

# **UNIVERSITÄTSKLINIKUM HAMBURG-EPPENDORF**

Zentrum für Innere Medizin

III. Medizinische Klinik und Poliklinik

Direktor: Prof. Dr. med. Rolf A. K. Stahl

## **Klinische Entscheidungsfindung – Etablierung und Bewertung eines neuen Seminarkonzeptes für PJ-Studierende**

### **Dissertation**

zur Erlangung des Grades eines Doktors der Medizin  
an der Medizinischen Fakultät der Universität Hamburg

vorgelegt von

Ulrike Wendt

aus Oldenburg (Oldb.)

Hamburg 2017

**Angenommen von der Medizinischen Fakultät der Universität Hamburg am: 11.09.2017**

**Veröffentlicht mit Genehmigung der Medizinischen Fakultät der Universität Hamburg.**

**Prüfungsausschuss, der/die Vorsitzende: Prof. Dr. S. Harendza**

**Prüfungsausschuss, zweite/r Gutachter/in: Prof. Dr. J. Kubitz**

# Inhaltsverzeichnis

<b>1 Einleitung</b> .....	<b>5</b>
1.1 Klinische Entscheidungsfindung: Begriff und Verwendung .....	5
1.2 Ärztliches Denken: Diagnostik als Problemlösevorgang .....	6
1.3 Intuition und Analyse: Die duale Prozesstheorie .....	8
1.4 Klinische Entscheidungsfindung: Der Weg zur Arbeitsdiagnose .....	9
1.4.1 Datensammlung, Fallpräsentation und innere Repräsentation .....	9
1.4.2 Die Generierung und Überprüfung von Differentialdiagnosen .....	9
1.4.3 Der Umgang mit Tests .....	10
1.5 Metakognition: Denken verstehen und Fehler erkennen .....	11
1.6 Ärztliches Denken lehren und lernen .....	12
<b>2 Arbeitshypothese und Fragestellung</b> .....	<b>15</b>
<b>3 Methoden</b> .....	<b>16</b>
3.1 Konzept des Seminars „Klinische Entscheidungsfindung“ .....	16
3.2 Durchführung des Seminars .....	17
3.3 Fragebogen zur Selbsteinschätzung .....	18
3.4 Bearbeitung eines Papierfalls .....	18
3.5 Datenerhebung .....	19
3.6 Statistische Auswertung .....	19
3.6.1 Auswertung der Selbsteinschätzungsbögen .....	19
3.6.2 Auswertung der Fallbearbeitung .....	20
3.6.3 Vergleich zwischen Selbsteinschätzung und Fallbearbeitung .....	20
<b>4 Ergebnisse</b> .....	<b>22</b>
4.1 Soziodemographische Daten der Seminarteilnehmenden .....	22
4.2 Fragebogen zur Selbsteinschätzung .....	22
4.2.1 Selbsteinschätzung für die Tertiale gesamt .....	22
4.2.2 Selbsteinschätzung von Männern und Frauen im Vergleich .....	24
4.2.3 Selbsteinschätzung innerhalb der Tertiale .....	26
4.2.4 Selbsteinschätzung der Tertiale im Vergleich: Delta Post minus Pre .....	29
4.2.5 Selbsteinschätzung der Tertiale im Vergleich: Pre, Retropre, Post .....	30
4.3 Bearbeitung des Papierfalls .....	33
4.3.1 Bearbeitung des Papierfalls für die Tertiale gesamt .....	33
4.3.2 Bearbeitung des Papierfalls Männer und Frauen im Vergleich .....	33
4.3.3 Bearbeitung des Papierfalls innerhalb der Tertiale und im Vergleich .....	34
4.4 Berechnung von Korrelationen .....	35

<b>5 Diskussion .....</b>	<b>36</b>
5.1 Fragebogen zur Selbsteinschätzung .....	36
5.1.1 Gesamtheit der Teilnehmenden .....	36
5.1.2 Vergleich Männer und Frauen .....	38
5.1.3 Vergleich der Tertiale .....	39
5.2 Papierfall .....	40
5.3 Korrelationen .....	41
5.4 Metakognition .....	41
5.5 Limitationen der Studie .....	42
<b>6 Zusammenfassung .....</b>	<b>44</b>
<b>7 Abkürzungsverzeichnis .....</b>	<b>46</b>
<b>8 Literaturverzeichnis .....</b>	<b>47</b>
<b>9 Anhang .....</b>	<b>52</b>
9.1 Lerneinheiten und Lernziele für das PJ-Seminar „Klinische Entscheidungsfindung“ .....	52
9.2 Fragebogen zur Selbsteinschätzung .....	54
9.3 Papierfall zur Erhebung des Seminarerfolgs .....	57
9.4 Aufgabenblatt zur Fallbearbeitung .....	58
<b>10 Danksagung .....</b>	<b>59</b>
<b>Eidesstattliche Erklärung .....</b>	<b>60</b>

# 1 Einleitung

Bei jeder Begegnung mit einem Patienten oder einer Patientin, in jeder klinischen Situation müssen Ärzte und Ärztinnen Entscheidungen treffen, die zum Teil wenig spektakulär erscheinen mögen, aber in Einzelfällen schwerwiegende Folgen haben können. Zu diesen Entscheidungen gehört u. a., welche Fragen sie stellen, wie sie Informationen gewichten, welche Untersuchungen sie anordnen; sie entscheiden auch, welche Untersuchungen, beispielsweise aus dem Bereich der Bildgebung, nicht durchgeführt werden, weil sie im entsprechenden Fall möglicherweise als zu belastend und wenig wegweisend angesehen werden. Ärzte und Ärztinnen wissen in der Regel jedoch wenig über die mit den tagtäglichen Entscheidungen verbundenen kognitiven Prozesse – obwohl genau dieses Wissen die Voraussetzung dafür wäre, das eigene ärztliche Tun und Entscheiden kritisch zu hinterfragen und dadurch Fehler ebenso wie Überdiagnostik zu minimieren.

Bei 19 Millionen stationären Behandlungen jährlich in Deutschland geht der AOK-Bundesverband in seinem Bericht 2014 davon aus, dass 19 000 Fehler in Krankenhäusern für den Patienten oder die Patientin tödlich enden. Das sind etwa fünfmal so viele Todesfälle wie im Straßenverkehr (AOK Bundesverband: Krankenhausbericht 2014). Erwartungsgemäß führen die operativen Therapien die Statistik der von Patienten vermuteten und gemeldeten Fehler an, nämlich mit 31 %. Auf die Befunderhebung entfallen 25 % der gemeldeten Fälle, auf die Diagnosestellung 4,6 %, was zusammengenommen fast dem Prozentsatz der chirurgischen Interventionen entspricht – wobei es sich bei Diagnose und Befund um die Grundlagen der ärztlichen Tätigkeit handelt. Darüber hinaus ist davon auszugehen, dass die Zahl der aufgedeckten Fehler nur einen Bruchteil der tatsächlich entstandenen darstellt. Keine Beachtung findet in dieser Statistik die Überdiagnostik, die einerseits dem Einzelnen schaden kann, andererseits die Gemeinschaft finanziell belastet.

Der Prozess der klinischen Entscheidungsfindung ist also nicht nur zentrales Moment im ärztlichen Alltag, sondern als solches folgenreich für das Wohlergehen der Patienten und Patientinnen. Die bewusste Auseinandersetzung damit und die Entwicklung der dazugehörigen Qualifikationen tragen entscheidend zu verantwortlicher, reflektierter Diagnostik und Therapie bei.

## 1.1 Klinische Entscheidungsfindung: Begriff und Verwendung

In englischsprachigen Fachpublikationen werden ärztliche Denkprozesse unter Clinical Reasoning gefasst. Dieser Begriff erschließt sich mühelos und ist dennoch nicht einfach ins

Deutsche übertragbar. Dem wörtlichen Verständnis nach wäre „klinisches Schlussfolgern“ eine treffende Übersetzung. Doch Clinical Reasoning wird weitgreifender verwendet als nur auf das rein logische Schließen bezogen. Es umfasst alle kognitiven Vorgänge, die das Denken im klinischen Zusammenhang auf die Probleme eines Patienten bezogen ausmachen. Daher setzen sich auch andere Gesundheitsberufe, etwa Pflege oder Physiotherapie, mit Clinical Reasoning auseinander (Higgs et al. 2008; Klemme et al. 2015). Für ärztliche Denkprozesse werden die Begriffe Medical Problem Solving und Medical Decision Making als die beiden Aspekte des Clinical Reasoning verwendet (Elstein und Schwartz 2002; Kiesewetter et al. 2013) oder Diagnostic Reasoning und Therapeutic Reasoning unter Clinical Reasoning subsumiert (Durning et al. 2011). Clinical Reasoning, Diagnostic Reasoning, Medical Problem Solving und Clinical Decision Making werden an anderer Stelle auch gleichbedeutend verwendet (Stieger et al. 2009).

In seiner grundlegenden Arbeit „Learning Clinical Reasoning“ stuft Jerome Kassirer den Prozess des Clinical Reasoning als die wesentliche Aufgabe von Ärzten und Ärztinnen ein:

„Indeed, the prime function of the physician is clinical reasoning: to suspect the cause of a patient's symptoms and signs, to gather additional relevant information, to select necessary tests, and to recommend therapy.“ (Kassirer et al. 2010, S. 4).

Klinische Entscheidungsfindung bedeutet jedoch mehr als die genannten Schritte der Diagnostik. Auf der Meta-Ebene umfasst sie das Wissen um die entsprechenden kognitiven Prozesse und ihre Strukturen. Erst diese umfassende Kenntnis befähigt zur nachvollziehbaren Begründung und zur kritischen Beurteilung des eigenen Handelns sowie der damit verbundenen Entscheidungen. Clinical Reasoning ist also auch eine wesentliche Voraussetzung ärztlicher Reflexionskompetenz (Goldszmidt et al. 2013).

## **1.2 Ärztliches Denken: Diagnostik als Problemlösevorgang**

Unabhängig von den Inhalten scheinen die kognitiven Schritte, die beim Denken als einem Lösen von Problemen vollzogen werden, im Prinzip die gleichen zu sein. So benannte John Dewey schon 1910 in seiner Schrift „How we think“ „fünf logisch verschiedene Stufen“, nachfolgend zitiert aus der neuesten deutschsprachigen Übersetzung (Dewey 2002, S. 56):

„(i) Man begegnet einer Schwierigkeit, (ii) sie wird lokalisiert und präzisiert, (iii) Ansatz einer möglichen Lösung, (iv) logische Entwicklung der Konsequenzen des Ansatzes, (v) weitere Beobachtung und experimentelles Vorgehen führen zur Annahme oder

Ablehnung, das heisst der Denkprozess findet seinen Abschluss, indem man sich für oder wider die bedingt angenommene Lösung entscheidet.“

Setzt man für „Lösung“ „diagnostische Hypothese“ oder „Differentialdiagnose“ ein, dann liest sich die Abfolge wie die Beschreibung des diagnostischen Prozesses (Kieseewetter et al. 2013):

- Informationen sammeln
- Informationen analysieren und Differentialdiagnosen formulieren
- mehr Informationen sammeln
- eine innere Repräsentation des Falles gewinnen
- die Differentialdiagnosen überprüfen
- Entscheidungen über weitere Untersuchungen treffen
- Entscheidung für eine Differentialdiagnose als Arbeitshypothese.

Beim Denken im Sinne der Problemlösung greifen Deduktion und Induktion ineinander: Der deduktive Vorgang, der sich vom Allgemeinen zum Besonderen, von der allgemeinen Regel zum Schluss auf den Einzelfall bewegt, und der induktive Vorgang, der aus den besonderen Gegebenheiten des Einzelfalles zur Regel als dem Allgemeinen vorstößt, sind nicht voneinander zu trennen (Dewey 1910/2002). Als eine Verbindung von Deduktion und Induktion wird außerdem die Abduktion beschrieben, worunter man die Bildung von Hypothesen oder Erklärungen für vorhandene Informationen versteht (Patel et al. 2005).

Verschiedene Problemlösestrategien unterscheiden sich in der Gewichtung ihrer Denkschritte und in der Art, die beim logischen Schließen führend ist, z. B. Raten, hypothetisch-deduktives Problemlösen, schematisch-induktives Problemlösen oder Mustererkennung (Harasym et al. 2008). Die Anwendung der jeweiligen Strategien ist davon abhängig, in welcher Form das vorhandene Wissen gespeichert ist: „reduced“, „dispersed“, „elaborated causal“, als „scheme“ oder als „scripts“ (Harasym et al. 2008). Diese fünf Klassen unterscheiden sich durch ansteigende Strukturiertheit und stellen insofern Stufen dar. Das Wissen ist von Stufe zu Stufe stärker vernetzt und zunehmend in sinnhaften Einheiten abrufbar (Bordage et al. 1990; Bordage 1994; Harasym et al. 2008). Unzusammenhängendes (dispersed) und ausgearbeitetes kausales (elaborated causal) Wissen ermöglichen ein hypothetisch-deduktives Vorgehen, Krankheitsskripte (scripts) dagegen stellen die Grundlage für Mustererkennung dar. Ihre Anwendung wird mit entsprechender Expertenschaft verbunden (Bordage und Lemieux 1991; Bordage 1994). Studierende oder wenig erfahrene

Ärzte und Ärztinnen gehen überwiegend hypothetisch-deduktiv vor, ebenso erfahrene Kliniker und Klinikerinnen, wenn es sich um komplexe, untypische Fälle handelt. In der Regel ist jedoch die Mustererkennung die Strategie des erfahrenen Arztes bzw. der erfahrenen Ärztin (Elstein und Schwartz 2002). Zum Repertoire der letzteren gehören auch Heuristiken im Sinne von Faustregeln und Schnellverfahren (Kassirer et al. 2010), während Medizin-studierende und junge Ärzte und Ärztinnen leicht heuristischen Irrtümern aufsitzen können (Rylander und Guerrasio 2016).

### **1.3 Intuition und Analyse: Die duale Prozesstheorie**

Als Kognitionspsychologe untersuchte Kahneman gemeinsam mit seinem Kollegen Tversky in den 70er Jahren die Rationalität von Entscheidungen. Er arbeitete mit Beispielen aus dem Alltag und der Wirtschaft, aber auch aus der Medizin und er beschrieb zwei unterschiedliche Arten der Entscheidungsfindung: eine intuitive und eine analytische (Kahneman 2012; Tversky und Kahneman 1974).

Die Auseinandersetzung damit, wie der Arzt bzw. die Ärztin zu einer Arbeitsdiagnose findet, führte dazu, kognitive Vorgänge, die dem diagnostischen Prozess unterliegen, ebenfalls in einen intuitiven, unbewussten und einen analytischen, bewussten Typ zu differenzieren (Croskerry 2009; Eva 2005; Kassirer et al. 2010). Bereits 1978 wurde die hypothetisch-deduktive Problemlösestrategie als Vorgang der Bildung diagnostischer Hypothesen und der Überprüfung derselben mittels entsprechender Datensammlung als analytisches Vorgehen beschrieben (Elstein et al. 1978). Das intuitive Vorgehen dagegen beruht im Wesentlichen auf dem Erkennen von Mustern, die sich im Laufe mannigfacher Begegnungen mit Patienten gebildet haben und die spontan ins Bewusstsein treten (Elstein und Schwartz 2002).

Dieses letztgenannte intuitive System 1 scheint phylogenetisch mit älteren Anteilen des Gehirns verbunden zu sein. Im Rückbezug auf frühere evolutionäre Entwicklungsstufen wird es angeregt und unterstützt durch Kontext, Bilder, Gefühle. Allerdings wird dieser intuitiven Reaktion bei höherer Schnelligkeit auch größere Fehleranfälligkeit zugeschrieben (Croskerry 2009). Das analytische System 2 dagegen wird als der schulungsfähige Teil beschrieben. Die Systeme 1 und 2 werden nicht dichotomisch gedacht, sondern interagierend. Folglich wird auch der Prozess der Diagnostik als Ergebnis der Zusammenarbeit und als Kontinuum beider Systeme verstanden (Croskerry 2009).



## **1.4 Klinische Entscheidungsfindung: Der Weg zur Arbeitsdiagnose**

Der zuweilen detektivisch anmutende Problemlösevorgang, den der diagnostische Prozess darstellt, umfasst verschiedene Elemente: die Datensammlung, zu der die Beobachtungen durch die Ärztin bzw. den Arzt ebenso gehören wie die Anamnese und die körperliche Untersuchung, die Bildung von Hypothesen zu den möglichen Ursachen der Beschwerden, die Formulierung der sich daraus ergebenden Differentialdiagnosen, die Überprüfung der Hypothesen, z. B. durch Tests, und schließlich die Entscheidung für eine Arbeitsdiagnose (Eva 2005; Kassirer et al. 2010). Alle Theorien des Clinical Reasoning beschreiben die analytische Entscheidungsfindung als einen zyklischen Prozess, bei dem zum Lösen von Problemen Hypothesen oder Differentialdiagnosen formuliert und im Weiteren überprüft werden (Patel et al. 2005).

### **1.4.1 Datensammlung, Fallpräsentation und innere Repräsentation**

Die diagnostische Datensammlung beginnt mit der Begegnung mit dem Patienten bzw. der Patientin. Sie umfasst neben dem subjektiven Eindruck des Arztes bzw. der Ärztin die klassische Anamnese und kann sich weiterhin speisen aus vielfältigen Quellen wie Fremdanamnese, Fremdbefunden, Übergaben oder Briefen. Die körperliche Untersuchung gehört ebenso dazu wie Laborwerte, Testergebnisse und weitere Untersuchungen. Informationen werden in einem fortlaufenden Prozess gesammelt, die Reihenfolge ist nicht festgelegt (Kassirer et al. 2010).

Die Fallpräsentation dient im klinischen Alltag der Vorstellung des Patienten bzw. der Patientin beispielsweise in der Visite oder bei der Übergabe. Hierbei ist es wichtig, die zum klinischen Denken wesentlichen Informationen prägnant und kurz zusammenzufassen, da das menschliche Kurzzeitgedächtnis nur fünf bis neun Elemente gleichzeitig speichern kann (Miller 1956). Die Patientenvorstellung dient außerdem dazu, sich selbst ein Bild von dem Fall zu machen und dieses auch in den Zuhörenden entstehen zu lassen. In Form einer solchen inneren Repräsentation ein Bild von dem Patienten oder der Patientin zu entwickeln, ist ein wichtiges Moment des ärztlichen Problemlösevorgangs (Kiesewetter et al. 2013).

### **1.4.2 Die Generierung und Überprüfung von Differentialdiagnosen**

Einen wesentlichen Bestandteil des klinischen Denkens stellt die Bildung von Hypothesen dar, also das Stellen von Differentialdiagnosen, die das Befinden eines Patienten bzw. einer Patientin, die Symptome und/oder die erhobenen Laborwerte erklären könnten. Im Englischen werden sie als Diagnostic Hypothesis bezeichnet. Die Hypothesen, die abhängig von den zugänglichen Informationen und vom Untersucher bzw. der Untersucherin als erste

gebildet werden, schaffen für das weitere Vorgehen den Rahmen. Um sie zu überprüfen, werden neue Fragen gestellt und/oder Untersuchungen angefordert. In einem fortlaufenden Prozess werden so Differentialdiagnosen überprüft, verworfen, neue generiert oder verworfene wieder aufgenommen (Kassirer et al. 2010). Der fortgesetzte, auch in Form einer Spirale beschriebene Abgleich der Differentialdiagnosen mit ermittelten Informationen führt schließlich über konkurrierende Verdachtsdiagnosen zu einer Arbeitshypothese, die als ausreichend wahrscheinlich angesehen wird, um weitere, auch invasivere Tests oder Therapieveruche zu unternehmen (Kassirer 2010). Differentialdiagnosen können unwahrscheinlich gemacht werden, die verbleibende Arbeitsdiagnose sollte möglichst wahrscheinlich gemacht werden. Bis dahin sind in den meisten Fällen schon eine Reihe Tests durchgeführt worden.

### **1.4.3 Der Umgang mit Tests**

Tests stellen ein wichtiges Mittel auf dem Weg der Erschließung einer Arbeitsdiagnose dar. Der Umgang mit Tests will allerdings gelernt sein, auch um belastende Überdiagnostik zu vermeiden, die im ärztlichen Alltag ein häufiges Problem darstellt (Hudzik et al. 2014). So forderten beispielsweise deutsche Medizinstudierende in einem Forschungsprojekt, das einen ersten Arbeitstag im Krankenhaus simulierte, signifikant mehr Laboruntersuchungen und Röntgendiagnostik an als niederländische Studierende, ohne dadurch zu besseren diagnostischen Ergebnissen zu kommen (Harendza et al. 2013).

Mit Hilfe eines Tests wird keine Diagnose gestellt, sondern die Wahrscheinlichkeit für die Richtigkeit einer Diagnose beschrieben. Dabei kommt der Pretest-Wahrscheinlichkeit, die unter anderem abhängig ist von der Prävalenz einer Erkrankung in einer bestimmten Bevölkerungsgruppe, eine besondere Bedeutung zu, da sich je nach Testergebnis eine unterschiedliche Posttest-Wahrscheinlichkeit ergibt, die leicht falsch eingeschätzt werden kann (Lyman und Balducci 1994). Tests werden also nach Einschätzung der Pretest-Wahrscheinlichkeit entweder eingesetzt, um eine Hypothese wahrscheinlicher oder um sie unwahrscheinlicher zu machen. Wenn Sensitivität und Spezifität eines Tests sowie die Prävalenz einer Erkrankung bekannt sind, lassen sich die Regeln von Bayes anwenden, die dabei helfen können, Tests zu interpretieren bzw. vor der Durchführung eines Tests abzuschätzen, ob die Anwendung sinnvoll ist (Sox et al. 2013). Dass Testergebnisse auch falsch positiv oder falsch negativ sein können, muss bei der diagnostischen Betrachtung ebenfalls berücksichtigt werden.

## 1.5 Metakognition: Denken verstehen und Fehler erkennen

Denken zu verstehen ist die Voraussetzung dafür, Fehler im kognitiven Prozess zu erkennen. Auf sie gehen falsche Diagnosen zurück, sofern nicht mangelndes Wissen oder sogenannte No-Fault-Fehler, also kaum zu vermeidende Fehler, die Ursache sind (Croskerry 2014). Studien im englischsprachigen Raum zufolge sind immerhin 10 % bis 15 % der Diagnosen, die gestellt werden, falsch (Cooper und Frain 2017). Wesentlichen Anteil daran haben typische Denkfehler oder Cognitive Biases, denen man leicht aufsitzen kann, wenn das Denken unreflektiert bleibt (Mamede et al. 2007). Vergleichbar optischen Täuschungen handelt es sich bei diesen Denkfehlern oder -verzerrungen um unbewusste, systematische Fehler, die dem intuitiven System 1 entspringen. Als „cognitive dispositions to respond“ (Croskerry 2003) entsprechen sie Automatismen, die schnell und oft hilfreich sind, aber nicht immer zur richtigen Lösung des Problems führen. Zu diesen Fehlern kann es in allen Phasen des diagnostischen Prozesses kommen (Kassirer et al. 2010). Einige Denkverzerrungen, die im medizinischen Bereich häufiger auftreten, sind Folgende (Croskerry 2003):

**Tabelle 1: Beispiele für Cognitive Biases oder Denkfehler**

<b>Anchoring</b> (Verankerung; Verhaftung)	Der ersten Information oder dem ersten Eindruck zu stark verhaftet bleiben, so dass neue Informationen in ihrer Bedeutung nicht zum Tragen kommen.
<b>Availability Bias</b> (Verfügbarkeit; Abrufbarkeit)	Eine Diagnose für wahrscheinlicher halten, weil ein entsprechendes Krankheitsbild sofort abrufbar ist, z. B. nachdem der Arzt bzw. die Ärztin kurz zuvor mit entsprechenden Fällen zu tun hatte oder sich in letzter Zeit besonders mit dieser Art von Fällen beschäftigt hat.
<b>Base Rate Neglect</b> (Prävalenzfehler)	Die Tendenz, die Prävalenz einer Krankheit nicht zu berücksichtigen oder die Pretest-Wahrscheinlichkeit falsch einzuschätzen, beispielsweise zu niedrig anzusetzen. Eine mögliche unbewusste Ursache hierfür könnte der Wunsch sein, eine bedrohliche Diagnose bei einem sympathischen Patienten oder einer sympathischen Patientin nicht weiter verfolgen zu müssen.
<b>Confirmation Bias</b> (Bestätigungsfehler)	Die Tendenz, eher nach Informationen zu suchen, die eine Differentialdiagnose bestätigen, als nach solchen, die sie widerlegen, z. B. wenn durch Kollegen oder Kolleginnen in einer anderen Klinik oder ambulant schon einmal eine entsprechende Verdachtsdiagnose formuliert wurde.
<b>Framing Effect</b> (Rahmungseffekt)	Die Beurteilung eines Falles wird davon beeinflusst, wie der Patient oder die Patientin beschrieben oder beispielsweise in der Übergabe vorgestellt wird. Die Art und Weise, wie der Rahmen gesetzt wird, beeinflusst die Wahrnehmung und die Erwartung.

<b>Overconfidence Bias</b> (Vermessenheitsverzerrung)	Überschätzen des eigenen Wissens und der eigenen Fähigkeiten; Selbstüberschätzung, die dazu führt, dass der eigenen Meinung mehr vertraut wird als den Fakten bzw. der Evidenz im Sinne der evidenzbasierten Medizin.
<b>Premature Closure</b> (Vorschneller Schluss)	Der Prozess der Suche nach einer Arbeitsdiagnose wird zu früh beendet, indem eine Differentialdiagnose als Arbeitsdiagnose angenommen wird, bevor diese oder andere diagnostische Hypothesen ausreichend untersucht und geprüft sind. Dies wird als der häufigste Fehler beschrieben (Graber 2005).
<b>Representativeness Restraint</b> (Repräsentativitätsverzerrung)	Tendenz des Arztes bzw. der Ärztin, Krankheitsbilder nur in ihrer typischen Darstellung anzunehmen, untypische Erscheinungsbilder von Erkrankungen zu übersehen.

Das analytische System 2 bietet eine Möglichkeit der Kontrolle des Systems 1 – wenn die Problematik der Denkverzerrungen bewusst ist. Dieses gezielte Umschalten in den analytischen Modus stellt den zentralen Aspekt des Debiasing dar. Dazu gehören außerdem eine Reihe von Strategien wie die Strukturierung von Anamnese und körperlicher Untersuchung, Checklisten, Mnemotechniken oder der Rückgriff auf Daten der evidenzbasierten Medizin (Croskerry 2017). Im Gegensatz zu ihren unerfahrenen Kolleginnen und Kollegen scheinen klinische Experten, deren intuitives Vorgehen und Entscheiden auf jahrelanger Erfahrung beruht, im analytischen Modus allerdings eher mehr Fehler zu machen, als wenn sie sich im intuitiven System bewegen (Norman und Eva 2010). Grundsätzlich gilt: Das reine Erklären von Denkfehlern reicht als Maßnahme nicht, damit diese vermieden werden. Eine Kombination des Unterrichtens von intuitivem und analytischem Weg des Clinical Reasoning dagegen scheint zu einer Reduktion von Denkfehlern zu führen (Norman und Eva 2010).

## 1.6 Ärztliches Denken lehren und lernen

Klinische Entscheidungsfindung wird bislang im Rahmen der Ausbildung von Medizinern und Medizinerinnen in Deutschland zumindest explizit kaum unterrichtet, soweit sich dies ohne systematische Erhebung beurteilen lässt. Sie ist auch nicht Bestandteil der ärztlichen Approbationsordnung. Im englischen Sprachraum dagegen werden schon seit Jahrzehnten unterschiedliche Ansätze zum Lehren und Lernen von Clinical Reasoning erprobt und umgesetzt. Eine Vielzahl von Unterrichtsformen wurde eingesetzt und beschrieben, z. B. problembasierte klinische Seminare, Diagnostic Grand-Rounds, integriertes fallbezogenes Lernen, Simulation und Nachbesprechung klinischer Einsätze oder die Arbeit mit virtuellen Patienten (Cooper et al. 2017). Es herrscht Übereinkunft, dass fallbasiertes Lernen die beste Form der Aneignung von Fähigkeiten der klinischen Entscheidungsfindung darstellt (Kassirer 2010).

Hier lassen sich auch Bezüge zu anderen Disziplinen herstellen, so zur allgemeinen Didaktik (Kaiser 1983) und zur Didaktik der Pflegeberufe (Barre 2014), in der die hermeneutische Fallkompetenz ins Zentrum der Lehre gerückt ist. Als besonders effektiv erwies sich in der Medizin die Verbindung der Arbeit an Fallbeispielen mit Key-Feature-Tests, also Prüfungen, deren Fragen und Aufgaben ebenfalls auf Fälle bezogen sind und bei denen die Antworten aus einem Menü elektronisch ausgewählt werden (Nikendei et al. 2009; Raupach et al 2016).

Schmidt und Mamede (2015) untersuchten die bis zum Juni 2014 zu Clinical-Reasoning-Unterricht erschienenen Studien und werteten 39 von 48 Studien aus. Sie differenzierten dabei zwischen zwei methodischen und zwei inhaltlichen Aspekten: Methodisch wurde unterschieden, ob Fälle im Ganzen präsentiert oder ob Informationen zum Fall schrittweise gegeben wurden. Das zuletzt genannte Vorgehen ist der realen Arzt-Patient-Begegnung nachempfunden und scheint auch deshalb die verbreitetste Form der Vermittlung zu sein. Die Idee, die hinter dieser Herangehensweise steht, leuchtet unmittelbar ein, erwies sich jedoch im Hinblick auf diagnostische Genauigkeit gegenüber der Arbeit mit vollständigen Fallvignetten von Anfang an als deutlich unterlegen (Nendaz et al. 2000). Eine weitere Kategorisierung wurde hinsichtlich der jeweiligen Seminarinhalte getroffen, nämlich zwischen Prozess- und Wissensvermittlung: Einerseits wurden im Unterricht Aspekte der Metakognition bearbeitet (way of thinking, process-oriented). In diesem prozessorientierten Ansatz wurden mit Stufen im Problemlöseprozess, Bayes' Theorem, Entscheidungsanalyse und Algorithmen Aspekte der klassischen klinischen Entscheidungsfindung bearbeitet (Rogers et al. 1991) oder ein die kognitiven Prozesse betreffendes systematisches Herangehen bei der Diagnostik vermittelt, also das hypothetisch-deduktive Vorgehen bei der Diagnosestellung genutzt (Beullens et al. 2006). Die Auseinandersetzung mit dem Prozess des Clinical Reasoning erhöhte die diagnostischen Fähigkeiten im ersten Fall, der Beschäftigung mit klassischen Aspekten des Clinical Reasoning, nicht, im zweiten, der Beschäftigung mit der systematischen Herangehensweise, nicht wesentlich. Dies wird als wenig überraschend dargestellt, da dieser Prozess von Fachwissen nicht zu isolieren sei, wenn es darum gehe, spezielle Erkrankungen zu verstehen und zu diagnostizieren (Glaser 1983). Zwei Studien befassten sich nicht mit der Vermittlung von Denkprozessen, sondern untersuchten die Bedeutung der jeweils angewandten Form des Denkens, also Mustererkennung oder analytisches Vorgehen, für die Diagnostik bzw. in den vorliegenden Fällen für die Beurteilung von EKGs. Hier zeigte sich, dass die Verbindung der beiden Formen kognitiver Prozesse zu den besten Ergebnissen führte (Ark et al. 2006). Andererseits stand in einer Gruppe von Studien inhaltlich die Vermittlung bzw. Anwendung von Fach- und Faktenwissen im Zentrum (knowledge-oriented). Diese wurde noch einmal dahingehend unterschieden, ob das Verständnis der pathophysiologischen Mechanismen einer Erkrän-

kung (Chamberland et al. 2011) oder die Unterscheidung zwischen Erkrankungen mit ähnlichen Erscheinungsbildern bearbeitet wurden (Mamede et al. 2012).

Schmidt und Mamede beziehen die Ergebnisse ihrer Metastudie auf Aspekte der Zwei-Prozess-Theorie bzw. auf die beschriebenen Stufen der zunehmend komplexen Speicherung von Inhalten im Gedächtnis. Sie favorisieren die Anwendung unterschiedlicher Methoden entsprechend dem Wissens- und Ausbildungsstand der angehenden Ärzte und Ärztinnen (Schmidt und Mamede 2015, S. 970):

„ .. we advocate a theory-based approach to the teaching of CR based on our understanding of how the knowledge of disease evolves in the mind of the medical student.“

Die Stufen des Lernens könnten in Anlehnung an Schmidt und Mamede (2015) beschrieben werden wie folgt:

**Tabelle 2: Stufen des Lernens**

Studienanfänger (vom unzusammenhängenden zum ausgearbeiteten, kausalen Wissen)	Sich die Anzeichen und Symptome eines Falles mittels der zugrunde liegenden pathophysiologischen Mechanismen selbst erklären, um zu einer Arbeitsdiagnose zu gelangen („self-explanation“)
Fortgeschrittene (vom kausalen zum „eingekapselten“/schematischen Wissen)	Auf der Basis von „eingekapseltem“, also in Form von Versatzstücken (Chunks) gespeichertem und abrufbarem Wissen unterschiedliche Krankheitsbilder vergleichen und kontrastieren, und zwar in Form einer abwägenden Reflexion („deliberate reflection“)
Weit Fortgeschrittene (Muster und Krankheitskripte)	Auf der Basis von Krankheitskripten die schrittweise Falllösung erproben

Ob mit realen oder virtuellen Patienten, mit Simulationspatienten oder Papier-Fällen gearbeitet wird, ob in Kleingruppen oder Großgruppen und mit oder ohne Rollenspiel, unangefochten ist, dass Studierende möglichst vielen Fällen begegnen sollten, nicht nur um analytisches Vorgehen einzuüben, sondern auch, um möglichst früh die Grundlagen für die Musterbildung zu legen und schließlich tatsächlich analytische und intuitive Entscheidungsfindung miteinander verbinden zu können (Eva 2005). Basierend auf diesen Erkenntnissen wurde im Rahmen des Pilotprojekts, das der vorliegenden Arbeit zugrunde liegt, ein Unterrichtskonzept für Clinical Reasoning für Studierende im Praktischen Jahr (PJ) an der Medizinischen Fakultät der Universität Hamburg erarbeitet und dessen Wirksamkeit untersucht.

## 2 Arbeitshypothese und Fragestellung

Ziele des Projekts waren die Etablierung eines Seminars zur klinischen Entscheidungsfindung im Praktischen Jahr (PJ) und eine lernzielbasierte Überprüfung des Erfolgs mittels Fragebogen zur Selbsteinschätzung und Bearbeitung eines Papierfalls. Die der Untersuchung zugrunde liegende Hypothese lautete: Medizinstudierende verbessern ihre Qualifikationen im Bereich der klinischen Entscheidungsfindung durch Intervention in Form eines achtwöchigen Seminars, das auf Lernzielen zum Thema klinische Entscheidungsfindung basiert. Bezogen auf die Bearbeitung eines Papierfalls sind sie in der Fallpräsentation fokussierter und erarbeiten mehr Differentialdiagnosen. Mit der vorliegenden Arbeit sollen daher folgende Fragestellungen untersucht werden:

1. Zeigte sich bei der Selbsteinschätzung von Qualifikationen hinsichtlich der acht bearbeiteten Aspekte der klinischen Entscheidungsfindung ein signifikanter Unterschied zwischen Pre- und Post-, zwischen Retropre- und Post- oder zwischen Pre- und Retropre-Befragung?
2. War die Problempräsentation des Papierfalls gemessen an der Zahl der im Verhältnis zur Nennung von Schlüsselbegriffen verwendeten Wörter nach dem Kurs fokussierter?
3. Zeigte sich nach dem Seminar ein Zuwachs bei der Nennung von Differentialdiagnosen im Hinblick auf den bearbeiteten Papierfall?
4. Zeigten sich Unterschiede in den unter 1. bis 3. genannten Aspekten abhängig davon, ob es sich um Studentinnen oder um Studenten handelte?
5. Zeigten sich Unterschiede in den unter 1. bis 3. genannten Aspekten abhängig davon, ob die Studierenden sich im ersten, zweiten oder dritten Tertial des PJs befanden?
6. Korrelierten die Selbsteinschätzung der Studierenden hinsichtlich der Problempräsentation (Item 1) und das Ergebnis der Fallbearbeitung (Wörterzahl und/oder Quotient) bzw. der Selbsteinschätzung hinsichtlich des Benennens von Differentialdiagnosen (Item 2) und das Ergebnis der Fallbearbeitung (Anzahl Differentialdiagnosen) vor und/oder nach dem Seminar?

## 3 Methoden

### 3.1 Konzept des Seminars „Klinische Entscheidungsfindung“

Das Konzept des Seminars wurde gemeinsam mit zwei in der Klinik und zwei in der Praxis tätigen Ärzten bzw. einer Ärztin entwickelt: Prof. Dr. Sigrid Harendza und Dr. Matthias Janneck, beide oberärztlich in der III. Medizinischen Klinik des Universitätsklinikums Hamburg-Eppendorf (UKE) tätig, Dr. Andreas Klinge und Dr. Ingo Krenz, beide ambulant in internistischen Praxen in Hamburg tätig. Im gemeinsamen Diskussionsprozess wurden, basierend auf den Schritten der klinischen Entscheidungsfindung (Kassirer et al. 2010), sechs Lerneinheiten definiert:

1. Datenerhebung, akkurate/abstrahierende Problempräsentation
2. Verdachtsfundierte Begründungen/Hypothesenbildung
3. Diagnosen auf dem Prüfstand/Überprüfung der Hypothesen und Testtheorie
4. Pattern Recognition (Mustererkennung)
5. Ärztliche Denkfehler
6. Umgang mit Ungewissheit

Bezogen auf diese sechs Lerneinheiten wurden insgesamt 18 Lernziele formuliert (siehe Anhang 9.1). Basierend darauf wurden die Lerninhalte der einzelnen Seminartermine festgelegt. Die Lerneinheiten (1), (2) und (3) wurden überlappend an insgesamt fünf Terminen Unterrichtsgegenstand, die Lerneinheiten (4), (5) und (6) wurden jeweils an einem Termin explizites Seminarthema.

Didaktisch-methodisch beruhte das Seminarkonzept auf dem Bearbeiten von Papierfällen, die von den vier Dozierenden aus ihrer eigenen ärztlichen Tätigkeit zusammengetragen wurden und jeweils für mindestens eine der Lerneinheiten geeignet schienen. Zu jedem dieser Fälle wurde eine Studierendenversion und eine Dozierendenversion, die zusätzliche Angaben enthielt, erstellt. Aus insgesamt 32 erstellten Fällen wurden im Seminarverlauf jeweils 17 unterschiedliche Papierfälle mit Passung zu den Lerneinheiten verwendet. Dieses fallbezogene Lernen ermöglichte es den Studierenden, den Lernprozess im Seminar vergleichbar der Auseinandersetzung zu gestalten, die angehende Ärzte und Ärztinnen alltäglich auf Station führen müssen (Kassirer 2010). Die Vermittlung der Qualifikationen fand also auf den Kontext bezogen statt, in dem sie auch zur alltäglichen Praxis werden sollten.



Die Relevanz und Aktualität der Auseinandersetzung wurde darüber hinaus vermittelt, indem Dozierende in den jeweiligen Seminarsitzungen mündlich Fälle präsentierten, mit denen sie tagesaktuell auf Station oder in ihrer Praxis zu tun hatten. Dadurch vermittelte sich manche Diskussion beispielsweise zu Differentialdiagnosen eher vergleichbar einer Diskussion unter Kollegen und Kolleginnen als wie ein Unterrichtsgespräch. Auch machten sich die Dozierenden in einzelnen Seminarsituationen wie bei einer Vorführung gegenseitig selbst zu Fragenden, die Hypothesen generierten und die Relevanz von Tests diskutierten. Die Studierenden konnten so im Sinne eines Vorbilds erleben, dass auch erfahrene Ärzte und Ärztinnen fragen und suchen – und die Antwort respektive die richtige Diagnose nicht sofort parat haben –, wodurch der Weg der klinischen Entscheidungsfindung als die notwendige Arbeit von Ärztinnen und Ärzten erfahrbar und erkennbar wurde.

In der Erarbeitung der Lernziele griffen thematische Einführungen durch die Dozierenden und eigenständige Bearbeitung von Papierfällen durch die Studierenden ineinander. Beispielsweise erarbeiteten die Studierenden im Rahmen der ersten Lerneinheit in Kleingruppen die Vorstellung eines Patienten wie für eine Visite und führten diese Fallpräsentation nachfolgend im Kurs durch. Schließlich stellten die Dozierenden darauf bezogen die Aspekte einer akkuraten Fallpräsentation zusammenfassend vor. Auf dieser Basis konnte die akkurate Fallpräsentation im weiteren Verlauf kritisch überprüft und die entsprechende Fähigkeit weiter entwickelt werden.

### **3.2 Durchführung des Seminars**

Am UKE findet der Seminarunterricht für Studierende im PJ in der Inneren Medizin – neben Lehrvisiten und weiteren Veranstaltungen – in Blöcken von acht Wochen statt, in denen jeweils Doppelstunden von insgesamt 90 Minuten einmal wöchentlich vorgesehen sind. Die Durchführung des neu konzipierten Seminars „Klinische Entscheidungsfindung“ übernahmen die oben genannten Dozierenden am ersten Termin gemeinsam. Den Unterricht der Termine zwei bis acht gestalteten ein einzelner Dozierender bzw. eine Dozierende oder ein Dozierenden-Team von zwei bis vier Lehrenden.

Der erste Durchlauf des Seminars fand ab Februar 2011 statt. Der Kurs wurde von da an dreimal jährlich, beginnend jeweils im Februar, Juni und Oktober angeboten. Die Innere Medizin können die Studierenden im Verlauf des PJs als erstes, zweites oder drittes Tertial absolvieren. Da das PJ im Februar und im Oktober begann, konnten jeweils Studierende des ersten und zweiten Tertials (Seminarbeginn im Februar), des zweiten und dritten Tertials (Seminarbeginn im Juni) und des ersten und dritten Tertials (Seminarbeginn im Oktober)

gleichzeitig den Kurs besuchen. Aus organisatorischen Gründen konnten vereinzelt Studierende erst zur vierten Seminarsitzung erstmalig zum Kurs kommen, die Selbsteinschätzung und Fallbearbeitung dieser Studierenden am Ende des Kurses ging nicht in die Auswertung des Seminars ein. Die Teilnahme am Kurs war, ebenso wie bei den übrigen Veranstaltungsangeboten in der Inneren Medizin, freiwillig. Im Rahmen dieses Projekts wurde der Kurs insgesamt siebenmal durchgeführt.

### **3.3 Fragebogen zur Selbsteinschätzung**

Der Erfolg des Seminars wurde zum einen überprüft, indem die Selbsteinschätzung der Studierenden im Hinblick auf verschiedene Aspekte der klinischen Entscheidungsfindung vor und nach dem Seminar erfragt wurde. Da es keinen validierten und auf die Seminarinhalte angepassten Bogen in der publizierten Literatur gab, wurde ein Instrument entwickelt, das auf der Basis der für die Lerninhalte formulierten Lernziele basierte und acht Items umfasste (siehe Anhang 9.2). Zum Beispiel wurde die Selbsteinschätzung zur Lerneinheit „Datenerhebung“ und „akkurate/abstrahierende Problempräsentation“ anhand der Aussage „Ich bin in der Lage, einen komplexen Patientenfall in zwei Sätzen wie für eine Visite zusammenzufassen.“ erfragt. Die Studierenden hatten die Möglichkeit, auf einer sechsstufigen Likert-Skala ihre Qualifikationen einzuordnen. In der Auswertung entsprach „gar nicht“ einem Punkt, „sehr gut“ sechs Punkten. Der Fragebogen wurde zu Beginn des ersten Seminartermins als Pre-Erhebung ausgegeben, dann wieder am Ende des Seminars. Der Fragebogen am Ende des Seminars umfasste die Selbsteinschätzung zweifach, nämlich als Post-Befragung und vergleichend als retrospektive Pre-Befragung, im weiteren Retropre-Befragung genannt, womit die Einschätzung der Studierenden hinsichtlich ihrer Fähigkeiten zu Beginn des Seminars aus der Perspektive nach dem Seminar erfragt wurde.

### **3.4 Bearbeitung eines Papierfalls**

Der Erfolg des Seminars wurde andererseits überprüft anhand der Bearbeitung eines Papierfalls, der vom Format her den im Seminar verwendeten Fällen entsprach (siehe Anhang 9.3). Zur Überprüfung der Qualifikationen hatten die Studierenden die Aufgabe, aus den vorgegebenen Informationen schriftlich eine Fallpräsentation in zwei bis drei Sätzen zu formulieren (Anhang 9.4). Darüber hinaus sollten sie Differentialdiagnosen nennen, die sie für den Fall für relevant hielten. Diese Differentialdiagnosen sollten nicht begründet, aber in eine Rangfolge gebracht werden. Außerdem sollten die Studierenden angeben, wie sicher sie sich der Diagnose waren, wobei die Zahlenwerte 1 bis 6 für „sicher bis unsicher“ vorgegeben waren. Zusätzlich wurde eine Einschätzung des Schwierigkeitsgrades des zu bearbeitenden Falles erfragt.

### **3.5 Datenerhebung**

Die Studierenden wurden in der ersten Stunde, unmittelbar mit Seminarbeginn und vor Erläuterung des Seminarkonzepts, gebeten, den Fragebogen zur Selbsteinschätzung und den genannten Papierfall zu bearbeiten. Zu diesem Fall folgten keine weiteren Erläuterungen und er wurde auch nicht während des Seminars verwendet. Die Unterlagen wurden eingesammelt, dann wurde das Seminarkonzept vorgestellt und mit der ersten Lerneinheit „Datenerhebung und Problempräsentation“ begonnen. Nach Abschluss des achten Seminartermins wurde der gleiche o. g. Fragebogen zur Selbsteinschätzung von den teilnehmenden Studierenden ausgefüllt, und zwar einmal als Retropre- und einmal als Post-Befragung. Auch wurde der schon vor Seminarbeginn verwendete Fall wieder bearbeitet. Hatten im Seminarverlauf mehrere Studierende vorab angegeben, dass sie an der letzten Seminarstunde nicht teilnehmen würden, wurden Fragebogen und Papierfall am Ende des siebten Seminartermins zur Bearbeitung ausgegeben.

### **3.6 Statistische Auswertung**

#### **3.6.1 Auswertung der Selbsteinschätzungsbögen**

Für alle statistischen Auswertungen wurde das Programm SPSS/PAWS 18 verwendet. Berechnet wurden zunächst auf Einzel-Item-Ebene die Mittelwerte (MW) und Standardabweichungen (SD) der Pre-, Retropre- und Post-Selbsteinschätzungsfragebögen aller Teilnehmenden, von denen komplette Datensätze vorlagen. Mittels Wilcoxon-Vorzeichen-Rang-Test wurde dann ermittelt, ob signifikante Unterschiede ( $p < 0,05$ ) zwischen Pre- bzw. Retropre- und Post-Befragung sowie zwischen Pre- und Retropre-Befragung bestanden. Anschließend wurden die Mittelwerte und Standardabweichung der einzelnen Items für Männer und Frauen getrennt ermittelt. Entsprechend wurde mittels Wilcoxon-Vorzeichen-Rang-Test jeweils für Männer und für Frauen als unterschiedlich große Gruppen untersucht, ob signifikante Unterschiede zwischen Pre- bzw. Retropre- und Post-Befragung sowie zwischen Pre- und Retropre-Befragung vorlagen. Auf Einzel-Item-Ebene wurden schließlich die Delta-Werte (Post-Werte minus Pre-Werte, Pre-Werte minus Retropre-Werte und Post-Werte minus Retropre-Werte) von Frauen und von Männern mittels Mann-Whitney-U-Test als Post-hoc-Test auf signifikante Unterschiede hin untersucht.

Außerdem wurden die Mittelwerte und Standardabweichungen auf Basis der Tertiale ermittelt und miteinander verglichen. Entsprechend wurde mittels Wilcoxon-Vorzeichen-Rang-Test jedes der drei unterschiedlich großen Tertiale in sich auf signifikante Unterschiede zwischen Pre- bzw. Retropre- und Post-Befragung sowie zwischen Pre- und Retropre-Befragung hin

untersucht. Auf Einzel-Item-Ebene wurden schließlich die Delta-Werte (Post-Werte minus Pre-Werte) von Teilnehmenden der drei verschiedenen Tertiale mittels Kruskal-Wallis-Test als Post-hoc-Test auf signifikante Unterschiede hin untersucht. Darüber hinaus wurden die drei Tertiale hinsichtlich ihrer Selbsteinschätzungen in Pre-, Retropre- und Post-Befragung mittels Kruskal-Wallis-Test verglichen.

Eine Berechnung der Effektstärke war aufgrund des Umfangs der Erhebung nicht sinnvoll und wurde entsprechend nicht vorgenommen. Mittels Cronbachs Alpha wurde die interne Konsistenz des Fragebogens berechnet.

### **3.6.2 Auswertung der Fallbearbeitung**

Der bearbeitete Fall wurde in zweierlei Hinsicht ausgewertet: Zum einen wurde die Präsentation des Falls untersucht im Hinblick auf

- die Anzahl der für die Fallpräsentation verwendeten Wörter,
- die Nennung von sieben zuvor für das Verständnis des Falls als relevant festgelegten Schlüsselwörtern, die bei korrekter Angabe mit jeweils einem Punkt bewertet wurden,
- fehlerhafte Angaben, die als ganzer oder halber Punkt von der für die Schlüsselwörter vergebenen Punktzahl abgezogen wurden.

Es wurde schließlich der Quotient aus der Anzahl der Schlüsselwörter minus Fehlerpunkten (SW = Dividend) und der Wörterzahl (WZ = Divisor) gebildet:  $SW/WZ$ . Verglichen wurden die Mittelwerte und Standardabweichungen der Wörterzahl pre mit der Wörterzahl post einerseits und des Quotienten pre mit dem Quotienten post andererseits. Zum anderen wurde die Anzahl der genannten Differentialdiagnosen pre und post miteinander verglichen, indem die Mittelwerte und die Standardabweichungen berechnet wurden. Die Signifikanz der Ergebnisse der Fallbeschreibung wurde ebenfalls mittels Wilcoxon-Vorzeichen-Rang-Test berechnet, sowohl in der Gesamtschau aller Teilnehmenden, als auch für die Gruppe der Männer und die Gruppe der Frauen getrennt sowie für die Tertiale getrennt. Es wurde ein Signifikanzniveau von  $p < 0,05$  festgelegt. Die Untersuchung auf signifikante Differenzen zwischen Männern und Frauen wurde wiederum mittels Mann-Whitney-U-Test durchgeführt, die auf signifikante Differenzen zwischen den Tertialen mittels Kruskal-Wallis-Test.

### **3.6.3 Vergleich zwischen Selbsteinschätzung und Fallbearbeitung**

Die zum Papierfall gestellten Aufgaben entsprachen den ersten beiden Items des Fragebogens zur Selbsteinschätzung, in denen es einmal um die Zusammenfassung eines komplexen Patientenfalls geht (Item 1 des Fragebogens), dann um die Nennung unterschied-

licher Differentialdiagnosen für ein individuelles Erkrankungsbild (Item 2 des Fragebogens). Die Korrelation zwischen den Ergebnissen der Auswertung der Selbsteinschätzung dieser beiden einzelnen Items und denen der Fallbearbeitung wurde nach Spearman-Rho für die Gesamtkohorte aller Teilnehmenden berechnet.

## 4 Ergebnisse

### 4.1 Soziodemographische Daten der Seminarteilnehmenden

Insgesamt beteiligten sich während der sechs in die Auswertung aufgenommenen Seminar-durchläufe zwischen Februar 2011 und Februar 2013 128 Teilnehmer und Teilnehmerinnen an der Datenerhebung, gaben also den Fragebogen und/oder die Fallbearbeitung vor und/oder nach dem Seminar ab, davon 3 Familiierende als Gäste. In die Seminauswertung wurden nur die 42 Studierenden mit kompletten Datensätzen eingeschlossen, also mit der Bearbeitung des Selbsteinschätzungsbogens und des Papierfalls sowohl vor als auch nach dem Seminar in Form von Pre-, Retropre- und Post-Befragungen. Die Teilnehmer und Teilnehmerinnen, deren Daten in die Studie eingingen, hatten an mindestens fünf von acht Seminarterminen teilgenommen. Von diesen 42 Studierenden waren 11 Männer und 31 Frauen. Sie befanden sich in unterschiedlichen PJ-Tertialen: 21 (7 männlich, 14 weiblich) absolvierten das erste, 13 (1 männlich, 12 weiblich) das zweite und 8 (3 männlich, 5 weiblich) das dritte Tertial.

### 4.2 Fragebogen zur Selbsteinschätzung

Die Cronbachs-Alpha-Werte wurden für den Pre-, Retropre- und Post-Fragebogen berechnet und liegen bei 0,80, 0,91 und 0,85.

#### 4.2.1 Selbsteinschätzung für die Tertiale gesamt

In Tabelle 3 ist dargestellt, wie die Gesamtheit der Teilnehmenden aller drei Tertiale sich in der Pre-, Retropre- und Post-Befragung selbst zu verschiedenen Qualifikationen der klinischen Entscheidungsfindung einschätzte. Nach dem Kurs bewerteten sich die Studierenden bei allen acht Items signifikant besser als vor dem Kurs (für Item 5:  $p = 0,01$ ; für alle anderen Items:  $p < 0,001$ ). Auch im Verhältnis zur Retropre-Befragung schätzten sich die Teilnehmenden nach dem Seminar mit  $p < 0,001$  für alle Items signifikant besser ein. Der größte Zuwachs von Pre nach Post zeigte sich bei Item 7 „Ich bin in der Lage, typische ärztliche Denkfehler zu erkennen und Risikosituationen für deren Entstehung zu identifizieren.“ (Pre:  $2,98 \pm 0,92$ ; Post:  $4,38 \pm 0,88$ ). Noch größer stellte sich der Zuwachs dieser Fähigkeit im Vergleich der Retropre- und Post-Selbsteinschätzung dar (Retropre:  $2,64 \pm 1,01$ ; Post:  $4,38 \pm 0,88$ ). Der geringste Zuwachs zeigte sich bei Item 5 „Ich bin in der Lage, medizinische Testergebnisse im Hinblick auf ihre Relevanz für den individuellen Patientenfall zu bewerten.“ (Pre:  $3,60 \pm 0,96$ ; Post:  $4,00 \pm 0,88$ ). Hier zeigte sich allerdings insgesamt der größte Unterschied zwischen Pre- und Retropre-Selbsteinschätzung (Pre:  $3,60 \pm 0,96$ ; Retropre:  $3,19 \pm 0,92$ ).

Das einzige Item, bei dem die Studierenden ihre Fähigkeit retropre, wenn auch nur geringfügig, besser einschätzten als vor dem Seminar, ist Item 6 „Ich bin in der Lage, typische Patientenpräsentationen unmittelbar einer Diagnose zuzuordnen.“ (Pre:  $3,67 \pm 0,87$ ; Retropre:  $3,69 \pm 1,02$ ). Dieses Item war in der Retropre-Befragung dasjenige mit dem höchsten Mittelwert. Die mit Ausnahme des genannten Items 6 im Vergleich von Pre- und Retropre-Befragung schlechtere Selbsteinschätzungen Retropre erwies sich für die Items 5 und 8 als signifikant (Item 5:  $p = 0,03$ ; Item 8:  $p = 0,038$ ). Den höchsten Mittelwert in der Pre- wie auch in der Post-Befragung wies das Item 1 auf, also die Fallpräsentation. Dieses ist auch das einzige Item, bei dem die Studierenden einen Punktwert über 4,5 erreichten, und zwar in der Post-Befragung.

**Tabelle 3: Selbsteinschätzung der Studierenden für die Tertiale insgesamt**

Ich bin in der Lage, ...	Pre MW $\pm$ SD	Retropre MW $\pm$ SD	Post MW $\pm$ SD
1) ... einen komplexen <b>Patientenfall</b> in zwei bis drei Sätzen <b>zusammenzufassen</b> .	3,81° $\pm$ 1,11	3,50 $\pm$ 1,08	4,76° $\pm$ 0,72
2) ... für ein individuelles Erkrankungsbild mehrere <b>Differentialdiagnosen</b> zu <b>benennen</b> .	3,52° $\pm$ 1,13	3,42 $\pm$ 1,06	4,21° $\pm$ 0,87
3) ... <b>Differentialdiagnosen</b> aus dem Gesamtkontext zu <b>begründen</b> und zu <b>gewichten</b> .	3,33° $\pm$ 1,05	3,31 $\pm$ 1,05	4,28° $\pm$ 0,74
4) ... Hypothesen-basiert <b>diagnostische Verfahren</b> in geeigneter Weise <b>anzuordnen</b> .	3,38° $\pm$ 0,96	3,21 $\pm$ 0,98	4,16° $\pm$ 0,88
5) ... medizinische <b>Testergebnisse</b> im Hinblick auf ihre Relevanz für den individuellen Patientenfall zu <b>bewerten</b> .	3,60* $\pm$ 0,96	3,19# $\pm$ 0,92	4,00° $\pm$ 0,88
6) ... typische <b>Patientenpräsentationen</b> <b>unmittelbar</b> einer Diagnose <b>zuzuordnen</b> .	3,67° $\pm$ 0,87	3,69 $\pm$ 1,02	4,30° $\pm$ 0,81
7) ... typische <b>ärztliche Denkfehler</b> zu <b>erkennen</b> und Risikosituationen für deren Entstehung zu identifizieren.	2,98° $\pm$ 0,92	2,64 $\pm$ 1,01	4,38° $\pm$ 0,88
8) ... die in einer medizinischen Entscheidung enthaltene <b>Ungewissheit</b> zu <b>erkennen</b> und Patienten zu <b>vermitteln</b> .	3,55° $\pm$ 1,06	3,19# $\pm$ 1,04	4,36° $\pm$ 0,88

Selbsteinschätzung der Studierenden zu ihren Qualifikationen hinsichtlich verschiedener Aspekte der klinischen Entscheidungsfindung in der Pre-, Retropre- und Post-Befragung: Tertiale 1, 2, 3 gesamt; Signifikanzen Pre versus Post (Spalte Pre, links), Signifikanzen Pre versus Retropre (Spalte Retropre, mittig) und Signifikanzen Retropre versus Post (Spalte Post, rechts):

#:  $p < 0,05$ , \*:  $p \leq 0,01$  und °:  $p \leq 0,001$ .

## 4.2.2 Selbsteinschätzung von Männern und Frauen im Vergleich

In Tabelle 4 sind die Ergebnisse der Geschlechter (G) im Vergleich dargestellt. Die Frauen (n = 31) schätzten ihre Qualifikationen in der Post-Befragung für alle acht Items signifikant besser ein als in der Pre-Befragung, während dies nur auf vier von acht Items für die kleinere Gruppe der Männer (n = 11) zutraf, und zwar auf die Items 1, 3, 4 und 7. Bei Item 5 beurteilten die Männer sich nach dem Seminar sogar schlechter als vor dem Seminar (Pre 4,09 ± 1,04; Post 3,91 ± 1,14), retropre schätzten sie ihre Qualifikation bezogen auf dieses Item um einen ganzen Skalen-Punkt schlechter ein als pre.

**Tabelle 4: Selbsteinschätzung von Männern und Frauen im Vergleich**

Ich bin in der Lage, ...	G	Pre MW ± SD	Retropre MW ± SD	Post MW ± SD
1) ... einen komplexen <b>Patientenfall</b> in zwei bis drei Sätzen <b>zusammenzufassen</b> .	w	3,81° ± 1,08	3,48 ± 1,03	4,71° ± 0,74
	m	3,82* ± 1,25	3,55 ± 1,29	4,91* ± 0,70
2) ... für ein individuelles Erkrankungsbild mehrere <b>Differentialdiagnosen</b> zu <b>benennen</b> .	w	3,48° ± 1,06	3,48 ± 0,93	4,29° ± 0,74
	m	3,64 ± 1,36	3,27 ± 1,42	4,00# ± 1,18
3) ... <b>Differentialdiagnosen</b> aus dem Gesamtkontext zu <b>begründen</b> und zu <b>gewichten</b> .	w	3,32° ± 1,08	3,29 ± 1,10	4,36° ± 0,76
	m	3,36# ± 1,03	3,36 ± 0,92	4,09* ± 0,70
4) ... Hypothesen-basiert <b>diagnostische Verfahren</b> in geeigneter Weise <b>anzuordnen</b> .	w	3,29° ± 0,94	3,10 ± 0,94	4,10° ± 0,83
	m	3,64# ± 1,03	3,55 ± 1,04	4,36* ± 1,03
5) ... medizinische <b>Testergebnisse</b> im Hinblick auf ihre Relevanz für den individuellen Patientenfall zu <b>bewerten</b> .	w	3,42# ± 0,89	3,23 ± 0,92	4,03° ± 0,80
	m	4,09 ± 1,04	3,09* ± 0,94	3,91# ± 1,14
6) ... typische <b>Patientenpräsentationen</b> <b>unmittelbar</b> einer Diagnose <b>zuzuordnen</b> .	w	3,68# ± 0,91	3,68 ± 0,95	4,26° ± 0,77
	m	3,64 ± 0,81	3,73 ± 1,27	4,46# ± 0,93
7) ... typische <b>ärztliche Denkfehler</b> zu <b>erkennen</b> und Risikosituationen für deren Entstehung zu identifizieren.	w	3,06° ± 0,96	2,52# ± 1,03	4,42° ± 0,85
	m	2,73* ± 0,79	3,00 ± 0,89	4,27# ± 1,01
8) ... die in einer medizinischen Entscheidung enthaltene <b>Ungewissheit</b> zu <b>erkennen</b> und Patienten zu <b>vermitteln</b> .	w	3,52° ± 1,06	3,16 ± 1,00	4,39° ± 0,76
	m	3,64 ± 1,12	3,27 ± 1,19	4,27* ± 1,19

Selbsteinschätzung der Studierenden zu ihren Qualifikationen hinsichtlich verschiedener Aspekte der klinischen Entscheidungsfindung in der Pre-, Retropre- und Post-Befragung: Tertiale 1, 2, 3 gesamt; die Geschlechter (G) im Vergleich (Frauen: w; n = 31) und (Männer: m; n = 11); Signifikanzen Pre versus Post (Spalte Pre, links), Signifikanzen Pre versus Retropre (Spalte Retropre, mittig) und Signifikanzen Retropre versus Post (Spalte Post, rechts): #: p < 0,05, \*: p ≤ 0,01 und °: p ≤ 0,001.



Der Zuwachs an Qualifikationen hinsichtlich verschiedener Aspekte der klinischen Entscheidungsfindung in der Selbsteinschätzung von der Retropre- zur Post-Befragung stellte sich bei beiden Geschlechtern für alle Items als signifikant dar, bei den Frauen (alle Items:  $p < 0,001$ ) ebenso wie bei der kleineren Gruppe der Männer (Items 1 und 4:  $p = 0,007$ ; Item 3:  $p = 0,005$ ; Item 8:  $p = 0,009$ ; Items 2, 6, 7:  $p = 0,011$ ; Item 5:  $p = 0,024$ ).

Die Differenz zwischen der Pre- und der Retropre-Selbsteinschätzung erscheint im Wilcoxon-Test für jedes Geschlecht für jeweils ein Item signifikant: bei den Männern für Item 5 ( $p = 0,005$ ), also die Bewertung medizinischer Testergebnisse im Hinblick auf ihre Relevanz für einen Patientenfall, und bei den Frauen für Item 7 ( $p = 0,047$ ), also das Erkennen typischer ärztlicher Denkfehler und die Identifikation von Risikosituationen für deren Entstehung. Bei der Beurteilung des letztgenannten Aspekts schätzten die Männer sich in der Retropre-Befragung sogar besser ein als vor dem Seminar. Dies trifft ebenfalls auf das Item 6 zu, bei dem die unmittelbare Zuordnung von typischen Patientenpräsentationen zu einer Diagnose beurteilt wird.

Insgesamt viermal, nämlich bei den Items 2, 3, 5 und 8, schätzten die Männer ihre Qualifikationen vor dem Seminar besser ein als die Frauen, nach dem Seminar jedoch schlechter. Auf das Item 7 bezogen bewerteten sich die Männer vor dem Seminar mit einem Punktwert von unter 3 auf der Likert-Skala ( $2,73 \pm 0,79$ ) und erreichten von Pre zu Post einen Zuwachs von mehr als 1,5 Punkten. Bei den Frauen zeigte sich bei diesem Item mit 1,36 Punkten ebenfalls die größte Differenz zwischen Pre- und Post-Werten. Im Vergleich der Pre-Werte von Männern und Frauen miteinander zeigten sich im Mann-Whitney-U-Test keine signifikanten Unterschiede zwischen den Geschlechtern, auch nicht im Vergleich der Retropre- oder Post-Werte.

In Tabelle 5 sind die Delta-Werte dargestellt: Post-Werte minus Pre-Werte, Pre-Werte minus Retropre-Werte und Post-Werte minus Retropre-Werte. Der Unterschied zwischen Männern und Frauen erweist sich im Vergleich der Selbsteinschätzung vor und nach dem Kurs im Mann-Whitney-U-Test einzig für Item 5, die Bewertung medizinischer Testergebnisse im Hinblick auf ihre Relevanz für den individuellen Patientenfall, als signifikant ( $p = 0,019$ ). Außerdem ist der Unterschied zwischen Männern und Frauen signifikant im Vergleich der Veränderungen von der Pre- zur Retropre-Selbsteinschätzung, und zwar ebenfalls für das Item 5 ( $p = 0,025$ ).

**Tabelle 5: Delta-Werte Männer und Frauen im Vergleich**

Ich bin in der Lage, ...	G	Post minus Pre $\Delta$ MW $\pm$ SD	Pre minus Retropre $\Delta$ MW $\pm$ SD	Post minus Retropre $\Delta$ MW $\pm$ SD
1) ... einen komplexen <b>Patientenfall</b> in zwei bis drei Sätzen <b>zusammenzufassen</b> .	w	0,90 $\pm$ 1,04	0,32 $\pm$ 1,05	1,23 $\pm$ 0,88
	m	1,09 $\pm$ 0,94	0,27 $\pm$ 1,19	1,36 $\pm$ 0,92
2) ... für ein individuelles Erkrankungsbild mehrere <b>Differentialdiagnosen</b> zu <b>benennen</b> .	w	0,81 $\pm$ 1,14	0,00 $\pm$ 0,97	0,81 $\pm$ 0,91
	m	0,36 $\pm$ 0,67	0,36 $\pm$ 0,924	0,73 $\pm$ 0,65
3) ... <b>Differentialdiagnosen</b> aus dem Gesamtkontext zu <b>begründen</b> und zu <b>gewichten</b> .	w	1,03 $\pm$ 0,95	0,03 $\pm$ -0,95	1,06 $\pm$ 0,89
	m	0,73 $\pm$ 0,79	0,00 $\pm$ 1,00	0,73 $\pm$ 0,47
4) ... Hypothesen-basiert <b>diagnostische Verfahren</b> in geeigneter Weise <b>anzuordnen</b> .	w	0,81 $\pm$ 1,01	0,19 $\pm$ 1,20	1,00 $\pm$ 0,93
	m	0,73 $\pm$ 1,01	0,09 $\pm$ 0,94	0,82 $\pm$ 0,60
5) ... medizinische <b>Testergebnisse</b> im Hinblick auf ihre Relevanz für den individuellen Patientenfall zu <b>bewerten</b> .	w	0,61 <sup>#</sup> $\pm$ 0,99	0,19 <sup>#</sup> $\pm$ 1,20	0,81 $\pm$ 0,75
	m	-0,18 <sup>#</sup> $\pm$ 0,60	1,00 <sup>#</sup> $\pm$ 0,63	0,82 $\pm$ 0,87
6) ... typische <b>Patientenpräsentationen</b> <b>unmittelbar</b> einer Diagnose <b>zuzuordnen</b> .	w	0,58 $\pm$ 0,96	0,00 $\pm$ 0,89	0,58 $\pm$ 0,77
	m	0,82 $\pm$ 1,25	-0,09 $\pm$ 1,58	0,73 $\pm$ 0,65
7) ... typische <b>ärztliche Denkfehler</b> zu <b>erkennen</b> und Risikosituationen für deren Entstehung zu identifizieren.	w	1,36 $\pm$ 1,23	0,55 $\pm$ 1,34	1,90 $\pm$ 1,08
	m	1,55 $\pm$ 1,13	-0,27 $\pm$ 1,20	1,27 $\pm$ 1,10
8) ... die in einer medizinischen Entscheidung enthaltene <b>Ungewissheit</b> zu <b>erkennen</b> und Patienten zu <b>vermitteln</b> .	w	0,87 $\pm$ 1,15	0,35 $\pm$ 1,36	1,23 $\pm$ 0,85
	m	0,64 $\pm$ 1,21	0,36 $\pm$ 1,36	1,00 $\pm$ 0,78

Delta-Werte ( $\Delta$ ; MW  $\pm$  SD) der Selbsteinschätzung der Studierenden: Post minus Pre, Pre minus Retropre, Post minus Retropre, die Geschlechter (G) im Vergleich (Frauen: w; n = 31 und Männer: m; n = 11); signifikante Differenzen zwischen w und m für Item 5;  $\Delta$  Post minus Pre und Pre minus Retropre: #: p < 0,05.

### 4.2.3 Selbsteinschätzung innerhalb der Tertiale

In den Tabellen 6, 7 und 8 ist nach Tertialen unterschieden dargestellt, wie die Studierenden ihre Qualifikationen im Hinblick auf die verschiedenen Aspekte der klinischen Entscheidungsfindung selbst einschätzten.

Im ersten Tertial (n = 21) (Tabelle 6) zeigte sich ein signifikanter Unterschied zwischen Pre- und Post-Befragung für die Items 1 (p = 0,001), 2 (p = 0,016), 3 (p = 0,003), 4 (p = 0,047), 6 (p = 0,007) und 7 (p < 0,001), jedoch nicht für die Items 5 und 8. Ein signifikanter Unter-

schied zwischen den Pre- und Retropre-Selbstbeurteilungen zeigte sich für die Items 5 ( $p = 0,022$ ) und 8 ( $p = 0,027$ ). Im Vergleich von Retropre- und Post-Werten zeigten sich für alle Items Signifikanzen (Items 1, 3, 7 und 8:  $p < 0,001$ ; Items 5 und 6:  $p = 0,002$ ; Item 2:  $p = 0,003$ ; Item 4:  $p = 0,004$ ).

Die Studierenden des ersten Tertials erreichten von Pre zu Post nur bei Item 7 „Ich bin in der Lage, typische ärztliche Denkfehler zu erkennen.“ einen Zuwachs von mehr als einem Punkt auf der Likert-Skala (Pre  $2,95 \pm 0,92$ ; Post  $4,57 \pm 0,98$ ). Der Zuwachs von absolut 1,62 Punkten von Pre nach Post ist im Vergleich der Tertiale der größte bezogen auf alle Items.

**Tabelle 6: Selbsteinschätzung der Teilnehmenden des ersten Tertials (n = 21)**

Ich bin in der Lage ...	Pre MW $\pm$ SD	Retropre MW $\pm$ SD	Post MW $\pm$ SD
1) einen Patientenfall ... zusammenzufassen.	3,38° $\pm$ 1,24	3,33 $\pm$ 1,02	4,62° $\pm$ 0,80
2) ... Differentialdiagnosen (DD) zu benennen.	3,38# $\pm$ 1,07	3,19 $\pm$ 1,17	4,00* $\pm$ 0,95
3) DD ... zu begründen und zu gewichten.	3,19* $\pm$ 1,03	3,14 $\pm$ 0,96	4,00° $\pm$ 0,63
4) ... diagnostische Verfahren ... anzuordnen.	3,33# $\pm$ 0,86	3,10 $\pm$ 0,89	3,81* $\pm$ 0,93
5) ... Testergebnisse ... zu bewerten.	3,62 $\pm$ 0,92	2,86# $\pm$ 0,96	3,62* $\pm$ 0,92
6) typische Patientenpräsentationen ... zuzuordnen.	3,43* $\pm$ 0,98	3,52 $\pm$ 1,21	4,19* $\pm$ 0,87
7) ... ärztliche Denkfehler zu erkennen ...	2,95° $\pm$ 0,92	2,71 $\pm$ 0,90	4,57° $\pm$ 0,98
8) ... Ungewissheit zu erkennen und zu vermitteln.	3,62 $\pm$ 0,97	3,10# $\pm$ 0,94	4,10° $\pm$ 0,94

Selbsteinschätzung der Studierenden zu ihren Qualifikationen hinsichtlich verschiedener Aspekte der klinischen Entscheidungsfindung in der Pre-, Retropre- und Post-Befragung: Terial 1; signifikante Zuwächse Pre versus Post (Spalte Pre, links), signifikant schlechtere Selbstbeurteilung Pre versus Retropre (Spalte Retropre, mittig) und signifikante Zuwächse Retropre versus Post (Spalte Post, rechts): #:  $p < 0,05$ , \*:  $p \leq 0,01$  und °:  $p \leq 0,001$ .

Auch im zweiten Tertial (n = 13) zeigte sich für sechs der acht Items zwischen der Pre- und der Post-Befragung ein signifikanter Zuwachs, und zwar für die Items 2 (p = 0,01), 3 und 5 (p = 0,002), 4 (p = 0,004), 8 (p = 0,007), 7 (p = 0,023), jedoch nicht für Item 1 und Item 6. Eine im Verhältnis zur Pre-Befragung signifikant schlechtere Retro-pre-Selbstbeurteilung zeigte sich für Item 1. Im Vergleich von Retro-pre- und Post-Werten zeigte sich für alle Items ein signifikanter Zuwachs bei der Bewertung der Qualifikationen (Item 5: p = 0,001; Items 4 und 7: p = 0,002; Items 3 und 8: p = 0,003; Item 1: p = 0,004; Item 2: p = 0,005; Item 6: p = 0,02).

Die Teilnehmenden des zweiten Tertials erreichten von Pre nach Post bei sechs Items einen absoluten Zuwachs von einem Punktwert oder mehr auf der Likert-Skala, nämlich bei den Items 2, 3, 4, 5, 7 und 8.

**Tabelle 7: Selbsteinschätzung der Teilnehmenden des zweiten Tertials (n = 13)**

Ich bin in der Lage ...	Pre M ± SD	Retro-pre M ± SD	Post M ± SD
1) einen Patientenfall ... zusammenzufassen.	4,31 ± 0,75	3,46* ± 1,13	4,69* ± 0,63
2) ... Differentialdiagnosen (DD) zu benennen.	3,54* ± 1,13	3,69 ± 0,95	4,54* ± 0,78
3) DD ... zu begründen und zu gewichten.	3,31* ± 0,95	3,15 ± 1,07	4,54* ± 0,88
4) ... diagnostische Verfahren ... anzuordnen.	3,23* ± 1,09	3,00 ± 0,82	4,38* ± 0,65
5) ... Testergebnisse ... zu bewerten.	3,23* ± 0,93	3,46 ± 0,66	4,38° ± 0,65
6) typische Patientenpräsentationen ... zuzuordnen.	4,00 ± 0,71	3,85 ± 0,80	4,38# ± 0,77
7) ... ärztliche Denkfehler zu erkennen ...	3,08# ± 0,95	2,31 ± 1,03	4,15* ± 0,90
8) ... Ungewissheit zu erkennen und zu vermitteln.	3,31* ± 1,32	3,15 ± 1,07	4,54* ± 0,78

Selbsteinschätzung der Studierenden zu ihren Qualifikationen hinsichtlich verschiedener Aspekte der klinischen Entscheidungsfindung in der Pre-, Retro-pre- und Post-Befragung: Tertial 2; signifikante Zuwächse Pre versus Post (Spalte Pre, links); signifikant schlechtere Selbstbeurteilung Pre versus Retro-pre (Spalte Retro-pre, mittig) und signifikante Zuwächse Retro-pre versus Post (Spalte Post, rechts): #: p < 0,05, \*: p ≤ 0,01 und °: p ≤ 0,001.

Im dritten Tertial (n = 8) zeigte sich bei der Hälfte der Items ein signifikanter Zuwachs in der Selbsteinschätzung von Pre zu Post, nämlich für die Items 1 (p = 0,024), 4 (p = 0,011), 7 (p = 0,008) und 8 (p = 0,02), also nicht für die Items 2, 3, 5 und 6. Ein signifikanter Unterschied zwischen den Pre- und Retro-pre-Selbstbeurteilungen zeigte sich nicht. Im Vergleich Retro-pre- und Post-Werte zeigten sich für fünf der acht Items signifikante Werte (Item 1: p = 0,023;

Item 2:  $p = 0,025$ ; Item 4:  $p = 0,02$ ; Item 7:  $p = 0,026$ ; Item 8:  $p = 0,014$ ). Die Teilnehmenden des dritten Tertials erreichten von Pre nach Post viermal einen Zuwachs von einem Punkt oder mehr auf der Likert-Skala, und zwar bei den Items 1, 4, 7 und 8.

**Tabelle 8: Selbsteinschätzung der Teilnehmenden des dritten Tertials (n = 8)**

Ich bin in der Lage ...	Pre MW $\pm$ SD	Retropre MW $\pm$ SD	Post MW $\pm$ SD
1) einen Patientenfall ... zusammenzufassen.	4,13 <sup>#</sup> $\pm$ 0,83	4,00 $\pm$ 1,20	5,25 <sup>#</sup> $\pm$ 0,46
2) ... Differentialdiagnosen (DD) zu benennen.	3,88 $\pm$ 1,36	3,63 $\pm$ 0,92	4,25 <sup>#</sup> $\pm$ 0,71
3) DD ... zu begründen und zu gewichten.	3,75 $\pm$ 1,28	4,00 $\pm$ 1,07	4,63 $\pm$ 0,52
4) ... diagnostische Verfahren ... anzuordnen.	3,75 <sup>#</sup> $\pm$ 1,04	3,88 $\pm$ 1,25	4,75 <sup>#</sup> $\pm$ 0,71
5) ... Testergebnisse ... zu bewerten.	4,13 $\pm$ 0,99	3,63 $\pm$ 0,92	4,38 $\pm$ 0,74
6) typische Patientenpräsentationen ... zuzuordnen.	3,75 $\pm$ 0,71	3,88 $\pm$ 0,83	4,50 $\pm$ 0,76
7) ... ärztliche Denkfehler zu erkennen ...	2,88* $\pm$ 0,99	3,00 $\pm$ 1,20	4,25 <sup>#</sup> $\pm$ 0,46
8) ... Ungewissheit zu erkennen und zu vermitteln.	3,75 <sup>#</sup> $\pm$ 0,89	3,50 $\pm$ 1,31	4,75 <sup>#</sup> $\pm$ 0,71

Selbsteinschätzung der Studierenden zu ihren Qualifikationen hinsichtlich verschiedener Aspekte der klinischen Entscheidungsfindung in der Pre-, Retropre- und Post-Befragung: Tertial 3; signifikante Zuwächse Pre versus Post (Spalte Pre, links), keine signifikanten Differenzen in der Selbstbeurteilung Pre versus Retropre (Spalte Retropre, mittig), signifikante Zuwächse Retropre versus Post (Spalte Post, rechts): #:  $p < 0,05$  und \*:  $p \leq 0,01$ .

#### 4.2.4 Selbsteinschätzung der Tertiale im Vergleich: Delta Post minus Pre

In Tabelle 9 sind die Delta-Werte dargestellt: Post-Werte minus Pre-Werte für das erste, zweite und dritte Tertial. Dabei stellen sich die Unterschiede zwischen den Tertialen die signifikanten Ergebnisse betreffend im Kruskal-Wallis-Test nicht nur als Effekte der Gruppengrößen, sondern im Falle der Items 1 und 5 auch als Effekte der Tertiale dar: Für diese Items zeigt sich mit Werten von  $p = 0,047$  respektive  $p = 0,002$  jeweils ein signifikanter Unterschied zwischen den Tertialen. Bei Item 1 ist der Zuwachs von der Pre- zur Post-Bearbeitung für das zweite Tertial im Verhältnis zum ersten Tertial signifikant geringer (Post-hoc-Test Tertial 2 zu Tertial 1:  $p = 0,018$ ), bei Item 5 ist der Zuwachs von der Pre- zur Post-Bearbeitung für das zweite Tertial signifikant größer als für das erste und dritte Tertial (Post-hoc-Test Tertial 1 zu Tertial 2:  $p < 0,001$ ; Tertial 3 zu Tertial 2:  $p = 0,044$ ).

**Tabelle 9: Delta-Werte Post minus Pre für die Tertiale im Vergleich**

Ich bin in der Lage ...	Post minus Pre Δ MW ± SD		
	Tertial 1 n = 21	Tertial 2 n = 13	Tertial 3 n = 8
1) einen Patientenfall ... zusammenzufassen.	1,24 <sup>#</sup> ± 1,14	0,39 ± 0,65	1,13 ± 0,84
2) ... Differentialdiagnosen (DD) zu benennen.	0,62 ± 1,02	1,00 ± 1,00	0,38 ± 1,19
3) DD ... zu begründen und zu gewichten.	0,81 ± 0,93	1,23 ± 0,83	0,88 ± 0,99
4) ... diagnostische Verfahren ... anzuordnen.	0,48 ± 1,08	1,15 ± 0,99	1,00 ± 0,54
5) ... Testergebnisse ... zu bewerten.	0° ± 0,89	1,15 <sup>#</sup> ± 0,69	0,25 ± 0,89
6) typische Patientenpräsentationen ... zuzuordnen.	0,76 ± 1,09	0,39 ± 0,87	0,75 ± 1,17
7) ... ärztliche Denkfehler zu erkennen ...	1,62 ± 1,20	1,08 ± 1,38	1,38 ± 0,74
8) ... Ungewissheit zu erkennen und zu vermitteln.	0,48 ± 1,17	1,23 ± 1,17	1,00 ± 0,93

Delta-Werte (Δ; MW ± SD) der Selbsteinschätzung der Studierenden: Post minus Pre, die Tertiale im Vergleich; signifikante Differenzen für Item 1 (p < 0,05) und Item 5 (p < 0,05); Tertial 1 : Tertial 2 (Spalte Tertial 1; links) und Tertial 2 : Tertial 3 (Spalte Tertial 2, mittig); #: p < 0,05 und °: p ≤ 0,001.

#### 4.2.5 Selbsteinschätzung der Tertiale im Vergleich: Pre, Retropre, Post

In den Tabellen 10, 11 und 12 ist jeweils für die Pre-, Retropre- und Post-Befragung dargestellt, wie die Teilnehmenden der verschiedenen Tertiale ihre Qualifikationen im Hinblick auf die verschiedenen Aspekte der klinischen Entscheidungsfindung bewerteten.

In der Pre-Befragung (Tabelle 10) zeigte das erste Tertial mit jeweils 3,62 Punkten auf der Likert-Skala die höchsten Werte für die Items 5 und 8. Das zweite Tertial dagegen zeigte für die Items 1 und 6 mit 4 oder mehr Punkten die höchsten Werte. Das dritte Tertial hat mit jeweils mehr als 4 Punkten auf der Likert-Skala die höchsten Werte bei den Items 1 und 5. Alle Tertiale schätzten ihre Qualifikationen vor dem Seminar am schlechtesten für Item 7 ein, für das dritte Tertial ergab sich mit 2,88 Punkten bei diesem Item der niedrigste Wert in der Pre-Befragung überhaupt. Im Kruskal-Wallis-Test ergibt sich hinsichtlich der Selbsteinschätzung vor dem Seminar kein signifikanter Unterschied zwischen den Tertialen.

Die Selbsteinschätzung der Teilnehmenden fiel in der Retropre-Bewertung (Tabelle 11) in allen Tertialen durchschnittlich schlechter aus als im Pre-Test. Während die Studierenden des ersten Tertials sich in der Retropre-Befragung mit Werten zwischen 2,71 und 3,56 auf

der Likert-Skala einschätzten, lag die Spanne in der Selbstbewertung bei den Teilnehmenden des zweiten Tertials zwischen 2,31 und 3,85. Und bei denen des dritten Tertials lagen die Werte zwischen 3,00 und 4,00 Punkten auf der Likert-Skala. Im Kruskal-Wallis-Test ergab sich kein signifikanter Unterschied zwischen den Tertialen.

In der Post-Befragung (Tabelle 12) erreichten die Teilnehmenden des dritten Tertials mit 5,25 Punkten den höchsten Punktwert überhaupt, und zwar für das Item 1. Auch die Teilnehmenden des ersten und zweiten Tertials erreichten bei diesem Item ihren höchsten Punktwert. Am schlechtesten schätzten sich in der Post-Befragung mit 3,62 Punkten für Item 5 die Teilnehmenden des ersten Tertials ein. Dies ist der einzige Fall, bei dem der Wert nach dem Seminar nicht höher als vor dem Seminar lag. Das zweite und dritte Tertial erreichte mit  $4,15 \pm 0,90$  bzw.  $4,25 \pm 0,46$  seinen jeweils schlechtesten Wert bei Item 7. Dieses ist das einzige Item, bei dem sich die Teilnehmenden des ersten Tertials mit  $4,57 \pm 0,98$  nach dem Seminar besser einschätzten als ihre Kommilitonen und Kommilitoninnen. Im Kruskal-Wallis-Test zeigt sich hinsichtlich der Post-Selbsteinschätzung der Teilnehmenden der drei Tertiale im Vergleich mit  $p < 0,05$  ein signifikanter Unterschied für Item 3 (Post-hoc-Test Tertial 1 zu Tertial 2 wie zu Tertial 3:  $p < 0,05$ ), für Item 4 (Post-hoc-Test Tertial 1 zu Tertial 3:  $p < 0,05$ ) und für Item 5 (Post-hoc-Test Tertial 1 zu Tertial 2 wie zu Tertial 3:  $p < 0,05$ ).

**Tabelle 10: Pre-Befragung, die Tertiale im Vergleich**

Ich bin in der Lage ...	<b>Tertial 1</b> (n = 21)	<b>Tertial 2</b> (n = 13)	<b>Tertial 3</b> (n = 8)
	MW $\pm$ SD	MW $\pm$ SD	MW $\pm$ SD
1) einen Patientenfall ... zusammenzufassen.	3,38 $\pm$ 1,24	4,31 $\pm$ 0,75	4,13 $\pm$ 0,83
2) ... Differentialdiagnosen (DD) zu benennen.	3,38 $\pm$ 1,07	3,54 $\pm$ 1,13	3,88 $\pm$ 1,36
3) DD ... zu begründen und zu gewichten.	3,19 $\pm$ 1,03	3,31 $\pm$ 0,95	3,75 $\pm$ 1,28
4) ... diagnostische Verfahren ... anzuordnen.	3,33 $\pm$ 0,86	3,23 $\pm$ 1,09	3,75 $\pm$ 1,04
5) ... Testergebnisse ... zu bewerten.	3,62 $\pm$ 0,92	3,23 $\pm$ 0,93	4,13 $\pm$ 0,99
6) typische Patientenpräsentationen ... zuzuordnen.	3,43 $\pm$ 0,98	4,00 $\pm$ 0,71	3,75 $\pm$ 0,71
7) ... ärztliche Denkfehler zu erkennen ...	2,95 $\pm$ 0,92	3,08 $\pm$ 0,95	2,88 $\pm$ 0,99
8) ... Ungewissheit zu erkennen und zu vermitteln.	3,62 $\pm$ 0,97	3,31 $\pm$ 1,32	3,75 $\pm$ 0,89

Fragebogen Pre, die Tertiale im Vergleich: Es zeigten sich keine signifikanten Unterschiede.

**Tabelle 11: Retropre-Befragung, die Tertiale im Vergleich**

Ich bin in der Lage ...	<b>Tertial 1</b> (n = 21)	<b>Tertial 2</b> (n = 13)	<b>Tertial 3</b> (n = 8)
	MW ± SD	MW ± SD	MW ± SD
1) einen Patientenfall ... zusammenzufassen.	3,33 ± 1,02	3,46 ± 1,13	4,00 ± 1,20
2) ... Differentialdiagnosen (DD) zu benennen.	3,19 ± 1,17	3,69 ± 0,95	3,63 ± 0,92
3) DD ... zu begründen und zu gewichten.	3,14 ± 0,96	3,15 ± 1,07	4,00 ± 1,07
4) ... diagnostische Verfahren ... anzuordnen.	3,10 ± 0,89	3,00 ± 0,82	3,88 ± 1,25
5) ... Testergebnisse ... zu bewerten.	2,86 ± 0,96	3,46 ± 0,66	3,63 ± 0,92
6) typische Patientenpräsentationen ... zuzuordnen.	3,52 ± 1,21	3,85 ± 0,80	3,88 ± 0,83
7) ... ärztliche Denkfehler zu erkennen ...	2,71 ± 0,90	2,31 ± 1,03	3,00 ± 1,20
8) ... Ungewissheit zu erkennen und zu vermitteln.	3,10 ± 0,94	3,15 ± 1,07	3,50 ± 1,31

Fragebogen Retropre, die Tertiale im Vergleich: Es zeigten sich keine signifikanten Unterschiede.

**Tabelle 12: Post-Befragung, die Tertiale im Vergleich**

Ich bin in der Lage ...	<b>Tertial 1</b> (n = 21)	<b>Tertial 2</b> (n = 13)	<b>Tertial 3</b> (n = 8)
	MW ± SD	MW ± SD	MW ± SD
1) einen Patientenfall ... zusammenzufassen.	4,62 ± 0,80	4,69 ± 0,63	5,25 ± 0,46
2) ... Differentialdiagnosen (DD) zu benennen.	4,00 ± 0,95	4,54 ± 0,78	4,25 ± 0,71
3) DD ... zu begründen und zu gewichten.	4,00 ± 0,63	4,54 <sup>#</sup> ± 0,88	4,63 <sup>#</sup> ± 0,52
4) ... diagnostische Verfahren ... anzuordnen.	3,81 ± 0,93	4,38 ± 0,65	4,75 <sup>#</sup> ± 0,71
5) ... Testergebnisse ... zu bewerten.	3,62 ± 0,92	4,38 <sup>#</sup> ± 0,65	4,38 <sup>#</sup> ± 0,74
6) typische Patientenpräsentationen ... zuzuordnen.	4,19 ± 0,87	4,38 ± 0,77	4,50 ± 0,76
7) ... ärztliche Denkfehler zu erkennen ...	4,57 ± 0,98	4,15 ± 0,90	4,25 ± 0,46
8) ... Ungewissheit zu erkennen und zu vermitteln.	4,10 ± 0,94	4,54 ± 0,78	4,75 ± 0,71

Fragebogen Post, die Tertiale im Vergleich; signifikante Differenzen für die Items 3, 4 und 5: Tertial 1 versus 2 (Spalte Tertial 2, mittig); Tertial 1 versus 3 (Spalte Tertial 3, rechts); #: p < 0,05.



## 4.3 Bearbeitung des Papierfalls

### 4.3.1 Bearbeitung des Papierfalls für die Tertiale gesamt

In Tabelle 13 ist die Auswertung der Bearbeitung des Papierfalls durch die Studierenden aller drei Tertiale als Gesamtheit vor und nach der Teilnahme an dem Seminar dargestellt. Die Anzahl der für die Fallpräsentation verwendeten Wörter sank von der Pre- zur Post-Befragung signifikant ( $p = 0,001$ ). Allerdings lag die Anzahl der genannten Schlüsselwörter im Post-Test sogar niedriger, wobei sich dies nicht als signifikant erwies. Auch der Quotient, also die Anzahl genannter Schlüsselwörter abzüglich von Fehlerpunkten für falsche Angaben geteilt durch die Wörterzahl, stieg signifikant an ( $p = 0,011$ ). Andere signifikante Änderungen konnten nicht beobachtet werden. Die Anzahl der Fehler lag nach dem Seminar zwar niedriger und die Anzahl der genannten Differentialdiagnosen höher, jedoch erwiesen sich diese Veränderungen nicht als signifikant.

**Tabelle 13: Bearbeitung des Papierfalls für die Gesamtheit der Studierenden**

Fall n = 42	WZ	SW	Quotient (SW/WZ)	Fehler	DD
<b>Pre</b> (MW ± SD)	37,67 ± 9,08	5,35 ± 1,23	0,15 ± 0,04	0,42 ± 0,63	3,43 ± 1,19
<b>Post</b> (MW ± SD)	31,64° ± 9,74	5,24 ± 1,65	0,17# ± 0,06	0,25 ± 0,50	3,79 ± 1,24

MW und SD als Ergebnisse der Bearbeitung des Papierfalls der Studierenden aller drei Tertiale gesamt; signifikante Werte: Wörterzahl Post < Pre, Quotient Post > Pre; #:  $p < 0,05$  und °:  $p \leq 0,001$ .

### 4.3.2 Bearbeitung des Papierfalls Männer und Frauen im Vergleich

In Tabelle 14 sind die Ergebnisse der Fallbearbeitung von Frauen ( $n = 31$ ) und Männern ( $n = 11$ ) vergleichend dargestellt. Signifikant bessere Ergebnisse zwischen der Bearbeitung des Papierfalls vor und nach dem Seminar zeigten sich ausschließlich für die Frauen, und zwar für die Wörterzahl ( $p = 0,001$ ) und für den Quotienten ( $p = 0,015$ ). Im Mann-Whitney-U-Test lässt sich kein signifikanter Unterschied zwischen Männern und Frauen nachweisen. Allerdings sank bei den Frauen die Zahl der verwendeten Schlüsselwörter, wenn auch nur geringfügig, wohingegen die Männer nach dem Seminar einen etwas höheren Anteil an Schlüsselwörtern benannten. Für beide Gruppen zeigte sich eine leichte Abnahme der Fehlerzahl. Die Frauen benannten nach dem Seminar durchschnittlich etwas mehr als vier Differentialdiagnosen. Die Männer gaben nach dem Seminar fast eine Differentialdiagnose weniger als die Frauen an.

**Tabelle 14: Bearbeitung des Papierfalls von Männern und Frauen im Vergleich**

Fall n = 42	G	WZ	SW	Quotient (SW/WZ)	Fehler	DD
Pre (MW ± SD)	w	37,19 ± 8,55	5,44 ± 1,21	0,15 ± 0,04	0,34 ± 0,45	3,65 ± 1,17
	m	39,00 ± 10,79	5,09 ± 1,30	0,14 ± 0,05	0,64 ± 0,98	2,82 ± 1,08
Post (MW ± SD)	w	29,87° ± 9,05	5,18 ± 1,49	0,18# ± 0,06	0,23 ± 0,44	4,03 ± 1,08
	m	36,64 ± 10,29	5,41 ± 2,13	0,15 ± 0,06	0,32 ± 0,64	3,09 ± 1,45

MW und SD als Ergebnisse der Bearbeitung des Papierfalls nach Geschlechtern (G) getrennt (Frauen: w, n = 31 und Männer: m, n = 11); signifikante Werte: Wörterzahl Post < Pre, Quotient Post > Pre (nur w): #: p < 0,05 und °: p ≤ 0,001.

#### **4.3.3 Bearbeitung des Papierfalls innerhalb der Tertiale und im Vergleich**

In Tabelle 15 sind die Ergebnisse der Fallbearbeitung nach Tertialen getrennt dargestellt. Für das erste Tertial erwies sich die Reduktion der verwendeten Wörter von Pre nach Post als signifikant (p = 0,037). Nach dem Seminar nannten die Teilnehmenden zwar weniger Schlüsselwörter, auch sank die Zahl der Fehler, jedoch erwiesen sich diese Änderungen ebenso wenig wie die von Quotient und Anzahl der genannten Differentialdiagnosen als signifikant. Die Ergebnisse der Fallbearbeitung von den Teilnehmenden des zweiten Tertials betrachtend ist signifikant einzig die Reduktion der verwendeten Wörter nach dem Seminar (p = 0,046). Die Teilnehmenden des zweiten Tertials nannten in der Fallbearbeitung nach dem Seminar geringfügig weniger Schlüsselwörter und auch die Zahl der Fehler sank. Die Zahl der genannten Differentialdiagnosen lag vor dem Seminar unter 3,5 und nach dem Seminar über 4, auch dieser Zuwachs ist nicht signifikant. In den unteren beiden Zeilen sind die Ergebnisse der Fallbearbeitung des dritten Tertials dargestellt. Bei dieser kleinsten Gruppe erwies sich keine Differenz zwischen Pre- und Post-Bearbeitung des Falls als signifikant. Im Kruskal-Wallis-Test zeigten sich im Vergleich der Tertiale für das Delta, also die Differenz zwischen Pre- und Post-Bearbeitung des Falls, keine signifikanten Unterschiede.

**Tabelle 15: Bearbeitung des Papierfalls von den Tertialen im Vergleich**

	Fall	WZ	SW	Quotient (SW/WZ)	Fehler	DD
<b>Tertial 1</b> n = 21 (MW ± SD)	<b>Pre</b>	39,62 ± 8,81	5,36 ± 1,23	0,14 ± 0,05	0,48 ± 0,72	3,33 ± 1,28
	<b>Post</b>	33,29 <sup>#</sup> ± 10,53	5,17 ± 1,73	0,16 ± 0,06	0,24 ± 0,52	3,67 ± 1,32
<b>Tertial 2</b> n = 13 (MW ± SD)	<b>Pre</b>	34,31 ± 8,53	5,12 ± 1,34	0,15 ± 0,03	0,38 ± 0,55	3,46 ± 1,13
	<b>Post</b>	28,15 <sup>#</sup> ± 9,43	4,96 ± 1,71	0,18 ± 0,06	0,35 ± 0,55	4,23 ± 1,17
<b>Tertial 3</b> n = 8 (MW ± SD)	<b>Pre</b>	38,00 ± 10,21	5,69 ± 1,10	0,16 ± 0,05	0,31 ± 0,59	3,63 ± 1,19
	<b>Post</b>	33,00 ± 7,31	5,88 ± 1,36	0,19 ± 0,06	0,13 ± 0,35	3,38 ± 1,06

MW und SD als Ergebnisse der Bearbeitung des Papierfalls nach Tertialen getrennt; signifikante Werte: Wörterzahl Post < Pre (nur Tertial 1 und 2): #: p < 0,05.

#### 4.4 Berechnung von Korrelationen

Es wurden folgende Korrelationen mittels Spearman-Rho jeweils für den Pre- und den Post-Test berechnet: Zwischen dem Item 1 des Fragebogens zur Selbsteinschätzung, nämlich „Ich bin in der Lage, einen komplexen Patientenfall in zwei bis drei Sätzen zusammenzufassen.“, und der Wörterzahl (Pre: p = 0,239; Post: p = 0,479) konnten keine signifikanten Korrelationen gezeigt werden, ebenso wenig zwischen dem genannten Item 1 und dem Quotienten der Bearbeitung des Papierfalls (Pre: p = 0,415; Post: p = 0,555). Auch zwischen dem Item 2 des Fragebogens zur Selbsteinschätzung, nämlich „Ich bin in der Lage, für ein individuelles Erkrankungsbild mehrere Differentialdiagnosen zu benennen.“, und der Anzahl der im Zuge der Bearbeitung des Papierfalls genannten Differentialdiagnosen zeigten sich keine signifikanten Werte (Pre: p = 0,862; Post: p = 0,184). Eine aussagekräftige Korrelation ließ sich also weder für die vor noch für die nach dem Seminar erhobenen Daten zeigen.

## 5 Diskussion

Die klinische Entscheidungsfindung umfasst Aspekte wie Datenerhebung, Problempäsentation, Hypothesenbildung, Überprüfung der Hypothesen, Testtheorie, Mustererkennung, ärztliche Denkfehler und Umgang mit Ungewissheit. Sie bildet den Kern der ärztlichen Tätigkeit, so dass der Grad der entsprechenden Qualifikationen wesentlichen Anteil am Gelingen der ärztlichen Intervention hat (Kassirer et al. 2010). Entsprechend bedeutsam ist die Vermittlung dieser Qualifikationen für die Ausbildung junger Mediziner und Medizinerinnen, ohne dass sich dies in den Curricula der medizinischen Ausbildung an deutschen Hochschulen spiegeln würde. Wie diese Vermittlung aussehen kann, wurde beispielhaft an der Entwicklung und Durchführung des Seminars „Klinische Entscheidungsfindung für Studierende im Praktischen Jahr“ erprobt und im Rahmen der vorliegenden Studie überprüft: Demzufolge verbesserten PJ-Studierende ihre Qualifikationen in diesem Bereich durch das achtwöchige Seminar mit lernzielbasierten Inhalten und den beschriebenen Vermittlungsformen, zu denen wesentlich die Arbeit mit Papierfällen gehörte.

### 5.1 Fragebogen zur Selbsteinschätzung

#### 5.1.1 Gesamtheit der Teilnehmenden

In der Selbsteinschätzung zeigte sich von der Pre-Befragung zur Post-Befragung ebenso wie von der Retropre- zur Post-Befragung für die Gesamtheit der Studienteilnehmer und -teilnehmerinnen ein signifikanter Zuwachs im Hinblick auf ihre Qualifikationen zu den acht bearbeiteten Aspekten der klinischen Entscheidungsfindung.

Item 1, „... einen komplexen Patientenfall in zwei bis drei Sätzen zusammenzufassen“, ist das Item, das in der Gesamtgruppe vor und nach dem Seminar den höchsten Punktwert überhaupt aufwies. Obwohl die Studierenden also einschätzten, dass sie mit den bisher im Studium erlebten Unterrichtsveranstaltungen die Patientenvorstellung gut beherrschten, beurteilten sie sich nach dem Kurs bezüglich der Patientenvorstellung, die in der Visite im Sinne einer akkuraten Problempäsentation eine der wichtigsten Fähigkeiten für den Berufsanfänger bzw. die -anfängerin ist (Williams und Surakanti 2016), signifikant besser. Insofern erwies es sich trotz der im Verhältnis zu anderen Aspekten der klinischen Entscheidungsfindung bei Seminarbeginn besseren Selbstbeurteilung für einen erfolgreichen Start in das Berufsleben auch in dem fortgeschrittenen Stadium der Ausbildung als sinnvoll, die akkurate Fallpräsentation explizit zum Gegenstand zu machen und wie beschrieben zu üben.

Der geringste Zuwachs in der Selbsteinschätzung zeigte sich bei Item 5 „Ich bin in der Lage, medizinische Testergebnisse im Hinblick auf ihre Relevanz für den individuellen Patientenfall zu bewerten.“. Bei diesem Item fand sich jedoch auch der größte Unterschied zwischen Pre- und Retropre-Selbsteinschätzung. Das Studium, zumindest der Regelstudiengang am UKE, in dem die Teilnehmenden des Seminars „Klinische Entscheidungsfindung“ noch ausgebildet wurden, scheint falsche Sicherheit in der Bewertung von Testergebnissen gegeben zu haben. In einer von Harendza et al. durchgeführten Studie, in der Studierende gegen Ende ihrer Ausbildung in nachgestellten Behandlungssituationen die Anamnese erheben und die weiteren diagnostischen Schritte durchführen bzw. veranlassen mussten, ordneten deutsche PJler und PJlerinnen wesentlich mehr Labor- und Röntgen-Untersuchungen an als ihre niederländischen Kommilitonen und Kommilitoninnen, deren Ausbildung auf der Basis eines sogenannten integrierten Curriculums stattfand, wobei die niederländischen Studierenden gleich gut waren im Stellen der korrekten Diagnose (2013). Die Weiterentwicklung von Qualifikationen zu diagnostischen Verfahren und Tests dient einerseits der Bildung des Bewusstseins für die Problematik der Überdiagnostik und damit deren Vermeidung und andererseits der Patientensicherheit. Ähnliches konnte für das Verschreiben von Medikamenten durch Studierende gezeigt werden (Celebi et al. 2009).

Der einzige Aspekt, bei dem die Studierenden ihre Fähigkeit retropre, wenn auch nur geringfügig, besser einschätzten als vor dem Seminar, ist Item 6 „Ich bin in der Lage, typische Patientenpräsentationen unmittelbar einer Diagnose zuzuordnen.“, das in der Retropre-Befragung den höchsten Mittelwert aufwies. Bei dieser Art der Zuordnung von Diagnosen handelt es sich um Mustererkennung (Croskerry 2009). Diese wird im Studium eingeübt in Form von Multiple-Choice-Fragen, die mit Mustererkennung arbeiten, sofern sie über eine reine Wissensabfrage hinausgehen (Freiwald et al. 2014). In der Diskussion, wieweit Mustererkennung als intuitiver Prozess der Diagnostik den Experten vorbehalten ist, würde sich damit zeigen, dass sie schon in der Ausbildung vermittelbar ist. Im Seminar „Klinische Entscheidungsfindung“ wurde mit der Thematisierung der Mustererkennung vor dem Hintergrund der dualen Prozesstheorie darüber hinaus ein Bewusstsein für die unterschiedlichen kognitiven Prozesse erarbeitet, die im diagnostischen Vorgang von Bedeutung sind. Erst dieses metakognitive Wissen um intuitives und analytisches Vorgehen ermöglicht es den Studierenden, die verschiedenen Arten des Denkens bewusst miteinander zu verbinden – was nicht nur Grundlage für elementare Debiasing-Strategien ist, sondern wodurch überhaupt die besten Ergebnisse im diagnostischen Prozess erzielt werden (Eva 2005, Ark et al. 2006). Dass die Zuschreibung von Mustern andererseits zu Denkfehlern wie Attribution Bias oder Framing führen kann, war im Kurs ebenfalls Thema (Cooper 2017).

Der größte Zuwachs zeigte sich sowohl von pre als auch von retropre zu post bei Item 7 „Ich bin in der Lage, typische ärztliche Denkfehler zu erkennen und Risikosituationen für deren Entstehung zu identifizieren.“. Denkfehler sind nicht Bestandteil des medizinischen Curriculums der Hamburger Universität, soweit überschaubar auch nicht anderer medizinischer Hochschulen in Deutschland. Die Teilnehmenden des Seminars „Klinische Entscheidungsfindung“ beschäftigten sich also erstmalig explizit und systematisch damit. Das Wissen um Cognitive Biases oder Denkverzerrungen stellt die Grundlage dafür dar, diese zu erkennen und ihnen entgegenzuarbeiten. Dies reduziert Diagnosefehler, wie Reilly et al. (2013) in einer Studie mit Assistenzärzten und -ärztinnen zeigten. Medizinische Fehler gehen zu 10 – 20 % auf diagnostische Fehler zurück (Reilly et al. 2013), mithin trägt die Auseinandersetzung mit ärztlichen Denkfehlern zur Patientensicherheit bei. Bemerkenswert ist, dass sich schon mit wenigen Unterrichtseinheiten signifikante Erfolge im Umgang mit Denkfehlern erreichen lassen. Ein Beitrag zur Reduktion von ärztlichen Denkfehlern scheint im Medizinstudium angebracht, da inzwischen jeder dritte Todesfall eines Patienten – zumindest in den USA – auf medizinische Fehler zurückzuführen ist (Markary und Daniel 2016).

Für Item 8, „die in einer medizinischen Entscheidung enthaltene Ungewissheit zu erkennen und Patienten zu vermitteln“, zeigt sich für die Gesamtgruppe aller Tertiale eine signifikante Veränderung in jedem Vergleich. Insbesondere aber stellt sich die Selbsteinschätzung im Hinblick auf diese Qualifikation retropre signifikant schlechter dar als pre. In ihrer Studie zum Selbstbild, Arztbild und Arztideal unter Medizinstudierenden zeigten Schrauth et al. (2009), dass das Selbstbild der befragten Medizinstudierenden „maßgeblich von Gefühlen der Machtlosigkeit und Unsicherheit geprägt“ war, was in einem Spannungsverhältnis zu ihren das Arztideal prägenden Vorstellungen stand. Auch die real erlebten Ärzte wurden von ihnen als deutlich sicherer wahrgenommen. Die Beschäftigung mit der medizinischen Entscheidungen inhärenten Ungewissheit dürfte bei hohem Selbstanspruch in puncto Sicherheit und bei bestehenden Gefühlen der Machtlosigkeit und Unsicherheit erheblich dazu beitragen, den Fokus von einer vermeintlichen eigenen Schwäche auf die Sache selbst zu legen. Das Thema Ungewissheit könnte auch im Rahmen von Prüfungen Bedeutung finden (Cooke und Lemay 2017), da sich Ärztinnen und Ärzte täglich damit auseinandersetzen müssen.

### **5.1.2 Vergleich Männer und Frauen**

Der Zuwachs an selbsteingeschätzter Qualifikation im Hinblick auf die Bewertung medizinischer Testergebnisse für den individuellen Patientenfall, also Item 5, ist in der vorliegenden Studie abhängig vom Geschlecht. Während die Steigerung von der Pre-Befragung zur Post-Befragung in der Selbsteinschätzung der Frauen signifikant war, bewerteten die Männer ihre

Qualifikation einzig bei diesem Item nach dem Seminar sogar niedriger als vor dem Seminar. Dies könnte verstanden werden als Selbstkorrektur des auch bei Medizinstudierenden vorkommenden Stereotyps „Männer haben mehr Selbstvertrauen in ihr Beherrschen von Technik“ (Martin et al. 2016), sofern Tests und deren Interpretation der Technik zugeordnet werden können. Die Männer hatten in der Pre-Befragung bei diesem Item mit über vier Punkten auf der Likert-Skala ihren höchsten Wert überhaupt, beurteilten sich selbst in der Retropre-Befragung jedoch um einen ganzen Skalen-Punkt niedriger. Dies erwies sich als signifikante Differenz.

Schrauth et al. konnten in ihrer Studie zu Selbstbild, Arztbild und Arztideal (2009) Selbstüberschätzungen von Studierenden zeigen, die sich und erfahrenen Ärzten ebenbürtige Fähigkeiten zuschrieben, wobei sich das Selbstbild der Medizinstudenten positiver ausgeprägt darstellte als das der Medizinstudentinnen. Eine Korrektur der Selbstüberschätzung könnte nach der Auseinandersetzung mit einem Thema in der signifikant schlechteren Selbsteinschätzung retropre gegen pre zum Ausdruck kommen. Tatsächlich wurde gezeigt, dass retrospektive Pre-Selbsteinschätzungen in Studien durchschnittlich schlechter ausfallen als die entsprechenden Pre-Befragungen (Levinson et al. 1990). Der im Hinblick auf die Bewertung medizinischer Testergebnisse beschriebene Unterschied allerdings geht über diesen Befund hinaus. Testverfahren werden im Studium zwar unterrichtet, dies aber meist nicht aus metakognitiver Perspektive mit entsprechend kritischem Impetus. Insofern unterstützte das Seminar „Klinische Entscheidungsfindung“ den kritischen Umgang und die Überprüfung der Selbsteinschätzung und ist mithin ein Instrument zur Reduktion der Vermessenheit als einer Quelle diagnostischer Fehler (Saposnik et al. 2016).

### **5.1.3 Vergleich der Tertiale**

Bei der Fallpräsentation, nämlich Item 1 „einen komplexen Patientenfall in zwei bis drei Sätzen zusammenzufassen“, ist der Zuwachs von der Pre- zur Postbearbeitung für das zweite Tertial im Verhältnis zum ersten Tertial signifikant geringer. Die Teilnehmenden des zweiten Tertials schätzten die eigene Qualifikation im Hinblick auf die Fallpräsentation vor dem Seminar im Verhältnis sehr hoch ein, in der Retropre-Bewertung dagegen wesentlich niedriger. Die Differenz zwischen Retropre- und Post-Selbsteinschätzung zeigte sich entsprechend auch in dieser Gruppe signifikant.

Vergleicht man die Ergebnisse der Teilnehmenden aus den drei unterschiedlichen Tertialen zu Item 5 „medizinische Testergebnisse im Hinblick auf ihre Relevanz für den individuellen Patientenfall zu bewerten“, so sticht Folgendes ins Auge: Der Zuwachs von der Pre- zur

Postbearbeitung war für das zweite Tertial signifikant größer als für das erste und dritte Tertial. Außerdem schätzte sich das zweite Tertial im Gegensatz zu den Teilnehmenden der beiden anderen Tertiale in der Retropre-Befragung besser ein als in der Pre-Befragung.

Zweimal also heben sich die Ergebnisse des zweiten Tertials gegen die der Kommilitonen ab. Wie schon bei der unterschiedlichen Selbsteinschätzung von Männern und Frauen lässt sich auch hier interpretieren, dass eine Selbstüberschätzung, die wesentlich zu Fehlern im diagnostischen Prozess beiträgt (Saposnik et al. 2016), korrigiert wird. Auf die Fallpräsentation, also Item 1, bezogen betraf dies das zweite Tertial. Warum die Selbstbeurteilung in der Pre-Prebefragung bei dieser Gruppe signifikant höher als in der Retropre-Befragung lag, lässt sich nur vermuten. Möglicherweise entsteht nach den ersten vier Monaten Arbeit im PJ der Eindruck, schon einige Erfahrung gesammelt zu haben, und damit ein gewisses Selbstbewusstsein. Gleichzeitig steht die Übernahme der Verantwortung auf Station als Assistenzärzte und -ärztinnen, also die Situation, in der das gewonnene Selbstbewusstsein auf die Prüfung gestellt wird, zu Beginn des zweiten Tertials noch nicht unmittelbar bevor. Auf die Bewertung von Testergebnissen, also das Item 5 bezogen sind die Teilnehmenden des zweiten Tertials in der Pre-Selbsteinschätzung näher an der Retropre-Selbsteinschätzung und wirken von daher realistischer als ihre Kommilitonen und Kommilitoninnen des ersten und dritten Tertials. Dabei ist auch zu bedenken, dass im zweiten Tertial mit einem einzigen Mann neben zwölf Frauen der Anteil der Männer, die sich in diesem Punkt retropre signifikant schlechter einschätzten als pre, am geringsten war.

## **5.2 Papierfall**

In der Bearbeitung des Papierfalls spiegelt sich das Ergebnis des Fragebogens zu Item 1, „einen komplexen Patientenfall in zwei bis drei Sätzen zusammenzufassen“, sofern sich der Blick auf die Wörterzahl und den Quotienten richtet. Auch hier zeigen sich in der Post-Bearbeitung signifikant bessere Ergebnisse, nämlich weniger Wörter und ein höherer Quotient. Bemerkenswert ist jedoch, dass die Zahl der korrekt genannten Schlüsselwörter nach dem Seminar nicht signifikant höher lag. Insofern stellte sich die Fallpräsentation nach dem Seminar zwar prägnanter, aber nicht präziser dar. Während sich die Studierenden im Fragebogen bei den auf die Differentialdiagnosen bezogenen Items 2 und 3 in der Selbsteinschätzung nach dem Seminar signifikant besser als davor beurteilten, zeigte sich in der Bearbeitung des Falls keine signifikante Zunahme der Anzahl der Differentialdiagnosen. Eine lange Liste möglicher Diagnosen ist nicht in jedem Fall ein Kriterium für starkes Clinical Reasoning (Bowen 2006), eine Reihe diagnostischer Hypothesen aufzustellen, ist jedoch die Voraussetzung dafür, auf analytischem Weg zu einer Arbeitshypothese zu gelangen. Im



Seminar wurde dies mittels unterschiedlicher Strategien geübt, zu denen auch Mnemotechniken gehörten. Eventuell konnten diese Strategien bei der Überprüfung im Rahmen der Fallbearbeitung nach dem Seminar nicht angewendet werden. Eine mögliche Erklärung dafür wäre, dass Debiasing-Strategien (Croskerry 2017) über einen längeren Zeitraum eingeübt werden müssten.

### **5.3 Korrelationen**

Es zeigten sich keine Korrelationen, weder zwischen den Ergebnissen der Selbsteinschätzung zu Item 1, also der Fallpräsentation und den Ergebnissen der Fallbearbeitung hinsichtlich der Wörterzahl einerseits und dem Quotienten andererseits, noch zwischen den Ergebnissen der Selbsteinschätzung zu Item 2, also der Benennung von Differentialdiagnosen und der Anzahl der bei der Fallbearbeitung genannten Diagnosen. Dies überrascht, da davon auszugehen wäre, dass ein signifikanter Zuwachs an Qualifikation in beiden Formaten in einem Zusammenhang stünde. Bei guter Erprobung des Fragebogens zur Selbsteinschätzung als Instrument und bei guten bis exzellenten Werten für die interne Konsistenz desselben besteht kein Anlass, die Ergebnisse des Selbsteinschätzungsbogens in Frage zu stellen. Es scheint wahrscheinlicher, dass die Fallbearbeitung als Instrument besser ausgearbeitet werden könnte. Dafür spricht, dass bei der Fallbearbeitung das Gewichten der Diagnosen als eine Unteraufgabe von einem Teil der Teilnehmenden nicht oder nur teilweise bearbeitet wurde. Es ist denkbar, dass die Studierenden zum Ende des Seminars den Eindruck hatten, viel gelernt zu haben, dass sie dies aber noch nicht gut anwenden konnten (Eva et al. 2004). Eventuell wurde auch flüchtig mit diesem Punkt der Fallbearbeitung umgegangen, was bedeuten würde, dass die Relevanz der Gewichtung von Differentialdiagnosen, also des Einschätzens von Wahrscheinlichkeiten, nicht vermittelt werden konnte. Diesbezüglich könnte es hilfreich sein, die Aufgaben noch stärker vorzustrukturieren, indem beispielsweise Vorgaben für die Gewichtung zur Auswahl gemacht werden: also ankreuzen zu lassen, ob die Diagnose für sehr wahrscheinlich (90 bis 100 %), wahrscheinlich (80 bis 90 %), möglich usw. gehalten wird. Es bliebe zu fragen, wie Begründungen für Differentialdiagnosen in die Fallarbeit integriert werden könnten.

### **5.4 Metakognition**

Kiesewetter et al. (2013) befassten sich in ihrer Studie mit Mustern kognitiven Problemlösens bei Studierenden der Humanmedizin bezogen auf diagnostische Vorgänge. Sie beschreiben einen Zusammenhang zwischen dem Durchlaufen eines „lower“ und eines „higher loops“, beide zusammen „complete model“ genannt, und der Korrektheit der Diagnose. Dabei scheint entscheidend, dass im Rahmen des „complete model pattern“ der „higher loop“ mit

Evaluation, Repräsentation und Integration höhere kognitive Aktivitäten umfasste, die die Fähigkeit zum abstrakten Denken erfordern. In dieser Studie zeigte sich weder eine Korrelation von Fachwissen mit dem Durchlaufen kompletter kognitiver Muster mit höheren kognitiven Aktivitäten, noch mit der korrekten Diagnose. Es ist also die – auch von Kiese-wetter et al. (2013) gestellte – Frage, in welchem Verhältnis Metakognition zu anderen Formen des Wissens (Faktenwissen, konzeptuelles und prozedurales Wissen) steht. Wenn ihr eine zentrale Bedeutung im ärztlichen Denken zukommt, dann leistet das Seminar „Klinische Entscheidungsfindung“ unabhängig von der Einzel-Item-Ebene einen grundlegenden Beitrag zur Entwicklung ärztlicher Qualifikationen: Es ist im Rahmen des Hamburger Medizinstudiums die einzigartige Situation, dass sich die Studierenden angeleitet und strukturiert auf der Ebene der Metakognition mit Aspekten ihres Fachs befassen. Beim „Denken über das Denken“ verändert sich das Bewusstsein von sich selbst und vom Gegenstand. Gefördert und entwickelt wird damit das abstrakte Denken als eine Fähigkeit, die Voraussetzung dafür ist, innere Repräsentationen auch von Patientenfällen zu bilden. Anders ausgedrückt: Wenn sich Medizinstudierende von der Beschäftigung mit dem Fach- und Faktenwissen lösen und eine Auseinandersetzung damit führen, was sie als angehende Ärzte und Ärztinnen wie tun – auch im Denken –, dann trägt das zur Entwicklung von Fähigkeiten bei, die jenseits des Faktenwissens zentral dafür sind, eine korrekte Diagnose zu stellen – und überhaupt Patienten und Patientinnen zu behandeln.

## **5.5 Limitationen der Studie**

Eine Schwäche der Studie ist die mit 42 Teilnehmenden geringe Fallzahl. Das ist darauf zurückzuführen, dass der Anteil kompletter Datensätze am Rücklauf ausgefüllter Erhebungsbögen mit einem Drittel (32,06 %) gering war. Einerseits hängt dies mit organisatorischen Besonderheiten des Seminars zusammen (überlappende Tertialgruppen von Studierenden konnten das Seminar erst nach der Hälfte der Laufzeit beginnen und damit keinen Pre-Fragebogen und -Fall bearbeiten), andererseits dürfte auch die Belastung der Studierenden während des PJs eine Rolle gespielt haben. Immer wieder wurde berichtet, dass die Teilnahme am Seminar schwierig sei, weil Studierende ihre Kollegen und Kolleginnen mit der Arbeit auf Station nicht allein lassen wollten. Auch Evaluationsmüdigkeit könnte eine Rolle gespielt haben.

Positiv hervorzuheben ist die interne Konsistenz der Fragebögen. Die Werte für Cronbachs Alpha liegen im Bereich zwischen gut und exzellent (George und Mallery 2002). Dass dieses Instrument explizit für die Studie entwickelt wurde, statt ein schon publiziertes zu verwenden,

erschwert allerdings die Vergleichbarkeit der vorliegenden Arbeit mit anderen Studien (Durning et al. 2012). Das Instrument wurde genau auf das Seminar zugeschnitten.

Als Einschränkung muss angeführt werden, dass Aspekte des Clinical Reasoning bearbeitet und in der Selbsteinschätzung wie in der Fallbearbeitung geprüft wurden, diagnostische Genauigkeit wurde jedoch nicht getestet. Dies betrifft die Korrektheit einer Arbeitsdiagnose, aber umfassender auch die Begründung diagnostischer Hypothesen. Darüber hinaus muss hinterfragt werden, ob es als Überprüfung für den Lernerfolg hinreicht, einen einzigen Papierfall bearbeiten zu lassen.

Als weiterer positiver Aspekt kann die Kürze der Intervention bei guten Ergebnissen genannt werden. Bei Be- und Überlastung von Studierenden wie Mitarbeitenden ist eine Ressourcen schonende Maßnahme leichter zu implantieren als eine umfassendere, die sich spiralartig über längere Phasen des Curriculums erstreckt, wofür in anderen Publikationen mit guten Gründen gestritten wird (Cooper et al. 2017). Ähnlich erscheint auch die Integration in unterschiedlichste Kurse im Sinne einer Grundlagenwissenschaft über die gesamte Ausbildung hinweg, wie sie Trowbridge et al. (2015) als eine Möglichkeit thematisieren, eher ein mittel- bis langfristiges Ziel. Ein einzelner Kurs kann von einer Fakultät bzw. von einem Dekanat leichter übernommen werden als ein sich über mehrere Semester erstreckender Studienplan. Entsprechend äußerte die Universität Marburg Interesse an dem Seminar.

Das Seminar wurde nach Abschluss der Studie am UKE fortgeführt. Für den Fall, dass am Ende des Kurses eine Überprüfung des Erfolgs durchgeführt werden soll, könnte die Form der Kontrolle überdacht und überarbeitet werden. Es spricht viel dafür, die explizite Auseinandersetzung mit Clinical Reasoning oder Aspekten davon früher im Studienverlauf einzuführen. Dies würde sich leichter ermöglichen lassen, wenn weitere Seminarkonzepte von begrenztem und überschaubarem Umfang entwickelt und überprüft würden.

## 6 Zusammenfassung

Die klinische Entscheidungsfindung stellt im ärztlichen Alltag eine grundlegende Kompetenz dar. Sie umfasst neben den kognitiven Prozessen bei der Bildung einer Arbeitsdiagnose und der Auseinandersetzung mit therapeutischen Entscheidungen auch die Ebene der Metakognition beispielsweise in Hinsicht auf das Erkennen von Denkfehlern. Dieses Clinical Reasoning ist jedoch weder Teil der Approbationsordnung, noch findet es in Deutschland explizit Berücksichtigung in den medizinischen Curricula. Ziel des Projekts war daher die Konzeption und Etablierung eines Seminars zur klinischen Entscheidungsfindung im PJ mit einer lernzielbasierten Kontrolle des Erfolgs. Es wurden sechs Seminare durchläufe aus den Jahren 2011 bis 2013 mit je acht doppelstündigen Terminen in die Auswertung aufgenommen. Zu Beginn des ersten und am Ende des letzten Seminartermins füllten die PJ-Studierenden im Tertiärs Innere Medizin einen Selbsteinschätzungsbogen zu acht Aspekten der klinischen Entscheidungsfindung mit einer sechsstufigen Likert-Skala aus. Zu denselben Zeitpunkten erhielten sie einen identischen Patientenfall, der im Hinblick auf Problempäsentation und Hypothesenbildung zu bearbeiten war. In die Auswertung des Seminars gingen 42 komplette Datensätze ein. Nach dem Kurs schätzte sich die Gesamtheit der Teilnehmenden in allen acht Aspekten der klinischen Entscheidungsfindung signifikant besser ein als vor dem Kurs. Der größte Zuwachs zeigte sich beim Item „Ich bin in der Lage, typische ärztliche Denkfehler zu erkennen.“. Frauen und Männer unterschieden sich signifikant in der Bewertung von Testergebnissen in Bezug auf einen individuellen Patientenfall im Hinblick auf die Differenz zwischen Pre- und Retropre- wie auch zwischen Pre- und Post-Beurteilung bei einer hohen Selbsteinschätzung der Männer vor dem Kurs. Für Studierende des zweiten Tertiärs ergab sich in der Beurteilung ihrer Fähigkeit, einen Patientenfall zusammenzufassen, ein signifikant geringerer Zuwachs als für das erste Tertiär und hinsichtlich der Bewertung von Testergebnissen ein signifikant größerer Zuwachs als für das erste und dritte Tertiär. Nach dem Kurs bewerteten sich Studierende des ersten Tertiärs signifikant schlechter als Studierende der beiden anderen Tertiäre für die Gewichtung und Begründung von Differentialdiagnosen und für die Bewertung von Testergebnissen bzw. für die Anordnung von diagnostischen Verfahren signifikant schlechter als das dritte Tertiär. Gemessen an der Zahl der im Verhältnis zur Nennung von Schlüsselbegriffen verwendeten Wörter war die Fallpräsentation nach dem Kurs signifikant fokussierter. Ein Zuwachs an Differentialdiagnosen fand sich nicht. Der Kurs stellt sich der Studie zufolge als sinnvolle und effektive Intervention zum Üben klinischer Entscheidungsfindung und zum Arbeiten auf der Ebene der Metakognition dar. Da bei der Kontrolle noch nicht alle Lernziele erreicht wurden, lässt sich der Kurs auf der Basis der Erkenntnisse dieser Arbeit weiterentwickeln.

## Summary

Clinical reasoning represents a basic competence in a doctor's everyday work. In addition to the cognitive processes involved in arriving at a working diagnosis and to the therapeutic decision-making, it also covers the meta-cognitive level, e.g. with regard to identifying cognitive biases in medical reasoning. However, clinical reasoning is neither included in the medical licensure regulations, nor is it explicitly considered in medical curricula in Germany. Thus the goal of the project was to establish a seminar in clinical reasoning in the practical year (PY), including success monitoring based on the learning goals. Six seminar blocks, featuring eight double hours in each case, were evaluated in the years 2011 to 2013. Before the first and after the last class of each seminar, the PY-students filled out a self-assessment form relating to eight aspects of clinical reasoning, with a six-point Likert scale. At the same times, they were given an identical patient case to work on with regard to presentation of the problem and to forming a hypothesis. 42 complete were included in the analysis of the seminar. After the course, the majority of the participants assessed themselves as being significantly better than before the course in all eight aspects. The greatest increase was found for the statement "I am able to identify typical errors in medical reasoning". Women and men differed significantly in their assessment of test results with respect to an individual patient case with regard to the difference between pre- and retrospective pre-assessment as well as between pre- and post-assessment while men showed a high level of self-assessment before the course. For students in the second third of the PY, there was a significantly smaller increase in the assessment of their own ability to summarise a patient's case compared to students in the first third, and a significantly larger increase regarding the evaluation of test results compared to students in the first and last thirds of the year. In the self-assessment after the course, the students in the first third of the PY assessed themselves as being significantly worse than those in the second and last third with respect to assessing and justifying differential diagnoses, and as significantly worse than students in the last third with respect to the evaluation of test results and the ordering of diagnostic tests. As measured by the number of words used in relation to the naming of key terms, the presentation of the patient's problem was significantly better focused after the course. There was no increase in the differential diagnoses after the course. According to the study, the course is a useful and effective intervention for practising clinical reasoning, and also for working on the meta-cognitive level. However, as assessment showed that not all learning goals were reached, the course can be developed further based on the findings of the study.

## 7 Abkürzungsverzeichnis

DD	Differentialdiagnose
EKG	Elektrokardiogramm
G	Geschlecht
m	Männer
MW	Mittelwert
n	Fallzahl
PJ	Praktisches Jahr
PJler	Medizinstudent im Praktischen Jahr
PJlerin	Medizinstudentin im Praktischen Jahr
SD	Standardabweichung
SW	Schlüsselwörter
UKE	Universitätsklinikum Hamburg-Eppendorf
w	Frauen
WZ	Wörterzahl

## 8 Literaturverzeichnis

- AOK Bundesverband. **Krankenhaus-Report 2014: Wege zu mehr Patientensicherheit.** Pressekonferenz/Pressemappe. Online im Internet: URL [http://aok-bv.de/imperia/md/aokbv/presse/pressemitteilungen/archiv/2014/krankenhaus\\_report\\_2014\\_pressemappe\\_210114.pdf](http://aok-bv.de/imperia/md/aokbv/presse/pressemitteilungen/archiv/2014/krankenhaus_report_2014_pressemappe_210114.pdf) (abgerufen am 22.06.2014)
- Ark TK, Brooks LR, Eva KW. **Giving learners the best of both worlds: do clinical teachers need to guard against teaching pattern recognition to novices?** *Acad Med.* 2006; 81(4): 405-409.
- Barre K. **Evidence-based Nursing in der pflegedidaktischen Vermittlung.** 2014. *Frankfurt am Main. Mabuse-Verlag.*
- Beullens J, Struyf E, van Damme B. **Diagnostic ability in relation to clinical seminars and extended-matching questions examinations.** *Med Educ.* 2006; 40(12): 1173-1179
- Bordage G. **Elaborated knowledge: a key to successful diagnostic thinking.** *Acad Med.* 1994; 69(11): 883-885
- Bordage G, Lemieux M. **Semantic structures an diagnostic thinking of experts and novices.** *Acad Med.* 1991; 66(9 Suppl): 70-72.
- Bordage G, Grant J, Marsden P. 1990. **Quantitative assessment of diagnostic ability.** *Med Educ.* 24(5): 413-425.
- Bowen JL. **Educational strategies to promote clinical diagnostic reasoning.** *N Engl J Med.* 2006; 355(21):2217-2225.
- Celebi N, Weyrich P, Riessen R, Kirchhoff K, Lammerding-Köppel M. **Problem-based training for medical students reduces common prescription errors: a randomised controlled trial.** *Med Educ.* 2009; 43(10):1010-1018.
- Chamberland M; St-Onge C, Setrakian J, Lanthier L, Bergeron L, Bourget A, Mamede S, Schmidt H, Rikers R. **The influence of medical students' self-explanations on diagnostic performance.** *Med Educ.* 2011; 45(7): 688-695.
- Cooke S, Lemay JF. **Transforming medical assessment: integrating uncertainty into the evaluation of clinical reasoning in medical education.** *Acad Med.* 2017. [Epub ahead of print].
- Cooper N. **Cognitive Biases.** In: Cooper N und Frain J (Hrsg.). *ABC of Clinical Reasoning.* 2017. *Chichester, West Sussex. John Wiley and Sons.* 22-26.

- Cooper N, Frain J. **Clinical Reasoning: An Overview**. In: Cooper N und Frain J (Hrsg.). ABC of Clinical Reasoning. 2017. *Chichester, West Sussex. John Wiley and Sons*. 1-5.
- Cooper N, Da Silva AL, Powell S. **Teaching Clinical Reasoning**. In: Cooper N und Frain J (Hrsg.). ABC of Clinical Reasoning. 2017. *Chichester, West Sussex. John Wiley and Sons*. 44-50.
- Croskerry P. **Metacognition and Cognitive Debiasing**. In: Cooper N und Frain J (Hrsg.). ABC of Clinical Reasoning. 2017. *Chichester, West Sussex. John Wiley and Sons*. 33-38.
- Croskerry P. **Bias: a normal operating characteristic of the diagnosing brain**. *Diagnosis*. 2014; 1(1): 23-27.
- Croskerry P. **A universal model of diagnostic reasoning**. *Acad Med*. 2009; 84(8): 1022-1028.
- Croskerry P. **The importance of cognitive errors in diagnosis and strategies to minimize them**. *Acad Med*. 2003; 78(8): 775-780
- Dewey J. **How we think**. 1910. *Lexington, Mass. D.C. Heath & Company*.
- Dewey J. **Wie wir denken**. 2002. *Zürich. Verlag Pestalozzianum*.
- Durning S, Artino AR Jr, Pangaro L, van der Vleuten CP, Schuwirth, L. **Context and clinical reasoning: Understanding the perspective of the expert's voice**. *Med Educ*. 2011; 45(9): 927-938.
- Durning SJ, Artino A, Boulet J, La Rochelle J, Van der Vleuten C, Arze B, Schuwirth L. **The feasibility, reliability, and validity of a post-encounter form for evaluating clinical reasoning**. *Med Teach*. 2012; 34(1): 30 -37.
- Elstein AS, Schwartz A. **Clinical problem solving and diagnostic decision making: selective review of the cognitive literature**. *BMJ*. 2002; 324(7339): 729-732.
- Elstein AS, Shulman LS, Sprafka SA. **Medical Problem Solving: An Analysis of Clinical Reasoning**. 1978. *Cambridge, MA. Harvard University Press*.
- Eva KW. **What every teacher needs to know about clinical reasoning**. *Med Educ*. 2005; 39(1): 98-106.
- Eva KW, Cunnington JP, Reiter HI, Keane DR, Norman GR. **How can I know what I don't know? Poor self assessment in a well-defined domain**. *Adv Health Sci Educ Theory Pract*. 2004, 9(3): 211-224.



- Freiwald T, Salimi M, Khaljani E, Harendza S. **Pattern recognition as a concept for multiple-choice questions in a national licensing exam.** BMC Med Educ. 2014, 14: 232 -232.
- George D, Mallery:P **SPSS for Windows Step by Step: A Simple Guide and Reference, 11.0 Update.** 2002. *Boston. Allyn & Bacon.*
- Glaser R. **Education and Thinking: The Role of Knowledge.** 1983. *Pittsburgh, PA. University of Pittsburgh Learning Research and Development Center.*
- Goldszmidt M, Minda JP, Bordage G. **Developing a unified list of physicians' reasoning tasks during clinical encounters.** Acad Med. 2013; 88(3): 390-397.
- Graber ML, Franklin N, Gordon R. **Diagnostic error in internal medicine.** Arch Intern Med 2005; 165(13): 1493-1499.
- Harasym HP, Tsai TC, Hemmati P. **Current trends in developing medical students' critical thinking abilities.** Kaohsiung J Med Sci. 2008; 24(7): 341-355.
- Harendza S, Alofs L, Huiskes J, Wijnen-Meijer M. **Ordering patterns for laboratory and radiology tests by students from different undergraduate medical curricula.** BMC Med Educ. 2013;13:109.
- Higgs J, Jones MA, Loftus S, Christensen N (Hrsg.). **Clinical Reasoning in the Health Professions.** 2008. *Amsterdam [u.a.]. Elsevier [u.a.].*
- Hudzik B, Hudzig M, Polonski L. **Choosing wisely: avoiding too much medicine.** Can Fam Physician. 2014. 60(10):873-876.
- Kahneman D. **Schnelles Denken, langsames Denken.** 2012. *München. Siedler.*
- Kaiser FJ. **Grundlagen der Fallstudiendidaktik – Historische Entwicklung- Theoretische Grundlagen – Unterrichtliche Praxis.** In: Kaiser, FJ (Hrsg.). Die Fallstudie – Theorie und Praxis der Fallstudiendidaktik. 1983. *Bad Heilbrunn. Verlag Julius Klinkhardt.* 9-34.
- Kassirer JP. **Teaching clinical reasoning: case-based and coached.** Acad Med. 2010; 85(7): 1118-1124.
- Kassirer JP, Wong JB, Kopelman RI. **Learning Clinical Reasoning.** 2010. *Baltimore, MD. Lippincott Williams & Wilkins Health.*
- Kiesewetter J, Ebersbach R, Görlitz A, Holzer M, Fischer MR, Schmidmaier, R. **Cognitive problem solving patterns of medical students correlate with success in diagnostic case solutions.** PLoS One. 2013; 8(8): e71486.
- Klemme B, Siegmann, G. **Clinical Reasoning: Therapeutische Denkprozesse lernen.** 2015. *Stuttgart. Thieme.*

- Levinson W, Gordon G, Skeff K. **Retrospective versus actual pre-course self-assessments.** *Eval and the health prof.* 1990; 13(4): 445-452.
- Lyman GH, Balducci L. **The effect of changing disease risk on clinical reasoning.** *J Gen Intern Med.* 1994; 9(9): 488-95.
- Makary MA, Daniel M. **Medical error-the third leading cause of death in the US.** *BMJ.* 2016; 353: 2139-2139.
- Mamede S, van Gog T, Moura AS, de Faria RMD, Peixoto JM, Rikers RMJP, Schmidt HG. **Reflection as a strategy to foster medical students' acquisition of diagnostic competence.** *Med Educ.* 2012; 46: 464-467.
- Mamede S, Schmidt HG, Rikers R. **Diagnostic errors and reflective practice in medicine.** *J Eval Clin Pract.* 2007; 13(1): 138-145.
- Martin AN, Hu Y, Le IA, Brooks KD, Mahmutovic A, Choi J, Kim H, Rasmussen SK, **Predicting surgical skill acquisition in preclinical medical students.** *Am J Surg.* 2016; 212(4): 596-601.
- Miller GA. **The magical number seven, plus or minus two: some limits on our capacity for processing information.** *Psychol Rev.* 1956; 63(2): 81-89.
- Nendaz MR, Raetzo MA, Junod AF, Vu NV. **Teaching diagnostic skills: clinical vignettes or chief complaints?** *Adv Health Sci Educ Theory Pract.* 2000; 5(1): 3-10.
- Nikendei C, Mennin S, Weyrich P, Kraus B, Zipfel S, Schrauth M, Jünger J. **Effects of a supplementary final year curriculum on students' clinical reasoning skills as assessed by key-feature examination.** *Med Teach.* 2009; 31(9): e438-e442.
- Norman GR, Eva KW. **Diagnostic error and clinical reasoning.** *Med Educ.* 2010; 44(1): 94-100.
- Patel VL, Arocha JF, Zhang J. **Thinking and Reasoning in Medicine.** In: Holyoak KJ und Morrison RG (Hrsg.). *The Cambridge handbook of thinking and reasoning.* 2005. *Cambridge. University Press.* 727-750.
- Raupach T, Andresen JC, Meyer K, Strobel L, Koziolok M, Jung W, Brown J, Anders S **Test-enhanced learning of clinical reasoning: a crossover randomised trial.** *Med Educ.* 2016; 50(7): 711-720.
- Reilly JB, Ogdie AR, Von Feldt JM, Myers JS. **Teaching about how doctors think: a longitudinal curriculum in cognitive bias and diagnostic error for residents.** *BMJ Qual Saf.* 2013; 22(12): 1044-1050.

- Rogers JC, Swee DE, Ullian JA. **Teaching medical decision making and students' clinical problem-solving skills.** Med Teach. 1991; 13(2): 157-164.
- Rylander M, Guerrasio J. **Heuristic errors in clinical reasoning.** Clin Teach. 2016; 13(4): 287-290.
- Saposnik G, Redelmeier D, Ruff CC, Tobler PN. **Cognitive biases associated with medical decisions: a systematic review.** BMC Med Inform Decis Mak. 2016; 16(1):138-138.
- Schmidt HG, Mamede S. **How to improve the teaching of clinical reasoning: a narrative review and a proposal.** Med Educ. 2015; 49(10): 961-973.
- Schrauth M, Kowalski A, Weyrich P, Begenau J, Werner A, Zipfel S, Nikendei C. **Selbstbild, Arztbild und Arztideal: Ein Vergleich Medizinstudierender 1981 und 2006.** Psychother Psych Med. 2009; 59(12): 446-453.
- Sox HC, Higgins MC, Owens DK. **Medical Decision Making.** 2013. *Chichester, West Sussex. John Wiley and Sons.*
- Stieger S, Praschinger A, Kletter K, Kainberger F. **Diagnostic grand rounds: A new teaching concept to train diagnostic reasoning.** Eur J Radiol. 2011; 78(3): 349-352.
- Trowbridge RL, Rencic JJ, Durning SJ. **Teaching Clinical Reasoning.** 2015. *Philadelphia, PA. American College of Physicians.* 31-50.
- Tversky A, Kahneman D. **Judgement under uncertainty: heuristics and biases.** Science. 1974; 185: 1124-1131.
- Williams DE, Surakanti S. **Developing oral case presentation skills: peer and self-evaluations as instructional tools.** Ochsner J. 2016; 16(1): 65-69.

## 9 Anhang

### 9.1 Lerneinheiten und Lernziele für das PJ-Seminar „Klinische Entscheidungsfindung“

#### Lerneinheit 1 „Datenerhebung, akkurate/abstrahierende Problempräsentation“

##### **Lernzieldefinition „Datenerhebung“:**

Der/die Studierende ist in der Lage,

- alle möglichen Quellen der Information in einem speziellen Fallbeispiel zu erkennen und zu benennen.
- den verfügbaren Quellen die relevanten Informationen zu entnehmen und sie neu formuliert schriftlich wiederzugeben.
- seinen/ihren subjektiven Eindruck vom Patienten/von der Patientin und dessen/deren Umgebung wahrzunehmen und zu beschreiben.

##### **Lernzieldefinition „akkurate/abstrahierende Problempräsentation“:**

Der/die Studierende ist in der Lage,

- aus den gesammelten Informationen eine für den individuellen Fall zutreffende, abstrahierende und auf allgemeine, objektive Kriterien zurückgreifende Vorstellung in zwei bis drei Sätzen zu formulieren.
- die Limitationen des Kurzzeitgedächtnisses zu erklären und Informationen unter deren Berücksichtigung zu präsentieren.

#### Lerneinheit 2 „Verdachtsfundierte Begründungen“

##### **Lernzieldefinition „Hypothesen“:**

Der/die Studierende ist in der Lage,

- individuelle Fälle dem Formenkreis einer Gruppe von Erkrankungen zuzuordnen (TANTE PIAV) und Differentialdiagnosen unter Einbeziehung von Symptomen, Untersuchungsergebnissen und anderen führenden Schlüsselinformationen kontextgerecht zu begründen.
- zu beschreiben, dass der Kontext die Hypothesengeneration positiv oder negativ beeinflussen kann, und diese Effekte in Betracht zu ziehen und zu diskutieren.

#### Lerneinheit 3 „Diagnosen auf dem Prüfstand und Testtheorie“

##### **Lernzieldefinition „Überprüfung der Hypothesen“:**

Der/die Studierende ist in der Lage,

- Hypothesen-basiert weitere diagnostische Verfahren in sinnvoller und kosteneffizienter Weise anzuordnen und aufgrund der Ergebnisse konkurrierende Verdachtsdiagnosen zu bearbeiten und einzugrenzen.
- geeignete diagnostische Testverfahren zu benennen und nach ihrer Relevanz sowie dem Grad der Belastung des Patienten (Invasivität) im aktuellen Fallbeispiel zu ordnen.
- für die von ihm vorgeschlagenen Testverfahren einen groben Anhalt für die Genauigkeit (Sensitivität und Spezifität) zu benennen.

#### **Lerneinheit 4 „Pattern Recognition“**

##### **Lernzieldefinition „Pattern Recognition“:**

Der/die Studierende ist in der Lage,

- den Diagnosestellungsprozess Mustererkennung zu definieren und an Fallbeispielen zu erklären.
- die Vor- und Nachteile der Mustererkennung und die Benutzung von Faustregeln zu benennen.

#### **Lerneinheit 5 „Ärztliche Denkfehler“**

##### **Lernzieldefinition „Ärztliche Denkfehler“:**

Der/die Studierende ist in der Lage,

- typische Ursachen von ärztlichen Denkfehlern zu benennen.
- anhand von Fallberichten typische ärztliche Denkfehler zu erkennen und zu erklären.

#### **Lerneinheit 6 „Umgang mit Ungewissheit“**

##### **Lernzieldefinition „Umgang mit Ungewissheit“:**

Der/die Studierende ist in der Lage,

- Ungewissheit von unzureichendem Wissen zu unterscheiden.
- die vielfältigen Erscheinungsformen der Ungewissheit in der Medizin zu erkennen.
- die verschiedenen Erscheinungsformen der Ungewissheit im diagnostischen Prozess zu definieren und an Beispielen zu erklären.
- die in einer medizinischen Entscheidung enthaltene Ungewissheit dem Patienten/der Patientin zu vermitteln.

## 9.2 Fragebogen zur Selbsteinschätzung



Universitätsklinikum  
Hamburg-Eppendorf

### Kompetenzfragebogen

	die ersten drei Buchstaben des Vornamens der Mutter (z. B. Julia, also JUL)	die ersten drei Buchstaben des Vornamens des Vaters (z. B. Peter, also PET)	die ersten zwei Ziffern des eigenen Geburtstags (z. B. 15.03.1980, also 15)
Code zur anonymen Zuordnung			

**Bitte geben Sie an, inwiefern die jeweilige Aussage für Sie zutrifft.  
Bitte malen Sie das entsprechende Oval mit einem dunklen Stift vollständig aus!**

Bitte so: ●

Ich bin im _____ Tertial.	0 männlich	0 weiblich
---------------------------	---------------	---------------

Ich bin in der Lage, ...	Gar nicht	sehr gut
... einen komplexen Patientenfall in zwei Sätzen wie für eine Visite zusammenzufassen.	0 0 0 0 0 0	
... für ein individuelles Erkrankungsbild unterschiedliche Differentialdiagnosen zu benennen.	0 0 0 0 0 0	
... diese Differentialdiagnosen aus dem Gesamtkontext zu begründen und zu gewichten.	0 0 0 0 0 0	
... Hypothesen-basiert weitere diagnostische Verfahren in geeigneter Weise anzuordnen.	0 0 0 0 0 0	
... medizinische Testergebnisse im Hinblick auf ihre Relevanz für den individuellen Patientenfall zu bewerten.	0 0 0 0 0 0	
... typische Patientenpräsentationen unmittelbar einer Diagnose zuzuordnen.	0 0 0 0 0 0	
... typische ärztliche Denkfehler bei mir zu erkennen und Risikosituationen für deren Entstehen zu identifizieren.	0 0 0 0 0 0	
... die in einer medizinischen Entscheidung enthaltene Ungewissheit zu erkennen und dem Patienten zu vermitteln.	0 0 0 0 0 0	



## Kompetenzfragebogen

	die ersten drei Buchstaben des Vornamens der Mutter (z. B. Julia, also JUL)	die ersten drei Buchstaben des Vornamens des Vaters (z. B. Peter, also PET)	die ersten zwei Ziffern des eigenen Geburtstags (z. B. 15.03.1980, also 15)
Code zur anonymen Zuordnung			

**Bitte geben Sie an, inwiefern die jeweilige Aussage für Sie zutrifft.  
Bitte malen Sie das entsprechende Oval mit einem dunklen Stift vollständig aus!**

So: ●

Nicht so: ○ ◐ ◑

Ich bin	0 männlich	0 weiblich
---------	---------------	---------------

Nachdem ich am Seminar „Clinical Reasoning“ teilgenommen habe, schätze ich nun meine Kompetenzen „retrospektiv vor“ und nach dem Seminar wie folgt ein:

		gar nicht					sehr gut
Ich bin in der Lage, einen komplexen Patientenfall in zwei Sätzen wie für eine Visite zusammenzufassen	vor	0	0	0	0	0	0
	nach	0	0	0	0	0	0
Ich bin in der Lage, für ein individuelles Erkrankungsbild unterschiedliche Differentialdiagnosen zu benennen.	vor	0	0	0	0	0	0
	nach	0	0	0	0	0	0
Ich bin in der Lage, diese Differentialdiagnosen aus dem Gesamtkontext zu begründen und zu gewichten.	vor	0	0	0	0	0	0
	nach	0	0	0	0	0	0
Ich bin in der Lage, Hypothesen-basiert weitere diagnostische Verfahren in geeigneter Weise anzuordnen.	vor	0	0	0	0	0	0
	nach	0	0	0	0	0	0

VOR und NACH dem Seminar		gar nicht			sehr gut		
Ich bin in der Lage, medizinische Testergebnisse im Hinblick auf ihre Relevanz für den individuellen Patientenfall zu bewerten.	vor	0	0	0	0	0	0
	nach	0	0	0	0	0	0
Ich bin in der Lage, typische Patientenpräsentationen unmittelbar einer Diagnose zuzuordnen.	vor	0	0	0	0	0	0
	nach	0	0	0	0	0	0
Ich bin in der Lage, typische ärztliche Denkfehler bei mir zu erkennen und Risikosituationen für deren Entstehen zu identifizieren.	vor	0	0	0	0	0	0
	nach	0	0	0	0	0	0
Ich bin in der Lage, die in einer medizinischen Entscheidung enthaltene Ungewissheit zu erkennen und dem Patienten zu vermitteln	vor	0	0	0	0	0	0
	nach	0	0	0	0	0	0



### 9.3 Papierfall zur Erhebung des Seminarerfolgs

<b>Kasuistik Nr. 24</b>	
<b>Beschwerdeschilderung durch die Patientin:</b>	
Seit drei Wochen fühle ich mich abgeschlagen. Ich habe häufiger Fieber gemessen und Temperaturen bis 38,5 °C. Nachts schwitze ich sehr stark und muss mich manchmal umziehen. Ich habe jetzt häufiger Kopfschmerzen im Stirnbereich und Schmerzen in Nacken und Schulter.	
<b>Anamnese:</b>	
72-jährige Patientin, vor 4 Wochen einmalig Gastroenteritis über zwei Tage, dann gastrointestinal beschwerdefrei. Keine Auslandsaufenthalte. Belastbarkeit sonst gut. Subjektiv keine Hautveränderungen bemerkt. Vorerkrankungen: Mitralklappenprolaps Malignes Melanom rechtes Auge vor 18 Jahren Gingko-Präparat seit 3 Monaten	
<b>Körperliche Untersuchung:</b>	
Patientin in gutem AZ. 70 kg bei 1,68 cm. RR 152/84 mmHg, Puls 82/min, regelmäßig. SPO2 99 %, Temperatur 37,4 °C. Cor, Pulmo und Abdomen o.p.B. bis auf systolisches Klickgeräusch. Keine Ödeme. Z.n. Enukleation des rechten Auges. Schläfen palpatorisch normal, Haut unauffällig.	

## 9.4 Aufgabenblatt zur Fallbearbeitung

### Fragen zur Kasuistik

	die ersten drei Buchstaben des Vornamens der Mutter (z. B. Julia, also JUL)	die ersten drei Buchstaben des Vornamens des Vaters (z. B. Peter, also PET)	die ersten zwei Ziffern des eigenen Geburtstags(z. B. 15.03.1980, also 15)
Code zur anonymen Zuordnung			

1. Fassen Sie das Wesentliche dieses Falles in 2 Sätzen, wie für eine Visite, zusammen.

2. Nennen Sie eine oder mehrere in Frage kommende Diagnose(n)	3. Bringen Sie Ihre Diagnosen in eine Rangfolge (1., 2., 3.,...)	4. Wie sicher sind Sie sich Ihrer Diagnose 1. ? 1-6 (sicher-unsicher)

(bei mehr Diagnosen bitte die Rückseite des Zettels verwenden)

5. Wie schwierig empfanden Sie die Bearbeitung dieses Falles? Skala 1 – 6 (leicht – schwer)

## 10 Danksagung

Ich danke meiner Doktormutter, Frau Professorin Dr. Sigrid Harendza, dass sie mir diese Arbeit überantwortet und mich bei der Umsetzung in jeder Hinsicht unterstützt hat. Ich danke ihr für ihre nahezu unermüdlich scheinende Geduld mit einer Doktorandin in der Mehrfachbelastung von Studium, Berufstätigkeit und Promotion. Ihre umgehenden Rückmeldungen und ihre anhaltende Ermutigung für meine Arbeit ließen mich den Faden der Auseinandersetzung auch neben dem Einstieg in die Berufstätigkeit als Ärztin in der Hand behalten bzw. immer wieder aufnehmen. Ohne ihre auch in schwierigen Situationen beständig konstruktive Haltung hätte ich diese Arbeit nicht fertiggestellt.

Mit Frau Professorin Dr. Sigrid Harendza möchte ich Herrn Dr. Matthias Janneck, Herrn Dr. Andreas Klinge und Herrn Dr. Ingo Krenz danken, dass sie mir die wissenschaftliche Begleitung und Auswertung des Projekts anvertraut haben. Ich danke ihnen, dass ich am Seminarprojekt Clinical Reasoning am UKE teilhaben konnte, und besonders danke ich ihnen für ihre begeisterte und motivierende Dozententätigkeit in diesem neuen Unterrichtsformat.

Mein Dank gilt auch Frau Anne Daubmann vom Institut für Medizinische Biomedizin und Epidemiologie am UKE für ihre Beratung und Unterstützung bei Fragen zur Statistik.

Frau Angelika Pohl danke ich für das Korrektorat und Herrn Clive Williams für die Übersetzung der Zusammenfassung ins Englische.

## **Eidesstattliche Erklärung**

Ich versichere ausdrücklich, dass ich die Arbeit selbstständig und ohne fremde Hilfe verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die aus den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen einzeln nach Ausgabe (Auflage und Jahr des Erscheinens), Band und Seite des benutzten Werks kenntlich gemacht habe.

Ferner versichere ich, dass ich die Dissertation bisher nicht einem Fachvertreter an einer anderen Hochschule zur Überprüfung vorgelegt oder mich anderweitig um Zulassung zur Promotion beworben habe.

Ich erkläre mich einverstanden, dass meine Dissertation vom Dekanat der Medizinischen Fakultät mit einer gängigen Software zur Erkennung von Plagiaten überprüft werden kann.

Unterschrift: .....