

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



**Case-based Learning –
Was macht einen Fall komplex?**

Dissertation
zum Erwerb des Doktorgrades der Medizin
an der Medizinischen Fakultät
der Ludwig-Maximilians-Universität München

vorgelegt von
Kim Sarah Öhler

aus
Laichingen

Jahr
2025

Mit Genehmigung der Medizinischen Fakultät der
Ludwig-Maximilians-Universität München

Erstes Gutachten: Prof. Dr. Martin Fischer
Zweites Gutachten: Prof. Dr. Ralf Schmidmaier
Drittes Gutachten: Priv. Doz. Dr. Thomas Geyer

Dekan: Prof. Dr. med. Thomas Gudermann

Tag der mündlichen Prüfung: 21.05.2025

Danksagung:

Mein großer Dank gilt in erster Linie meinem Doktorvater Herrn Prof. Dr. Martin Fischer. Seine fachliche Expertise, genauso wie sein interprofessioneller Weitblick, waren mir eine große Stütze und haben dieses Projekt erst ermöglicht. Ich möchte mich im Besonderen für seine dauerhafte, zielgerechte Unterstützung, genauso wie für sein Vertrauen, bedanken.

Im Weiteren möchte ich mich bei meinem Betreuer Prof. Dr. Ralf Schmidmaier für die Teilhabe an seinem Erfahrungsschatz bedanken. Seine klinische Expertise zusammen mit seiner organisatorischen Unterstützung waren mir eine große Hilfe. Mein weiterer Dank gilt meinem Betreuer Dr. Dr. Marc Weidenbusch für seine unermüdliche Begeisterung für das Fach der Medizindidaktik. Er hat dieses Projekt in großem Maße unterstützt und hat mich menschlich, sowie fachlich begleitet.

Darüber hinaus gilt mein Dank Dr. Jan Zottmann. Sein großer Einsatz für das Projekt war für mich Motivation und Hilfe zugleich. Seine Unterstützung in wissenschaftlichen Angelegenheiten und seine Perspektive außerhalb der Medizin haben diese Arbeit bereichert. Ein weiteres Dankeschön möchte ich Prof. Dr. Matthias Stadler widmen. Sein besonderes Fachwissen hat dieses Projekt auf ihrer Zielgeraden maßgeblich unterstützt und begleitet.

Ich möchte mich bei allen früheren, sowie aktuellen Mitarbeiterinnen und Mitarbeitern des Instituts für Didaktik und Ausbildungsforschung der LMU, die mich bei diesem Projekt unterstützt haben, bedanken. Mein ausdrücklicher Dank gilt Dr. Leah Braun, Katharina Borrmann, sowie Dr. Benedikt Lenzer für die Bereitstellung ihrer Forschungsergebnisse und die Unterstützung in sämtlichen Anliegen. Im Weiteren danke ich Prof. Dr. Josef Eberle für seine Beratung in Bezug auf das Scoring-System. Elisabeth Hilger und Dr. Frank Papa möchte ich für die gemeinsame, produktive Arbeit an diesem, sowie anderen, spannenden Projekten, danken.

Zuletzt bedanke ich mich von Herzen bei meinen Freunden und bei meiner Familie. Sie haben dieses Projekt durch Blickwinkel außerhalb der Medizin herausgefordert und meine Arbeit mit Leben gefüllt.

Für Karl

Zusammenfassung

Einleitung: Das Lernen mit Fällen nimmt mittlerweile eine entscheidende Rolle in der Medizinischen Ausbildung ein. Genau wie andere Lernkonzepte sollte auch die fallbasierte Unterrichtsmethode an die Fähigkeiten des Lernenden angepasst werden. Dabei kann es anspruchsvoll sein, die Schwierigkeit eines Falls zu bestimmen (1). Um sich dieser Aufgabe zu widmen, kann man sich eines empirisch wenig erforschten Konzeptes, der Fallkomplexität, bedienen (2). Die beiden Ziele dieser Arbeit sind, ein universell anwendbares Scoring-System für die Fallkomplexität zu entwickeln und dieses anschließend zu validieren. Der zweite Teil widmet sich der Frage, ob Komplexitätswerte aus dem neuen System ein Prädiktor für die Fallschwierigkeit sein können.

Methoden: In einem ersten Schritt wurde ein universell anwendbares Scoring-System **CASE: ComplexityAssessmentStructure** entwickelt. Dieses wurde auf 338 Fälle unterschiedlicher Autoren und Fallpräsentationen angewandt und die Interrater-Reliabilität von 59 Fällen anhand von Cohen's Kappa bestimmt. Anschließend wurden die Daten von 342 Studienteilnehmenden, die 45 Fälle bearbeiteten, anhand eines linear-logistischen Testmodells (LLTM) in Verbindung mit einem Rasch-Modell (RM) ausgewertet.

Ergebnisse: Die Werte für die Interrater-Reliabilität zeigten eine moderate, bis nahezu perfekte Übereinstimmung bei den Fallbearbeitungen. Es konnte eine mittlere bis starke Korrelation zwischen den geschätzten Schwierigkeiten des RM und des LLTM beobachtet werden. Eine hohe Korrelation zwischen den geschätzten Schwierigkeiten der beiden Modelle deutet auf eine valide Repräsentation der Item-Schwierigkeiten durch das festgelegte Item-Charakteristikum: Komplexität hin.

Diskussion: Diese Arbeit stellt einen Vorschlag für ein Punktesystem der Fallkomplexität vor, das auf ein breites Spektrum von Fällen anwendbar ist. Basierend auf den Korrelationen zwischen der Komplexität und der Schwierigkeit von Fällen können Komplexitätsscores voraussichtlich als ein Instrument verwendet werden, um die fallbasierte Ausbildung individuell zu steuern. Weitere Forschung auf diesem Gebiet sollte sich in Zukunft vor allem der Fallkomplexität im Kontext der Schwierigkeit widmen, um Studierenden das richtige Lernmaterial zum richtigen Zeitpunkt im Studium anbieten zu können.

Inhaltsverzeichnis

Zusammenfassung	V
1 Fallbasiertes Lernen als zentrales Element der medizinischen Aus- und Weiterbildung	1
1.1 Geschichte des fallbasierten Lernens	1
1.2 Definitionen: was genau ist ein Fall – und was nicht?	4
1.3 Wirkmechanismen fallbasierten Lernens	10
1.4 Chancen und Risiken beim Einsatz fallbasierten Lernens	13
2 Schwierigkeit und Komplexität als zentrale Fallcharakteristika	16
2.1 Fallschwierigkeit	16
2.1.1 Definitionen von Fallschwierigkeit	16
2.1.2 Faktoren, die die Fallschwierigkeit beeinflussen	17
2.1.3 Möglichkeiten der Erfassung von Fallschwierigkeit	19
2.2 Fallkomplexität	21
2.2.1 Definitionen von Fallkomplexität	21
2.2.2 Faktoren, die die Fallkomplexität beeinflussen	23
2.2.3 Möglichkeiten der Erfassung von Fallkomplexität	28
2.3 Fallschwierigkeit und Fallkomplexität im Verhältnis	31
3 Problemstellung und Forschungsfragen	33
3.1 Inwiefern kann die Fallkomplexität im medizinischen Kontext reliabel und valide erfasst werden?	33
3.2 Inwiefern lässt sich die Fallschwierigkeit mit Hilfe der Fallkomplexität vorhersagen?	34
4 Methoden	35
4.1 Qualitative Betrachtung der Fallkomplexität	35
4.1.1 Anwendung eines bestehenden Modells	35
4.1.2 Weiterentwicklung zu einem neuen Scoring-System CASE	37
4.1.3 Prüfung des neuen Scoring-Systems	43
4.2 Quantitative Datenanalyse	46
4.2.1 Akquise und Aufbereitung der Daten	46
4.2.2 Das Rasch-Modell	49
4.2.3 Das Linear-logistische Testmodell	50
4.2.4 Überprüfung der Modellgültigkeit	51
4.2.5 Schätzung der Fallschwierigkeit	52
4.2.6 Der Zusammenhang zwischen Fallkomplexität und Fallschwierigkeit	53

5 Ergebnisse	54
5.1 Resultate der qualitativen Analyse der Fallkomplexität.....	54
5.1.1 Notwendigkeit der Weiterentwicklung bisheriger Konzepte	54
5.1.2 Vorschlag eines Scoring-Systems CASE zur standardisierten Erfassung von Fallkomplexität.....	59
5.2 Ergebnisse der quantitativen Untersuchung	69
5.2.1 RM und LLTM ermöglichen die Analyse der Daten	69
5.2.2 Die Komplexität eines Falls beeinflusst seine Schwierigkeit	69
5.2.3 Komplexität verschiedener Komponenten mit unterschiedlichem Einfluss auf die Fallschwierigkeit	71
6 Diskussion	74
6.1 Diskussion der Modelle zur Erfassung von Fallkomplexität	74
6.2 Betrachtung der beobachteten Korrelation zwischen Komplexität und Schwierigkeit von Fällen	81
6.3 Einordnung der Ergebnisse in den Kontext der Medizindidaktik.....	83
6.4 Limitationen der Forschung.....	85
6.5 Empfehlungen für weiterführende Untersuchungen.....	86
Literaturverzeichnis	90
Abbildungs- und Tabellenverzeichnis	98
Abkürzungen	101
Anhang.....	102
Lebenslauf	115
Eidesstattliche Erklärung.....	116

Hinweis: Innerhalb dieser Arbeit wird, soweit möglich, eine genderneutrale Sprache benutzt. Zu Zwecken des Leseflusses und da für einige Fachbegriffe keine exakten deutschen Übersetzungen bekannt sind, wird teilweise nur die männliche Form der Personenbezeichnung genutzt. Diese gilt natürlich für alle Geschlechter.

1 Fallbasiertes Lernen als zentrales Element der medizinischen Aus- und Weiterbildung

1.1 Geschichte des fallbasierten Lernens

Die Dokumentation von Fallbeschreibungen ist eine der ältesten Arten, medizinisches Wissen weiterzugeben. Es bestehen Belege, dass bereits im 16. Jahrhundert vor Christus Patientengeschichten in ägyptischen Abhandlungen festgehalten wurden (3). Auch von den griechischen Ärzten Galen und Hippokrates sind Schriften bekannt, die Fälle ihrer Laufbahn enthalten (4). Während Fallbeschreibungen vermutlich primär dafür genutzt wurden, um Wissen an nachfolgende Gelehrte weiterzugeben, hat sich erst im Laufe der Zeit ein systematischer Ansatz für das Lernen entwickelt.

Das System des Lernens mit Fällen geht auf die juristische Ausbildung des späten 19. Jahrhunderts zurück. Hierbei ist bekannt, dass der Dekan der Harvard Law School Christopher Columbus Langdell 1870 erstmals den Vorschlag vorbrachte, Studierende mit Fällen zu konfrontieren. Obwohl diese Methode anfangs auf Skepsis stieß, wurde sie zwischen Ende des 19. und Anfang des 20. Jahrhunderts in etlichen Bildungseinrichtungen der Vereinigten Staaten adaptiert (5). Nach und nach erfreute sich die Methode des Lernens mit Fällen zunehmender Beliebtheit und so orientierten sich auch Einrichtungen anderer Fachbereiche wie die Harvard Graduate School of Business Administration an dieser Methode und bauten ihr Curriculum darauf auf (6). Bis zum 20. Jahrhundert verlief die Entwicklung der medizinischen Ausbildung ähnlich zu den Anfängen der juristischen Ausbildung. Bevor das Lernen mit Fällen implementiert wurde, bestand die Grundausbildung beider Disziplinen aus dem Studium von Faktenwissen. Nach und nach erkannte man, dass auch der Unterricht in der Klinik ein wichtiger Baustein für die Ausbildung eines Arztes darstellt und schloss ihn als feste Komponente in das Curriculum ein (7). 1901 fand in Deutschland eine strukturierte Reform des

Medizinstudiums statt. Zentrale Forderung war der vermehrte Einbezug praktischer Aspekte wie beispielsweise der Unterricht am Krankenbett. Auch das Studium anhand von klinischen Fallpräsentationen sollte gefördert werden (8). Obwohl das fallbasierte Lernen noch wenig konkretisiert wurde, ließen sich schon damals Tendenzen für die feste Implementierung dieser Unterrichtsform erkennen. Für die Ergänzung klassischer Lehrmethoden um den fallbasierten Ansatz wurde die Harvard Law School oft als Vorbild gesehen und andere Institutionen griffen auf ihre Methoden zurück (9). In der ärztlichen Approbationsordnung von 1970 wird der Fokus stärker denn je auf die praktisch orientierte Ausbildung gelegt und klinische Fallbesprechungen als Instrument für deren Realisierung hervorgehoben (10). Die Methode fallbasierten Lernens hat nach und nach festen Einzug in das medizinische Curriculum gehalten. Die aktuelle Approbationsordnung für Ärzte (ÄApprO) von 2002 baut in dieser Hinsicht auf den Vorherigen auf und definiert den heutigen Stellenwert fall- und problembasierter Lernstrategien innerhalb der medizinischen Lehre (11). Diese hebt hervor, dass die Lehre fächerübergreifend und besonders problemorientiert gestaltet werden sollte. Seminargruppen, die ein festgelegtes Lernziel verfolgen, sollen insbesondere dazu dienen, anhand von Fällen problembasiertes Denken zu fördern. Diese Gruppen zeichnen sich durch eine geringe Teilnehmerzahl und die Leitung durch Dozierende aus. Auch der Stellenwert der Vorlesungen innerhalb des Studiums wird definiert. In der aktuellen Approbationsordnung werden diese als Ergänzung oder Vorbereitung der Seminare eingestuft (11) und nicht wie in der Vergangenheit als primäres Element der medizinischen Ausbildung (8).

Die Begriffe fallbasiert und problembasiert werden häufig synonym verwendet. Obwohl sich die Ziele fall- und problembasierter Lehre häufig ähneln, nehmen einige Autoren bewusst eine Trennung der beiden Begriffe vor (12, 13). Problembasiertes Lernen findet meist in Kleingruppen statt, in denen die Teilnehmenden mit einer bestimmten Thematik konfrontiert werden. Ziel ist die differenzierte Auseinandersetzung mit einem Problem und die Erarbeitung einer Lösungsstrategie (12). Fallbasiertes Lernen zeichnet sich oft durch zusätzliche Interventionen aus, um den Fokus der Lernenden auf das Lernziel zu richten. Darüber hinaus haben Studierende innerhalb fallbasierter Lernumgebungen meist die Möglichkeit, Fragen zu stellen und sich im Voraus vorzubereiten (12). Fallbasiertes Lernen beinhaltet häufig problembasierte Aspekte. Daraus ergeben sich einige

Berührungspunkte beider Methoden (13). Obwohl diese Gemeinsamkeiten nicht abschließend untersucht sind, könnte problembasiertes Lernen eine Komponente fallbasierten Lernens darstellen (13).

Die Regularien der Approbationsordnung sind diesbezüglich zwar wegweisend, die konkrete Umsetzung dagegen obliegt jeder einzelnen Fakultät. „Medizin ist so spannend wie ein Krimi“ (14), findet der Leiter des Zentrums für unerkannte und seltene Krankheiten (ZUSE) in Marburg, Prof. Dr. Jürgen Schäfer. Aus diesem Grund war er einer der ersten in Deutschland, die Arztsendungen wie „Dr. House“¹ für die medizinische Lehre nutzten. In seinem Buch vergleicht er den Prozess der Aufklärung eines Verbrechens mit dem der Diagnosefindung einer medizinischen Fallpräsentation. In beiden Situationen muss detektivisches Gespür angewandt werden, um alle Informationen in Einklang miteinander zu bringen und letztendlich aus dieser Erfahrung lernen zu können (14). Diese Methode erfreute sich schnell zunehmender Beliebtheit und machte fallbasierte Lehre populärer. Immer mehr deutsche Universitäten verwendeten ähnliche Lehrmethoden wie Schäfer und nutzten fiktiven Fälle, um Wissen spannend und anschaulich zu vermitteln.

Nicht nur die Relevanz des Lernens mit Fällen hat sich über die Zeit geändert, sondern auch das Format. Die klassische Art, Lehrmaterial zu präsentieren, ist der Aufschrieb oder Ausdruck auf Papier. Aus den Überlieferungen zur Reform des deutschen Medizinstudiums geht hervor, dass im vergangenen Jahrhundert vorwiegend Wert darauf gelegt wurde, Studierenden Zugang zu Schriften in Bibliotheken für das Selbststudium zu gewähren (8). Diese Methode war neben Präsentationen von Dozierenden und der eigenen Erfahrung am Krankenbett lange Zeit die einzige Möglichkeit, Fälle für das Lernen zu nutzen. Erst durch die Digitalisierung und globale Vernetzung etablierten sich neue Methoden. Heutzutage besitzen einige Fakultäten ihr eigenes Online-Lernprogramm für medizinische Fälle. Für die Ludwig-Maximilians-Universität München (LMU) bietet die Plattform CASUS die Möglichkeit, Patientenfälle digital zu bearbeiten (15). Das Programm kann von Lehrbeauftragten genutzt werden, um zu steuern, welche Fälle zu welchem Zeitpunkt im Studium bearbeitet werden sollen. Diese sind so fester

¹ https://de.wikipedia.org/wiki/Dr._House

Bestandteil des Curriculums und deren Bearbeitung kann zentral überprüft werden (15). Die Fallbearbeitungen sind dabei meist interaktiv und ermöglichen beispielsweise das Abrufen von Untersuchungsergebnissen, Röntgenbildern und Laborwerten, analog zu den zukünftigen klinischen Abläufen. Gleichzeitig kann direkt Feedback innerhalb der Fallbearbeitungen generiert oder die Daten für Forschungszwecke ausgewertet werden (15). Ein anderes Projekt der Charité – Universitätsmedizin Berlin bietet die Möglichkeit, deutschlandweit digitale Lernressourcen zentral und fakultätsübergreifend zur Verfügung zu stellen. Auf der Plattform LOOP share können Fallbeschreibungen verschiedenster Autoren gefunden und eingesehen werden (16). Heutzutage ist fallbasiertes Lernen demnach nicht mehr primär von dem Erfahrungsschatz der Lehrenden abhängig, sondern durch die Fülle an verfügbarem Fallmaterial vergleichsweise einfach umzusetzen. Auch Technologien wie Videosimulationen werden bereits für fallbasiertes Lernen genutzt (17); in Zukunft werden dafür vermutlich noch weitere Optionen (z.B. Virtual Reality) zur Verfügung stehen.

1.2 Definitionen: was genau ist ein Fall – und was nicht?

Kolodner führt 1992 im Zusammenhang mit dem Fallbegriff das Alltagsbeispiel der Zubereitung einer Mahlzeit für mehrere Personen an (1). Diese verfügen über verschiedene Vorlieben und Allergien, was die Aufgabe nicht einfach gestaltet. Sie erfordert vielfältige Überlegungen und Denkprozesse, um die passende Lösung (in diesem Fall: das passende Gericht) zu finden. Um diese Problemstellung zu bewältigen, bedient sich die Gastgeberin fallbasierter Denkprozesse. Ein wichtiger Schritt ist dabei, sich an vorherige Erlebnisse zu erinnern und somit das Problem zu charakterisieren (1). Im Beispiel hat eine Bekannte bei einem früheren Treffen keinen Fisch gegessen. Die Folgeüberlegung für die aktuelle Situation ist demnach, ob eine andere Art von Fisch serviert werden sollte oder überhaupt keiner. Kommt die Gastgeberin unter Betrachtung ihrer früheren Erlebnisse und zusätzlichem Wissen, z.B. zu Allergien ihrer Gäste, zu einem Urteil, welches Gericht für alle Teilnehmer passend ist, so wurde fallbasiertes Lernen genutzt, um ein definiertes Problem zu lösen (1). Im täglichen Leben zieht das Lernen mit Fällen nicht immer eine Handlungsfolge nach sich. Oft wird es genutzt, um sich verschiedene Situationen erklären zu können, z.B. wieso eine bestimmte Aktivität

nicht zum gewünschten Ergebnis geführt hat. Auch hierbei kann es hilfreich sein, auf vergleichbare eigene oder fremde Erlebnisse zurückzugreifen, um eine Lösung zu finden (1). Das Beispiel veranschaulicht, dass nicht nur im akademischen Kontext, sondern vor allem im Alltag fallbasierte Überlegungen angestellt werden. Was allerdings konkret als Fall gilt und was nicht, ist nicht sofort klar zu definieren.

Im deutschen Sprachgebrauch hat der Begriff des Falls unzählige verschiedene Bedeutungen. Die Definition, die für den Kontext dieser Arbeit am ehesten passend ist, charakterisiert einen Fall als „bestimmte (zu klärende) Angelegenheit“ (18). Darüber hinaus besteht eine separate Definition innerhalb der Medizin als „das Auftreten, Vorhandensein einer Krankheit, Krankheitsfall“ (18). Besonders in der Medizin wird mit dem Begriff des Falls meist eine Gegebenheit beschrieben, die sich bereits ereignet oder präsentiert hat und nun zusammenfassend betrachtet wird. Diese gewisse Vorzeitigkeit wird beispielsweise auch beim Begriff des Kriminalfalls deutlich, bei dem ein Verbrechen bereits stattgefunden hat.

Ein Fall kann vor allem je nach Anwendungsgebiet unterschiedlich definiert werden. Generell wird der Begriff oft verwendet, wenn es um frühere Erfahrungen geht. Dabei kann ein Fall dazu dienen, neue Situationen im Kontext vorheriger Erlebnisse zu interpretieren. In der Medizin beinhalten Fälle meist noch einen zusätzlichen Aspekt: das Endergebnis einer früheren Problemstellung. Bei dieser Verwendung von Fällen wird vor allem die Lösung einer Problematik herangezogen und auf die aktuelle Situation angewandt (1). Die Bearbeitung eines Falls charakterisiert sich somit durch das Erinnern an bisherige Erlebnisse, den individuellen Umgang damit und mögliche Lösungswege.

Betrachtet man die unzähligen Anwendungsgebiete von Fällen, ist eine präzise Definition schwer möglich. In der Medizin werden Studierende meist früh an das Lernen mit Fällen herangeführt, einheitliche Maßstäbe gibt es hierfür jedoch kaum. Wenn im medizinischen Kontext von Fällen die Rede ist, sind damit meist sogenannte Case Studies, zu deutsch Fallstudien, gemeint (19). Darunter versteht man didaktische Beispiele, die für ähnliche Probleme in der Zukunft hilfreich sein können (19). Medizinische Fälle ähneln sich meist in der Form und beinhalten allgemeine Informationen zu einem Patienten und sämtliche Beobachtungen, die der Behandler während der Konsultation macht. Um den Lernprozess zu

unterstützen, gibt es oftmals Anmerkungen, was in diesem konkreten Fall nicht beachtet wurde oder wie man ihn in Zukunft besser lösen könnte. Meist werden verschiedene Differentialdiagnosen herausgearbeitet und kontrastiert, um beim Leser differenzierte Denkprozesse anzustoßen (19). Es geht bei medizinischen Fällen somit weniger darum, dem Leser eine Musterlösung mit richtiger Diagnose inklusive Behandlungsweg zu präsentieren, sondern vielmehr darum, auf praxisorientierte Weise den Lernprozess zu fördern. Je nach Lernziel werden Fälle allerdings sehr unterschiedlich präsentiert. In renommierten Zeitschriften wie beispielsweise dem New England Journal of Medicine (NEJM) werden oft sehr ausführliche Fallbeschreibungen veröffentlicht (20–22). Sie beinhalten eine detailreiche Krankengeschichte, generelle Informationen über den Patienten, Untersuchungsergebnisse, Behandlungsansätze und eine Fülle an möglichen Differentialdiagnosen. Diese Fälle präsentieren realistische Patientengeschichten, die sich oft durch seltene Krankheiten oder eine unübliche Manifestierung von Symptomen auszeichnen. Sie können von Studierenden und (zur Weiterbildung) von erfahrenen Ärzten genutzt werden (20–22), (vgl. Anhang #1). Für das Medizinstudium werden derart umfassende Fallpräsentationen jedoch selten genutzt. Für die Gestaltung der Lehre innerhalb des Medizinstudiums ist es möglich, durch die Auswahl von Fällen ein oder mehrere spezifische Lernziele abzubilden. Dabei muss ein Fall nicht notwendigerweise einen realen Patienten darstellen, sondern kann auch einen prototypischen Charakter haben. Dozierende orientieren sich bei der Erstellung von Fällen meist an Ereignissen aus ihrer klinischen Erfahrung und passen diese an die spezifischen Anforderungen der Studierenden an.

Der Großteil von Fällen, die für das Medizinstudium genutzt werden, orientiert sich an einer festen Struktur. Diese Fälle beinhalten meist folgende Informationen: eine Anamnese mit Informationen zu aktuellen Beschwerden, Risikofaktoren, Medikamente, Erkrankungen und psychosoziale Aspekte, eine körperliche Untersuchung und technische Befunde (23). Fälle werden besonders in der Medizin auch häufig für Studien genutzt (2, 13, 17). Je nach Forschungszweck ist es allerdings auch denkbar, dass Fälle nur aus einer einzelnen Information bestehen.

Aus diesen Überlegungen ergibt sich die Frage, welche Arten von Informationen für einen medizinischen Fall unverzichtbar sind und was nicht als Fall gilt. Für die

Beantwortung kann man sich beispielsweise an der CARE-Leitlinie für die Erstellung klinischer Fallberichte orientieren (24). Diese charakterisiert einen Fallbericht als die Darstellung einer bestimmten Problematik eines oder mehrerer Patienten, die für didaktische, medizinische oder theoretische Vorhaben genutzt wird (24). Was diese und andere Beschreibungen eines Falls gemeinsam haben, ist der Fokus auf ein (medizinisches) Problem. Dieses kann als Hauptkomponente eines Falls angesehen werden und wird meist durch eine Fülle an Informationen dargestellt. Je nach Fachgebiet oder Anwendungsgebiet innerhalb der Medizin kann es allerdings Unterschiede geben, durch welche Art von Information das Problem charakterisiert wird. So existieren Beschreibungen, die neben Patientendaten nur eine Aufzählung von verschiedenen Symptomen beinhalten. Obwohl keine weiteren Informationen bekannt sind, könnte man dies dennoch als Fall definieren. Durch die Symptome wird eine medizinische Thematik dargestellt, was den Leser vermutlich zu Überlegungen bezüglich einer Diagnose anregen wird. Auf der Plattform CASUS finden sich aber auch Vignetten, bei denen nicht eindeutig ist, ob sie die bisher genannten Charakteristika eines (medizinischen) Falls erfüllen. Innerhalb dieser Vignetten werden unter anderem didaktische Vorgehensweisen und Instruktionsansätze für die Ausbildung von Lehrern thematisiert (25). Obwohl dadurch zwar ein Problem adressiert wird, sind die klassischen Charakteristika eines medizinischen Falls hier nicht zu finden. Durch dieses Beispiel wird deutlich, dass für verschiedene Anwendungsgebiete voraussichtlich unterschiedliche Falldefinitionen gelten müssen.

Innerhalb der Medizin gibt es neben Fällen mit charakteristischer Struktur (2, 23) auch Vignetten ohne Bezug zu einem konkreten Patienten. Hierbei handelt es sich beispielsweise um Darstellungen von Diagnosen und den dazu passenden Symptomen. Diese Form der Präsentation wird häufig in Lehrbüchern oder Onlineprogrammen für die Lehre genutzt. Eine Problemstellung oder weitere wichtige Charakteristika eines Falls sind daraus allerdings nicht ersichtlich. Diese Art von Information würde somit nicht als medizinischer Fall klassifiziert werden, da hier die Vermittlung von Wissen im Vordergrund steht und nicht dessen Anwendung.

Neben den inhaltlichen Aspekten eines Falls kann man zwei weitere Charakteristika von Fallbearbeitungen identifizieren: formale Faktoren und individuelle

Faktoren. Formale Faktoren beschreiben in diesem Kontext Aspekte, die sich weniger auf den Inhalt eines Falls, sondern auf die fallbasierte Lernumgebung beziehen. Diese Merkmale sollen am Beispiel von virtuellen Patienten (VP) dargestellt werden. VP werden definiert als „eine interaktive Computersimulation von realen klinischen Szenarien für den Zweck des Trainings innerhalb des Gesundheitswesens und der Medizin, der Ausbildung oder der Beurteilung“ (26). VP sind eine spezielle Art, Fälle zu präsentieren, die besonders durch den technischen Fortschritt immer mehr Aufmerksamkeit erlangt. VP bieten die Möglichkeit, Fälle online und interaktiv darzustellen (26). Da innerhalb des Curriculums der LMU ein großer Teil von Fällen auf diese Weise anhand der Plattform CASUS online bearbeitet wird, soll deren Charakteristika besondere Aufmerksamkeit geschenkt werden. Huwendiek et al. benennen 19 Hauptmerkmale eines VP, die vier Kategorien zugeordnet werden können (27) (vgl. Tabelle 1).

Tabelle 1: Merkmale eines VP, gekürzt aus Huwendiek et al. 2009 (27)

Kategorie	Faktor	Beschreibung
Generell	Titel	Name des Falls
	Beschreibung	Freitext
	Sprache	Standardsprache
	Identifikation	Individuelle ID oder Adresse
	Herkunft	Autor und Mitwirkende
	Übliche Lernzeit	Durchschnittliche Fallbearbeitungszeit
Ausbildung	Ausbildungsstand	Z.B. absolvierte Fachsemester, Oberarzt, etc.
	Lernmethode	Lehren, Lernen und Bewertung. Formative und Zusammenfassende Aspekte
	Themenumfang	Abgedeckte Themen
	Ziele und Ergebnisse	Ziele des Falls und auf welche Resultate sie sich beziehen
Instruktionsdesign	Pfad	Linear oder verzweigt
	Benutzermodalität	Zahl der involvierten Anwender und ihre Rollen
	Medien und Ressourcen	Verwendung von Bildern, Audio, Video, Animationen, etc.

	Geschichte und Patientenfokus	Gleichgewicht zwischen dargestellter Information (eine Geschichte oder Laborresultate lesen) und einen Patientenfall darstellen (Teilhabe an der Konversation, am Charakter oder Motiv). Ist Behandlung des Patienten Hauptfokus oder ist sie Mittel für andere Aufgaben wie z.B. Zusammenarbeit im Team, Grundlagenwissenschaft oder Professionalität?
	Interaktivität	Art der Fragen und Aufgaben. Gesamtheit der kognitiven Interaktionen
	Feedback	Arten von Feedback während, nach der Fallbearbeitung oder beides
Technisch	Programm	Welches Programm wurde für die Fallbearbeitungen genutzt?
	Format	Welches technische Format am Ende genutzt wird - z.B. text/HTML für Web und application/EXE für disc basierte Medien
	Integration und Schnittstellen	Welche anderen Formate, Werkzeuge oder Kontextfaktoren benötigt das Programm?

Diese Charakteristika stellen Faktoren dar, die nicht nur bei der Analyse, sondern auch bei der Konstruktion von Fällen eine Rolle spielen. Nichtsdestotrotz gibt es möglicherweise weitere Parameter im Zusammenhang mit fallbasierten Lernumgebungen, auf die Dozierende, die fallbasiertes Lernen einsetzen, nur bedingt Einfluss haben. Diese individuellen Aspekte beziehen sich weniger auf das Fallmaterial, sondern mehr auf Umwelt- und persönliche Faktoren. Hier könnte z.B. die Motivation der Studierenden eine Rolle spielen (17).

Diese Arbeit soll sich weniger auf formale und individuelle Faktoren konzentrieren, sondern sich inhaltlichen Aspekten widmen. Obwohl gewisse inhaltliche Charakteristika eines Falls definiert werden können, erlaubt die Fülle an verschiedenen Fachdisziplinen, Anwendungsgebieten und Zielsetzungen es nicht, eine allgemeingültige Definition zu formulieren. Innerhalb der Medizin kann man sich beispielsweise an festen Richtlinien zur Erstellung von Fallberichten orientieren (24). Dennoch sollte es möglich bleiben, die Form von Fällen für individuelle Zwecke anzupassen und eine eigene Definition, gemäß der jeweiligen Zielsetzung, aufzustellen.

1.3 Wirkmechanismen fallbasierten Lernens

Fallbasiertes Lernen beinhaltet das Zurückgreifen auf frühere Erlebnisse, um eine neue Aufgabe meistern zu können. Ein Ziel dieser Lernstrategie ist, dass sich eine Person an ein ähnliches Problem in der Vergangenheit erinnert und das dort Gelernte auf die aktuelle Situation anwenden kann. Dies kann nicht nur das Lösen des aktuellen Problems, sondern auch die Interpretation oder Bewertung neuer Situationen und Lösungen beinhalten. Der grundlegende Vorgang im Umgang mit Fällen wird auch als fallbasiertes Argumentieren, beziehungsweise case-based reasoning bezeichnet. Dieser Prozess ist die Voraussetzung für fallbasiertes Lernen (1). Da die kognitiven Vorgänge dabei oft unterbewusst ablaufen, werden sie im Folgenden erläutert.

Fallbasiertes Lernen kann als eine Abfolge bestimmter Aktivitäten betrachtet werden. In einem ersten Schritt muss eine Verbindung des aktuellen Problems und früheren Erfahrungen hergestellt werden. Dabei ist es wichtig, die passenden Beispiele zu finden, die in der Folge eine Vorhersage für die neue Situation erlauben (1). Anschließend sollte die Auswahl der Fälle auf den passendsten oder einige wenige begrenzt werden. Im nächsten Schritt wird die Lösung bzw. der Lösungsansatz des alten Falls in Erinnerung gerufen, um zu einem Ergebnis für das aktuelle Problem zu kommen (1). Da frühere und neue Fälle in der Medizin selten vollkommen identisch sind, muss im Anschluss der Lösungsansatz der früheren Situation oft an die neue angepasst werden. Bevor die Lösung angewandt werden kann, sollte sie erst kognitiv bewertet werden. Hierbei kann sie erneut mit ähnlichen Situationen abgeglichen werden, um einzuschätzen, ob eine bestimmte Lösung auch in diesem Fall wirklich funktionieren kann (1). In einem vorletzten Schritt wird das Ergebnis nun am realen Problem erprobt. Danach kann evaluiert werden, ob der vorangegangene Prozess erfolgreich war, oder ob in Zukunft Anpassungen vorgenommen werden müssen. Dieser Schritt ist sehr wichtig, um Lernprozesse zu ermöglichen. Letztlich wird auch die neue Lösung im Gedächtnis verankert, um sie für zukünftige Fälle nutzen zu können (1) (vgl. Abbildung 1). Wie gut eine Lösung letztendlich ist, hängt von einigen Faktoren ab. Hierbei spielen frühere Erfahrungen mit Fällen und den Prozessen der Anpassung und der Evaluation eine wichtige Rolle. Darüber hinaus ist auch die individuelle Fähigkeit für logisches Denken essenziell (1).

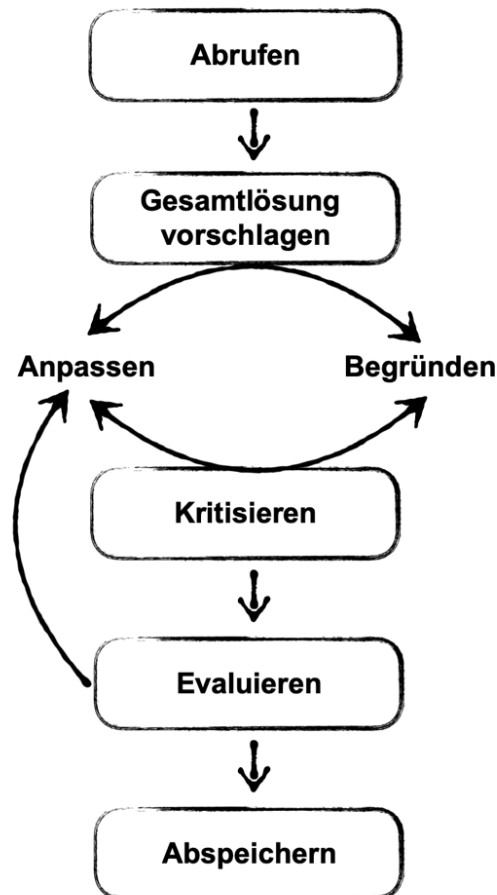


Abbildung 1: Prozesse fallbasierten Denkens aus Kolodner 1992 (1)

Lernen mit Fällen findet oft in kleinen Gruppen statt (28), die von Dozierenden geleitet werden. Dies ermöglicht nicht nur, dass Studierende jederzeit Feedback von einem Experten erhalten können, sondern fördert ebenso den Meinungsaustausch und den generellen Einsatz der Teilnehmenden (29, 30). Die Motivation ist sicherlich ein wichtiger Faktor, wenn man betrachtet, wie fallbasiertes Lernen genau wirkt. Studierende geben an, dass diese Art zu lernen ihre Aufmerksamkeit und ihren Ideenreichtum fördert (31). Fallbasierte Lernumgebungen schaffen eine kreative Lernatmosphäre. Hierbei besteht Raum für Fehler und vor allem die Akzeptanz für individuelle Problemlösestrategien. Um dennoch einen effektiven Lernprozess zu gewährleisten, können Moderatoren mit Hinweisen eingreifen (12). Fallbasiertes Lernen erleichtert allerdings nicht nur den allgemeinen Lernprozess, sondern stellt den Bezug zu realen Situationen her (32). Dies könnte ein weiterer Grund sein, warum diese Methode so beliebt ist. Durch die direkte Anwendung des gelernten Wissens auf Patientenfälle wird der Transfer zwischen Lehre und Praxis geschaffen (28, 33).

Relevant sind darüber hinaus auch die grundlegenden Voraussetzungen, die zur Wirksamkeit fallbasierten Lernens beitragen. Die Authentizität eines Falls ist dabei ein wichtiges Charakteristikum (34). Diese wird z.B. dadurch erreicht, dass sich der Inhalt von Fällen an tatsächlichen Erlebnissen orientiert (35). Weitere wichtige Aspekte eines guten Falls sind, dass er detailliert und realistisch ist (7). Realitätsgetreue Fälle können kreiert werden, indem Komponenten verwendet werden, die unter anderem irrelevante Aspekte als Instrumente der Ablenkung beinhalten. Darüber hinaus wird eine reale Atmosphäre geschaffen, wenn das Material erst nach und nach präsentiert wird (36). Es ist davon auszugehen, dass realistische fallbasierte Szenarien das Interesse der Studierenden fördern, da sie spannend sind und vor allem direkt ersichtlich ist, wofür das Erlernte hilfreich sein kann (35).

Für die Wirkung fallbasierten Lernens spielen allerdings nicht nur die Leistung des Individuums oder der Inhalt des Falls eine Rolle, sondern auch die Art, wie dieser präsentiert wird. Studierende bevorzugen dabei eher Onlineformate im Vergleich zum klassischen Frontalunterricht. Die Anwenderfreundlichkeit internetbasierter Lehr-Lernformate spielt dabei eine große Rolle für die Teilnehmenden (37). Abgesehen von Onlineprogrammen, die den Lernenden automatisiert Feedback geben, sind Dozierenden-geleitete Gruppendiskussionen ein gängiges Lehr-Lernformat. Der Dozent nimmt dabei eine wichtige Rolle ein und fungiert als eine Art Tutor, der dafür sorgt, dass das Lernziel abgedeckt wird. Dabei kann er auf implizite Fragen zurückgreifen, Studierenden Hinweise geben und dazu animieren, den Inhalt zusammenzufassen. Diese Methode ist Teil eines Konzepts, das als Scaffolding bezeichnet wird (7).

Eine theoretische Grundlage für dieses Vorgehen bietet Vygotskys Zone of Proximal Development (38). Nach diesem Modell hängt die Wahrscheinlichkeit, mit der eine Aufgabe gelöst wird, nicht nur von der Schwierigkeit der Aufgabe selbst ab, sondern auch von der Unterstützung, die währenddessen bereitgestellt wird (39). Diese Hilfestellungen bzw. Scaffolds können dazu dienen, die Begeisterung der Lernenden zu wecken, die Aufmerksamkeit auf bestimmte Aspekte zu lenken oder das Lernmaterial bei Bedarf zugänglicher zu gestalten (39). Die Instrumente, die für die Methode des Scaffolding genutzt werden können, sind vielfältig (40). Sie können einerseits durch die Teilnehmenden selbst umgesetzt werden,

indem diese beispielsweise dazu aufgefordert werden, wichtige Aspekte des Falls zusammenzutragen (41) oder andererseits als externe Hilfestellungen angeboten werden (7). Als Scaffolds werden darüber hinaus nicht nur Vorgänge bezeichnet, sondern auch Modifikationen des Lernmaterials selbst. Hierbei können Scaffolds verwendet werden, die z.B. die Typizität oder auch die Komplexität der Lernumgebung betreffen (40).

Die generellen Effekte von Scaffolds und der Einfluss ihrer verschiedenen Instrumente sind nicht abschließend geklärt. Es gibt Hinweise darauf, dass Scaffolds die diagnostische Effizienz Medizinstudierender verbessern können (42). Die Studienlage hierzu ist jedoch zum Teil widersprüchlich (41, 42).

Zusammenfassend lässt sich festhalten, dass es viele Anhaltspunkte und Erklärungsansätze für die Effekte fallbasierten Lernens gibt. Dennoch basieren viele der genannten Wirkmechanismen auf der Rückmeldung von beteiligten Personen und weisen oftmals einen geringen Evidenzgrad auf (28). Insgesamt gibt es nur wenige Analysen zu Wirkmechanismen fallbasierten Lernens. Diese charakterisieren die Lernmethode allerdings meist als effektives Format, was den Zeitaufwand und die Zielorientierung angeht. Sowohl Teilnehmende als auch Dozierende geben an, diese Form des Lernens besonders zu schätzen (12). Die zugrundeliegenden Prozesse fallbasierten Lernens und die Faktoren, die zu dessen Effektivität beitragen, müssen jedoch auch in Zukunft näher untersucht werden.

1.4 Chancen und Risiken beim Einsatz fallbasierten Lernens

Die Methode, anhand von Fällen zu lernen, bietet zahlreiche Vorteile gegenüber anderen Ansätzen. Das vielleicht größte Potential liegt dabei in dem problembasierten Ansatz. Dieser ist insbesondere in der medizinischen und juristischen Ausbildung sehr beliebt (7). Sowohl angehende Juristen als auch Mediziner werden durch diese Lernform bereits früh an die Art von Denkprozessen herangeführt, die sie letztendlich im Beruf benötigen. Durch die Methode des Lernens mit Fällen wird nicht nur die Relevanz des Lerngegenstandes deutlich (43), sondern es werden Prozesse erprobt, die so auch im späteren Berufsleben stattfinden. In den meisten Fachrichtungen besteht die ärztliche Konsultation aus einem Patienten, der sich mit einer gewissen Problematik präsentiert. Ein Arzt, der bereits früh Problemlöse- und Behandlungsstrategien in seiner Ausbildung trainiert hat,

kann diese Fähigkeiten voraussichtlich auch auf die Behandlung von Patienten anwenden. Ein weiterer wichtiger Aspekt ist die mögliche Zeitersparnis in der Praxis. Wurde man in der Vergangenheit bereits mit einem ähnlichen Problem konfrontiert, erscheint es logisch, dass die Lösung für den aktuellen Fall schneller gefunden werden kann, ohne den Prozess von vorne zu durchlaufen (1). Hierbei ist es sogar möglich, zu einem Ergebnis zu kommen, obwohl nicht alle Komponenten des Problems vollständig verstanden werden. Dabei ist es außerdem sehr hilfreich, sich an frühere Probleme oder Fehler zu erinnern, damit diese nicht ein zweites Mal auftreten (1). Das Lernen mit Fällen ermöglicht somit zu erkennen, welche Aspekte eines Problems die relevanten sind; es hilft dabei, sich auf das Wichtige zu fokussieren (1), wodurch Ressourcen gespart werden können.

Fallbasiertes Lernen hat das Potential, einen weitreichenden und nachhaltigen Lernprozess zu fördern. Dies kann gelingen, indem Studierende innerhalb einer Gruppe agieren und dabei Unterstützung von Dozierenden erhalten (28). Die Lernziele beziehen sich dabei nicht nur auf das reine Fachwissen, sondern haben viele weitere Bestandteile. Durch die meist stattfindende Interaktion mit anderen Studierenden oder Experten kann die Fähigkeit zur Zusammenarbeit und Kommunikation gefördert werden (43, 44). Die fallbasierte Lernmethode ist generell nicht zwingend an bestimmte örtliche oder personelle Voraussetzungen gebunden. Nicht nur die Interaktion innerhalb einer Gruppe, sondern auch die Selbstverantwortung von Studierenden wird gefördert, indem sie auf individuelle Weise an eine Problemstellung herangehen können. Diese Eigenschaft ist auch langfristig wichtig, um einer selbstständigen, regelmäßigen Weiterbildung im späteren Beruf nachzukommen (43). Fallbasiertes Lernen ist nicht auf die Anwendung bei der studentischen Ausbildung begrenzt. Die Methode bietet das Potential für lebenslanges Lernen, indem Fälle an den jeweiligen Wissensstand angepasst werden können. Angehende Fachärzte und erfahrende Kliniker könnten von fallbasiertem Lernen ebenfalls profitieren. Moderne, internetbasierte Lernprogramme werden dabei heutzutage oft bevorzugt (37). Diese Formate gestalten das Lernen mit Fällen zeitgemäß und gleichzeitig flexibel.

Feedback ist ein wichtiges Instrument während des Lernprozesses (45). Egal, ob fallbasiertes Lernen in Präsenz oder online stattfindet, es ermöglicht auf vergleichsweise einfachem Weg, Feedback zu implementieren. In

Lehrveranstaltungen kann der Experte, bzw. die Lehrperson dafür zur Verfügung stehen. In Onlineszenarien werden bestimmte Funktionen von Programmen genutzt, um automatisiert Feedback oder Hinweise zu generieren (15, 46). Dadurch können die Bearbeitenden direkt eine Rückmeldung erhalten und sich auf die wichtigen Merkmale eines Falls konzentrieren.

Obwohl das Lernen mit Fällen vielfältige Möglichkeiten bietet, birgt es auch einige Risiken. Die Fähigkeiten, die durch fallbasiertes Lernen vermittelt werden, sind zwar vielfältig, oft allerdings auch unorganisierter im Vergleich zu klassisch angeeignetem Wissen (47). Vor allem die Anwendung fallbasierten Lernens als einzige Lernmethode sollte kritisch gesehen und möglicherweise durch Strategien ergänzt werden, die zusammenhängend und strukturiert Wissen vermitteln. Es ist zu beachten, dass Studierende erst eine Wissensbasis aufbauen müssen, bevor fallbasiertes Lernen wirksam sein kann (47). Die Auswahl geeigneter Fälle für den richtigen Zeitpunkt in der medizinischen Ausbildung stellt dabei eine Herausforderung dar. Ist ein Fall zu einfach, resultiert daraus möglicherweise Langeweile und der gewünschte Lerneffekt tritt nicht ein. Ist eine Problemstellung zu fortgeschritten, kann Frustration oder Motivationsverlust daraus resultieren und das Lernziel voraussichtlich selbst mit Hilfestellung nicht erreicht werden.

Wird diese Lernmethode in einem kontrollierten Umfeld (z.B. innerhalb Kleingruppen unter der Leitung einer Lehrperson) abgehalten, erfordert es erhebliche Ressourcen. Es benötigt Zeit und vor allem die nötige Expertise der Dozierenden (47). Dies kann unter Umständen erhebliche Umstände und Kosten für die jeweilige Fakultät verursachen. Teilweise kann es sich schwierig darstellen, fallbasierte Seminare in das ohnehin schon volle Curriculum einzubauen. Einige Fakultäten bieten daher freiwillige Kurse oder digitale Lehrkonzepte an. Die LMU setzt diese z.B. in Form von Clinical Case Discussions (CCD) (48) um.

Die Inanspruchnahme solcher Angebote erfordert Eigeninitiative auf Seiten der Studierenden, die allerdings nicht immer vorausgesetzt werden kann. Wenn fallbasierte Lernmethoden zu einem gewissen Zeitpunkt im Studium eingeführt werden, kann es sein, dass Studierende sich nicht auf Anrieb innerhalb des neuen Formats zurechtfinden. Besonders der vorklinische Abschnitt besteht aus viel Frontalunterricht und der Vermittlung von Faktenwissen, in dem die Lehrperson eine klassische vortragende Rolle einnimmt. Bei dem Lernen mit Fällen hingegen

verändert sich diese Rolle. Dozierende tragen nicht vor, sondern treten in den Hintergrund und stehen im Optimalfall für Fragen und Anregungen zur Verfügung (47). Diese Konstellation erfordert eine vorangehende Einführung in das Lehr-Lernformat und Vorbereitung vor allem auf Seiten der lehrenden Person. Werden im Voraus keine klaren Prinzipien festgelegt oder kommuniziert, besteht die Gefahr, dass die Bearbeitung von Fällen unstrukturiert und nicht zielführend abläuft (47).

Eine Bedingung für effektives fallbasiertes Argumentieren ist, dass ein passender Fall aus der Vergangenheit herangezogen wird. Dabei kann es vorkommen, dass Fälle ausgewählt werden, ohne zu evaluieren, ob diese für die aktuelle Situation anwendbar sind (1). Die Selektion geeigneter Fälle ist unmöglich, wenn in der Vergangenheit noch keine ähnlichen Situationen erlebt wurden. Eine weitere Schwierigkeit besteht, wenn zwar passende Fälle erlebt wurden, man sich allerdings nicht an diese erinnern kann oder sie nicht in Verbindung mit dem aktuellen Problem bringt (49). Diese Problematik könnte beispielsweise bei wenig erfahrenen Studierenden auftreten.

2 Schwierigkeit und Komplexität als zentrale Fallcharakteristika

2.1 Fallschwierigkeit

2.1.1 Definitionen von Fallschwierigkeit

Die Schwierigkeit ist ein Parameter, der bei vielen täglichen Problemstellungen eine Rolle spielt. Im allgemeinen Gebrauch wird der Begriff verwendet, um zu beschreiben, wie leicht oder schwer uns eine bestimmte Tätigkeit fällt. Schwierigkeit wird demzufolge von jedem Einzelnen subjektiv empfunden und ist oft schlecht vergleichbar. Für die Ausbildungsforschung dagegen ist es wichtig, gewisse Größen zu objektivieren, um sie zu messen und anpassen zu können.

In der psychologischen Forschung wird der Begriff der Schwierigkeit oft verwendet, wenn es um die Beurteilung von Tests geht. Schwierigkeit bezeichnet in diesem Kontext meist die Wahrscheinlichkeit, mit der eine Person eine Aufgabe richtig löst (50). In der Medizin ist die quantitative Analyse der Schwierigkeit eine übliche Methode, um das Niveau von Prüfungsaufgaben zu kontrollieren und

sicherzustellen. Die Aufgabenschwierigkeit bezeichnet dabei die „mittlere bei dieser Aufgabe erreichte Punktzahl“ (51). Für eine Problemstellung, bei der entweder ein Punkt für die richtige Lösung oder kein Punkt für die falsche Lösung erreicht werden kann, ist die Schwierigkeit gleichzusetzen mit der relativen Menge an Studierenden, die die richtige Lösung angegeben haben. Demzufolge stehen hohe Werte für einfachere und niedrige Werte für schwierigere Aufgaben (51). Von einigen Autoren wird betont, dass die Schwierigkeit nicht alleine durch die Aufgabe selbst bedingt ist, sondern von anderen Faktoren abhängt (2). Bei der Interpretation der Schwierigkeit ist deshalb zu berücksichtigen, dass diese vor dem Hintergrund der untersuchten Gruppe beurteilt werden muss (51). Während eine bestimmte Problemstellung bei Studierenden im ersten Semester ein schlechtes Ergebnis erzielt, ist sie möglicherweise gegen Ende des Studiums für die meisten korrekt lösbar.

Für die Schwierigkeit selbst existieren empirische Richtwerte, an denen man sich bei der Erstellung einer Aufgabe oder einer ganzen Prüfung orientieren sollte. Dabei kann es sinnvoll sein, schon während des Planungsprozesses eine erste Kalkulation der Schwierigkeit durchzuführen, um eine Einschätzung der Lösbarkeit vorzunehmen (51). Für die Definition der Schwierigkeit eines Falls lassen sich in der Literatur nur wenige Ansätze finden. Einige Autoren charakterisieren die Schwierigkeit einer fallbasierten Lernumgebung über deren Typizität. Dieses Konzept befasst sich damit, inwiefern die Informationen eines Falls mit der klassischen bzw. typischen Präsentation einer Diagnose übereinstimmen (52) und wird im folgenden Absatz näher erklärt. Innerhalb dieser Arbeit soll die Falltypizität als eigenständiges Konzept angesehen und nicht per se mit der Fallschwierigkeit gleichgesetzt werden.

2.1.2 Faktoren, die die Fallschwierigkeit beeinflussen

Ein Aspekt mit möglichem Einfluss auf die Schwierigkeit eines Falls ist die Typizität. Ein Beispiel für eine typische Fallpräsentation eines Myokardinfarktes ist die Symptomkonstellation: retrosternaler Schmerz, ausstrahlend in den linken Arm, Dyspnoe und Todesangst. Die gleiche Diagnose kann allerdings auch klinisch stumm verlaufen oder sich lediglich durch Unwohlsein äußern. Studien geben Hinweise darauf, dass es Probanden leichter fällt, die richtige Diagnose eines typischen Falls anzugeben, als die eines atypischen (52, 53). Diese

Erkenntnisse deuten darauf hin, dass die Schwierigkeit eines Falls unter anderem davon abhängt, wie typisch dieser präsentiert wird. Einfache Fälle sind demnach durch typische Symptome und Befunde gekennzeichnet, während schwierige Fälle so dargestellt werden, dass sie nicht direkt an die richtige Diagnose denken lassen. Hierbei könnten Erfahrungen mit in der Vergangenheit bearbeiteten Fällen eine wichtige Rolle spielen. Je nachdem, wie sich frühere Fälle präsentiert haben, fällt die Assoziation mit dem aktuellen Problem leichter oder schwerer.

Neben der Typizität kann auch die Fülle und die Verteilung des Informationsgehalts einen Einfluss auf die Fallschwierigkeit haben. Wenn beispielsweise eine ausführliche Krankengeschichte, die Schilderung des aktuellen Problems, Hinweise zur Patientenversorgung und Angaben zu präventiven Maßnahmen vorliegen, gelingt es vermutlich eher, eine richtige Lösung zu finden, als wenn elementare Teile fehlen. Liegt die Anamnese nur bruchstückhaft vor oder fehlen Teile der körperlichen Untersuchung, könnte es dem Lernenden schwerfallen, den Fall in seiner Gesamtheit zu erfassen und einzuordnen (54). Die Art der Darstellung eines medizinischen Problems könnte demnach ein wichtiger Faktor mit möglichem Einfluss auf die Schwierigkeit und somit auf die Lösbarkeit von Fällen sein.

Bei der Bearbeitung von Fällen konnte beobachtet werden, dass individuelle Faktoren einen Einfluss auf die Fallbearbeitung haben können. Die Motivation der Studierenden ist dabei ein Merkmal, das sehr variabel ist (17). Der generelle Lerneffekt einer Methode ist maßgeblich von dem individuellen Maß an Motivation der Lernenden abhängig (55). Zusammen mit vorherigen Erlebnissen und der generellen Vorerfahrung des Einzelnen (1) könnte sie beeinflussen, wie schwierig eine Aufgabe für eine bestimmte Person ist (2).

Je nachdem, welche Fähigkeiten der jeweilige Fall erfordert und abhängig davon, ob diese vorhanden sind, kann er sich den Lernenden einfach oder schwierig präsentieren. Die für die Bearbeitung eines Falls benötigten Fähigkeiten reichen dabei von der Erinnerung an Lerninhalte oder dem Verstehen eines Problems bis hin zu der Fähigkeit, Wissen anzuwenden, zu organisieren und weitere Schritte zu planen (54). Fälle, die differenzierte Denkprozesse erfordern, könnten demnach schwieriger sein als solche, bei denen die Bearbeitung weniger kognitive Ressourcen beansprucht.

2.1.3 Möglichkeiten der Erfassung von Fallschwierigkeit

Für den Bereich des fallbasierten Lernens gibt es verschiedene Methoden, um die Fallschwierigkeit zu bestimmen. Dennoch wurde diesem Aspekt in der Vergangenheit wenig Bedeutung beigemessen (56). Der Amsterdam Clinical Challenge Scale (ACCS) wurde eingeführt, um ein einheitliches Instrument zu etablieren, mit dem die Schwierigkeit eines medizinischen Falls gemessen werden kann. Die Struktur, anhand derer hier die Punkte vergeben werden, besteht aus folgenden sechs Komponenten, die in der Patientenvorstellung eine Rolle spielen: Frühere Krankengeschichte, aktuelles Problem, Kommunikation des Patienten, körperliche Untersuchung, Patientenmanagement und Prävention (56) (vgl. Tabelle 2).

Tabelle 2: Struktur eines klinischen Falls nach dem ACCS nach Gercama et al. 2000 (56)

1.	<i>Krankengeschichte</i>
2.	<i>Aktuelles Problem</i>
3.	<i>Kommunikation des Patienten</i>
4.	<i>Körperliche Untersuchung</i>
5.	<i>Patientenmanagement</i>
6.	<i>Prävention</i>

Als erstes soll der gesamte Fall hinsichtlich seiner Schwierigkeit beurteilt werden. Im Anschluss wird jede einzelne der sechs Komponenten separat bewertet. Dabei können jeweils bis zu fünf Punkte für die Schwierigkeit vergeben werden. Grundlage für die Punktevergabe ist die Annahme, dass ein Arzt den Fall bearbeitet, ohne sich bereits mit dem Patienten auseinandergesetzt zu haben. Dieses System hat den Vorteil, dass es nicht nur für Patientensimulationen, sondern auch für reale Patienten angewandt werden kann. Gleichzeitig kann das Modell sowohl für die Lehre genutzt werden als auch die Forschung auf diesem Gebiet erleichtern (56).

Ein weiteres Modell, dass bereits für die Beurteilung der Schwierigkeit von Fällen genutzt wurde, ist die SOLO Taxonomie (Structure of the Observed Learning Outcome) (54). Dieses Konzept kategorisiert verschiedene geistige Fähigkeiten der Lernenden und kann z.B. dafür verwendet werden, gewisse Lernziele

festzulegen (54). Koskinen adaptiert dieses Modell für die Bestimmung der Schwierigkeit von Fällen, bei denen sich eine Aufgabe bzw. Frage anschließt. Diesem Vorgang liegt die Annahme zu Grunde, dass verschiedene Arten von Informationen auf unterschiedliche Weise prozessiert werden müssen (57). Daraus können Kriterien für die Beurteilung der Fallschwierigkeit abgeleitet werden (54) (vgl. Tabelle 3).

Tabelle 3: Kriterien für die Beurteilung der Schwierigkeit von Aufgaben auf Basis der SOLO-Taxonomie, gekürzt aus Koskinen 2007 (54)

Stufe 1	Die Frage ist irrelevant
Stufe 2	Die Frage erfordert den Fokus auf einen Aspekt
Stufe 3	Die Frage erfordert den Fokus auf mehrere relevante Aspekte, deren Integration nicht gelingt
Stufe 4	Der Student muss verschiedene Aspekte integrieren und eine stimmige Lösung finden
Stufe 5	Die Frage erfordert ein hohes Maß an Abstraktionsfähigkeit

Nach diesem Modell ist eine Aufgabe, die lediglich den Fokus auf einen Aspekt legt, voraussichtlich einfacher als eine solche, bei der zahlreiche Aspekte beachtet und integriert werden müssen. Die Schwierigkeit lässt sich anhand der dargestellten Stufen klassifizieren. Die Autorin schlägt dafür ein Punktesystem vor. Dieses reicht von zwei Punkten für eine Aufgabe der Stufe zwei, bis zu fünf Punkten für eine vergleichsweise schwere Thematik der fünften Stufe. Stufe eins konnte nicht für die Bewertung der Fallschwierigkeit verwendet werden, ist allerdings trotzdem mitdargestellt, da sie von der ersten Kategorie der SOLO Taxonomie abgeleitet wurde (54).

Zusammengefasst bietet dieses Modell zwar die Möglichkeit, die Schwierigkeit von Fällen zu beurteilen, es wurde aber vorrangig zur Kategorisierung von Fällen und deren anschließenden Fragen bzw. Aufgaben, entwickelt. Inwiefern dies auf den reinen Inhalt des Falls angewandt werden kann, bleibt offen. Interessant ist

zweifelloso der Zusammenhang zwischen der Schwierigkeit von Fällen, den zugrundeliegenden Anforderungen an Lernende und deren kognitiven Prozessen.

In Studien anderer Autoren wurde für die Beurteilung von Fällen kein Punktesystem für die Schwierigkeit genutzt, sondern der Fokus auf die diagnostische Richtigkeit und Effizienz der Lernenden gelegt. Liegen Daten von Fallbearbeitungen vor, kann z.B. beurteilt werden, ob die richtige Lösung vom Probanden selbst gefunden wurde und wie viele Versuche dafür benötigt wurden. Beendet ein Teilnehmer eine unvollständige Fallbearbeitung frühzeitig, könnte das daran liegen, dass die Aufgabe für ihn zu schwierig war (2, 58). Diese Parameter werden oft bei Fallbearbeitungen erhoben (41, 42, 58) und sind gut vergleichbar, in ihrer Aussagekraft allerdings eingeschränkt. Um ein vollständiges Bild davon zu erhalten, wie schwierig ein Fall ist, könnte es unter Umständen erforderlich sein, qualitative Aspekte anderer Analysemethoden miteinzubeziehen.

Die Bewertung der Schwierigkeit von Fällen ist nicht nur von hoher Relevanz für die Medizin, sondern auch für andere Domänen wie die Rechtswissenschaften. (7). Dies liegt nahe, wenn man die zahlreichen Möglichkeiten betrachtet, die ein verlässliches Bewertungsschema mit sich bringen würde. Unterteilt man einen Fall in einzelne Kategorien, könnte mit einem solchen Bewertungssystem in Abhängigkeit vom Lernziel die generelle Schwierigkeit des Falls oder die Schwierigkeit einzelner Aspekte gezielt verändert werden. Auch reale Patientenfälle, die sich tatsächlich so ereignet haben, könnten für die Lehre vereinfacht oder schwieriger gestaltet werden, wenn die individuellen Lernvoraussetzungen der Studierenden dies erfordern. Aber auch für die Forschung ist dieses Thema interessant: Existiert ein einheitliches Bewertungsschema für die Fallschwierigkeit, können Ergebnisse fallbasierten Lernens einfacher ausgewertet und verglichen werden.

2.2 Fallkomplexität

2.2.1 Definitionen von Fallkomplexität

Das Wort *komplex* wird im alltäglichen und besonders im akademischen Kontext häufig verwendet. In der Definition des Wortes wird es als „umfassend, allseitig“, „vielfältig miteinander verflochten“ und „zusammengesetzt“ (59) charakterisiert. Das Konzept der Komplexität spielt bei verschiedenen Lehransätzen eine wichtige Rolle. Innerhalb des fallbasierten Lernens kann es relevant sein, die

Komplexität eines Falls zu charakterisieren. In der Literatur bestehen dazu verschiedene Ansätze mit dem Ziel, Fälle selbst zu kategorisieren, um sie für Forschungs- und Lehrzwecke zielgerecht einsetzen zu können (2, 60).

Der Begriff *komplex* wird im Kontext verschiedener Lernumgebungen häufig synonym mit *schwierig* verwendet. Für die nähere Betrachtung innerhalb der Ausbildungsforschung ist es allerdings wichtig, die Komplexität als eigenständiges Merkmal zu definieren. Diese ist im Gegensatz zur Fallschwierigkeit lediglich durch das reine Fallmaterial bedingt und unabhängig von anderen Faktoren (2). Eine Person könnte beispielsweise mit einer bestimmten Aufgabe Schwierigkeiten haben, obwohl diese nicht komplex ist und von anderen Lernenden leicht gelöst werden kann.

Oft wird die Komplexität eines Falls durch dessen allgemeinen Informationsgehalt definiert. Ein Fall ist demnach komplexer, wenn er mehr Informationen enthält, die prozessiert werden müssen. Diese Definition ist zwar sehr allgemein gehalten, sie bietet allerdings die Möglichkeit, in Studienszenarien gezielt Aspekte hinzuzufügen oder wegzulassen, um die Komplexität zu steuern (36). Ein weiteres Modell orientiert sich ebenfalls an diesem Grundsatz. Nach Payne wird die Komplexität einer bestimmten Aufgabe durch die verschiedenen Facetten, die sie beinhaltet, charakterisiert. Grundlage für diesen Ansatz ist ebenfalls die Annahme, dass die Auseinandersetzung mit verschiedenen Aspekten einer Problemstellung einen Einfluss auf die Komplexität haben kann (61). Diese Auseinandersetzung besteht konkret aus der Berücksichtigung verschiedener Alternativen, wenn es um die Lösung eines Problems geht. Wenn Personen mit einer Aufgabe konfrontiert werden, bei der sie sich zwischen zwei möglichen Optionen entscheiden müssen, werden sie vermutlich eine ähnliche Fülle an Informationen zu beiden heranziehen und sich anschließend entscheiden. Werden ihnen allerdings komplexere Problemstellungen präsentiert, die aus vielen Alternativen bestehen, werden sie wahrscheinlich nicht für jede Option die gleiche Menge an Informationen betrachten (61). In komplexen Lernumgebungen versuchen Personen, die große Fülle an Informationen herunterzubrechen, um einfacher damit umgehen zu können. Dies kann z.B. dadurch gelingen, dass Problemlösestrategien angewandt werden, um irrelevante Informationen von vornherein auszuschließen und damit den Denkprozess effektiver zu gestalten (62).

2.2.2 Faktoren, die die Fallkomplexität beeinflussen

Ein Ansatz, um die Komplexität von Patienten zu charakterisieren, ist das Vektormodell von Safford et al. (63). Ein komplexer Fall zeichnet sich hier durch zahlreiche Informationen über einen Patienten aus. Diese Merkmale werden im Modell auf verschiedene Achsen verteilt, die dabei gleichzeitig in Verbindung miteinander stehen. Die Achsen repräsentieren die Sozioökonomie, Kultur, Genetik, Umgebung und das Verhalten eines Patienten (63) (vgl. Abbildung 2).

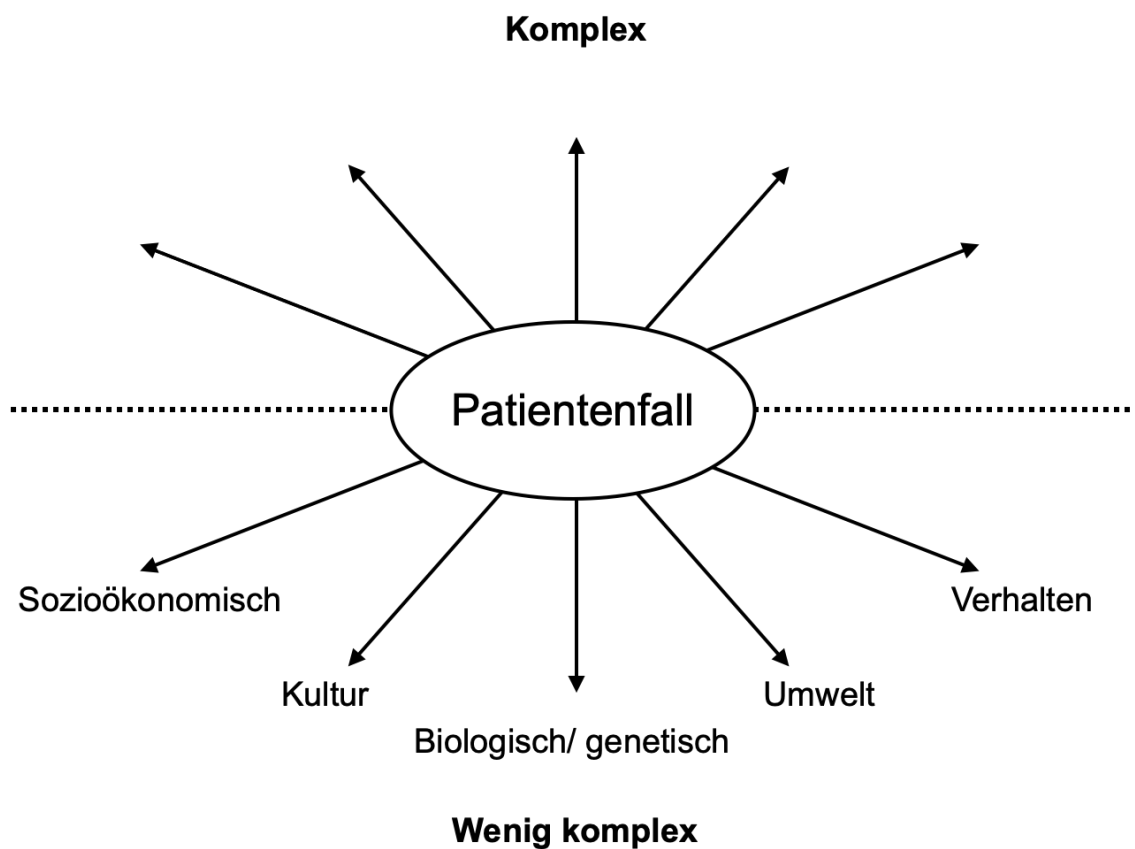


Abbildung 2: Verschiedene Komponenten mit möglichem Einfluss auf die Komplexität eines Patienten aus Safford et al. 2007 (63)

Eine wichtige Schlussfolgerung, die innerhalb dieses Systems gezogen wird, ist, dass Komplexität weit mehr umfasst als die reine Diagnose oder Krankengeschichte eines Patienten (63). Hierbei wird besonders deutlich, dass z.B. die soziale Situation zur Komplexität eines Falls beitragen kann. Angenommen ein Arzt behandelt zwei Patienten mit identischen Diagnosen, die nun entlassen werden sollen – sicherlich fällt es leichter, Patient A in ein stabiles, soziales Umfeld mit finanziellem Rückhalt zu entlassen, als Patient B, der selbst eine kranke Ehefrau

pflegt und durch die sozioökonomische Situation schlechteren Zugang zu medizinischen Leistungen hat. Patient B erfordert demzufolge andere Überlegungen und voraussichtlich auch andere Maßnahmen als Patient A (63). Dieses Modell liefert zwar keine konkrete Definition, charakterisiert Komplexität aber über gewisse Faktoren, die diese beeinflussen. Besonders sozioökonomische und psychosoziale Aspekte scheinen einen relevanten Einfluss auf die Fallkomplexität zu haben und spielen auch in anderen Quellen eine wichtige Rolle (2).

Ein weiterer Ansatz, der sich mit der Komplexität von Lernumgebungen auseinandersetzt, ist die Cognitive Load Theory (CLT), die in der Ausbildungsforschung weit verbreitet ist. Das Modell beschäftigt sich im Allgemeinen damit, wie menschliche Denkprozesse funktionieren und wie Wissen prozessiert wird. Das Erklärungsmodell dient als Basis, um an Lernmethoden zu forschen und diese zielgerichtet anpassen zu können (62).

Ein Ziel der CLT ist, die Fülle an Informationen, mit denen Lernende in komplexen Lernumgebungen umgehen müssen, zu ordnen und letztendlich den sogenannten Cognitive Load zu minimieren, um effektiv lernen zu können (62). Eine Grundannahme ist dabei, dass Kompetenzen dadurch gefördert werden, dass Wissen in verschiedenen Schubladen im Langzeitgedächtnis abgespeichert wird. Diese Schemata bestehen sowohl aus der grundlegenden Ursache eines medizinischen Problems, der Art wie sich die Thematik präsentiert, als auch deren Folgen. Dieses Konzept wird als sogenanntes Illness Script bezeichnet (62, 64) (vgl. Abbildung 3).

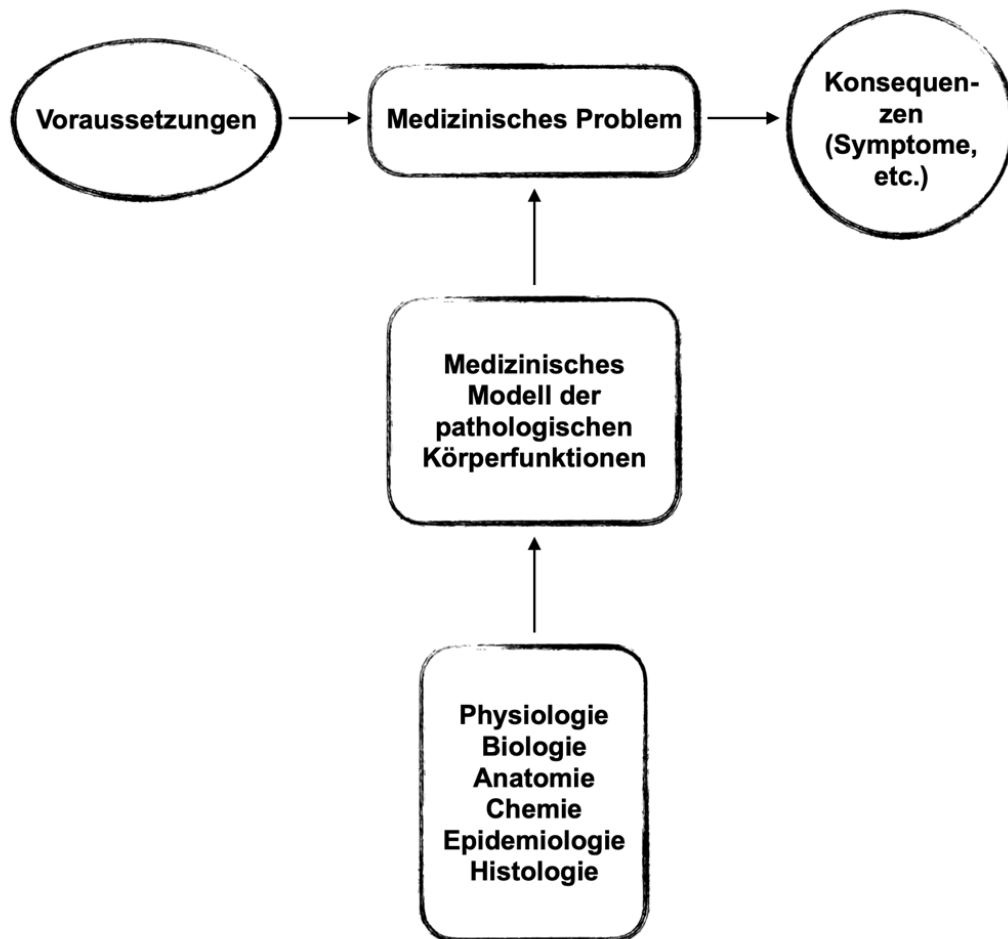


Abbildung 3: Illness Script aus Custers et al. 1998 (64)

Lernen bezeichnet der CLT zufolge das Abspeichern und Prozessieren bestimmter Aspekte innerhalb dieser Illness Scripts. Eine wichtige Annahme dieser Theorie ist, dass die Kapazität des menschlichen Arbeitsgedächtnisses begrenzt ist. Informationen gehen nach kurzer Zeit verloren, es sei denn, sie werden kontinuierlich wiederholt. Haben sie erst einmal das Langzeitgedächtnis erreicht, können Komponenten anhand dieser Schemata, in denen sie verankert wurden, abgerufen werden. Die Schemata selbst sind dabei sehr vielfältig und variieren in ihrer Vielschichtigkeit. Die daraus folgende Automation, mit der Wissen abgerufen wird, ist wichtig, da das Arbeitsgedächtnis selbst überhaupt nicht in der Lage ist, diese Fülle an Informationen zu verarbeiten (62, 64). Für die Praxis bedeutet dies konkret, dass Wissen dadurch entsteht, dass grundlegende Erkenntnisse mit komplexeren Denkprozessen kombiniert werden müssen. Im Medizinstudium wird so z.B. das Wissen über Ursache-Wirkungsbeziehungen in Illness Scripts abgespeichert und so in einen umfangreichen Kontext eingebettet. Diese

Schemata ermöglichen es den Lernenden, in der Zukunft ähnliche Krankheiten zu unterscheiden und so die passende Diagnose zu finden (62).

In Studien konnte gezeigt werden, dass Studierende mit gut ausgearbeiteten Illness Scripts weniger Cognitive Load erleben und bessere Ergebnisse erzielen. Dies könnte daran liegen, dass durch die Illness Scripts schwierige Aufgaben nicht gänzlich neu erarbeitet werden müssen, sondern auf ein einzelnes Modell zurückgegriffen werden kann, was die Anforderung an das Arbeitsgedächtnis reduziert (65).

Innerhalb der CLT werden drei verschiedene Arten der kognitiven Belastung unterschieden: intrinsic load, germane load und extraneous load (62) (vgl. Abbildung 4).

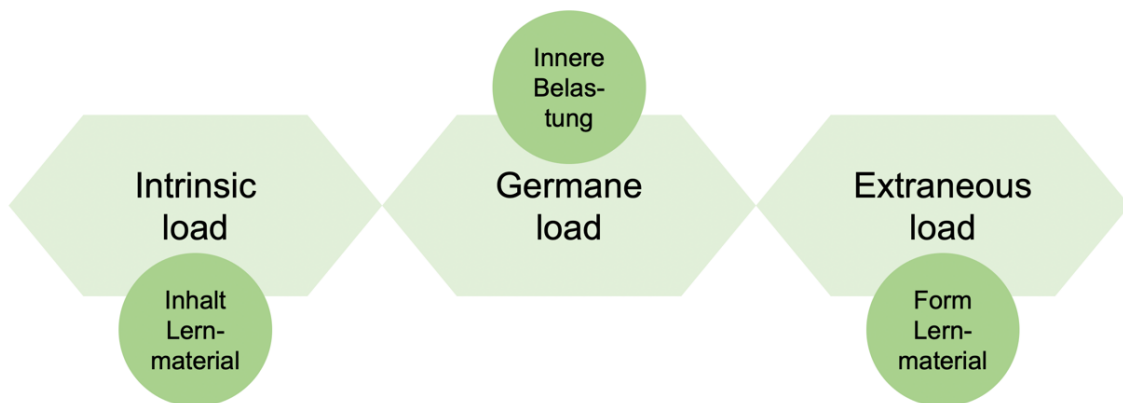


Abbildung 4: Komponenten der CLT

Die Komplexität eines Problems wird dabei vor allem dem intrinsic load zugeordnet. Dieser wird durch das Lernmaterial selbst verursacht und ist demnach durch Instruktoren, die Fälle konstruieren, steuerbar. Der intrinsic load und somit auch die Komplexität können so durch weniger komplexe bis sehr komplexe Fallpräsentationen beeinflusst werden (62). Dies ist für die Ausbildungsforschung von Bedeutung, da die CLT eine Hilfestellung bei der individuellen Anpassung von Lernmaterial darstellen kann.

Die Komplexität hat allerdings nicht nur einen direkten Einfluss auf den intrinsic load, sondern steht möglicherweise auch mit dem germane load in Verbindung. Germane load wird allgemein durch den Lernprozess und den Umgang mit dem intrinsic load verursacht. Er beschreibt vor allem, in welchem Ausmaß das

Arbeitsgedächtnis beansprucht wird, um den intrinsic load zu prozessieren. Der germane load ist dabei direkt mit dem Lernvorgang verbunden. Inwiefern neue Informationen prozessiert und in vorhandene Illness Scripts implementiert werden, beeinflusst den germane load und hat eine Auswirkung auf den Lernprozess (62). Generell könnte das Maß an Komplexität innerhalb eines Falls beeinflussen, inwiefern Informationen prozessiert werden (2) und so auch den germane load verändern.

Um den Lernprozess möglichst effektiv zu gestalten, sollte bei der Auswahl von Lehrmaterial das Ziel verfolgt werden, den germane load zu optimieren. Dies kann dadurch erreicht werden, dass der Lernstoff durch möglichst unterschiedliche Aufgabenstellungen und Fälle präsentiert wird. Durch den Vergleich verschiedener Fallpräsentationen können Illness Scripts auf Seiten der Lernenden etabliert werden (62). Besonders für die Medizin ist dieses Vorgehen wichtig, da so die Fähigkeit gefördert wird, ähnliche Fälle in Verbindung miteinander zu bringen oder neue Fallpräsentationen von bisherigen abzugrenzen. Dadurch ist es möglich, die Fähigkeit zu erlernen, zwischen verschiedenen Differentialdiagnosen zu unterscheiden. Dies stellt eine Grundlage für die korrekte Behandlung von Patienten dar. Praktizierende Mediziner werden mit zunehmender Berufserfahrung mit immer mehr klinischen Fällen konfrontiert – so steigt auch ihre Expertise.

Bei Studierenden, die über wenig praktische Erfahrung verfügen, kann die Variierung des Lernmaterials zu einer Zunahme des intrinsic load führen. Dies liegt daran, dass sie zunächst jede Information einzeln prozessieren müssen, bevor sie gelernt haben, welcher Aspekt welchem Schema zuzuordnen ist. Somit kann erzielt werden, dass Denkprozesse angestoßen werden und der germane load optimiert wird (62). Diese Erkenntnisse tragen einen großen Teil dazu bei, sich dem Konzept der Fallkomplexität zu nähern. Nach der CLT kann diese zusammengefasst als Ursache von intrinsic load eingeordnet werden und steht gleichzeitig in Verbindung mit dem germane load (62) (vgl. Abbildung 5).



Abbildung 5: Komplexität innerhalb der CLT

Für eine zielgerichtete Ausbildung könnte es interessant sein, den intrinsic load durch Steuerung der Komplexität zu variieren. Versteht man diese als Fülle der verschiedenen Aspekte eines Falls, hat das Weglassen oder Hinzufügen von Eigenschaften möglicherweise ebenfalls einen Einfluss auf den germane load und somit die Effektivität des Lernprozesses.

Trotz einiger Hinweise auf den Zusammenhang von Komplexität und bestehenden Konzepten innerhalb der Ausbildungsforschung (60, 62) ist nicht abschließend geklärt, inwiefern die Komplexität konkret gesteuert oder gemessen werden kann (2).

In einem anderen Ansatz wird die Komplexität durch die Typizität eines Falls operationalisiert. Konkret bedeutet dies, dass wenig komplexe Fälle Merkmale aufweisen, die typisch für eine bekannte Diagnose sind. Im Gegensatz dazu zeichnen sich komplexe Fälle dadurch aus, dass sie über Informationen verfügen, die sich nicht unter eine bestimmte Diagnose subsummieren lassen (65). Dieser Ansatz ist durchaus kritisch zu sehen, zumal in anderen Quellen die Falltypizität als ein eigenständiges Merkmal dargestellt wird (66) und so keine eindeutige Definition der Komplexität möglich ist. Die Autoren stellen in der Publikation selbst klar, dass die Komplexität von Fällen ein vielschichtiges Konzept ist und andere Modelle überprüft werden sollten, um diese zu charakterisieren (65).

2.2.3 Möglichkeiten der Erfassung von Fallkomplexität

Bereits 1984 führte Hennen ein Punktesystem für die Bestimmung der Komplexität von Fällen ein. Die Annahme, die seinem Vorschlag zugrunde liegt, ist, dass die Komplexität aus der Summe verschiedener Merkmale des Falls resultiert. Sie wird somit als ein eigenständiges Modell angesehen, das durch den Fall selbst verursacht wird. Der Autor sieht dies als Basis für den anschließenden Prozess der Bearbeitung des Problems (60).

Das Bewertungssystem basiert auf fünf Komponenten mit möglichem Einfluss auf die Fallkomplexität: Symptome, körperliche Untersuchung und technische Befunde, Sozioökonomie und Verhalten, Diagnose und zuletzt das Management des Patienten. Für jeden dieser Komponenten können bis zu drei Punkte vergeben werden. Je höher die Punktzahl, desto komplexer sind die Angaben. Die Summe ergibt den sogenannten Komplexitätsindex. Für die Punktevergabe gibt es für jede einzelne der fünf Kategorien bestimmte Kriterien (60) (vgl. Tabelle 4).

Tabelle 4: Punktesystem für die Fallkomplexität aus Hennen 1984 (60)

Symptome	Einzelnes Symptom	Mehrere Symptome	Mehrere unabhängige Symptome
Punkte	1	2	3
(Körperliche) Untersuchungen	Nicht für Diagnose relevant/ kongruent zu Symptomen	Befunde nicht wegweisend	Matrix aus Befunden notwendig
Punkte	1	2	3
Sozioökonomie und Verhalten	Patient und Familie unauffällig	Patient oder Familie auffällig	Kontext hat relevanten Einfluss auf Patienten
Punkte	1	2	3
Diagnose	Eindeutige Diagnose	Eine Diagnose mit Unstimmigkeiten oder mehrere Diagnosen	Mehrere Diagnosen, Unstimmigkeiten bei mind. einer Diagnose
Punkte	1	2	3
Management	Unidimensional (1 Medikament/ 1 Operation...)	Kombination erforderlich	Kombination und Management von Komorbiditäten erforderlich
Punkte	1	2	3

Dieses System verfolgt somit nicht nur einen quantitativen Ansatz, sondern erfordert, dass die einzelnen Komponenten eines Falls qualitativ analysiert werden. Insgesamt spielen bei der Anwendung dieses Systems verschiedene Dimensionen der Informationen eine Rolle. Für den Autor ist ein Fall dann besonders komplex, wenn er nicht nur viele Komponenten enthält, sondern auch Aspekte, die nicht mit den vorherigen verknüpft sind. Dadurch werden zusätzliche Informationsebenen bzw. Dimensionen eröffnet (60).

Im Gegensatz zu anderen Modellen, die auf retrospektiven Ansätzen beruhen, bot der Ansatz von Hennen erstmals die Möglichkeit, die Komplexität von Fällen a priori zu bestimmen (60). Dies deutet auf das große Potential beim Einsatz für die Lehre und innerhalb der Ausbildungsforschung hin. Aus diesem Grund haben weitere Autoren diesen Ansatz aufgegriffen und adaptiert.

In einem weiteren Vorschlag für ein Punktesystem von Braun et al. wird die Fallkomplexität ebenfalls als eigenständiger Parameter interpretiert, der direkt durch das Fallmaterial bedingt ist. Aus diesem Grund ist die Komplexität unabhängig vom jeweiligen Bearbeiter eines Falls. Für die Etablierung eines Punktesystems darf sie vor allem nicht synonym mit anderen Faktoren (insbesondere der Fallschwierigkeit) gebraucht werden (2). Um die Fallkomplexität in simulierten Lernumgebungen bestimmen zu können, schlagen die Autoren ähnlich wie Hennen ein System vor, das auf Basis bestimmter Kategorien Punkte vergibt. Ziel soll sein, den Parameter der Fallkomplexität messbar und vor allem in Forschungsszenarien vergleichbar zu machen. In diesem Modell werden Punkte für fünf Kategorien eines medizinischen Falls vergeben (2) (vgl. Tabelle 5).

Tabelle 5: Scoring-System zur Bestimmung der Fallkomplexität aus Braun et al. 2019 (2)

Anamnese	Informationen vorhanden	1 Punkt
Körperliche Untersuchung	Informationen vorhanden	1 Punkt
Technische Untersuchungen	Informationen vorhanden	1 Punkt
Psychosoziale Aspekte	Informationen vorhanden	1 Punkt
Nebendiagnose	Vorhanden	1 Punkt

Pro Information, die einer Kategorie zugeordnet werden kann, wird ein Punkt in dieser vergeben. Zusätzliche Punkte können vergeben werden, wenn Informationen das Kriterium der Linearität nicht erfüllen. Für die Autoren bedeutet dies, dass Aspekte, die nicht zu vorherigen passen und zusätzliche Informationsebenen eröffnen, die Komplexität eines Falls erhöhen. Dies kommt zustande, wenn neue Informationen mit anderen Befunden interagieren. Dabei spielt es keine Rolle, ob es sich um physiologische oder auffällige Aspekte handelt. Maximal

kann ein Fall so zehn Punkte für die Komplexität erreichen. Diese Begrenzung basiert auf der Annahme, dass ein Fall nicht unendlich komplex sein und mehr als sechs zusätzliche Ebenen enthalten kann (2).

Obwohl die vorgestellten Ansätze und die entsprechenden Punktesysteme die Erfassung der Komplexität teilweise unterschiedlich umsetzen, stützen sie alle die Forderung nach einem einheitlichen Konzept für das Lernen mit Fällen. Leider bestehen aktuell wenig Ansätze, die sich diesem Problem widmen. Sicher ist jedenfalls, dass ein einheitliches Konzept die Erstellung von Fällen, die Interpretation von Forschungsergebnissen und die Vergleichbarkeit gewisser Standards innerhalb der Lehre erleichtern würde. Diese Arbeit soll zu diesem Vorhaben beitragen.

2.3 Fallschwierigkeit und Fallkomplexität im Verhältnis

Obwohl die Beziehung zwischen der Komplexität und der Schwierigkeit eines medizinischen Falls bislang wenig erforscht ist, legen frühere Abhandlungen eine mögliche Wirkbeziehung nahe (2). Eine Studie, die unter anderem die Schwierigkeit einzelner Fälle miteinander vergleicht, nennt die Komplexität als eine Ursache der beobachteten Unterschiede (67). Hierbei wurden vier Fälle anhand ihrer Lösungswahrscheinlichkeit (von leicht bis schwierig) kategorisiert. Die richtigen Diagnosen der Fälle waren: Aortenstenose, Urämie, Hyperventilation und AV-Knoten Reentrytachykardie. Die vierte Fallpräsentation wurde lediglich von fünf Prozent der Teilnehmenden richtig gelöst und war somit am schwierigsten. Dieser Fall beinhaltete ein komplexes Elektrokardiogramm, das interpretiert werden musste. Gleichzeitig ist die AV-Knoten Reentrytachykardie vergleichsweise selten, dadurch wird ebenfalls die Schwierigkeit beeinflusst (67). Dieses Beispiel zeigt, dass es durchaus möglich ist, dass ein einzelner komplexer Bestandteil eines Falls dessen generelle Schwierigkeit erhöht. Trotzdem vermuten die Autoren, dass die Lösungswahrscheinlichkeit besonders durch das Vorwissen der Studierenden beeinflusst wird (67). Diese Studie macht zwar wichtige Beobachtungen, lässt aber keine Aussagen über die Kausalität oder genaue Wirkbeziehungen zu.

Im Rahmen psychologischer Forschung wurde bereits nachgewiesen, dass Komplexität und Schwierigkeit miteinander in Verbindung stehen (68, 69). Computersimulationen, sogenannte microworlds, werden dabei oft verwendet, um ein

Abbild der realen Welt zu konstruieren. Microworlds bestehen aus bestimmten Variablen, die miteinander interagieren und somit in Verdacht stehen, die Komplexität einer konstruierten Umgebung zu beeinflussen (50). Es konnte gezeigt werden, dass hierbei die Schwierigkeit gewisser Szenarien mit deren Komplexität steigt (68, 69). Da dieses Konzept der Simulation auch mit medizinischen Fällen verfolgt wird (40), ist es denkbar, dass auch hierbei eine solche Wirkbeziehung besteht.

Diese Quellen lassen vermuten, dass die Komplexität einen Einfluss auf die Schwierigkeit eines Falls ausübt. Trotzdem gibt es Anhaltspunkte dafür, dass auch noch weitere Faktoren zur Fallschwierigkeit beitragen (vgl. Abbildung 6).

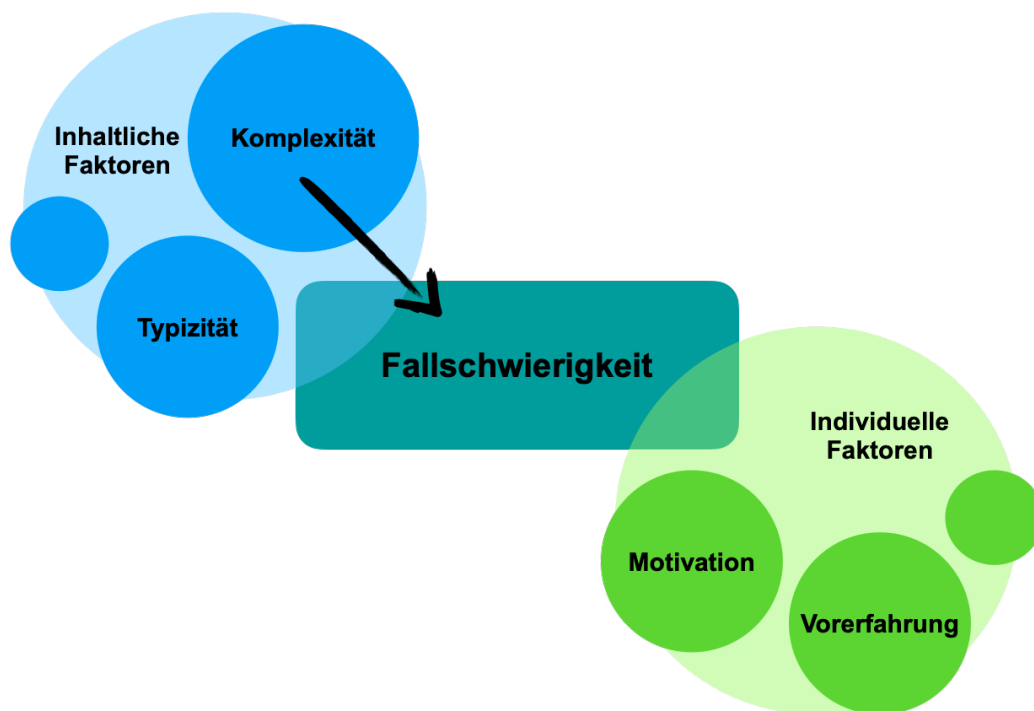


Abbildung 6: Faktoren mit möglichem Einfluss auf die Fallschwierigkeit

Ein Grund, wieso dieser Zusammenhang in der Vergangenheit nicht zufriedenstellend geklärt wurde, ist sicherlich, dass Schwierigkeit und Komplexität je nach Autoren und Kontext unterschiedlich definiert werden. Eine Operationalisierung der Fallschwierigkeit und der Fallkomplexität soll deshalb Voraussetzung für die Bearbeitung der Forschungsfragen dieser Arbeit darstellen.

3 Problemstellung und Forschungsfragen

3.1 Inwiefern kann die Fallkomplexität im medizinischen Kontext reliabel und valide erfasst werden?

Das Lernen anhand von realen Fällen und Fallbeispielen ist ein etabliertes Konzept in der medizinischen Ausbildung (70). Die Relevanz fallbezogener Lehre wird vor allem durch die Prüfungsinhalte des zweiten Staatsexamens (M2) deutlich. Die Approbationsordnung legt fest, dass die schriftliche Prüfung fallorientiert gestaltet werden soll, um somit besonderen Fokus auf problembasierte Aufgabenstellungen legen zu können (11). Für die Umsetzung und Konstruktion fallbezogener Prüfungsszenarien gibt es viele verschiedene Ansätze. Eine dieser Möglichkeiten stellen sogenannte Key-Feature (KF) Fragen dar. Diese bestehen meist aus einem kurzen medizinischen Fall, gefolgt von mehreren Fragen, die sich auf die entscheidenden Aspekte des Problemlöseprozesses konzentrieren. Vorteil dieser Methode ist, dass sie vergleichsweise einfach an die Anforderungen der Lernenden angepasst und formal verändert werden kann (71). Außerdem gibt es Hinweise darauf, dass diese Methode im Vergleich zu langen Fällen, die lediglich eine multiple choice Antwort abfragen, überlegen sein könnte (72). Prof. Dr. med. Martin Fischer, Leiter des Instituts für Didaktik und Ausbildungsforschung in der Medizin (DAM) an der LMU München begründet das Potential des KF-Formats insbesondere darin, dass die Fragen nicht nur theoretische Inhalte prüfen, sondern die Handlungskompetenz stärken können (73). Ein Ziel der medizinischen Ausbildung soll es schließlich sein, zukünftige Ärzte auf den Berufsalltag vorzubereiten. Die vermehrte Einbindung dieses Prüfungsformats könnte demnach in Zukunft relevant sein. Aus diesem Grund arbeitet das Institut für medizinische und pharmazeutische Prüfungsfragen (IMPP) seit 2018 daran, KF-Aufgaben vermehrt in den zweiten Abschnitt der Ärztlichen Prüfung einzubauen (73).

Fallbasierte Formate sind jedoch nicht nur für die Prüfung von Kompetenzen in Examina, sondern auch zu deren Erwerb wichtig. Obwohl diese Formate von zahlreichen Disziplinen für das Training und Assessment in der Ausbildung verwendet werden, ist oft unklar, in welchem Ausmaß diese dem Stand aktueller Forschung entsprechen. Seit Jahren bestehen deshalb Forderungen, dass besonders in komplexen Lernumgebungen wie der Medizin Wert auf

wissenschaftlich basierte, moderne Lehr-Lernmethoden gelegt werden sollte (52). Fallbasiertes Lernen sollte zum richtigen Zeitpunkt im Studium und in der richtigen Form präsentiert werden. Oft ist dabei der wichtigste Aspekt für die Auswahl von Fällen die Thematik oder das Fachgebiet, das aktuell unterrichtet wird. Fälle, die Studierende während ihrer Ausbildung bearbeiten sind entweder fest in das Curriculum integriert oder stellen freiwillige Zusatzangebote dar. Diese Zusatzangebote werden zwar zum Teil angenommen, oft allerdings auch nicht bearbeitet. Ein möglicher Grund dafür könnte sein, dass das dargebotene Fallmaterial nicht optimal an die Fähigkeiten der Lernenden angepasst wurde.

Um dieses Problem zu vermeiden, müssen bestimmte Kriterien bei der Auswahl eines medizinischen Falls beachtet werden. Der Aspekt, dem bei dieser Arbeit die größte Aufmerksamkeit geschenkt werden soll, ist die Fallkomplexität. Diese wurde zwar in früheren Abhandlungen diskutiert (2, 60, 63), dennoch besteht bisher kein etabliertes System, um die Komplexität zu quantifizieren.

Das erste Ziel dieser Arbeit ist es, ein Scoring-System zu entwickeln, mit dem die Komplexität verschiedenster medizinischer Fälle reliabel erfasst werden kann und dieses im Anschluss zu validieren.

3.2 Inwiefern lässt sich die Fallschwierigkeit mit Hilfe der Fallkomplexität vorhersagen?

Während des Medizinstudiums werden Studierende mit einer Vielzahl an Fällen unterschiedlicher Schwierigkeit konfrontiert. Um den Lernprozess optimal zu fördern, wäre dabei eine systematische Anpassung der Fallschwierigkeit an den jeweiligen Lernstand der Studierenden wünschenswert (40). Dieser kann allerdings selbst zwischen Absolventen des gleichen Semesters deutlich variieren. Gründe dafür könnten individuelle Erfahrungen außerhalb des Studiums, die persönliche Eignung oder auch die individuelle Motivation sein. Diese Faktoren verdeutlichen, dass es eine Herausforderung darstellt, die angemessene Fallschwierigkeit für jeden individuell auszuwählen.

Insgesamt bestehen eine Vielzahl an Faktoren, die potenziell einen Einfluss auf die Schwierigkeit von Fällen ausüben. Hinzu kommt, dass die Begriffe Schwierigkeit und Komplexität häufig synonym verwendet werden (2).

Aus diesem Grund widmet sich diese Arbeit einer Operationalisierung beider Konzepte und grenzt sie voneinander ab. Dabei soll insbesondere untersucht werden, ob eine Wirkbeziehung beider Komponenten besteht und ob die Fallkomplexität als Prädiktor für die Fallschwierigkeit dienen kann.

4 Methoden

Im Folgenden wird der Weg zur Entwicklung eines neuen Scoring-Systems, seine Anwendung und die Methodik zur Überprüfung des Modells dargestellt. Die Konzeption des Systems wurde im Sommersemester 2020 und Wintersemester 2020/2021 durchgeführt. Für die Anwendung und die statistische Auswertung erstreckte sich der zeitliche Rahmen von Sommersemester 2021 bis Sommersemester 2022.

4.1 Qualitative Betrachtung der Fallkomplexität

4.1.1 Anwendung eines bestehenden Modells

Wie zuvor dargestellt (vgl. S. 30, f.), haben Braun et al. 2019 ein Modell zur Messung der Fallkomplexität in einer Lernumgebung entwickelt. Dieses Scoring-System wird in der Originalarbeit anhand eines beispielhaften Falls erläutert, es folgt jedoch keine Anwendung auf weitere Fälle. Die Autoren weisen allerdings darauf hin, dass die Prüfung der Anwendbarkeit des Modells Gegenstand zukünftiger Forschung sein sollte (2).

Um dieses System im Rahmen der Untersuchung der Fallkomplexität zu erproben, wurde es nun auf 12 Fallvignetten angewandt. Diese stammen von vorherigen Studien am DAM (2, 41, 42, 67, 74) und wurden im online Lernprogramm CASUS bearbeitet. Alle diese Fälle weisen das Leitsymptom Atemnot auf und haben die folgenden Diagnosen: Aortenstenose, Urämie, Hyperventilation, AV-Knoten-Reentrytachykardie, COPD, Tuberkulose, Myokarditis, Pneumothorax mit Infekt, Asthma, Legionellose, Bronchialkarzinom und Sarkoidose. Diese von den Autoren der Fälle vorgegebenen Diagnosen waren zum Zeitpunkt der Fallbearbeitung nicht bekannt und wurden nachträglich recherchiert. Die Fälle sind einheitlich strukturiert, indem bei allen die gleichen Untersuchungen, wie z.B. Röntgen, Urinuntersuchung oder Lungenfunktion abrufbar sind. Abbildung 7 zeigt

einen dieser Fälle, dargestellt auf der Lernplattform CASUS. Sämtliche Informationen des Falls konnten der Reihenfolge nach über die linke Navigationsleiste abgerufen werden.

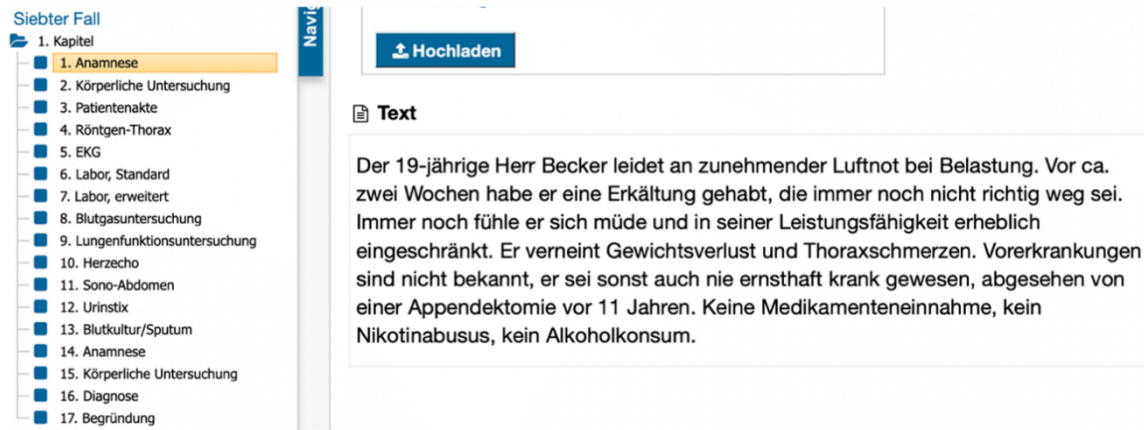


Abbildung 7: Fall (Myokarditis) in der Lernumgebung CASUS (41, 74), vollständiger Fall in Anhang #2

Diese 12 Fälle wurden von zwei Personen unabhängig voneinander bearbeitet. Person bzw. Rater Nummer eins, war Dr. Dr. Marc Weidenbusch, während das zweite Rating von Kim Öhler vorgenommen wurde. Diese Konstellation wurde gewählt, um zwei verschiedene Erfahrungsschätze und Wissensniveaus (Facharzt und Studentin) abzubilden. Dadurch sollte ein erster Eindruck entstehen, ob das Scoring-System von verschiedenen Personen, unabhängig ihres Fachwissens und ihrer Berufserfahrung, angewandt werden kann. Das Modell wurde bereits während der Anwendung auf bestimmte inhaltliche Aspekte hin untersucht. Besondere Aufmerksamkeit galt dabei der Übersichtlichkeit und Anwendbarkeit des Scoring-Systems, sowie der Vergleichbarkeit der Ergebnisse.

Um diese Kriterien zu betrachten, wurde bei der Anwendung darauf geachtet, die Vorgaben der Autoren genau einzuhalten. In einem ersten Schritt wurden Punkte für die vier ersten Kategorien des Systems: Anamnese, körperliche Untersuchung, technische Untersuchungen und psychosoziale Aspekte vergeben. Dabei konnte sich anhand der Navigationsleiste innerhalb des Falls orientiert werden und Informationen somit leicht gefunden und zugeordnet werden, ohne dass für jede einzelne Kategorie der ganze Fall abgesucht werden musste. Gemäß den Vorgaben der Autoren war die Punktevergabe durch einen zweistufigen Prozess gekennzeichnet. Zuerst wurde beurteilt, ob Informationen, passend zu der

jeweiligen Kategorie, identifiziert werden konnten. War dies der Fall, wurde ein Punkt in dieser Kategorie vergeben. In einem zweiten Schritt wurde festgelegt, ob Informationen in derselben Kategorie folgen, die die erste Linearitätsebene verlassen. Zusätzliche Punkte wurden somit beispielsweise vergeben, wenn Informationen auf neue Differentialdiagnosen hinwiesen oder nicht dem vorherigen Symptomkomplex zugeordnet werden konnten. Anhand dieses Schemas wurde Kategorie für Kategorie beurteilt und die Punkte notiert. Im Anschluss daran wurde für jeden der Fälle untersucht, ob dieser eine Nebendiagnose enthält. Dafür mussten alle Informationen, besonders die Symptome und Untersuchungsergebnisse, herangezogen werden, um eine Verdachtsdiagnose, sowie mögliche Nebendiagnosen zu stellen. Konnte eine Nebendiagnose identifiziert werden, erhielt ein Fall einen zusätzlichen Punkt. Durch Addition der einzelnen Punktwerte wurde für jeden der 12 Fälle eine Gesamtpunktzahl für die Komplexität generiert. Dieses Vorgehen diente primär, um das Scoring-System zu erproben und mögliche Hürden bei der Anwendung zu identifizieren. Im Anschluss wurde die Übereinstimmung bei der Nutzung des Scoring-Systems durch die beiden Rater untersucht. Hierfür wurde die Interrater-Reliabilität als ein Maß der Übereinstimmung verwendet. Diese wurde sowohl für jede einzelne der Kategorien (Anamnese, körperliche Untersuchung, technische Untersuchungen, psychosoziale Aspekte und Nebendiagnose) als auch für die Gesamt-Scores anhand von Cohen's Kappa berechnet.

4.1.2 Weiterentwicklung zu einem neuen Scoring-System CASE

Auf Basis der Anwendung des Systems nach Braun et al. (2) wurde ein neues Scoring-System entwickelt. Dieses soll vorrangig dazu dienen, die Komplexität bestehender Fälle zu beurteilen, daran orientiert sich auch dessen Namensgebung **CASE: ComplexityAssessmentStructure**.

Das strukturelle Design des neuen Scoring-Systems wurde anhand etablierter Strukturen der medizinischen Ausbildung und der klinischen Praxis entwickelt. Bei der Behandlung eines Patienten nehmen zwei Komponenten eine elementare Rolle ein: Das Finden der richtigen Diagnose und das Bereitstellen der passenden Therapie (70). Um diese Fertigkeiten zu erlernen, hebt die ÄApprO besonders die Anamnese und die körperliche Untersuchung als elementare Bausteine der Lehre hervor (11). Von diesem Gesichtspunkt aus erscheint es

sinnvoll, einen medizinischen Fall auf Basis dieser beiden Elemente zu entwickeln. Da ein wichtiges Ziel bei der Konstruktion eines Scoring-Systems die einfache Anwendbarkeit darstellt, wurde auch dieses in einen Anamnese- und Untersuchungsteil gegliedert. In der modernen Medizin stehen neben der körperlichen Untersuchung eine Vielzahl an technischen und apparativen Methoden, wie beispielsweise Ultraschall, funktionelle Tests oder Laboruntersuchungen zur Verfügung. Diese finden ebenfalls in gesonderten Untergruppen Platz im Scoring System.

Um den Informationsgehalt eines medizinischen Falls zu ordnen und bewerten zu können, wurde ein multi-dimensionales System entwickelt. Dieses ermöglicht, den Inhalt eines Falls auf drei verschiedenen Ebenen einzuordnen: Dimensionen, Kategorien und Klassen (vgl. Abbildung 8).

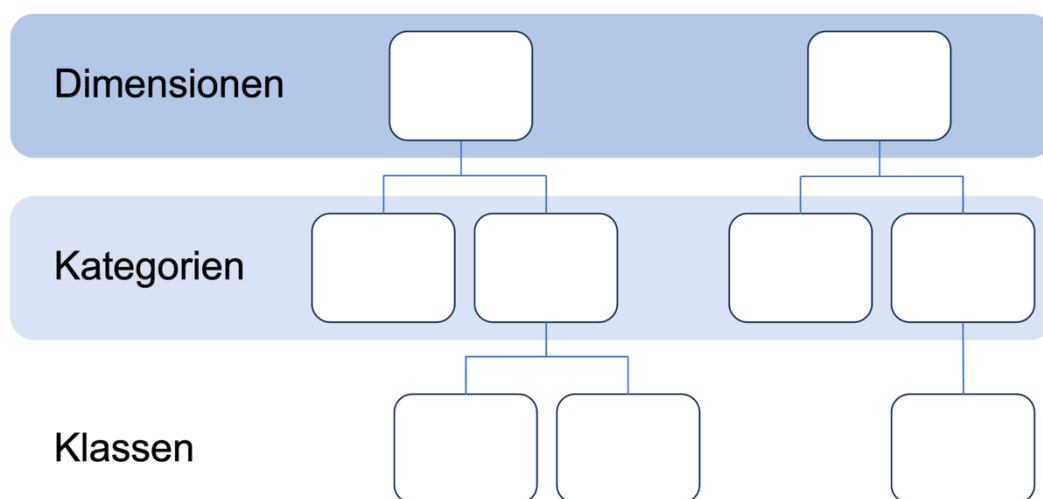


Abbildung 8: Hierarchische Gliederung des neuen Scoring-Systems CASE

Die erste Dimension stellt die Anamnese dar. Eine Analyse deutscher Dokumentationsbögen in der klinischen Ausbildung zeigt, dass in der Mehrzahl der Fälle die folgenden sechs Komponenten eine wichtige Rolle in der Anamnese spielen: die Geschichte der aktuellen Erkrankung, Vordiagnosen, Medikamente, die psychosoziale Anamnese, Allergien und eine Familienanamnese (23). Diese Aspekte wurden komplementiert durch die Vorbehandlung und eine Reiseanamnese. Diese acht Komponenten stellen die zur Anamnese zugehörigen Kategorien dar. In der psychosozialen Anamnese, Reiseanamnese und Familiengeschichte wird hierbei die jeweilige Information genauer kategorisiert und

zusätzlich zwischen einer beliebigen und einer relevanten Information unterschieden. Wird eine Person im Fall beispielsweise als nervös und ängstlich beschrieben, erfordert dies voraussichtlich differenziertere Denkprozesse beim Lernenden, verglichen mit einem Patienten, der als freundlich charakterisiert wird. Dementsprechend werden durch auffällige Befunde auch mehr Punkte im Scoring-System erzielt. Nicht nur das Verhalten eines Patienten, sondern auch emotionale Beziehungen und der Lebensstil sollten in die psychosoziale Anamnese miteinbezogen werden. Hierbei spielen besonders krankheitsbezogene Risikofaktoren eine wichtige Rolle. Um zwischen beliebigen und relevanten Aspekten zu unterscheiden, kann evidenzbasierte Literatur herangezogen werden, um bestimmte Risikofaktoren und Verhaltensweisen zu klassifizieren.

Für die Kategorie der Reiseanamnese ist es hilfreich, sich bei dem Auswärtigen Amt oder einem Tropeninstitut zu informieren, welche potenziell gefährdenden Bedingungen in der jeweiligen Region vorherrschen.

Die Familiengeschichte ist hauptsächlich relevant, um genetische Risikofaktoren zu identifizieren. Dabei kann ein festes Schema helfen, um zwischen beliebigen und relevanten Informationen zu unterscheiden. Abbildung 9 stellt einen Vorschlag für ein solches Schema dar. In einem ersten Schritt müssen die Erkrankungen innerhalb der Familie erfasst werden. Anschließend wird entschieden, ob genetische Risikofaktoren für eine bestimmte Erkrankung bekannt sind. Wird dies mit nein beantwortet, erhält die Information einen Punkt für die Familienanamnese. Bei einem ja oder bei Unsicherheit sollte evidenzbasierte Literatur herangezogen werden, um sich zu informieren. Besteht Konsens über die genetischen Risikofaktoren, erhält der Aspekt insgesamt zwei Punkte für die Familiengeschichte. Ein Beispiel für einen Punkt ist die durchgemachte Varizelleninfektion der Großmutter. Diese Infektionskrankheit ist häufig und beinhaltet keinerlei Hinweise auf eine genetische Disposition. Zwei Punkte dagegen würden bei einer Chorea Huntington Erkrankung der Mutter vergeben werden.

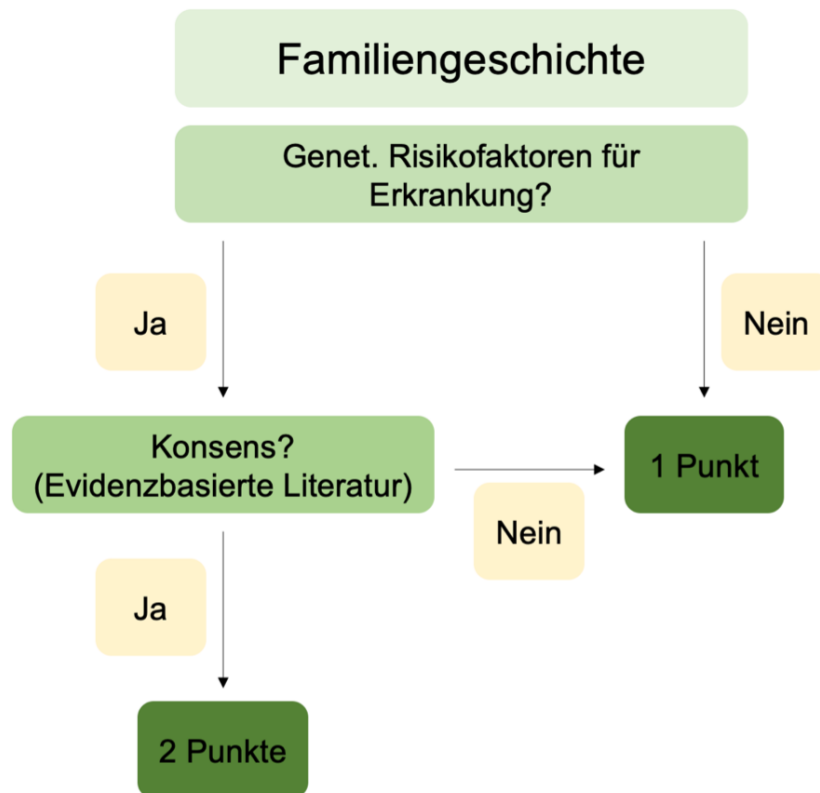


Abbildung 9: Schema zur Identifikation relevanter familiärer Risikofaktoren

Die 9. Kategorie in der Dimension der Anamnese stellt die Zuordnung der Symptome zu den jeweiligen Organsystemen dar. Diese bilden die dritte Ebene des Scorings, die Klassen. Um das Manual einheitlich zu gestalten, finden sich die folgenden Klassen im gesamten Scoring-System: Allgemein, Kopf und Hals, kardiovaskuläres System, respiratorisch, gastrointestinal, nephro und urogenital, muskuloskelettal, neurologisch, psychiatrisch, endokrin, hämato-onkologisch und Haut. Viele dieser Organsysteme werden beispielsweise in Dokumentationsbögen für die medizinische Ausbildung verwendet (23) oder sind Teil der klinischen Dokumentation der Anamnese und anderen Befunden. Die Adaptation dieser Organsystemklassen dient einerseits der einheitlichen Strukturierung des Systems, andererseits wird so eine Anpassung an bekannte Strukturen der Praxis gewährleistet. Die vorgeschlagenen Organsysteme sollen vorrangig dazu dienen, die Informationen eines Falls sinnvoll kategorisieren zu können. Dabei kommt es vor, dass gewisse Symptome nicht nur durch eine Anamnese, sondern auch durch die körperlichen Untersuchung identifiziert werden können. Für eine einfache Zuordnung innerhalb des Scoring-Systems gilt der Grundsatz: Berichtet der Patient von Beschwerden, werden diese unter der Anamnese

zusammengefasst. Erhält man dagegen Befunde während der körperlichen Untersuchung, erfolgt die Zuordnung zu dieser Kategorie.

Die zweite Dimension, der alle Arten von Untersuchungen zuzuordnen sind, beinhaltet folgende vier Kategorien: Bildgebung, Laboruntersuchungen inklusive klinischer Chemie, Mikrobiologie und Pathologie, die körperliche Untersuchung und funktionelle Tests. Jede dieser Kategorien wird wie die vorherigen durch die 12 Organsystemklassen gegliedert und beinhaltet ebenfalls, wie einzelne Kategorien der Anamnese, eine zusätzliche Unterscheidung. Diese wird bei der jeweiligen Untersuchung zwischen einer Untersuchung, die angeordnet wurde und auffälligen Ergebnissen getroffen (vgl. Tabelle 6).

Während der Entwicklung des Manuals wurden regelmäßig zufällig ausgewählte Fälle herangezogen, um die Anwendbarkeit, insbesondere der Organsystemklassen, einzuschätzen.

Tabelle 6: Übersicht über das neue Scoring-System CASE

Dimensionen	Kategorien	Klassen
Anamnese	Geschichte aktuelle Erkrankung	
	Vordiagnosen	
	Vorbehandlung	
	Medikamente	
	Psychosoziale Anamnese	Beliebig/ relevant?
	Allergien	
	Familienanamnese	Beliebig/ relevant?
	Reiseanamnese	Beliebig/ relevant?
	Symptome nach	Organsystemen
Untersuchungen	Bildgebung	Organsysteme Beliebig/ relevant?
	Labor	
	Körperliche Untersuchung	
	Funktionelle Tests	

Die erste Version des Scoring-Systems wurde anhand eines binären Zahlenschemas entwickelt. Hierbei wurde lediglich differenziert, ob eine oder mehrere zu einer bestimmten Klasse zugehörige Informationen genannt wurden oder nicht. Für die Praxis bedeutet dies, dass in einer Klasse maximal zwei Punkte

erzielt werden können, unabhängig von der Anzahl an Untersuchungen oder pathologischen Ergebnissen. Im Beispiel: *Sie erheben bei Herrn Mustermann folgende Befunde: Cor rein und rhythmisch, Strömungsgeräusche über den Karotiden auskultierbar, Herzfrequenz 124/min, Blutdruck 145/90* würde somit ein Punkt für die Untersuchung des kardiovaskulären Systems und ein zusätzlicher für diverse Auffälligkeiten erreicht werden. Ein bedeutsames Merkmal dieser Methode stellt die Begrenzung der Maximalpunktzahl dar. Daraus lässt sich der Schluss ziehen, dass der komplexeste vorstellbare Fall mindestens eine Information in jeder Kategorie plus mindestens einer Auffälligkeit in derselben enthält. Die theoretisch erreichbare Maximalpunktzahl für die Komplexität liegt hierbei bei 167.

Anhand der Anwendung des Manuals auf einen Discovery Datensatz, der 15 zufällig ausgewählte Fälle enthielt, zeigte sich schnell, dass die oben genannte Version in der Praxis nicht anwendbar ist. Der erste wichtige Aspekt ist, dass das Scoring nicht durch eine Maximalpunktzahl begrenzt werden darf. Für die Praxis der medizinischen Lehre sollte es möglich bleiben, immer komplexere Fälle zu entwickeln, angepasst an den Lernfortschritt des Individuums. Selbst langjährig klinisch tätige Mediziner sollten vom Lernen mit Fällen profitieren können. Der zweite Fehlschluss der genannten Methode ist die Definition der Komplexität, die diese impliziert. Benutzt man das Schema der Nullen und Einsen, so ist der Fall am komplexesten, dessen Informationsgehalt sich über die meisten Kategorien erstreckt. Beachtet man im Weiteren den Grundsatz, dass der Bearbeiter eines Falls jede Information, ungeachtet der zugehörigen Kategorie prozessieren muss, liegt es nahe, jeden Aspekt in die Bestimmung der Fallkomplexität miteinzubeziehen.

In der überarbeiteten und endgültigen Scoring-Version erfolgt die Punkteverteilung durch Addition. Jede Information, die in das Manual eingeordnet werden kann, erhält einen Punkt. Zusätzliche Punkte in einer Klasse können durch auffällige oder pathologische Befunde erzielt werden. Je mehr Informationen eines Falls gepunktet werden können, desto höher wird der Wert für die Komplexität im Scoring-System. Eine Sonderstellung nehmen die Organsystemklassen innerhalb der Dimension der Anamnese ein. Hierfür wird jedes innerhalb der Anamnese genannte und zuvor bereits bepunktete Symptom einem Organsystem zugeordnet. Punkte werden nicht erneut für jedes Symptom, sondern für jedes

genannte Organsystem vergeben. Berichtet ein Patient beispielsweise über Brustschmerz und Herzstolpern, würden diese Symptome innerhalb der Organsysteme nicht zwei zusätzliche Punkte, sondern lediglich einen für das kardiovaskuläre System generieren. Diese Einteilung soll zum einen dazu dienen, die Informationen der Anamnese inhaltlich sinnvoll einordnen zu können. Anhand dieser Zuordnung können somit z.B. qualitative Analysen eines Falls vorgenommen werden. Zum anderen wird mit diesem Vorgehen die These verfolgt, dass ein Fall für den Bearbeiter komplexer wird, wenn dieser Informationen aus verschiedenen Organsystemen kombinieren und kontrastieren muss, da dadurch meist verschiedene Differentialdiagnosen in Betracht gezogen werden müssen. Für ein exemplarisches Scoring vgl. S. 64.

4.1.3 Prüfung des neuen Scoring-Systems

Um Anwenderunabhängigkeit des Scoring-Systems CASE zu überprüfen, wurde ein Interrater Datensatz mit 59 Fällen von zwei Medizinstudierenden im letzten klinischen Ausbildungsjahr (Elisabeth Hilger und Kim Öhler) unabhängig voneinander nach dem neuen Scoring-System bewertet. Diese Fallbeschreibungen setzten sich zusammen aus 15 Fällen der bundesweiten Lernressource LOOOP share der Charité in Berlin (16), 12 Fällen aus vorherigen Studien des Institutes für Didaktik und Ausbildungsforschung der LMU (2, 41, 42, 67, 74), 29 KF-Aufgaben, die für Studien am Institut für Didaktik und Ausbildungsforschung der LMU verwendet wurden (17) und 3 Fällen, die im NEJM erschienen sind (20–22). Letztere wurden gezielt ausgewählt, um das Manual an besonders langen Fallpräsentationen mit einer Vielzahl an Untersuchungen und möglichen Differentialdiagnosen zu erproben. Das häufigste Leitsymptom dieser 59 Fälle war Dyspnoe, gefolgt von Thoraxschmerz und anderen (vgl. Tabelle 7).

Tabelle 7: Leitsymptome der 59 Fälle zur Untersuchung der Interrater-Reliabilität

Leit-symptom	Dyspnoe	Thorax-schmerz	Fieber	Erbre-chen	Schwä-che	Andere
Fallzahl	26	6	4	2	2	19

Im Anschluss an die Bewertung der Komplexität durch das Scoring-System wurde die Interrater-Reliabilität der 59 Fälle anhand von Cohen's Kappa berechnet. Diese wurde sowohl für die Gesamt Scores als auch für einzelne Komponenten des Scoring-Systems bestimmt. Die Unterteilung wurde gewählt, um nicht nur Aussagen über die Übereinstimmung des Endwertes, sondern auch über die der einzelnen Komponenten innerhalb des Systems treffen zu können. Um die Übersichtlichkeit und Interpretierbarkeit zu gewährleisten, wurden dabei folgende Parameter definiert: Anamnese, Bildgebung, Laborergebnisse, körperliche Untersuchung und Funktionsuntersuchungen. Die letzten vier Komponenten stellen gleichzeitig Kategorien des Scoring-Systems dar. In der Anamnese wurden die dazugehörigen Kategorien des Systems zusammengefasst, um eine größere Gruppe zu bilden. Zur Vorbereitung der Berechnung der Interrater-Reliabilität wurden die einzelnen Punktwerte der Komplexität einer Gruppe addiert und somit jeweils ein Endwert für die Anamnese, Bildgebung, Laborergebnisse, die körperliche Untersuchung und die Funktionsuntersuchungen generiert. Dies wurde für beide Rater durchgeführt und anschließend für jede Gruppe Cohen's Kappa berechnet. Bei der Betrachtung des Scoring-Systems (vgl. Tabelle 6) wird deutlich, dass die Endwerte der Gruppen durch eine unterschiedliche Anzahl an Einzelwerten generiert wurden. Dies ist für die thematische Betrachtung des Interrater-Koeffizienten nötig, um inhaltlich sinnvolle Gruppen zu bilden. Eine zu feine oder zu grobe Gliederung würde die Aussagekraft der Ergebnisse begrenzen.

Da die 59 Fälle aufgrund ihrer unterschiedlichen Form und Länge voraussichtlich eine breite Spanne an Komplexitätswerten aufweisen, scheint es sinnvoll, die Interrater-Reliabilität nicht nur für die Gesamt-Scores, sondern auch für verschiedene Wertebereiche innerhalb dieser zu untersuchen. Geprüft werden soll hierbei, ob sich die Übereinstimmung beider Rater sowohl für wenig komplexe als auch sehr komplexe Fallszenarien in einem ähnlichen Bereich bewegt, oder, ob dabei Unterschiede beobachtet werden können. Aufgrund des Umstandes, dass bei komplexen Fällen mehr Informationen von beiden Ratern prozessiert werden müssen und sich somit die Wahrscheinlichkeit für Fehler oder Irrtümer bei der Bearbeitung erhöht, ergibt sich die These, dass der Interrater-Koeffizient für eine Gruppe von sehr komplexen Fällen niedriger ausfallen könnte als in der Gruppe wenig komplexer Fallbeschreibungen. Diese Theorie wird im Folgenden durch die Betrachtung von Cohen's Kappa für zwei Gruppen an Komplexitätswerten

überprüft. Diese Einteilung richtet sich nach der Höhe und der Varianz der ermittelten Komplexitätsscores.

Um das Manual in einem letzten Schritt an einem Validation Datensatz zu erproben, wurde es auf 264 weitere Fälle unterschiedlicher Formate (Kurzfälle, KF-Fälle, ausführliche Fallbeschreibungen, Videosimulationen) angewandt. Die meisten dieser Fälle wurden in der online Lernumgebung CASUS bearbeitet und stammen von Autoren vorheriger Studien am Institut für Didaktik und Ausbildungsforschung der LMU oder aus Kooperationen mit anderen Instituten. Die Fälle wurden vorrangig für Forschungs-, oder Lehrzwecke konzipiert (2, 16, 17, 20–22, 41, 42, 58, 67, 74–78).

Anhand des Discovery, des Interrater und des Validation Datensatzes wurden so insgesamt 338 Fälle bearbeitet und deren Komplexität nach dem neuen Scoring-System bestimmt. Tabelle 8 stellt die drei Datensätze und ihre Verwendung innerhalb des Forschungsvorhabens dar.

Tabelle 8: Übersicht über Inhalt und Zweck der verwendeten Falldatensätze

Datensatz	<i>Discovery</i>	<i>Interrater</i>	<i>Validation</i>
Anzahl Fälle	15	59	264
Zweck	<i>Entwicklung Scoring-System</i>	<i>Prüfung Anwender- unabhängigkeit</i>	<i>Anwendung System</i>

Zur Bearbeitung dieser Fälle wurde auf unterschiedliche Formate zurückgegriffen, um das neue Scoring-System nicht nur an thematisch unterschiedlichen Problemstellungen, sondern auch an formal unterschiedlichen Fallszenarien zu erproben. Am häufigsten handelte es sich bei den Fallbearbeitungen um interaktive Onlinefälle, gefolgt von KFs, Kurzfällen, Videosimulationen und drei langen Case Records (vgl. Tabelle 9).

Tabelle 9: Übersicht über den Anteil aller Fälle pro verwendetem Fallformat

Fallart	Interakt. Online- fälle	Key-Fea- tures	Kurzfälle	Videosi- mulatio- nen	Case Re- cords
Anzahl	217	91	18	9	3

4.2 Quantitative Datenanalyse

4.2.1 Akquise und Aufbereitung der Daten

Um das Modell der Fallkomplexität empirisch zu prüfen und den Zusammenhang zwischen der Fallkomplexität und der Fallschwierigkeit zu untersuchen, wurden drei Datensätze analysiert. Die Daten wurden im Rahmen früherer Studien am DAM erhoben und wurden von den Autoren für diesen Forschungszweck zur Verfügung gestellt.

In einem ersten Schritt wurde beurteilt, ob die innerhalb der Studien bearbeiteten Szenarien sich für die Anwendung des Scoring-Systems und die anschließende Auswertung eignen. Die meisten der zu bearbeitenden Fälle wurden bereits in den Schritten der Anwendung und Überprüfung des Systems gescored. Die noch nicht vorliegenden Werte für die Komplexität der restlichen Fälle wurden nun anhand des Scoring-Systems bestimmt. Da die Kriterien für die Anwendung des Scoring-Systems erfüllt waren, konnte für jeden der Fälle eine Punktzahl generiert werden.

Datensatz 1 bildet die Ergebnisse von 88 Teilnehmenden ab, die jeweils 12 Fälle anhand der Lernplattform CASUS bearbeitet haben. Die Falldarstellungen beinhalten neben einer Patientenvorstellung eine körperliche Untersuchung, Röntgenbilder, ein EKG, Laborergebnisse, eine Lungenfunktionsuntersuchung, eine Echokardiografie, sowie eine Abdomen-Sonografie, deren Ergebnisse von den Teilnehmenden abgerufen werden konnten (vgl. Abb. 7, S. 36) (41, 42, 74). Im Anschluss an die Fallbearbeitung wurde von dem Programm die Eingabe einer Diagnose gefordert.

Die anhand von Datensatz 1 erhobenen Werte wurden unter anderem für eine Interventionsstudie genutzt, die die Rolle von Scaffolds in Bezug auf die diagnostischen Fähigkeiten der Teilnehmenden untersucht hat (42). Dieser Studie wurden weitere, qualitative Untersuchungen, angeschlossen (74). Die Autoren

widmen sich dabei einer qualitativen Analyse von Fehlern, die Studierende während des Diagnoseprozesses machen.

Für die geplante Auswertung der Daten soll lediglich die Diagnoserichtigkeit der Fallbearbeitungen als Parameter für die Analyse herangezogen werden. Diese wurde bereits im Rahmen der Originalstudien erhoben und durch ein dichotomes Schema kodiert. Wurde die richtige Diagnose durch einen Teilnehmer eingegeben, wurde dies durch eine Eins dokumentiert. Bei einer falschen Antwort dagegen steht eine Null (42). Die Musterlösungen der Autoren bzw. die richtigen Diagnosen der 12 Fälle sind: Aortenstenose, Urämie, Hyperventilation, AV-Knoten-Reentrytachykardie, COPD, Tuberkulose, Myokarditis, Pneumothorax (75), Asthma, Legionellenpneumonie, Lungenkarzinom und Sarkoidose. Innerhalb von Datensatz 1 wurde jeder Fall von jeder einzelnen Versuchsperson bearbeitet, so dass keine fehlenden Daten vorlagen.

Datensatz 2 beinhaltet 15 Fälle und 148 Versuchspersonen. Die virtuellen Patienten wurden hierbei ebenfalls auf der Plattform CASUS dargestellt. Der Aufbau der Fälle entspricht dem, wie er für die Erhebung des ersten Datensatzes genutzt wurde (vgl. Abb. 7, S. 36). Innerhalb der Originalstudie wurde hierbei untersucht, wie sich verschiedene Arten von Scaffolding auf die „Diagnoserichtigkeit, Effizienz und Fehler“ Medizinstudierender auswirken können (41). Hierfür wurden die 148 Versuchspersonen fünf Gruppen zugeteilt: einer Repräsentation Scaffold Gruppe, einer Repräsentation Scaffold Gruppe mit Feedback, zwei Gruppen mit strukturierter Reflektion, einmal mit- und einmal ohne Feedback und einer Kontrollgruppe (41).

Die Diagnoserichtigkeit innerhalb dieser Studie wurde ebenfalls dichotom (richtig oder falsch) erfasst. Die Expertenlösungen der 15 Fälle lauten: COPD (dreimal), Tuberkulose (zweimal), Pneumothorax (zweimal), Myokarditis, Asthma, Hyperventilation, Herzinsuffizienz, Choledocholithiasis, Lungenembolie, Lungenkarzinom und Divertikulitis (41). Für fünf Fälle lagen lediglich die Ergebnisse von 118 Teilnehmenden vor. Die restlichen Fallvignetten wurden von allen 148 Personen bearbeitet.

Datensatz 3 bezieht sich auf 18 Fälle und die zugehörigen Antworten von 106 Studienteilnehmenden. Bei diesen Bearbeitungen handelt es sich nicht um vollständige Fallpräsentationen, sondern um neun KFs und neun Problem Solving

Tasks. Erstere wurden innerhalb der Originalstudie dazu verwendet, die strategische Expertise der Teilnehmenden zu prüfen. Dies wurde durch kurze Fallpräsentationen, die ohne viele Zusatzinformationen und Untersuchungsergebnisse auskommen, durchgeführt. Im Anschluss wurden z.B. Differentialdiagnosen oder die Relevanz gewisser Laborparameter per Freitextantwort erfragt (17) (vgl. Abbildung 10).

1-1 Ein 56 Jahre alter Mann stellt sich in Begleitung des Rettungsdienstes in Ihrer Notaufnahme vor. Er berichtet über einmaliges Bluterbrechen vor ca. 2 Stunden, ansonsten seien keine Vorerkrankungen bekannt. Die Rettungssanitäter berichten von mehreren leeren Schnapsflaschen in der Wohnung des Patienten. Der Blutdruck des Patienten ist 100/70 mmHg, die Pulsfrequenz 111/min.
Welcher Laborparameter hat die größte therapeutische Konsequenz in dieser Situation?

Abbildung 10: KF mit Aufgabe (17)

Die Problem Solving Tasks dagegen zielen darauf ab, konditionales Wissen zu überprüfen. Im Anschluss an eine kurze Fallpräsentation sollten die Versuchspersonen im Freitext argumentieren, begründen und Hypothesen generieren. Alle Freitextantworten der Fragen wurden bereits innerhalb der Originalstudie von Experten ausgewertet und durch Zahlen kodiert (17). Auch die Fallbearbeitungen innerhalb dieser Studie wurden anhand von CASUS durchgeführt. Im Gegensatz zu Datensatz 1 und 2 liegen hier keine Interventionen vor, da die Daten zum Zeitpunkt T_0 im Rahmen des klinischen Wissenstest vor Intervention herangezogen wurden. Die Antwortrichtigkeit wurde innerhalb der ursprünglichen Studie durch Werte zwischen null und eins erfasst. Hierfür wurden die Freitextantworten von zwei Ärzten anhand eines Manuals ausgewertet (17).

Für die geplante Analyse wurden nun durch eine Dummy-Kodierung ganze Zahlen generiert. Werte $\geq 0,5$ wurden durch eine Eins ersetzt, somit als richtig gewertet. Alle Werte $< 0,5$ wurden als falsch beurteilt und erhielten eine Null.

Ein Fall wurde von 21 Teilnehmenden nicht bearbeitet. Die restlichen Daten lagen vollständig vor.

4.2.2 Das Rasch-Modell

Die Auswertung der Daten basiert auf der Anwendung des Rasch-Modells (RM). Dieses Modell wird meist innerhalb der sozialwissenschaftlichen Testtheorie genutzt, kann jedoch auch Anwendung in anderen Bereichen finden (79). Die Anwendung des RM beschränkt sich nicht auf einzelne Fächer wie die Psychologie, sondern ist interdisziplinär weit verbreitet (80). Zunehmende Popularität hat das RM durch die Anwendung innerhalb der PISA-Studien erlangt. Hierbei wird es z.B. für die Analyse der Anforderungen und Kompetenzen verwendet (81).

Das RM wird als das bedeutendste Modell innerhalb der probabilistischen Testtheorie angesehen. Häufig wird es verwendet, um die Güte von Tests zu evaluieren. Psychologische Tests werden heutzutage in vielen Bereichen eingesetzt. Sie können beispielsweise dazu dienen, die Fähigkeiten von Schülern oder potenziellen Mitarbeitern zu überprüfen. Solche Tests können jedoch nicht ohne weiteres konzipiert werden, sondern müssen gewissen Voraussetzungen entsprechen. Diese können anhand dieses Modells geprüft werden (80). Es dient dabei allerdings nicht nur zur Erstellung von Fragebögen, sondern auch zur Vorhersage zukünftigen Verhaltens oder zur Differenzierung verschiedener Personengruppen für Vermarktungszwecke (79).

Die theoretische Grundlage für die Anwendung des RM basiert auf einer Datenmatrix. Das bedeutet, dass für jeden Teilnehmer bestimmt wird, ob eine Aufgabe, auch Item genannt, richtig oder falsch beantwortet wurde (80). Eine Voraussetzung für die Anwendung des Modells ist die Form, in der die Daten vorliegen. In unserem Fall wurden diese dichotom kodiert. Innerhalb der Datensätze wurde kategorisiert, ob ein Teilnehmer die richtige oder die falsche Antwort gegeben hat. Dies wurde entweder durch die Originaldaten vorgegeben oder durch die durchgeführte Dummy-Kodierung festgehalten. Ein Item kann in Bezug auf die vorliegenden Daten sowohl eine geforderte Diagnose als auch die Antwort auf eine Key-Feature Aufgabe oder ein Problem Solving Task sein. Die beispielhafte Datenmatrix in Tabelle 10 beinhaltet die Antworten von fünf Personen, die jeweils drei Aufgaben bearbeitet haben. Im Beispiel hat Person zwei die Aufgaben eins und drei richtig beantwortet.

Tabelle 10: Beispielhafte Datenmatrix

Person	Aufgabe		
	1	2	3
1	1	1	0
2	1	0	1
3	0	1	0
4	0	0	1
5	1	1	0

Für jeden der drei Datensätze und dessen zugehöriger Werte wurde solch eine Matrix generiert.

Neben den Antworten, die die Personen geben, spielen noch andere Faktoren für die Anwendung des RM eine Rolle. Eine Grundannahme ist dabei, dass das Ergebnis eines Teilnehmers unter anderem von seiner Fähigkeit abhängt. Die Fähigkeit ist dabei ein personenspezifischer Parameter, der sich nicht verändert. Gleichzeitig haben auch andere zufällige Faktoren wie beispielsweise die Tagesform der Person einen Einfluss auf die Lösungswahrscheinlichkeit. Für eine generell sehr fähige Person kann man zwar annehmen, dass sie eine Aufgabe richtig löst, sicher kann man jedoch erst sein, wenn sie bearbeitet wurde. Ob eine Person eine Aufgabe richtig löst, hängt somit teilweise von ihrer Fähigkeit ab. Darüber hinaus entscheidet auch die Schwierigkeit einer Aufgabe, ob sie richtig gelöst wird. Diese beiden Parameter müssen somit auch für die Modellgleichung verwendet werden. Das RM beschreibt die Wahrscheinlichkeit, dass eine Person j mit einem bestimmten Fähigkeitsparameter θ_j ein Item i auf Basis dessen Schwierigkeitsparameter σ_i korrekt löst (50).

$$P(X_{ij} = 1 | \theta_j, \sigma_i) = \frac{e^{\theta_j - \sigma_i}}{1 + e^{\theta_j - \sigma_i}} \quad (1)$$

4.2.3 Das Linear-logistische Testmodell

Das Linear-logistische Testmodell (LLTM) wurde als Erweiterung des ursprünglichen Rasch-Modells entwickelt. Mit diesem Konzept kann die Auswirkung

einzelner Komponenten auf die Lösungswahrscheinlichkeit eines Items untersucht werden (80). Dies gelingt durch eine Linearisierung des allgemeinen RM. Bei der Anwendung des LLTM werden die Items nach diesen Eigenschaften gescored. Dabei ist q_{ik} der score des Items i in Bezug auf Eigenschaft k . Das Modell beinhaltet η_k , die Gewichtung von k innerhalb der Item Schwierigkeit und θ_j , die Fähigkeit von Person j . Die Item Schwierigkeit σ_i stellt dabei eine additive Linearfunktion der Charakteristika q_{ik} und der Gewichtung η_k , dar (50):

$$\sigma_i = \sum_{k=1}^k q_{ik} \eta_k \quad (2)$$

Setzt man Gleichung (2) nun in die des allgemeinen RM (1) ein, erhält man die Wahrscheinlichkeit der Person j , ein Item i auf Basis des Fähigkeitsparameter θ_j , des scores q und der Gewichtung η , im LLTM richtig zu beantworten (50).

$$P(X_{ij} = 1 | \theta_j, q, \eta) = \frac{e^{\theta_j - \sum_{k=1}^K q_{ik} \eta_k}}{1 + e^{\theta_j - \sum_{k=1}^K q_{ik} \eta_k}} \quad (3)$$

4.2.4 Überprüfung der Modellgültigkeit

Spricht man vom sogenannten Rasch-Modell, wird damit vorwiegend auf die Modellgleichung hingewiesen. Neben dieser gibt es einige statistische Voraussetzungen und Modellannahmen (80).

Eine wichtige Voraussetzung für die Anwendung des RM ist die stochastische Unabhängigkeit. Dies bedeutet konkret, dass die Wahrscheinlichkeit, dass ein Teilnehmer eine bestimmte Aufgabe richtig beantwortet, unabhängig davon sein muss, ob er eine zweite Aufgabe korrekt lösen kann (80). Diese Voraussetzung ist z.B. verletzt, wenn Aufgaben aufeinander basieren oder sich ähneln. Bei einigen Studien werden die Items so konzipiert, dass sie einen Lernprozess begünstigen und die jeweils folgende Aufgabe den neuen Wissensstand abprüft. Werden ähnliche oder sogar gleiche Items verwendet, wäre die Unabhängigkeit ebenfalls nicht erfüllt und je nach Reihenfolge der Aufgaben könnten bestimmte Teilnehmer einen Vorteil gegenüber anderen erlangen (80).

Bevor die vorliegenden Daten analysiert wurden, wurde diese Voraussetzung für jeden Datensatz geprüft. Die inhaltliche Untersuchung der einzelnen Datensätzen in Hinblick auf die Zielsetzung der Originalstudien zeigte keine Abhängigkeiten zwischen den Items. Ebenso war kein Lerneffekt innerhalb dieser Datensätze beabsichtigt.

Im Rahmen des dritten Datensatzes wurden vier Fälle identifiziert, die keine Varianz aufweisen. Diese wurden von der Auswertung ausgeschlossen.

4.2.5 Schätzung der Fallschwierigkeit

Die Schwierigkeit einer bestimmten Aufgabe oder eines medizinischen Falls ist ein Konzept, das nicht einfach zu bestimmen ist. Es sind unzählige Möglichkeiten vorstellbar, um sich diesem zu nähern. Je nach Lernumgebung bestehen verschiedene Faktoren mit möglichem Einfluss auf die generelle Schwierigkeit. Aus diesem Grund und, da kein universell anwendbares System existiert, muss die Schwierigkeit in diesem Fall geschätzt werden.

Anhand des RM und des LLTM können bestimmte Parameter, die die Aufgaben oder Personen betreffen, geschätzt werden (79). Für die Bearbeitung der zweiten Forschungsfrage ist besonders ein Parameter interessant: die Schwierigkeit der verschiedenen Fälle, bzw. Aufgaben.

Durch das RM, sowie das LLTM, werden normalerweise die Wahrscheinlichkeit einer Antwort x_{ij} bei den bekannten Größen θ_j und σ_i ermittelt. Für die folgenden Untersuchungen soll allerdings, nach einem umgekehrten Prinzip, durch die Antworten bestimmt werden, wie schwierig die jeweiligen Aufgaben sind. Deshalb werden anhand der gegebenen Antworten x_{ij} , die Werte der unbekannten Variablen θ_j und σ_i gesucht. Dieser Vorgang wird als Schätzen von Parametern bezeichnet. Dafür werden die Größen für die Parameter gesucht, die für falsche Antworten eine niedrige Lösungswahrscheinlichkeit und für richtige Antworten eine hohe Wahrscheinlichkeit anzeigen. Günstig ist es dabei, wenn Aufgaben von möglichst vielen Probanden bearbeitet wurden, um einen Trend zu erkennen. Wird z.B. ein Item oft richtig beantwortet, ist es eher einfach und hat daher ein niedriges σ_i (79).

Die Schwierigkeiten der Items aller drei Datensätze wurden zuerst durch das allgemeine Rasch-Modell und im Anschluss ebenfalls anhand des LLTM geschätzt.

Diese Schätzung erfolgte aus einer Matrix von Item Eigenschaften und den gewichteten Scoring-Werten.

Eine hohe Korrelation zwischen der geschätzten Schwierigkeit des Rasch-Modells und dem LLTM zeigt eine valide Repräsentation der Item Schwierigkeiten durch die theoretischen Item Charakteristika (50).

4.2.6 Der Zusammenhang zwischen Fallkomplexität und Fallschwierigkeit

Das LLTM basiert auf der Annahme, dass die beobachteten Unterschiede der Aufgabenschwierigkeiten ausschließlich durch die eingeschlossenen Parameter bedingt sind (82). Konkret bedeutet das, dass für die Analyse lediglich die Fallkomplexität und deren Einfluss auf die Fallschwierigkeit untersucht wurde.

Nachdem untersucht worden ist, inwiefern die geschätzten Schwierigkeiten aus dem LLTM und RM korrelieren, wurden die Gewichtungen der einzelnen Komponenten des Scoring-Systems ermittelt. Diese werden durch die Eta (η) Werte repräsentiert. Positive Werte von Eta implizieren, dass das untersuchte Charakteristikum die Schwierigkeit eines Items verringert, während negative Werte darauf hindeuten, dass ein Item dadurch schwieriger wird (50). Wichtig ist hierbei, dass die Werte in dieser Form nicht standardisiert sind und somit in ihrer Höhe nicht interpretiert werden dürfen. In Relation zueinander können die Werte dennoch betrachtet werden.

Dadurch wurde bestimmt, welchen Einfluss die Komplexität der einzelnen Komponenten des Systems auf die Fallschwierigkeit ausübt. Hierfür wurden folgende fünf Features definiert: Anamnese, Bildgebung, Laborergebnisse, körperliche Untersuchung und funktionelle Tests.

Alle Analysen wurden unter Nutzung von R4.0.3 (R Core Team, 2022) durchgeführt (83).

5 Ergebnisse

5.1 Resultate der qualitativen Analyse der Fallkomplexität

5.1.1 Notwendigkeit der Weiterentwicklung bisheriger Konzepte

Die Komplexitätswerte, die mit dem System von Braun et al. (2) durch beide Rater (Rater 1: Dr. Dr. Marc Weidenbusch, Rater 2: Kim Öhler) generiert wurden, reichten von vier bis zehn Punkten. Eine Punktzahl von zehn stellt gleichzeitig die von den Autoren theoretisch festgelegte Maximalpunktzahl dar (2).

Da in jedem der 12 Fälle Informationen zu den Komponenten Anamnese, körperliche Untersuchung, technische Untersuchung und psychosoziale Aspekte genannt wurden, erhielt jeder der Fälle mindestens einen Punkt pro Kategorie für die Komplexität. Für einige Fälle wurden von den beiden Ratern Nebendiagnosen erkannt und somit zusätzliche Punkte in dieser Kategorie vergeben. Beide Anwender vergaben die geringsten Werte für die Gesamtkomplexität an einen Fall mit der Diagnose Myokarditis (Fall 7). Im Schnitt am meisten Punkte erhielten die beiden Fälle mit den Diagnosen Urämie (Fall 2) und Pneumothorax mit Infekt (Fall 8) (vgl. Tabelle 11). Tabelle 11 ordnet den 12 Fällen die jeweils von den Ratern vergebene Endpunktzahl der Komplexität zu. Hierbei ist zu beachten, dass die dargestellten Diagnosen von den jeweiligen Autoren der Fälle stammen und erst im Anschluss an die Beurteilung der Komplexität recherchiert wurden. Diese Diagnosen stimmen nicht in allen Fällen mit der Einschätzung beider Rater überein, da diese bei den meisten Vignetten mögliche Nebendiagnosen identifizieren konnten. Um die Fälle zu charakterisieren, sind die vorgesehenen Diagnosen der Autoren als eine Art Musterlösung in Tabelle 11 mitdargestellt.

Tabelle 11: Musterdiagnosen der 12 Fälle und die durch beide Rater vergebenen Punktzahlen für die Gesamtkomplexität

Fallnummer	Diagnose 1	Diagnose 2	Punkte Rater 1	Punkte Rater 2
Fall 1	Aortenstenose	-	8	4
Fall 2	Urämie	-	10	6
Fall 3	Hyperventilation	-	8	5
Fall 4	AV-Knoten- Reent- rytachykardie	-	7	6
Fall 5	COPD	-	6	4
Fall 6	Tuberkulose	-	9	6
Fall 7	Myokarditis	-	5	4
Fall 8	Pneumothorax	Grippaler Infekt	8	8
Fall 9	Asthma	-	8	5
Fall 10	Legionellen- pneumonie	-	8	6
Fall 11	Lungenkarzi- nom	-	7	6
Fall 12	Sarkoidose	-	6	5

Bis auf das Scoring des achten Falls zeigten sich keine Übereinstimmungen für die mit dem Modell nach Braun et al. (2) bestimmte Gesamtkomplexität der 12 Fälle (vgl. Tabelle 11). Den achten Fall nutzen die Autoren im Anhang ihrer Veröffentlichung, um ein exemplarisches Scoring zu präsentieren (2). Da die beiden Rater (Dr. Dr. Marc Weidenbusch und Kim Öhler) bereits zum Zeitpunkt des

Scorings über Kenntnis dieser Abhandlung und der für den achten Fall vorgesehenen Diagnosen verfügten, muss im Fall acht eine Beeinflussung durch dieses Material in Betracht gezogen werden.

Abbildung 11 stellt die Ergebnisse beider Rater für die Gesamtkomplexität graphisch dar. Obwohl Rater 1 die Komplexität der Fälle in elf von 12 Fällen höher als Rater 2 beurteilte, lassen sich dennoch Trends für die Werte beider Rater in Relation zueinander erkennen.

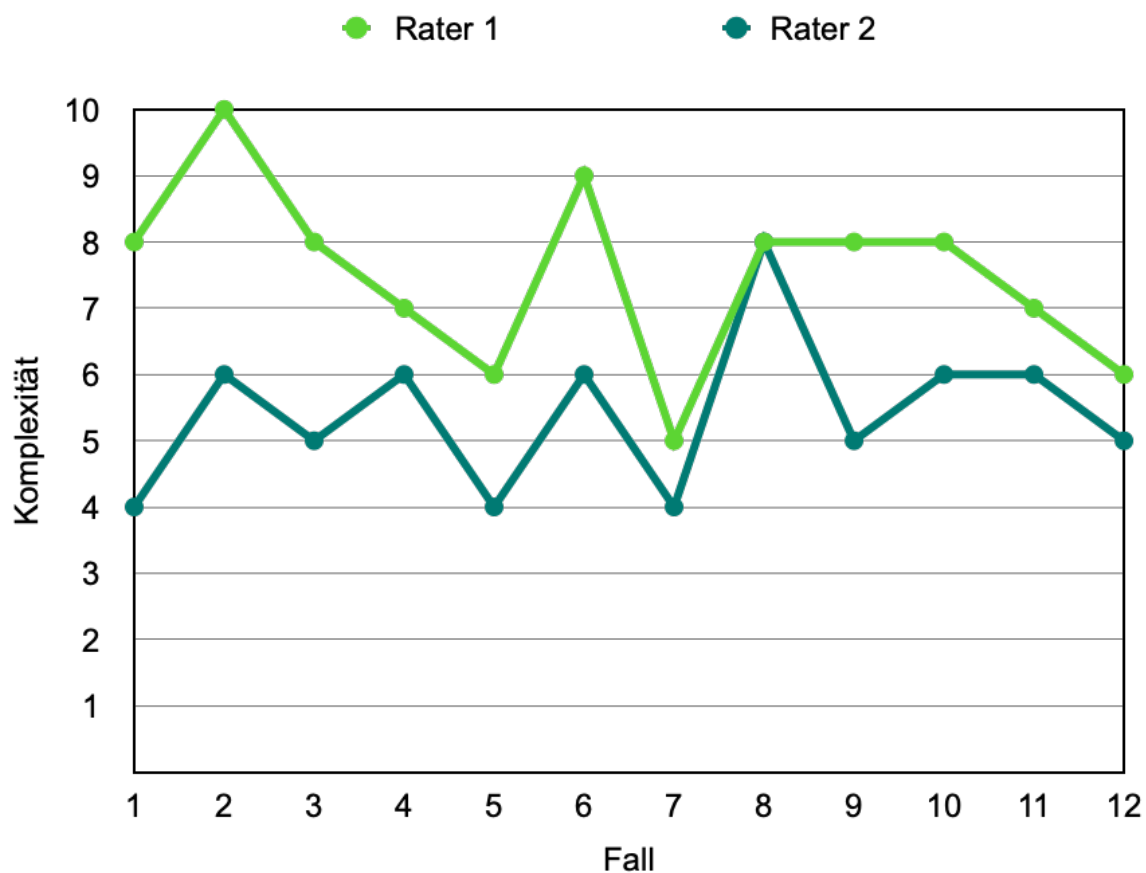


Abbildung 11: Gesamtwerte beider Rater für die Komplexität (Für eine Fallübersicht vgl. Tabelle 11)

Neben den Gesamtwerten wurden die einzelnen Kategorien auf die Übereinstimmung beider Rater untersucht. Für die Anamnese und die Gesamtwerte zeigte sich eine schlechte Übereinstimmung beider Anwender, gemessen durch den Interraterkoeffizient Cohen's Kappa. Die technischen Untersuchungen und die Kategorie der Nebendiagnosen wiesen eine geringe Übereinstimmung für die Komplexität auf. Bei der körperlichen Untersuchung bestand eine ausreichende

Einigung und die Kategorie der psychosozialen Faktoren erreichte eine moderate Übereinstimmung (vgl. Tabelle 12).

Tabelle 12: Cohen's Kappa der Gesamtwerte und der fünf Kategorien bei Verwendung des Systems nach Braun et al. (2). (1: Anamnese, 2: Körperl. US, 3: Techn. US, 4: Psychosozial, 5: Nebendiagnosen)

Kategorien	Gesamt	1	2	3	4	5
Cohen's Kappa	-0.05	-0.02	0.40	0.07	0.46	0.05

Die Anwendung dieses Scoring-Systems schuf die Basis für die Weiterentwicklung des Verfahrens und die Konzeption eines neuen Systems. Das neue Scoring-System wurde primär auf Basis empirischer Abläufe und verbreiteter Strukturen in der Medizin konzipiert. Die Weiterentwicklung bis zur endgültigen Version hat unter Betrachtung der bereits dargestellten Ergebnisse und der daraus abgeleiteten Anforderungen stattgefunden.

Braun et al. schlagen als letzte ihrer fünf Kategorien das Scoring einer Nebendiagnose vor (2). Besonders diese Kategorie bereitete den beiden Ratern Schwierigkeiten bei der Bearbeitung der Fälle. In neun von 12 Fällen identifizierte Rater 1 mehr Nebendiagnosen innerhalb der Fallbeschreibung und vergab dementsprechend mehr Punkte für diese Kategorie als Rater 2. Bei der anschließenden qualitativen Analyse fiel auf, dass Rater 1 deutlich mehr klinische Befunde und Untersuchungsergebnisse als Hinweise auf eine mögliche Differentialdiagnose wertete als Rater 2. Aufgrund einer inhaltlichen Diskussion erschienen diese Aspekte Rater 2 durchaus plausibel, in der ersten Einschätzung hatte er jedoch meist andere Schlüsse für die Beurteilung des Falls gezogen.

Durch die Nutzung des Systems nach Braun et al. (2) und die Hürden, die dabei beobachtet wurden, ließen sich folgende generelle Anforderungen an das neue Scoring-System festhalten:

- Die Bedingung der Anwenderunabhängigkeit sollte erfüllt werden
- Der Gebrauch sollte einfach und standardisiert ablaufen können

- Für die Benutzung sollte kein großes Maß an klinischer Erfahrung notwendig sein
- Anhand des Systems sollen alle möglichen Informationen eines Falls erfasst werden können

Um die Anwenderunabhängigkeit für das neue Scoring-System zu optimieren, erfährt deshalb besonders die Kategorie der Nebendiagnosen einen anderen Stellenwert als bei Braun et al. (2). Dieser Aspekt stellt im neuen System keine eigenständige Kategorie dar. Ein Grund hierfür liegt in den vielfältigen Darstellungsmöglichkeiten und der unterschiedlichen Relevanz einer Nebendiagnose. Eine Möglichkeit ist es, dass verschiedene Vordiagnosen eines Patienten innerhalb der Anamnese in Erfahrung gebracht werden. Da hierbei die Unterscheidung in Haupt- und Nebendiagnosen oft nicht eindeutig vorgenommen werden kann, werden alle Diagnosen, die durch eine Befragung in Erfahrung gebracht werden, unter Vorerkrankungen in der Dimension der Anamnese berücksichtigt. Diese sind insofern wichtig, da sie Denkprozesse und den folgenden Entscheidungsprozess bzw. die Diagnosefindung des Bearbeiters beeinflussen können. In anderen Fällen wiederum kommt es vor, dass der Bearbeiter gegen Ende nicht nur eine mögliche Hauptdiagnose, sondern eine oder mehrere Nebendiagnosen stellt. Grund dafür kann beispielsweise sein, dass die Symptomkonstellation nicht durch eine einzelne Diagnose erklärt werden kann. Dieser Schritt ist allerdings individuell und kann somit in Bezug auf die Anwenderunabhängigkeit und die verschiedenen Fähigkeiten des Einzelnen nicht in das neue Scoring-System miteinbezogen werden. Das System soll dazu dienen, die Komplexität eines Falls zu beurteilen. Da im praktischen Gebrauch von Fällen selten Musterdiagnosen der Autoren vorliegen oder diese hin- und wieder infrage gestellt werden können, wird im neuen System dieser Aspekt gezielt nicht berücksichtigt. Teilweise kommt es vor, dass im Laufe einer Fallbearbeitung Nebendiagnosen genannt werden, die die aktuelle Symptomkonstellation erklären oder auch die Hauptdiagnose am Ende eines Falls aufgelöst wird. Diese Aspekte erhalten ebenfalls keine zusätzlichen Punkte, da unklar ist, ob sie nachträglich überhaupt zur Komplexität eines Falls beitragen können.

Neben der Kategorie der Nebendiagnosen bot der Ablauf der Punktevergabe durch die Betrachtung der Linearität der Information Anlass zur Diskussion. Beide Rater waren sich in der Einschätzung der Kongruenz der Informationen

eines Falls oft uneinig. Sobald Fälle mehrere auffällige Untersuchungsergebnisse oder verschiedene Symptome aufwiesen, fiel es schwer, zu beurteilen, ob die jeweilige Information eine neue Ebene im Vergleich zur Vorherigen eröffnet oder nicht. Aus diesem Grund und um den Gebrauch eines Scoring-Systems standardisierter zu gestalten, wurde auch der Vorgang der Punktevergabe geändert und der Aspekt der Linearität nicht aufgegriffen.

Die komplette Struktur des weiterentwickelten Scoring-Systems zielt unter anderem darauf ab, die Anwenderunabhängigkeit zu optimieren. Mit seinen übersichtlichen Kategorien und den einheitlichen Klassen soll ein standardisierter Gebrauch gewährleistet werden. Diese sind derart feingliedrig gestaltet, um möglichst alle Informationen eines Falls, die möglicherweise die Komplexität beeinflussen, einordnen zu können.

Diese Ergebnisse stützen nicht nur die Notwendigkeit der Weiterentwicklung bisheriger Konzepte, sondern definieren gleichzeitig die Anforderungen an ein neues System. Diese werden auf den folgenden Seiten für das neue Scoring-System geprüft und die Ergebnisse dargestellt.

5.1.2 Vorschlag eines Scoring-Systems CASE zur standardisierten Erfassung von Fallkomplexität

Anhand der mehrfach überarbeiteten Version des neuen Scoring-Systems CASE konnten Werte für die Komplexität für eine breite Spanne von Fällen generiert werden. Bei der Anwendung auf die 59 Fälle des Interrater Datensatzes wurden von beiden Anwendern Ergebnisse für die Komplexität zwischen 2 (für einen Kurzfall) und 185 (für einen sehr detailreichen Fall) erreicht (vgl. Abbildung 12).

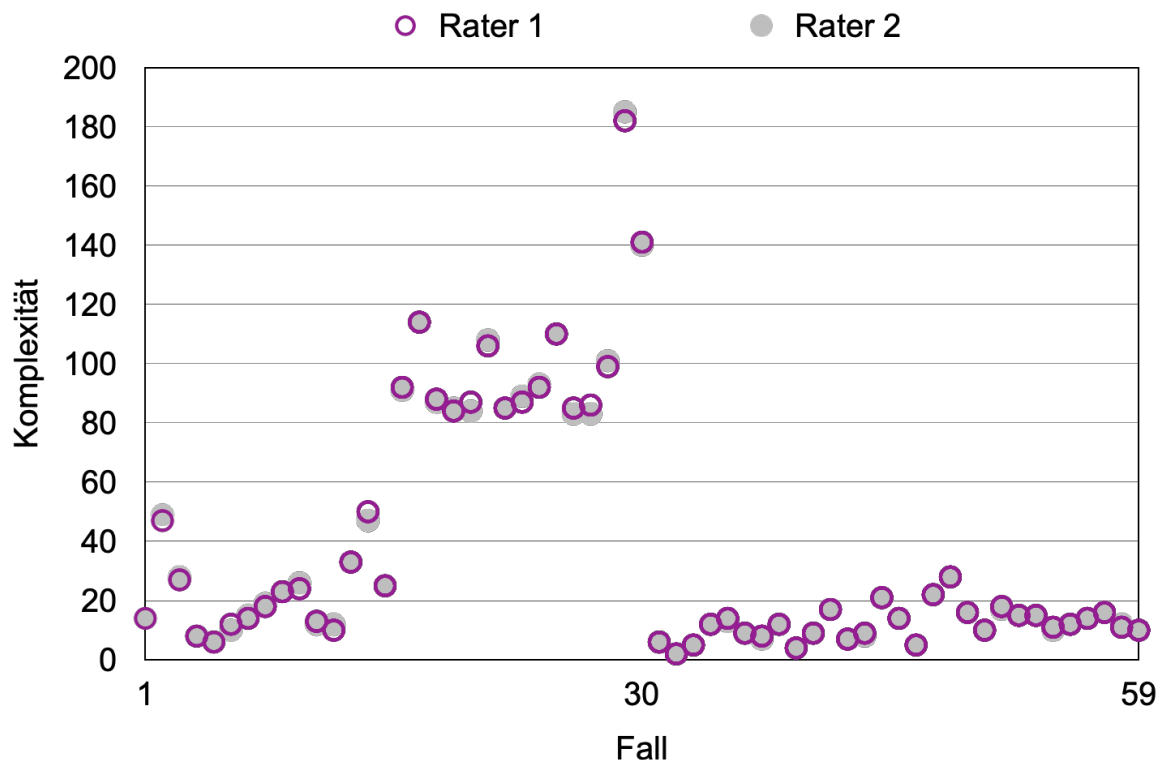


Abbildung 12: Gesamtscores beider Rater nach dem neuen System CASE (Rater 1: Elisabeth Hilger, Rater 2: Kim Öhler)

Cohen's Kappa für die Gesamtwerte lag bei 0.53. Für die einzelnen Untergruppen innerhalb des Systems: Anamnese, Bildgebung, Labor, körperliche Untersuchung und funktionelle Tests betrug der Interraterkoeffizient mindestens 0.57 (vgl. Tabelle 13).

Tabelle 13: Cohen's Kappa der Gesamtwerte und der fünf Untergruppen bei der Verwendung des neuen Scoring-Systems (1: Anamnese, 2: Bildgebung, 3: Labor, 4: Körperl. US, 5: Funktionell)

Untergruppen	Gesamt	1	2	3	4	5
Cohen's Kappa	0.53	0.57	0.80	0.94	0.77	0.93

44 der 59 Fälle wurden von beiden Ratern mit einer Komplexität beurteilt, die weniger als 60 Punkte betrug. Bei den restlichen 15 Fällen wurden Werte über 60, bis hin zu 185 vergeben. Aufgrund der großen Varianz der Komplexitätswerte

wurde der Interrater-Koeffizient der Gesamtwerte im Anschluss jeweils für beide Gruppen separat berechnet (vgl. Tabelle 14).

Tabelle 14: Cohen's Kappa der Gesamtwerte, berechnet für 44 Fälle mit einer Komplexität <60 und 15 Fälle mit einer Komplexität >60

Kategorien	Gesamt <60	Gesamt >60
Cohen's Kappa	0.64	0.17

Für die Gruppe der Gesamtwerte <60 zeigte sich eine gute Übereinstimmung bei der Anwendung des Systems durch die beiden Rater. In der Gruppe der 15 Fälle mit einer Komplexität von 60 und höher konnte dagegen lediglich eine geringe Übereinstimmung beobachtet werden. Dieses Ergebnis stützt die erwartbare Vermutung, dass Cohen's Kappa für Fälle mit hoher Komplexität nach dem Scoring-System eine geringere Übereinstimmung beider Rater aufweist (vgl. S. 44, f.).

Insgesamt zeigte sich somit sowohl für die Gesamtwerte als auch für die einzelnen Untergruppen des Systems (Anamnese, Bildgebung, Labor, körperliche Untersuchung, Funktionsuntersuchungen) eine moderate bis nahezu perfekte Übereinstimmung bei der Nutzung des Scoring-Systems durch diese beiden Anwender. Auf Basis dieses Ergebnisses wurde entschieden, dass die Punkte für die folgenden Fälle nur durch einen der beiden Rater (Kim Öhler) vergeben werden.

Die meisten der 279 im Anschluss bearbeiteten Fälle des Validation Datensatzes waren interaktive Onlinefälle aus der online Lernumgebung CASUS. Anhand des Manuals ließ sich darüber hinaus die Komplexität von Videosimulationen mit Schauspielpatienten, Kurzfällen, KFs und besonders langen Case Records mit Fließtext bestimmen. Das häufigste Leitsymptom der bearbeiteten Fälle war Dyspnoe, gefolgt von orthopädischen Beschwerden (vgl. Abb. 13). Die Diagnosen der Fälle beziehen sich entweder auf die Musterlösungen, die viele Fälle am Ende angegeben hatten oder wurden dem Material der zugehörigen Studien entnommen. Die mitgelieferte Diagnose des Autors war dabei nicht in allen Fällen nachvollziehbar und deckte sich nicht immer mit den Einschätzungen der Bearbeiter. In Hinblick auf die Zielsetzung dieser Arbeit ist dieses Ergebnis jedoch von

untergeordneter Bedeutung. Für 119 der 279 Fälle konnte dem Material keine durch den Autor vorgegebene Diagnose entnommen werden.

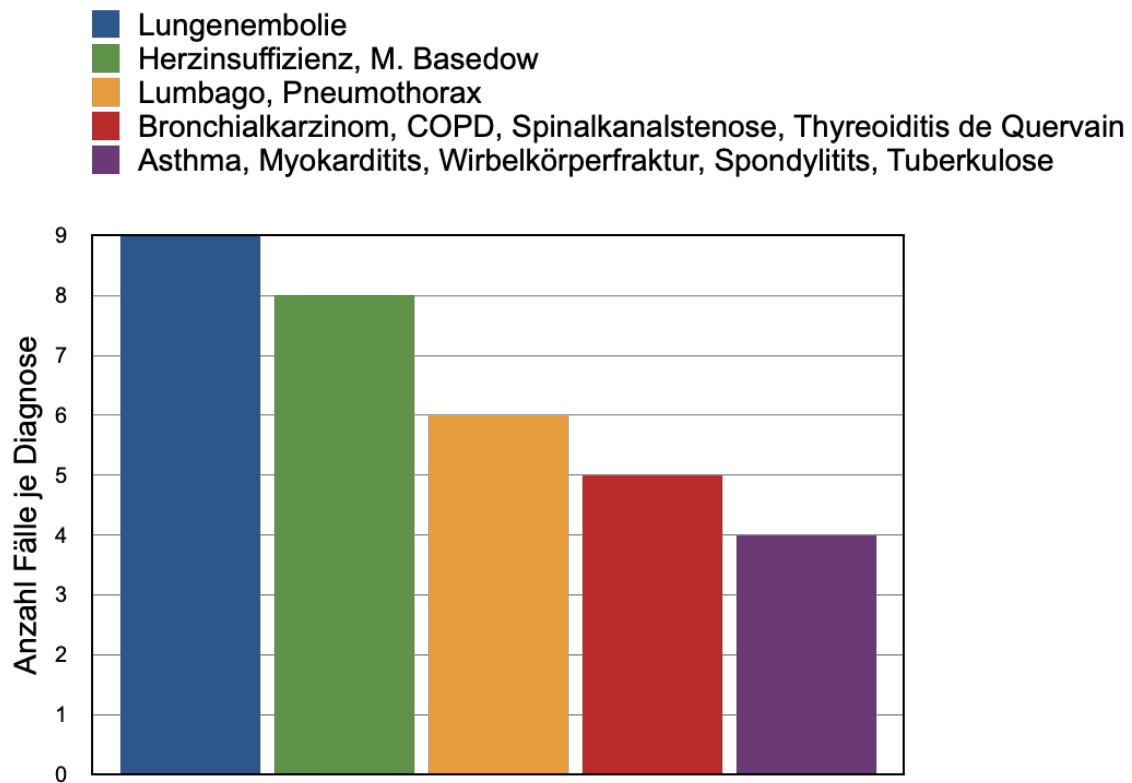


Abbildung 13: Häufigste Diagnosen von 160 Fällen aus dem Validation Datensatz

Für die Komplexität dieser Fälle wurden anhand des Scoring-Systems CASE Werte zwischen 2 bei einem Kurzfall (vgl. Abbildung 14) und 246 für einen Fall mit der Diagnose einer bakteriellen Endokarditis bestimmt. Der Kurzfall mit der niedrigsten Komplexität ist in Abbildung 14 dargestellt. Die Informationen, die die beiden Punkte generiert haben, sind dabei hervorgehoben. In diesem Beispiel werden die beiden Aspekte der Anamnese zugeordnet. Zum einen wird eine Vor-diagnose der Patientin, die perniziöse Anämie, genannt, zum anderen erhält der Leser mit dem zweiten Aspekt die Information, dass die Patientin verstorben ist.

Sie haben bei einer ansonsten gesunden, 35-jährigen Patientin vor einer Woche eine **perniziöse Anämie** diagnostiziert. Nach Beginn einer leitliniengerechten Therapie haben Sie die Patientin zur Wiedervorstellung in 3 Wochen in die Ambulanz einbestellt. Sie erhalten vom Hausarzt einen Anruf, dass die Patientin in der vergangenen Nacht **verstorben** ist. Der Kollege ist erheblich aufgebracht, als er erfährt, dass Sie die Patientin nach Therapiebeginn nicht engmaschiger überwacht haben und weist auf die notwendige Kontrolle eines Laborwertes hin.

Abbildung 14: Kurzfall (17) mit der niedrigsten Komplexität, Informationen werden lediglich innerhalb der ersten Dimension (Anamnese) genannt

Die meisten der bearbeiteten Fälle erhielten geringe Werte für die Komplexität. Nur einige wenige wurden mit einer Komplexität von 200 aufwärts bewertet (vgl. Abbildung 15).

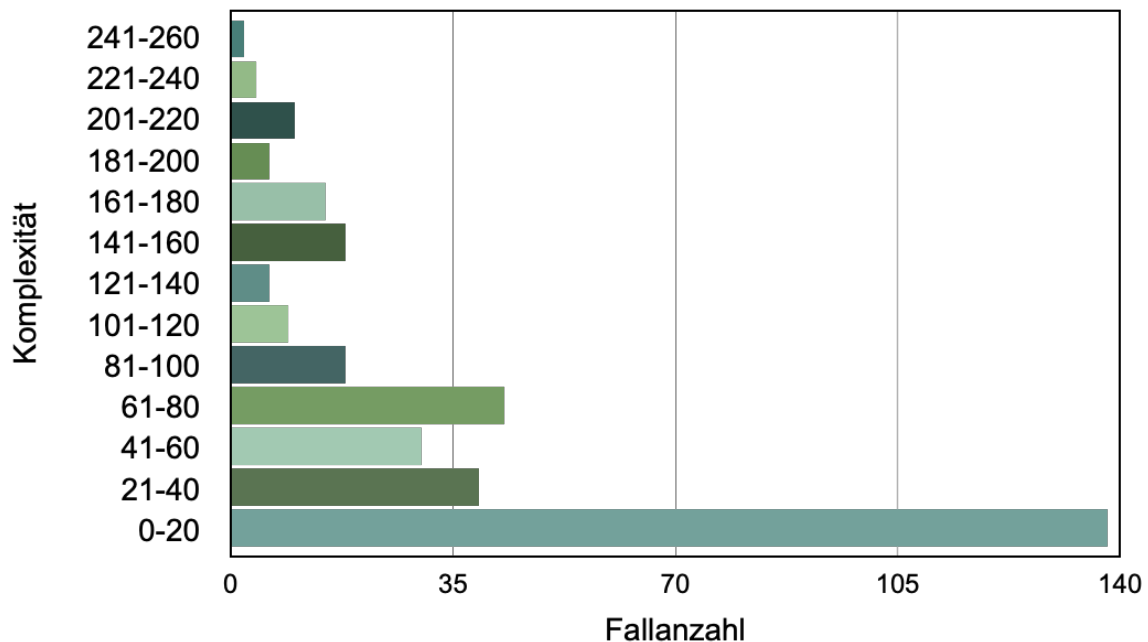


Abbildung 15: Anzahl Fälle je Komplexitätsspanne

Abbildung 16 zeigt einen Kurzfall (17), der im Scoring eine Komplexität von 21 Punkten erreicht hat. Die Informationen konzentrieren sich hierbei im Vergleich zu Abbildung 14 nicht nur auf eine Dimension, sondern sind vielfältiger. Es werden Aspekte zu beiden Dimensionen (Anamnese: blau und Untersuchungen: grün und orange) genannt. Tabelle 15 ordnet die Informationen den verschiedenen Dimensionen, Kategorien und Organsystemklassen zu und beschreibt das exemplarische Scoring dieses Falls. Für eine Darstellung dieses Falls innerhalb der verwendeten Scoring Tabelle vgl. Anhang #3.

Eine 22 Jahre alte Patientin mit Typ 1 Diabetes mellitus stellt sich in Ihrer Notaufnahme vor. Sie hat seit mehreren Tagen Übelkeit, Erbrechen, Polyurie, Polydipsie und abdominelle Schmerzen. In der körperlichen Untersuchung finden sich sehr trockene Schleimhäute, außerdem atmet die Patient sehr tief. Im Labor finden Sie: Natrium 147 mM (Normwert 135-145 mM), Kalium 6 mM (Normwert 3,5-5,1 mM), Chlorid 93 mM (Normwert 98-106 mM), Bicarbonat 11 mM (Normwert 22-29 mM). In der arteriellen Blutgasanalyse hat die Patientin einen pH von 7,40 (Normwert 7,38-7,42)

Abbildung 16: Kurzfall (17) mit Informationen in beiden Dimensionen und verschiedenen Kategorien

Tabelle 15: Exemplarisches Scoring: Fall aus Abbildung 16

Dimensionen	Kategorien	Punkte
Anamnese	Geschichte aktuelle Erkrankung: Übelkeit, Erbrechen, Polyurie, Polydipsie, abdominelle Schmerzen	5
	Vordiagnosen: DM Typ 1	1
	Vorbehandlung	
	Medikamente	
	Psychosoziale Anamnese	
	Allergien	
	Familienanamnese	
	Reiseanamnese	
	Organsysteme: Gastrointestinal, nephro/ urogenital	2
Untersuchungen	Bildgebung	
	Labor: Natrium 147 mM, Kalium 6 mM, Chlorid 93 mM, Bicarbonat 11mM, pH: 7,40	5 Werte 4 auffällig
	Körperliche Untersuchung: Sehr trockene Schleimhäute, atmet sehr tief	2 untersucht 2 auffällig
	Funktionelle Tests	
		21 Punkte

Obwohl viele verschiedene Arten von Fällen untersucht wurden, stellt Abbildung 17 einen exemplarischen Fall dar, wie er in Bezug auf die Länge und Art der Informationen häufig zur Anwendung kam. Er verfügt über eine mittlere Falllänge und lässt sich inhaltlich in zwei verschiedene Teile gliedern. Teil eins umfasst die Anamnese und wird im Scoring somit der ersten Dimension zugeordnet. In Abbildung 17 sind die Informationen dieses Abschnitts blau markiert. Die körperlichen Untersuchungsergebnisse im zweiten Teil sind grün markiert. Diese Informationen können der zweiten Dimension innerhalb des Scoring-Systems, den Untersuchungen, zugeordnet werden. Bei Betrachtung der Informationen innerhalb der zweiten Dimension wird deutlich, dass dieser Fall lediglich eine Kategorie (körperliche Untersuchung) innerhalb der Dimension der Untersuchungen abbildet. Diese Struktur wird oft von den Autoren kurzer oder mittellanger Fälle gewählt. In der klinischen Praxis stellen die Anamneseerhebung und die körperliche Untersuchung oft die wichtigsten Bausteine innerhalb der Diagnosefindung dar und finden meist als erstes statt, da sie schnell und ressourcensparend durchzuführen sind. In vielen Szenarien liefern diese Schritte bereits entscheidende Hinweise auf eine mögliche Diagnose und müssen nicht notwendigerweise um apparative Untersuchungen oder Laborergebnisse komplettiert werden. Fälle, die diesen Aspekt abbilden, können somit realistisch und je nach verfügbaren Methoden, in einem spezifischen Setting dargestellt werden.

Der Fall in Abbildung 17 hat durch das Scoring-System CASE eine Komplexität von 57 Punkten erreicht. Er repräsentiert somit nicht nur einen Fall mittlerer Länge, sondern auch mittlerer Komplexität.

Bei der Erprobung des Scoring-Systems gab es wenige Fälle, deren Informationen lediglich einer Dimension zugeordnet werden konnten. In den meisten Fällen, die untersucht wurden, erstreckten sich die Informationen über beiden Dimensionen, diverse Kategorien und verschiedene Organsysteme.

Der 67-jährige Herr Pohl leidet unter zunehmender Luftnot bei Belastung. Diese falle ihm in den letzten Wochen immer mehr auf. Zudem verliere er immer mehr an Gewicht (5 kg in den letzten zwei Monaten). Nachtschweiß und Fieber verneint er, nur gelegentliches Husten, insbesondere morgens, habe er. Insgesamt fühlt er sich deutlich im Alltag eingeschränkt. Vorerkrankungen habe er außer einer chronischen Sinusitis nicht. Gelegentlicher Alkoholkonsum, Nikotinabusus mit 80 pack years.

In der körperlichen Untersuchung erheben Sie folgende Befunde:

Wacher, voll orientierter, kognitiv unauffälliger 67-jähriger Patient in reduziertem Allgemeinzustand und Ernährungszustand (63 kg, 1,85 m, BMI 18,5 kg/m²), zugewandt und freundlich, adäquate verbale Reaktion.

Vitalparameter: Blutdruck 120/80 mmHg, Herzfrequenz 95/min, Atemfrequenz 17/min, Körpertemperatur (Ohr) 36,8 °C

Kardiovaskuläres System: Herztöne rein und rhythmisch auskultierbar, keine Herzgeräusche, keine Strömungsgeräusche über den Carotiden auskultierbar. Keine Jugularvenenstauung, keine peripheren Ödeme. Periphere Pulse (A. radialis bds., A. dorsalis pedis bds., A. tibialis posterior bds.) tastbar.

Respiratorisches System: Keine Thoraxdeformitäten. Keine Lippenzyanose. Hypersonorer Klopfeschall beidseits, atemabhängig bds. schlecht verschiebbliche Lungengrenzen. Leises Atemgeräusch, Bronchialatmung, keine feuchten oder trockenen Atemnebengeräusche. Brustwirbelsäule nicht klopfeschmerzhaft.

Abdomen: Inspektion unauffällig. Darmgeräusche lebhaft über allen vier Quadranten auskultierbar, keine abdominellen Strömungsgeräusche. Bauchdecke weich, Palpation nicht schmerzhaft, keine palpablen Resistenzen. Perkutorisch normaler Klopfeschall. Leber bei Inspiration mit weichem Leberrand und ohne Schmerzhaftigkeit 5 cm kaudal des Rippenbogens tastbar, perkutorisch 11 cm in der MCL. Milz nicht tastbar. Keine Hernien. Kein Nierenlagerklopfeschmerz. Lendenwirbelsäule nicht klopfeschmerzhaft.

Lymphknoten: keine pathologisch vergrößerten zervikalen, axillären oder inguinalen Lymphknoten tastbar.

Abbildung 17: Fall mittlerer Länge (41) mit dem Leitsymptom Dyspnoe und einer Komplexität von 53

Durch das Scoring aller 338 Fälle der drei Datensätze wurde deutlich, dass gewisse Spalten des Scoring-Systems CASE keine Verwendung fanden. Während innerhalb der Dimension der Anamnese Informationen zu allen Kategorien und Klassen genannt wurden, wurden folgende Spalten innerhalb der Untersuchungsergebnisse nicht verwendet:

- Kategorie: Bildgebung - Organsysteme: Psychiatrisch, endokrin und Haut
- Kategorie: Labor, Klinische Chemie - Organsysteme: Kopf/ Hals, psychiatrisch und Haut
- Kategorie: Labor, Mikrobiologie - Organsysteme: Allgemein, Kopf/ Hals, muskuloskelettal, psychiatrisch und endokrin
- Kategorie: Labor, Pathologie - Organsysteme: Allgemein, muskuloskelettal, neurologisch, psychiatrisch, endokrin und Haut
- Kategorie: Körperliche Untersuchung - Organsysteme: Endokrin

- Kategorie: Funktionsuntersuchungen - Organsysteme: Allgemein, gastro-intestinal und psychiatrisch

Die fehlende Verwendung der Organsystemklasse: Allgemein lässt sich darauf zurückführen, dass diese im näheren Sinne kein eigenes Organsystem darstellt, sondern verwendet wurde, wenn ein Befund mehrere Organsysteme einschließen kann oder er keinem eindeutig zugeordnet werden konnte. Für die mangelnde Verwendung der anderen Spalten können verschiedene Erklärungen herangezogen werden. Sicher ist, dass die Auswahl der 338 Fälle möglichst vielfältig gestaltet wurde aber dennoch nicht alle möglichen Themen abgedeckt werden konnten. Die pathologische Untersuchung der Haut ist ein Verfahren, das in der Praxis häufig angewandt wird, wenn beispielsweise ein suspekter Nävus untersucht werden soll. Dass die Vignetten keinen rein dermatologischen Fall beinhalteten, kann diese Beobachtung erklären. Im Gegensatz dazu beinhaltet das Scoring-System Kategorien wie die Bildgebung des psychiatrischen Organsystems, die wie ein Widerspruch erscheinen. In der klinischen Praxis wird ebenso kein Laborwert regelhaft verwendet, der spezifisch dem System Kopf/ Hals zugeordnet werden kann. Nichtsdestotrotz wurden zu Zwecken der Kongruenz alle Organsysteme für alle Kategorien beibehalten.

Beide Rater hatten bei einigen Falldarstellungen Schwierigkeiten, eine einheitliche Zuordnung von Symptomen zu den Organsystemen vorzunehmen. Dies lag daran, dass verschiedene Symptome wie z.B. Schwindel verschiedene Ursachen (neurologisch, psychiatrisch, kardiovaskulär...) haben können. Da die Musterlösungen der Fälle zum Zeitpunkt des Scorings nicht miteinbezogen wurden, war die Einordnung eines Symptomkomplexes in Hinblick auf die Enddiagnose nicht möglich. Aus diesem Grund wurde die methodische Gliederung in Organsysteme als Vorschlag dargestellt und kann individualisiert werden. Für eine beispielhafte Symptomzuordnung zu den einzelnen Organsystemen vgl. Tabelle 16. Obwohl diese Zuordnung nur als Vorschlag zu werten ist, hat die Verwendung des Scoring-Systems gezeigt, dass es für den standardisierten Gebrauch sinnvoll erscheint, eine solche Zuordnung festzulegen.

Tabelle 16: Vorschlag einer Zuordnung von Symptomen zu Organsystemen

Organsystem	Symptome			
Allgemein	Fieber	Schwitzen	Unwohlsein	Müdigkeit
Kopf-Hals	Kopfschmerzen	Halsschmerzen	Ohrenschmerzen	Schluckbeschwerden
Kardio-Vask	Brustschmerz	Brustenge	Palpitationen	Synkopen
Respi	Atemnot	Husten	Auswurf	Atemabhängiger Schmerz
GI	Übelkeit	Erbrechen	Bauchschmerzen	Obstipation
Nephro-Uro	Dysurie	Nykturie	Harnverhalt	Hämaturie
Musk	Rückenschmerzen	Muskelschwäche	Knieschwellung	Muskelkrampf
Neuro	Schwindel	Amnesie	Hemiparese	Parästhesien
Psych	Abgeschlagenheit	Angst	Unruhe	Wahnvorstellungen
Endokrin	Polydipsie	Hirsutismus	Exophthalmus	Polyurie
Häm-Onko	Nachtschweiß	Ungewollter Gewichtsverlust	Lymphadenopathie	s. Allgemein
Haut	Ekzem	Trockene Haut	Petechien	Hämatome

5.2 Ergebnisse der quantitativen Untersuchung

Nachdem gezeigt werden konnte, dass die Anwendung des neuen Scoring-Systems in Hinblick auf qualitative Merkmale einen Mehrwert bieten kann, konzentriert sich die folgende Darstellung auf die quantitative Analyse. Während im vorherigen Teil Merkmale, wie die Anwendbarkeit und die praktische Handhabung des Scoring-Systems relevant waren, soll die quantitative Auswertung der Daten zeigen, inwiefern die mit dem System bestimmten Komplexitätswerte die Schwierigkeit eines Falls vorhersagen können. Ebenso relevant ist die Frage, welche Komponenten einen Fall schwieriger und welche ihn möglicherweise einfacher machen.

5.2.1 RM und LLTM ermöglichen die Analyse der Daten

5.2.2 Die Komplexität eines Falls beeinflusst seine Schwierigkeit

Bei der Auswertung von Datensatz 1 zeigte sich eine hohe Korrelation zwischen den geschätzten Schwierigkeiten des RM und des LLTM ($r = .70$) (vgl. Abbildung 18). Dies weist darauf hin, dass für diesen Datensatz die Item Schwierigkeiten valide durch das festgelegte Item Charakteristikum Komplexität repräsentiert werden können.

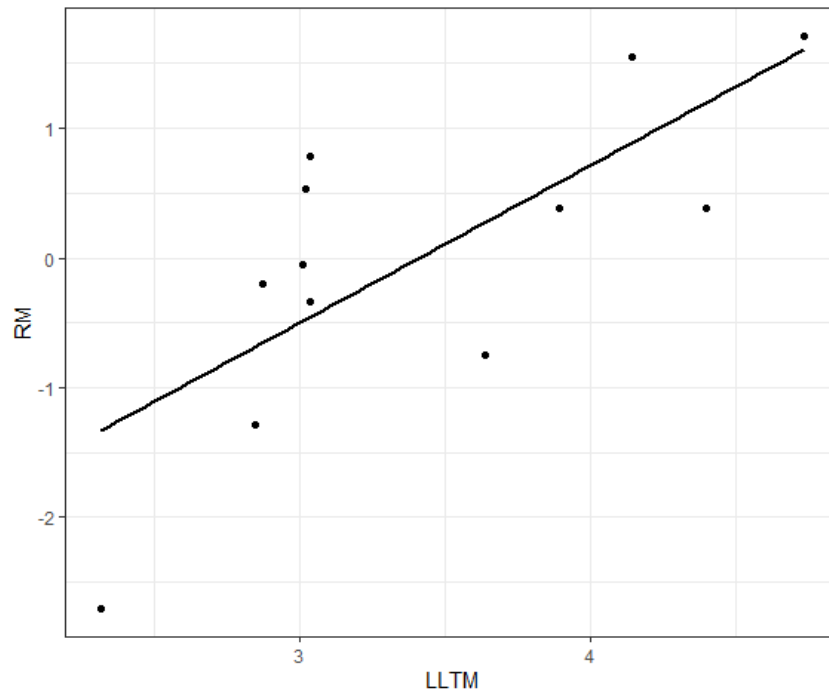


Abbildung 18: Korrelation der Schwierigkeiten aus RM und LLTM für Datensatz 1; Datenerhebung in Originalstudien (42, 74)

Datensatz 2 ergab ebenfalls eine hohe Korrelation zwischen den Schwierigkeiten des RM und des LLTM ($r = .70$) (vgl. Abbildung 19).

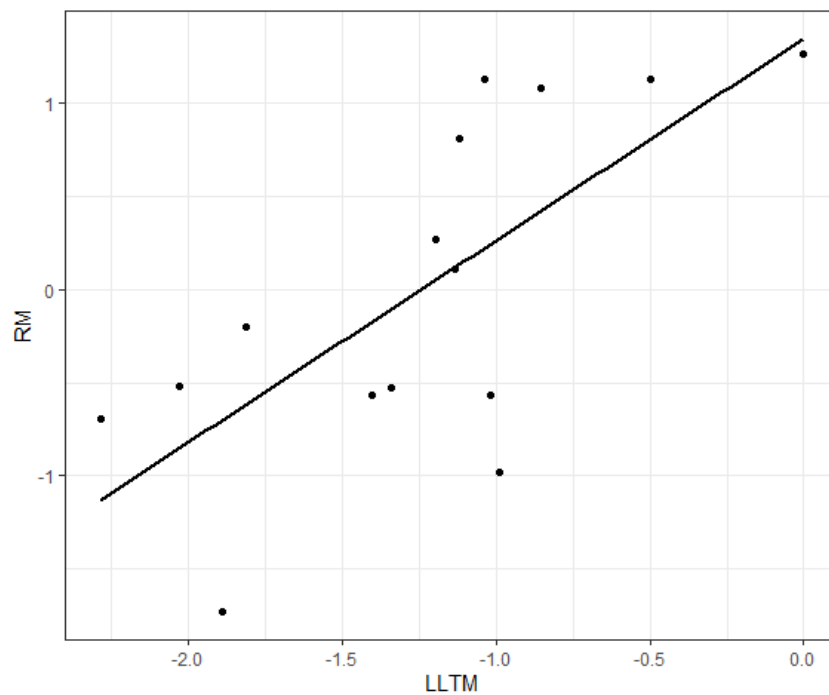


Abbildung 19: Korrelation der Schwierigkeiten aus RM und LLTM für Datensatz 2; Datenerhebung in Originalstudie (41)

Anhand von Datensatz 3 präsentierte sich eine mittlere Korrelation zwischen dem RM und dem LLTM ($r = .54$) (vgl. Abbildung 20).

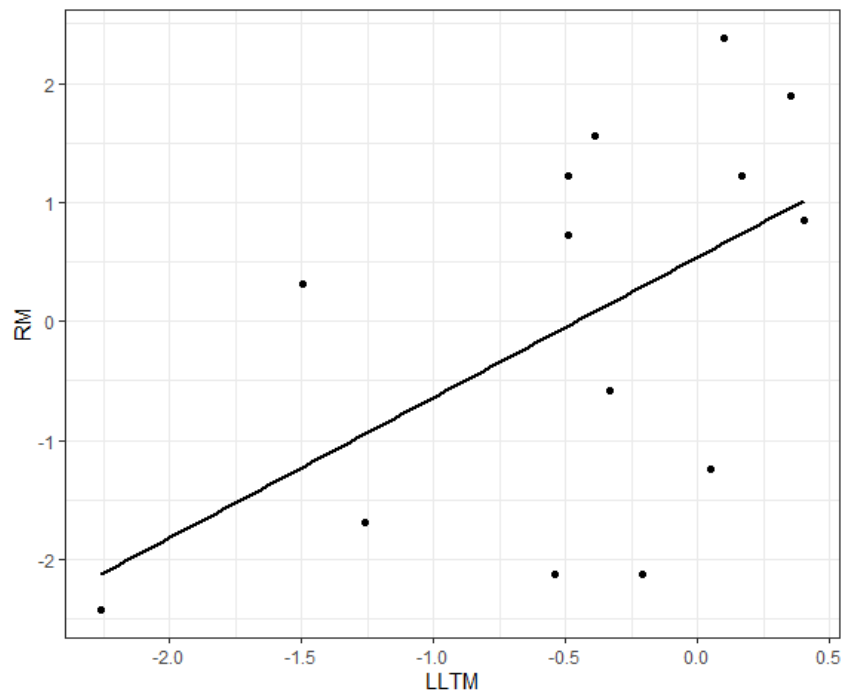


Abbildung 20: Korrelation der Schwierigkeiten aus RM und LLTM für Datensatz 3; Datenerhebung in Originalstudie (17)

5.2.3 Komplexität verschiedener Komponenten mit unterschiedlichem Einfluss auf die Fallschwierigkeit

Bei der Analyse der Gewichtungen der einzelnen Features zeigten sich Unterschiede, sowohl innerhalb der jeweiligen Datensätze als auch im Vergleich mit den anderen beiden Datensätzen.

Für Datensatz 1 ergaben sich positive Werte für die Anamnese, die Bildgebung und die Funktionsuntersuchungen. Komplexität innerhalb dieser Features macht die jeweiligen Fälle somit leichter, während Laborergebnisse und die körperliche Untersuchung einen Fall schwieriger machen (vgl. Tabelle 17).

Tabelle 17: Gewichtungen der Features für Datensatz 1

Dimension	Eta
Anamnese	0.21
Bildgebung	0.71
Labor	-0.07
Körperliche US	-0.23
Funktions US	0.04

Für Datensatz 2 ergibt sich ein ähnliches Bild. Im Vergleich zu Datensatz 1 tragen Informationen innerhalb der Anamnese jedoch dazu bei, dass ein Fall schwieriger wird (vgl. Tabelle 18).

Tabelle 18: Gewichtungen der Features für Datensatz 2

Dimension	Eta
Anamnese	-0.10
Bildgebung	0.23
Labor	-0.02
Körperliche US	-0.13
Funktions US	0.08

Die Gewichtungen von Datensatz 3 weisen keine Gemeinsamkeiten mit den anderen beiden Datensätzen auf. Lediglich die Laborergebnisse verringern die Schwierigkeit eines Falls während sie durch die anderen Features erhöht wird. Keiner der Fälle dieses Datensatzes enthielt Informationen, die den Funktionsuntersuchungen hätten zugeordnet werden können (vgl. Tabelle 19).

Tabelle 19: Gewichtungen der Features für Datensatz 3

Dimension	Eta
Anamnese	-0.07
Bildgebung	-0.23
Labor	0.12
Körperliche US	-0.01
Funktions US	k.A.

6 Diskussion

6.1 Diskussion der Modelle zur Erfassung von Fallkomplexität

Die Anwendung des Scoring-Systems von Braun et al. (2) hat einen großen Teil zur Annäherung an das Konstrukt der Fallkomplexität beigetragen. Besonders die Kategorisierung fallbezogener Faktoren ist ein Konzept mit viel Potential auf diesem Gebiet. Die Autoren haben damit die Basis für die vorliegende Weiterentwicklung des Modells zur Erfassung der Fallkomplexität geschaffen. Im Folgenden soll das Modell von Braun et al. (2) hinsichtlich der Konstruktion des neuen Scoring-Systems beleuchtet werden. Im Anschluss daran wird das neue System in Bezug auf inhaltliche Anforderungen sowie auf Qualitätsmerkmale diskutiert.

Bei der konkreten Anwendung des Scoring-Systems nach Braun et al. (2) auf 12 Fälle hatten beide Bearbeitenden Schwierigkeiten. Diskussionspunkte boten sich sowohl bei der Beurteilung der Linearität der Informationen als auch in Bezug auf die Kategorie der Nebendiagnosen. Beide Rater schätzten die Linearität der Komponenten eines Falls häufig unterschiedlich ein. Ein Grund dafür könnte sein, dass sich das Vorwissen und die klinische Erfahrung beider Rater (1: Facharzt, 2: Studentin) erheblich unterschied. Rater 1 schätzte in elf von 12 Fällen die Komplexität höher ein als Rater 2. Ein Faktor, der dieses Ergebnis vermutlich bedingt, ist die unterschiedliche Vorerfahrung auf dem Gebiet der Medizindidaktik und bei der Behandlung von Patienten. Mit vergleichsweise geringer klinischer Erfahrung gelang es der Studentin kaum, komplexe Linearitätsebenen innerhalb der Fälle zu erkennen. Darüber hinaus scheint es plausibel, dass ein Facharzt durch jahrelange klinische Erfahrung neue Linearitätsebenen oder Differentialdiagnosen entdeckt, die womöglich vom Autor des Falls überhaupt nicht beabsichtigt waren. Ein Beispiel eines Falls, bei dem sich beide Bearbeitenden uneinig waren, ist der einer Cholesterinwerterhöhung bei einem Patienten mit AV-Knoten-Reentrytachykardie. Einerseits ist eine Erhöhung des Cholesterins über den Normwert hinaus ein häufiger Zufallsbefund, andererseits ist er nicht wegweisend für die richtige Diagnose und könnte vom Autor auch als möglicher Distraktor verwendet worden sein. Während diese Information aus studentischer Sicht nicht gescored wurde, hat sie innerhalb der Beurteilung des Facharztes einen Punkt generiert. Diese Schwierigkeiten könnten einen Grund für die unbefriedigende Interrater-Reliabilität bei der Anwendung dieses Scoring-Systems darstellen.

Besonders innerhalb der Anamnese gibt es häufig eine Fülle verschiedener Informationen, die meist nur schwer nach ihrer Linearität beurteilt werden konnten. Teilweise lassen sich Aspekte der Anamnese bei der Bearbeitung eines Falls erst durch zusätzliche Laborbefunde oder eine Bildgebung einordnen. Da in der Praxis oft keine Musterlösungen der Fälle vorliegen, wurden elf von 12 Fällen ohne Kenntnis der vorgesehenen Diagnose bearbeitet. Diese wurde erst später für die inhaltliche Betrachtung herangezogen. Ein Ziel bei der Bearbeitung eines Falls ist es, dass am Ende verschiedene Puzzleteile ein Gesamtbild ergeben und zu der richtigen Diagnose führen. Die Frage, die sich zunächst daraus ergab, war, ob die Beurteilung der Linearität möglicherweise erst nach Betrachtung des gesamten Falls mit Kenntnis der anschließenden Diagnose zuverlässig durchgeführt werden kann.

Die Beurteilung und das Scoring möglicher Nebendiagnosen knüpfen thematisch an die Diskussion der Linearitätsebenen an. Analog zu den Problemen bei der Entscheidung, ob Informationen linear zueinander sind, gestaltete sich das Scoring von Nebendiagnosen schwierig. Auch für diesen Schritt ist eine tiefergreifende Beurteilung des Falls hinsichtlich einer etwaigen Diagnose und möglicher Nebendiagnosen erforderlich. Genauso, wie die individuelle Schwierigkeit eines Falls von vielen Faktoren abhängen kann, so ist auch die Fähigkeit zur klinischen Entscheidungsfindung, bzw. die Diagnosekompetenz, individuell unterschiedlich. Da diese Fertigkeiten z.B. durch die Vorerfahrung und den Kenntnisstand des Einzelnen beeinflusst werden, konnten sie nicht als Voraussetzungen zur Benutzung des neuen Scoring-Systems implementiert werden. Dieses soll primär dazu dienen, die reine fallbezogene Komplexität zu messen und von variablen Faktoren größtenteils unabhängig sein.

Anhand der Anwendung des Systems nach Braun et al. (2) wurde deutlich, dass das Messen der Komplexität eines Falls der Medizindidaktik einen großen Mehrwert bieten kann. Nichtsdestotrotz hat die Arbeit mit diesem Modell (2) gezeigt, dass nicht zuletzt aus Gründen der Intersubjektivität eine Weiterentwicklung erforderlich ist.

Das neue Scoring-System CASE für die Messung von Fallkomplexität wurde unter Beachtung verschiedener Anforderungen an ein Scoring-System entwickelt. Wenn man z.B. die Glasgow-Coma-Scale (GCS) als einen der bekanntesten Scores im klinischen Alltag heranzieht, werden bestimmte Voraussetzungen

deutlich, die ein Punktesystem erfüllen muss. Um anwenderunabhängig, zuverlässig und schnell eine Aussage über den Zustand eines Schädel-Hirn-Verletzten Patienten zu treffen, muss die GCS die Kriterien der Objektivität, Reliabilität und Validität erfüllen (84). Das Anwendungsgebiet für das vorgestellte Scoring-System weicht zwar von dem, typischer klinischer Scores ab, sollte aber dennoch nach üblichen Gütekriterien beurteilt werden.

Damit die Verwendung in der Praxis sinnig ist, sollte die Objektivität des Systems erfüllt sein. Ein Messinstrument ist objektiv, sofern es den gemessenen Parameter unabhängig von den verschiedenen Anwendern und deren Interpretation erhebt (85). Bei einem Scoring-System für Fälle bedeutet das, dass im Optimalfall für den gleichen Fall immer der gleiche Score von verschiedenen Personen erhoben wird. Gleichzeitig, neben der Objektivität, ist es wichtig, dass ein Modell auch reliabel ist. Damit wird die Zuverlässigkeit eines Systems beschrieben. Ist dieses reliabel, so misst die Methode ein bestimmtes Merkmal genau und ohne Irrtümer (85).

Für das Scoring-System CASE wurden diese beiden Parameter anhand der Interrater-Reliabilität überprüft. Während die Interrater-Reliabilität von manchen Autoren als Maß der Reliabilität verwendet wird (86), dient sie, je nach Forschungszweck, auch zur Untersuchung der Objektivität eines Verfahrens (87). Als statistisches Maß für die Bestimmung der Interrater-Reliabilität wurde Cohen's Kappa verwendet. Eine Betrachtung der Übereinstimmung zweier Rater ist an bestimmte Bedingungen, bzw. Voraussetzungen, geknüpft. Eine davon ist, dass die Benutzung des Scoring-Systems genau nach den festgelegten Vorgaben stattfindet und die Anwender in der Benutzung geübt sind. Für die Endpunktzahl spielt es in CASE eine untergeordnete Rolle, ob eine gewisse Information beispielsweise in der Kategorie „psychosoziale Anamnese“ oder unter „Familienanamnese“ eingruppiert wird. Wichtiger ist, dass die Kriterien eingehalten werden, wann wie viele Punkte für eine bestimmte Information vergeben werden. Bevor beide Rater die 59 Fälle bearbeiteten, haben sie an der Entwicklung des Scoring-Systems aktiv mitgewirkt und dessen Kategorien definiert. Durch diese Zusammenarbeit wurde beabsichtigt, möglichst genaue Standards für die Bearbeitung der Fälle zu schaffen. Weitere Absprachen für das Scoring wurden nicht getroffen, um das Ergebnis darüber hinaus nicht zu beeinflussen. Ein weiterer Faktor, der das Interrater-Ergebnis beeinflussen kann, ist der Fall selbst. Cohen's

Kappa der Gesamtwerte lag mit 0.53 in einem Bereich, der zwar für eine moderate Übereinstimmung spricht, im Vergleich zu den inhaltlichen Untergruppen allerdings geringer ausfällt. Um dieses Ergebnis näher zu beleuchten, wurde der Interrater-Koeffizient für eine Gruppe von Fällen mit niedriger Komplexität unter 60 und für eine mit hoher Komplexität über 60 bestimmt. Dies hat gezeigt, dass ein maßgeblicher Unterschied für beide Gruppen besteht (0.64 für die Gruppe niedriger Komplexität zu 0.17 für die Gruppe hoher Komplexität). Dieses Ergebnis kann verschiedene Gründe haben. Im Beispiel von Fallszenarien mit niedriger Komplexität müssen bei der Bearbeitung meist weniger Informationen prozessiert oder gescored werden als bei sehr komplexen Fällen. Durch diesen Umstand ergeben sich bei komplexen Falldarstellungen a priori mehr Möglichkeiten, als Rater abzuweichen. Trotz aller Sorgfalt beider Rater bei der Bearbeitung der Fälle kann es vorkommen, dass bei komplexen Fallbeschreibungen, die viele Komponenten enthalten, bestimmte Informationen überlesen und somit nicht gescored werden. Die geringe Übereinstimmung bei komplexen Fällen kann ebenfalls darauf hindeuten, dass das Scoring-System möglicherweise die Komplexität für diese Gruppe von Fällen weniger genau misst, bzw. eventuell überarbeitet werden muss. Innerhalb der Untergruppen weist die Dimension der Anamnese mit 0.57 den geringsten Wert für den Interrater-Koeffizient auf. Diese Gruppe umfasst sowohl das Scoring der einzelnen Kategorien wie beispielsweise die Symptome und Vorerkrankungen, als auch die Eingruppierung der Symptome in Organsystemklassen. Diese Einteilung könnte eine Erklärung für das Interrater-Ergebnis darstellen. Obwohl die Organsystemklassen vorher durch eine beispielhafte Symptomzuordnung definiert wurden, lassen sie aktuell viel Interpretationsspielraum. So können Symptome je nach Ursprung verschiedenen Organsystemen zugeordnet werden. Werden mehrere Symptome zu einem Organsystem zugeordnet, generieren diese jedoch keine zusätzlichen Punkte. Ein Ansatz, der zukünftig geprüft werden sollte, ist, dass die Organsystemklassen innerhalb der Anamnese genauer definiert oder je nach Anwendungsgebiet angepasst werden, um dort Irrtümer bei der Kategorisierung zu minimieren.

Trotz dieser Beobachtungen zeigte sich für sowohl für die Gesamtwerte als auch für die Untergruppen des Scoring-Systems eine moderate bis nahezu perfekte Übereinstimmung in der Anwendung durch die beiden Rater. Dies deutet darauf hin, dass das Scoring-System im Rahmen der vorliegenden Untersuchungen

zuverlässig und objektiv angewandt werden konnte. Beide Rater befanden sich zum Zeitpunkt der Beurteilung gegen Ende des klinischen Abschnittes an der Ludwig-Maximilians-Universität. Trotz einiger Unterschiede in der Vorerfahrung mit der Betreuung von Patienten (Arbeit im Rettungsdienst, etc.) gelang es beiden, das Scoring-System standardisiert anzuwenden. Dies lässt vermuten, dass Medizinstudierende im klinischen Abschnitt bereits über die nötigen Fähigkeiten verfügen, um die Komplexität von Fällen anhand des Manuals zu bestimmen. Eine Bedingung für die Anwendung des vorgestellten Systems ist jedoch sicherlich ein gewisses Maß an medizinischem Wissen und ein Überblick über die verschiedenen Strukturen des jeweiligen Gesundheitssystems. Praktische Vorerfahrungen im medizinischen Bereich sind ebenso hilfreich. In der Praxis soll das Scoring-System vorrangig dazu dienen, bestehende Fälle innerhalb des Curriculums zu beurteilen. Da dies meist von erfahrenen Ärzten geleistet wird, sollten sie über einen ausreichenden Kenntnisstand verfügen, um das System anzuwenden.

Bereits Braun et al. (2) heben für die Etablierung eines Scoring-Systems zwei Voraussetzungen hervor: „Generalisierbarkeit und Praktikabilität“ (2). CASE als Vorschlag für ein solches System konnte auf Fälle verschiedenster Autoren und Formate angewendet werden. Ein Grund dafür ist, dass viele Fälle einen ähnlichen Aufbau haben und die gleichen Komponenten beinhalten. Meist finden sich Angaben zur Anamnese, einer körperlichen Untersuchung und technisch erhobenen Befunden (z.B. Laborwerte oder EKG). Einige Fälle beinhalten zusätzliche Informationen wie etwa die vorangegangene Krankengeschichte oder schildern die aktuelle Lebenssituation des Patienten. Solche Aspekte werden teilweise bewusst innerhalb einer Lernumgebung eingesetzt, um einen Fall glaubwürdig und realitätsgetreu darzustellen (7). In dem vorgeschlagenen Scoring-System finden sich eine Vielzahl an Kategorien, um möglichst jede Information, die in den untersuchten Fällen enthalten ist und relevant für den Bearbeitungsprozess sein kann, einordnen zu können. Aus diesem Grund sind die Kategorien in dem Scoring-System derart feingliedrig gestaltet (vgl. Tabelle 6, S. 41). Es finden sich allerdings auch Fälle, die z.B. nur die Symptome eines Patienten und keine weiteren Informationen enthalten. Diese können genauso im System eingruppiert werden, indem lediglich Punkte in einer Rubrik vergeben werden.

Generell soll das System CASE, besonders in Bezug auf seine Kategorien, lediglich einen Vorschlag darstellen, der für die innerhalb dieser Arbeit bearbeiteten Fälle anwendbar war. Obwohl gewisse Klassen oder ganze Kategorien überhaupt nicht bis wenig Verwendung bei der Anwendung fanden, wurde die Struktur des Systems bewusst beibehalten, um „Generalisierbarkeit und Praktikabilität“ (2) zu gewähren. Dies ist ein Grund, wieso das gesamte Manual die gleichen Klassen aufweist. Ein weiterer Grund ist das Ziel, Raum für zukünftige Forschung und die Nutzung innerhalb verschiedener medizinischer Disziplinen zu schaffen. Die genannten Klassen sollen als ein Vorschlag betrachtet werden, sind aber gleichzeitig allgemein genug gehalten, um bei Bedarf eine Individualisierung des Scorings zu ermöglichen. Innerhalb der Medizin gibt es sicherlich spezielle Fachgebiete, für die die ein oder andere Kategorie oder Klasse in diesem System ergänzt werden müsste. Passt ein Detail in keine der beschriebenen Klassen, gibt es die Möglichkeit, eine neue zu ergänzen oder es unter die Klasse „Allgemein“ zu subsummieren. Das eigentliche Scoring kann trotzdem nach dem gleichen Prinzip erfolgen.

Obwohl es für die Erhebung der Endpunktzahl meist zweitrangig ist, in welche der Kategorien eine Information eingruppiert wird, ist es dennoch zu empfehlen, sich nicht nur quantitativ, sondern auch inhaltlich an den Strukturen des Systems zu orientieren. Diese können hilfreich sein, wenn inhaltliche Analysen einzelner Kategorien vorgenommen werden sollen.

Das vorgestellte System zur Messung der Fallkomplexität bietet zwar viele Möglichkeiten, hat jedoch auch einige Beschränkungen. Zur Validierung eines Scoring-Systems gehört neben den behandelten Aspekten ebenso die Frage nach der Vergleichbarkeit und der Aussagekraft der Ergebnisse. Mit dem neuen System ist es möglich, eine Punktzahl für die Fallkomplexität zu generieren. Je höher der Wert, desto komplexer ist ein Fall. Bereits Braun et al. empfehlen für die Weiterentwicklung ihres Scoring-Systems eine Unterteilung der Ergebnisse in wenig komplex bis sehr komplex, um den Wert einordnen zu können (2). Da CASE allerdings keine Maximalpunktzahl vorgibt, ist es schwierig, Komplexitätsgrade zu vergeben. Theoretisch kann die Punktzahl bei der Anwendung des vorliegenden Scoring-Systems unendlich hoch ausfallen, da der Informationsgehalt eines Falls nicht begrenzt ist. Die Komplexität eines realen Patienten ist ebenso wenig eingrenzbar, auch wenn eine Überschreitung von einer bestimmten Anzahl an

Symptomen oder Untersuchungen unwahrscheinlich erscheint. Innerhalb der Lehre besteht somit die Möglichkeit für das Angebot und das Scoring besonders komplexer Fälle, wie sie wahrscheinlich nicht im Medizinstudium, sondern vorrangig in der berufsbezogenen Weiterbildung Verwendung finden. Anhand des Manuals bleibt somit die Möglichkeit bestehen, die Komplexität an die individuellen Bedürfnisse unterschiedlich vorwissensstarker Lernender anzupassen. Die fehlende Vorgabe von Komplexitätsgraden erschwert allerdings die Interpretation der einzelnen Komplexitätswerte und könnte deren Aussagekraft begrenzen. Um das Scoring-System dennoch sinnvoll zu nutzen, könnte es hilfreich sein, individuelle Maßstäbe, je nach Anwendungsgebiet festzulegen. Zur Definition von Lernzielen im Medizinstudium kann der Nationale Kompetenzbasierte Lernzielkatalog Medizin (NKLM) herangezogen werden. Dieser definiert Lernziele durch verschiedene, sowohl theoretische als auch praktische Kompetenzen, die Studierende erlernen sollen (88). Bei dem Einsatz fallbasierter Lehr-Lernmethoden könnte die Vorgabe von Komplexitätsgraden sinnvoll sein, um die erforderliche Komplexität des Lernmaterials, je nach Lernziel, bzw. Kompetenz, anzupassen.

Um die erste Forschungsfrage aufzugreifen, inwiefern die Komplexität medizinischer Fälle reliabel und valide erfasst werden kann, müssen die diskutierten Aspekte herangezogen werden. Die Frage ist insofern zu bejahen, als dass es gelungen ist, ein Scoring-System zu entwickeln, mit dem die Komplexität im Rahmen der vorliegenden Untersuchungen standardisiert und anwenderunabhängig erfasst werden konnte. Während dieses System einige Anforderungen an ein Scoring-System erfüllt (z.B. eine solide Interrater-Reliabilität), waren andere wiederum schwer zu überprüfen. Eines dieser Kriterien ist die Validität, also die Frage, ob eine Methode das misst, was sie zu messen beabsichtigt (89). Da es bislang kein etabliertes Messinstrument für Fallkomplexität gibt (2), kann für die Validierung des neuen Scoring-Systems lediglich auf wenige Definitionen früherer Forschung zurückgegriffen werden. Komplexität wird hierbei meist als Folge der Informationen innerhalb eines Falls angesehen. Sie wird direkt durch das Material verursacht und ist weitgehend unabhängig von externen Faktoren, die z.B. die Person betreffen, die den Fall bearbeitet (2, 63). Das Grundkonzept von CASE basiert auf dieser Definition und hat sich im Rahmen der durchgeführten Untersuchungen als praktikabel dargestellt.

In manchen Quellen wird die Validität eines Systems durch die Interpretation des Punktwertes gewährleistet (90). Anhand des Scoring-Systems kann zwar eine Punktzahl generiert werden, daraus folgt allerdings noch keine Möglichkeit zur Einordnung der Werte. Da die Vergleichbarkeit mit anderen Scoring-Systemen und die Aussagekraft der Komplexitätswerte ohne weiterführende Forschung aktuell noch begrenzt sind, kann diese Komponente der Forschungsfrage nicht abschließend beantwortet werden. Um dennoch eine Einschätzung vornehmen zu können, konzentriert sich die Validierung des Scorings-Systems CASE darauf, die beobachtete Fallschwierigkeit anhand der Komplexitätswerte vorherzusagen.

6.2 Betrachtung der beobachteten Korrelation zwischen Komplexität und Schwierigkeit von Fällen

Ein Ziel dieser Arbeit war es, zu prüfen, ob die Schwierigkeit von Fällen anhand ihrer Komplexität vorhergesagt werden kann. Die Ergebnisse legen nahe, dass die Item-Schwierigkeit tatsächlich in zwei von drei Datensätzen durch das Item-Charakteristikum Komplexität vorhergesagt werden kann (für Datensatz 3 ist dieser Zusammenhang ebenso erkennbar, wenn auch in schwächerem Ausmaß als bei den Datensätzen 1 und 2). Die These: „je komplexer ein Fall, desto schwieriger ist er“, kann somit bezogen auf die Gesamtkomplexität und die hier angewandten Methoden mit Vorbehalt bestätigt werden.

Betrachtet man den Einfluss der Komplexitätswerte einzelner Komponenten auf die Gesamtschwierigkeit, ergibt sich ein heterogeneres Bild. Die Ergebnisse zeigen, dass die fünf untersuchten Features (Anamnese, Bildgebung, Labor, Körperliche Untersuchung, Funktionsuntersuchungen) unterschiedlichen Einfluss auf die Gesamtschwierigkeit eines Falls ausüben. Diese Beziehungen sind nicht einheitlich und unterscheiden sich innerhalb der drei Datensätze. Innerhalb des ersten Datensatzes trägt die Komplexität der Features Anamnese, Bildgebung und Funktionsuntersuchungen zur Verringerung der Schwierigkeit bei, während die Komplexität der Laborergebnisse und der körperlichen Untersuchung die Gesamtschwierigkeit erhöht. Vergleicht man diese Ergebnisse mit denen des zweiten Datensatzes, ergibt sich für alle Features mit Ausnahme der Anamnese der gleiche Zusammenhang. Komplexität der Anamnese macht einen Fall innerhalb des zweiten Datensatzes nicht leichter, sondern schwieriger. Für den dritten Datensatz zeigen sich dagegen völlig abweichende Ergebnisse. Dieses heterogene

Bild der Gewichtungen der einzelnen Features stellt die zweite Forschungsthese infrage. Die Aussage „je komplexer ein Fall, desto schwieriger ist er“ lässt sich somit vorerst nicht auf die einzelnen Features anwenden.

Diese Ergebnisse sollten jedoch nicht per se als widersprüchliche Beobachtungen interpretiert werden. Eine mögliche Erklärung ergibt sich bei einer qualitativen Betrachtung der Fälle. Während die Datensätze 1 und 2 auf einheitlich strukturierten Fällen aus der Lernumgebung CASUS basieren, deren Gewichtungen sich Großteils ähneln, basiert Datensatz 3 auf kurzen Fällen, die abweichend strukturiert sind. Dies legt nahe, dass der unterschiedliche Einfluss verschiedener Features auf die Fallschwierigkeit in Zusammenhang mit qualitativen Aspekten eines Falls stehen könnte. Während manche Informationen innerhalb eines Falls die Lernenden zu Denkprozessen anregen oder sogar bewusst als Distraktoren verwendet werden, gibt es sicherlich andere Arten von Informationen, die einen Fall sogar einfacher machen und zu einer korrekten Lösung beitragen. Dieser variable Einfluss unterscheidet sich möglicherweise nicht nur zwischen unterschiedlich konstruierten Fällen, sondern auch zwischen jedem einzelnen Fall.

Diese Ergebnisse müssen auch unter Beachtung der Art der Datenerhebung innerhalb der Originalstudien differenziert betrachtet werden. Für Datensatz 1 und 2 haben im Unterschied zu Datensatz 3 Interventionen, überwiegend in Form von Scaffolding, bei der Bearbeitung der Fälle stattgefunden (17, 41, 42, 74). Diese Interventionen könnten sich auf die Richtigkeit der Diagnosen der Studierenden ausgewirkt haben, die für die Analyse innerhalb dieser Arbeit herangezogen wurde. Die zu Datensatz 1 zugehörige Studie hat ergeben, dass kein signifikanter Einfluss des Scaffolding auf die Diagnoserichtigkeit nachgewiesen werden konnte. Lediglich die Effizienz in Bezug auf die Dauer der Fallbearbeitung wurde durch die Intervention positiv beeinflusst (42).

Für Datensatz 2 ergibt die zugehörige Studie ein ähnliches Resultat. Auch hier wurde die Diagnoserichtigkeit durch keine der Interventionen signifikant beeinflusst. Innerhalb dieser Studie konnte gleichermaßen kein Einfluss der Maßnahmen auf die Diagnoseeffektivität, verglichen mit der Kontrollgruppe, gezeigt werden (41).

Da aufgrund der Ergebnisse der Originalstudien nicht per se von einer Beeinflussung der Diagnoserichtigkeit durch die vorliegenden Interventionen

ausgegangen werden kann, scheint es weiterhin möglich, die Resultate der vorliegenden Analyse der drei Datensätze miteinander zu vergleichen. Trotz allem kann eine Beeinflussung der Diagnoserichtigkeit durch gewisse Parameter nicht ausgeschlossen werden. Welche Faktoren Einfluss auf die Richtigkeit einer Fallbearbeitung haben, muss in Zukunft weitergehend untersucht werden. Da diese Wirkbeziehungen im Falle der vorgenommenen Analyse nicht abschließend geklärt sind, sind die dargestellten Ergebnisse sicherlich mit Vorbehalt und differenziert zu betrachten. In der Realität gestaltet sich die Betrachtung dieser Wirkbeziehungen aufgrund der Vielzahl möglicher geplanter und ungeplanter Interventionen schwierig. Dies muss für zukünftige Studien beachtet und untersucht werden.

Zusammenfassend konnte im Rahmen der vorgenommenen Analyse ein positiver Zusammenhang zwischen der Gesamtkomplexität eines Falls und seiner Schwierigkeit gezeigt werden. Wichtig ist dabei die Beobachtung, dass die Komplexität einer einzelnen Kategorie einen Fall nicht notwendigerweise schwieriger macht. Hierbei könnte z.B. die Art der Information, die Typizität oder auch die Kongruenz der Komponenten eine Rolle spielen. Weiterführende Betrachtungen sollten deshalb besonders den Zusammenhang zwischen der Komplexität und der Typizität eines Falls untersuchen und eventuelle Trends bei der Gewichtung einzelner Kategorien identifizieren. Diese Überlegungen werfen die Frage auf, ob beispielsweise eine Beurteilung der Linearität (2) als zusätzliche Komponente neben dem reinen Informationsgehalt in das Scoring-System implementiert werden sollte. Voraussetzung für zukünftige Aussagen diesbezüglich sollte jedoch ein tiefergreifendes Verständnis der Fallkomplexität, der Fallschwierigkeit und ihrer Wirkbeziehungen sein.

6.3 Einordnung der Ergebnisse in den Kontext der Medizindidaktik

Das vorgestellte Scoring-System CASE stellt eine Möglichkeit dar, Fälle standardisiert nach ihrer Komplexität zu beurteilen. In der Vergangenheit wurde bereits von verschiedenen Autoren betont, dass solch ein System einen didaktischen Mehrwert bieten kann (2, 60). Die Forschungsergebnisse dieser Arbeit unterstützen diese These.

Für die Gestaltung von Hochschullehre sollte es möglich sein, das genutzte Fallmaterial an zukünftige Entwicklungen anzupassen und um sinnvolle Methoden zu ergänzen. Das neue Scoring-System bietet die Möglichkeit, die Komplexität von Fällen an die Bedürfnisse des Einzelnen anzupassen. Es leistet somit einen Beitrag zur häufig geforderten Standardisierung fallbasierter Lehre (2, 40).

Lernen endet nicht mit Abschluss des Studiums. Die Weiterbildung von Ärztinnen und Ärzten ist wichtig, um eine gute Qualität der medizinischen Behandlung zu sichern. Umso mehr Berufserfahrung ein Einzelner gesammelt hat, desto mehr Patienten bzw. Fälle hat er behandelt. Für weniger erfahrene Berufsanfänger könnten Fälle ein wirksames Instrument sein, um sie auf den Berufsalltag vorzubereiten. Geübte Kliniker können vermutlich ebenso von einer fallbasierten Weiterbildung profitieren, allerdings müssten Fälle für diese Zielgruppe anspruchsvoller und somit komplexer gestaltet werden. Durch die Beurteilung der Fallschwierigkeit könnten Fälle an einen bestimmten Erfahrungsstand angepasst werden.

Das in dieser Arbeit vorgestellte System bietet nicht nur die Möglichkeit, bereits vorhandene Fälle einzuordnen, sondern könnte auch für die Erstellung von Fallvignetten nützlich sein. Hierfür bietet CASE nicht nur die Möglichkeit einer quantitativen Bewertung der Fallkomplexität, sondern erlaubt auch die inhaltliche Einordnung, sowie zielgerichtete Gestaltung eines Falls. Fallbasierte Lernumgebungen verfolgen oft ein bestimmtes Lernziel oder sollen ein gewisses Thema abdecken. Mithilfe des Manuals können die Informationen eines Falls entweder bewusst über verschiedene Kategorien verteilt werden oder sich auf eine ganz bestimmte Kategorie konzentrieren. Soll ein Fall beispielsweise die Diagnosekompetenz anhand von Röntgenbildern fördern, so kann der Informationsgehalt der restlichen technischen und körperlichen Untersuchung geringgehalten werden und der Fokus auf einer aussagekräftigen Bildgebung liegen. Um Fälle anhand des Systems zu konstruieren, kann das Scoring nach einem umgekehrten Prinzip erfolgen. Zunächst kann die gewünschte Anzahl an Punkten in den jeweiligen Kategorien vergeben werden, die durch den Informationsgehalt abgedeckt werden sollen. Die Komplexität kann ebenfalls im Voraus anhand der Endpunktzahl festgelegt werden. Im Anschluss könnten die Zahlenwerte in konkrete Informationen umgewandelt und somit ein Fall mit vordefinierter Komplexität erstellt werden.

Fischer et al. betonen die Relevanz der sorgfältigen Auswahl des richtigen Fallmaterials, je nach Lernziel und gewünschter Schwierigkeit (40). Dass dies in Zukunft eine Herausforderung für die Medizindidaktik darstellt, wird auch durch die Ergebnisse der Untersuchungen anhand von CASE deutlich. Besonders die Erkenntnis, dass Komplexität verschiedener Features einen unterschiedlichen Einfluss auf die Schwierigkeit eines Falls ausüben könnte, hat möglicherweise Bedeutung für die Verwendung von Fallszenarien. Darüber hinaus können neben der Komplexität auch andere Faktoren wie z.B. verschiedene Arten von Scaffolds die Schwierigkeit einer Lernumgebung beeinflussen (40). Diese Beobachtungen zeigen, dass es für die Näherung an das Gesamtkonzept Fallschwierigkeit nicht ausreichend ist, einzelne Parameter mit möglichem Einfluss zu identifizieren. Einerseits sollten diese Faktoren näher beleuchtet werden und andererseits ihre Wirkbeziehungen mit der Fallschwierigkeit, sowie Interaktionen untereinander, untersucht werden.

6.4 Limitationen der Forschung

Das vorgestellte Scoring-System stellt letztendlich ein Modell dar, dass die Fallkomplexität auf Basis einiger Grundannahmen quantifizieren kann. Es ist anzunehmen, dass das Konzept der Komplexität jedoch weitaus vielschichtiger ist, als es mit einem System abgebildet werden kann.

Obwohl das Scoring-System auf eine Vielzahl verschiedener Fälle angewandt werden konnte, stellen diese nur einen Auszug möglicher Fallpräsentationen dar. Aufgrund ihrer Verfügbarkeit und Zugänglichkeit stammen die meisten dieser Falldarstellungen aus dem Repertoire der LMU und wurden auf ihrer internen Lernplattform CASUS bearbeitet. Diese Fallauswahl deckt sicherlich nicht alle Arten von Fällen oder Fachgebieten ab. Besonders die Untersuchung der zweiten Forschungsfrage wurde durch drei Datensätze vorgenommen, von denen sich zwei in Bezug auf den Aufbau der bearbeiteten Fälle sehr ähneln. Bislang ist unklar, inwiefern die Komplexität einzelner Kategorien und die Gesamtschwierigkeit durch die Struktur verschiedener Fälle beeinflusst werden kann. Aus diesem Grund sollten zukünftige Untersuchungen Fallbearbeitungen mit Fokus auf unterschiedliche Arten von Fällen analysieren, um diesen Zusammenhang zu klären. Bei der Bearbeitung von Fällen mit hoher Komplexität zeigte sich eine niedrigere Interrater-Reliabilität als bei Fällen mit niedriger Komplexität. Dieses

Ergebnis könnte darauf hindeuten, dass das Scoring-System an besonders komplexen Fällen weitergehend geprüft werden sollte. Fokus sollte hierbei auf möglichen Fehlerquellen während des Scoring Prozesses liegen, da bislang nicht klar ist, ob die Fülle an Informationen selbst für eine geringere Übereinstimmung beider Rater sorgt, oder, ob das Scoring-System für derart komplexe Fälle angepasst werden muss. Für diese Untersuchungen eignen sich z.B. Fälle aus dem NEJM (20–22), da diese im Gegensatz zu Fällen in Casus (vgl. Anhang #2) meist über viele Informationsebenen und einem komplexen Krankheitsverlauf verfügen.

Die Basis eines jeden Scoring-Systems sollte eine genaue Definition des Parameters sein, den es zu messen beabsichtigt. Anhand dieser Arbeit wurde das Konzept der Fallkomplexität beleuchtet und eingegrenzt. Dennoch bleibt die Komplexität von Fällen vorerst ein Konstrukt, dass nicht ausreichend erforscht ist. Darüber hinaus oder vielleicht genau aus diesem Grund bestehen bislang wenige Vorschläge für Systeme zur Erfassung der Fallkomplexität, wovon keines in der Praxis etabliert ist (2, 60). Demzufolge gestaltete sich der Prozess der Validierung des neuen Systems schwierig. Dieses wurde zwar unter Berücksichtigung früherer Forschungsergebnisse (2, 60, 63) konstruiert und erprobt, dennoch lassen sich wenige Aussagen zu seiner Messgenauigkeit treffen. Eine Art der Validierung, die Vorhersage von Fallschwierigkeit anhand von Komplexitätsscores, konnte dennoch durchgeführt werden.

6.5 Empfehlungen für weiterführende Untersuchungen

Ein Schwerpunkt zukünftiger Untersuchungen sollte auf der weiteren Erprobung des vorgestellten Scoring-Systems liegen. Ein wichtiger Baustein ist dabei die Anwendbarkeit von CASE durch verschiedene Personen, wie z.B. Studierende zu Beginn der Ausbildung, erfahrene praktizierende Ärzte und Lehrbeauftragte. Durch die Überprüfung der Anwendbarkeit könnten sich weitere Aspekte für die Weiterentwicklung des Systems ergeben.

Ein Ansatzpunkt für die Überarbeitung des Systems ist der Vorgang der Punktevergabe. Aktuell erfolgt das Scoring linear, das bedeutet, dass jede Information eines Falls mit gleichem Gewicht zur Endpunktzahl beiträgt. Ob dieses System die Realität abbildet, muss allerdings Gegenstand zukünftiger Forschung sein. Es ist denkbar, dass verschiedene Arten von Informationen einen unterschiedlich

starken Einfluss auf die Gesamtkomplexität haben. Ist dies tatsächlich der Fall, wäre die Einführung einer Gewichtung der jeweiligen Kategorien erforderlich. Dies könnte beispielsweise anhand von Exponentialfaktoren umgesetzt werden. Die Bestimmung der Gesamtwerte würde somit nicht aus der reinen Addition der einzelnen Punkte, sondern aus einer Anwendung einer zusammengesetzten Formel bestehen.

Eine weitere Thematik, die daran anknüpft, ist die Heterogenität der Komplexität als möglicher Forschungsgegenstand. Die Anwendung von CASE auf verschiedene Fälle macht deren unterschiedlichen Aufbau deutlich. Daraus ergibt sich eine zum Teil völlig variable Verteilung der Informationen auf verschiedene Dimensionen und Kategorien und somit auch eine unterschiedliche Verteilung der Komplexität. Während manche Fälle lediglich aus einer Anamnese Sequenz bestehen (homogen), erstreckt sich der Informationsgehalt anderer Fälle gleichmäßig auf nahezu alle Bausteine des Systems (heterogen). Eine Analyse, ob die Heterogenität der Komplexität Einfluss auf die Schwierigkeit eines Falls nimmt und inwiefern sie in Verbindung mit anderen Parametern wie z.B. der Typizität steht, wäre wünschenswert.

Für diesen Zweck könnte eine Optimierung der Darstellung von Komplexitätswerten etabliert werden. Wünschenswert wäre es, ein System zu entwickeln, mit dem anhand des endgültigen Scores sofort ersichtlich ist, wenn ein Fall z.B. ausschließlich Informationen innerhalb der Anamnese enthält. Um den quantitativen Wert der Komplexität durch eine inhaltliche Komponente zu ergänzen, wäre es denkbar, dem Endwert weitere Informationen hinzuzufügen. Ein Beispiel dafür stellt die Unterteilung in die beiden Dimensionen der Anamnese und der Untersuchung dar. Der Punktwert für einen Fall mit der Komplexität 45, 22 Punkten für die Anamnese und 23 Punkten für die Untersuchungen, könnte folgendermaßen dargestellt werden: 45 (22/23). Darüber hinaus besteht die Möglichkeit, diese Aufteilung noch feingliedriger, für einzelne Kategorien oder Organsysteme zu gestalten.

Anhand des Scoring-Systems können zwar Punktwerte für die Komplexität vergeben werden, es lässt jedoch in seiner jetzigen Form noch viel Interpretationsspielraum. Wünschenswert für zukünftige Forschung wäre die Weiterentwicklung des Systems, die den einzelnen Werten mehr Aussagekraft verleiht. Ein möglicher Ansatz könnte sein, Kategorien bzw. Ranges für die Komplexität zu

schaffen. Genauso, wie beispielsweise die Glasgow-Coma-Scale (GCS) eine Beurteilung des Grades des Schädel-Hirn-Traumas (SHT) ermöglicht, könnte auch für das vorgestellte System eine Legende entwickelt werden. Im Falle der GCS kann durch den Punktwert eine Einordnung in ein leichtes SHT (GCS 15-13), ein mittelschweres SHT (GCS 12-9) oder ein schweres SHT (GCS 8-3) (91) vorgenommen werden. Denkbar wäre solch eine überschaubare Skalierung auch für die Fallkomplexität. Hierbei könnte die Komplexität in niedrig, mittel und hoch unterteilt werden. Diese Einteilung würde es ermöglichen, verschiedene Werte in Relation zueinander zu setzen und somit die Interpretation der Komplexität deutlich vereinfachen. Diese Klassifikation könnte gleichzeitig eine Basis für die Anwendung innerhalb der Lehre schaffen.

Eine Forschungsfrage, die zukünftig in den Fokus rücken sollte, ist, wie die Fallkomplexität in Bezug auf den Lernerfolg sinnvoll eingesetzt werden kann. Während die Anpassung der Aufgabenschwierigkeit an den Lernstand der Lernenden bei der Gestaltung von Lehre oftmals vorausgesetzt wird (38), wurde in der Vergangenheit in Bezug auf die Komplexität des Lernmaterials wenig geforscht. Innerhalb dieser Arbeit konnte gezeigt werden, dass unter gewissen Bedingungen Fallschwierigkeit durch Fallkomplexität vorhergesagt werden kann. Demzufolge scheint die gezielte Anpassung von Fallkomplexität auch zukünftig ein Parameter mit viel Potential auf dem Gebiet fallbasierter Lehr-Lernmethoden zu sein. Hierbei ist es vorstellbar, dass die bereits genannten Kategorien der Komplexität dazu verwendet werden können, standardisierte Empfehlungen für die Lehre zu erstellen. Um dieses Ziel zu verfolgen, ist weiterführende Forschung auf diesem Gebiet nötig.

Die Adaptivität des Lernmaterials stellt hierbei einen Aspekt mit viel Potential dar, um die Komplexität von Fällen an die Bedürfnisse der Lernenden anzupassen. Dies könnte bei Programmen wie z.B. CASUS in Zukunft automatisiert ablaufen. Wünschenswert wäre, dass das Programm während der Fallbearbeitung erkennt, wie komplex die nächsten Fälle sein müssen, damit ein Lernender davon profitieren kann. Es wählt je nach Niveau einen entsprechenden Fall aus und könnte im Idealfall sogar einen solchen generieren. Darüber hinaus könnte auch der Einsatz in Gebieten außerhalb der Ausbildungsforschung diskutiert werden. Nicht nur theoretische Lehr-Lernfälle, sondern auch Patientenvorstellungen im beruflichen Alltag verfügen über ein gewisses Maß an Komplexität (63). Daraus ergibt

sich die Frage, ob ein Scoring-System für die Komplexität auch im klinischen Alltag eingesetzt werden kann. Tatsache ist, dass manche Patienten einfacher zu diagnostizieren und zu behandeln sind als andere. Ein Grund dafür könnte die Komplexität des Patienten selbst sein. Für die Anwendung des vorgestellten Punktesystems spielt es keinerlei Rolle, ob ein Fall dazu verwendet wird, zu lernen, oder ob die Behandlung des Patienten vordergründiges Ziel ist. Die Verwendung von Komplexitätswerten in der klinischen Praxis könnte einige Vorteile bieten, wenn es um die Forschung in Bezug auf das Patientenmanagement geht. Es gibt Anhaltspunkte dafür, dass Patienten mit multiplen Symptomen und generell vielen verschiedenen Informationen schwierig zu diagnostizieren sind (92) und dies ein möglicher Grund für die hohe Zahl an Fehldiagnosen im klinischen Alltag ist (93). Mit einem Scoring-System könnte genau diese Komplexität erfasst und für die Optimierung von Diagnosestrategien und Behandlungsplänen genutzt werden.

Innerhalb dieser Arbeit wurde der Zusammenhang zwischen der Komplexität und der Schwierigkeit eines Falls untersucht. Sie soll zu dem übergeordneten Ziel, einer verbesserten Standardisierung von Fällen, um diese bedarfsgerecht einsetzen zu können, beitragen. Neben den bereits genannten Einsatzgebieten innerhalb der Lehre könnte dies für weitere Bereiche des universitären Betriebs, wie z.B. die Gestaltung von Prüfungen oder den Einsatz innerhalb der Forschung relevant sein. Studien könnten durch die Verwendung von Fällen mit vordefinierter und objektiv überprüfbarer Komplexität – etwa als Prä- und Posttest (94), profitieren.

Um das Potential fallbasierter Lehr-Lernmethoden in Zukunft ausschöpfen zu können, sollte der Fokus weiterführender Untersuchungen sowohl auf der inhaltlichen Gestaltung und Adaptivität des Lernmaterials als auch auf formalen Anforderungen liegen. Eine Herausforderung der Medizindidaktik wird es künftig sein, Fälle zeitgemäß und vor allem spannend darzustellen. Wünschenswert wäre deshalb, dass in Zukunft Onlineprogramme in ihrer Gestaltung optimiert und durch Komponenten wie z.B. digitale realitätsgetreue Simulationen ergänzt werden. Nur durch die Anpassung an neue Forschungsergebnisse und digitale Standards kann fallbasierte Lehre aktuell bleiben.

Literaturverzeichnis

1. Kolodner, J. L. (1992). An introduction to case-based reasoning. *Artificial Intelligence Review*, 6(1), 3-34. <https://doi.org/10.1007/BF00155578>
2. Braun, L. T., Lenzer, B., Fischer, M. R., Schmidmaier, R. (2019). Complexity of clinical cases in simulated learning environments: proposal for a scoring system. *GMS Journal for Medical Education*, 36(6), 6-10. <https://doi.org/10.3205/zma001288>
3. Major, R. H. (1930). The Papyrus Ebers. *Annals of Medical History*, 2(5), 547-555.
4. Jouanna, J. (2012). *Greek Medicine from Hippocrates to Galen: Selected Papers*, 40. BRILL. <https://doi.org/10.1163/9789004232549>
5. Stone, H. F. (1916). Dr. Redlich on the Case Method in American University Law Schools. *St. Louis Law Review*, 1(2), 111-121. https://openscholarship.wustl.edu/law_lawreview/vol1/iss2/2
6. Merseeth, K. K. (1991). The Early History of Case-Based Instruction: Insights for Teacher Education Today. *Journal of Teacher Education*, 42(4), 243-249. <https://doi.org/10.1177/002248719104200402>
7. Williams, S. M. (1992). Putting case-based instruction into context: Examples from legal and medical education. *Journal of the Learning Sciences* 2(4), 367-427. https://doi.org/10.1207/s15327809jls0204_2
8. Pietsch, C. D. (2010). *Die Reform des deutschen Medizinstudiums 1901* (Dissertation Medizin, Martin-Luther-Universität Halle-Wittenberg). <https://doi.org/10.25673/340>
9. Patterson, E. W. (2022). THE CASE METHOD IN AMERICAN LEGAL EDUCATION: ITS ORIGINS AND OBJECTIVES. *Journal of Legal Education* 4(1), 1-24. <https://www.jstor.org/stable/42890542>
10. Approbationsordnung für Ärzte vom 28. Oktober 1970. https://download.uni-mainz.de/verwaltung-sl/ordnungen/AppO_Medizin_bis_09_03.pdf
11. ÄApprO 2002 - Approbationsordnung für Ärzte. https://www.gesetze-im-internet.de/_appro_2002/BJNR240500002.html
12. Srinivasan, M., Wilkes, M., Stevenson, F., Nguyen, T., Slavin, S. (2007). Comparing Problem-Based Learning with Case-Based Learning: Effects of a Major Curricular Shift at Two Institutions. *Academic Medicine* 82(1), 74-82. <https://doi.org/10.1097/01.ACM.0000249963.93776.aa>
13. Zottmann, J., Stegmann, K., Strijbos, J-W., Vogel, F. (2013). Computer-supported collaborative learning with digital video cases in teacher education: The impact of teaching on knowledge convergence. *Computers in Human Behavior*, 29(5), 2100-2108. <https://doi.org/10.1016/j.chb.2013.04.014>
14. Schäfer, J. R. (2012). *HOUSE MEDIZIN*. Wiley-VCH
15. Fischer, M. (1999). CASUS: An authoring and learning tool that support diagnostic reasoning. *Zeitschrift für Hochschuldidaktik* 3(1), 87-98.
16. LOOP Share. <https://loop-share.charite.de/Zend/menu>

17. Lenzer, B. (2019). *Clinical Case Discussions – ein vielversprechender Ansatz zur Förderung des Clinical Reasoning in der ärztlichen Ausbildung?* (Dissertation Medizin, Ludwig-Maximilians-Universität München). <https://doi.org/10.5282/edoc.25394>
18. Klappenbach, R. (Hrsg.). (1976). Fall. In *WÖRTERBUCH DER DEUTSCHEN GEGENWARTSSPRACHE* (Band 2: DEUTSCH-GLAUBEN). De Gruyter. <https://doi-org.emedien.ub.uni-muenchen.de/10.1515/9783112578162>
19. Solomon, J. (2006). Case Studies: why are they important? *Nature Clinical Practice Cardiovascular Medicine*, 3(11), 579-579.
20. Uyeki, T. M., Sharma, A., Branda, J. A. (2009). Case 40-2009: A 29-Year-Old Man with Fever and Respiratory Failure. *The New England Journal of Medicine*, 361, 2558-2569. <https://doi.org/10.1056/NEJMcpc0905545>
21. Marks, P. W., Zukerberg, L. R. (2004). Case 30-2004: A 37-Year-Old Woman with Paresthesias of the Arms and Legs. *The New England Journal of Medicine*, 351(13), 1333-1341. <https://doi.org/10.1056/NEJMcpc040921>
22. Kotton, D. N., Muse, V. V., Nishino, M. (2012). Case 2-2012: A 63-Year-Old Woman with Dyspnea and Rapidly Progressive Respiratory Failure. *The New England Journal of Medicine*, 366(3), 259-269. <https://doi.org/10.1056/NEJMcpc1109274>
23. Bornemann, B. (2016). *Dokumentationsbögen der Inneren Medizin und der Chirurgie für Anamnese und körperliche Untersuchung für die studentische Lehre in Deutschland* (Dissertation Medizin, Ludwig-Maximilians-Universität München). <https://doi.org/10.5282/edoc.19166>
24. Gagnier, J. J., Riley, D., Altman, D. G., Moher, D., Sox, H., Kienle, G. S. (2013). The CARE guidelines: Consensus-based clinical case reporting guideline development. *Deutsches Ärzteblatt*, 110(37), 603-608. <https://doi.org/10.3238/arztebl.2013.0603>
25. Fischer, F., Opitz, A., (Ed.). (2022). *Learning to Diagnose with Simulations: Examples from Teacher Education and Medical Education*. Springer. <https://doi.org/10.1007/978-3-030-89147-3>
26. Ellaway, R., Candler, C., Greene, P. (2006, 11. September). An Architectural Model for MedBiquitous Virtual Patients. <http://groups.medbiq.org/medbiq/display/VPWG/MedBiquitous+Virtual+Patient+Architecture>
27. Huwendiek, S., De leng, B. A., Zary, N., Fischer, M. R., Ruiz, J. G., Ellaway, R. (2009). Towards a typology of virtual patients. *Medical Teacher*, 31(8), 743-748. <https://doi.org/10.1080/01421590903124708>
28. Thistlethwaite, J. E, Davies, D., Ekeocha, S., Kidd, J. M., MacDougall, C., Matthews, P., Purkis, J., Clay, D. (2012). The effectiveness of case-based learning in health professional education. A BEME systematic review: BEME Guide No. 23. *Medical Teacher*, 34(6), e421-e444. <https://doi.org/10.3109/0142159X.2012.680939>
29. Sutyak, J. P., Spotnitz, A. J., Mehne, P. R. (1996). Role of Case Structure and Prior Experience in a Case-based Surgical Clerkship. *The*

- American Journal of Surgery*, 172, 286-290.
[https://doi.org/10.1016/S0002-9610\(96\)00108-0](https://doi.org/10.1016/S0002-9610(96)00108-0)
30. Rybarczyk, B. J., Baines, A. T., McVey, M., Thompson, J. T., Wilkins, H. (2007). A Case-based Approach Increases Student Learning Outcomes and Comprehension of Cellular Respiration Concepts. *Biochemistry and Molecular Biology Education*, 35(3), 181-186.
<https://doi.org/10.1002/bmb.40>
 31. Mayo, J. A. (2004). Using case-based instruction to bridge the gap between theory and practice in psychology of adjustment. *Journal of Constructivist Psychology*, 17(2), 137-146.
<https://doi.org/10.1080/1072053049273917>
 32. Ye Hong Lai Wah Chan Lean Huat Lim, C. (1998). Case studies in community pharmacy: the contribution of family physicians to the teaching of community pharmacy. *Medical Teacher*, 20(5), 476-9.
<https://doi.org/10.1080/01421599880625>
 33. Herreid, C. F. (1997/1998). Some Basic Rules of Good Storytelling Help Teachers Generate Student Excitement in the Classroom. *Journal of College Science Teaching*, 27(3), 163-165. <http://www.ecsb.org/wp-content/uploads/2016/09/What-Makes-a-Good-Case.pdf>
 34. Kolodner, J. L., Owensby, J. N., Guzdial, M. (2004). CASE-BASED LEARNING AIDS. In Jonassen, D., Driscoll, M. (Hrsg.), *Handbook of Research on Educational Communications and Technology* (2. Aufl., S. 829-861). Routledge. <https://doi.org/10.4324/9781410609519>
 35. Kim, S., Phillips, W. R., Pinsky, L., Brock, D., Phillips, K., Keary, J. (2006). A conceptual framework for developing teaching cases: a review and synthesis of the literature across disciplines. *Medical Education*, 40(9), 867-876. <https://doi.org/10.1111/j.1365-2929.2006.02544.x>
 36. Johnson, J. F., Bagdasarov, Z., Connelly, S., Harkrider, L., Devenport, L. D., Mumford, M. D. (2012). Case-based ethics education: the impact of cause complexity and outcome favorability on ethicality. *Journal of Empirical Research on Human Research Ethics*, 7(3), 63-77.
<https://doi.org/10.1525/jer.2012.7.3.63>
 37. Krockenberger, M. B., Bosward, K. L., Canfield, P. J. (2007). Integrated Case-Based Applied Pathology (ICAP): A Diagnostic-Approach Model for the Learning and Teaching of Veterinary Pathology. *Journal of Veterinary Medical Education*, 34(4), 396-408. <https://doi.org/10.3138/jvme.34.4.396>
 38. Vygotsky, L. S. (1978). Interaction between learning and development. In M. Cole, V. John-Steiner, S. Scribner, E. Souberman (Eds.), *Mind and society: The development of higher psychological processes* (79-91). Cambridge, MA: Harvard University Press
 39. Chernikova, O., Heitzmann, N., Fink, M. C., Timothy, V., Seidel, T., Fischer, F., Cosima, DFG Research Group (2020). Facilitating Diagnostic Competences in Higher Education-a Meta-Analysis in Medical and Teacher Education. *Educational Psychology Review*, 32(1), 157-196.
<https://doi.org/10.1007/s10648-019-09492-2>
 40. Fischer, F., Bauer, E., Seidel, T., Schmidmaier, R., Radkowitsch, A., Neuhaus, B.J., Hofer, S.I., Sommerhoff, D., Ufer, S., Kuhn,

- J., Küchemann, S., Sailer, M., Koenen, J., Gartmeier, M., Berberat, P., Frenzel, A., Heitzmann, N., Holzberger, D., Pfeffer, J., Lewalter, D., Niklas, F., Schmidt-Hertha, B., Gollwitzer, M., Vorholzer, A., Chernikova, O., Schons, C., Pickal, A.J., Bannert, M., Michaeli, T., Stadler, M. and Fischer, M.R. (2022). Representational scaffolding in digital simulations – learning professional practices in higher education. *Information and Learning Sciences*, 123(11/12), 645-665. <https://doi.org/10.1108/ILS-06-2022-0076>
41. Braun, L. T., Borrmann, K. F., Lottspeich, C., Heinrich, D. A., Kiesewetter, J., Fischer, M. R., Schmidmaier, R. (2019). Scaffolding clinical reasoning of medical students with virtual patients: effects on diagnostic accuracy, efficiency, and errors. *Diagnosis*, 6(2), 137-149. <https://doi.org/10.1515/dx-2018-0090>
 42. Braun, L. T., Zottmann, J. M., Adolf, C., Lottspeich, C., Then, C., Wirth, S., Fischer, M. R., Schmidmaier, R. (2017). Representation scaffolds improve diagnostic efficiency in medical students. *Medical Education*, 51(11), 1118-1126. <https://doi.org/10.1111/medu.13355>
 43. Jones R. W. (2006). Problem-based Learning: Description, Advantages, Disadvantages, Scenarios and Facilitation. *Anaesthesia and Intensive Care*, 34(4), 485-488. <https://doi.org/10.1177/0310057X0603400417>
 44. Abdelkarim, A., Schween, D., Ford, T. (2018). Advantages and Disadvantages of Problem-Based Learning from the Professional Perspective of Medical and Dental Faculty. *EC Dental Science*, 17(7), 1073-1079. https://www.researchgate.net/publication/326519231_Advantages_and_disadvantages_of_problem-based_learning_from_the_professional_perspective_of_medical_and_dental_faculty
 45. Heitzmann, N., Fischer, F., Kühne-Eversmann, L., Fischer, M. R. (2015). Enhancing diagnostic competence with self-explanation prompts and adaptable feedback. *Medical Education*, 49(10), 993–1003. <https://doi.org/10.1111/medu.12778>
 46. Heitzmann, N., Fischer, F., Fischer, M. R. (2018). Worked examples with errors: when self-explanation prompts hinder learning of teachers diagnostic competences on problem-based learning. *Instructional Science*, 46(2), 245-271. <https://doi.org/10.1007/s11251-017-9432-2>
 47. Clyne, A. M., Billiar, K. L. (2016). Problem-Based Learning in Biomechanics: Advantages, Challenges, and Implementation Strategies. *Journal of Biomechanical Engineering*, 138(7), 1-9. <https://doi.org/10.1115/1.4033671>
 48. Koenemann, N., Lenzer, B., Zottmann, J. M., Fischer, M. R., Weidenbusch, M. (2020). Clinical Case Discussions – a novel, supervised peer-teaching format to promote clinical reasoning in medical students. *GMS Journal for Medical Education*, 37(5). <https://doi.org/10.3205/zma001341>
 49. Voskoglou, M. G., Subbotin, I. Y. (2015). Application of the Triangular Fuzzy Model to Assessment of Analogical Reasoning Skills. *American Journal of Applied Mathematics and Statistics*, 3(1), 1–6. <https://doi.org/10.12691/ajams-3-1-1>

50. Stadler, M., Niepel, C., Greiff, S. (2016) Easily too difficult: Estimating item difficulty in computer simulated microworlds. *Computers in Human Behavior*, 65, 100-106. <https://doi.org/10.1016/j.chb.2016.08.025>
51. Möltner, A., Schellberg, D., Jünger, J. (2006). Grundlegende quantitative Analysen medizinischer Prüfungen. *GMS Journal for Medical Education*, 23(3), 1-11. <http://www.egms.de/de/journals/zma/2006-23/zma000272.shtml>
52. Papa, F. J. (2014). Learning sciences principles that can inform the construction of new approaches to diagnostic training. *Diagnosis*, 1(1), 125-129. <https://doi.org/10.1515/dx-2013-0013>
53. Papa, F. J., Oglesby, M. W., Aldrich, D. G., Schaller, F., Cipher, D. J. (2007). Improving diagnostic capabilities of medical students via application of cognitive sciences-derived learning principles. *Medical Education*, 41(4), 419-425. <https://doi.org/10.1111/j.1365-2929.2006.02693.x>
54. Koskinen, H. (2007). Evaluation of the Level of Difficulty of Patient Cases for Veterinary Problem-Solving Examination: A Preliminary Comparison of Three Taxonomies of Learning. *Journal of Veterinary Medical Education*, 34(2), 106-111. <https://doi.org/10.3138/jvme.34.2.106>
55. Tokan, M. K., Imakulata, M. M. (2019). The effect of motivation and learning behaviour on student achievement. *South African Journal of Education*, 39(1). <https://doi.org/10.15700/saje.v39n1a1510>
56. Gercama, A. J., de Haan, M., van der Vleuten, C. P. M. (2000). Reliability of the Amsterdam Clinical Challenge Scale (ACCS): a new instrument to assess the level of difficulty of patient cases in medical education. *Medical Education*, 34(7), 519-524. <https://doi.org/10.1046/j.1365-2923.2000.00663.x>
57. Biggs, J. B., Collis, K. F. (1982). Evaluation the Quality of Learning-The SOLO Taxonomy (Structure of the Observed Learning Outcome). Elsevier. <https://doi.org/10.1016/C2013-0-10375-3>
58. Hege, I., Kononowicz, A. A., Kiesewetter, J., Foster-Johnson, L. (2018). Uncovering the relation between clinical reasoning and diagnostic accuracy – An analysis of learner’s clinical reasoning processes in virtual patients. *PLoS ONE*, 13(10), 1-12. <https://doi.org/10.1371/journal.pone.0204900>
59. Klappenbach R., Kempcke, G. (Hrsg.) (1981). komplex. In *WÖRTER-BUCH DER DEUTSCHEN GEGENWARTSSPRACHE* (Band 3: GLAUBEN-LYZEUM). De Gruyter. <https://doi-org.emedien.ub.uni-muenchen.de/10.1515/9783112541623>
60. Hennen, B. K. (1984). Measuring the complexity of clinical problems. *Journal of Medical Education*, 59(6), 487-493. <https://doi.org/10.1097/00001888-198406000-00005>
61. Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Performance*, 16(2), 366-387. [https://doi.org/10.1016/0030-5073\(76\)90022-2](https://doi.org/10.1016/0030-5073(76)90022-2)
62. van Merriënboer, J. J. G., Sweller, J. (2010). Cognitive load theory in health professional education: design principles and strategies. *Medical*

- Education*, 44(1), 85–93. <https://doi.org/10.1111/j.1365-2923.2009.03498.x>
63. Safford, M. M., Allison, J. J., Kiefe, C. I. (2007). Patient Complexity: More Than Comorbidity. The Vector Model of Complexity. *Journal of General Internal Medicine*, 22(S3), 382-290. <https://doi.org/10.1007/s11606-007-0307-0>
 64. Custers, E. J. F. M., Boshuizen, H. P. A., Schmidt, H. G. (1998). The Role of Illness Scripts in the Development of Medical Diagnostic Expertise: Results from an Interview Study. *Cognition and Instruction*, 16(4), 367-398. https://doi.org/10.1207/s1532690xci1604_1
 65. Young, J. Q., van Dijk, S. M., O'Sullivan, P. S., Custers, E. J., Irby, D. M., ten Cate, O. (2016). Influence of learner knowledge and case complexity on handover accuracy and cognitive load: results from a simulation study. *Medical Education*, 50(9), 969-978. <https://doi.org/10.1111/medu.13107>
 66. Custers, E. J. F. M., Boshuizen, H. P. A., Schmidt, H. G. (1996). The influence of medical expertise, case typicality, and illness script component on case processing and disease probability estimates. *Memory & Cognition*, 24(3), 384-399. <https://doi.org/10.3758/BF03213301>
 67. Braun, L. T., Lenzer, B., Kiesewetter, J., Fischer, M. R., Schmidmaier, R. (2018). How case representations of medical students change during case processing – Results of a qualitative study. *GMS Journal for Medical Education*, 35(3), Doc41. <https://doi.org/10.3205/zma001187>
 68. Funke, J. (1983). Einige Bemerkungen zu Problemen der Problemlöseforschung oder: Ist Testintelligenz doch ein Prädiktor?. *Diagnostica*, 29(4), 283–302. <https://doi.org/10.11588/heidok.00008131>
 69. Funke, J. (1992). *Wissen über dynamische Systeme: Erwerb, Repräsentation und Anwendung*. Springer. <http://link.springer.com/10.1007/978-3-642-77346-4>
 70. Kassirer, J. P. (2010). Teaching Clinical Reasoning: Case-Based and Coached. *Academic Medicine*, 85(7), 1118-1124. <https://doi.org/10.1097/acm.0b013e3181d5dd0d>
 71. Page, G., Bordage, G., Allen, T. (1995). Developing key-feature problems and examinations to assess clinical decision-making skills. *Academic Medicine*, 70(3), 194-201. <https://doi.org/10.1097/00001888-199503000-00009>
 72. Andresen, J. C. (2020). *Effektivität von Key-Feature-Prüfungen beim Erwerb der Kompetenz Clinical Reasoning in der medizinischen Ausbildung*. (Dissertation Medizin, Georg-August-University Göttingen). <https://ediss.uni-goettingen.de/handle/21.11130/00-1735-0000-0005-1452-8>
 73. Vogel, Tim. (2018). Ärztliche Prüfung: Das Stex im Wandel. *Deutsches Ärzteblatt International*, 14(2). <https://www.aerzteblatt.de/int/article.asp?id=200878>
 74. Braun, L. T., Zwaan, L., Kiesewetter, J., Fischer, M. R., Schmidmaier, R. (2017). Diagnostic errors by medical students: results of a prospective

- qualitative study. *BMC Medical Education*, 17(1), 191.
<https://doi.org/10.1186/s12909-017-1044-7>
75. Fink, M. C., Heitzmann, N., Siebeck, M., Fischer, F., Fischer, M. R. (2021). Learning to diagnose accurately through virtual patients: do reflection phases have an added benefit? *BMC Medical Education*, 21(1), 523. <https://doi.org/10.1186/s12909-021-02937-9>
 76. Fink, M. C., Reitmeier, V., Stadler, M., Siebeck, M., Fischer, F., Fischer, M. R. (2021). Assessment of Diagnostic Competences With Standardized Patients Versus Virtual Patients: Experimental Study in the Context of History Taking. *Journal of Medical Internet Research*, 23(3), e21196. <https://doi.org/10.2196/21196>
 77. Radkowsch, A., Sailer, M., Schmidmaier, R., Fischer, M. R., Fischer, F. (2021). Learning to diagnose collaboratively – Effects of adaptive collaboration scripts in agent-based medical simulations. *Learning and Instruction*, 75, 1-11. <https://doi.org/10.1016/j.learninstruc.2021.101487>
 78. Heitzmann, N., Fischer, F., Kühne-Eversmann, L., Fischer, M. R. (2015). Enhancing diagnostic competence with self-explanation prompts and adaptable feedback. *Medical Education*, 49(10), 993–1003. <https://doi.org/10.1111/medu.12778>
 79. Koller, I., Alexandrowicz, R., Hatzinger, R. (2012). *Das Rasch-Modell in der Praxis: eine Einführung in eRm*. Facultas. <https://doi.org/10.36198/9783838537863>
 80. Strobl, C., Matiaske, W., Fantapié Altobelli, C. (2015) *Das Rasch-Modell: eine verständliche Einführung für Studium und Praxis*. (3. erweiterte Auflage). Reiner Hampp.
 81. Prenzel, M., Sällzer, C., Klieme, E., Köller O. (Hrsg.) (2013). *PISA 2012. Fortschritte und Herausforderungen in Deutschland*. Waxmann. https://www.pedocs.de/volltexte/2020/18838/pdf/Prenzel_et_al_2013_PISA_2012_Fortschritte_und_Herausforderungen.pdf
 82. Baghaei, P., Kubinger, K. D. (2015). Linear Logistic Test Modeling with R. *Practical Assessment, Research & Evaluation*, 20(1). <https://doi.org/10.7275/8f33-hz58>
 83. R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
 84. Lazarovici, M. *Anwenderperformanz und- variabilität der Glasgow-Koma-Skala - Prospektive Anwenderstudie zur Prozeßqualität der GCS-Anwendung*. (Dissertation Medizin, Ludwig-Maximilians-Universität München). <https://doi.org/10.1007/s10049-002-0446-0>
 85. Moosbrugger, H., Kelava, A. (2008). Qualitätsanforderungen an einen psychologischen Test (Testgütekriterien). In Moosbrugger, H., Kelava, A. (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (1. Aufl., S. 7-26). Springer. https://doi.org/10.1007/978-3-540-71635-8_2
 86. Neuschwander, M., In-Albon, T., Adornetto, C., Roth, B., Schneider, S. (2013). Interrater-Reliabilität des Diagnostischen Interviews bei psychischen Störungen im Kindes- und Jugendalter (Kinder-DIPS). *Zeitschrift*

- für Kinder- und Jugendpsychiatrie und Psychotherapie*, 41(5), 319-334.
<https://doi.org/10.1024/1422-4917//a000247>
87. Bülte, D., Schnitker, R. (2017). Interrater-Reliabilität der Aachener Materialien zur Diagnostik Neurogener Sprechstörungen.
<https://doi.org/10.13140/RG.2.2.26244.01926>
 88. Fischer, M. R., Bauer, D., Mohn, K., NKLM-Projektgruppe. (2015). Endlich fertig! Nationale Kompetenzbasierte Lernzielkataloge Medizin (NKLM) und Zahnmedizin (NKLZ) gehen in die Erprobung. *GMA Zeitschrift für medizinische Ausbildung*, 32 (3).
<https://doi.org/10.3205/zma000977>
 89. Jenßen, L., Dunekacke, S., Blömeke, S. (2018). Qualitätssicherung in der Kompetenzforschung. Empfehlungen für den Nachweis von Validität in Testentwicklung und Veröffentlichungspraxis. *Zeitschrift für Pädagogik*, 61, 11-31. <https://doi.org/10.25656/01:15501>
 90. Kane, M. T. (2013). Validating the Interpretations and Uses of Test Scores: Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement*, 50(1), 1-73.
<https://doi.org/10.1111/jedm.12000>
 91. Teasdale, G., Maas, A., Lecky, F., Manley, G., Stocchetti, N., Murray, G. (2014). The Glasgow Coma Scale at 40 years: standing the test of time. *The Lancet Neurology*, 13(8), 844-854. [https://doi.org/10.1016/S1474-4422\(14\)70120-6](https://doi.org/10.1016/S1474-4422(14)70120-6)
 92. Kostopoulou, O., Oudhoff, J., Nath, R., Delaney, B. C., Munro, C. W., Harries, C, Holder, R. (2008). Predictors of Diagnostic Accuracy and Safe Management in Difficult Diagnostic Problems in Family Medicine. *Medical Decision Making*, 28(5), 668-680.
<https://doi.org/10.1177/0272989X08319958>
 93. Graber, M. L. (2013). The incidence of diagnostic error in medicine. *BMJ Quality & Safety*, 22, ii21-ii27. <https://doi.org/10.1136/bmjqs-2012-001615>
 94. Zottmann, J. M., Stegmann, K., Strijbos, J. W., Vogel, F., Wecker, C., Fischer, F. (2013). Computer-supported collaborative learning with digital video cases in teacher education: The impact of teaching experience on knowledge convergence. *Computers in Human Behaviour*, 29(5), 2100-2108.

Abbildungs- und Tabellenverzeichnis

Abbildung 1:	Prozesse fallbasierten Denkens aus Kolodner 1992 (1)	11
Abbildung 2:	Verschiedene Komponenten mit möglichem Einfluss auf die Komplexität eines Patienten aus Safford et al. 2007 (63)	23
Abbildung 3:	Illness Script aus Custers et al. 1998 (64)	25
Abbildung 4:	Komponenten der CLT	26
Abbildung 5:	Komplexität innerhalb der CLT	28
Abbildung 6:	Faktoren mit möglichem Einfluss auf die Fallschwierigkeit	32
Abbildung 7:	Fall (Myokarditis) in der Lernumgebung CASUS (41, 74), vollständiger Fall in Anhang #2	36
Abbildung 8:	Hierarchische Gliederung des neuen Scoring-Systems CASE	38
Abbildung 9:	Schema zur Identifikation relevanter familiärer Risikofaktoren	40
Abbildung 10:	KF mit Aufgabe (17)	48
Abbildung 11:	Gesamtwerte beider Rater für die Komplexität (Für eine Fallübersicht vgl. Tabelle 11)	56
Abbildung 12:	Gesamtscores beider Rater nach dem neuen System CASE (Rater 1: Elisabeth Hilger, Rater 2: Kim Öhler)	60
Abbildung 13:	Häufigste Diagnosen von 160 Fällen aus dem Validation Datensatz	62
Abbildung 14:	Kurzfall (17) mit der niedrigsten Komplexität, Informationen werden lediglich innerhalb der ersten Dimension (Anamnese) genannt	63
Abbildung 15:	Anzahl Fälle je Komplexitätsspanne	63
Abbildung 16:	Kurzfall (17) mit Informationen in beiden Dimensionen und verschiedenen Kategorien	64
Abbildung 17:	Fall mittlerer Länge (41) mit dem Leitsymptom Dyspnoe und einer Komplexität von 53	66

Abbildung 18:	Korrelation der Schwierigkeiten aus RM und LLTM für Datensatz 1; Datenerhebung in Originalstudien (42, 74)	70
Abbildung 19:	Korrelation der Schwierigkeiten aus RM und LLTM für Datensatz 2; Datenerhebung in Originalstudie (41)	70
Abbildung 20:	Korrelation der Schwierigkeiten aus RM und LLTM für Datensatz 3; Datenerhebung in Originalstudie (17)	71

Tabelle 1:	Merkmale eines VP, gekürzt aus Huwendiek et al. 2009 (27)	8
Tabelle 2:	Struktur eines klinischen Falls nach dem ACCS nach Gercama et al. 2000 (56)	19
Tabelle 3:	Kriterien für die Beurteilung der Schwierigkeit von Aufgaben auf Basis der SOLO-Taxonomie, gekürzt aus Koskinen 2007 (54)	20
Tabelle 4:	Punktesystem für die Fallkomplexität aus Hennen 1984 (60)	29
Tabelle 5:	Scoring-System zur Bestimmung der Fallkomplexität aus Braun et al. 2019 (2)	30
Tabelle 6:	Übersicht über das neue Scoring-System CASE	41
Tabelle 7:	Leitsymptome der 59 Fälle zur Untersuchung der Interrater-Reliabilität	43
Tabelle 8:	Übersicht über Inhalt und Zweck der verwendeten Falldatensätze	45
Tabelle 9:	Übersicht über den Anteil aller Fälle pro verwendetem Fallformat	46
Tabelle 10:	Beispielhafte Datenmatrix	50
Tabelle 11:	Musterdiagnosen der 12 Fälle und die durch beide Rater vergebenen Punktzahlen für die Gesamtkomplexität	55

Tabelle 12:	Cohen's Kappa der Gesamtwerte und der fünf Kategorien bei Verwendung des Systems nach Braun et al. (2). (1: Anamnese, 2: Körperl. US, 3: Techn. US, 4: Psycho-sozial, 5: Nebendiagnosen)	57
Tabelle 13:	Cohen's Kappa der Gesamtwerte und der fünf Untergruppen bei der Verwendung des neuen Scoring-Systems (1: Anamnese, 2: Bildgebung, 3: Labor, 4: Körperl. US, 5: Funktionell)	60
Tabelle 14:	Cohen's Kappa der Gesamtwerte, berechnet für 44 Fälle mit einer Komplexität <60 und 15 Fälle mit einer Komplexität >60	61
Tabelle 15:	Exemplarisches Scoring: Fall aus Abbildung 16	64
Tabelle 16:	Vorschlag einer Zuordnung von Symptomen zu Organsystemen	68
Tabelle 17:	Gewichtungen der Features für Datensatz 1	72
Tabelle 18:	Gewichtungen der Features für Datensatz 2	72
Tabelle 19:	Gewichtungen der Features für Datensatz 3	73

Abkürzungen

ÄApprO	Approbationsordnung für Ärzte
ACCS	Amsterdam Clinical Challenge Scale
CARE	CAse REports
CASE	ComplexityAssessmentStructurE
CCD	Clinical Case Discussions
CLT	Cognitive Load Theory
DAM	Institut für Didaktik und Ausbildungsforschung in der Medizin
GCS	Glasgow-Coma-Scale
IMPP	Institut für medizinische und pharmazeutische Prüfungsfragen
KF	Key-Feature
LLTM	Linear-logistisches Testmodell
LMU	Ludwig-Maximilians-Universität München
M2	Zweiter Abschnitt der Ärztlichen Prüfung
NEJM	New England Journal of Medicine
NKLM	Nationaler Kompetenzbasierter Lernzielkatalog Medizin
PISA	Programme for International Student Assessment
RM	Rasch-Modell
SHT	Schädel-Hirn-Trauma
SOLO	Structure of the Observed Learning Outcome
VP	Virtueller Patient

#1 Detaillierte Fallpräsentation, erschienen im NEJM (21)

The NEW ENGLAND JOURNAL of MEDICINE

CASE RECORDS of the MASSACHUSETTS GENERAL HOSPITAL

Founded by Richard C. Cabot
Nancy Lee Harris, M.D., Editor

Jo-Anne O. Shepard, M.D., Associate Editor
Sally H. Ebeling, Assistant Editor

Stacey M. Ellender, Assistant Editor
Christine C. Peters, Assistant Editor



Case 30-2004: A 37-Year-Old Woman with Paresthesias of the Arms and Legs

Peter W. Marks, M.D., Ph.D., and Lawrence R. Zukerberg, M.D.

PRESENTATION OF CASE

A 37-year-old woman was seen in the clinic because of numbness in the arms and legs.

Four months earlier, she had noticed tingling and numbness in the fingertips, without weakness, three weeks after having last ridden a bicycle. She consulted her primary care physician. The temperature was 37.2°C and the blood pressure 115/85 mm Hg. The height was 1.75 m and the weight 69.1 kg. The lungs were clear to auscultation, and the heart sounds were normal. No rash or lymphadenopathy was found. Examination of the breasts, abdomen, and arms and legs and a pelvic examination revealed no abnormalities. The range of motion of the neck and the hand-grip strength were normal; the sensation in the hands also seemed to be intact. No treatment was given, and the symptoms resolved after three weeks.

Two and a half months later she noticed fresh blood on the toilet tissue after passing a stool. Rectal examination revealed a hemorrhoid. Shortly thereafter she had a brief illness, characterized by nausea, possible fever, and leg fatigue on exertion. During the next several weeks, she noticed the onset of numbness and tingling that extended from the fingertips to the upper arms; subsequently, numbness and tingling developed in the legs as well. She also noticed weakness in her legs, to the extent that she could no longer run a mile and was losing her balance. She recalled seeing apparent insect bites on her lower legs after riding a bicycle in tall grass in the spring, in the suburban areas west of Boston. She had not recently traveled outside the Boston area. A review of her diet revealed that she ate meat and dairy products as well as vegetables and grains. She was referred to a neurologist.

Hypothyroidism had been discovered when the patient was 15 years old, and more recently depression had developed. A physical examination four months before the first episode of numbness and tingling had revealed no abnormalities. At that time, the levels of glucose, urea nitrogen, creatinine, calcium, total bilirubin, total protein, albumin, globulin, electrolytes, aspartate aminotransferase, alanine aminotransferase, and alkaline phosphatase were normal. Tests for IgG antibodies to herpes simplex virus types 1 and 2 were positive; tests for IgM antibodies and antibodies to hepatitis A virus were negative. Other laboratory-test results obtained at the time of that examination are shown in Tables 1 and 2. She took levothyroxine (200 µg) and sertraline (100 mg) daily. She was not married, and she worked in an office. She did not smoke or drink alcohol.

From the Department of Medicine, Division of Hematology, Brigham and Women's Hospital (P.W.M.); the Departments of Medicine (P.W.M.) and Pathology (L.R.Z.), Harvard Medical School; and the Department of Pathology, Massachusetts General Hospital (L.R.Z.) — all in Boston.

N Engl J Med 2004;351:1333-41.

Copyright © 2004 Massachusetts Medical Society.

Table 1. Hematologic Laboratory Values.

Variable	4 Mo before Onset of Numbness	4 Mo after Onset of Numbness
Hemoglobin (%)	35.7	27.6
Hemoglobin (g/dl)	12.6	10.2
Red cells (per mm ³)	3.61×10 ⁶	2.42×10 ⁶
Mean corpuscular volume (μm ³)	99	114
Mean corpuscular hemoglobin (pg/red cell)	34.9	42.2
Mean corpuscular hemoglobin concentration (g/dl)	35.2	37.1
White cells (per mm ³)	5,900	3,200
Differential count (%)		
Neutrophils		37
Lymphocytes		62
Monocytes		1
White-cell morphology		Occasional hypersegmented neutrophils
Red-cell morphology		
Anisocytosis		1+
Macrocytosis		3+
Schistocytosis		1+
Teardrop forms		1+
Platelets (per mm ³)	342,000	288,000

Table 2. Blood Chemical Values.

Variable	4 Mo before Onset of Numbness	4 Mo after Onset of Numbness
Thyrotropin (μU/ml)	10.28	1.82
Thyroxine (pg/dl)*	8.6	
Thyroid hormone-binding index	1.23	
Free thyroxine index	10.6	
Aspartate aminotransferase (U/liter)		38
Alanine aminotransferase (U/liter)		33

* To convert the value for thyroxine to nanomoles per liter, multiply by 12.87.

her cranial-nerve and motor functions. There was a slight delay in the relaxation of the deep-tendon reflexes; the plantar responses were flexor. Vibratory sensation was reduced in the feet, as was proprioception; other sensory responses were intact. The result of Romberg's test was normal. Cortical sensation (including graphesthesia, stereognosis, and tactile extinction) was normal.

Radiographs of the lumbar spine revealed no abnormalities. Cranial and cervical magnetic resonance imaging (MRI) and magnetic resonance angiographic studies, performed after the administration of gadolinium, disclosed no intracranial abnormalities. On T₂-weighted images, there was apparent hyperintensity of the cervical spinal cord at the C2-to-C3 level, but at that level, the cord was incompletely evaluated. A subsequent, dedicated MRI study of the cervical spine revealed no abnormalities. Laboratory values obtained at the time of this visit are shown in Tables 1 and 2.

A diagnostic procedure was performed.

DIFFERENTIAL DIAGNOSIS

Dr. Peter W. Marks: This young woman, who had a history of hypothyroidism, had a gradual onset of paresthesias of the hands and arms, followed by paresthesias of the legs, and weakness that affected her ability to exercise. Although I am aware of the diagnosis in this case, it provides an opportunity to review the differential diagnosis of rather non-specific initial symptoms.

PARESTHESIAS

The term "paresthesia" is used to describe an abnormal burning or pricking sensation. It is often described by patients as numbness or tingling in a part of the body. Inadvertent compression of a nerve by pressure on an arm or leg may cause the transient paresthesia with which most of us are familiar. However, many diseases may be characterized by paresthesia as one of their initial features (Table 3).

Entrapment neuropathies, such as carpal tunnel syndrome, are relatively common. At the time of the patient's initial presentation, when she had paresthesias in the fingers after cycling, carpal tunnel syndrome may have been given serious consideration.¹ However, the subsequent development of paresthesias in both the upper and lower limbs indicates the presence of a systemic disorder. Patients with central nervous system disorders may present

Table 3. Possible Causes of Paresthesias.

Central nervous system disorders
Cerebrovascular accidents
Multiple sclerosis
Multiple myeloma
Peripheral nervous system disorders
Entrapment neuropathies (e.g., carpal tunnel syndrome)
Herededitary conditions affecting peripheral nerves (e.g., Charcot-Marie-Tooth disease)
Polyneuropathies
Trauma
Metabolic disorders
Diabetes mellitus
Thyroid dysfunction
Porphyrria
Uremia
Vitamin B ₁₂ deficiency
Toxins
Alcohol abuse
Heavy metal poisoning
Side effects of medications (e.g., vincristine and medications for human immunodeficiency virus infection)
Infections
Encephalitis
Lyme disease
Rheumatologic disorders
Systemic lupus erythematosus or disease
Polyarteritis nodosa
Systemic lupus erythematosus
Rheumatoid arthritis
Vasculitis
Cancer and related disorders
Compression of neurologic structures by tumor
Infiltration of nerves by tumor
Paraneoplastic syndromes

with paresthesias, and in a young woman, multiple sclerosis is a concern. However, this patient's MRI examination revealed no abnormalities. No aspect of her history implicated agents such as alcohol, drugs, or other toxins. Many rheumatologic disorders can cause paresthesias, either through central nervous system effects, as in systemic lupus erythematosus, or through vasculitis that affects the peripheral nerves. Cancers may result in nerve compression, may invade or infiltrate the nerves, or may be associated with paraneoplastic phenomena. However, this patient does not have other signs or symptoms suggesting that she has a rheumatologic disease or cancer.

Infectious causes of paresthesia include Lyme disease, which is endemic in several areas of the United States. The presumed insect bites on the pa-

tient's legs raise the possibility of a vector-borne disease. Metabolic disorders such as diabetes mellitus, hypothyroidism, porphyria, uremia, and vitamin B₁₂ deficiency can cause paresthesias through damage to central or peripheral nervous system structures. The patient had a history of hypothyroidism, but her thyrotropin level was normal at the time of the most recent measurement, although previously it had been abnormally elevated. Vitamin B₁₂ (cobalamin) deficiency may cause symmetric paresthesias and disorders of proprioception and vibratory sensation affecting the upper and lower limbs, even in the absence of overt anemia; this patient may have been at increased risk for this disorder, given her history of hypothyroidism. Though categorized as metabolic causes of paresthesias, both hypothyroidism and vitamin B₁₂ deficiency may be caused by autoimmune processes. Thus, in broad terms, the most likely causes of paresthesia in this patient are infectious or metabolic.

MACROCYTIC ANEMIA

At the time the patient visited the neurologist because of her symptoms, a complete blood count revealed macrocytic anemia. Macrocytic anemias have a number of causes and may be categorized as megaloblastic or nonmegaloblastic (Table 4). Megaloblastic anemias, which result from defects in DNA synthesis and cell maturation, include anemias due to folate or vitamin B₁₂ deficiency, chemotherapy-related agents, or myelodysplasia. Although the patient subsequently had an episode of rectal bleeding from a hemorrhoid, she did not appear to be losing blood rapidly, so a brisk reticulocyte response is unlikely to be the cause of the macrocytosis. Although myelodysplasia would be unusual in a 37-year-old woman, it may occur at this age. The patient was not taking any medication that could cause a macrocytic anemia, such as zidovudine or a folate antagonist such as methotrexate. She did not use alcohol and did not appear to have overt signs of liver disease. Although hypothyroidism may also cause macrocytic anemia, the increase in mean corpuscular volume in this condition is generally limited. Folate or vitamin B₁₂ deficiency, on the other hand, may be associated with a substantial increase in the size of red cells. The mean corpuscular volume in this patient — 114 μm³ — would be consistent with such a deficiency.

In this case, the correct diagnosis is suggested by the interconnection of causes of paresthesia and causes of marked macrocytosis: vitamin B₁₂ (co-

Table 4. Possible Causes of Macrocytic Anemia.

Reticuloocyte responses
Normal response to blood loss
Response to hemolysis
Bone marrow failure
Aplastic anemia
Myelodysplasia
Liver disease
Toxins
Alcohol
Chemotherapeutic agents
Thyroid disease
Vitamin deficiencies
Folate deficiency
Vitamin B ₁₂ deficiency

balamin) deficiency. Since this patient reported eating a diet containing some meat and dairy products, malabsorption of vitamin B₁₂ seems more likely than poor intake. In the absence of a history of gastrointestinal surgery or achylia, the most likely cause of this patient's inability to absorb vitamin B₁₂ is pernicious anemia. This disorder would explain the patient's progressive paresthesias as well as her other neurologic findings. Although the prevalence of pernicious anemia increases with advancing age, it may occur in young persons, particularly if it is a manifestation of a heritable syndrome.

AUTOIMMUNE POLYENDOCRINE SYNDROME

Pernicious anemia is an autoimmune disorder.² In this patient, in whom hypothyroidism and probable pernicious anemia had developed at an early age and who had a family history of anemia, we must consider one of the syndromes of polyglandular failure. At least three major constellations of findings have been identified; they are termed autoimmune polyendocrine syndrome types I, II, and III (Table 5).³ In type III disease, autoimmune thyroiditis is associated with another organ-specific autoimmune disease. Three subcategories of type III disease have been described, according to the disorder associated with the autoimmune thyroiditis. One of these, type IIB, which consists of autoimmune thyroiditis and pernicious anemia, is consistent with this patient's characteristics on presentation.

PERNICIOUS ANEMIA

Pathophysiologic Features
Dietary vitamin B₁₂ binds to intrinsic factor, which is produced by cells in the gastric fundus and body,

after it dissociates from protein carriers in the acidic environment of the stomach. It then travels through the gut to the terminal ileum, where it is absorbed through a receptor-mediated process.⁴ Deficiency can be caused by a variety of circumstances: dietary insufficiency (which is rare), food cobalamin deficiency (i.e., an inability to split cobalamin from food), achylia (which may be caused by aging, drugs that inhibit gastric acid secretion, or surgery), loss of ileal receptors (which may be caused by resection of the terminal ileum or inflammatory bowel disease), biologic competition (from *Diphyllobothrium latum* [the fish tapeworm] or bacterial overgrowth syndrome), rare congenital syndromes (such as transcobalamin II deficiency), and lack of production of intrinsic factor in the stomach because of pernicious anemia, a process that reflects autoimmune destruction of the gastric parietal and zymogenic cells.

Vitamin B₁₂ is required for two key sets of reactions (Fig. 1). In nucleic acid metabolism, it is involved in the folate cycle, and it serves as a methyl group acceptor in the conversion of methyltetrahydrofolate to tetrahydrofolate, which then goes on to participate in purine synthesis. In this process, homocysteine is converted to methionine. The other reaction is the conversion of propionyl-coenzyme A (CoA) to succinyl-CoA through the intermediate methylmalonyl-CoA. Although it has been hypothesized that vitamin B₁₂ deficiency affects the central nervous system through this mechanism, findings in patients with enzymatic deficiencies in this pathway do not support this idea. Thus, the cause of the neurologic manifestations of vitamin B₁₂ deficiency remains obscure.⁵

Whereas the mechanism, vitamin B₁₂ deficiency can be associated with permanent neurologic damage, so appropriate diagnosis and management are critical. Folate supplementation can partially overcome the anemia and the effect of vitamin B₁₂ deficiency on the peripheral nerves, but it does not affect the manifestations in the central nervous system.⁶

Diagnostic Testing

Diagnosing vitamin B₁₂ deficiency is straightforward when the deficiency is profound and there is a hypoproliferative anemia characterized by marked macrocytosis, hypersegmentation of neutrophils, pancytopenia, and signs of ineffective erythropoiesis (such as elevated levels of lactate dehydrogenase and indirect bilirubin) on laboratory testing. Diag-

nosis a subtle deficiency, which appears to be increasingly common, is more challenging. In such cases, vitamin B₁₂ levels are often in the range of 200 to 300 pg per milliliter (150 to 220 pmol per liter) — low but often not below the lower limit of normal. Liver disease and myeloproliferative diseases can also lead to falsely low-normal or normal levels of vitamin B₁₂ in patients with pernicious anemia.⁷ In such situations, measurement of homocysteine and methylmalonic acid levels may be useful for documenting true vitamin B₁₂ deficiency and for distinguishing it from folate deficiency (Fig. 2). In vitamin B₁₂ deficiency, the levels of both homocysteine and methylmalonic acid are elevated, whereas in most cases of folate deficiency only the homocysteine level is increased. In the absence of renal insufficiency, which can also elevate the homocysteine level, a methylmalonic acid level above the normal range is highly suggestive of vitamin B₁₂ deficiency. The combination of a significant elevation in both methylmalonic acid and homocysteine virtually confirms the diagnosis.

Although the Schilling test, which involves the sequential administration of radiolabeled vitamin B₁₂ and intrinsic factor, was once commonly used in the evaluation of vitamin B₁₂ deficiency, it has fallen out of favor because of its complexity and the need to administer radioactive materials. A test for antibodies against parietal cells may be helpful in some cases, as may a test for antibodies against intrinsic factor, but the former is relatively nonspecific and the latter relatively insensitive (though very specific).

In the patient under discussion, the initial diagnostic procedures should have included a review of the peripheral-blood smear as well as an evaluation of the vitamin B₁₂ levels. The necessity of additional diagnostic procedures would then depend on the findings.

Dr. Nancy Lee Harris (Pathology): Dr. Zarghamne-Garzanti, you followed this patient during and after her initial evaluation. Would you summarize your thinking before the diagnostic testing?

Dr. Mengni Zarghamne-Garzanti (Medicine): My first concern in this young woman with numbness of the arms and legs was multiple sclerosis, and I ordered the MRI study to evaluate this possibility. Vitamin B₁₂ deficiency was my next consideration. Since her thyroid disease was well controlled, I did not think it was the cause of her problems.

Table 5. Autoimmune Polyendocrine Syndromes.

Major Features	Potential Associations
Type I (at least two of the following three features) Chronic mucocutaneous candidiasis or Hypoparathyroidism or Hypoadrenalism	Diabetes Hypogonadism Pernicious anemia, vitiligo
Type II Adrenal insufficiency plus either Autoimmune thyroid disease or Insulin-dependent diabetes mellitus	Celiac disease Hypogonadism Pernicious anemia
Type III Autoimmune thyroiditis plus Type I (diabetes mellitus [type IIA]) or Pernicious anemia (type IIB) or Vitiligo, alopecia, or both (type IIC)	Celiac disease Hypogonadism Mastitis Sarcoidosis

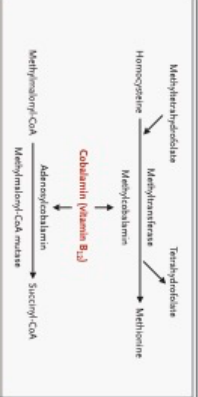


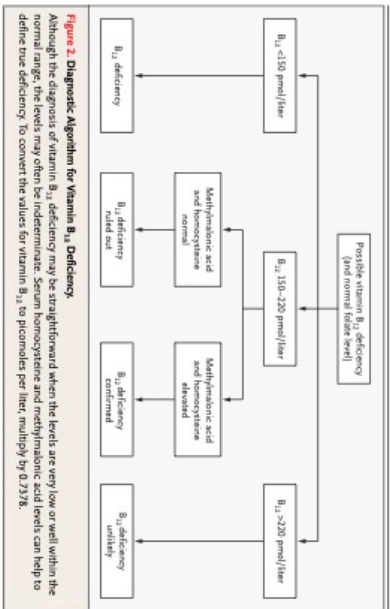
Figure 1. Metabolic reactions involving Vitamin B₁₂.
Vitamin B₁₂ is involved in both nucleic acid metabolism and lipid metabolism. Its function in nucleic acid metabolism is intimately intertwined with that of folate. Thus, folate supplementation can partially reverse the peripheral effects of vitamin B₁₂ deficiency. CoA denotes coenzyme A.

CLINICAL DIAGNOSIS

Pernicious anemia.

DR. PETER W. MARKS'S DIAGNOSIS

Pernicious anemia, possibly associated with the autoimmune polyendocrine syndrome, type IIB.



PATHOLOGICAL DISCUSSION

Dr. Lawrence R. Zolothoff: The peripheral-blood smear showed that approximately 10 percent of the mature neutrophils were hypersegmented, with six or more lobes (Fig. 3). A few hypersegmented neutrophils may be seen in a variety of disorders, but when more than 5 percent of the neutrophils are hypersegmented, the probability of either vitamin B₁₂ deficiency or folate deficiency is very high. Hypersegmented eosinophils with three lobes were also present. In addition, there were numerous oval macrocytes and small and fragmented erythrocytes. The high mean corpuscular volume, in conjunction with the findings on the peripheral blood smear, are evidence of megaloblastic anemia. Laboratory studies showed that the level of vitamin B₁₂ was very low, at 68 pg per milliliter (50 pmol per liter) (normal, greater than 250 pg per milliliter [184 pmol per liter]). The folate level was slightly above normal. Therefore, the diagnosis was megaloblastic anemia due to vitamin B₁₂ deficiency.

To evaluate the cause of the vitamin B₁₂ deficiency, an upper gastrointestinal endoscopy was performed, and biopsy specimens were obtained from the duodenum, antrum, and fundus. The duodenum and antrum were normal, with no evidence of chronic gastritis. Low-power microscopic examination of the gastric body and fundus showed extensive atrophy and marked thinning of the mucosa (Fig. 4A). Closer examination showed severe atrophy, with complete absence of fundic glands and of chief and parietal cells. The lamina propria contained a dense lymphocytic infiltrate, and the mucosa showed intestinal metaplasia with goblet cells (Fig. 4B).

The histologic features of severe chronic atrophic gastritis limited to the gastric body and fundus are those of autoimmune gastritis. A test for antibodies to intrinsic factor was positive. Therefore, the diagnosis was pernicious anemia due to autoimmune gastritis, with vitamin B₁₂ deficiency.

The pathophysiology of autoimmune gastritis is centered on the hydrogen-potassium ATPase gastric proton pump.² This pump is present only in the parietal cells of the gastric body. The initial event appears to be a CD4⁺ T-cell reaction against the proton pump. The T cells cause parietal-cell injury and result in exposure of both intrinsic factor and the hydrogen-potassium ATPase to antigen-presenting cells, activation of the immune response with secretion of interferon and cytokines, and formation of antibodies against intrinsic factor and the gastric proton pump.

Although the immune response is directed only against the components of the parietal cell, over

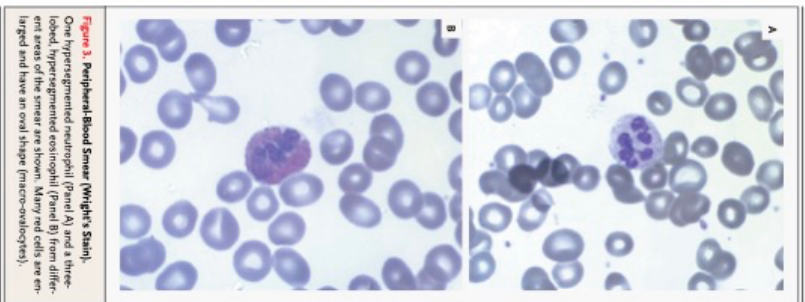


Figure 3. Peripheral-Blood Smear (Wright's Stain).
One hypersegmented neutrophil (Panel A) and a three-lobed hypersegmented eosinophil (Panel B) from different areas of the smear are shown. Many red cells are enlarged and have an oval shape (macrocytosis).

and carcinoma. In addition, the lack of acid leads to increased gastrin production and endocrine-cell hyperplasia and, in some patients, the development of multiple carcinoid tumors.

There are additional considerations. First, it has been reported that in some patients both the gastritis and anemia can be reversed by treatment with corticosteroids or azathioprine.^{8,9} In some patients treated with immunosuppression, the chronic gastritis disappeared, the gastric mucosa returned to normal, and acid output increased. Thus, immunosuppression may be one way to decrease the risk of dysplasia and carcinoma in these patients, although data showing a reduction in the risk of cancer with such therapy are lacking. *Helicobacter pylori* gastritis may lead to multifocal gastric atrophy, which in some patients leads to hypochlorhydria and the development of pernicious anemia. In a study of 138 patients with vitamin B₁₂ deficiency, *H. pylori* was detected in 56 percent of the patients, and eradication of *H. pylori* improved the anemia and vitamin B₁₂ levels in 40 percent of this subgroup.¹⁰ Thus, *H. pylori* appears to be a causative agent in some cases of vitamin B₁₂ deficiency in adults; overall, this may be a more common cause than autoimmune gastritis. In this case, there was no evidence of *H. pylori* infection, and the absence of the antral inflammation that is typical of this infection is further evidence against this diagnosis.

DISCUSSION OF MANAGEMENT

Dr. Mark: A hundred years ago, pernicious anemia was a major cause of illness and death. Thanks to the efforts of pioneers in the field of hematology, a treatment for this disorder was developed before its pathophysiology was understood.¹¹ George Minot, William Murphy, and George Whipple demonstrated that a daily diet that included about 100 to 200 g of beef liver (which we now know contains about 100 to 200 µg of vitamin B₁₂) could rapidly reverse the anemia. In an elegant series of experiments in which red meat was incubated with gastric juice for various lengths of time and then was administered to patients, William Cantle demonstrated that gastric juice appeared to contain an "intrinsic factor" that was required for absorption of the "extrinsic factor" (vitamin B₁₂). Vitamin B₁₂ was isolated almost simultaneously by two independent groups in 1948, and the nature of its structure was determined by x-ray crystallography.

Today, the treatment of pernicious anemia in-

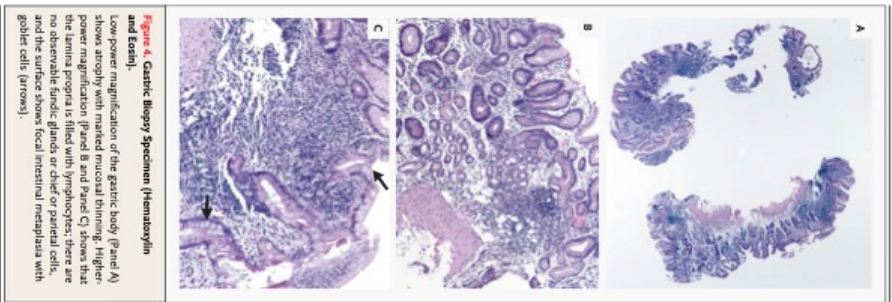


Figure 4. Gastric Biopsy Specimen (Hematoxylin and Eosin). Low-power magnification of the gastric body (Panel A) shows atrophy with marked mucosal thinning. Higher-power magnification (Panel B and Panel C) shows that the lamina propria is filled with lymphocytes; there are no observable fundic glands or chief or parietal cells, and the surface shows focal intestinal metaplasia with goblet cells (arrows).

whenever either parenteral or oral administration of vitamin B₁₂. Although monthly intramuscular injection of 100 to 1000 µg of vitamin B₁₂ has long been a standard treatment for vitamin B₁₂ deficiency, properly administered oral replacement is equally effective.¹² In patients with mild vitamin B₁₂ deficiency, oral administration of 1 to 2 mg of vitamin B₁₂ daily can be effective, even as initial therapy. In practice, it is prudent when treating patients with severe deficiency to give parenteral vitamin B₁₂ initially. After initial restoration of normal vitamin B₁₂ levels, a transition to oral therapy is not unreasonable, although periodic monitoring of the vitamin B₁₂ levels in such patients may be desirable. Those who doubt that vitamin B₁₂ can be absorbed by passive action in the absence of intrinsic factor need only look to the work of Minot and colleagues, by administering large quantities of liver, they accomplished the same goal.

A lingering question with respect to pernicious anemia is whether surveillance for gastric adenocarcinoma, carcinoids, and colonic polyps is warranted. Although some studies have shown that patients with pernicious anemia have a several-fold increase in the rate of gastric and colorectal cancers, others have not. At this time, there are no formal recommendations in the literature for screening patients with pernicious anemia for gastrointestinal cancers, although periodic screening of younger persons may be reasonable.

This patient could well have an autoimmune polyendocrine syndrome, probably type 11B. Autoimmune polyendocrine syndrome type 11B is associated with risks of other disorders, including hypogonadism, myasthenia gravis, rheumatoid arthritis, and sarcoidosis. The patient currently has no evidence of these disorders, but her increased risk for them should be kept in mind as she is followed.

Dr. Deborah J. Wexler (Medicine): What is the prognosis for the recovery of neurologic function?

Dr. Morke: Unfortunately, about 50 percent of patients are left with at least a mild neurologic deficit after the vitamin B₁₂ deficiency is corrected. Such residual deficits are more common in patients with long-standing or severe neurologic symptoms before therapy than in those with symptoms of recent onset or those with mild symptoms.

Dr. Lloyd Axelrod (Endocrinology): Could you comment on the risk of hypokalemia in patients with pernicious anemia who are treated with vi-

tamin B₁₂? When is potassium replacement indicated?

Dr. Morke: Parenteral vitamin B₁₂ replacement can result in a dramatic proliferation of bone marrow cells, which take up potassium, and patients can become hypokalemic very quickly. In a person whose vitamin B₁₂ deficiency is severe, I would monitor potassium levels and provide replacement therapy as necessary. Sudden death may occur in patients after treatment for vitamin B₁₂ deficiency is initiated; in addition to hypokalemia, volume overload may lead to death. When one sees a patient with severe vitamin B₁₂ deficiency and a hematocrit of 12 to 15 percent, one may be tempted to transfuse immediately to raise the hematocrit to a normal level. However, most of these patients have arrived at a low hematocrit over the course of years and tolerate it well; a transfusion volume sufficient

to raise the hematocrit to a normal level can cause volume overload and cardiac complications.

Dr. Zeligman-Gavril: In the case under discussion, we gave the patient 1 mg of parenteral vitamin B₁₂ daily for one week, followed by weekly injections for two months and monthly injections thereafter. The response was dramatic. Within 7 to 10 days, she was completely free of symptoms and had no neurologic deficits. I last saw her six months after the diagnosis was made, and she had completely recovered. She subsequently moved away from Massachusetts.

ANATOMICAL DIAGNOSIS

Pernicious anemia with autoimmune gastritis and vitamin B₁₂ deficiency.

Dr. Morke reports that he is now a senior clinical research physician at Harvard.

REFERENCES

1. Bal S, Pavithrasa. BMJ 2002;324:1301.
2. Tob MJ, van Driel RJ, Gerson DA, et al. Pernicious anemia. N Engl J Med 1999;351:1441-4.
3. Mitr A, Madhavan NK. Autoimmune diseases of the adrenal glands, parathyroid glands, pituitary, and hypothalamic-pituitary axis. Endocrinol Metab Clin North Am 1999;139:1299-348.
4. Morke D, Axelrod L. Pernicious anemia. N Engl J Med 1994;330:1444-50.
5. Cammell R, McIntyre S, Jones SJ. Cobalamin deficiency with and without neurologic abnormalities: differences in homocysteine, thymine on gastric mucosal histology and acid secretion in chronic gastritis. Gut 1973; 14:104-6.
6. Cammell R, Bryant C, Urdal AL, et al. Is cobalamin deficiency a disease? A new agent in vitamin B12 deficiency? Arch Intern Med 2000;160:1349-53.
7. Snow CE. Laboratory diagnosis of vitamin B12 and folate deficiency: a guide for the primary care physician. Arch Intern Med 1999;159:1299-348.
8. Morke D, Axelrod L, Thompson S, Madhavan NK. Sarcoidosis and pernicious anemia. Clin Exp Immunol 1984;63:359-66.
9. Cammell R, McIntyre S, Jones SJ. Cobalamin deficiency with and without neurologic abnormalities: differences in homocysteine, thymine on gastric mucosal histology and acid secretion in chronic gastritis. Gut 1973; 14:104-6.
10. Cammell R, Bryant C, Urdal AL, et al. Is cobalamin deficiency a disease? A new agent in vitamin B12 deficiency? Arch Intern Med 2000;160:1349-53.
11. Chazotte L. A history of pernicious anemia. Br J Haematol 2000;111:407-15.
12. Krumholz RM, De Groot LJ, Allen RH, et al. Guidelines for the diagnosis and treatment of vitamin B12 deficiency: a practice guideline. Blood 1998;92:1391-8.
13. Jones SJ, Cammell R, McIntyre S, et al. Pernicious anemia: a review of the clinical and laboratory features. Blood 1984;63:359-66.
14. Jones SJ, Cammell R, McIntyre S, et al. Pernicious anemia: a review of the clinical and laboratory features. Blood 1984;63:359-66.
15. Jones SJ, Cammell R, McIntyre S, et al. Pernicious anemia: a review of the clinical and laboratory features. Blood 1984;63:359-66.
16. Jones SJ, Cammell R, McIntyre S, et al. Pernicious anemia: a review of the clinical and laboratory features. Blood 1984;63:359-66.
17. Jones SJ, Cammell R, McIntyre S, et al. Pernicious anemia: a review of the clinical and laboratory features. Blood 1984;63:359-66.
18. Jones SJ, Cammell R, McIntyre S, et al. Pernicious anemia: a review of the clinical and laboratory features. Blood 1984;63:359-66.
19. Jones SJ, Cammell R, McIntyre S, et al. Pernicious anemia: a review of the clinical and laboratory features. Blood 1984;63:359-66.
20. Jones SJ, Cammell R, McIntyre S, et al. Pernicious anemia: a review of the clinical and laboratory features. Blood 1984;63:359-66.

35-MILLIMETER SLIDES FOR THE CASE RECORDS

Any reader of the Journal who uses the Case Records of the Massachusetts General Hospital as a medical teaching exercise or reference material is eligible to receive 35-mm slides, with identifying legends, of the pertinent x-ray films, electrocardiograms, gross specimens, and photomicrographs of each case. The slides are 2 in. by 2 in., for use with a standard 35-mm projector. These slides, which illustrate the current cases in the Journal, are mailed from the Department of Pathology to correspond to the week of publication and may be retained by the subscriber. Each year approximately 250 slides from 40 cases are sent to each subscriber. The cost of the subscription is \$450 per year. Application forms for the current subscription year, which began in January, may be obtained from Lantern Slides Service, Department of Pathology, Massachusetts General Hospital, Boston, MA 02114 (telephone 617-26-2749).

Slides from individual cases may be obtained at a cost of \$35 per case.

#2 Fall (Myokarditis) in CASUS (41, 74)

Der 19-jährige Herr Becker leidet an zunehmender Luftnot bei Belastung. Vor ca. zwei Wochen habe er eine Erkältung gehabt, die immer noch nicht richtig weg sei. Immer noch fühle er sich müde und in seiner Leistungsfähigkeit erheblich eingeschränkt. Er verneint Gewichtsverlust und Thoraxschmerzen. Vorerkrankungen sind nicht bekannt, er sei sonst auch nie ernsthaft krank gewesen, abgesehen von einer Appendektomie vor 11 Jahren. Keine Medikamenteneinnahme, kein Nikotinabusus, kein Alkoholkonsum.

In der körperlichen Untersuchung erheben Sie folgende Befunde:

Wacher, voll orientierter, kognitiv unauffälliger 19-jähriger Patient in reduziertem Allgemeinzustand und normalem Ernährungszustand, sportlicher Körperbau (85 kg, 1,90 m, BMI 24 kg/m²), zugewandt und freundlich, adäquate verbale Reaktion

Vitalparameter: Blutdruck 120/80 mmHg, Herzfrequenz 115/min, Atemfrequenz 20/min, Körpertemperatur (Ohr) 37,5 °C

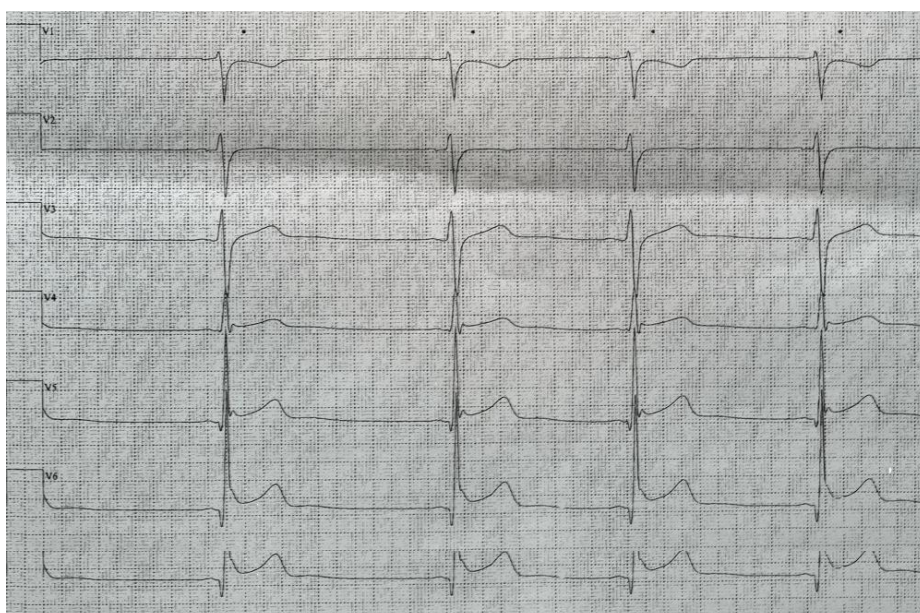
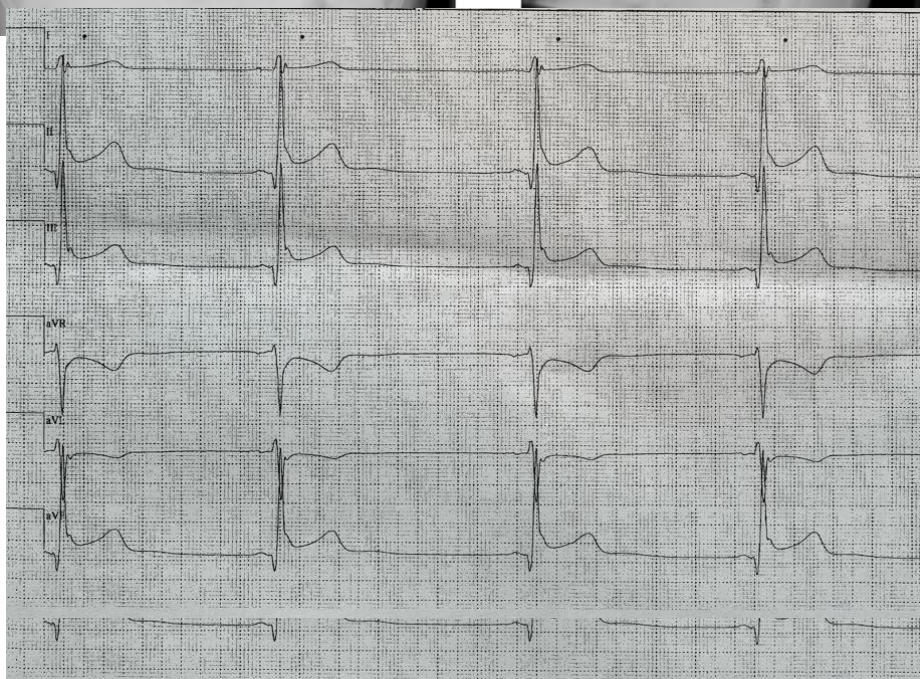
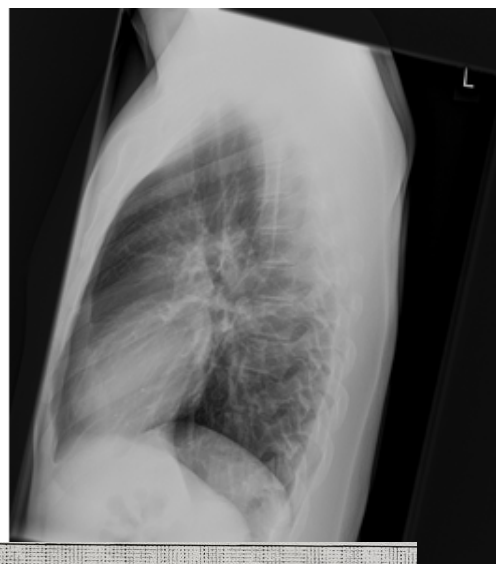
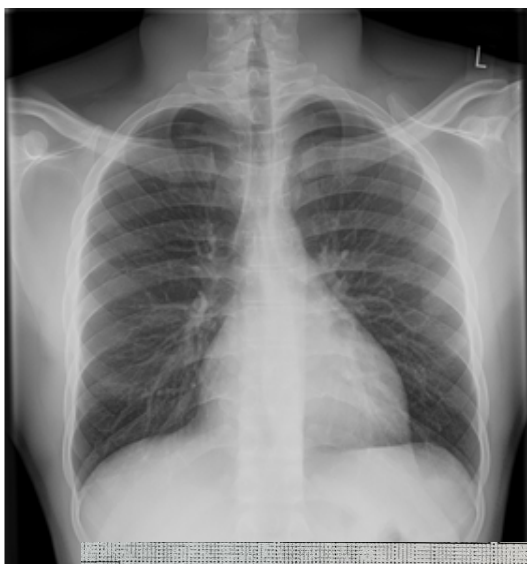
Kardiovaskuläres System: Herztöne rein und rhythmisch auskultierbar, 3. Herzton auskultierbar, keine Herzgeräusche, keine Strömungsgeräusche über den Carotiden auskultierbar. Keine Jugularvenenstauung, deutlich eindrückbare Ödeme an den Unterschenkeln bds.. Periphere Pulse (A. radialis bds., A. dorsalis pedis bds., A. tibialis posterior bds.) tastbar.

Respiratorisches System: Keine Thoraxdeformitäten. Keine Lippenzyanose. Sonorer Klopfeschall, atemabhängig bds. um 2 Querfinger verschiebbliche Lungengrenzen. Vesikuläres Atemgeräusch, keine feuchten oder trockenen Atemnebengeräusche. Brustwirbelsäule nicht klopfschmerzhaft.

Abdomen: Appendektomienarbe, Inspektion ansonsten unauffällig. Darmgeräusche lebhaft über allen vier Quadranten auskultierbar, keine abdominellen Strömungsgeräusche. Bauchdecke weich, Palpation nicht schmerzhaft, keine palpablen Resistenzen. Perkutorisch normaler Klopfeschall. Leber bei Inspiration mit weichem Leberrand und ohne Schmerzhaftigkeit tastbar, perkutorisch 11cm in der MCL. Milz nicht tastbar. Keine Hernien. Kein Nierenlagerklopfeschmerz. Lendenwirbelsäule nicht klopfschmerzhaft.

Lymphknoten: keine pathologisch vergrößerten zervikalen, axillären oder inguinalen Lymphknoten tastbar.

Neurologie: Pupillen mittelweit, isokor und prompt lichtreagibel, seitengleich auslösbare Reflexe (PSR, ASR, BSR), Gangbild unauffällig, kein fokalneurologisches Defizit.



Basislabor

Referenzwerte

Erythrozyten	5,0 l/pl	4,2-5,1/pl
Hämoglobin	15,8 g/dl	12,0-16,0 g/dl
Hämatokrit	0,45	0,36-0,46
Leukozyten	13,6 G/l	4,0-11,0 G/l
Thrombozyten	210 G/l	150-440 G/l
Kreatinin	0,8 mg/dl	0,5-1,0 mg/dl
Natrium	137 mmol/l	135-145 mmol/l
Kalium	4,5 mmol/l	3,5-5,0 mmol/l
CRP	1,5 mg/dl	< 0,5 mg/dl
GOT	28 U/l	< 33 U/l
GPT	33 U/l	< 35 U/l
γ-GT	34 U/l	< 38 U/l
Quick	110 %	70-130%
INR	0,9	0,8-1,2
aPTT	24 sec	23-35 sec
Glucose	85 mg/dl	70-115 mg/dl
Harnstoff	4,5 mmol/l	2,0 – 8,0 mmol/l
GFR	>60	>60
TSH	2,2 mU/l	0,4 – 2,5 mU/l

Erweitertes Labor

Chlorid	102 mmol/l	96-110 mmol/l
Magnesium	0,85 mmol/l	0,70-1,20 mmol/l
Calcium	2,55 mmol/l	2,15-2,60 mmol/l
BSG	24 mm/h	Bis 15 mm bzw. 20 mm/h
Bilirubin	1,6 mg/dl	< 1,0 mg/dl
Cholesterin/TAG	180 mg/dl	150-220 mg/dl
Ferritin	123 µg/l	23- 217 µg/l
CK-MB	8 µg/l	<5 µg/l
Troponin T	0,65 ng/ml	<0,1 ng/ml
Harnsäure	4,7 mg/dl	2,5 – 7,0 mg/dl
D-Dimer	210 ng/ml	<500 ng/ml
PCT	0,04 µg/l	<0,05 µg/l
HbA1C	4,2 %	3,0 – 6,0 %
Lactat	1,82 mmol/l	0,63- 2,44 mmol/l
Differentialblutbild:		
Neutrophile	72 %	40-75 %
Eosinophile	4 %	1-6 %
Basophile	0 %	0 – 1 %
Monozyten	4 %	2 – 8 %
Lymphozyten	20 %	20 – 45 %

BGA

Wert	Befund	Referenzbereich
-------------	---------------	------------------------

pH	7,39	7,36- 7,44
pO ₂	82 mmHg	75-98 mmHg
s _a O ₂	95 %	95-97 %
pCO ₂	32 mmHg	35-45 mmHg
Standardbikarbonat	27 mmol/l	20-28 mmol/l
BE	0,5 mmol/l	-2- +2 mmol/l
Probenart	arteriell	
FiO ₂	Raumluft	
AF	20/min	

Lufu

	Soll	Ist	%(Ist1/Soll)
VC MAX [l]	4.27	4.67	109.37
FEV 1 [l]	3.19	3.75	117.55
FEV 1 % VC MAX[%]	75.51	80.02	105.97
MEF 50 [l/s]	4.31	5.05	117.17
MEF 25 [l/s]	1.56	1.64	105.13
TLC [l]	6.98	7.08	101.4
RV [l]	2.51	1.77	70.5
TLCoc/VA[mmol/min/kPa/l]	1.32	1.17	88.8

Legende

VC	Vitalkapazität
FEV	forciertes expiratorisches Volumen
MEF	Maximaler expiratorischer Fluss
TLC	Totale Lungenkapazität
RV	Residualvolumen
TLCoc/VA	CO-Transferkoeffizient

Echokardiographie

Aortenwurzel normal weit - AK: altersentsprechend, Öffnungsbewegung normal, keine Insuffizienz, keine Stenose ($V_{\max} = 1,9 \text{ m/s}$, $dp_{\max} = 14 \text{ mmHg}$) - RA: normale Größe (visuell beurteilt) - LA: normale Größe - RV: normale Größe, Wanddicke normal, normale RV-Funktion - LV: normale Größe, Globale mäßige systolische Global-Funktionseinschränkung (planimetrische EF biplan~40%) ohne Wandbewegungsstörungen. Diastolische Funktionseinschränkung Grad 2 (Pseudonormalisierung).- MK: altersentsprechend, Bewegung normal, geringe Insuffizienz, keine Stenose - TK: altersentsprechend, Bewegung normal, geringe Insuffizienz, $dp_{\max} \text{ RV/RA} = 22 \text{ mmHg}$ - V. cava 23 mm, ausreichend atemmoduliert – gering gradiger, nicht punktionswürdiger Perikarderguß nachweisbar

Legende

AK	Aortenklappe
RA	Rechter Vorhof
LA	Linker Vorhof
RV	Rechter Ventrikel
LV	Linker Ventrikel
MK	Mitralklappe
TK	Trikuspidalklappe
KÖF nach Kont	Klappenöffnungsfläche
EF	Ejektionsfraktion
VCI	Vena cava inferior

Sonographie

Leber: Gut beurteilbar, normal groß, glatte Organkontur, Unterrand spitzwinklig, homogenes echonormales Binnenreflexmuster, unauffällige Gefäßarchitektur, keine fokalen Läsionen.

Gallenwege: Gut beurteilbar, kein Nachweis einer intra- oder extrahepatischen Gallenwegserweiterung, kein Steinnachweis, kein Hinweis für intraluminale Raumforderung. Max. Weite des DHC 5 mm.

Gallenblase: gut beurteilbar, Abbildungsqualität sehr gut.

Regelrechte Lage, normal groß, normale Form, Wand: akzentuiert, Wanddicke: 2 mm, Palpation: kein Druckschmerz.

Pankreas: gut beurteilbar mit Ausnahme des Schwanzes, Abbildungsqualität sehr gut, Pankreasgang unauffällig. Soweit einsehbar unauffälliger Organbefund, scharf abgrenzbar zur Umgebung.

Milz: Gut beurteilbar, normal groß, echonormales homogenes Binnenreflexmuster, keine fokalen Veränderungen, Hilus frei.

Rechte Niere: gut beurteilbar, Abbildungsqualität sehr gut.

Lage: orthotop, normal groß, normale Form, Parenchym: normal breit. Metrik: Größe: 108 mm x 54 mm x 33 mm, Vol. 101 ml, Parenchymbreite 19 mm.

Linke Niere: Gut beurteilbar, orthotope Lage, normale Organgröße, normale Form, Parenchymsaum altersentsprechend, glatte Organkontur, kein Harnstau, keine Konkreme. Größe: 114 mm x 50 mm x 35 mm, Vol. 104 ml, Parenchymbreite 13 mm.

Harnblase: Eingeschränkt beurteilbar, soweit erkennbar orthotop gelegen, unauffällige Wandverhältnisse, intraluminal keine pathologischen Echostrukturen, normale Organgröße.

Pleura: Beidseits kein Pleuraerguss.

Perikard: Kleiner Perikarderguss.

Peritoneum: Kein Aszites.

ZUSAMMENFASSENDE BEURTEILUNG

Kleiner Perikarderguss (ICD10 I30.1)

Urinuntersuchung

Leukozyten	negativ
Nitrit	negativ
Protein	negativ
pH	5,5
Blut	negativ
Ketone	negativ
Glukose	negativ

Bakteriologie

1. Material: Sputum

Mikroskopie

grampositive Kokken	-	Erythrozyten	-
Detritus	-	Leukozyten	-

Kultur

+++	Mundflora

Beurteilung:
negativer mikrobiologischer Befund

2. Material: Blutkultur

Kultur

Kein Wachstum

Beurteilung:
negativer mikrobiologischer Befund

Mengenangaben bedeuten: +++ viel, ++ mäßig viel, + wenig, (+) sehr wenig

#3 Auszug aus der Scoring-Tabelle: Kurzfall (17)

Dimensionen		Anamnese												
Kategorien	Gesch. akt. Erkr	Vordiagnosen	Vorbehandlung	Medikamente	Psychosoziale Anamnese	Allergien	Familienanamnese	Reiseanamnese						
Klassen					beliebig	relevant	beliebig	relevant	beliebig	relevant	beliebig	relevant	Punkte	
Punkte	5	1	0	0	0	0	0	0	0	0	0	0	6	
Anamnese														
Organsysteme														
allgemein	Kopf/ Hals	kard-vas.	respi	GI	nephro/ uro	musk	neuro	psych	endokrin	häm-onko	Haut			
0	0	0	0	1	1	0	0	0	0	0	0	2		
Untersuchungen														
Bildgebung														
allgemein	Kopf/ Hals	kard-vas.	respi	GI	nephro/ uro	musk	neuro	psych	endokrin	häm-onko	Haut	0		
0	0	0	0	0	0	0	0	0	0	0	0			
Untersuchungen														
Labor														
allgemein	Kopf/ Hals	kard-vas.	respi	GI	nephro/ uro	musk	neuro	psych	endokrin	häm-onko	Haut	9		
0	0	0	1	0	0	0	0	0	4 4	0	0			
Untersuchungen														
Körperliche Untersuchung														
allgemein	Kopf/ Hals	kard-vas.	respi	GI	nephro/ uro	musk	neuro	psych	endokrin	häm-onko	Haut	4		
0	1 1	0	1 1	0	0	0	0	0	0	0	0			
Untersuchungen														
Funktionelle Tests														
allgemein	Kopf/ Hals	kard-vas.	respi	GI	nephro/ uro	musk	neuro	psych	endokrin	häm-onko	Haut	0		
0	0	0	0	0	0	0	0	0	0	0	0			
21														

Eidesstattliche Erklärung

Ich, Kim Sarah Öhler, erkläre hiermit an Eides statt, dass ich die vorliegende Dissertation *Case-based Learning – Was macht einen Fall komplex?* selbstständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

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

München, den 19.02.2024

Kim Sarah Öhler