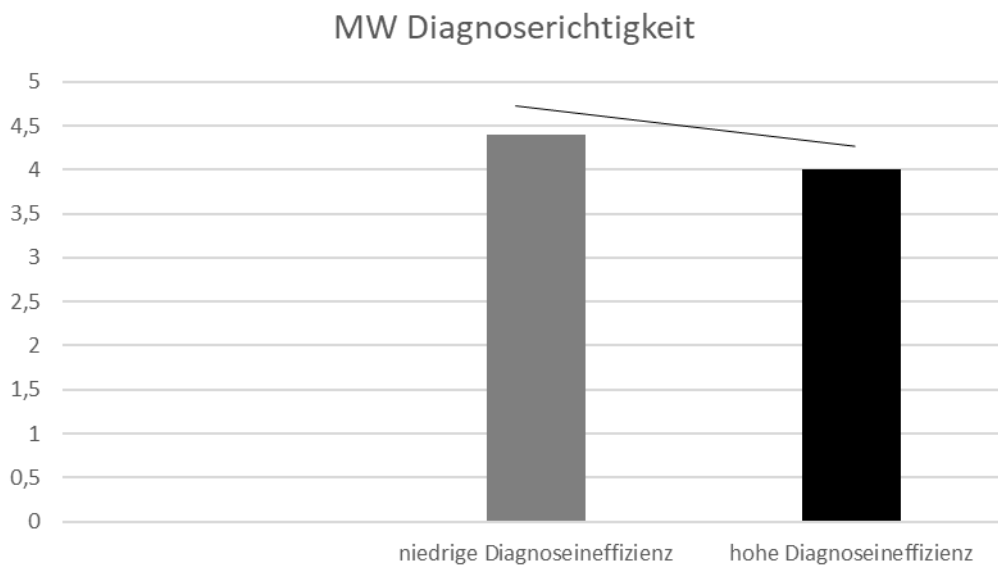


Abbildung 10: Diagramme der Diagnoserichtigkeit der hohen und niedrigen Diagnoseineffizienz, Tendenzen erkennbar, jedoch **nicht signifikant**

### 4.2.3 Auswirkung des Falldarbietungsformats und der Vorwissensstufe auf den Wissenszuwachs

Eine einstufige messwiederholte Varianzanalyse wurde mit den Mittelwerten der Ergebnisse des prozeduralen Wissens im Prä- und Posttest zur Überprüfung von H5 und H6 durchgeführt. Der Messzeitpunkt hatte einen signifikant mittleren Effekt auf die Ergebnisse des prozeduralen Wissens ( $F(1,138) = 6.79, p = .01, \eta^2 = .05$ ). Die Ergebnisse des Prä- und Posttests unterschieden sich also signifikant. Die Analysen des Messzeitpunkt mit dem Falldarbietungsformat ( $F(1,138) = .75, p = .40, n.s.$ ) und des Messzeitpunkt mit der Vorwissensstufe und dem Falldarbietungsformat ( $F(1,138) = .75, p = .39, n.s.$ ) ergaben keine signifikanten Effekte. Das heißt, die Mittelwerte der unterschiedlichen Falldarbietungsformate unterschieden sich nicht signifikant im Prä- und Posttest und auch nicht in der Kombination mit den Vorwissensstufen. Also zeigten sich keine signifikanten Unterschiede des prozeduralen Wissens, ob nun ein Studierender mit niedrigem Vorwissen einen Onlinefall im *serial cue*-Format oder im *whole case*-Format bearbeitet hatte und das Gleiche galt für Studierende mit hohem Vorwissen. H5 und H6 können daher nicht bestätigt werden.

Jedoch zeigte sich, dass die Vorwissensstufe allein mit dem Messzeitpunkt sehr wohl einen signifikanten Effekt ausübte ( $F(1,138) = 8.08, p = .01, \eta^2 = .06$ ). Das bedeutete, dass sich die Mittelwerte der Prä- und Posttests im Vergleich der beiden Vorwissensstufen signifikant unterschieden. Bei Betrachtung der Messwerte, siehe Tabelle 10 und Abbildung 10, zeigte sich, dass nur Studierende mit hohem Vorwissen einen signifikanten Unterschied im prozeduralen Wissen aufwiesen. Das prozedurale Wissen im Posttest war bei ihnen signifikant höher als im Prätest. Studierende mit niedrigem Vorwissen hingegen zeigten im Vergleich des Prä- zum Posttest keine Verbesserung im prozeduralen Wissen.

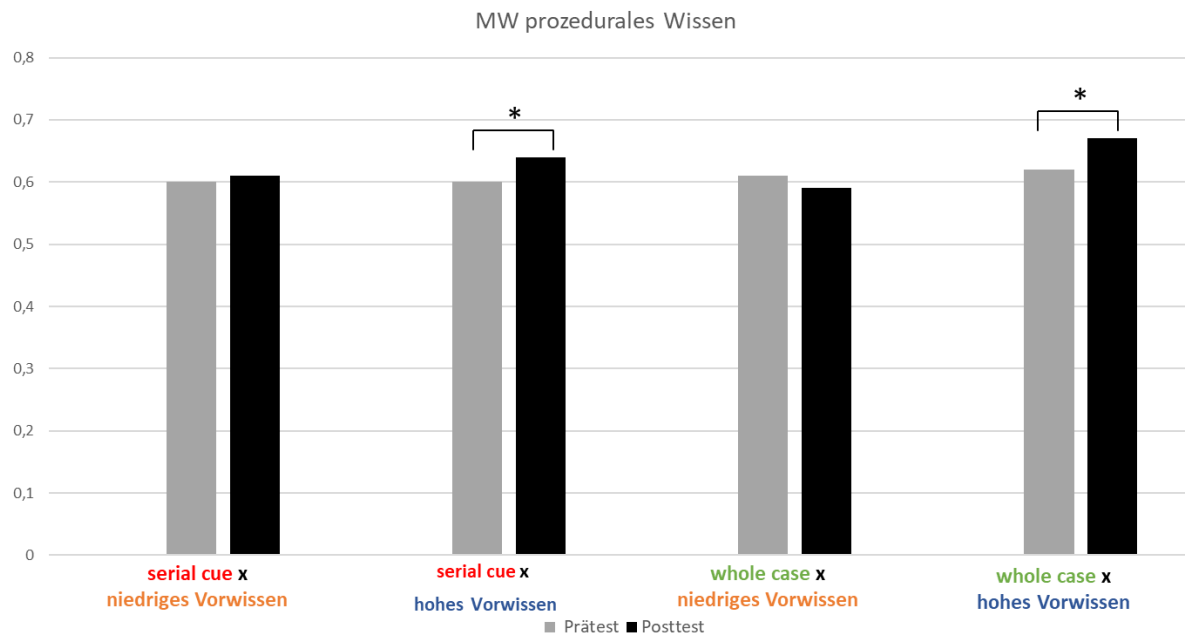


Abbildung 11: Diagramm der Mittelwerte des prozeduralen Wissens der vier experimentellen Gruppen im Prä- und Posttest, \* signifikanter Unterschied gekennzeichnet

# 5 Diskussion

---

## 5.1 ZUSAMMENFASSUNG DER HAUPTERGEBNISSE

Hypothese 1 formulierte die Annahme, dass Studierende mit höherem Vorwissen im *serial cue*-Format eine höhere Diagnoserichtigkeit erreichen als im *whole case*-Format. Hypothese 2 besagte, dass Studierende mit niedrigerem Vorwissen im *whole case*-Format eine höhere Diagnoserichtigkeit erreichen als im *serial cue*-Format. Unsere Ergebnisse konnten diese beiden Annahmen jedoch nicht bestätigen, siehe Kapitel 4.2.1. Studierende mit niedrigem oder hohem Vorwissen zeigten also keine signifikanten Unterschiede in der Diagnoserichtigkeit, ob sie nun einen Onlinefall im *serial cue*-Format oder im *whole case*-Format bearbeitet hatten. Die Diagnoserichtigkeit der Falldarbietungsformate, unabhängig von der Vorwissensstufe, zeigte ebenso keine signifikanten Unterschiede. Ob ein Studierender nun einen Fall im *serial cue*- oder *whole case*-Format bearbeitet hatte, machte also keinen signifikanten Unterschied. Die Vorwissensstufe hingegen hatte einen signifikanten Interaktionseffekt auf die Diagnoserichtigkeit. Studierende mit hohem Vorwissen diagnostizierten demnach signifikant mehr Onlinefälle richtig als Studierende mit niedrigem Vorwissen.

Die Auswahl der diagnostischen Schritte der Studierenden im *serial cue*-Format wurde untersucht. Diagnoseeffizienz war gegeben, wenn ein Studierender nur die für die Diagnosestellung nötigen diagnostischen Schritte ausgewählt hatte. Diagnoseineffizienz war gegeben, wenn ein Studierender alle diagnostischen Schritte, darunter auch irrelevante, ausgewählt hatte. Hypothese 3 behauptete, dass Studierende, die effiziente Diagnostik im *serial cue*-Format durchführen, eine höhere

Diagnoserichtigkeit erreichen als Studierende, die keine effiziente Diagnostik durchführen. Hypothese 4 formulierte die dazu passende Annahme, dass Studierende, die keine ineffiziente Diagnostik im *serial cue*-Format durchführen, eine höhere Diagnoserichtigkeit erreichen als Studierende, die ineffiziente Diagnostik durchführen. Beide Hypothesen konnten durch die Ergebnisse dieser Studie nicht bewiesen werden, vergleiche Kapitel 4.2.2. Die Diagnoseeffizienz und die Diagnoseineffizienz hatten jeweils keinen signifikanten Einfluss auf die Diagnoserichtigkeit. Das hieß, dass weder Studierende, die nur die für die Diagnosefindung nötigen diagnostischen Schritte angewandt hatten, noch Studierende, die alle diagnostischen Schritte angewandt hatten, mehr Onlinefälle richtig gelöst hatten.

Angelehnt an die Hypothesen 1 und 2, formulierte Hypothese 5 die Annahme, dass Studierende mit höherem Vorwissen im *serial cue*-Format einen größeren Wissenszuwachs haben als im *whole case*-Format. Hypothese 6 besagte, auch angelehnt an Hypothese 1 und 2, dass Studierende mit niedrigerem Vorwissen im *whole case*-Format einen größeren Wissenszuwachs haben als im *serial cue*-Format. Unsere Ergebnisse konnten dies nicht bestätigen, siehe auch Kapitel 4.2.3. Das Falldarbietungsformat mit dem Messzeitpunkt und der Vorwissensstufe wies keinen signifikanten Effekt auf. Das heißt, ob ein Studierender mit niedrigem oder hohem Vorwissen nun einen Onlinefall im *serial cue*- oder *whole case*-Format bearbeitet hatte, übte keinen signifikanten Einfluss auf das prozedurale Wissen aus. Die Vorwissensstufe allein übte jedoch einen signifikanten Interaktionseffekt auf das prozedurale Wissen aus. Es zeigte sich, dass Studierende mit hohem Vorwissen einen signifikanten Wissenszuwachs aufwiesen, während Studierende mit niedrigem Vorwissen dies nicht taten.

## 5.2 DISKUSSION DER ERGEBNISSE

### 5.2.1 Diskussion der Auswirkung des Falldarbietungsformats und der Vorwissensstufe auf die Diagnoserichtigkeit

In Hypothese 1 wurde die Annahme formuliert, dass Studierende mit höherem Vorwissen im *serial cue*-Format eine höhere Diagnoserichtigkeit erreichen als im *whole case*-Format. Hypothese 2 formulierte die dazugehörige Annahme, dass Studierende mit niedrigerem Vorwissen im *whole case*-Format eine höhere Diagnoserichtigkeit erreichen als im *serial cue*-Format. Unsere Studie konnte diese Annahmen jedoch nicht bestätigen. Eine Betrachtung der Mittelwerte zeigte, dass die Diagnoserichtigkeit der Studierenden im *whole case*-Format zwar höher war, doch war dies nicht signifikant. Das Falldarbietungsformat hatte in unserer Studie also keinen signifikanten Effekt auf die Diagnoserichtigkeit. Die Studie von Schubach, Goos, Fabry, Vach und Boeker (2017) verglich das *serial cue*-Format mit dem Key-Feature-Format, siehe 5.3.1. Sie beschreiben wie unsere Studie ebenfalls keine signifikanten Unterschiede des Falldarbietungsformats. Doch zeigt sich in deren Studie ein numerischer, wenn auch nicht signifikanter, Vorteil im *serial cue*-Format. Schmidt und Mamede (2015) hingegen behaupten, dass das Falldarbietungsformat doch einen Unterschied machen würde. Die Diagnoserichtigkeit ist in dem von ihnen durchgeführten Review im *serial cue*-Format im Vergleich zum *whole case*-Format signifikant kleiner. Bei Betrachtung der von ihnen verglichenen Studien ist jedoch fraglich, ob man diese Studien wirklich alle in den Vergleich ziehen konnte. Die Studiendesigns, die Interventionen und die gemessenen Variablen unterscheiden sich in den verglichenen Studien sehr. Einige dieser Studien verglichen zum Beispiel nur die Vorwissensstufen. Die einzige Studie, die in dem Review von Schmidt und Mamede (2015) wirklich das Falldarbietungsformat ähnlich dem *serial cue*- mit dem

*whole case*-Format verglich, war die Studie von Nendaz et al. (2000). Die ganze Aussage von Schmidt und Mamede (2015) beruht daher auf der Studie von Nendaz et al. (2000). So entspricht die Aussagekraft des Reviews von Schmidt und Mamede (2015) der einer einzigen Studie. Die eben schon erwähnte Studie von Nendaz et al. (2000) kommt zu dem Ergebnis, dass die Diagnoserichtigkeit im *serial cue*-Format generell signifikant kleiner ist als im *whole case*-Format und widerspricht somit unseren Ergebnissen. Es gibt jedoch im Vergleich zu unserer Studie einige Unterschiede. Nendaz et al. (2000) legten ausbildungsspezifisch drei Vorwissensstufen fest, Medizinstudierende verschiedener Semester, Assistenzärzte verschiedener Jahre und Fachärzte, siehe auch 5.3.3. In unserer Studie waren zwei Vorwissensstufen aus Medizinstudierenden gebildet worden. Dies war mit Absicht so gewählt, da untersucht werden sollte, wie sich Falldarbietungsformate auf die diagnostischen Kompetenzen Diagnoserichtigkeit bei Medizinstudierenden mit unterschiedlichen Vorwissensstufen auswirken. Ein Vergleich mit Assistenz- und Fachärzten wie bei Nendaz et al. (2000) erscheint da wenig. Im Unterschied zu unserer Studie, in der acht Onlinefälle bearbeitet wurden, haben bei Nendaz et al. (2000) die Teilnehmer nur einen einzelnen Onlinefall bearbeitet. Ein einzelner Fehler fällt bei Nendaz et al. (2000) somit stärker ins Gewicht und es könnte eine mögliche Konfundierung der Ergebnisse durch den Inhalt des einzelnen Falls entstanden sein, siehe dazu auch 5.3.8.

Nendaz et al. (2000) versuchen die signifikant besseren Ergebnisse des *whole case*-Formats damit zu erklären, dass die Teilnehmer im *serial cue*-Format Schwierigkeiten gehabt hätten, wichtige Informationen zu sammeln. Das Sammeln von wichtigen Informationen scheint wichtiger zu sein als die Interpretation dieser (Gruppen, Wolf, & Billi, 1991). Ein breiteres Angebot an Informationen im *whole case*-Format schien das Sammeln von wichtigen Informationen vereinfacht zu haben.

Wird das *serial cue*-Format jedoch um einen Hypothesenausschlussansatz ergänzt, zeigen sich Vorteile gegenüber dem *whole case*-Format mit Selbsterklärungsansatz (Al Rumayyan et al., 2018), vergleiche 5.3.1. Dies könnte bedeuten, dass das gezielte Ausschließen von Hypothesen die Schwächen des *serial cue*-Formats gegenüber dem *whole case*-Format ausgleichen könnte. Jedoch erweist sich der Ansatz des Ausschlusses von Hypothesen generell besser als der Selbsterklärungsansatz (Al Rumayyan et al., 2018). Studierende, die Patientenfälle im *whole case*-Format bei Al Rumayyan et al. (2018) bearbeitet hatten, könnten also durch die Aufforderung zum Selbsterklärungsansatz behindert worden sein.

Vor allem Medizinstudierende haben eine geringere Diagnoserichtigkeit im *serial cue*-Format aufgewiesen (Nendaz et al., 2000). Hierbei sollte nun beachtet werden, dass sie Medizinstudierende mit Assistenzärzten der inneren Medizin und Fachärzten der inneren Medizin verglichen. Das Studierende schlechtere Ergebnisse im Diagnostizieren erreichen als diese, erscheint nur logisch. Nendaz et al. (2000) versuchen dies damit zu erklären, dass Studierende am wenigsten in der Lage wären, wichtige Informationen auszuwählen. Assistenzärzte und Fachärzte hingegen verfügen schon über *illness scripts*, die ihnen dabei geholfen hätten (Schmidt et al., 1990).

Unsere Studie zeigte, dass das Falldarbietungsformat keinen signifikanten Effekt auf die Diagnoserichtigkeit ausübt. Es schien auch keinen Unterschied zu machen, in welcher Vorwissensstufe sich die Medizinstudierenden dabei befanden. Eine mögliche Erklärung für diese Unterschiede zu den Ergebnissen von Nendaz et al. (2000) könnte sein, dass bei Nendaz et al. (2000) jeder Teilnehmer nur eine Diagnose stellen musste und wie oben schon erwähnt der Effekt auch durch den Inhalt des Falls entstanden sein konnte, während in unserer Untersuchung acht Diagnosen gestellt werden mussten, siehe dazu auch 5.3.1 und 5.3.5. Die Schwierigkeit, wichtige Informationen



sammeln zu können, könnte von Onlinefall zu Onlinefall und von Teilnehmer zu Teilnehmer bei acht Onlinefällen mehr variiert haben als bei einem Fall. Die Schwierigkeit wichtige Informationen zu sammeln, scheint also auch von den Onlinefällen und deren Inhalten abzuhängen. Dies könnte daraufhin deuten, dass Medizinstudierende bereits teilweise über *illness scripts* verfügen. Wenn auch wahrscheinlich über weniger ausführliche wie Assistenzärzte und Fachärzte dies tun. Je mehr Erfahrungen Ärzte mit Patienten einer bestimmten Diagnose gesammelt hatten, über mehr *illness scripts* verfügten sie (Eugène JFM Custers et al., 1998). Biomedizinisches Wissen scheint dabei eine wichtige Rolle bei der Entwicklung von *illness scripts* zu spielen (Schmidt & Rikers, 2007).

Eine weitere Erklärung könnte sein, dass Studierende, die Onlinefälle im *serial cue*-Format bearbeiteten, den Hypothesenausschlussansatz, wie bei Al Rumayyan et al. (2018) verwendet, von selbst anwandten und so die Schwächen des *serial cue*-Formats gegenüber dem *whole case*-Format automatisch ausglich. Eine weitere mögliche Erklärung wäre, dass die Darstellung und die Benutzung der Falldarbietungsformate auf CASUS® nicht unterschiedlich genug verwirklicht worden war. Zwar konnten die Probanden im *whole case*-Format keine diagnostischen Schritte auswählen, doch konnten diese den Bildschirm entlang scrollen und so das *whole case*-Format teilweise wie ein *serial cue*-Format bearbeiten. Es war also nicht garantiert, dass keine diagnostischen Schritte übersprungen wurden. Ebenso war es möglich, im *serial cue*-Format alle diagnostischen Schritte auszuwählen, da es keine Begrenzung der diagnostischen Schritte gegeben hatte und das *serial cue*-Format so wie ein *whole case*-Format anzuwenden.

Unsere Studie zeigt, dass die Vorwissensstufe allein einen signifikanten Effekt mittlerer Stärke auf die Diagnoserichtigkeit ausübte. Das bedeutet, Studierende mit hohem

Vorwissen diagnostizierten signifikant mehr Fälle richtig als Studierende mit niedrigem Vorwissen. In diesem Punkt stimmt unsere Studie mit den Ergebnissen von Nendaz et al. (2000) überein, die besagen, dass die höheren Vorwissenstufen höhere Diagnoserichtigkeit erreichen. Auch wenn diese wie oben erwähnt andere Vorwissenstufen definiert hatten, siehe auch Tabelle 11. Die Diagnoserichtigkeit hängt sehr vom Vorwissen über eine Diagnose und einen Fall ab (Nendaz et al., 2000). Die Erklärung für diese Beobachtung erschien zunächst sehr simpel. Mehr Vorwissen über ein Thema beinhaltet wohl auch mehr Wissen über Krankheiten und deren diagnostische Abklärung. Auch Braun, Zwaan, Kiesewetter, Fischer und Schmidmaier (2017) stellen fest, dass die Diagnoserichtigkeit vom Vorwissen der Medizinstudierenden abhängt. Sie führen auf, dass es dabei auch auf die Fähigkeiten der Studierenden ankommt, Befunde richtig zu interpretieren und den richtigen Kontext zu generieren. Die Studierende mit hohem Vorwissen unserer Studie scheinen diese Fähigkeiten beherrscht zu haben.

Andere Studien hingegen zeigen, dass mehr Vorwissen nicht unbedingt zur richtigen Diagnose führte (Jan Kiesewetter et al., 2016; Schmidmaier et al., 2013). Nicht das Vorwissen, sondern dessen zielgerichtete Anwendung helfe dabei, richtige Diagnosen zu stellen (Jan Kiesewetter et al., 2016). Es käme vor allem auf die Evaluation, die Repräsentation und die Integration von Informationen an (Jan Kiesewetter et al., 2013). Wenn Studierende dank ihrem Vorwissen verstehen, warum ein Symptom auftritt und welche Mechanismen dahinter stehen, wiesen sie eine höhere Diagnoserichtigkeit auf (Jan Kiesewetter et al., 2016). Es scheint, dass wenn in unserer Untersuchung Studierende mehr Vorwissen besaßen, sie dies auch gleichzeitig zielgerichtet anzuwenden wussten.

Abschließend könnten die Ergebnisse dieser Studie für die Förderung von diagnostischen Kompetenzen und die Lehre mittels Onlinefällen folgendes bedeuten: Es sollte weniger auf das Falldarbietungsformat geachtet werden und mehr auf die Vorwissensstufe in welcher sich der Lernende befindet. Studierende in unterschiedlichen Phasen des Wissenserwerbs, bzw. Studierende unterschiedlicher Vorwissensstufen, haben keinen Vorteil von unterschiedlicher Falldarbietung mit *serial cue*- und *whole case*-Formaten. Um die diagnostische Kompetenz Diagnoserichtigkeit zu fördern, sollte den Studierenden mehr Wissen über Krankheiten, deren Pathomechanismen und deren diagnostische Abklärung vermittelt werden. Ein Schwerpunkt sollte dabei sein, Studierenden die richtige Anwendung von Wissen beizubringen.

Pinnock et al. (2012) geben eine Anleitung zur Erstellung virtueller Patienten für Studierende mit niedrigem Wissen. Sie gehen dabei nicht auf das Falldarbietungsformat ein, sondern empfehlen lernunterstützende Maßnahmen wie *Feedback*. Neue Arten des medizinischen *social learnings* könnten ebenso miteinbezogen werden (Hautz, Kämmer, Schaubert, Spies, & Gaissmaier, 2015; Jan Kiesewetter, Fischer, & Fischer, 2017). Zudem sollte auf die Erzählweise des Inhalts achtgegeben werden, da Fälle, die negative Emotionen hervorrufen, schlechtere Resultate erzielen (Hege, Kononowicz, Berman, Lenzer, & Kiesewetter, 2018; Mamede et al., 2016) .

Die optimalen Falldarbietungen für Studierende mit niedrigem und hohem Vorwissen konnten also noch nicht gefunden werden. Zukünftige Studien könnten die oben erwähnten Konzepte miteinbeziehen, um sie zu finden.

	Ergebnisse Nendaz et al. (2000)	Ergebnisse dieser Studie
<b>Medizinstudierende</b>	whole case 36 % serial cue 10 %	<b>niedriges Vorwissen:</b> whole case 50 % serial cue 48 %  <b>hohes Vorwissen:</b> whole case 62 % serial cue 56 %
<b>Assistenzärzte der Inneren Medizin</b>	whole case 81 % serial cue 47 %	-
<b>Fachärzte der Inneren Medizin</b>	whole case 100 % serial cue 59 %	-

Tabelle 12: Vergleich der Ergebnisse mit den Ergebnissen von Nendaz et al. (2000), Anzahl der richtig gestellten Diagnosen in Prozent

### 5.2.2 Diskussion der Auswirkung der diagnostischen Schritte im serial cue-Format auf die Diagnoserichtigkeit

Hypothese 3 besagte, dass Studierende, die effiziente Diagnostik im *serial cue*-Format durchführen, eine höhere Diagnoserichtigkeit erreichen als Studierende, die keine effiziente Diagnostik durchführen. Hypothese 4 formulierte die dazu passende Annahme, dass Studierende, die keine ineffiziente Diagnostik im *serial cue*-Format durchführen, eine höhere Diagnoserichtigkeit erreichen als Studierende, die ineffiziente Diagnostik durchführen. Die Ergebnisse unserer Studie bestätigen diese Annahmen jedoch nicht. Diagnoseeffizienz und Diagnoseineffizienz im *serial cue*-Format haben jeweils keinen signifikanten Effekt auf die Diagnoserichtigkeit. Es sollte beachtet werden, dass hier nur die halbe Studienpopulation untersucht werden konnte, da nur die halbe Population die Onlinefälle im *serial cue*-Format bearbeitet hatte. Das

*whole case*-Format konnte in Bezug auf diese Variable nicht untersucht werden, da Probanden in diesem Format keine diagnostischen Schritte auswählen konnten.

Die Variable Diagnoseineffizienz wurde eingeführt, um den Umgang der Studierenden mit irrelevanten diagnostischen Schritten zu untersuchen. In keinem der Onlinefälle waren alle diagnostischen Schritte für die Diagnosefindung nötig gewesen, daher galt die Auswahl von allen diagnostischen Schritten als ineffizient. Auch in Bezug auf die Untersuchung der Falldarbietungsformate war die Einführung nötig, da so untersucht werden sollte, ob Teilnehmer, die das *serial cue*-Format wie ein *whole case*-Format angewandt hatten, bessere oder schlechtere Ergebnisse erzielen würden. Unsere Studie zeigt, dass Studierenden, die eine hohe Diagnoseineffizienz aufwiesen, keine signifikant höhere Diagnoserichtigkeit aufwiesen. Das bedeutet, dass Probanden im *serial cue*-Format keinen Vorteil hatten, wenn sie den Onlinefall wie ein *whole case*-Format anwandten und alle ihnen gebotenen Informationen einsahen. Mehr Informationen allein bedeuten also noch keine höhere Diagnoserichtigkeit bei der Bearbeitung von Onlinefällen. Zudem zeigt sich eine Tendenz, dass Studierende mit hoher Diagnoseineffizienz eine niedrigere Diagnoserichtigkeit aufwiesen als Studierende mit niedriger Diagnoseineffizienz. Vereinfacht bedeutet das, dass Studierende, die sich alle Informationen angesehen hatten, tendenziell schlechter diagnostizierten als Studierende, die dies nicht getan hatten. Studierende, die alle diagnostischen Schritte im *serial cue*-Format gewählt hatten, taten dies möglicherweise aus Ratlosigkeit. Schmidt und Mamede (2015) meinen, dass es Studierenden schwerfällt, die richtige Diagnostik auszuwählen, um ihre Hypothesen zu überprüfen.

Nur ca. 23 % der Teilnehmer führten in keinem der Onlinefälle alle Untersuchungen durch. Das heißt 77 % führten in mindestens einem oder mehreren Onlinefällen alle Untersuchungen durch. Eine Erklärung dafür könnte sein, dass die Hemmschwelle,

bei einem virtuellen Patienten unnötige Diagnostik durchzuführen, generell geringer sein könnte. Im Gegensatz zum echten Patienten sind die Ergebnisse im Onlinefall sofort einsehbar, keine Budgets werden belastet, Untersuchungen müssen nicht vor einem Vorgesetzten gerechtfertigt werden und dem virtuellen Patienten entstehen keine Unannehmlichkeiten.

Elstein und Schwarz (2002) zeigen, dass die Diagnoserichtigkeit vom effizienten Umgang mit Informationen abhängt. Passend dazu erkannten bei Nendaz et al. (2000) Teilnehmer mit richtig gestellter Diagnose signifikant mehr kritische Informationen als Teilnehmer mit falsch gestellter Diagnose. Kritische Informationen können hier annähernd mit nötigen diagnostischen Schritten gleichgesetzt werden. Sie erklärten sich den Effekt dadurch, dass die Teilnehmer wohl sehr früh an die richtigen Differentialdiagnosen gedacht hatten (Barrows & Feltovich, 1987). Elstein und Schwarz (2002) meinen zudem, dass Fehler in der klinischen Entscheidungsfindung durch falsche Hypothesengeneration entstanden. Kurz erklärt bedeutet das, dass wenn am Anfang an die falsche Hypothese gedacht wurde, also an die falsche Diagnose, auch die falschen diagnostischen Schritte ausgewählt wurden, um diese zu überprüfen. Unsere Studie widerspricht den Studien von Nendaz et al. (2000) und Elstein und Schwarz (2002) jedoch. Unsere Studie zeigt, dass Studierende, die eine höhere Diagnoseeffizienz aufwiesen, keine signifikant höhere Diagnoserichtigkeit erreichen konnten als Studierende mit niedriger Diagnoseeffizienz. Eine Erklärung für diesen Unterschied könnte sein, dass zu wenig Studierende effizient diagnostiziert hatten, um einen signifikanten Effekt aufzeigen zu können. Eine deskriptive Analyse der Daten zeigt, dass insgesamt nur 7 % der Studierenden in mindestens einem Onlinefall effizient diagnostiziert hatten. Es scheint den Studierenden also schwerzufallen, effiziente Diagnostik anzuwenden. Die Studierenden wissen gegebenenfalls einfach auch noch nicht, was effiziente Diagnostik ausmacht (Schmidt

& Mamede, 2015). Dass effizientes Diagnostizieren in unserer Studie nicht automatisch zur richtigen Diagnose führte, könnte auch darauf hinweisen, dass es Medizinstudierende noch überfordert gleichzeitig effizient und richtig zu diagnostizieren. Effizientes Diagnostizieren muss gegebenenfalls separat trainiert und gefördert werden. Braun et al. (2017) zeigen zum Beispiel, dass durch eine gezielte Anleitung zur Repräsentation von Informationen die Diagnoseeffizienz signifikant gesteigert werden konnte. Hier sollte jedoch beachtet werden, dass dort die Diagnoseeffizienz anders definiert wurde. Braun et al. (2017) definieren die Diagnoseeffizienz als die Anzahl der richtig gelösten Onlinefälle geteilt durch die Bearbeitungszeit der Onlinefälle. Die Diagnoseeffizienz war somit abhängig von der Zeit und der Diagnoserichtigkeit. Sie begründeten die Zeitabhängigkeit damit, dass Ärzte im Berufsleben auch unter zeitlichem Druck stünden. Die Diagnoseeffizienz in unserer Studie ist unabhängig von der Bearbeitungszeit und der Diagnoserichtigkeit definiert. Sie ist die Anzahl der Onlinefälle mit nur den notwendig durchgeführten diagnostischen Schritten geteilt durch die Gesamtzahl der Onlinefälle. Unsere Studie sollte untersuchen, ob Medizinstudierende mit effizienter Diagnostik auch mehr Onlinefälle richtig lösen könnten. Die Diagnoseeffizienz musste daher unabhängig von der Diagnoserichtigkeit definiert sein.

Abschließend kann gesagt werden, dass die Diagnoseeffizienz kein Mittel zur Förderung der Diagnoserichtigkeit ist. Sie ist den Ergebnissen unserer Studie nach eher eine eigene diagnostische Kompetenz. Um sie zu fördern, sollte der Fokus darauf gerichtet sein, den Studierenden beizubringen, früh richtige Hypothesen aufzustellen (Barrows & Feltovich, 1987) und diese Hypothesen zu überprüfen (Schmidt & Mamede, 2015). Für die weitere Förderung könnte dies auch bedeuten, Diagnoseeffizienz anfangs getrennt von Diagnoserichtigkeit zu lehren. Gemeinsam

sollten Diagnoserichtigkeit und Diagnoseeffizienz nur unter richtiger Anleitung gelehrt werden, wie zum Beispiel bei Braun et al. (2017), um der möglichen Überforderung, die Medizinstudierende beim effizienten Diagnostizieren erleben, entgegenzuwirken.

### *5.2.3 Diskussion der Auswirkung des Fallformats und der Vorwissensstufe auf den Wissenszuwachs*

Hypothese 5 formulierte die Behauptung, dass Studierende mit höherem Vorwissen im *serial cue*-Format einen größeren Wissenszuwachs haben als im *whole case*-Format. Passend dazu besagte Hypothese 6, dass Studierende mit niedrigerem Vorwissen im *whole case*-Format einen größeren Wissenszuwachs haben als im *serial cue*-Format. Unsere Studie konnte die Annahmen nicht bestätigen. Die optimale Darbietung von Informationen kann allein schon einen großen Wissenszuwachs erzeugen (Avrahami et al., 1997; Hatala et al., 2003). Als solch eine optimale Problemdarbietung hätte das Falldarbietungsformat fungieren können, hat es in unserer Studie aber nicht. Das Falldarbietungsformat hat keine signifikante Wirkung auf den Wissenszuwachs. Das Falldarbietungsformat scheint also keine unterstützenden Effekte auf Lernprozesse zu haben und auch keinen Einfluss auf die optimale Darbietung von Informationen zu haben. Die Testung des Wissenszuwachses ergab, dass der Messzeitpunkt einen signifikant mittleren Effekt ausübte. Also unterscheiden sich die Ergebnisse des Prä- und Posttests signifikant. Bei Betrachtung zeigt sich, dass Studierende mit hohem Vorwissen einen Wissenszuwachs aufwiesen, während Studierende mit niedrigem Vorwissen keinen aufwiesen. Prozedurales Wissen schien aber bisher unabhängig vom deskriptiven Wissen zu sein (Schmidmaier et al., 2013). In unserer Studie scheint es aber doch zusammenzuhängen. Eine Erklärung für diesen Unterschied könnte sein, dass in der Studie von Schmidmaier et al. (2013) Studierende zuerst eine Lernphase für



deskriptives Wissen durchliefen, bevor dieses abgefragt wurde. Das deskriptive Wissen war also vorher beeinflusst worden. In unserer Studie wurde deskriptives Wissen durch Multiple-Choice-Fragen ohne vorherige Lernphase direkt abgefragt. Es entspricht damit dem aktuellen Wissensstand der Studierenden ohne vorherige Beeinflussung, vergleiche dazu 5.3.6.

Schmidmaier et al. (2013) vermuten, dass prozedurales Wissen durch praktische klinische Erfahrungen zunehmen könnte. Der in unserer Studie aufgetretene Wissenszuwachs könnte bedeuten, dass Studierende mit höherem deskriptivem Vorwissen aus Onlinefällen vergleichbare Erfahrungen sammeln konnten. Andere Studien zeigen, dass die Anwendung von Repräsentation (Braun et al., 2017), Reflexion (Mamede et al., 2014) und Selbsterklärung (Heitzmann, Fischer, Kühne-Eversmann, & Fischer, 2015) erfolgreich Lernprozesse unterstützt. In unserer Studie sind solche Methoden nicht verwendet worden. Der Wissenszuwachs bei Studierenden mit niedrigerem Vorwissen könnte dadurch eventuell behindert worden sein. Dass Studierende mit hohem Vorwissen trotzdem einen signifikanten Wissenszuwachs aufweisen, könnte bedeuten, dass Studierende mit hohem Vorwissen eigenständig solche lernunterstützenden Prozesse angewandt haben.

Klinische Schlussfolgerungen hängen vom Vorwissen (Norman, 2005; Schmidt & Boshuizen, 1993), vom Erlernen der gedanklichen Prozesse des Diagnostizierens (Elstein et al., 1978) und vom analytischen Vorgehen des Diagnostizierens ab (Croskerry & Nimmo, 2011). Studierende mit hohem Vorwissen hätten demnach schon die richtigen gedanklichen Prozesse und die richtigen analytischen Vorgehensweisen erlernt, während Studierende mit niedrigem Vorwissen diese gedanklichen Prozesse und analytischen Vorgehensweisen noch nicht erlernt hätten.

Dies könnte für die Förderung des Wissenszuwachses bedeuten, lernunterstützende Prozesse bei Studierenden weiterhin zu fördern. Wissen sollte verstärkt vermittelt werden und dabei auch Prozesse wie Repräsentation (Braun et al., 2017), Selbsterklärung (Heitzmann et al., 2015) und Reflexion (Mamede et al., 2014) beinhalten. Vor allem Studierenden mit geringem Vorwissen sollte aufgezeigt werden, was für Methoden es gibt und wie diese anzuwenden sind.

## 5.3 DISKUSSION DER METHODEN

### 5.3.1 Studiendesign

Die durchgeführte Studie bestand aus einem 2 x 2-Design. Falldarbietungsformat *serial cue* und *whole case* und Vorwissenstufe hohes und niedriges Vorwissen wurden in vier experimentellen Gruppen variiert. Die Diagnoserichtigkeit und der Wissenszuwachs im prozeduralen Wissen wurden erhoben. Nendaz et al. (2000) verglichen auch das *serial cue*- mit dem *whole case*-Format, auch wenn sie die Falldarbietungsformate anders bezeichneten. Im Gegensatz zu unserer Studie wurden hier drei Vorwissenstufen verglichen, siehe auch 5.3.3. Eine solche Einteilung könnte als realitätsnäher betrachtet werden. Unsere Studie bezieht sich aber explizit auf Medizinstudierende. Eine Einteilung dieser in drei Vorwissenstufen könnte eine neue Möglichkeit der Untersuchung des Vorwissens sein, ist aber nicht so klar abzugrenzen wie bei Nendaz et al. (2000). Pinnock et al. (2012) verglichen ebenfalls zwei Vorwissenstufen, benutzen dabei aber nur das *serial cue*-Format und die Teilnehmer erhielten gleichzeitig *Feedback* auf ihre Auswahl diagnostischer Schritte. Es ist dabei also unklar, ob das *serial cue*-Format als Intervention einen Effekt zeigte oder das *Feedback* oder beides zusammen. Unsere Studie hingegen hat durch eine statistische Analyse mögliche Konfundierungen ausgeschlossen und untersucht gezielt die Effekte der Falldarbietungsformate und des Vorwissens. Al Rumayyan et al. (2018) verglichen

wie unsere Studie das *whole case*-Format mit dem *serial cue*-Format. Sie unterschieden dabei aber keine unterschiedlichen Vorwissensstufen, so wie in unserer Studie. Doch wurden die Studierenden bei Al Rumayyan et al. (2018) im *whole case*-Format angewiesen, einen Selbsterklärungsansatz zu verfolgen, während die Studierenden im *serial cue*-Format angewiesen wurden, den Hypothesenausschlussansatz anzuwenden. Unsere Studie hat auf solche Anweisungen verzichtet um einer Konfundierung zu entgehen. Schuhbach et al. (2017) verglichen das *serial cue*-Format mit Key-Feature Problems, ohne Vorwissensstufen zu unterscheiden. Da Key-Feature Problems prozedurales Wissen abfragen (Kopp et al., 2006), sind diese weniger geeignet, um die Diagnoserichtigkeit zu vergleichen. Daher scheint ein Vergleich von *serial cue*- und *whole case*-Format, wie in unserer Studie durchgeführt, sinnvoller um eine mögliche Auswirkung von auf die Diagnoserichtigkeit zu untersuchen.

Schmidt und Mamede (2015) führten ein Review von 24 empirischen Studien durch. Jedoch ist fraglich, ob man diese Studien alle in den Vergleich ziehen konnte. Die Studiendesigns, die Interventionen und die gemessenen Variablen unterschieden sich, wie schon in 5.2.1 erwähnt, in den einzelnen Studien sehr. Manche verglichen nur die Vorwissensstufe, andere verglichen nur das Falldarbietungsformat und noch andere verfolgten einen gänzlich anderen Ansatz. Die Frage, wie Falldarbietungsformat und Vorwissensstufe sich auf die Diagnoserichtigkeit bei Medizinstudierenden auswirken, konnte damit nicht eindeutig beantwortet werden. Letztlich beziehen sich die Aussagen von Schmidt und Mamede (2015) dabei hauptsächlich auf die Ergebnisse von Nendaz et al. (2000).

Unsere Studie untersuchte die Fragestellung der Auswirkung von Falldarbietungsformat und Vorwissensstufe daher gezielt bei Medizinstudierenden im geeigneten Kontext.

### *5.3.2 Studienteilnehmer*

Es nahmen Studierende des siebten bis zwölften Semesters teil, das heißt des klinischen Abschnitts. Studierende im siebten Semester mussten Modul 23 Teil L6 schon abgeschlossen haben, um teilnehmen zu können, damit Basiswissen zu den Themen Fieber und Rückenschmerzen vorausgesetzt werden konnte. Hier könnte man diskutieren, dass man auch Studierende des vorklinischen Abschnitts hätte miteinbeziehen können. Pinnock et al. (2012) z.B. verglichen Studierende des vorklinischen Abschnitts mit Studierenden des klinischen Abschnitts. Jedoch kann man im vorklinischen Abschnitt kein Basiswissen zu den Themen Fieber und Rückenschmerzen voraussetzen. Als nächstes wäre zu diskutieren, ob eine Teilnahmebedingung, wie der Abschluss eines Studienabschnittes, gerechtfertigt war. Teilnehmer bei Schubach et al. (2017) waren Medizinstudierende des siebten bis zehnten Semesters, die ein einwöchiges Blockpraktikum der Chirurgie abgeschlossen haben mussten. Bei Schubach et al. (2017) scheint dies gerechtfertigt, da es sich in der Studie um chirurgische Themen handelte. Bei uns ebenso, da ein Basiswissen über Erkrankungen und deren diagnostische Abklärung vorhanden sein musste, um Diagnosen stellen zu können. Weitere Studien untersuchten Medizinstudierende des dritten und vierten Semesters (Al Rumayyan et al., 2018; Windish, Price, Clever, Magaziner, & Thomas, 2005), das heißt nur Studierende des vorklinischen Abschnitts. Montaldo und Herskovic (2013) untersuchten Studierende des fünften und sechsten Semesters. Nendaz et al. (2000) untersuchten Studierende des siebten bis zwölften Semesters und zusätzlich noch Assistenzärzte der Inneren Medizin und Fachärzte der

Innenen Medizin. Es wurden in diesen Studien keine weiteren Teilnahmebedingungen gestellt. Eventuell mussten diese gar nicht explizit gestellt werden und waren mit dem Abschluss eines vorherigen Semesters verbunden. Für zukünftige Studien könnte es sinnvoll sein, Teilnahmebedingungen zu begründen, um sie nachvollziehbarer zu machen, wie z.B. auch die Semesteranzahl der Studierenden. Eine Begrenzung der Semesteranzahl nach Basiswissen zu den abgefragten Inhalten wie in unserer Studie erschien sinnvoll.

### *5.3.3 Vorwissensstufe*

In dieser Studie wurden die Teilnehmer durch einen deskriptiven Wissenstest über die Themen Fieber und Rückenschmerzen in zwei Vorwissensstufen eingeteilt, niedriges Vorwissen und hohes Vorwissen. Die Vorwissensstufe orientiert sich hier also nicht an der Anzahl der Semester und war so themenspezifisch und individuell. Ein individueller Ansatz ist vorteilhaft, da eine hohe Semesteranzahl nicht zwingend hohes Vorwissen eines Studierenden zu einem Thema bedingt. Pinnock et al. (2012) unterschieden ebenfalls zwei Vorwissensstufen. Sie unterteilten in Studierende des vorklinischen und des klinischen Abschnitts. Die Ergebnisse ihrer Untersuchung waren somit semesterspezifisch anwendbar. Eine solche Einteilung erscheint bei unserer Fragestellung jedoch wenig sinnvoll, da Studierende der Vorklinik durch den Mangel an klinischen Wissen klar im Nachteil wären. Nendaz et al. (2000) legten ausbildungsspezifisch drei Vorwissensstufen fest. Sie unterteilten in Medizinstudierende, Assistenzärzte der Inneren Medizin und Fachärzte der Inneren Medizin. Die Ergebnisse unserer Studie beziehen sich auf die medizinische Ausbildung von Studierenden, eine Einteilung mit Assistenzärzten und Fachärzten erscheint da ungeeignet. Unsere Ergebnisse könnten jedoch semesterspezifisch und individuell für Studierende angewandt werden. Kurze Prätests, die Studierende vor

Bearbeitung eines Onlinefalls automatisch je nach Ergebnis zu einem *whole case*- oder *serial cue*-Format Onlinefall weiterleiten, wären technisch machbar. Jeder Medizinstudierende könnte so je nach seinem Vorwissen die optimale Falldarbietung erhalten, auch wenn sein Vorwissen dabei das seines Semesters übersteigt oder darunter zurückbleibt. Die Einteilung der Vorwissensstufen unserer Studie war damit gerechtfertigt und vorteilhaft.

#### 5.3.4 Plattform

Diese Studie verwendete das *e-learning* Portal CASUS® (Version 2017, Instruct AG, München) zur Entwicklung und Darstellung von virtuellen Patienten in Onlinefällen. In CASUS® konnte genau bestimmt werden, was der Bildschirm auf sogenannten Karten anzeigte und wann diese angezeigt werden sollten. Ein Nachteil von Plattformen wie CASUS® ist, dass sie von Strom, Internet und Betreiber abhängig sind. Bei technischen Problemen hätte eine ganze Erhebung ausfallen können. Ein Vorteil hingegen ist, dass sie die Daten jedes Teilnehmers genauestens erhoben und automatisch in exportierbarer Form speicherten. Pinnock et al. (2012) verwendeten ebenfalls virtuelle Patienten und benützten dafür ein eigenes, extra entwickeltes System, evPaeds. EvPaeds war ein System mit Decision tree Software und *Feedback*komponenten (Pinnock et al., 2012). CASUS® hatte keine Decision tree Software. Doch konnten in Casus® auch diagnostische Entscheidungen abgefragt werden, da es die Auswahl diagnostischer Schritte im *serial cue*-Format anzeigen konnte. Eine *Feedback*-Komponente hätte zudem in CASUS® auch eingearbeitet werden können. Unsere Studie verzichtete jedoch mit Absicht auf *Feedback*, da es als eigene Intervention gesehen werden könnte. Andere Studien verwendeten keine virtuellen Patienten. Bei Kiesewetter et al. (2016) lagen die Patienteninformationen in Papierform vor und auf Nachfrage bekam der Proband ein Blatt mit der nächsten

Untersuchung gereicht. Die Daten stammten aus echten Patientenfällen. Noch andere Studien verwendeten echte Patienten zur Darbietung ihrer Fälle (Goss, 1996), (Linn et al., 2012), (Eugène JFM Custers, Stuyt, & Robbé, 2000). Littlefield et al. (1999) benützten Schauspieler als Patienten. Bei Windish (2000) und Windish et al. (2005) spielten andere Studierende die Patienten. Echte Patienten sind natürlich realitätsnaher als virtuelle Patienten. Sie bieten nicht nur reine Informationen, sondern auch Sinneseindrücke und emotionale Erlebnisse. Doch ist bei Erhebungen dieser Art zu beachten, dass Daten von Hand genau protokolliert werden mussten. Es ist eine zeitaufwendige und fehleranfällige Methode. Ebenso kann die Darstellung eines Patienten von einem Tag zum anderen sehr variieren. Virtuelle Patienten hingegen werden jedes Mal gleich dargestellt. Wenn weder echte noch virtuelle Patienten verwendet wurden, wie zum Beispiel bei Kassirer (2010) und Dequeker und Jaspaert (1998) und die Patienteninformationen von einem Versuchsleiter vorgetragen wurden, handelte es sich um eine Mischform. Eine Mischform beider Varianten mit Stärken und Schwächen beider Seiten. Eine Stärke dieser gemischten Methode war, dass die Untersuchungen so technisch und personell unabhängig waren. Wir entschieden uns für virtuelle Patienten, da diese Methode standardisiert durchführbar war und zudem personell unabhängig war. Daten konnten auch exakter erfasst werden, da das System von CASUS® diese genau aufzeichnete. Zudem gaben Studierende bei echten Patienten mehr Probleme an, wenn es darum ging, wichtige Informationen herauszufiltern (Lysaght & Bent, 2005). Abschließend kann gesagt werden, dass ein Fokus dieser Studie auf der Förderung von diagnostischen Kompetenzen im *e-learning* lag. Die Ergebnisse sollten über Präsenzveranstaltungen hinaus auch beim selbstständigen Lernen Zuhause anwendbar sein. Echte Patienten zu instrumentalisieren erschien in dieser Situation zu Recht ungeeignet. In technischer Hinsicht war CASUS® EvPaeds bei der Fragestellung unserer Studie überlegen, da

*Feedback*-Komponenten abgeschaltet werden konnten. Im gesamten Erhebungszeitraum traten kein einziges Mal Probleme mit der technischen Infrastruktur auf. CASUS® war somit die optimale Plattform für die Entwicklung, die Darstellung und die Datenerhebung unserer Studie.

### *5.3.5 Diagnoserichtigkeit*

Die Diagnoserichtigkeit wurde durch acht zu lösende Onlinefälle mit den Themen Fieber und Rückenschmerzen gemessen. Eine richtig gestellte Diagnose ergab einen Punkt. Maximal waren acht Punkte und minimal null Punkte zu erreichen. Es handelte sich dabei nicht um echte Patientenfälle. Die Onlinefälle wurden aus Lehrbuchinformationen zu den jeweiligen Diagnosen zusammengestellt. Andere Studien benutzten Daten aus echten Patientenfällen, z.B. (Al Rumayyan et al., 2018; Nendaz et al., 2000; Pinnock et al., 2012). Daten aus Lehrbüchern erschienen vielleicht zu passend für die jeweiligen Diagnosen. Daten aus echten Patientenfällen sind realitätsnäher, könnten aber auch ablenkende Zusatzinformationen beinhalten. Ein Vorteil von aus Lehrbuchinformationen zusammengestellten Onlinefällen war, dass die Daten keine Begrenzung aufwiesen. So viele diagnostische Untersuchungen wie nötig konnten erstellt werden. Fehlte z.B. eine Untersuchung bei einem echten Patientenfall, konnte man diese nicht nachfordern. Cook und Triola (2009) empfehlen ausdrücklich, virtuelle Patienten zu erstellen, daran hat sich unsere Studie gehalten.

Bei Nendaz et al. (2000) mussten die Teilnehmer keine endgültige Diagnose stellen, sondern drei Differentialdiagnosen benennen und nach ihrer Wahrscheinlichkeit ordnen. Nach dieser Ordnung wurden dann Punkte vergeben. Die Messung der Diagnoserichtigkeit bezog sich dabei auf ein akutes Problem (Nendaz et al., 2000) und man konnte sehen in welche diagnostische Richtung gedacht wurde. Nachteilig dabei war jedoch, dass dadurch keine finale Diagnose festgelegt werden musste. Ein



einfaches Punktesystem für richtig gestellte Diagnosen, wie in unserer Studie verwendet, ist leichter auszuwerten als eine Rangordnung wie bei Nendaz et al. (2000). Ebenso ist die Festlegung auf eine Diagnose realitätsnaher. Rangordnungen von Diagnosen spielen im ärztlichen Alltag eine kleine Rolle. Ab einem bestimmten Punkt muss eine Diagnose festgelegt werden, um eine spezifische Therapie einleiten zu können. Die Variable Diagnoserichtigkeit war in unserer Studie somit ideal definiert und abgefragt worden.

Braun et al. (2017) definierten Diagnoseeffizienz als die Anzahl der richtig gelösten Onlinefälle dividiert durch die Bearbeitungszeit aller Onlinefälle. Sie begründeten die Zeitabhängigkeit damit, dass Ärzte im Berufsleben unter zeitlichem Druck arbeiten müssten. Doch die Diagnoseeffizienz war damit auch abhängig von der Diagnoserichtigkeit. Unsere Studie definierte Diagnoseeffizienz unabhängig von Zeit und Diagnoserichtigkeit. Sie war definiert als der Mittelwert der Anzahl der Onlinefälle, in denen nur die zur Diagnosefindung nötigen diagnostischen Untersuchungen durchgeführt wurden. Unnötig durchgeführte Untersuchungen sind nicht nur ein Mehraufwand an Kosten an das Gesundheitssystem, sondern belasten auch die betroffenen Patienten. Es sollte untersucht werden, ob Medizinstudierende, die effiziente Diagnostik durchführten, auch mehr Onlinefälle richtig diagnostizierten. Die Diagnoseeffizienz musste also unabhängig von der Diagnoserichtigkeit sein. Zudem definierte unsere Studie auch Diagnoseineffizienz. Diagnoseineffizienz war der Mittelwert der Anzahl der Onlinefälle, in denen alle Untersuchungen durchgeführt wurden. Die Untersuchung der Auswirkung von Diagnoseeffizienz und Diagnoseineffizienz im *serial cue*-Format hatte den Schwachpunkt, dass nur die Hälfte der Teilnehmer miteinbezogen werden konnte. Das heißt, hier war die Teststärke geringer. Doch sollte der Umgang von Studierenden mit einer Auswahl an

diagnostischen Schritten untersucht werden. Im *whole case*-Format gab es keine solche Auswahl. Der Umgang mit der Auswahl diagnostischer Schritte war auch wichtig in Bezug auf die Untersuchung der Falldarbietungsformate. Sollten Teilnehmer das *serial cue*-Format wie ein *whole case*-Format angewandt haben, wäre dies damit nachzuweisen gewesen. Die Diagnoseeffizienz war für die Zwecke unserer Studie angemessen definiert und wurde durch eine konträre Variable, die Diagnoseineffizienz, noch ergänzt.

### 5.3.6 Messinstrumente

In dieser Studie wurden Multiple-Choice-Fragen verwendet um deskriptives Wissen abzufragen und acht Key-Feature Problems, vier in einem Prätest und vier in einem Posttest, um prozedurales Wissen abzufragen. Kiese Wetter et al. (2016) verwendeten eine Lautdenkentechnik und ordneten das gesagte drei Wissenskategorien zu, deskriptives Wissen, prozedurales Wissen und metakognitives Wissen. Die Lautdenkentechnik ist zeitaufwendiger in der Auswertung und nicht bei Gruppenerhebungen am Einzelnen geeignet. Schubach et al. (2017) verwendeten in ihrer Studie auch Key-Feature Problems. Doch verwendeten sie diese als Intervention. Die eigentliche Messung des prozeduralen Wissens geschah durch einen Script-Concordance-Test, der auch klinische Entscheidungsfindungen überprüfen sollte (Charlin, Roy, Brailovsky, Goulet, & Van Der Vleuten, 2000). Die Validität des Script-Concordance-Tests ist gegeben, wenn klare Richtlinien zur Gestaltung vorlagen und Expertenmeinungen übereinstimmten (Lubarsky, Charlin, Cook, Chalk, & van der Vleuten, 2011). Der Script-Concordance-Test hängt sehr von der klinischen Erfahrung eines Probanden ab (Dory, Gagnon, Vanpee, & Charlin, 2012). Da in dieser Studie das prozedurale Wissen von Studierenden mit unterschiedlichen Vorwissenstufen untersucht werden sollte und diese so auch unterschiedlich viel Erfahrung hatten,

erschien der Script-Concordance-Test damit weniger geeignet. Studien belegen zudem die hohe Validität von Key-feature Problems (Bordage, Brailovsky, Carretier, & Page, 1995; R. Hatala & Norman, 2002; Trudel, Bordage, & Downing, 2008). Die Verwendung von Key-Feature Problems als Messinstrument für erfahrungsabhängiges prozedurales Wissen ist in unserer Studie gerechtfertigt. Schmidmaier et al. (2013) benützten ebenso Key-Feature Problems, um prozedurales Wissen abzufragen. Deskriptives Wissen wurde bei Schmidmaier et al. (2013) erfragt, indem die Studierenden dieses bei Hinweisen aktiv wiedergeben mussten. Studierende durchliefen bei Schmidmaier et al. (2013) jedoch zuerst eine Lernphase für deskriptives Wissen, bevor dieses abgefragt wurde. Deskriptives Wissen hatte dort also eine Verknüpfung mit dem Kurzzeitgedächtnis und war zuvor beeinflusst worden. In unserer Studie wurde deskriptives Wissen ohne Lernphase direkt abgefragt und entsprach damit dem individuellen Wissensstand der Studierenden ohne vorherige Beeinflussung. Da die Auswirkungen von Falldarbietungsformaten und Vorwissen untersucht werden sollte und nicht die Auswirkung von vorherigen Lernphasen, war die Abfrage des deskriptiven Wissens in unserer Studie somit gut gewählt.

### *5.3.7 Pilotierung*

Die Onlinefälle und Messinstrumente wurden von acht freiwilligen Teilnehmern pilotiert. Es nahmen zwei Assistenzärztinnen im ersten Jahr und sechs Medizinstudierende teil. Zwei Studierende beendeten die Pilotierung nicht, deshalb gehen nur die Ergebnisse von sechs Teilnehmern in die Auswertung mit ein. Eine Stärke dieser Vorgehensweise war, dass die Pilotierung unter nahezu gleichen Bedingungen stattfand, wie die Datenerhebung selbst. Etwaige Probleme mit CASUS® wären aufgefallen und eine zeitliche Einordnung der Bearbeitungszeit konnte ermittelt werden. Die Miteinbeziehung von Assistenzärztinnen im ersten Jahr

brachte Vorteile. Zum einen war es ein zusätzliches Review der Onlinefälle und Messinstrumente, zum anderen wurden damit auch zwei Vorwissenstufen pilotiert. Eine Schwäche war jedoch, dass die Pilotierung eben doch nicht unter exakt gleichen Bedingungen stattfand. Die Pilotierung war eine Erhebung am Einzelnen Zuhause, die Datenerhebung eine Gruppenerhebung am Einzelnen in Computerräumen. Freilich wurden die Teilnehmer gebeten, in einem ausgeruhten Zustand zu arbeiten und sich nur auf die Bearbeitung des Pilotkurses zu konzentrieren, doch konnte nicht garantiert werden, dass sie es auch einhielten. Die Teilnehmerzahl von acht, bzw. letztendlich sechs, ist gering und damit auch die Aussagekraft. Doch zeigen Pilotierungen in der Regel schnell einen Sättigungseffekt (Sandelowski, 1995), das bedeutet, man braucht nur wenige Teilnehmer für eine Pilotierung. In der Studie von Pinnock et al. (2012) ging ein Versuchsleiter mit einer kleinen Gruppe von freiwilligen Studierenden die verwendeten Patientenfälle durch, um diese zu pilotieren. Die Studierenden wurden bei der Bearbeitung der Fälle beobachtet und mussten Fragen dazu beantworten (Pinnock et al., 2012). Eine genaue Anzahl der Teilnehmer ist nicht genannt worden. So kann man die Aussagekraft beider Pilotierungen nicht vergleichen. Pinnock et al. (2012) wichen in der Phase der Pilotierung, eine Gruppenerhebung, von der Situation der Datenerhebung, eine Erhebung am Einzelnen, ab. Ein Vorteil jedoch war, dass die Studierenden durch die Fragen *Feedback* zu den Patientenfällen gaben. In unserer Studie bestand für Pilotteilnehmer ebenso die Möglichkeiten *Feedback* zu den bearbeiteten Onlinefällen zu geben. Es stand ihnen offen am Ende eines Onlinefalles einen Freitext mit Anmerkungen zu verfassen. Vier Teilnehmer nutzen dieses Angebot. Eine der Versuchssituation am nächsten kommenden Situation der Pilotierung mit zusätzlicher Möglichkeit der *Feedback*-Gabe ist für unsere Studie daher optimal.

### 5.3.8 Studiendurchführung

Es fand eine Gruppenerhebung am Einzelnen statt. Probanden zogen einen anonymen Pin und bearbeiteten daraufhin einen Prätest, acht Onlinefälle im *serial cue*- oder *whole case*-Format und einen Posttest. Bei Schubach et al. (2017) erfolgten die Interventionen an zwei Seminartagen und die Datenerhebung an einem dritten. Ein Vorteil gegenüber unserer Vorgehensweise ist, dass somit ein längerfristiger Wissenszuwachs hätte abgebildet werden können. Doch ist der Stundenplan eines Medizinstudierenden im klinischen Abschnitt der Ludwigs-Maximilians-Universität kompakt gefüllt und recht individuell. Drei Erhebungstage für erhoffte 120 Teilnehmer zu organisieren hätte nicht funktioniert, ohne dabei einen großen Teil der Teilnehmer aus zeitlichen Gründen zu verlieren. Windish (2000) erhob Daten an zweier Paaren, die gemeinsam einen Fall bearbeiteten. Bei Al Rumayyan et al. (2018) mussten sechs Studierende gemeinsam eine Diagnose für acht Patientenfälle stellen. Dies sind kooperative Vorgehensweisen, die wieder selbst als Interventionen gesehen werden könnten. In der Untersuchung von Nendaz et al. (2000) bearbeitete die eine Hälfte der Teilnehmer Patientenfall 1 und die andere Patientenfall 2. Dadurch verringert sich die Aussagekraft der Diagnoserichtigkeit. Falls eine Diagnose falsch gestellt wurde, fiel das bei Nendaz et al. (2000) stark ins Gewicht. Dagegen kann wie in unserer Studie die Bearbeitung beziehungsweise die Lösung von acht Onlinefälle die Diagnoserichtigkeit robuster gegen einzelne Fehler abgebildet haben. Konfundierungen durch *Feedback* oder kooperatives Arbeiten wurden durch die Vermeidung dieser Methoden ausgeschlossen. Die gewählte Studiendurchführung war somit gut geeignet, um die Fragestellungen unserer Studie zu untersuchen.

## 6 Kurzzusammenfassung

---

**Einleitung:** Diagnostische Kompetenzen sind kontextspezifische, kognitive Leistungen (Koeppen et al., 2008). Sie sind unabdingbar für die Arztprofession und sollten während des Medizinstudiums vermittelt werden (Nationaler Kompetenz-basierter Lernzielkatalog Medizin (NKLM) 2015). Virtuelle Patienten bieten sich an, um diagnostische Kompetenzen bei Studierenden zu trainieren (Lang et al., 2013). Die Wirkung der Unterstützung von Lernprozessen, wie zum Beispiel das Falldarbietungsformat von virtuellen Patienten, scheint von der Phase des Kompetenzerwerbs und dem Vorwissen abzuhängen (Kalyuga, 2007). Bisher ist unklar wie sich die Falldarbietungsformate dabei auf den Wissenszuwachs auswirken (Schmidt & Mamede, 2015).

**Methoden:** Die Falldarbietungsformate der virtuellen Patienten *serial cue* und *whole case* (experimentell) und die Vorwissenstufen niedriges und hohes Vorwissen (quasi-experimentell) wurden in einem 2 x 2-Design variiert. Die diagnostische Kompetenz Diagnoserichtigkeit wurde mittels acht zu lösenden Onlinefällen erhoben. Eine Vorwissenstufe wurde durch einen deskriptiven Wissenstest (Prätest/Multiple-Choice-Fragen) und ein prozeduraler Wissenserwerb (Prä-Posttest/Key-Feature Problems) durch einen Ergebnisvergleich bestimmt. Die Kompetenzen Diagnoseeffizienz und Diagnoseineffizienz wurden anhand der verwendeten diagnostischen Schritte im *serial cue*-Format bestimmt. Eine univariante Varianzanalyse (ANOVA) des Falldarbietungsformats und der Vorwissenstufe wurde mit der Diagnoserichtigkeit als abhängige Variable durchgeführt. Eine weitere univariante Varianzanalyse (ANOVA) mit der abhängigen Variable Diagnoserichtigkeit wurde mit der Diagnoseeffizienz und der Diagnoseineffizienz berechnet. Eine

einstufige messwiederholte Varianzanalyse (repeated ANOVA) des Falldarbietungsformats und der Vorwissensstufe wurde mit den Mittelwerten des prozeduralen Wissens als abhängige Variablen bestimmt.

**Ergebnisse:** Das Falldarbietungsformat hatte **keinen signifikanten** Einfluss auf die Diagnoserichtigkeit oder den Wissenszuwachs, weder allein (Diagnoserichtigkeit:  $F(1,138) = 1.52, p = .22, n.s.$ ; Wissenszuwachs:  $F(1,138) = .75, p = .40, n.s.$ ) noch abhängig von der Vorwissensstufe ( $F(1,138) = .55, p = .46, n.s.$ ; Wissenszuwachs:  $F(1,138) = .75, p = .39, n.s.$ ). Die Vorwissensstufe jedoch hatte sowohl auf die Diagnoserichtigkeit ( $F(1,138) = 9.54, p = .002, \eta^2 = .07$ ) wie auf den Wissenszuwachs ( $F(1,138) = 8.08, p = .01, \eta^2 = .06$ ) einen **signifikanten** Effekt. Diagnoseeffizienz ( $F(1,67) = .46, p = .50, n.s.$ ) und Diagnoseineffizienz ( $F(1,67) = 2.8, p = .09, n.s.$ ) übten **keinen signifikanten** Einfluss auf die Diagnoserichtigkeit aus.

**Fazit:** Die vorliegende Dissertation liefert Argumente, dass eine vorwissensabhängige Falldarbietung mit Formaten wie *serial cue* und *whole case* für Medizinstudierende nicht geeignet ist. Zwar gab es in der Literatur Hinweise, dass Studierende mit unterschiedlichem Vorwissen von verschiedenen Falldarbietungsformaten profitieren könnten (Nendaz et al., 2000; Schmidt & Mamede, 2015), doch widerspricht unsere Studie diesen Vermutungen. Die optimale vorwissensabhängige Falldarbietungsform konnte in dieser Studie nicht gefunden werden. Um diagnostische Kompetenzen und die Diagnoserichtigkeit zu fördern, sollte den Studierenden mehr Vorwissen über Krankheiten und deren diagnostische Abklärung vermittelt werden. Die Diagnoseeffizienz scheint dabei eine eigene diagnostische Kompetenz zu sein, die separat gelehrt werden sollte.

## 7 Literaturverzeichnis

---

- Abs, H. J. (2007). Überlegungen zur Modellierung diagnostischer Kompetenz bei Lehrerinnen und Lehrern. *Forschung zur Lehrerbildung. Kompetenzentwicklung und Programmevaluation*, 63-84.
- Al Rumayyan, A., Ahmed, N., Al Subait, R., Al Ghamdi, G., Mahzari, M. M., Mohamed, T. A., . . . Schmidt, H. G. (2018). Teaching clinical reasoning through hypothetico-deduction is (slightly) better than self-explanation in tutorial groups: An experimental study. *Perspectives on Medical Education*, 1-7.
- Anderson, K. J. (2006). *Factors affecting the development of undergraduate medical students' clinical reasoning ability*.
- Association, of, American, Medical, & Colleges. ( 2007). Effective Use of Educational Technology in Medical Education: Summary Report of the 2006 AAMC Colloquium on Educational Technology. Washington, DC. *AAMC*, 7.
- Avrahami, J., Kareev, Y., Bogot, Y., Caspi, R., Dunaevsky, S., & Lerner, S. (1997). Teaching by examples: Implications for the process of category acquisition. *The Quarterly Journal of Experimental Psychology: Section A*, 50(3), 586-606.
- Barrows, H. S., & Feltovich, P. J. (1987). The clinical reasoning process. *Medical education*, 21(2), 86-91.
- Berman, N. B., Durning, S. J., Fischer, M. R., Huwendiek, S., & Triola, M. M. (2016). The Role for Virtual Patients in the Future of Medical Education. *Academic Medicine*, 91(9), 1217-1222. doi:10.1097/acm.0000000000001146
- Boeker, M., Andel, P., Vach, W., & Frankenschmidt, A. (2013). Game-based e-learning is more effective than a conventional instructional method: a randomized controlled trial with third-year medical students. *PLoS One*, 8(12), e82328. doi:10.1371/journal.pone.0082328
- Bordage, G., Brailovsky, C., Carretier, H., & Page, G. (1995). Content validation of key features on a national examination of clinical decision-making skills. *Academic Medicine*.
- Braun, L., Zottmann, J., Adolf, C., Lottspeich, C., Then, C., Wirth, S., . . . Schmidmaier, R. (2017). Representation scaffolds improve diagnostic efficiency in medical students. *Medical education*.
- Braun, L., Zwaan, L., Kiesewetter, J., Fischer, M., & Schmidmaier, R. (2017). Diagnostic errors by medical students: Results of a prospective qualitative study. 17. doi:10.1186/s12909-017-1044-7
- Brendel, T. (2017). „Schriftliche Prüfungen in der medizinischen Lehre“.
- Charlin, B., Roy, L., Brailovsky, C., Goulet, F., & Van Der Vleuten, C. (2000). The Script Concordance test: a tool to assess the reflective clinician. *Teaching and learning in medicine*, 12(4), 189-195.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* 2nd edn: Erlbaum Associates, Hillsdale.
- Cook, D. A., & Triola, M. M. (2009). Virtual patients: a critical literature review and proposed next steps. *Medical education*, 43(4), 303-311.
- Croskerry, P., & Nimmo, G. (2011). Better clinical decision making and reducing diagnostic error. *The journal of the Royal College of Physicians of Edinburgh*, 41(2), 155-162.
- Custers, E. J., Boshuizen, H. P., & Schmidt, H. G. (1998). The role of illness scripts in the development of medical diagnostic expertise: Results from an interview study. *Cognition and instruction*, 16(4), 367-398.
- Custers, E. J., Stuyt, P. M., & Robbé, P. F. D. V. (2000). Clinical problem analysis (CPA): a systematic approach to teaching complex medical problem solving. *Academic Medicine*, 75(3), 291-297.



- Dequeker, J., & Jaspaert, R. (1998). Teaching problem-solving and clinical reasoning: 20 years experience with video-supported small-group learning. *MEDICAL EDUCATION-OXFORD-*, 32, 384-389.
- Deshpande, A. A., & Huang, S. H. (2011). Simulation games in engineering education: A state-of-the-art review. *Computer Applications in Engineering Education*, 19(3), 399-410.
- Dhaliwal, G., & Ilgen, J. (2016). Clinical Reasoning: Talk the Talk or Just Walk the Walk? *Journal of graduate medical education*, 8(2), 274-276.
- Dory, V., Gagnon, R., Vanpee, D., & Charlin, B. (2012). How to construct and implement script concordance tests: insights from a systematic review. *Medical education*, 46(6), 552-563.
- Ellaway, R., Poulton, T., Fors, U., McGee, J. B., & Albright, S. (2008). Building a virtual patient commons. *Medical Teacher*, 30(2), 170-174.
- Elstein, A. S., & Schwarz, A. (2002). Evidence base of clinical diagnosis: Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. *BMJ: British Medical Journal*, 324(7339), 729.
- Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). Medical problem solving an analysis of clinical reasoning.
- Fischer, M. R., Hege, I., Hörnlein, A., Puppe, F., Tönshoff, B., & Huwendiek, S. (2008). Virtuelle Patienten in der medizinischen Ausbildung: Vergleich verschiedener Strategien zur curricularen Integration. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, 102(10), 648-653.
- Fleming, M. H. (1991). Clinical reasoning in medicine compared with clinical reasoning in occupational therapy. *American Journal of Occupational Therapy*, 45(11), 988-996.
- Gagne, R., Briggs, L., & Wagner, W. (1992). Introduction. Principles of instructional design (pp. 3-19): Toronto, ON: Harcourt Brace Jovanich College.
- Goss, J. R. (1996). Teaching clinical reasoning to second-year medical students. *Academic medicine: journal of the Association of American Medical Colleges*, 71(4), 349-352; discussion 348.
- Gruppen, L. D., Wolf, F. M., & Billi, J. E. (1991). Information gathering and integration as sources of error in diagnostic decision making. *Medical Decision Making*, 11(4), 233-239.
- Hasegawa, T., Ogasawara, C., & Katz, E. C. (2007). Measuring diagnostic competency and the analysis of factors influencing competency using written case studies. *International Journal of Nursing Knowledge*, 18(3), 93-102.
- Hatala, R., & Norman, G. R. (2002). Adapting the key features examination for a clinical clerkship. *Medical education*, 36(2), 160-165.
- Hatala, R. M., Brooks, L. R., & Norman, G. R. (2003). Practice makes perfect: the critical role of mixed practice in the acquisition of ECG interpretation skills. *Advances in Health Sciences Education*, 8(1), 17-26.
- Hautz, W. E., Kämmer, J. E., Schaubert, S. K., Spies, C. D., & Gaissmaier, W. (2015). Diagnostic performance by medical students working individually or in teams. *Jama*, 313(3), 303-304.
- Hege, I., Kononowicz, A. A., Berman, N. B., Lenzer, B., & Kiesewetter, J. (2018). Advancing clinical reasoning in virtual patients—development and application of a conceptual framework. *GMS Journal for Medical Education*, 35(1).
- Heitzmann, N., Fischer, F., Kühne-Eversmann, L., & Fischer, M. R. (2015). Enhancing diagnostic competence with self-explanation prompts and adaptable feedback. *Medical education*, 49(10), 993-1003.
- Herold, G., & Mitarbeiter, u. (2016). *Innere Medizin* Gerd Herold, Köln.
- Ilgen, J. S., Humbert, A. J., Kuhn, G., Hansen, M. L., Norman, G. R., Eva, K. W., . . . Sherbino, J. (2012). Assessing diagnostic reasoning: a consensus statement summarizing theory, practice, and future needs. *Academic Emergency Medicine*, 19(12), 1454-1461.
- Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational psychology review*, 19(4), 509-539.
- Kassirer, J. P. (2010). Teaching clinical reasoning: case-based and coached. *Academic Medicine*, 85(7), 1118-1124.

- Kiesewetter, J., Ebersbach, R., Görlitz, A., Holzer, M., Fischer, M. R., & Schmidmaier, R. (2013). Cognitive problem solving patterns of medical students correlate with success in diagnostic case solutions. *PLoS One*, *8*(8), e71486.
- Kiesewetter, J., Ebersbach, R., Tsalas, N., Holzer, M., Schmidmaier, R., & Fischer, M. (2016). Knowledge is not enough to solve the problems – The role of diagnostic knowledge in clinical reasoning activities. *16*. doi:10.1186/s12909-016-0821-z
- Kiesewetter, J., Fischer, F., & Fischer, M. R. (2017). Collaborative clinical reasoning—a systematic review of empirical studies. *Journal of Continuing Education in the Health Professions*, *37*(2), 123-128.
- Kiesewetter, J., Kiessling, C., & Fischer, M. (2013). Simulationsbasierte Lehre in der Medizin. Beitrag zur Patientensicherheit und ethischer Imperativ. *Fehler und Ethik in der Medizin Neue Wege für Patientenrechte*. Würzburg: Königshausen & Neumann.
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational psychologist*, *41*(2), 75-86.
- Koeppen, K., Hartig, J., Klieme, E., & Leutner, D. (2008). Current issues in competence modeling and assessment. *Zeitschrift für Psychologie/Journal of Psychology*, *216*(2), 61-73.
- Kogan, J. R., Holmboe, E. S., & Hauer, K. E. (2009). Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. *Jama*, *302*(12), 1316-1326.
- Kopp, V., Möltner, A., & Fischer, M. (2006). Key-Feature-Probleme zum Prüfen von prozeduralem Wissen: Ein Praxisleitfaden. *GMS Z Med Ausbild.*, *23*(23):Doc50.
- .
- Lang, V. J., Kogan, J., Berman, N., & Torre, D. (2013). The evolving role of online virtual patients in internal medicine clerkship education nationally. *Academic Medicine*, *88*(11), 1713-1718.
- Linn, A., Khaw, C., Kildea, H., & Tonkin, A. (2012). Clinical reasoning: A guide to improving teaching and practice. *Australian family physician*, *41*(1/2), 18.
- Littlefield, J. H., Hahn, H. B., & Meyer, A. S. (1999). Evaluation of a role-play learning exercise in an ambulatory clinic setting. *Advances in Health Sciences Education*, *4*(2), 167-173.
- Lubarsky, S., Charlin, B., Cook, D. A., Chalk, C., & van der Vleuten, C. P. (2011). Script concordance testing: a review of published validity evidence. *Medical education*, *45*(4), 329-338.
- Ludwig-Maximilians-Universität (2015, Oktober 22). *Studium*. Retrieved from [https://www.uni-muenchen.de/studium/studienangebot/studiengaenge/studienfaecher/medizin\\_2/staatsexam/index.html](https://www.uni-muenchen.de/studium/studienangebot/studiengaenge/studienfaecher/medizin_2/staatsexam/index.html)
- Lysaght, R., & Bent, M. (2005). A comparative analysis of case presentation modalities used in clinical reasoning coursework in occupational therapy. *American Journal of Occupational Therapy*, *59*(3), 314-324.
- Mamede, S., Schmidt, H. G., Rikers, R. M., Penaforte, J. C., & Coelho-Filho, J. M. (2007). Breaking down automaticity: case ambiguity and the shift to reflective approaches in clinical reasoning. *Medical education*, *41*(12), 1185-1192.
- Mamede, S., Van Gog, T., Sampaio, A. M., De Faria, R. M. D., Maria, J. P., & Schmidt, H. G. (2014). How can students' diagnostic competence benefit most from practice with clinical cases? The effects of structured reflection on future diagnosis of the same and novel diseases. *Academic Medicine*, *89*(1), 121-127.
- Mamede, S., Van Gog, T., Schuit, S. C., Van den Berge, K., Van Daele, P. L., Bueving, H., . . . Schmidt, H. G. (2016). Why patients' disruptive behaviours impair diagnostic reasoning: a randomised experiment. *BMJ Qual Saf*, bmjqs-2015-005065.
- Mattingly, C. (1991). What is clinical reasoning? *American Journal of Occupational Therapy*, *45*(11), 979-986.
- Modi, J. N., Gupta, P., & Singh, T. (2015). Teaching and assessing clinical reasoning skills. *Indian pediatrics*, *52*(9), 787-794.
- Montaldo, G., & Herskovic, P. (2013). Teaching of clinical reasoning to medical students using prototypical clinical cases. *Revista medica de Chile*, *141*(7), 823-830.

- Mutnick, A., & Barone, M. (2014). Assessing and remediating clinical reasoning *Remediation in Medical Education* (pp. 85-101): Springer.
- Nationaler\_Kompetenz-basierter\_Lernzielkatalog, M., & (NKLM). (2015).
- Nendaz, M. R., Raetz, M. A., Junod, A. F., & Vu, N. V. (2000). Teaching diagnostic skills: clinical vignettes or chief complaints? *Advances in Health Sciences Education*, 5(1), 3-10.
- Niethard, F. U., Pfeil, J., & Biberthaler, P. (2014). *Duale Reihe Orthopädie und Unfallchirurgie* (7. Auflage ed.): Georg Thieme Verlag KG, Stuttgart.
- Page, G., Bordage, G., & Allen, T. (1995). Developing key-feature problems and examinations to assess clinical decision-making skills. *Academic Medicine*, 70(3), 194-201.
- Patel, V. L., Evans, D. A., & Groen, G. J. (1989). Biomedical knowledge and clinical reasoning. *Cognitive science in medicine: Biomedical modeling*, 53-112.
- Pinnock, R., Spence, F., Chung, A., & Booth, R. (2012). evPaeds: undergraduate clinical reasoning. *The clinical teacher*, 9(3), 152-157.
- Prensky, M. (2001). Digital natives, digital immigrants part 1. *On the horizon*, 9(5), 1-6.
- Ryle, G. (1949). *The Concept of Mind* (London: Hutchinson's University Library). Cited on, 25.
- Sandelowski, M. (1995). Sample size in qualitative research. *Research in nursing & health*, 18(2), 179-183.
- Schmidmaier, R., Eiber, S., Ebersbach, R., Schiller, M., Hege, I., Holzer, M., & Fischer, M. R. (2013). Learning the facts in medical school is not enough: which factors predict successful application of procedural knowledge in a laboratory setting? *BMC medical education*, 13(1), 28.
- Schmidt, H., Norman, G., & Boshuizen, H. (1990). A cognitive perspective on medical expertise: theory and implications. *Academic Medicine*, 65(10), 611-621.
- Schmidt, H. G., & Boshuizen, H. P. (1993). On acquiring expertise in medicine. *Educational psychology review*, 5(3), 205-221.
- Schmidt, H. G., & Mamede, S. (2015). How to improve the teaching of clinical reasoning: a narrative review and a proposal. *Medical education*, 49(10), 961-973. doi:10.1111/medu.12775
- Schmidt, H. G., Mamede, S., Van Den Berge, K., Van Gog, T., Van Saase, J. L., & Rikers, R. M. (2014). Exposure to media information about a disease can cause doctors to misdiagnose similar-looking clinical cases. *Academic Medicine*, 89(2), 285-291.
- Schmidt, H. G., & Moust, J. (2000). Towards a taxonomy of problems used in problem-based learning curricula. *Journal on Excellence in College Teaching*, 11(2/3), 57-72.
- Schmidt, H. G., & Rikers, R. M. (2007). How expertise develops in medicine: knowledge encapsulation and illness script formation. *Medical education*, 41(12), 1133-1139.
- Schubach, F., Goos, M., Fabry, G., Vach, W., & Boeker, M. (2017). Virtual patients in the acquisition of clinical reasoning skills: does presentation mode matter? A quasi-randomized controlled trial. *BMC Med Educ*, 17(1), 165. doi:10.1186/s12909-017-1004-2
- Simmons, B. (2010). Clinical reasoning: concept analysis. *Journal of Advanced Nursing*, 66(5), 1151-1158.
- Stegmann, K., Fischer, Frank. (2016). Auswirkungen digitaler Medien auf den Wissens- und Kompetenzerwerb an der Hochschule. epub:38264.
- Trudel, J. L., Bordage, G., & Downing, S. M. (2008). Reliability and validity of key feature cases for the self-assessment of colon and rectal surgeons. *Annals of surgery*, 248(2), 252-258.
- Van De Wiel, M. W., Boshuizen, H. P., & Schmidt, H. G. (2000). Knowledge restructuring in expertise development: Evidence from pathophysiological representations of clinical cases by students and physicians. *European Journal of Cognitive Psychology*, 12(3), 323-356.
- Van Merriënboer, J. J., & Sweller, J. (2010). Cognitive load theory in health professional education: design principles and strategies. *Medical education*, 44(1), 85-93.
- Windish, D. M. (2000). Teaching medical students clinical reasoning skills. *Academic Medicine*, 75(1), 90.
- Windish, D. M., Price, E. G., Clever, S. L., Magaziner, J. L., & Thomas, P. A. (2005). Teaching medical students the important connection between communication and clinical reasoning. *Journal of General Internal Medicine*, 20(12), 1108-1113.

## 8 Abbildungsverzeichnis

---

Abbildung 1: 2 x 2 faktorielles Studiendesign und die vier resultierenden experimentellen Gruppen	17
Abbildung 2: Darstellung eines Onlinefalles im <i>serial cue</i> -Format in Casus®	22
Abbildung 3: Darstellung eines Onlinefalles im <i>whole case</i> -Format in Casus®	22
Abbildung 4: Darstellung einer Multiple-Choice-Frage in Casus®	26
Abbildung 5: Darstellung eines Key-Feature Problems in Casus®	29
Abbildung 6: Bildliche Darstellung des Studienablaufs	32
Abbildung 7: Q-Q-Diagramm der Residuen der Diagnoserichtigkeit	38
Abbildung 8: Boxplots der Residuen der Diagnoserichtigkeit	38
Abbildung 9: Diagramm der Diagnoserichtigkeit der vier experimentellen Gruppen	48
Abbildung 10: Diagramme der Diagnoserichtigkeit der hohen und niedrigen Diagnoseineffizienz	49
Abbildung 11: Diagramm der Mittelwerte des prozeduralen Wissens der vier experimentellen Gruppen im Prä- und Posttest	51

## 9 Tabellenverzeichnis

---

Tabelle 1: Reihenfolge der Onlinefälle in der Studiendurchführung	25
Tabelle 2: Paarung der Fragen im Prä- und Posttest nach Schwierigkeit	27
Tabelle 3: Zuordnung der Key-Feature Problems in Prä- und Posttest	29
Tabelle 4: Variablendefinition und Berechnungsvorschriften, Tabelle zweigeteilt über S.34 und 35	35
Tabelle 5: Stichprobeneigenschaften der vier experimentellen Gruppen	41
Tabelle 6: Diagnoserichtigkeit aller Teilnehmer gesamt	42
Tabelle 7: Diagnoserichtigkeit der vier experimentellen Gruppen gesamt	42
Tabelle 8: Lösung der einzelnen Onlinefälle	43
Tabelle 9: Diagnoseeffizienz und Diagnoseineffizienz	44
Tabelle 10: Ergebnisse der Diagnoserichtigkeit des Median-Splits der Diagnoseeffizienz und der Diagnoseineffizienz	44
Tabelle 11: Ergebnisse des prozeduralen Wissenstestes	45
Tabelle 12: Vergleich der Ergebnisse mit den Ergebnissen von Nendaz et al. (2000)	60

# 10 Anhang

## 10.1 Anhang 1: Erläuterung des Modulsystems des klinischen Abschnitts.

„Das Studium setzt sich aus sechs Modulen zusammen, die die Studierenden in einer bestimmten Reihenfolge durchlaufen.“ (Ludwig-Maximilians-Universität, 2015). Die Reihenfolge kann dabei individuell variieren. Nach dem ersten Staatsexamen folgt automatisch das Modul 1. Nach Modul 1 folgt das Modul 23 Teil 1 L6 oder Teil 2 L7. Danach folgt der andere jeweils noch nicht absolvierte Teil von Modul 23. Danach folgt entweder das Modul 4 oder das Modul 5 und im nächsten Semester das noch fehlende der jeweiligen. Dann folgt das reguläre Modul 6. Jedoch kann Modul 6 jederzeit nach Modul 1 vorgezogen werden, es handelt sich dann um ein irreguläres Modul 6. Nach einem irregulären Modul 6, folgt das Modul, das eigentlich an dessen Stelle hätte absolviert werden sollen und die Modulreihenfolge läuft anschließend regulär weiter.

5. Semester	Modul 1
6. Semester	Modul 23 Teil 1 <b>oder</b> Modul 23 Teil 2 <b>oder</b> irreguläres Modul 6
7. Semester	Modul 23 Teil 2 <b>oder</b> Modul 23 Teil 1 <b>oder</b> irreguläres Modul 6
8. Semester	Modul 4 <b>oder</b> Modul 5 <b>oder</b> irreguläres Modul 6
9. Semester	Modul 5 <b>oder</b> Modul 4 <b>oder</b> irreguläres Modul 6
10. Semester	Reguläres Modul 6
	2. Staatsexamen, schriftlicher Teil nach neuer ÄAppO von 2012
11. Semester	Praktisches Jahr
12. Semester	Praktisches Jahr
	2. Staatsexamen, mündlicher Teil nach neuer ÄAppO von 2012

Modul	Inhalt
Modul 1 Grundlagen der klinischen Medizin	Humangenetik, Hygiene, Mikrobiologie, Virologie, Klinische Chemie, Laboratoriumsdiagnostik, Pathologie, Pharmakologie, Toxikologie, Bildgebende Verfahren, Radiologie, Strahlenbehandlung, Strahlenschutz, Infektiologie, Immunologie
<u>Modul 23</u> <u>Interdisziplinäres Basisjahr</u>	<u>Innere Medizin, Chirurgie</u>

Teil 1 L6	Nephrologie, Urologie, Rheuma-Woche, Trauma-Woche, Pulmologie, Blut und Infektiologie, Orthopädie, Unfallchirurgie, Physikalische Medizin
Teil 2 L7	Gastroenterologie, Endokrinologie, Kardiologie, Herzchirurgie, Anästhesiologie, Intensivmedizin, Notfallmedizin, Schmerztherapie
Modul 4 Nervensystem und Sensorium	Dermatologie, Augenheilkunde, HNO, Neurologie und Neurochirurgie, Psychiatrie und Psychotherapie, Psychosomatische Medizin, klinische Pharmakologie (Querschnittsbereich Q 9)
Modul 5 Gezeiten - die Lebensabschnitte	Kinderheilkunde, Frauenheilkunde und Geburtshilfe, Rehabilitation, Geriatrie, Blockpraktikum Allgemeinmedizin, Wahlfach.
Modul 6 Projektsemester	Definierter Freiraum für wissenschaftliches Arbeiten

**10.2 Anhang 2: Tabelle der Onlinefälle mit zur Verfügung stehender Diagnostik und Literaturverweis**

Diagnose	Fallname	Zur Verfügung stehende Diagnostik	Literaturverweis
<b>Bakterielle Endokarditis</b>	Herr Schneider  (ehem. Engelhardt)	<ul style="list-style-type: none"> <li>• Anamnese</li> <li>• Körperliche Untersuchung</li> <li>• Standard Labor</li> <li>• Blutkulturen</li> <li>• EKG</li> <li>• Stuhluntersuchung</li> <li>• TPO-AK, TR-AK Bestimmung</li> <li>• TSH-, fT3-, fT4 Bestimmung</li> <li>• Echokardiographie</li> <li>• Röntgen Thorax</li> </ul>	(Herold & Mitarbeiter, 2016, pp. 158-162)
Bandscheibenprolaps	Herr Baudisch	<ul style="list-style-type: none"> <li>• Anamnese</li> <li>• Körperliche Untersuchung</li> <li>• Standard Labor</li> <li>• Rheumaserologie</li> <li>• EKG</li> <li>• Röntgen LWS</li> <li>• MRT LWS</li> </ul>	(Niethard, Pfeil, & Biberthaler, 2014, p. 392)
chronisch lymphatische Leukämie	Herr Clement	<ul style="list-style-type: none"> <li>• Anamnese</li> <li>• Körperliche Untersuchung</li> <li>• Standard Labor</li> <li>• Rheumaserologie</li> <li>• Hepatitis-Serologie</li> <li>• Blutausstrich</li> <li>• Blutkulturen</li> <li>• EKG</li> <li>• Echokardiographie</li> <li>• Abdomen Sonographie</li> <li>• Röntgen Thorax</li> </ul>	(Herold & Mitarbeiter, 2016, pp. 88-91)
<b>Hepatitis A Infektion</b>	Frau Hoffman  (ehem. Herrmann)	<ul style="list-style-type: none"> <li>• Anamnese</li> <li>• Körperliche Untersuchung</li> <li>• Standard Labor</li> <li>• Dicker Tropfen</li> <li>• Blutausstrich</li> <li>• HIV-Test</li> <li>• Hepatitis-Serologie</li> <li>• Blutkulturen</li> <li>• EKG</li> <li>• Echokardiographie</li> <li>• Röntgen Thorax</li> </ul>	(Herold & Mitarbeiter, 2016, pp. 527, 539-540)

<b>HIV-Infektion</b>	Herr Schwarz (ehem. Hiller)	<ul style="list-style-type: none"> <li>• Anamnese</li> <li>• Körperliche Untersuchung</li> <li>• Standard Labor</li> <li>• EBV-Serologie</li> <li>• Blutkulturen</li> <li>• EKG</li> <li>• Echokardiographie</li> <li>• HIV-Test</li> <li>• Röntgen Thorax</li> </ul>	(Herold & Mitarbeiter, 2016, pp. 898-907)S.898-907
<b>Hyperthyreosis factitia</b>	Frau Klein (ehem. Huber)	<ul style="list-style-type: none"> <li>• Anamnese</li> <li>• Körperliche Untersuchung</li> <li>• Standard Labor</li> <li>• TPO-AK, TR-AK Bestimmung</li> <li>• TSH-, fT3-, fT4 Bestimmung</li> <li>• Blutkulturen</li> <li>• EKG</li> <li>• Stuhluntersuchung</li> <li>• Echokardiographie</li> <li>• Röntgen Thorax</li> <li>• Schilddrüsen Sonographie</li> <li>• Schilddrüsen Szintigraphie</li> </ul>	(Herold & Mitarbeiter, 2016, pp. 762-764)
<b>Lumbago</b>	Herr Müller (ehem. Ludwig)	<ul style="list-style-type: none"> <li>• Anamnese</li> <li>• Körperliche Untersuchung</li> <li>• Standard Labor</li> <li>• Rheumaserologie</li> <li>• Urinstix</li> <li>• Röntgen LWS</li> <li>• MRT LWS</li> <li>• DXA</li> <li>• EKG</li> </ul>	(Niethard et al., 2014, pp. 426-427)
<b>Malaria tertiana Infektion</b>	Herr Maul	<ul style="list-style-type: none"> <li>• Anamnese</li> <li>• Körperliche Untersuchung</li> <li>• Standard Labor</li> <li>• Blutkulturen</li> <li>• EKG</li> <li>• Echokardiographie</li> </ul>	(Herold & Mitarbeiter, 2016, pp. 911-916)



		<ul style="list-style-type: none"> <li>• Dicker Tropfen</li> <li>• Blutausstrich</li> <li>• HIV-Test</li> <li>• Röntgen Thorax</li> <li>• Stuhluntersuchung</li> <li>• Urinstix</li> </ul>	
Nephrolithiasis	Frau Neptun	<ul style="list-style-type: none"> <li>• Anamnese</li> <li>• Körperliche Untersuchung</li> <li>• Standard Labor</li> <li>• TPO-AK, TR-AK Bestimmung</li> <li>• TSH-, fT3-, fT4 Bestimmung</li> <li>• EKG</li> <li>• Röntgen LWS</li> <li>• MRT LWS</li> <li>• Urinstix</li> <li>• Abdomen Sonographie</li> </ul>	(Herold & Mitarbeiter, 2016, pp. 661-664)
<b>Osteoporotische Wirbelkörperfraktur</b>	Frau Fischer (ehem. Oswald)	<ul style="list-style-type: none"> <li>• Anamnese</li> <li>• Körperliche Untersuchung</li> <li>• Standard Labor</li> <li>• EBV-Serologie</li> <li>• EKG</li> <li>• Röntgen LWS</li> <li>• MRT LWS</li> <li>• Urinstix</li> <li>• DXA</li> </ul>	(Herold & Mitarbeiter, 2016, pp. 779-782)  (Niethard et al., 2014, pp. 420-421)
<b>Spinalkanalstenose</b>	Herr Wagner (ehem. Spind)	<ul style="list-style-type: none"> <li>• Anamnese</li> <li>• Körperliche Untersuchung</li> <li>• Standard Labor</li> <li>• Rheumaserologie</li> <li>• Röntgen LWS</li> <li>• MRT LWS</li> <li>• Urinstix</li> </ul>	(Niethard et al., 2014, pp. 429-430)
<b>Spondylitis ankylosans</b>	Herr Maier (ehem. Becker)	<ul style="list-style-type: none"> <li>• Anamnese</li> <li>• Körperliche Untersuchung</li> <li>• Standard Labor</li> <li>• Rheumaserologie</li> <li>• HIV-Test</li> <li>• Röntgen LWS</li> <li>• MRT LWS</li> <li>• Urinstix</li> <li>• EKG</li> </ul>	(Herold & Mitarbeiter, 2016, pp. 675-677)

### 10.3 Anhang 3: Auswertung der statistischen Daten

#### 10.3.1 Test auf Normalverteilung der Residuen

##### Verarbeitete Fälle

	Gültig		Fälle Fehlend		Gesamt	
	N	Prozent	N	Prozent	N	Prozent
Residuum für Summe_Diagnoserichtigkeit	142	95,9%	6	4,1%	148	100,0%

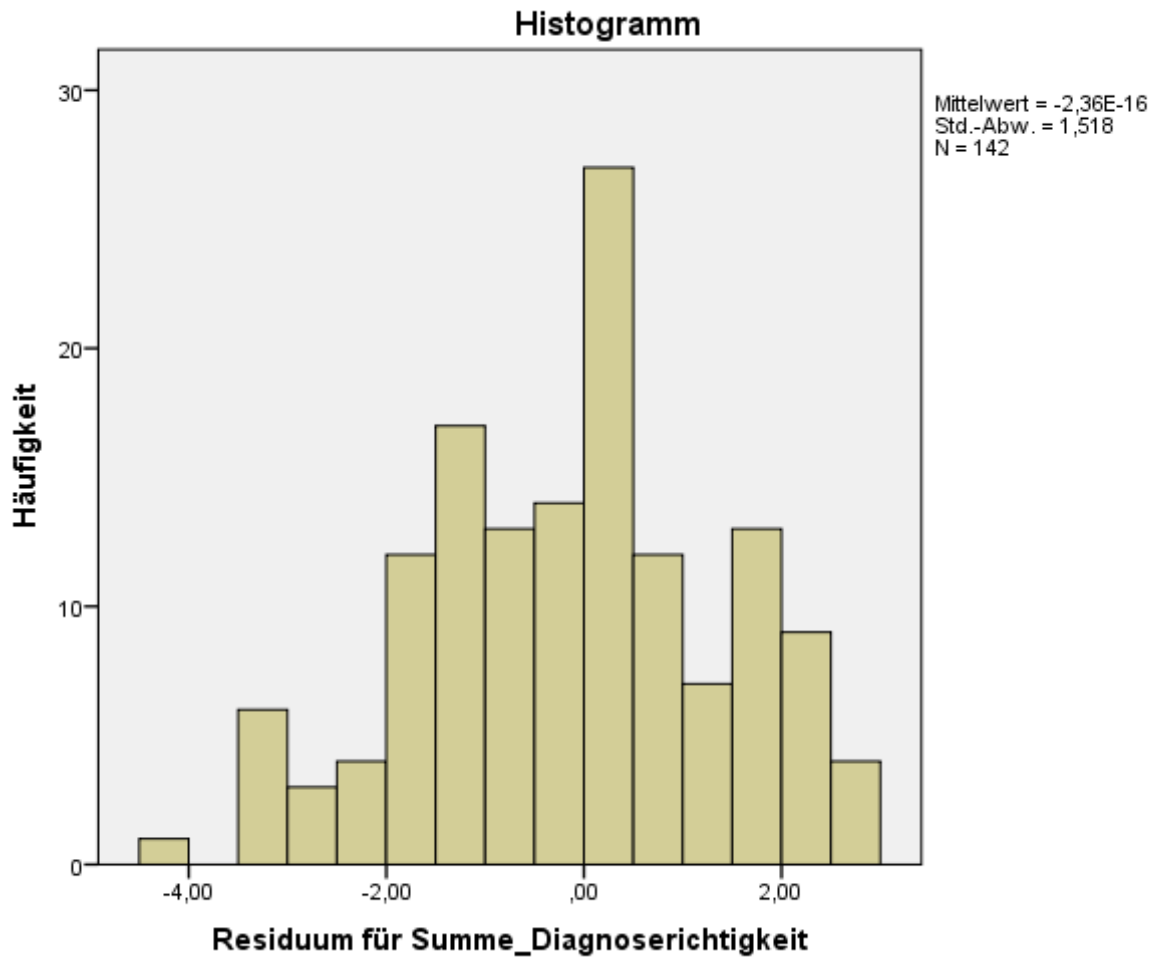
##### Deskriptive Statistik

		Statistik	Standardfehler	
Residuum für	Mittelwert	,0000	,12742	
Summe_Diagnoserichtigkeit	95% Konfidenzintervall des Mittelwerts	Untergrenze Obergrenze	-,2519 ,2519	
	5% getrimmtes Mittel		,0241	
	Median		,0909	
	Varianz		2,305	
	Standardabweichung		1,51838	
	Minimum		-4,04	
	Maximum		2,96	
	Spannweite		7,01	
	Interquartilbereich		2,01	
	Schiefe		-,144	,203
	Kurtosis		-,420	,404

##### Tests auf Normalverteilung

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
Residuum für Summe_Diagnoserichtigkeit	,094	142	,004	,978	142	,020

a. Signifikanzkorrektur nach Lilliefors



### Tests auf Normalverteilung

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
Residuum für Bearbeitungszeit_Mean	,095	142	,003	,961	142	,000
Residuum für Mean_KF_pre	,068	142	,200*	,987	142	,205
Residuum für Mean_KF_post	,068	142	,200*	,987	142	,194

\*. Dies ist eine untere Grenze der echten Signifikanz.

a. Signifikanzkorrektur nach Lilliefors

### 10.3.2 Median des Vorwissens

Häufigkeiten		Median Vorwissen
N	Gültig	142
	Fehlend	6

#### Median Vorwissen

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1,00	6	4,1	4,2	4,2
	2,00	18	12,2	12,7	16,9
	3,00	35	23,6	24,6	41,5
	4,00	34	23,0	23,9	65,5
	5,00	30	20,3	21,1	86,6
	6,00	15	10,1	10,6	97,2
	7,00	3	2,0	2,1	99,3
	8,00	1	,7	,7	100,0
	Gesamt		142	95,9	100,0
Fehlend	System	6	4,1		
Gesamt		148	100,0		

#### Median Vorwissen

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	0,00 (niedriges Vorwissen)	59	39,9	41,5	41,5
	1,00 (hohes Vorwissen)	83	56,1	58,5	100,0
	Gesamt	142	95,9	100,0	
Fehlend	System	6	4,1		
Gesamt		148	100,0		

### 10.3.3 MANOVA der Stichprobenverteilung

#### Zwischensubjektfaktoren

		N
1=sn 2=sh 3=wn 4=wh	1	33
	2	38

3	25
4	45

### Multivariate Tests<sup>a</sup>

Effekt		Wert	F	Hypothese df	Fehler df	Sig.
Konstanter Term	Pillai-Spur	,990	2291,563 <sup>b</sup>	6,000	132,000	,000
	Wilks-Lambda	,010	2291,563 <sup>b</sup>	6,000	132,000	,000
	Hotelling-Spur	104,162	2291,563 <sup>b</sup>	6,000	132,000	,000
	Größte charakteristische Wurzel nach Roy	104,162	2291,563 <sup>b</sup>	6,000	132,000	,000
Studiengruppe.acc	Pillai-Spur	,104	,800	18,000	402,000	,701
	Wilks-Lambda	,899	,794	18,000	373,838	,708
	Hotelling-Spur	,109	,788	18,000	392,000	,715
	Größte charakteristische Wurzel nach Roy	,062	1,387 <sup>c</sup>	6,000	134,000	,224

a. Design: Konstanter Term + Studiengruppe.acc

b. Exakte Statistik

c. Die Statistik ist eine Obergrenze auf F, die eine Untergrenze auf dem Signifikanzniveau ergibt.

### Tests der Zwischensubjekteffekte

Quelle	Abhängige Variable	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
Korrigiertes Modell	0= m, 1= w	,235 <sup>a</sup>	3	,078	,377	,769
	Alter	45,010 <sup>b</sup>	3	15,003	1,811	,148
	0=d, 1=nd	,113 <sup>c</sup>	3	,038	,332	,802
	Abinote	,539 <sup>d</sup>	3	,180	,563	,640
	Physikumsnote	,804 <sup>e</sup>	3	,268	,367	,777
	23=0, 4=1, 5=2, 6=3, PJ=4	7,863 <sup>f</sup>	3	2,621	1,304	,276
Konstanter Term	0= m, 1= w	70,113	1	70,113	338,013	,000
	Alter	79730,395	1	79730,395	9625,975	,000
	0=d, 1=nd	2,350	1	2,350	20,719	,000

	Abinote	334,032	1	334,032	1047,946	,000
	Physikumsnote	922,227	1	922,227	1264,330	,000
	23=0, 4=1, 5=2, 6=3, PJ=4	289,855	1	289,855	144,206	,000
Studiengruppe.acc	0= m, 1= w	,235	3	,078	,377	,769
	Alter	45,010	3	15,003	1,811	,148
	0=d, 1=nd	,113	3	,038	,332	,802
	Abinote	,539	3	,180	,563	,640
	Physikumsnote	,804	3	,268	,367	,777
	23=0, 4=1, 5=2, 6=3, PJ=4	7,863	3	2,621	1,304	,276
Fehler	0= m, 1= w	28,418	137	,207		
	Alter	1134,749	137	8,283		
	0=d, 1=nd	15,542	137	,113		
	Abinote	43,669	137	,319		
	Physikumsnote	99,930	137	,729		
	23=0, 4=1, 5=2, 6=3, PJ=4	275,371	137	2,010		
Gesamt	0= m, 1= w	101,000	141			
	Alter	85106,000	141			
	0=d, 1=nd	18,030	141			
	Abinote	395,000	141			
	Physikumsnote	1061,710	141			
	23=0, 4=1, 5=2, 6=3, PJ=4	605,000	141			
Korrigierte Gesamtvariation	0= m, 1= w	28,652	140			
	Alter	1179,759	140			
	0=d, 1=nd	15,655	140			
	Abinote	44,207	140			
	Physikumsnote	100,734	140			
	23=0, 4=1, 5=2, 6=3, PJ=4	283,234	140			

a. R-Quadrat = ,008 (korrigiertes R-Quadrat = -,014)

b. R-Quadrat = ,038 (korrigiertes R-Quadrat = ,017)

c. R-Quadrat = ,007 (korrigiertes R-Quadrat = -,015)

d. R-Quadrat = ,012 (korrigiertes R-Quadrat = -,009)

e. R-Quadrat = ,008 (korrigiertes R-Quadrat = -,014)

f. R-Quadrat = ,028 (korrigiertes R-Quadrat = ,006)

### 10.3.4 ANOVA KF prä Gruppen

#### Zwischensubjektfaktoren

	N	
1=sn 2=sh 3=wn 4=wh	1	33
	2	38
	3	26
	4	45

#### Tests der Zwischensubjekteffekte

Abhängige Variable: Mean\_KF\_pre

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
Korrigiertes Modell	,012 <sup>a</sup>	3	,004	,756	,520
Konstanter Term	49,689	1	49,689	9514,208	,000
Studiengruppe.acc	,012	3	,004	,756	,520
Fehler	,721	138	,005		
Gesamt	52,572	142			
Korrigierte Gesamtvariation	,733	141			

a. R-Quadrat = ,016 (korrigiertes R-Quadrat = -,005)

### 10.3.5 ANOVA der demographischen Variablen

#### Tests der Zwischensubjekteffekte

Abhängige Variable: Summe\_Diagnoserichtigkeit

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.	Partielles Eta-Quadrat
Korrigiertes Modell	109,306 <sup>a</sup>	36	3,036	1,296	,156	,308
Konstanter Term	398,540	1	398,540	170,150	,000	,618
Geschlechtz	,007	1	,007	,003	,957	,000
Alter	26,453	14	1,889	,807	,660	,097
Muttersprachez	20,234	2	10,117	4,319	,016	,076
Geschlechtz * Alter	45,210	8	5,651	2,413	,020	,155
Geschlechtz * Muttersprachez	2,210	1	2,210	,943	,334	,009
Alter * Muttersprachez	26,904	8	3,363	1,436	,190	,099
Geschlechtz * Alter * Muttersprachez	5,902	2	2,951	1,260	,288	,023
Fehler	245,940	105	2,342			
Gesamt	3177,000	142				
Korrigierte Gesamtvariation	355,246	141				

a. R-Quadrat = ,308 (korrigiertes R-Quadrat = ,070)

### Tests der Zwischensubjekteffekte

Abhängige Variable: Summe\_Diagnoserichtigkeit

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.	Partielles Eta-Quadrat
Korrigiertes Modell	292,650 <sup>a</sup>	112	2,613	1,174	,320	,824
Konstanter Term	1226,289	1	1226,289	551,141	,000	,952
Abinote	28,236	18	1,569	,705	,778	,312
Physikumsnote	14,429	10	1,443	,648	,760	,188
Modulz	2,904	4	,726	,326	,858	,045
Abinote * Physikumsnote	56,568	23	2,459	1,105	,396	,476
Abinote * Modulz	44,538	19	2,344	1,054	,440	,417
Physikumsnote * Modulz	37,923	9	4,214	1,894	,095	,378
Abinote * Physikumsnote * Modulz	,000	0	.	.	.	,000
Fehler	62,300	28	2,225			
Gesamt	3152,000	141				
Korrigierte Gesamtvariation	354,950	140				

a. R-Quadrat = ,824 (korrigiertes R-Quadrat = ,122)

### 10.3.6 Korrelationen nach Pearson

		Korrelationen							
		Diagnose- Neffizienz	Diagnose- effizienz	Summe_Diagn- oserichtigkeit	0= m, 1= w	Alter	0=d, 1=nd	Abinote	Physikum- snote
Diagnose- Neffizienz	Korrelatio- n nach Pearson	1	-,540**	-0,11	-0,041	-0,05	,322**	-0,067	0,074
	Signifikan- z (2-seitig)		0	0,361	0,734	0,702	0,006	0,58	0,54
	N	71	71	71	71	71	71	71	71
Diagnose- effizienz	Korrelatio- n nach Pearson	-,540**	1	0,183	-0,044	0,082	-0,203	0,108	0,038
	Signifikan- z (2-seitig)	0		0,126	0,715	0,495	0,09	0,371	0,754
	N	71	71	71	71	71	71	71	71
Summe_Diagnos- erichtigkeit	Korrelatio- n nach Pearson	-0,11	0,183	1	-0,096	-0,12	-,166*	-,183*	-,287**
	Signifikan- z (2-seitig)	0,361	0,126		0,256	0,173	0,049	0,03	0,001
	N	71	71	142	142	142	142	141	142
0= m, 1= w	Korrelatio- n nach Pearson	-0,041	-0,044	-0,096	1	-0	-0,04	0,025	0,148



	Signifikanz (2-seitig)	0,734	0,715	0,256		0,966	0,638	0,768	0,079
	N	71	71	142	142	142	142	141	142
Alter	Korrelation nach Pearson	-0,046	0,082	-0,115	-0,004	1	0,079	,676**	,204*
	Signifikanz (2-seitig)	0,702	0,495	0,173	0,966		0,351	0	0,015
	N	71	71	142	142	142	142	141	142
0=d, 1=end	Korrelation nach Pearson	,322**	-0,203	-,166*	-0,04	0,079	1	-,166*	0,128
	Signifikanz (2-seitig)	0,006	0,09	0,049	0,638	0,351		0,049	0,129
	N	71	71	142	142	142	142	141	142
Abinote	Korrelation nach Pearson	-0,067	0,108	-,183*	0,025	,676*	-,166*	1	,221**
	Signifikanz (2-seitig)	0,58	0,371	0,03	0,768	0	0,049		0,009
	N	71	71	141	141	141	141	141	141
Physikumsnote	Korrelation nach Pearson	0,074	0,038	-,287**	0,148	,204*	0,128	,221**	1
	Signifikanz (2-seitig)	0,54	0,754	0,001	0,079	0,015	0,129	0,009	
	N	71	71	142	142	142	142	141	142
23=0, 4=1, 5=2, 6=3, PJ=4	Korrelation nach Pearson	-0,224	0,117	0,08	-0,077	,375*	0,04	0,091	-0,12
	Signifikanz (2-seitig)	0,061	0,33	0,344	0,361	0	0,64	0,286	0,153
	N	71	71	142	142	142	142	141	142
Summe_MC_pre	Korrelation nach Pearson	-0,182	0,094	,268**	-0,147	0,037	0,011	-0,012	-0,133
	Signifikanz (2-seitig)	0,128	0,437	0,001	0,08	0,665	0,894	0,892	0,116
	N	71	71	142	142	142	142	141	142
Summe_KF_pre	Korrelation nach Pearson	-,268*	0,172	,301**	,182*	0,023	-,200*	-0,047	-,219**
	Signifikanz (2-seitig)	0,024	0,15	0	0,03	0,787	0,017	0,584	0,009
	N	71	71	142	142	142	142	141	142
Summe_MC_pos	Korrelation nach Pearson	-0,069	0,09	,363**	-0,126	0,036	0,107	-0,008	-,191*
	Signifikanz (2-seitig)	0,569	0,455	0	0,136	0,675	0,204	0,926	0,022
	N	71	71	142	142	142	142	141	142
Summe_KF_post	Korrelation nach Pearson	-,250*	,267*	,324**	-0,154	0,036	-0,077	0,04	-,344**
	Signifikanz (2-seitig)	0,035	0,024	0	0,068	0,67	0,365	0,64	0
	N	71	71	142	142	142	142	141	142

Bearbeitungszeit _Summe_korr	Korrelation nach Pearson	,384**	-0,245	-0,122	-0,013	,181*	0,132	0,123	0,063
	Signifikanz (2-seitig)	0,002	0,053	0,174	0,889	0,043	0,141	0,172	0,484
	N								

\*\* . Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

\* . Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

		23=0, 4=1, 5=2, 6=3, PJ=4	Summe_ MC_pre	Summe_KF _pre	Summe_ MC_post	Summe_KF _post	Bearbeitungszeit_Summe_ korr
DiagnoseIneffizienz z	Korrelation nach Pearson	-,224	-,182	-,268*	-,069	-,250*	,384**
	Signifikanz (2-seitig)	,061	,128	,024	,569	,035	,002
	N	71	71	71	71	71	63
Diagnoseeffizienz	Korrelation nach Pearson	,117	,094	,172	,090	,267*	-,245
	Signifikanz (2-seitig)	,330	,437	,150	,455	,024	,053
	N	71	71	71	71	71	63
Summe_Diagnoserichtigkeit	Korrelation nach Pearson	,080	,268**	,301**	,363**	,324**	-,122
	Signifikanz (2-seitig)	,344	,001	,000	,000	,000	,174
	N	142	142	142	142	142	126
0=m, 1=w	Korrelation nach Pearson	-,077	-,147	,182*	-,126	-,154	-,013
	Signifikanz (2-seitig)	,361	,080	,030	,136	,068	,889
	N	142	142	142	142	142	126
Alter	Korrelation nach Pearson	,375**	,037	,023	,036	,036	,181*
	Signifikanz (2-seitig)	,000	,665	,787	,675	,670	,043
	N	142	142	142	142	142	126
0=d, 1=nd	Korrelation nach Pearson	,040	,011	-,200*	,107	-,077	,132
	Signifikanz (2-seitig)	,640	,894	,017	,204	,365	,141

	N	142	142	142	142	142	126
Abinote	Korrelation nach Pearson	,091	-,012	-,047	-,008	,040	,123
	Signifikanz (2-seitig)	,286	,892	,584	,926	,640	,172
	N	141	141	141	141	141	125
Physikumsnote	Korrelation nach Pearson	-,120	-,133	-,219**	-,191*	-,344**	,063
	Signifikanz (2-seitig)	,153	,116	,009	,022	,000	,484
	N	142	142	142	142	142	126
23=0, 4=1, 5=2, 6=3, PJ=4	Korrelation nach Pearson	1	,197*	,126	,155	,375**	-,157
	Signifikanz (2-seitig)		,019	,136	,066	,000	,079
	N	142	142	142	142	142	126
Summe_MC_pre	Korrelation nach Pearson	,197*	1	,076	,181*	,230**	-,027
	Signifikanz (2-seitig)	,019		,369	,032	,006	,765
	N	142	142	142	142	142	126
Summe_KF_pre	Korrelation nach Pearson	,126	,076	1	,054	,334**	-,100
	Signifikanz (2-seitig)	,136	,369		,525	,000	,266
	N	142	142	142	142	142	126
Summe_MC_post	Korrelation nach Pearson	,155	,181*	,054	1	,065	-,067
	Signifikanz (2-seitig)	,066	,032	,525		,443	,458
	N	142	142	142	142	142	126
Summe_KF_post	Korrelation nach Pearson	,375**	,230**	,334**	,065	1	-,165
	Signifikanz (2-seitig)	,000	,006	,000	,443		,064
	N	142	142	142	142	142	126
Bearbeitungsort	Korrelation nach Pearson	-,157	-,027	-,100	-,067	-,165	1
	Signifikanz (2-seitig)	,079	,765	,266	,458	,064	

N	126	1	126	1	1	126	1	1	1	126
		2		2	2		2	2	2	
		6		5	6		6	6	6	

### 10.3.7 Diagnoseeffizienz und Diagnoseineffizienz

#### Deskriptive Statistik

	N	Minimum	Maximum	Mittelwert	Standardabweichung
Diagnoseeffizienz	71	,00	3,00	,6056	,90204
Diagnoseineffizienz	71	,00	7,00	3,1549	2,83674
Gültige Werte (Listenweise)	71				

#### Diagnoseeffizienz

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	,00	43	29,1	60,6	60,6
	1,00	18	12,2	25,4	85,9
	2,00	5	3,4	7,0	93,0
	3,00	5	3,4	7,0	100,0
	Gesamt	71	48,0	100,0	
Fehlend	System	77	52,0		
Gesamt		148	100,0		

#### Diagnoseineffizienz

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	,00	16	10,8	22,5	22,5
	1,00	14	9,5	19,7	42,3
	2,00	8	5,4	11,3	53,5
	3,00	5	3,4	7,0	60,6
	4,00	4	2,7	5,6	66,2
	5,00	2	1,4	2,8	69,0
	6,00	1	,7	1,4	70,4
	7,00	21	14,2	29,6	100,0
	Gesamt	71	48,0	100,0	
Fehlend	System	77	52,0		
Gesamt		148	100,0		

### Deskriptive Statistik

		Statistik	Standardfehler	
Diagnoseeffizienz	Mittelwert	,6056	,10705	
	95% Konfidenzintervall des Mittelwerts	Untergrenze	,3921	
		Obergrenze	,8191	
	5% getrimmtes Mittel	,5063		
	Median	,0000		
	Varianz	,814		
	Standardabweichung	,90204		
	Minimum	,00		
	Maximum	3,00		
	Spannweite	3,00		
	Interquartilbereich	1,00		
	Schiefe	1,477	,285	
	Kurtosis	1,321	,563	
	DiagnoseIneffizienz	Mittelwert	3,1549	,33666
95% Konfidenzintervall des Mittelwerts		Untergrenze	2,4835	
		Obergrenze	3,8264	
5% getrimmtes Mittel		3,1166		
Median		2,0000		
Varianz		8,047		
Standardabweichung		2,83674		
Minimum		,00		
Maximum		7,00		
Spannweite		7,00		
Interquartilbereich		6,00		
Schiefe		,372	,285	
Kurtosis		-1,551	,563	

### 10.3.8 Hypothesentestung H1 und H2

#### Zwischensubjektfaktoren

		N
1=sn 2=sh 3=wn 4=wh	1	33
	2	38
	3	26
	4	45

## Tests der Zwischensubjekteffekte

Abhängige Variable: Summe\_Diagnoserichtigkeit

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.	Partielles Eta-Quadrat
Korrigiertes Modell	30,173 <sup>a</sup>	3	10,058	4,270	,006	,085
Konstanter Term	2616,236	1	2616,236	1110,643	,000	,889
Studiengruppe.acc	30,173	3	10,058	4,270	,006	,085
Fehler	325,074	138	2,356			
Gesamt	3177,000	142				
Korrigierte Gesamtvariation	355,246	141				

a. R-Quadrat = ,085 (korrigiertes R-Quadrat = ,065)

## Schätzer

Abhängige Variable: Summe\_Diagnoserichtigkeit

1=sn 2=sh 3=wn 4=wh	Mittelwert	Standardfehler	95%-Konfidenzintervall	
			Untergrenze	Obergrenze
1	3,909	,267	3,381	4,437
2	4,526	,249	4,034	5,019
3	4,038	,301	3,443	4,634
4	5,044	,229	4,592	5,497

## Paarweise Vergleiche

Abhängige Variable: Summe\_Diagnoserichtigkeit

(I)1=sn 2=sh 3=wn 4=wh	(J)1=sn 2=sh 3=wn 4=wh	Mittlere Differenz (I-J)	Standardfehler	Sig. <sup>b</sup>	95% Konfidenzintervall für die Differenz <sup>b</sup>	
					Untergrenze	Obergrenze
1	2	-,617	,365	,093	-1,339	,105
	3	-,129	,402	,748	-,925	,666
	4	-1,135*	,352	,002	-1,831	-,440
2	1	,617	,365	,093	-,105	1,339
	3	,488	,391	,214	-,285	1,260
	4	-,518	,338	,128	-1,187	,150
3	1	,129	,402	,748	-,666	,925
	2	-,488	,391	,214	-1,260	,285
	4	-1,006*	,378	,009	-1,754	-,258
4	1	1,135*	,352	,002	,440	1,831
	2	,518	,338	,128	-,150	1,187
	3	1,006*	,378	,009	,258	1,754

Basiert auf den geschätzten Randmitteln

\*. Die mittlere Differenz ist auf dem ,05-Niveau signifikant.

b. Anpassung für Mehrfachvergleiche: Geringste signifikante Differenz (entspricht keinen Anpassungen).

### Tests auf Univariate

Abhängige Variable: Summe\_Diagnoserichtigkeit

	Quadratsumme	df	Mittel der Quadrate	F	Sig.	Partielles Eta-Quadrat
Kontrast	30,173	3	10,058	4,270	,006	,085
Fehler	325,074	138	2,356			

F prüft den Effekt von 1=sn 2=sh 3=wn 4=wh. Dieser Test basiert auf den linear unabhängigen paarweisen Vergleichen zwischen den geschätzten Randmitteln.

### Zwischensubjektfaktoren

	Wertelabel	N
Median Vorwissen	,00	59
	1,00	83
Case Format	1 Serial Cue	71
	2 Whole Case	71

### Tests der Zwischensubjekteffekte

Abhängige Variable: Summe\_Diagnoserichtigkeit

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.	Partielles Eta-Quadrat
Korrigiertes Modell	30,173 <sup>a</sup>	3	10,058	4,270	,006	,085
Konstanter Term	2616,236	1	2616,236	1110,643	,000	,889
MC_Vorwissen_Median	22,462	1	22,462	9,535	,002	,065
Gruppe	3,574	1	3,574	1,517	,220	,011
MC_Vorwissen_Median * Gruppe	1,288	1	1,288	,547	,461	,004
Fehler	325,074	138	2,356			
Gesamt	3177,000	142				
Korrigierte Gesamtvariation	355,246	141				

a. R-Quadrat = ,085 (korrigiertes R-Quadrat = ,065)

## Parameterschätzer

Abhängige Variable: Summe\_Diagnoserichtigkeit

Parameter	Regressionskoeffizient	Standardfehler	T	Sig.	95%-Konfidenzintervall		Partielles Eta-Quadrat
					Untergrenze	Obergrenze	
Konstanter Term	5,044	,229	22,048	,000	4,592	5,497	,779
[MC_Vorwissen_Median=,00]	-1,006	,378	-2,661	,009	-1,754	-,258	,049
[MC_Vorwissen_Median=1,00]	0 <sup>a</sup>	.	.	.	.	.	.
[Gruppe=1]	-,518	,338	-1,532	,128	-1,187	,150	,017
[Gruppe=2]	0 <sup>a</sup>	.	.	.	.	.	.
[MC_Vorwissen_Median=,00] * [Gruppe=1]	,389	,526	,740	,461	-,651	1,428	,004
[MC_Vorwissen_Median=,00] * [Gruppe=2]	0 <sup>a</sup>	.	.	.	.	.	.
[MC_Vorwissen_Median=1,00] * [Gruppe=1]	0 <sup>a</sup>	.	.	.	.	.	.
[MC_Vorwissen_Median=1,00] * [Gruppe=2]	0 <sup>a</sup>	.	.	.	.	.	.

a. Dieser Parameter wird auf Null gesetzt, weil er redundant ist.

### 1. Gesamtmittelwert

Abhängige Variable: Summe\_Diagnoserichtigkeit

Mittelwert	Standardfehler	95%-Konfidenzintervall	
		Untergrenze	Obergrenze
4,380	,131	4,120	4,639

### Schätzer

Abhängige Variable: Summe\_Diagnoserichtigkeit

Median Vorwissen	Mittelwert	Standardfehler	95%-Konfidenzintervall	
			Untergrenze	Obergrenze
,00	3,974	,201	3,576	4,372
1,00	4,785	,169	4,451	5,120



### Paarweise Vergleiche

Abhängige Variable: Summe\_Diagnoserichtigkeit

(I)Median Vorwissen	(J)Median Vorwissen	Mittlere Differenz (I- J)	Standardfehler	Sig. <sup>b</sup>	95% Konfidenzintervall für die Differenz <sup>b</sup>	
					Untergrenze	Obergrenze
,00	1,00	-,812*	,263	,002	-1,331	-,292
1,00	,00	,812*	,263	,002	,292	1,331

Basiert auf den geschätzten Randmitteln

\*. Die mittlere Differenz ist auf dem ,05-Niveau signifikant.

b. Anpassung für Mehrfachvergleiche: Geringste signifikante Differenz (entspricht keinen Anpassungen).

### Tests auf Univariate

Abhängige Variable: Summe\_Diagnoserichtigkeit

	Quadratsumme	df	Mittel der Quadrate	F	Sig.	Partielles Eta- Quadrat
Kontrast	22,462	1	22,462	9,535	,002	,065
Fehler	325,074	138	2,356			

F prüft den Effekt von Median Vorwissen. Dieser Test basiert auf den linear unabhängigen paarweisen Vergleichen zwischen den geschätzten Randmitteln.

### Schätzer

Abhängige Variable: Summe\_Diagnoserichtigkeit

Case Format	Mittelwert	Standardfehler	95%-Konfidenzintervall	
			Untergrenze	Obergrenze
Serial Cue	4,218	,183	3,857	4,579
Whole Case	4,541	,189	4,168	4,915

### Paarweise Vergleiche

Abhängige Variable: Summe\_Diagnoserichtigkeit

(I)Case Format	(J)Case Format	Mittlere Differenz (I-J)	Standardfehler	Sig. <sup>a</sup>	95% Konfidenzintervall für die Differenz <sup>a</sup>	
					Untergrenze	Obergrenze
Serial Cue	Whole Case	-,324	,263	,220	-,843	,196
Whole Case	Serial Cue	,324	,263	,220	-,196	,843

Basiert auf den geschätzten Randmitteln

a. Anpassung für Mehrfachvergleiche: Geringste signifikante Differenz (entspricht keinen Anpassungen).

## Tests auf Univariate

Abhängige Variable: Summe\_Diagnoserichtigkeit

	Quadratsumme	df	Mittel der Quadrate	F	Sig.	Partielles Eta-Quadrat
Kontrast	3,574	1	3,574	1,517	,220	,011
Fehler	325,074	138	2,356			

F prüft den Effekt von Case Format. Dieser Test basiert auf den linear unabhängigen paarweisen Vergleichen zwischen den geschätzten Randmitteln.

## 4. Median Vorwissen \* Case Format

Abhängige Variable: Summe\_Diagnoserichtigkeit

Median Vorwissen	Case Format	Mittelwert	Standardfehler	95%-Konfidenzintervall	
				Untergrenze	Obergrenze
,00	Serial Cue	3,909	,267	3,381	4,437
	Whole Case	4,038	,301	3,443	4,634
1,00	Serial Cue	4,526	,249	4,034	5,019
	Whole Case	5,044	,229	4,592	5,497

### 10.3.9 Hypothesentestung H3 und H4

#### Zwischensubjektfaktoren

		N
Diagnoseeffizienz; 0=	,00	43
niedrige 1=hohe	1,00	28
DiagnoseIneffizienz 1= hohe	,00	38
0= niedrige	1,00	33

#### Deskriptive Statistiken

Abhängige Variable: Summe\_Diagnoserichtigkeit

Diagnoseeffizienz	DiagnoseIneffizienz	Mittelwert	Standardabweichung	N
,00	,00	3,7500	1,21543	12
	1,00	4,1613	1,39276	31
	Gesamt	4,0465	1,34436	43
1,00	,00	4,6923	1,25759	26
	1,00	2,5000	,70711	2
	Gesamt	4,5357	1,34666	28
Gesamt	,00	4,3947	1,30569	38
	1,00	4,0606	1,41287	33
	Gesamt	4,2394	1,35715	71

### Tests der Zwischensubjekteffekte

Abhängige Variable: Summe\_Diagnoserichtigkeit

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.	Partielles Eta-Quadrat
Korrigiertes Modell	14,448 <sup>a</sup>	3	4,816	2,818	,046	,112
Konstanter Term	348,777	1	348,777	204,120	,000	,753
Mediansplit_diagnoeff	,790	1	,790	,463	,499	,007
Mediansplit_diagnoineff	4,850	1	4,850	2,838	,094	,041
Mediansplit_diagnoeff * Mediansplit_diagnoineff	10,364	1	10,364	6,066	,016	,083
Fehler	114,482	67	1,709			
Gesamt	1405,000	71				
Korrigierte Gesamtvariation	128,930	70				

a. R-Quadrat = ,112 (korrigiertes R-Quadrat = ,072)

### 10.3.10 Hypothesentestung H5 und H6

#### Innersubjektfaktoren

Maß: MEASURE\_1

Faktor1	Abhängige Variable
1	Mean_KF_pre
2	Mean_KF_post

#### Zwischensubjektfaktoren

	Wertelabel	N
Median Vorwissen	,00	59
	1,00	83
Case Format	1 Serial Cue	71
	2 Whole Case	71

**Multivariate Tests<sup>a</sup>**

Effekt	Wert	F	Hypothese		Fehler		Sig.	Partielles	Dezentr. Parameter	Beobachtete Schärfe <sup>c</sup>
			df	df	Eta-Quadrat					
Faktor1	Pillai-Spur	,047	6,790 <sup>b</sup>	1,000	138,000	,010	,047	6,790	,735	
	Wilks-Lambda	,953	6,790 <sup>b</sup>	1,000	138,000	,010	,047	6,790	,735	
	Hotelling-Spur	,049	6,790 <sup>b</sup>	1,000	138,000	,010	,047	6,790	,735	
	Größte charakteristische Wurzel nach Roy	,049	6,790 <sup>b</sup>	1,000	138,000	,010	,047	6,790	,735	
Faktor1 * MC_Vorwissen_Median	Pillai-Spur	,055	8,083 <sup>b</sup>	1,000	138,000	,005	,055	8,083	,806	
	Wilks-Lambda	,945	8,083 <sup>b</sup>	1,000	138,000	,005	,055	8,083	,806	
	Hotelling-Spur	,059	8,083 <sup>b</sup>	1,000	138,000	,005	,055	8,083	,806	
	Größte charakteristische Wurzel nach Roy	,059	8,083 <sup>b</sup>	1,000	138,000	,005	,055	8,083	,806	
Faktor1 * Gruppe	Pillai-Spur	,005	,700 <sup>b</sup>	1,000	138,000	,404	,005	,700	,132	
	Wilks-Lambda	,995	,700 <sup>b</sup>	1,000	138,000	,404	,005	,700	,132	
	Hotelling-Spur	,005	,700 <sup>b</sup>	1,000	138,000	,404	,005	,700	,132	
	Größte charakteristische Wurzel nach Roy	,005	,700 <sup>b</sup>	1,000	138,000	,404	,005	,700	,132	
Faktor1 * MC_Vorwissen_Median * Gruppe	Pillai-Spur	,005	,746 <sup>b</sup>	1,000	138,000	,389	,005	,746	,138	
	Wilks-Lambda	,995	,746 <sup>b</sup>	1,000	138,000	,389	,005	,746	,138	
	Hotelling-Spur	,005	,746 <sup>b</sup>	1,000	138,000	,389	,005	,746	,138	
	Größte charakteristische Wurzel nach Roy	,005	,746 <sup>b</sup>	1,000	138,000	,389	,005	,746	,138	

a. Design: Konstanter Term + MC\_Vorwissen\_Median + Gruppe + MC\_Vorwissen\_Median \* Gruppe

Innersubjektdesign: Faktor1

b. Exakte Statistik

c. Unter Verwendung von Alpha = ,05 berechnet

### Tests der Innersubjekteffekte

Maß: MEASURE\_1

Quelle		Quadratsumme vom Typ III	df	Mittel der Quadrat e	F	Sig.	Partielle s Eta- Quadrat	Dezent. Paramet er	Beobachtet e Schärfe <sup>a</sup>
Faktor1	Sphärizität angenommen	,037	1	,037	6,790	,010	,047	6,790	,735
	Greenhouse-Geisser	,037	1,000	,037	6,790	,010	,047	6,790	,735
	Huynh-Feldt	,037	1,000	,037	6,790	,010	,047	6,790	,735
	Untergrenze	,037	1,000	,037	6,790	,010	,047	6,790	,735
Faktor1 * MC_Vorwissen _Median	Sphärizität angenommen	,044	1	,044	8,083	,005	,055	8,083	,806
	Greenhouse-Geisser	,044	1,000	,044	8,083	,005	,055	8,083	,806
	Huynh-Feldt	,044	1,000	,044	8,083	,005	,055	8,083	,806
	Untergrenze	,044	1,000	,044	8,083	,005	,055	8,083	,806
Faktor1 * Gruppe	Sphärizität angenommen	,004	1	,004	,700	,404	,005	,700	,132
	Greenhouse-Geisser	,004	1,000	,004	,700	,404	,005	,700	,132
	Huynh-Feldt	,004	1,000	,004	,700	,404	,005	,700	,132
	Untergrenze	,004	1,000	,004	,700	,404	,005	,700	,132
Faktor1 * MC_Vorwissen_ Median * Gruppe	Sphärizität angenommen	,004	1	,004	,746	,389	,005	,746	,138
	Greenhouse-Geisser	,004	1,000	,004	,746	,389	,005	,746	,138
	Huynh-Feldt	,004	1,000	,004	,746	,389	,005	,746	,138
	Untergrenze	,004	1,000	,004	,746	,389	,005	,746	,138
Fehler(Faktor1)	Sphärizität angenommen	,749	138	,005					
	Greenhouse-Geisser	,749	138,000	,005					

Huynh-Feldt	,749	138,000	,005					
Untergrenze	,749	138,000	,005					

a. Unter Verwendung von Alpha = ,05 berechnet

### Tests der Innersubjektkontraste

Maß: MEASURE\_1

Quelle	Faktor1	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.	Partielles Eta-Quadrat	Dezentrierter Parameter	Beobachtete Schärfe <sup>a</sup>
Faktor1	Linear	,037	1	,037	6,790	,010	,047	6,790	,735
Faktor1 * MC_Vorwissen_Median	Linear	,044	1	,044	8,083	,005	,055	8,083	,806
Faktor1 * Gruppe	Linear	,004	1	,004	,700	,404	,005	,700	,132
Faktor1 * MC_Vorwissen_Median * Gruppe	Linear	,004	1	,004	,746	,389	,005	,746	,138
Fehler(Faktor1)	Linear	,749	138	,005					

a. Unter Verwendung von Alpha = ,05 berechnet

### Levene-Test auf Gleichheit der Fehlervarianzen<sup>a</sup>

	F	df1	df2	Sig.
Mean_KF_pre	1,539	3	138	,207
Mean_KF_post	2,027	3	138	,113

Prüft die Nullhypothese, daß die Fehlervarianz der abhängigen Variablen über Gruppen hinweg gleich ist.

a. Design: Konstanter Term + MC\_Vorwissen\_Median + Gruppe +  
 MC\_Vorwissen\_Median \* Gruppe  
 Innersubjektdesign: Faktor1

**Parameterschätzer**

Abhängig e Variable	Parameter	Regressionsko effizientB	Standard fehler	T	Si g.	95%- Konfidenzinterv all		Parti elles Eta- Quad rat	Deze ntr. Para meter	Beobac htete Schärfe <sup>b</sup>
						Untergr enze	Obergr enze			
Mean_K F_pre	Konstanter Term	,616	,011	57, 211	,0 0 0	,595	,638	,960	57,21 1	1,000
	[MC_Vorwissen_M edian=,00]	-,010	,018	- ,56 2	,5 7 5	-,045	,025	,002	,562	,086
	[MC_Vorwissen_M edian=1,00]	0 <sup>a</sup>	.	.	.	.	.	.	.	.
	[Gruppe=1]	-,021	,016	- 1,3 34	,1 8 4	-,053	,010	,013	1,334	,263
	[Gruppe=2]	0 <sup>a</sup>	.	.	.	.	.	.	.	.
	[MC_Vorwissen_M edian=,00] *	,011	,025	,46 1	,6 4 5	-,038	,060	,002	,461	,074
	[MC_Vorwissen_M edian=,00] *	0 <sup>a</sup>	.	.	.	.	.	.	.	.
	[MC_Vorwissen_M edian=1,00] *	0 <sup>a</sup>	.	.	.	.	.	.	.	.
	[MC_Vorwissen_M edian=1,00] *	0 <sup>a</sup>	.	.	.	.	.	.	.	.
Mean_K F_post	Konstanter Term	,665	,015	45, 267	,0 0 0	,636	,694	,937	45,26 7	1,000
	[MC_Vorwissen_M edian=,00]	-,076	,024	- 3,1 36	,0 0 2	-,124	-,028	,067	3,136	,876
	[MC_Vorwissen_M edian=1,00]	0 <sup>a</sup>	.	.	.	.	.	.	.	.

[Gruppe=1]	-,022	,022	- ,3 1,0 00	1 1 9	-,065	,021	,007	1,000	,168
[Gruppe=2]	0 <sup>a</sup>	.	.	.	.	.	.	.	.
[MC_Vorwissen_M edian=,00] *	,042	,034	1,2 51	,2 1	-,025	,109	,011	1,251	,237
[Gruppe=1]				3					
[MC_Vorwissen_M edian=,00] *	0 <sup>a</sup>	.	.	.	.	.	.	.	.
[Gruppe=2]									
[MC_Vorwissen_M edian=1,00] *	0 <sup>a</sup>	.	.	.	.	.	.	.	.
[Gruppe=1]									
[MC_Vorwissen_M edian=1,00] *	0 <sup>a</sup>	.	.	.	.	.	.	.	.
[Gruppe=2]									

a. Dieser Parameter wird auf Null gesetzt, weil er redundant ist.

b. Unter Verwendung von Alpha = ,05 berechnet

### Schätzer

Maß: MEASURE\_1

Median Vorwissen	Mittelwert	Standardfehler	95%-Konfidenzintervall	
			Untergrenze	Obergrenze
,00	,600	,009	,582	,618
1,00	,630	,008	,615	,645

### Paarweise Vergleiche

Maß: MEASURE\_1

(I)Median Vorwissen	(J)Median Vorwissen	Mittlere Differenz (I- J)	Standardfehler	Sig. <sup>b</sup>	95% Konfidenzintervall für die Differenz <sup>b</sup>	
					Untergrenze	Obergrenze
,00	1,00	-,030*	,012	,013	-,053	-,006
1,00	,00	,030*	,012	,013	,006	,053

Basiert auf den geschätzten Randmitteln

\*. Die mittlere Differenz ist auf dem ,05-Niveau signifikant.

b. Anpassung für Mehrfachvergleiche: Geringste signifikante Differenz (entspricht keinen Anpassungen).



### Tests auf Univariate

Maß: MEASURE\_1

	Quadratsumme	df	Mittel der Quadrate	F	Sig.	Partielles Eta-Quadrat	Dezentr. Parameter	Beobachtete Schärfe <sup>a</sup>
Kontrast	,030	1	,030	6,308	,013	,044	6,308	,703
Fehler	,656	138	,005					

F prüft den Effekt von Median Vorwissen. Dieser Test basiert auf den linear unabhängigen paarweisen Vergleichen zwischen den geschätzten Randmitteln.

a. Unter Verwendung von Alpha = ,05 berechnet

### Schätzer

Maß: MEASURE\_1

Case Format	Mittelwert	Standardfehler	95%-Konfidenzintervall	
			Untergrenze	Obergrenze
Serial Cue	,611	,008	,595	,627
Whole Case	,619	,008	,602	,636

### Paarweise Vergleiche

Maß: MEASURE\_1

(I)Case Format	(J)Case Format	Mittlere Differenz (I-J)	Standardfehler	Sig. <sup>a</sup>	95% Konfidenzintervall für die Differenz <sup>a</sup>	
					Untergrenze	Obergrenze
Serial Cue	Whole Case	-,008	,012	,495	-,031	,015
Whole Case	Serial Cue	,008	,012	,495	-,015	,031

Basiert auf den geschätzten Randmitteln

a. Anpassung für Mehrfachvergleiche: Geringste signifikante Differenz (entspricht keinen Anpassungen).

### Tests auf Univariate

Maß: MEASURE\_1

	Quadratsumme	df	Mittel der Quadrate	F	Sig.	Partielles Eta-Quadrat	Dezentr. Parameter	Beobachtete Schärfe <sup>a</sup>
Kontrast	,002	1	,002	,467	,495	,003	,467	,104
Fehler	,656	138	,005					

F prüft den Effekt von Case Format. Dieser Test basiert auf den linear unabhängigen paarweisen Vergleichen zwischen den geschätzten Randmitteln.

a. Unter Verwendung von Alpha = ,05 berechnet

### Schätzer

Maß: MEASURE\_1

Faktor1	Mittelwert	Standardfehler	95%-Konfidenzintervall	
			Untergrenze	Obergrenze
1	,604	,006	,591	,616
2	,627	,008	,610	,643

### Paarweise Vergleiche

Maß: MEASURE\_1

(I)Faktor1	(J)Faktor1	Mittlere Differenz (I-J)	Standardfehler	Sig. <sup>b</sup>	95% Konfidenzintervall für die Differenz <sup>b</sup>	
					Untergrenze	Obergrenze
1	2	-,023*	,009	,010	-,041	-,006
2	1	,023*	,009	,010	,006	,041

Basiert auf den geschätzten Randmitteln

\*. Die mittlere Differenz ist auf dem ,05-Niveau signifikant.

b. Anpassung für Mehrfachvergleiche: Geringste signifikante Differenz (entspricht keinen Anpassungen).

### Multivariate Tests

	Wert	F	Hypothese	Fehler	Sig.	Partielles	Dezentr.	Beobachtete
			df	df		Eta-Quadrat	Parameter	Schärfe <sup>b</sup>
Pillai-Spur	,047	6,790 <sup>a</sup>	1,000	138,000	,010	,047	6,790	,735
Wilks-Lambda	,953	6,790 <sup>a</sup>	1,000	138,000	,010	,047	6,790	,735
Hotelling-Spur	,049	6,790 <sup>a</sup>	1,000	138,000	,010	,047	6,790	,735
Größte charakteristische Wurzel nach Roy	,049	6,790 <sup>a</sup>	1,000	138,000	,010	,047	6,790	,735

Jedes F prüft den multivariaten Effekt von Faktor1. Diese Tests basieren auf den linear unabhängigen, paarweisen Vergleichen bei den geschätzten Randmitteln.

a. Exakte Statistik

b. Unter Verwendung von Alpha = ,05 berechnet

### 4. Median Vorwissen \* Case Format

Maß: MEASURE\_1

Median Vorwissen	Case Format	Mittelwert	Standardfehler	95%-Konfidenzintervall	
				Untergrenze	Obergrenze
,00	Serial Cue	,603	,012	,579	,627
	Whole Case	,598	,014	,571	,624
1,00	Serial Cue	,619	,011	,597	,641
	Whole Case	,641	,010	,620	,661

### 5. Median Vorwissen \* Faktor1

Maß: MEASURE\_1

Median Vorwissen	Faktor1	Mittelwert	Standardfehler	95%-Konfidenzintervall	
				Untergrenze	Obergrenze
,00	1	,601	,009	,583	,620
	2	,599	,013	,574	,625
1,00	1	,606	,008	,590	,621
	2	,654	,011	,633	,676

### 6. Case Format \* Faktor1

Maß: MEASURE\_1

Case Format	Faktor1	Mittelwert	Standardfehler	95%-Konfidenzintervall	
				Untergrenze	Obergrenze
Serial Cue	1	,596	,009	,579	,613
	2	,627	,012	,603	,650
Whole Case	1	,611	,009	,594	,629
	2	,627	,012	,603	,651

### 7. Median Vorwissen \* Case Format \* Faktor1

Maß: MEASURE\_1

Median Vorwissen	Case Format	Faktor1	Mittelwert	Standardfehler	95%-Konfidenzintervall	
					Untergrenze	Obergrenze
,00	Serial Cue	1	,597	,013	,572	,621
		2	,610	,017	,576	,643
	Whole Case	1	,606	,014	,578	,634
		2	,589	,019	,551	,627
1,00	Serial Cue	1	,595	,012	,572	,618
		2	,643	,016	,612	,675
	Whole Case	1	,616	,011	,595	,638
		2	,665	,015	,636	,694

# 11 Eidesstaatliche Versicherung

---

Jung, Valentina Manuela

---

Name, Vorname

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Titel „Die Auswirkung des Falldarbietungsformats und des Vorwissens auf diagnostische Kompetenzen bei Medizinstudierenden“ selbstständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

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

Geislingen an der Steige, 24.07.2021

---

Ort, Datum

Valentina Manuela Jung

---

Unterschrift Doktorandin bzw. Doktorand

## 12 Danksagung

---

Ich danke meinem Doktorvater, Prof. Dr. Martin Fischer. Herzlichst danke ich PD Dr. Jan Kieseewetter, der mich mit großem Engagement betreute und mein wissenschaftliches Arbeiten förderte. Mein ganz besonderer Dank gilt meiner Freundin Leah Braun, die mir stets mit einem offenen Ohr und guten Ratschlägen zur Seite stand. Dank gebührt auch Hannah Gerstenkorn und Christian Adolf für ihre Mitarbeit bei der Entwicklung der Onlinefälle. Ich danke Prof. Dr. med. Ralf Schmidmaier für die zur Verfügungsstellung der verwendeten EKGs. Dr. med. Hanna Zimmerman danke ich für die Bereitstellung der verwendeten radiologischen Bildmaterialien. Weiter gilt mein besonderer Dank Andrea Keller-Sauter. Vielen Dank an meine Freunde und Pilotteilnehmer, Timmy Honecker, Romina Jöbstel und Julia Zimmermann. Meiner Familie danke ich für ihre Unterstützung und ihre Liebe. Zuletzt danke ich meinem Partner Wolfgang, der mir mit Liebe und Freundschaft beiseite stand.

# 13 Lebenslauf

---

## VALENTINA MANUELA JUNG

---

### *Persönliche Daten*

**Geburtsdatum** 20.05.1992 in Schwäbisch Gmünd

**Familienstand**

**Nationalität**

**Religion**

**Anschrift**

**Telefonnummer**

**E-Mail-Adresse**

---

### *Berufserfahrung*

**20.01.2020 - xxx** Ärztin in Weiterbildung, Innere Medizin/ Medizinische Klinik,  
Helfenstein Klinik Geislingen- Alb Fils Kliniken Gmbh

**15.03.2017- 04.10.2018** Wissenschaftliche Hilfskraft am Institut für Didaktik und  
Ausbildungsforschung in der Medizin  
Klinikum der Universität München, LMU München,  
Pettenkoferstraße 8a, 80336 München

**07/2011 – 10/2011** Hilfskraft auf 400 € Basis bei Praxis Dres. Jung Deggingen,  
Ditzenbacher Str. 33, 73326 Deggingen

---

### *Studium*

**15.11.2019** Ärztliche Prüfung Gesamtnote 2,16

Mündliches Staatsexamen M3 mit der Note 2,0

**11.10.2018** Schriftliches Staatsexamen M2 mit der Note 2,0

**19.08.2015** Physikum mit der Note 2,5

**01.10.2012 –  
31.03.2020** Studium der Humanmedizin an der Ludwig-Maximilians-  
Universität München

---

### *Praktisches Jahr*

---

<b>01.07.2019-20.10.2019</b>	Chirurgie – Helfenstein Klinik Geislingen - Alb Fils Kliniken, Eybstraße 16, 73312 Geislingen an der Steige
<b>11.03.2019-30.06.2019</b>	Dermatologie - Klinik und Poliklinik für Dermatologie und Allergologie der LMU München, Frauenlobstraße 9 - 11, 80337 München
<b>19.11.2018-10.03.2019</b>	Innere Medizin - Medizinische Klinik und Poliklinik IV, Ziemssenstraße 1, 80337 München

---

### ***Famulaturen***

<b>14.08.2017-12.09.2017</b>	Klinik und Poliklinik für Dermatologie und Allergologie der LMU München, Frauenlobstraße 9 - 11, 80337 München
<b>27.02.2017-28.03.2017</b>	Ambulante Medizin- Praxis Dres. Jung Deggingen, Ditzenbacher Str. 33, 73326 Deggingen
<b>08.08.2016-08.09.2016</b>	Klinik und Poliklinik für Dermatologie und Allergologie der LMU München, Frauenlobstraße 9 - 11, 80337 München
<b>15.02.2016-15.03.2016</b>	Allgemeinmedizin- Praxis Dres. Jung Deggingen, Ditzenbacher Str. 33, 73326 Deggingen

---

### ***Schulbildung***

<b>06.06.2011</b>	Abitur mit der Note 1,8
<b>2002 - 2011</b>	Helfenstein-Gymnasium Geislingen an der Steige
<b>1998 - 2002</b>	Hiltensburg Grundschule Bad Ditzenbach

---

### ***Praktika***

<b>17.02.2014- 24.03.2014</b>	Medizinischer Krankenpflegedienst Helfenstein Klinik Geislingen- Alb Fils Kliniken
<b>05.08.2013- 06.10.2013</b>	Medizinischer Krankenpflegedienst Helfenstein Klinik Geislingen- Alb Fils Kliniken

---

### ***Qualifikationen***

<b>Zertifikate</b>	TEAM-G Zertifikat: German Traumata Evaluation and Management Course
<b>Fremdsprachen</b>	Englisch (erweiterte Grundkenntnisse) Französisch (Grundkenntnisse) Spanisch (Grundkenntnisse) Deutsche Gebärdensprache (Grundkenntnisse)
<b>EDV</b>	MS Office: Word, Powerpoint, Excel (in ständiger Anwendung) IBM SPSS Statistics (gute Kenntnisse) Endnote (gute Kenntnisse)

---

## ***Publikationen/***

### ***Kongressbeiträge***

**BMC Medical Education** Learning clinical reasoning: how virtual patient case format and prior knowledge interact. *Jan Kiesewetter, Michael Sailer, Valentina M. Jung, Regina Schönberger, Elisabeth Bauer, Jan M. Zottmann, Inga Hege, Hanna Zimmermann, Frank Fischer & Martin R. Fischer*  
**Article number 73(2020)**  
**veröffentlicht**  
**14.03.2020:**

**AMEE annual conference Basel 2018** Evaluation of virtual case simulations considering the extraneous cognitive load and emotions of medical students. *Regina Schönberger, Valentina Jung, Michael Sailer, Elisabeth A. Bauer, Frank Fischer, Martin R. Fischer, Jan Kiesewetter*

Vorgetragen am 29.8.2018 von Valentina Jung

**GMA-Jahrestagung Wien 2018** Der Effekt von Vorwissen und Falldarbietungsformat auf die Diagnoserichtigkeit und den prozeduralen Wissenserwerb Medizinstudierender. *Valentina Jung, Regina Schönberger, Michael Sailer, Elisabeth A. Bauer, Frank Fischer, Martin R. Fischer, Jan Kiesewetter*

Vorgetragen am 21.9.2018 von Valentina Jung

---

## ***Referenz***



---

PD Dr. Jan Kieseewetter, Dipl.-Psych.  
(Jan.kieseewetter@med.uni-muenchen.de)

---

### ***Ehrenamtliches***

- |                  |   |
|------------------|---|
| <b>2019</b>      | Aktives Mitglied im Sportsverein TSV Eybach   |
| <b>2009-2011</b> | Referatssprecher Schülermitverwaltung des Helfenstein-Gymnasiums                                      |
|                  | Mitglied im Kammerchor des Helfenstein-Gymnasiums   |
| <b>Seit 1999</b> | Mitglied im Sportverein FTSV Bad Ditzgenbach- Gosbach<br>Abteilung Tennis (derzeit passives Mitglied) |