Erfolg in Mathematikklausuren ingenieurwissenschaftlicher Studiengänge unter besonderer Berücksichtigung prozeduralen Wissens

Dissertation

zur Erlangung des akademischen Grades eines

Doktors der Pädagogik

(Dr. paed.)

der Fakultät für Mathematik der

Technischen Universität Dortmund

vorgelegt von

Mike Altieri

am 15. Juni 2016

Dissertation

Erfolg in Mathematikklausuren ingenieurwissenschaftlicher Studiengänge unter besonderer Berücksichtigung prozeduralen Wissens

Fakultät für Mathematik
Technische Universität Dortmund

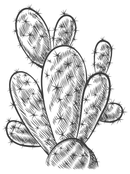
Erstgutachterin: Prof. Dr. Susanne Prediger

Zweitgutachter: Prof. Dr. Reinhard Hochmuth

Tag der mündlichen Prüfung: 31.8.2016

für Gábor

1947-2014



in jedem Anfang liegt ein Ende
in jedem Ende liegt ein Anfang

Inhaltsverzeichnis

Einleitung

„Die Mathematik ist ***die*** *Hürde. Das ist nicht eine Hürde,*
es ist ***die*** *Hürde im MINT-Bereich.*
Auch bei Ingenieuren.“

Prof. Dr. Metin Tolan
Prorektor Studium der TU Dortmund

In einer internen Untersuchung aller Klausuren eines Semesters mit einer Durchfallquote von über 50 % konstatiert Tolan (2013), dass es sich bei 18 von 20 Klausuren um Mathematikklausuren handelt. Dies gilt vermutlich nicht nur an der TU Dortmund. Zwar veröffentlichen Hochschulen keine offiziellen Zahlen zu Durchfallquoten in Mathematikklausuren, erfahrungsgemäß liegt diese Quote aber selten unterhalb von 30 %, sondern tendenziell bei 50 % – auch an anderen Hochschulen[1].
Dass viele Studierende in Mathematik an der Hochschule scheitern, ist allgemein bekannt: Darauf deuten die vielen populärwissenschaftlichen Artikel hin, die die Problematik thematisieren (z. B. Doyé, 2013; DPA, 2012; Vollmers, 2011). Aber egal, wo man hinsieht: Sowohl in den öffentlichen Medien, in privaten Gesprächen und natürlich auch in der Hochschule – überall wird kritisch und oft verwundert gefragt, warum denn so viele Studierende an den Mathematikklausuren im Studium scheitern. Umgekehrt lässt sich natürlich auch fragen: Welchen Hintergrund hat Erfolg in Mathematikklausuren an der Hochschule? Betrachtet man so die notwendigen Ressourcen für Erfolg, ist davon auszugehen, dass Ansätze für sinnvolle, diagnostische Aussagen gefunden werden können, an welchen Hindernissen manche Studierende konkret scheitern.

Während über Erfolg und Misserfolg in der Schulmathematik insbesondere im Bereich der Diagnose und Förderung bereits viel geforscht worden ist (Prediger & Leuders, 2016; Moser Opitz & Nührenbörger, 2015; Scherer & Moser Opitz, 2010; Moser Opitz, 2007; Freesemann, 2014), ist dieser Aspekt in der Hochschulmathematik aus didaktischer Sicht bisher wenig untersucht. Zu nennen sind hier Ansätze von Tynjälä, Salminen, Sutela, Nu-

[1] Diese Beobachtung leitet sich aus persönlichen Gesprächen mit Dozentinnen und Dozenten anderer Hochschulen ab. Das Thema Durchfallquoten wird sehr sensibel behandelt, sodass keine frei zugängliche, verallgemeinerbare Diskussionsgrundlage existiert.

utinen und Pitkänen (2005) sowie Rach und Heinze (2013), die sich mit Lernstrategien als Schlüssel zum Erfolg befassen - wie beispielsweise Selbsterklärungen beim Mathematiklernen. Ferner sind einige durch das BMBF geförderte Projekte zu nennen, wie beispielsweise „Verbund StuFo: Der Studieneingang als formative Phase für den Studienerfolg. Analysen zur Wirksamkeit von Interventionen“ (Mauermeister, Zylla & Wagner, 2015). Das Ziel des breit aufgestellten Verbunds ist die Analyse von Wirkungszusammenhängen zwischen Modellen der Studieneingangsphase und dem Erfolg des Studiums zur Studienmitte.

Eine systematische Analyse, woran Studierende in Mathematikklausuren scheitern, fehlt bisher allerdings. Diese Frage ist insbesondere für Studierende der ingenieurwissenschaftlichen Studiengänge relevant, da das Studium äußerst mathematikbetont ist und Vorlesungen im Verlauf des Studiums auf dem Mathematikwissen aus vorangegangenen Semestern aufbauen. Zudem sind die Studienabbruchquoten in ingenieurwissenschaftlichen Fächern mit bis zu 51 % im Studiengang Bauingenieurwesen nach wie vor hoch[2] (Heublein et al., 2014, S. 4). In 70 % dieser Fälle werden Leistungsprobleme als ein Motiv für den Studienabbruch genannt (Heublein, Hutzsch, Schreiber, Sommer & Besuch, 2010, S. 20). Nach dem Eingangszitat von Tolan (2013) sowie den Interviews in Doyé (2013), DPA (2012) und Vollmers (2011) stellen Mathematikklausuren hierbei **die** Hürde dar. Eine detaillierte Analyse der Scheiterungsgründe erscheint daher gerade für Mathematikklausuren der ingenieurwissenschaftlichen Studiengänge hoch relevant. Diese Relevanz gilt zwar auch für andere mathematikbetonte Studiengänge wie etwa das Mathematikstudium selbst, die vorliegende Arbeit fokussiert jedoch auf die größte Studierendengruppe in den ingenieurwissenschaftlichen Studiengängen, denn hier liegt das langjährige Arbeitsfeld des Autors. Aus der Analyse sollten sich dann Ansatzpunkte für zukünftige Unterstützungsmaßnahmen ableiten, die mittelfristig zu einer Reduzierung von Durchfallquoten und Dropouts beitragen können.

Daher werden in der vorliegenden Arbeit die Hintergründe von Erfolg und Misserfolg in Mathematikklausuren am Beispiel der Gruppe der Studierenden aus den ingenieurwissenschaftlichen Fächern untersucht. Die Arbeit leistet damit einen Beitrag zur Grundlagenforschung im diagnostischen Bereich der Hochschuldidaktik. Die übergreifende Fragestellung der Arbeit lautet dementsprechend:

Welche Hintergründe haben Erfolg und Misserfolg in Mathematikklausuren der ingenieurwissenschaftlichen Studiengänge?

[2] Die Abbruchquoten in den anderen erfassten ingenieurwissenschaftlichen Studiengängen an Universitäten lauten nach Heublein, Richter, Schmelzer und Sommer (2014): Maschinenbau (36 %) und Elektrotechnik (37 %).

Zur Annäherung stellt sich zunächst die Frage, welche Anforderungen Klausuren der Ingenieurmathematik abprüfen. In der Fachliteratur findet sich diesbezüglich der Vorschlag, zwischen Anforderungen an prozedurales Wissen und konzeptuelles Wissen zu unterscheiden (Engelbrecht, Harding & Potgieter, 2005; Hiebert, 1986). Eine Analyse der Literatur zu diesen Konstrukten zeigt die wiederholt kritische Einschätzung, dass in den ingenieurwissenschaftlichen Studiengängen Mathematik im Wesentlichen prozedural gelehrt, gelernt und geprüft wird (Bergqvist, 2007; Engelbrecht, Bergsten & Kågesten, 2012; Zerr, 2009; Schoenfeld, 1989; Hallett, 2006). Unter prozeduralem Wissen wird in der vorliegenden Arbeit die Verknüpfung von Kalkülkenntnis und Kalkülfertigkeit verstanden, vgl. Abbildung 0.1:

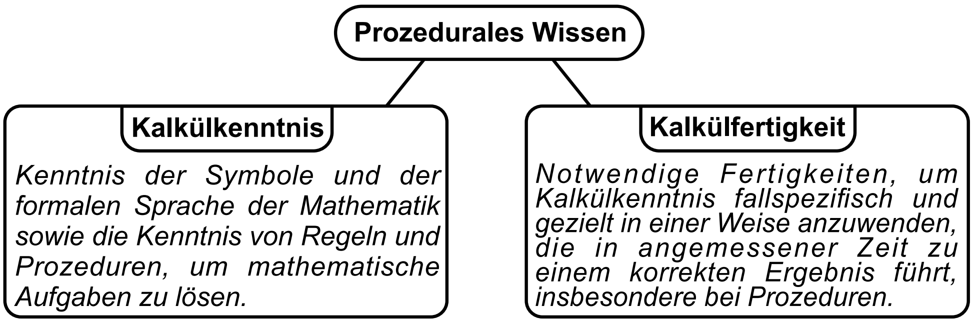


Abbildung 0.1 Aufschlüsselung prozeduralen Wissens in Anlehnung an Hiebert und Lefevre (1986)

Prozedurales Wissen scheint also eine hohe Relevanz in den Mathematikklausuren der ingenieurwissenschaftlichen Fächer zu haben. Doch enthalten diese Klausuren nicht nur leicht routinisierbare Basisprozeduren, sondern komplexe Anforderungsprofile mit einer Verknüpfung zahlreicher Wissenselemente. Aus der Forschung liegen allerdings bisher keine Befunde darüber vor, inwieweit schon das Beherrschen einfacher Basisprozeduren eine Prognosekraft für den Erfolg in Klausuren der Ingenieurmathematik hat.

Die empirischen Analysen zur Prognose von Klausurerfolg von Studierenden haben sich hingegen auf andere Faktoren konzentriert (Trost, Klieme & Nauels, 1997; Kobrin, Patterson, Shaw, Mattern & Barbuti, 2008). So erwiesen sich beispielsweise die Abiturnote und Einstufungstests als sinnvolle Prädiktoren für den Erfolg in Klausuren der Studieneingangsphase. Ebenso sind Zusammenhänge zwischen Klausurleistung und allgemein kognitiven Faktoren - wie z. B. kognitiver Grundfertigkeit und Konzentration, sowie affektiven Faktoren - wie beispielsweise Prüfungsangst und Prokrastination, nachgewiesen (Rindermann & Neubauer, 2000; Jensen, 1981; Seipp, 1991; Kim & Seo, 2015). Die vorgenannten Prädiktoren haben allerdings weder mit den Inhalten einer Mathematikvorlesung für ingenieurwissenschaftliche Studiengänge noch mit dem Klausurinhalt viel gemeinsam. Abgesehen von fachspezifischen Einstufungstests, die in der Regel *vor* Veranstaltungsbeginn durchgeführt werden, handelt es sich um Prädiktoren, die per se nichts mit Ingenieurmathematik zu tun haben. In der Literatur finden sich keine Untersuchungen zu *fachspezifischen* Prädiktoren, die zur Prognose des Erfolgs in Klausuren der Ingenieurmathematik dienen, obwohl Analysen von Anforderungsprofilen dieser Klausuren durchaus existieren (Bergqvist, 2007). Entsprechend fehlen Untersuchungen zu den Zusammenhängen zwischen fachspezifischen Prädiktoren und Klausurerfolg. Die vorliegende Arbeit versucht, diese Lücke zu schließen, indem sie prozedurales und konzeptuelles Wissen als Vertreter sowohl der fachspezifischen Prädiktoren als auch der (Grund-)Anforderungen im Anforderungsprofil von Klausuren der Ingenieurmathematik in entsprechende Analysen einschließt.

Für die Analysen ist sicherzustellen, dass die untersuchte Studierendenkohorte repräsentativ ist. Zu diesem Zweck werden bekannte affektive und allgemein kognitive Prädiktoren mit in die Untersuchungen einbezogen und letztere ergebnisoffen durchgeführt. Es wird also die Vorhersagbarkeit von *Erfolg in Mathematikklausuren ingenieurwissenschaftlicher Studiengänge unter besonderer Berücksichtigung prozeduralen Wissens* untersucht, denn prozedurales Wissen gilt wie erwähnt als dominant in Klausuren der Ingenieurmathematik. Ziel ist es, aus den Resultaten Ansatzpunkte für gezielte und empirisch begründete Unterstützungsangebote abzuleiten. Umgekehrt soll sich durch Einbezug verschiedenartiger Prädiktoren auch beurteilen lassen, in welchen Bereichen geringer oder kein Unterstützungsbedarf besteht. Die Untersuchungen können also zu einer Priorisierung von Unterstützungsangeboten beitragen. Dabei wird zunächst überprüft, ob das prozedurale Wissen tatsächlich als Prädiktor für den Klausurerfolg (bestanden versus nicht bestanden) und Klausurleistung (erreichte Punktzahl) der Studierenden der ingenieurwissenschaftlichen Fächer taugt. Hierzu wurden Studien durchgeführt, die in der vorliegenden Arbeit unter den folgenden beiden Forschungsfragen analysiert werden:

- Forschungsfrage 1: *Wie groß sind die Zusammenhänge zwischen Klausurleistung und fachspezifischen, allgemein kognitiven sowie affektiven Faktoren?*
- Forschungsfrage 2: *In welchem Maße lassen sich Klausurerfolg und Klausurleistung vorhersagen?*

Es ist bereits an dieser Stelle zu erwähnen, dass in den Untersuchungen der beiden Forschungsfragen die These bestätigt werden kann, dass es einen signifikanten und großen Zusammenhang zwischen prozeduralem Wissen und Klausurerfolg und -leistung gibt. Diese Feststellung alleine reicht allerdings noch nicht aus, um Studierende bei ihrem Klausurerfolg zu unterstützen. Zwar ist eine Förderung prozeduralen Wissens auch aus fachdidaktischer Sicht naheliegend, da prozedurales Wissen ein – möglicherweise sogar dominanter – Bestandteil des Anforderungsprofils von Klausuren der Ingenieurmathematik ist. Noch unklar ist jedoch, *welche* Studierenden *wann* und *worin* unterstützt werden sollten. Das fehlende Verbindungsstück zwischen obiger Feststellung und der eben formulierten Anschlussfrage wird in der vorliegenden Arbeit aus folgenden drei Teilen zusammengesetzt:

- Das erste Teil besteht in der Spezifikation des „*worin*“. Dazu ist ein Aufschlüsseln des prozeduralen Wissens erforderlich, das zeigt, was dieses im Detail ist bzw. was die notwendigen Ressourcen innerhalb des prozeduralen Wissens ausmacht. Zwar existiert mit dem grundlegenden Standardwerk von Hiebert (1986) ein solider theoretischer Bezugspunkt, dieser reicht allerdings nicht aus, um beispielsweise typische Schwächen von Studierenden bei einer fehlerhaften Lösung prozeduraler (Klausur-)Aufgaben zu beschreiben. Auch weitere Literatur zu prozeduralem Wissen im Hochschulbereich (z. B. Engelbrecht, Bergsten & Kågesten, 2009; Engelbrecht et al., 2005; Mahir, 2009; Chappell & Killpatrick, 2003; Arslan, 2010) liefert keinen Hinweis auf diesen Teil des Verbindungsstücks, d. h. *worin* Studierende spezifisch unterstützt werden müssen, um ihr prozedurales Wissen auszubauen. Die vorliegende Arbeit schließt diese Lücke und zerlegt prozedurales Wissen in bedeutsame Bestandteile, die sich als (Grund-)Anforderungen im komplexen Anforderungsprofil von Klausuren der Ingenieurmathematik wiederfinden. Dadurch soll prozedurales Wissen Analysen zugänglich werden, die beispielsweise auf die Untersuchung von Gelingensbedingungen zur Lösung prozeduraler Aufgaben unter anderem in Klausuren ausgerichtet sind. Die Aufschlüsselung wird im Theorieteil der Arbeit geleistet, die empirische Absicherung ergibt sich als Teilergebnis bei der Diskusssion von Forschungsfrage 2.

- Das zweite Teil spezifiziert, *welche* Studierenden unterstützt werden sollen. Aus der Literatur ist bekannt, dass Studierende mit einem Leistungskurs in Mathematik leistungsstärker (in Mathematik) sind und seltener ein ingenieurwissenschaftliches Studium abbrechen als Studierende mit einem Grundkurs in Mathematik (Maag Merki, 2012; Büning, 2004; Heublein et al., 2010). Aufgrund des Zusammenhangs zwischen Klausurleistung und prozeduralem Wissen müssten folglich Studierende mit einem Leistungskurs in Mathematik über ein größeres prozedurales Wissen verfügen und damit hier weniger Unterstützungsbedarf aufweisen als Studierende mit einem Grundkurs in Mathematik. Ob dies tatsächlich der Fall ist, wurde bisher noch nicht untersucht. Um diese Forschungslücke zu schließen, wird die genannte These in Forschungsfrage 3 (*Welche Merkmalsverteilungen weisen Gruppen prozedural schwacher, prozedural durchschnittlicher und prozedural starker Studierender auf?*) überprüft. Ferner werden hier analoge Analysen für weitere Hintergrund- und Leistungsmerkmale durchgeführt, für die aus der Literatur Zusammenhänge mit der Leistung im Studium bekannt sind.

- Gegenstand des dritten Teils ist, *wann* Studierende unterstützt werden sollten. In der Literatur wird das erste Studienjahr als maßgebend für den Erfolg im weiteren Studienverlauf angesehen (Ortiz & Dehon, 2013; Henn & Polaczek, 2007; Kolb, Kraus, Pixner & Schüpbach, 2006). Es existieren allerdings keine Untersuchungen, wie sich prozedurales Wissen in verschiedenen Leistungsgruppen von Studierenden innerhalb des ersten Jahres und darüber hinaus entwickelt. Die vorliegende Arbeit versucht, diese Lücke im Rahmen von Forschungsfrage 4 zu schließen (*Wie entwickelt sich prozedurales Wissen bei ausgewählten Studierendengruppen während der ersten Studiensemester?*). Es wird analysiert, ob sich auch hier der Erfolg im ersten Studienjahr auf die weitere Entwicklung prozeduralen Wissens auswirkt. Dabei wurde bewusst auf systemische Analysen oder Veränderungen der Lehr-Lern-Arrangements verzichtet, denn es sollte – auch mit Blick auf die Übertragbarkeit der Resultate auf andere Hochschulen – die Entwicklung prozeduralen Wissens im aktuellen, von speziellen didaktischen Maßnahmen weitgehend unberührten System untersucht werden.

Aus der Literatur sind qualitative Analysen typischer Fehlermuster im prozeduralen Wissen von Studierenden insbesondere im grundlagenmathematischen Bereich bekannt (Kersten, 2015). Es fehlen allerdings Längsschnittstudien im Hochschulbereich, die untersuchen, wie sich Ressourcen und Schwächen mit der Zeit verändern und worin im Detail Studierende bei der Bearbeitung prozeduraler Aufgaben scheitern. Die vorliegende Arbeit leistet hierzu im Rahmen von Forschungsfrage 5 (*Woran scheitern prozedural schwache Studierende bei der Durchführung von Basisprozeduren und wie entwickeln sich individuelle Ressourcen und Schwächen?*) einen Beitrag, denn das zugrundeliegende Erhebungsinstrument zur Messung prozeduralen Wissens bietet auch die Möglichkeit für qualitative Analysen. Im Fokus von Forschungsfrage 5 stehen daher qualitative Untersuchungen. In den qualitativen Auswertungen wird exemplarisch untersucht, wie sich Ressourcen und Schwächen in der Kalkülkenntnis und -fertigkeit bei prozedural schwachen Studierenden in den ersten Semestern entwickeln. Dadurch können wichtige Erkenntnisse für Unterstützungsangebote gewonnen werden, denn die Analysen liefern Hinweise auf Förderbedarfe.

Wie bereits angedeutet, wird zuweilen kritisiert, dass Mathematik in den ingenieurwissenschaftlichen Fächern im Wesentlichen prozedural geprüft wird (z. B. Bergqvist, 2007). Im deutschsprachigen Raum existieren keine entsprechenden Untersuchungen bezüglich der Anforderungen an prozedurales Wissen in Klausuren der Ingenieurmathematik. Zur Schließung dieser Forschungslücke sollen die im Rahmen von Forschungsfrage 6 (*Inwiefern sind die in der Referenzklausur gestellten Anforderungen mit den entsprechenden Anforderungen analoger Klausuren an anderen Hochschulen vergleichbar?*) durchgeführten Analysen von Klausuren anderer Hochschulen beitragen. Forschungsfrage 6 ist damit relevant für die weitere Diskussion, insbesondere inwiefern prozedurales Wissen in Mathematikklausuren für ingenieurwissenschaftliche Studiengänge über den lokalen Rahmen hinaus eine wichtige Rolle spielt und wie aus der vorliegenden Arbeit resultierende Anknüpfpunkte für weitere Forschungsarbeiten hinsichtlich ihrer Generalisierbarkeit zu bewerten sind.

Bei der Untersuchung der Forschungsfragen werden verschiedene Korrelations- und Regressionsanalysen durchgeführt, insbesondere zur Beurteilung von Prognosegüten für Klausurerfolg und -leistung. Hinzu kommen deskriptiv- und inferenzstatistische Methoden, um Aussagen über Signifikanz und Bedeutsamkeit der Ergebnisse zu treffen. Zur Untersuchung des prozeduralen Wissens bei Studierenden werden Quer- und Längsschnittanalysen sowie eine qualitative Analyse von Studierendenbearbeitungen durchgeführt.

Aufbau der Arbeit

Die ersten beiden Kapitel stellen den Theorieteil der Arbeit dar: In *Kapitel 1* werden prozedurales und konzeptuelles Wissen eingeführt und definiert. Nach einer Erörterung der Zusammenhänge zwischen prozeduralem und konzeptuellem Wissen in der Ingenieurmathematik wird die Relevanz prozeduralen Wissens an Hochschulen diskutiert und dabei dessen besondere Berücksichtigung in der vorliegenden Arbeit begründet.
Kapitel 2 stellt die Faktoren *kognitive Grundfertigkeit* und *Konzentration* sowie *Prüfungsangst* und *Prokrastination* vor. Als allgemein kognitive bzw. affektive Faktoren grenzen sich diese klar von den beiden fachspezifischen Faktoren aus Kapitel 1 ab. Neben der Konzeptualisierung der einzelnen Konstrukte wird jeweils auf deren Operationalisierung und die zu erwartenden Zusammenhänge mit der Klausurleistung eingegangen.
Kapitel 3 diskutiert aufbauend auf dem Theorieteil die übergeordnete Fragestellung der vorliegenden Arbeit und beschreibt die Forschungsfragen, die anhand der vorausgegangenen Überlegungen ausdifferenziert werden. Ferner werden bereits praktische Konsequenzen angedeutet, die sich aus der Diskussion der Forschungsfragen ergeben.
Den Methodenteil der Arbeit bildet *Kapitel 4*: Zunächst erfolgt in Abschnitt *4.1* ein Überblick darüber, zu welchen Zeitpunkten welche Erhebungen durchgeführt wurden und wie sich die Stichproben typischerweise zusammensetzten. Anschließend wird einheitlich für alle Erhebungsinstrumente deren Objektivität erläutert, gefolgt von einer Beschreibung der Referenzklausur, auf die sich die durchgeführten Korrelations- und Regressionsanalysen beziehen. In den Abschnitten *4.2* bis *4.4.* werden die Instrumente der Datenerhebung vorgestellt und insbesondere auf deren psychometrische Qualität eingegangen. Der abschließende Abschnitt *4.5* geht auf die Methoden der Datenauswertung ein.

In *Kapitel 5* werden auf Grundlage der vorangegangenen Kapitel zu jeder Forschungsfrage jeweils die Fragestellung und methodische Vorgehensweise zusammengefasst, die empirischen Ergebnisse präsentiert und im Hinblick auf die jeweilige Forschungsfrage interpretiert. Dabei wird auch auf Grenzen des Forschungsprojekts hingewiesen und auf mögliche Konsequenzen für die Lehre und Anschlussfragen eingegangen.
Geschlossen wird die Arbeit in *Kapitel 6* mit einer zusammenfassenden Formulierung der Befunde und Konsequenzen für die Lehre, Anschlussfragen sowie einem Ausblick auf mögliche anschließende Forschungsvorhaben.

1 Theoretischer Rahmen: Prozedurales und konzeptuelles Wissen - Komplexe Prüfungsinhalte und prädiktiv isolierbare Anforderungen

Prozedurales und konzeptuelles Wissen stehen in einer langen Tradition mit verwandten Konstrukten, wie beispielsweise „instrumental understanding“ und „relational understanding“ (Skemp, 1979) sowie „procedural knowledge“ und „declarative knowledge“ (Anderson, 1983)[1]. Eine ebenso lange Tradition hat die Diskussion, in welchem Verhältnis prozedurales und konzeptuelles Wissen unterrichtet werden sollten (Wu, 1999). Die Frage nach dem Verhältnis zwischen prozeduralem und konzeptuellem Wissen ist schließlich auch an Mathematikklausuren ingenieurwissenschaftlicher Studiengänge zu stellen, wenn deren Anforderungsprofil sichtbar gemacht werden soll. Wie bereits in der Einleitung erwähnt, stellen Mathematikklausuren für viele Studierenden eine Hürde im Studium dar. Um infolgedessen geeignete Unterstützungsangebote entwickeln zu können, ist es aus fachdidaktischer Sicht hochrelevant, ein entsprechendes Anforderungsprofil für die genannten Klausuren skizzieren zu können. In Abschnitt 1.2.2 werden dazu exemplarisch Anforderungen anhand einiger Klausuraufgaben diskutiert. Dabei wird einerseits deutlich, dass in der Tat sowohl Anforderungen an konzeptuelles als auch prozedurales Wissen gestellt werden, andererseits zeigt sich, dass in den Aufgaben die Anforderungen an prozedurales Wissen hinsichtlich ihrer Komplexität ein breites Spektrum von gering bis hochkomplex abdecken. Das Ziel dieses Kapitels ist daher, einen geeigneten Rahmen bereitzustellen, um diese unterschiedlichen Anforderungen an prozedurales Wissen in Mathematikklausuren ingenieurwissenschaftlicher Studiengänge genauer beschreiben zu können. Dadurch ist es möglich, die „atomaren Bestandteile“ prozeduralen Wissens zu identifizieren und zu benennen. Im weiteren Verlauf der Arbeit werden diese atomaren Bestandteile als

[1] Für einen historischen Überblick über diese Konstruktpaare siehe Haapasalo und Kadijevich (2000).

prädiktiv isolierbare Grundanforderung in Mathematikklausuren (ingenieurwissenschaftlicher Studiengänge), deren Anforderungsprofil insgesamt weit über diese Grundanforderungen hinausreicht, eingehend untersucht. Diesbezüglich wurde in der Einleitung bereits ausgeführt, dass bisher keine entsprechenden Untersuchungen über die Prognosegüte solcher fachspezifischen Einzelprädiktoren existieren. Die vorliegende Arbeit schließt daher auf Grundlage einer differenzierteren Beschreibung der Anforderungen an prozedurales Wissen in Klausuren diese Forschungslücke.
Für prozedurales und konzeptuelles Wissen existiert mit dem Standardwerk von Hiebert (1986) ein solider theoretischer Bezugspunkt, der allerdings insbesondere sowohl für die Beschreibung der unterschiedlichen Anforderungen an prozedurales Wissen in Klausuren als auch für die Isolierung eines fachspezifischen Prädiktors auf Grundlage dieser Anforderungen weiter ausdifferenziert werden muss. Dies geschieht in Abschnitt 1.1.1, der den theoretischen Rahmen für prozedurales Wissen aufspannt. Der anschließende Abschnitt 1.1.2 stellt analog den theoretischen Rahmen für konzeptuelles Wissen bereit.
In der Literatur wird kontrovers darüber diskutiert, ob prozedurales und konzeptuelles Wissen unabhängig voneinander zu sehen sind (Star, 2005; Kieran, 2013). In Abschnitt 1.1.3 wird daher der Zusammenhang zwischen den beiden Typen von Wissen aus der Hochschulperspektive betrachtet. Hier wird argumentiert, dass in der Ingenieurmathematik prozedurales und konzeptuelles Wissen über zahlreiche Verbindungen miteinander verknüpft sind, sodass die beiden Typen von Wissen in der Ingenieurmathematik nicht als unabhängig voneinander gelten können.
Im letzten Abschnitt 1.2 wird schließlich die Relevanz prozeduralen und konzeptuellen Wissens in der Ingenieurmathematik diskutiert. Dabei wird insbesondere die bereits in der Einleitung angedeutete These ausgeschärft, dass prozeduralem Wissen in der Ingenieurmathematik in den Bereichen Lehre, Prüfungen und in der Wahrnehmung von Mathematik bei Studierenden eine dominante Rolle zukommt.

1.1 Klärung der Begriffe

1.1.1 Theoretischer Rahmen für prozedurales Wissens

Es gibt verschiedene Definitionen für prozedurales Wissen. Diese sind vor allem mit Blick auf die Mathematik der Primar- und Sekundarstufe formuliert worden. Ziel des vorliegenden Abschnitts ist es, eine für den Hochschulbereich adäquate Definition zu finden. Die Definition sollte Charakteristika prozeduralen Wissens in der Ingenieurmathematik einerseits gut wiedergeben, andererseits sollten sich mit ihr typische Probleme von Studierenden im Bereich des prozeduralen Wissens erfassen lassen. Darüber hinaus sollte es möglich sein, zu der Definition passende Operationalisierungen für prozedurales Wissen zu finden. Um deutlich zu machen, wie existierende Definitionen dazu gegebenenfalls zu ergänzen sind, werden im Folgenden zwei Ansätze für Definitionen aus der Literatur vorgestellt und anhand realer Aufgaben- und Bearbeitungsbeispiele diskutiert. Anschließend werden die Definitionen so verfeinert und ergänzt, dass sie den oben formulierten Ansprüchen für die Ingenieurmathematik genügen.

Ansätze für die Definition prozeduralen Wissens

Star et al. (2015) definieren prozedurales Wissen wie folgt: „Procedural knowledge refers to having knowledge of action sequences for solving a problem (e.g., an algorithm for solving linear equations)“ (S. 45).
Eine alternative Definition von Rittle-Johnson und Schneider (2014) lautet: „Overall, there is a general consensus that procedural knowledge is the ability to execute action sequences (i.e. procedures) to solve problems“ (S. 5).
Ein wesentlicher Unterschied zwischen den Definitionen besteht darin, dass Star et al. (2015) prozedurales Wissen im Sinne von *Kenntnis von* (Algorithmen, Prozeduren) definieren, während die Definition von Rittle-Johnson und Schneider (2014) den praktischen Aspekt der *Fertigkeit* (Prozeduren auszuführen) betont.

Anhand der in Abbildung 1.1 dargestellen Aufgabe aus dem Speed- und Leistungstest prozedurales Wissen sowie der beiden prozeduralen Aufgaben aus der Mathematikklausur „Höhere Mathematik I“ für ingenieurwissenschaftliche Studiengänge aus dem Wintersemester 2013/14 (TU Dortmund) in Abbildung 1.2 sollen nun zwei Aspekte bezüglich dieser Definitionen im Detail betrachtet werden.

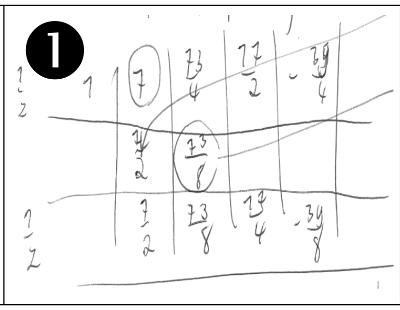
Der erste Aspekt betrifft die Beschreibbarkeit verschiedener Fehlerquellen bei der Bearbeitung prozeduraler Aufgaben auf Grundlage der angegebenen Definitionen. Eine Beschreibung der Fehlerquellen ist zum Beispiel wichtig, um Studierende mit Blick auf eine erfolgreiche Bearbeitung von Klausuraufgaben auf systematische Fehler aufmerksam machen zu können (vgl. Kersten, 2015; Prediger & Wittmann, 2009). In Abbildung 1.1 sind zwei Bearbeitungen von Studierenden zu sehen, die die Anwendung des Horner-Schemas im Rahmen einer Nullstellenberechnung zeigen. Legt man die oben angegebenen Definitionen prozeduralen Wissens zugrunde, so verfügen beide Studierenden nicht über ein ausreichendes prozedurales Wissen, um die Aufgabe zu lösen, denn beide Studierenden scheitern bei der Durchführung des Horner-Schemas. Nach Rittle-Johnson und Schneider (2014) verfügen sie im vorliegenden Fall nicht über die *Fertigkeit*, die Aufgabe erfolgreich zu bearbeiten. Gemäß Star et al. (2015) mangelt es ihnen hingegen an *Kenntnissen* der auszuführenden Regeln und Prozeduren, um zu einem korrekten Ergebnis zu gelangen. Bei genauerer Betrachtung der Bearbeitungen stellt sich allerdings heraus, dass die beiden

Das Polynom

$$P(z) = z^4 + 7z^3 + \frac{73}{4}z^2 + \frac{17}{2}z - \frac{39}{4}$$

hat die Nullstellen $z_1 = \frac{1}{2}$, $z_2 = -\frac{3}{2}$ und $z_3, z_4 \in \mathbb{C}$.

Bestimmen Sie z_3 und z_4 Mit Hilfe des Horner-Schemas und berechnen Sie $z_1 - z_2 - z_3 - z_3$.



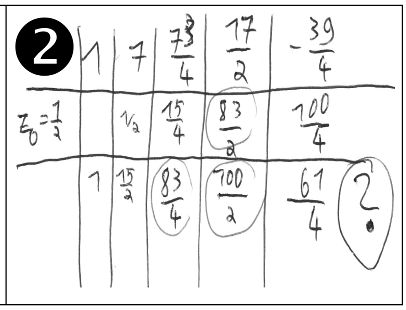


Abbildung 1.1 Aufgabe aus dem Speed- und Leistungstest prozedurales Wissen mit verschiedenen Bearbeitungsfehlern bei der Durchführung des Horner-Schemas zur Nullstellenbestimmung

Studierenden aus verschiedenen Gründen scheitern. Während bei der ersten Bearbeitung das Horner-Schema falsch verwendet wird und der Ablauf nicht internalisiert wurde, wird in der zweiten Bearbeitung lediglich ein Rechenfehler begangen, der natürlich nicht an das Horner-Schema gebunden ist. Im ersten Fall ist somit eine unzureichende *Kenntnis* des Horner-Schemas der Scheiterungsgrund, im zweiten Fall mangelnde *Fertigkeit* während der formal korrekten Ausführung innerhalb der zur Verfügung stehenden Zeit. Somit sind für die Analyse von Fehlerquellen bei der Bearbeitung prozeduraler Aufgaben *beide* Facetten relevant, d. h. sowohl die Kenntnis von Prozeduren als auch die Fertigkeit, Prozeduren erfolgreich innerhalb einer vorgegebenen Zeit ausführen zu können. Eine Verfeinerung der Definition prozeduralen Wissens im Hinblick auf eine Integration von Kenntnis *und* Fertigkeit in die Definition ist daher einer von zwei Aspekten, die im nachfolgenden Abschnitt aufgegriffen werden.

Der zweite aufzugreifende Aspekt betrifft die Beschreibung der *Komplexität* prozeduraler Aufgaben. Eine solche Beschreibung ist beispielsweise wichtig, um das eingangs erwähnte Anforderungsprofil einer Klausur skizzieren und prädiktiv isolierbare Grundanforderungen identifizieren zu können.

Aufgabe ❶: Gegeben sei die Matrix $A = \begin{pmatrix} 1 & 0 & 4 \\ 0 & -3 & 3 \\ 4 & 3 & 12 \end{pmatrix}$.

Berechnen Sie $\det(A)$.

Aufgabe ❷: Gegeben sei die Matrix $A = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix}$.

Berechnen Sie eine Orthonormalbasis des $\mathbb{R}^3$ aus Eigenvektoren von A.

Abbildung 1.2 Prozedurale Aufgaben aus einer Klausur für ingenieurwissenschaftliche Studiengänge

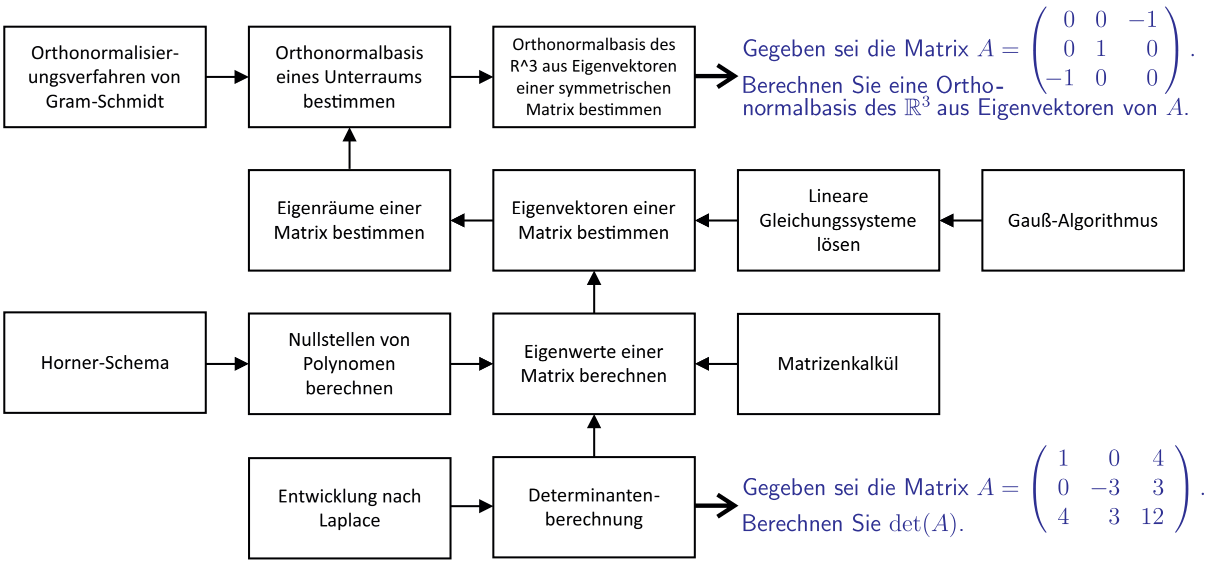


Abbildung 1.3 Netzwerk benötigter Prozeduren zur Bestimmung einer Determinante (Aufgabe 1) bzw. einer Orthonormalbasis des $\mathbb{R}^3$ aus Eigenvektoren einer reellen, symmetrischen 3×3-Matrix (Aufgabe 2)

In Abbildung 1.3 ist veranschaulicht, welche Abläufe zur Lösung der beiden Klausuraufgaben aus Abbildung 1.2 nötig sind. Pfeile, die zu einem Kasten führen, geben dabei an, welche Prozeduren zur Ausführung der im Kasten genannten Prozedur erforderlich sind. Demnach müssen zur Lösung von Aufgabe 1 wesentlich weniger Schritte ausgeführt werden. Für die Determinantenberechnung muss nur eine Prozedur bekannt sein, nämlich die Entwicklung nach Laplace. Die Prozedur der Determinantenberechnung selbst ist auf dem Pfad zur Lösung von Aufgabe 2 relevant. Folglich bilden die zur Lösung von Aufgabe 1 benötigten Prozeduren einen Teil der zur Lösung von Aufgabe 2 benötigten Prozeduren. Aufgabe 2 ist damit wesentlich komplexer als Aufgabe 1. Beiden Aufgaben gemeinsam ist allerdings, dass sie allein durch prozedurales Wissen gemäß der oben angegebenen Definitionen von Star et al. (2015) und Rittle-Johnson und Schneider (2014) zu lösen sind. Bei beiden Autorenteams zählen die durchzuführenden Prozeduren *Determinantenberechnung* (Aufgabe 1) und *Orthonormalbasis des* $\mathbb{R}^3$ *aus Eigenvektoren einer symmetrischen Matrix bestimmen* (Aufgabe 2) zu „action sequences". Um aber direkt mithilfe der Definitionen deutlich zu machen, dass sich die beiden eben genannten Prozeduren in ihrer Komplexität unterscheiden und Prozeduren hierarchisch geschachtelt sein können, um also im Hinblick auf ein Anforderungsprofil für Klausuraufgaben die *Feinstruktur* prozeduralen Wissens adäquat beschreiben zu können, ist eine Differenzierung erforderlich, die Abstufungen in der Komplexität von „action sequences" ermöglicht. Eine entsprechende Verfeinerung der Definition ist der zweite Aspekt, der im nächsten Abschnitt aufgegriffen wird.

Definition prozeduralen Wissens für die Ingenieurmathematik

Zur Verfeinerung der Definition prozeduralen Wissens sind einige Vorbereitungen erforderlich. Hierbei geht es, wie im letzten Abschnitt angedeutet, im Wesentlichen um eine Differenzierung des Begriffs *Prozedur*, wodurch sich anschließend bereits Unterschiede in der Komplexität prozeduraler Aufgaben beschreiben lassen. Zum Begriff Prozedur bemerken Hiebert und Lefevre (1986):

> „A key feature of procedures is that they are executed in a predetermined linear sequence ... The only relational requirement for a procedure to run is that prescription n must know that it comes after prescription $n-1$“ (Hiebert & Lefevre, 1986, S. 6).

Sie definieren Prozeduren wie folgt:

Definition (Prozedur): *„Eine Prozedur ist eine Schritt-für-Schritt-Anweisung, die vorschreibt, wie eine Aufgabe zu lösen ist“ (Hiebert und Lefevre, 1986, Übersetzung MA).*

Aus der Darstellung der Prozeduren zur Berechnung der Determinante einer Matrix und zur Berechnung einer Orthonormalbasis des $\mathbb{R}^3$ aus Eigenvektoren einer symmetrischen Matrix in Abbildung 1.3 geht hervor, dass in der Ingenieurmathematik Prozeduren eine sehr unterschiedliche Komplexität haben und ineinander geschachtelt sein können. Hiebert und Lefevre (1986) formulieren Letzteres wie folgt:

> „An important feature of the procedural system is that it is structured. Procedures are hierarchically arranged so that some procedures are embedded in others as subprocedures. An entire sequence of step-by-step prescriptions or subprocedures can be characterized as a superprocedure“ (Hiebert & Lefevre, 1986, S. 7).

Hiebert und Lefevre (1986) unterscheiden bei Prozeduren also zwischen Superprozeduren und Subprozeduren. Eine vertiefte Charakterisierung von Prozeduren findet sich bei Lithner (2006) bzw. Bergqvist (2007): Lithner führt in seinem Kategorisierungsschema zur Beschreibung der Anforderungen an Mathematikaufgaben „algorithmic reasoning“ als Kategorie ein. Dieses umfasst nach Lithner (2006) den Einsatz einer Menge von Algorithmen und Regeln in einer durch die Aufgabe(nstellung) implizierten Weise und entspricht damit dem Ausführen einer Prozedur bei Hiebert und Lefevre (1986). Darauf aufbauend adaptiert und verfeinert Bergqvist (2007) Lithners Schema speziell zur Typisierung von Mathematikaufgaben im Hochschulbereich. Insbesondere untergliedert sie „algorithmic reasoning“ in vier weitere Typen (vgl. Abbildung 1.4): (1) Den ersten Untertyp stellen „basic algorithms“ dar. Bergqvist (2007) definiert diese wie folgt:

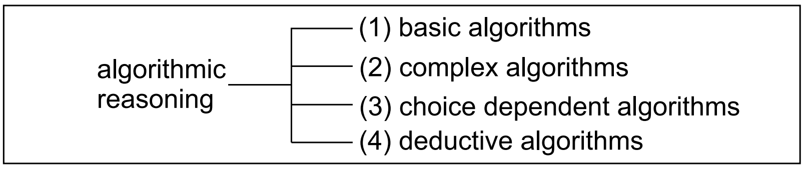


Abbildung 1.4 Unterteilung von „algorithmic reasoning“ als Anforderung an Mathematikaufgaben (Bergqvist, 2007, S. 359)

> „The basic algorithms are very direct, uncomplicated, and usually short algorithms. They are also often used as subalgorithms of longer solutions of other tasks. Some typical members of this subcategory are the differentiation algorithm, the algorithm for solving a second-degree equation, and the algorithms for solving differential equations of standard types“ (Bergqvist, 2007, S. 359).

(2) Den zweiten Untertyp bilden „complex algorithms“, die Bergqvist (2007) wie folgt beschreibt:

> „The tasks solvable with complex algorithms have solutions that are longer and more complicated than the basic algorithms. Each solution is based on a global algorithm familiar to the students, but contains also one or several subalgorithms. The separate subalgorithms are trivial for the students, or familiar in themselves. A subalgorithm to a complex algorithm is typically a basic algorithm“ (Bergqvist, 2007, S. 359).

Als Beispiel für einen „complex algorithm“ führt Bergqvist die Diskussion einer Kurve an. (3) „[C]hoice dependent algorithms“ erfordern zum Lösen der Aufgabe eine adäquate Strategiewahl, wobei die zur Wahl stehenden Strategien selbst wieder bekannte Algorithmen darstellen. Ein Beispiel ist die Grenzwertberechnung einer Funktion in einem Punkt, bei der die weitere Vorgehensweise von der Funktion und dem Punkt abhängt. (4) Zu „deductive algorithms“ zählen Beweise, die in routinierter Form ablaufen. Ein Beispiel ist der Nachweis der Stetigkeit einer Funktion in einem Punkt ihres Definitionsbereichs. Dabei können (3) und (4) als Spezialfälle von (2) aufgefasst werden, die lediglich inhaltlich weiter ausdifferenziert sind.

Auf der Grundlage von Hiebert und Lefevre (1986) und Bergqvist (2007) kann nun eine Unterscheidung in Basis- und Superprozeduren eingeführt werden, die angelehnt ist an Bergqvists basic algorithms bzw. complex algorithms.

Basisprozeduren sind in der Regel kurze und direkte Prozeduren. **Superprozeduren** sind aus mehreren Basisprozeduren zusammengesetzt, sie sind daher in der Regel länger und komplexer als Basisprozeduren.

Basis- und Superprozeduren sind immer nur relative Unterscheidungen, denn jede Basisprozedur lässt sich theoretisch bis aufs Zählen und die Grundrechenarten zerlegen. Im Kontext dieser Arbeit wird die Grenze zwischen Basis- und Superprozeduren wie folgt angelegt: Eine Basisprozedur ist eine in der Ingenieurmathematik eingeführte Prozedur, in die im Wesentlichen Schulwissen einfließt. Superprozeduren hingegen sind aus Basisprozeduren zusammengesetzt. Die Übergänge zwischen Basis- und Superprozeduren sind trotz dieser Abgrenzung fließend. Einige Beispiele sind in Abbildung 1.5 angegeben. Bereits in dem dort gezeigten Ausschnitt an Prozeduren wird sichtbar, dass diese in Form einer „ist enthalten in"-Beziehung miteinander verbunden sind. In der Abbildung ist diese Verbindung durch Pfeile gekennzeichnet.

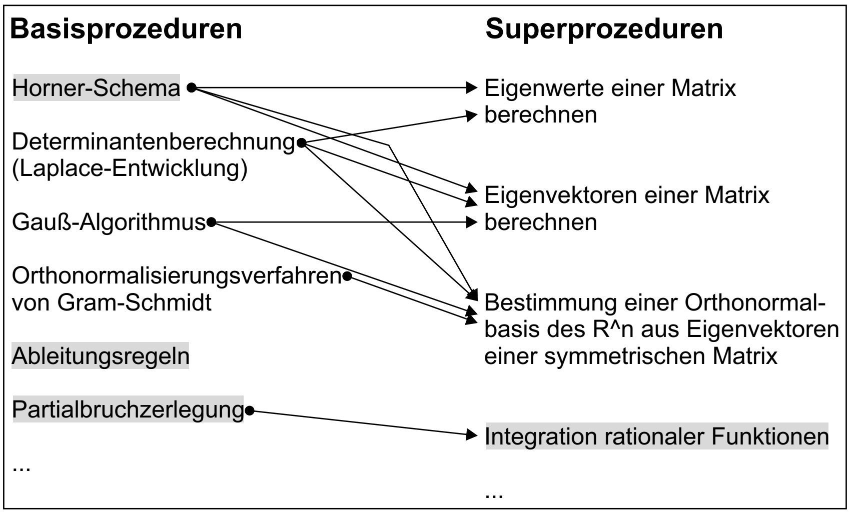


Abbildung 1.5 Einige Basis- und Superprozeduren in der Ingenieurmathematik aus den Gebieten der Analysis (grau hinterlegt) und linearen Algebra. Die Pfeile geben an, welche Prozeduren zur Ausführung von Superprozeduren erforderlich sind.

Die Menge aller Basis- und Superprozeduren mit allen Verknüpfungen, auf die eine Person in einer spezifischen Situation (z. B. während einer Klausur) zugreift, wird in der vorliegenden Arbeit in Anlehnung an Marshall (1980, S. 130) und Sacerdoti (1975, S. 10) als prozedurales Netzwerk dieser Person bezeichnet:

Definition (prozedurales Netzwerk einer Person in einer spezifischen Situation): *Die Gesamtheit aller Prozeduren, d. h. alle Basis- und Superprozeduren, und ihrer Verknüpfungen, auf die eine Person in einer spezifischen Situation zugreift, heißt prozedurales Netzwerk dieser Person in dieser Situation.*

Wie sich bereits bei den beiden diskutierten prozeduralen Aufgaben in Abbildung 1.2 angedeutet hat, gibt es in der Hochschulmathematik ein breites Spektrum sowohl an Basis- als auch an Superprozeduren. Beide gehören zur übergeordneten Kategorie der Prozeduren, auf deren Grundlage prozedurales Wissen im Folgenden definiert wird.

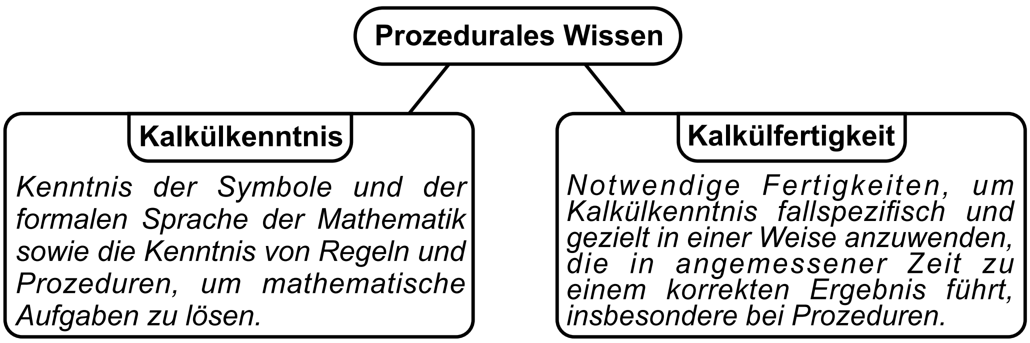
Dazu seien nochmals die Definitionen prozeduralen Wissens von Star et al. (2015) und Rittle-Johnson und Schneider (2014) betrachtet. Ein wesentlicher Unterschied in den beiden Definitionen besteht darin, dass Star et al. (2015) auf die *Kenntnis* von Prozeduren etc. fokussieren, während Rittle-Johnson und Schneider die *Fertigkeit* betonen, Prozeduren ausführen zu können. Die Diskussion der Bearbeitungsbeispiele aus Abbildung 1.1 hat gezeigt, dass beides, Kenntnis von Prozeduren und Fertigkeit bei der Ausführung von Prozeduren, zur Analyse von Fehlerquellen wichtig ist. Um diese Differenzierung innerhalb des prozeduralen Wissens zu ermöglichen, dienen die nachfolgenden Definitionen.

Definition (Kalkülkenntnis) *in Anlehnung an Hiebert und Lefevre (1986, S. 6): Kalkülkenntnis bezeichnet die Kenntnis der Symbole und der formalen Sprache der Mathematik sowie die Kenntnis von Regeln und Prozeduren, um mathematische Aufgaben zu lösen.*

Definition (Kalkülfertigkeit)*: Kalkülfertigkeit bezeichnet die notwendigen Fertigkeiten, um Kalkülkenntnis fallspezifisch und gezielt in einer Weise anzuwenden, die in angemessener Zeit zu einem korrekten Ergebnis führt, insbesondere bei Prozeduren.*

An die vorangegangenen Definitionen anknüpfend ergibt sich die folgende Definition von prozeduralem Wissen für die Ingenieurmathematik:

Definition (prozedurales Wissen)*: Prozedurales Wissen ist die Verknüpfung von Kalkülkenntnis und Kalkülfertigkeit.*



Kalkülkenntnis bezieht sich jeweils auf konkrete Themenbereiche und spezifische Prozeduren, während Kalkülfertigkeit meist vielfältige themenübergreifende Fertigkeiten umfasst und insbesondere auch elementarmathematische Rechenfertigkeiten (zur Bruchrechnung oder Termumformung) einschließt. Eine Schwäche in der Kalkülkenntnis würde sich etwa in systematischen Fehlern durch fehlende oder fehlerhafte Prozedurschritte zeigen, eine Schwäche in der Kalkülfertigkeit zum Beispiel in Vorzeichenfehlern, aber auch in einem zu geringen Arbeitstempo bei der Durchführung der Prozedur.

Prozedurales Wissen beinhaltet damit sowohl die Kenntnis von Basis- und Superprozeduren als auch die Fertigkeit, diese so ausführen zu können, dass in angemessener Zeit ein korrektes Ergebnis resultiert.

Fazit

In Abbildung 1.6 sind die in diesem Abschnitt eingeführten Begriffe nochmals zusammengefasst. Mit der Untergliederung von Prozeduren in Basis- und Superprozeduren kann der Unterschied zwischen den beiden prozeduralen Klausuraufgaben in Abbildung 1.2 nun benannt werden. Bei Aufgabe 1 handelt es sich um eine Basisprozedur, während Aufgabe 2 eine Superprozedur darstellt, die aus mehreren Super- und Basisprozeduren besteht. Aufgabe 2 ist folglich eine wesentlich komplexere prozedurale Klausuraufgabe als Aufgabe 1. Mit der Unterscheidung zwischen Basis- und Superpozeduren können nun klausurrelevante Charakteristika prozeduralen Wissens in Form von unterschiedlichen Anforderungen bezüglich der Komplexität von Prozeduren der Ingenieurmathematik gut wiedergegeben werden. Damit ist die erste der eingangs formulierten Forderungen an eine Definition prozeduralen Wissens erfüllt. Die Kenntnis von Basis- und Superprozeduren gehört somit zum Anforderungsprofil der Klausur, aus der die beiden diskutierten Aufgaben stammen. Hierbei sind Basisprozeduren als die „atomaren Bestandteile“ prozeduralen Wissens anzusehen, die im weiteren Verlauf der Arbeit als prädiktiv isolierbare Grundanforderungen in Klausuren der Ingenieurmathematik detaillierter betrachtet werden. Dabei wird untersucht, inwiefern diese Grundanforderungen als fachspezifischer Prädiktor geeignet ist, um den Erfolg in Klausuren, deren Anforderungsprofil weit über die Beherrschung von Basisprozeduren hinausreicht, vorauszusagen. Im Methodenteil wird daher das korrekte Ausführen von Basisprozeduren in angemessener Zeit als Operationalisierung prozeduralen Wissens eingeführt. Mit dieser zur eingeführten Definition passenden Operationalisierung prozeduralen Wissens ist eine weitere der eingangs gestellten Forderungen erfüllt.

<table>
<tr><td>Prozedur
Definition: Schritt-für-Schritt-Anweisungen, die vorschreiben, wie eine Aufgabe zu lösen ist.</td><td colspan="2">Prozedurales Wissen
setzt sich zusammen aus:</td></tr>
<tr><td>Basisprozeduren und Superprozeduren
Basisprozeduren sind in der Regel kurze und direkte Prozeduren. Superprozeduren sind aus mehreren Basisprozeduren zusammengesetzt, sie sind daher in der Regel länger und komplexer als Basisprozeduren.</td><td rowspan="2">Kalkülkenntnis
Definition:
Kenntnis der Symbole und der formalen Sprache der Mathematik sowie die Kenntnis von Regeln und Prozeduren, um mathematische Aufgaben zu lösen.</td><td rowspan="2">Kalkülfertigkeit
Definition:
Notwendige Fertigkeiten, um Kalkülkenntnis fallspezifisch und gezielt in einer Weise anzuwenden, die in angemessener Zeit zu einem korrekten Ergebnis führt, insbesondere bei Prozeduren.</td></tr>
<tr><td>Prozedurales Netzwerk (einer Person in einer spezifischen Situation)
Definition: Die Gesamtheit aller Prozeduren, d. h. alle Basis- und Superprozeduren, und ihrer Verknüpfungen, auf die eine Person in einer spezifischen Situation zugreift, heißt prozedurales Netzwerk dieser Person in dieser Situation.</td></tr>
</table>

Abbildung 1.6 Überblick über die im vorliegenden Abschnitt eingeführten Begriffe

Rückblickend auf die Bearbeitungsbeispiele in Abbildung 1.1 gelingt nun auch eine genauere Charakterisierung der Bearbeitungsfehler. In der ersten Bearbeitung wird die Basisprozedur „Horner-Schema" falsch verwendet, sodass eine Schwäche in der Kalkülkenntnis zum Scheitern führt. In der zweiten Bearbeitung wird die Prozedur formal korrekt ausgeführt, d. h. die Kalkülkenntnis ist vorhanden. Der Rechenfehler ist der Kalkülfertigkeit zuzuordnen. Somit können typische Probleme von Studierenden bei der Anwendung prozeduralen Wissens benannt werden, womit auch die letzte der eingangs formulierten Forderungen an eine Definition prozeduralen Wissens für die Ingenieurmathematik erfüllt ist.

1.1.2 Theoretischer Rahmen für konzeptuelles Wissens

Prozedurales Wissen ist mit konzeptuellem Wissen eng verknüpft, da letzteres unter anderem die Begründungen liefert, warum Prozeduren funktionieren (Baroody, Feil & Johnson, 2007; Hiebert & Lefevre, 1986, S. 10). Im nächsten Abschnitt zeigt sich darüber hinaus, dass gerade im Hochschulbereich die Verbindungen zwischen konzeptuellem und prozeduralem Wissen noch vielfältiger sind und diese Verbindungen zur Festigung prozeduralen Wissens beitragen. Zudem wird in der vorliegenden Arbeit auch konzeptuelles Wissen als Prädiktor für Klausurerfolg und -leistung untersucht. Aus diesen Gründen ist die Thematisierung konzeptuellen Wissens an dieser Stelle wichtig.

Als theoretischer Bezugspunkt zur Definition von konzeptuellem Wissen dient hier erneut das Standardwerk von Hiebert (1986). Hierin verstehen Hiebert und Lefevre (1986) unter konzeptuellem Wissen ein Netzwerk von Basiskonzepten: „Conceptual knowledge is characterized most clearly as knowledge that is rich in relationships. It can be thought of as a connected web of knowledge, a network in which the linking relationships are as prominent as the discrete pieces of information" (S. 4). Basiskonzepte beschrieben die Autoren vor 30 Jahren noch im Paradigma der Informationsverarbeitung als „discrete pieces of information". Heute werden sie breiter verstanden als die kleinsten mathematischen Inhaltselemente und ihre Bedeutungsgehalte, zum Beispiel ein Konzept wie Funktion oder eine Operation wie Multiplikation oder Verkettung sowie ihre Bedeutungen (vom Hofe, 1992). Auch nach 30 Jahren noch aktuell ist Hiebert und Lefevres Modell, als relevante größere Einheiten konzeptuellen Wissens („units of conceptual knowledge") die Verknüpfung von mehreren Basiskonzepten zu betrachten (vgl. Abbildung 1.7). Analog zu den Superprozeduren können diese als Superkonzepte bezeichnet werden. Hiebert und Lefevre (1986) verwenden viele Begriffe für konzeptuelles Wissen synonym[2] In der vorliegenden Arbeit wird hierfür jedoch nur der Begriff konzeptuelles Netzwerk („conceptual network") verwendet, was zu

[2] z. B. „connected web of knowledge" (S. 3), „conceptual network" (S. 9), „underlying rationale on which they [procedures] are based" (S. 10), „conceptual underpinning [of procedures]" (S. 11), „conceptual knowledge base" (S. 17).

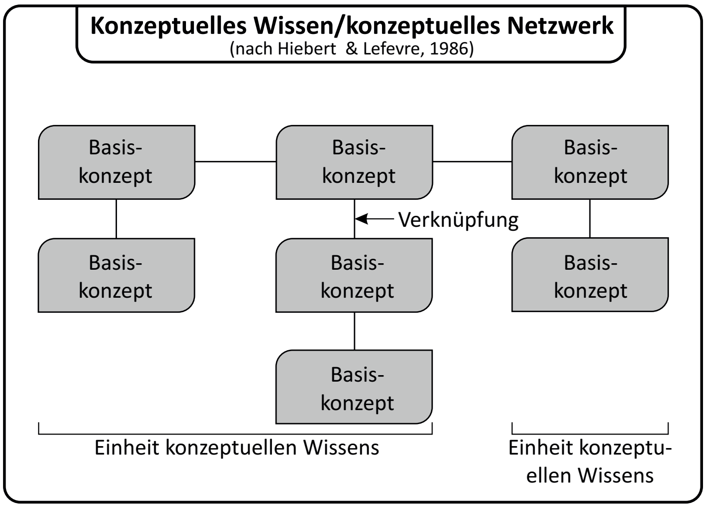


Abbildung 1.7 Konzeptuelles Wissen nach Hiebert und Lefevre (1986)

folgender Definition für konzeptuelles Wissen bzw. konzeptuelles Netzwerk einer Person in einer spezifischen Situation führt:

Definition (konzeptuelles Wissen/konzeptuelles Netzwerk einer Person in einer spezifischen Situation): *Unter konzeptuellem Wissen einer Person in einer spezifischen Situation sind alle Basiskonzepte mit allen Verknüpfungen zwischen diesen Basiskonzepten zu verstehen, auf die die Person in dieser Situation zugreift.*
Synonym zu konzeptuellem Wissen wird hier auch der Begriff konzeptuelles Netzwerk (einer Person in einer spezifischen Situation) verwendet.

Im konzeptuellen Netzwerk entsprechen Basiskonzepte den Basisprozeduren und die Einheiten konzeptuellen Wissens den Superprozeduren des prozeduralen Netzwerks. Im Gegensatz zum prozeduralen Netzwerk können im konzeptuellen Netzwerk die Verbindungen auch ungerichtet sein, d. h. es ist nicht immer vorgeschrieben, in welcher Reihenfolge Basiskonzepte miteinander zu verknüpfen sind, um eine Aufgabe zu lösen. Hiebert und Lefevre (1986) formulieren diesen Sachverhalt wie folgt:

> „Perhaps the biggest difference between procedural knowledge and conceptual knowledge is that the primary relationship in procedural knowledge is 'after,' which is used to sequence subprocedures and superprocedures linearly. In contrast, conceptual knowledge is saturated with relationships of many kinds“ (Hiebert & Lefevre, 1986, S. 8).

Sowohl das prozedurale als auch das konzeptuelle Netzwerk verfügen für sich betrachtet über zahlreiche Verknüpfungen. Über diese internen Verknüpfungen hinaus sind aber auch beide Netzwerke miteinander verbunden, was im nachfolgenden Abschnitt näher beschrieben wird.

1.1.3 Zusammenhänge zwischen konzeptuellem und prozeduralem Wissen

Nach Hiebert und Lefevre (1986, S. 10) stellt konzeptuelles Wissen die konzeptuelle Basis von prozeduralem Wissen dar. Die gerichteten Verbindungen innerhalb des prozeduralen Netzwerkes werden über die bereits zu Anfang des Kapitels angedeuteten Verbindungen zum konzeptuellen Netzwerk ebenso erklärt wie das Funktionieren der Prozeduren selbst. Die Verbindungen zum konzeptuellen Netzwerk stellen das „knowing why“ (Mason & Spence, 1999) dar, während das prozedurale Netzwerk das „knowing how“ (Scheffler, 1965) umfasst.

Für Aufgabe 2 in Abbildung 1.2 lässt sich dies wie in Abbildung 1.8 aufgezeigt veranschaulichen. Zu sehen sind die beteiligten Basis- und Superprozeduren, die im Rahmen der übergeordneten Superprozedur zur Lösung der Aufgabe ausgeführt werden müssen. Die gerichteten Kanten geben die Reihenfolge an, in der die beteiligten Prozeduren auszuführen sind. Die ungerichteten Kanten stellen die Verbindungen zum konzeptuellen Netzwerk dar. Dabei handelt es sich entweder um Verbindungen, die von einer Prozedur ausgehen, oder um Verbindungen, die von einer gerichteten Kante ausgehen. Erstere erklären mithilfe konzeptuellen Wissens, warum die jeweilige Prozedur funktioniert. Letztere rechtfertigen mithilfe konzeptuellen Wissens die gerichtete Verbindung zwischen zwei Prozeduren. Exemplarisch sei dies an den in Abbildung 1.2 dicker markierten Verbindungen erläutert: Um zu verstehen, *warum* zur Eigenwertberechnung sowohl die Berechnung einer Determinante als auch die Berechnung von Nullstellen eines Polynoms erforderlich sind, müssen innerhalb des konzeptuellen Netzwerkes folgende Basiskonzepte miteinander verknüpft werden:

1. Eigenwerte (der symmetrischen Matrix $A \in \mathbb{R}^{3\times 3}$) sind alle $\lambda \in \mathbb{R}$, für die ein $\vec{v} \in \mathbb{R}^3 \setminus \left\{\vec{0}\right\}$ existiert, sodass $A\vec{v} = \lambda\vec{v}$.

2. Obige Gleichung lässt sich wie folgt umformen: $A\vec{v} = \lambda\vec{v} \Leftrightarrow (A - \lambda I)\vec{v} = \vec{0}$. Gesucht sind also Lösungen eines homogenen linearen Gleichungssystems, die vom Nullvektor verschieden sind. Solche Lösungen existieren nur dann, wenn das lineare Gleichungssystem nicht eindeutig lösbar ist. Dies ist genau dann der Fall, falls $\det(A - \lambda I) = 0$ gilt. Daher ist zur Eigenwertberechnung die Bestimmung einer Determinante erforderlich.

3. Der Ausdruck $\det(A - \lambda I)$ ist ein Polynom vom Grad 3 in λ. Die Nullstellen sind die Eigenwerte von A. Daher ist zur Eigenwertberechnung die Bestimmung von Nullstellen eines Polynoms erforderlich.

Die Durchführung einer Prozedur ist grundsätzlich ohne Verbindungen zum konzeptuellen Netzwerk möglich. White und Mitchelmore (1996) unterscheiden in diesem Zusammen-

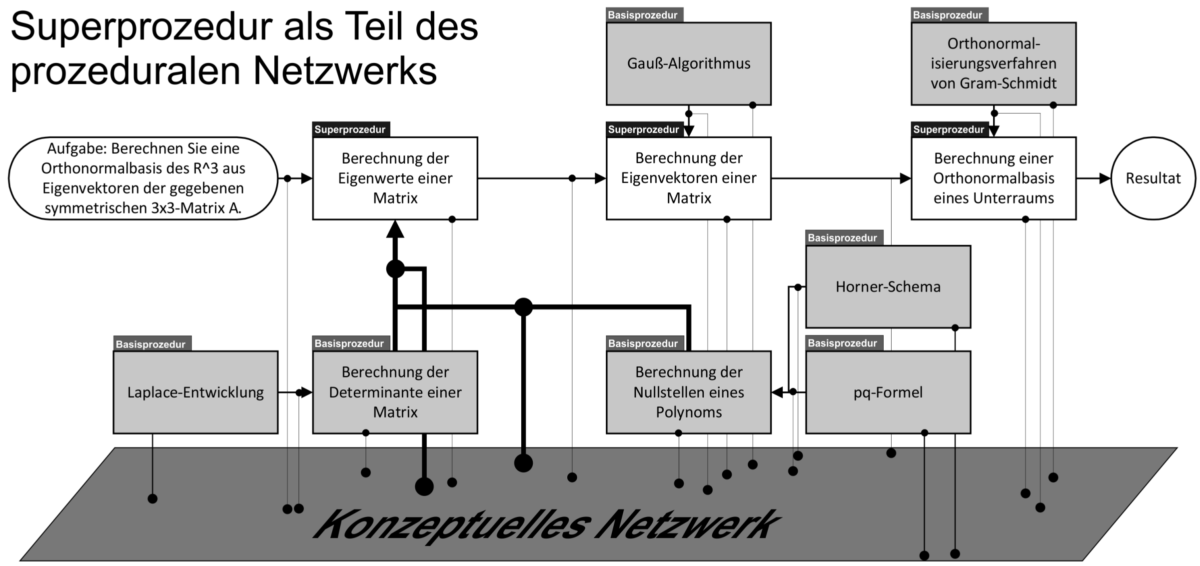


Abbildung 1.8 Superpozedur und ihre Verbindungen zum konzeptuellen Netzwerk zur Bestimmung einer Orthornomalbasis des $\mathbb{R}^3$ aus Eigenvektoren einer symmetrischen Matrix. Basisprozeduren sind zur besseren Unterscheidbarkeit grau hinterlegt.

hang zwischen unterstütztem („supported“) und nicht unterstütztem („unsupported“) prozeduralen Wissen: „Procedural knowledge may or may not be supported by conceptual knowledge. Unsupported procedural knowledge is similar to Skemp's (1976) 'instrumental understanding,' which he describes as knowing rules without knowing why they work“ (S. 81). Dabei gilt unterstütztes prozedurales Wissen als wesentlich stabiler und nachhaltiger: „Studies showed that procedural knowledge is the one that faster deteriorates with time (e.g. Allen et al., 2005). If there is no conceptual meaning, this kind of knowledge is very fragile in the long term memory and usually remembered inappropriately“ (Matic, 2014, S. 20). Das konzeptuelle Netzwerk dient also nicht nur zur Begründung prozeduralen Wissens, sondern auch der gestützten Navigation in Superprozeduren, sodass diese nachhaltiger beherrscht werden. Dieser Sachverhalt erscheint umso relevanter, je mehr Verbindungen es zwischen einer Prozedur und dem konzeptuellen Netzwerk gibt. Da die Verbindungen in der Hochschulmathematik oft zahlreich sind (vgl. Abbildung 1.8), erscheint die Verbindung von prozeduralem mit konzeptuellem Wissen gerade hier wichtig, um prozedurales Wissen zu festigen (vgl. hierzu auch die Arbeit von Frerich & Pitsch, 2016).

Für die Ingenieurmathematik wäre es daher interessant zu untersuchen, ob sich eine bessere Verbindung von prozeduralem und konzeptuellem Wissen in Form einer starken Verankerung prozeduralen Wissens im konzeptuellen Netzwerk positiv auf die Facetten prozeduralen Wissens, d. h. auf Kalkülkenntnis und Kalkülfertigkeit, sowohl bei Basisprozeduren als auch bei Superprozeduren auswirkt. Eine entsprechende Analyse liegt allerdings außerhalb des Rahmens der vorliegenden Arbeit und bleibt nachfolgenden Forschungsvorhaben vorbehalten.

Die vorliegende Arbeit wird sich zunächst auf die Relevanz der prozeduralen Atome, der Basisprozeduren, beschränken und argumentieren, dass ihre Beherrschung (insbesondere in Bezug auf Kalkülkenntnis) stark prädiktiv für die Bewältigung der komplexen Anforderungen in Klausuraufgaben ist.

1.2 Relevanz prozeduralen und konzeptuellen Wissens im Hochschulbereich

Theoretisch ist also plausibel herzuleiten, dass prozedurales und konzeptuelles Wissen in engem Zusammenhang stehen und beide auch für die Ingenieurmathematik relevant sind. Gleichwohl ergibt sich in verschiedenen Bereichen jedoch ein unterschiedliches Ungleichgewicht, das im Folgenden kurz vorgestellt werden soll. Die Bereiche betreffen die hochschuldidaktische Literatur zur Gestaltung der Ingenieurmathematik (Abschnitt 1.2.1), Anforderungen in Mathematikklausuren (Abschnitt 1.2.2) und die Wahrnehmung von Mathematik durch Studierende der ingenieurwissenschaftlichen Studiengänge (Abschnitt 1.2.3).

1.2.1 Prozedurales und konzeptuelles Wissens in der Mathematikausbildung ingenieurwissenschaftlicher Studiengänge

Ein wesentlicher Diskussionspunkt der in den 1980er-Jahren beginnenden „math wars“ (Schoenfeld, 2004) war die Frage nach dem Verhältnis, in dem prozedurales und konzeptuelles Wissen vermittelt werden sollen. Inzwischen sind die „math wars“ weitgehend eingestellt, denn es scheint ein Kompromiss gefunden, den sowohl die Vertreterinnen und Vertreter eines eher konzeptorientierten als auch die Befürworterinnen und Befürworter eines traditionellen, d. h. eher prozedural ausgerichteten, Unterrichts akzeptieren. Der Kompromiss besteht international in der Auffassung, dass Lehre prozedurales und konzeptuelles Wissens in einem ausgewogenen Verhältnis vermitteln sollte:

> „Balancing the acquisition of conceptual understanding and procedural proficiency is far from being strictly an American concern. Standards documents and curriculum frameworks from around the world, such as the Australian (Leonelli & Schmitt, 2001) and British frameworks (ATM, 2006) are quite pronounced in their calls for such a balance“ (Bosse & Bahr, 2008, S. 3).

Dennoch wird nach Aussagen vieler Autorinnen und Autoren Mathematik in ingenieurwissenschaftlichen Studiengängen weltweit auch heute noch überwiegend traditionell, d. h.

prozedural, unterrichtet (Engelbrecht et al., 2012; Binti, Zeynivandnezhad, Binti & David, 2014). Zerr (2009) fasst diese Erkenntnis wie folgt zusammen:

> „It has traditionally been the case that procedures have been emphasized in the teaching of mathematics [12], and my approach has often been no different. It is no surprise then that students tend to leave introductory mathematics courses with a narrow view of the subject, and consequently are lacking the full set of skills that they should have acquired“ (Zerr, 2009, S. 1).

Die Kritik an der fehlenden Balance war der Hintergrund für die Etablierung vieler Innovationsprojekte, die die hochschuldidaktische Aufmerksamkeit kompensatorisch auf die Seite des konzeptuellen Wissens gelegt haben (Chappell & Killpatrick, 2003; Bergsten, Engelbrecht & Kågesten, 2015; Huang, 2010; Zerr, 2009; Arslan, 2010; White & Mitchelmore, 1996; Grundmeier, Hansen & Sousa, 2006; Dreyfus & Eisenberg, 1990): „Most research tries to focus on conceptual knowledge, although currently the emphasis is on procedural knowledge“ (Binti et al., 2014, S. 164). Dadurch ist allerdings ein neues Ungleichgewicht zugunsten der Erforschung konzeptuellen Wissens im Hochschulbereich entstanden. Deshalb und aufgrund der nach wie vor prozedural ausgerichteten Lehre in der Ingenieurmathematik legt die vorliegende Arbeit einen Schwerpunkt auf die Untersuchung prozeduralen Wissens. Sie trägt damit zu einer Verbesserung der Balance in der Forschung über prozedurales und konzeptuelles Wissen in der Ingenieurmathematik bei.

1.2.2 Prozedurales und konzeptuelles Wissen in den Mathematikklausuren ingenieurwissenschaftlicher Studiengänge

In Abbildung 1.9 ist eine Mathematikklausur für Studienanfänger ingenieurwissenschaftlicher Studiengänge an der TU Dortmund aus dem Wintersemester 2013/14 zu sehen. Im rechten Teil der Abbildung sind einige Aufgaben vergrößert wiedergegeben. Im Folgenden werden exemplarisch einige der Anforderungen diskutiert, die zur Lösung der gezeigten Aufgaben erforderlich sind. Abschließend wird der Frage nachgegangen, ob einige dieser Anforderungen besonders häufig in Klausuren der Ingenieurmathematik anzutreffen sind.

In Aufgabe 1 wird ein Abruf von Faktenwissen verlangt. Die Aufgaben 2 und 4 wurden bereits in Abschnitt 1.1.1 diskutiert: Aufgabe 2 erfordert die Durchführung einer Basisprozedur und Aufgabe 4 die Durchführung einer Superprozedur. Zur Lösung der beiden Aufgaben ist ausschließlich prozedurales Wissen erforderlich. Bei den verbleibenden Aufgaben sind zur Lösung zahlreiche Verbindungen herzustellen. Prozedurales Wissen oder Faktenwissen reichen hier nicht aus; es ist ein Zugriff auf konzeptuelles Wissen erforderlich. Natürlich können zwar bei entsprechenden Vorkenntnissen auch diese Aufgaben

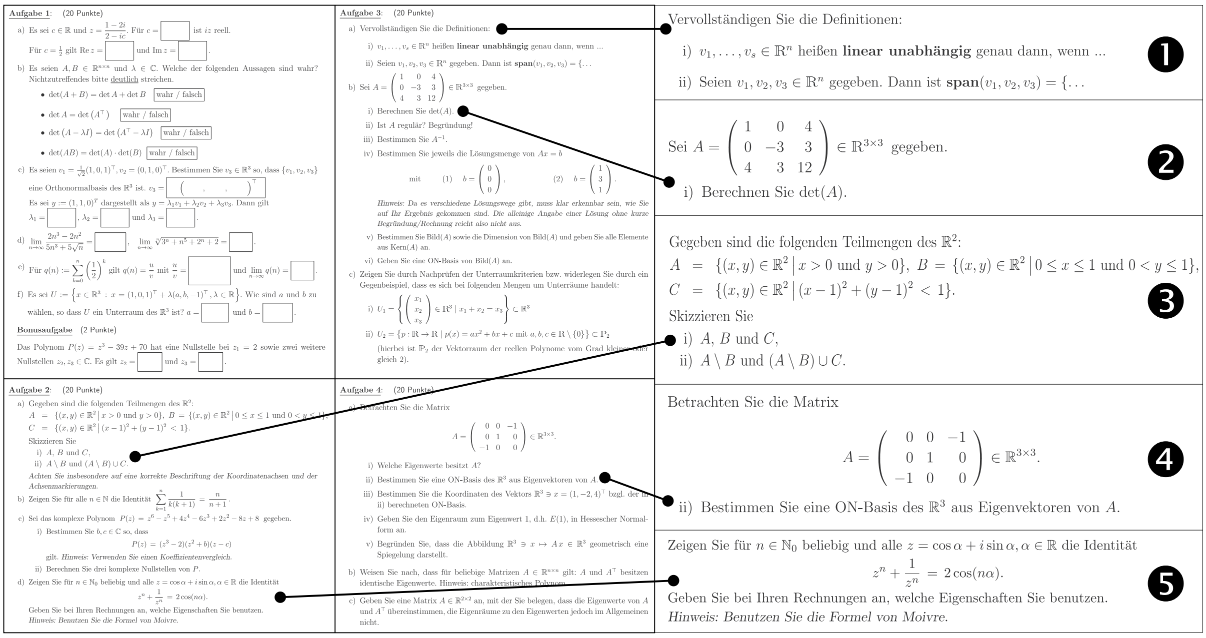


Abbildung 1.9 Klausur „Höhere Mathematik I“ aus dem Wintersemester 2013/14 (links) und ausgewählte Teilaufgaben

routinisiert gelöst werden, beispielsweise durch ein vorgeschaltetes „teaching to the test“-Arrangement. Nichtsdestotrotz sind zahlreiche Themen zu vernetzen, sodass die Definition konzeptuellen Wissens hier zutrifft.

Die vorliegende Klausur stellt damit neben einem Abruf von Faktenwissen sowohl Anforderungen an prozedurales als auch an konzeptuelles Wissen. Bezüglich der Gewichtung von prozeduralem und konzeptuellem Wissen in Mathematikklausuren für ingenieurwissenschaftliche Studiengänge gelangt Bergqvist (2007) in ihrer Analyse von 16 Klausuren des ersten Kurses in „Calculus“ an vier schwedischen Universitäten zu folgenden Ergebnissen: Knapp 70 % aller Aufgaben lassen sich allein durch den Abruf von Faktenwissen (17 %) oder „algorithmic reasoning“ (52 %) lösen. Die Klausuren sind also sehr prozedural und auf die Abfrage von Faktenwissen ausgerichtet. 15 von 16 Klausuren können weitgehend ohne konzeptuelles Wissen bestanden werden. In diesen Fällen ist es sogar möglich, fast ohne konzeptuelles Wissen die Bestnote zu erzielen. Zur Repräsentativität der Studie bemerkt Bergqvist (2007):

> „The result is not possible to completely generalize to all exams in mathematics at Swedish universities, but it is reasonable to believe that the results are fairly representative for the situation at calculus courses during the academic year in question“ (Bergqvist, 2007, S. 32).

Bergqvist (2007) stellt also fest, dass an schwedischen Hochschulen prozedurales Wissen im Anforderungsprofil von Klausuren zur Vorlesung „Calculus“ dominiert. In Deutschland gibt es noch keine vergleichbare Untersuchung. Daher wird die Vermutung, dass auch in Deutschland Mathematikklausuren ingenieurwissenschaftlicher Studiengänge vorrangig prozedural ausgerichtet sind, im Rahmen der vorliegenden Arbeit überprüft. Dabei wird auch die in Abbildung 1.9 dargestellte Klausur einbezogen, die Grundlage einiger Untersuchungen der vorliegenden Arbeit ist. Es wurde bereits festgestellt, dass diese Klausur unter anderem Anforderungen an prozedurales Wissen stellt. Ob dies in einem ähnlichen Umfang wie bei Bergqvist (2007) der Fall ist, werden die Untersuchungen in Abschnitt 5.6 zeigen.

Der vorliegende und vorangegangene Abschnitt hat gezeigt, dass viele Indizien für eine dominante Rolle prozeduralen Wissens sowohl in der Ausbildung als auch in Klausuren der Mathematik für ingenieurwissenschaftliche Studiengänge sprechen. Im nachfolgenden Abschnitt wird abschließend diskutiert, ob prozedurales Wissen auch in der Wahrnehmung von Mathematik bei Studierenden dominiert.

1.2.3 Prozedurales und konzeptuelles Wissens in der Wahrnehmung von Mathematik bei Studierenden

Untersuchungen aus verschiedenen Ländern zeigen, dass Studierende Mathematik (eher) als prozedural ansehen (Engelbrecht et al., 2012; Aspinwall & Miller, 1997; Zerr, 2009; Carpenter, Lindquist, Matthews & Silver, 1983; Schoenfeld, 1989). Hallett (2006) formuliert dies wie folgt: „In addition, many students came to college believing that mathematics centers on manipulative techniques, rather than interpretation and understanding“ (S. 1). Einige Autorinnen und Autoren vertreten die Ansicht, dass Studierende Konzepte als weniger wichtig erachten und ihr Lernen häufig auf das Auswendiglernen von Prozeduren reduzieren, mit denen sich Standardprobleme routinisiert lösen lassen (Mahir, 2009). Zerr (2009) konstatiert diesbezüglich: „In fact, I have found that students frequently come to view calculus as strictly procedural, and give little thought to the arguably more important conceptual ideas that they are also learning“ (S. 1).

Es ist davon auszugehen, dass es mindestens die folgenden drei möglichen Gründe gibt, warum Studierende Mathematik eher für prozedural halten und ihr Lernen entsprechend ausrichten:
Als erster möglicher Grund sind die institutionellen Rahmenbedingungen zu nennen. In den vorangegangenen Abschnitten wurde deutlich, dass sowohl Unterrichtspraxis als auch Klausuren in Mathematik für ingenieurwissenschaftliche Studiengänge prozedural ausgerichtet sein könnten. Beides hat Auswirkungen auf das Lernverhalten und die Einstellungen der Studierenden (Brown, Collins & Duguid, 1989; Hallett, 2006; Kane, Crooks & Cohen, 1999; Lithner, 2004). El Gaidi und Ekholm (2015) resümieren: „However, if both

teaching and examination of university mathematics is geared towards procedural knowledge, then the students cannot be blamed for not developing the conceptual knowledge hoped for“ (S. 8).

Ein zweiter möglicher Grund, warum bei Studierenden eine Fokussierung auf prozedurales Wissen überwiegen könnte, lautet: In der Hochschulmathematik sind die Themen komplex geschachtelt. Wer ein Thema nicht versteht, hat in der Regel Schwierigkeiten, das darauf aufbauende Folgethema zu verstehen:

> „In calculus, new concepts are built up by using previously established concepts. A new concept is understood only when it is related and reconciled with previous concepts [17]. Specifically, one cannot understand differentiation without knowing limit and one cannot understand integration without knowing differentiation“ (Mahir, 2009, S. 202).

Tall (1996) vermutet, dass Studierende das Bemühen um Verständnis einstellen, sobald Schlüsselthemen nicht verstanden werden. Stattdessen wird versucht, über ein Auswendiglernen von Prozeduren zum Prüfungserfolg zu gelangen. Gemäß den Ergebnissen von Bergqvist (2007) kann dies eine erfolgreiche Strategie für den kurzfristigen Klausurerfolg sein, die deshalb möglicherweise auch an die nachfolgende Studierendengeneration weitergegeben wird. Das kann dazu beitragen, die Dominanz prozeduralen Wissens in der Wahrnehmung von Mathematik bei Generationen von Studierenden aufrechtzuerhalten.

Einen dritten möglichen Grund schreiben einige Autorinnen und Autoren der Vorprägung aus der Schule zu (Engelbrecht et al., 2012; Ferrini-Mundy & Gaudard, 1992). Sie verweisen darauf, dass Schulunterricht und Aufgabenformat prozedural ausgerichtet sind (Foster, 2014; Boesen, 2006). Jordan et al. (2008) und Drüke-Noe (2014) bestätigen dies für Klausuren, die am Ende der Sekundarstufe I geschrieben werden:

> „Schon dieses globale empirische Ergebnis bestätigt die aus theoretischer Sicht diskutierte Kalkülorientierung von Klassenarbeitsaufgaben (vgl. Kap. 1.4) und es ist zu resümieren, dass der Anspruch an das Umgehen mit Kalkülen einen sehr wesentlichen Anteil am insgesamt in diesen Aufgaben abgebildeten kognitiven Anspruchsniveau hat“ (Drüke-Noe, 2014, S. 243).

Fazit

Die vorangegangenen Abschnitte haben aufgezeigt, dass nach Auffassung einiger Autorinnen und Autoren in den ingenieurwissenschaftlichen Studiengängen sowohl der Mathematikunterricht als auch Mathematikklausuren und die Wahrnehmung der Studierenden von Mathematik prozedural ausgerichtet sind. Aufgrund dieser perzipierten Dominanz des prozeduralen Wissens richten Studien mehrheitlich ihren Fokus kompensatorisch auf die

Förderung konzeptuellen Wissens. Dies mag für die Reform der Ingenieurmathematik eine relevante Strategie sein, doch hat die einseitige Fokussierung dazu geführt, dass über die Herausforderungen und Hintergründe im Bereich des prozeduralen Wissens bislang wenig differenzierte Erkenntnisse vorliegen. Zur Wiederherstellung des Gleichgewichts zwischen dem Forschungsfokus auf prozeduralem und demjenigen auf konzeptuellem Wissen rückt daher die vorliegende Arbeit die Untersuchung prozeduralen Wissens in den Vordergrund (siehe hierzu auch Altieri & Prediger, voraussichtlich 2016). Dafür lässt sich neben der Herstellung eines Gleichgewichts zwischen prozeduralem und konzeptuellem Wissen noch ein weiterer pragmatischer Grund sowie ein theoretischer Grund anführen: Aufgrund der in der Einleitung erwähnten hohen Durchfallquoten in Klausuren der Ingenieurmathematik ist es hochrelevant, die Gründe hierfür zu erforschen. Anhand der Klausur in Abbildung 1.9 wurde exemplarisch gezeigt, dass hier sowohl Basis- als auch Superprozeduren durchzuführen sind und somit hinsichtlich der Komplexität sehr unterschiedliche Anforderungen an prozedurales Wissen gestellt werden. Dies ist ein pragmatischer Grund, prozedurales Wissen im Hochschulbereich eingehender zu beforschen. Der theoretische Grund lautet wie bereits ausgeführt, dass prozedurales Wissen im Hochschulbereich wenig erforscht ist und sich daher zahlreiche Forschungsfragen an prozedurales Wissen stellen lassen. Aufgrund des mangelnden deskriptiven Forschungsstands, ohne den theoretisch begründete und empiriebasierte Innovationen nicht zu planen sind, beschränkt sich die Arbeit auf die Erfassung und Analyse des Status quo, während ein Eingreifen in den Veranstaltungsbetrieb darauf aufbauenden Folgeprojekten vorbehalten bleibt.

2 Theoretischer Rahmen: Allgemein kognitive und affektive Prädiktoren für Klausurerfolg

In Kapitel 1 wurden mit prozeduralem und konzeptuellem Wissen zwei Konstrukte eingeführt, die sich direkt auf Fachwissen in Mathematik beziehen. Im Rahmen der vorliegenden Arbeit werden unter anderem die Korrelationen zwischen diesen fachspezifischen Faktoren und der Klausurleistung untersucht. Damit soll insbesondere analysiert werden, ob und ggf. wie gut sich prozedurales und konzeptuelles Wissen als Prädiktoren für Klausurerfolg eignen.
Um die Stärke der Korrelation sowie die Prognosegüte der fachspezifischen Faktoren besser einordnen zu können, werden vier weitere Prädiktoren in die Untersuchungen einbezogen, deren Zusammenhang mit Prüfungsleistungen in der Literatur belegt ist. Hierbei handelt es sich um die beiden allgemein kognitiven Faktoren (1) kognitive Grundfertigkeit und (2) Konzentration sowie um die affektiven Faktoren (3) Prüfungsangst und (4) Prokrastination (vgl. Abbildung 2.1). Die Einbeziehung dieser Prädiktoren mit bekannten Zusammenhängen zur Prüfungsleistung dient auch dazu, die Repräsentativität der Stichprobe zu verifizieren. In diesem Fall sollten sich die in der vorliegenden Arbeit gefundenen Korrelationen zwischen Prädiktor und Prüfungsleistung im Rahmen der aus der Literatur bekannten Korrelationen bewegen.

Das Ziel des Kapitels ist, die genannten allgemeinen kognitiven und affektiven Faktoren vorzustellen und darzulegen, mit welcher Korrelation zur Prüfungsleistung zu rechnen ist.

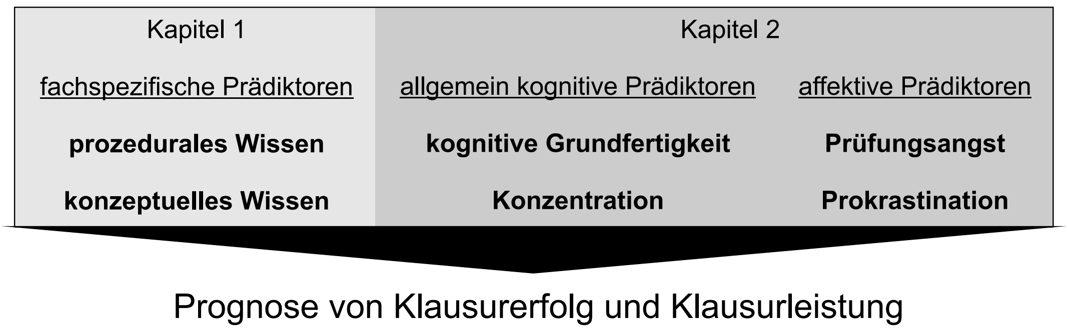


Abbildung 2.1 Überblick über die in der vorliegenden Arbeit untersuchten Prädiktoren

Ferner wird jeweils auf die Konzeptualisierung und Operationalisierung der Konstrukte eingegangen. Dabei werden zunächst die allgemein kognitiven Faktoren kognitive Grundfertigkeit (2.1) und Konzentration (2.2) eingeführt. Anschließend folgt die Beschreibung der affektiven Faktoren Prüfungsangst (2.3) und Prokrastination (2.4).

2.1 Kognitive Grundfertigkeit

Im nachfolgenden Abschnitt wird zunächst beschrieben, was in der vorliegenden Arbeit unter kognitiver Grundfertigkeit zu verstehen ist. Anschließend wird gezeigt, dass in der Literatur kognitive Grundfertigkeit als bester Einzelprädiktor für allgemeine Intelligenz angesehen wird. Allgemeine Intelligenz wiederum gilt als bester Prädiktor für Leistung sowohl in der Schule als auch an der Hochschule insgesamt (Rost und Sparfeldt, 2008, S. 309; Holling, Preckel und Vock, 2004, S. 47). Folglich ist davon auszugehen, dass kognitive Grundfertigkeit ein guter Prädiktor für Erfolg in Mathematikklausuren ist. Aus diesem Grund wurde dieser Prädiktor in die Untersuchungen der vorliegenden Arbeit aufgenommen. Ein weiterer Grund für seine Berücksichtigung ist, im Rahmen der Korrelationsanalysen den Anteil kognitiver Grundfertigkeit aus der Korrelation anderer Prädiktoren mit der Klausurleistung herauspartialisieren zu können. Dies ist wichtig, um kognitive Grundfertigkeit als Moderatorvariable auszuschließen.

2.1.1 Konzeptualisierung und Einordnung in psychometrische Intelligenztheorien

In der vorliegenden Arbeit wird kognitive Grundfertigkeit durch das in der Psychologie sogenannte schlussfolgernde Denken (engl. „reasoning“) repräsentiert. Damit ist die Fähigkeit gemeint, „Gesetzmäßigkeiten oder logisch zwingende Zusammenhänge zu erkennen und zweckvoll verwerten zu können“ (Kubinger, 2009, S. 198), oder anders formuliert: „Reasoning pertains to the process of drawing conclusions from principles and from evidence ... In reasoning, we move from what is already known to infer a new conclusion or to evaluate a proposed conclusion“ (Sternberg, 2011, S. 456). Schlussfolgerndes Denken in diesem Sinne ist abzugrenzen von dem eng gefassten „Schlussfolgern“ als mathematische Methode im Rahmen rein mathematischer Fragestellungen.

In den Intelligenztheorien spielt schlussfolgerndes Denken bis heute eine zentrale Rolle. Bereits Binet, der „Vater“ der Intelligenzforschung, benannte ursprünglich schlussfolgerndes Denken als einen Faktor, der die Intelligenz bestimmt (nach Wilhelm & Engle, 2004, S. 377).
Mit der Erfindung der Faktorenanalyse durch Spearman (1904) wurde schlussfolgerndes

Denken psychometrischen Methoden zugänglich.[1] Spearman untersuchte verschiedene Intelligenztests und bemerkte, dass diese positiv miteinander korrelierten. Er schloss daraus auf die Existenz eines Faktors, der an vielen verschiedenen Intelligenzleistungen beteiligt ist. Dieser Faktor wird als Generalfaktor der Intelligenz oder kurz g-Faktor bezeichnet. Er repräsentiert die allgemeine Intelligenz einer Person. Spearman (1927) fand heraus, dass jene Tests am höchsten auf dem g-Faktor laden, die ein Ableiten von Beziehungen und Schlussfolgern („eduction of relations“, „eduction of correlates“) erfordern. Dabei handelt es sich nach Jensen (1998) um Tests, die Folgendes beinhalten: „inductive and deductive reasoning, grasping relations, inferring rules, generalizing, seeing the similarity in things that differ ... or the difference between things that are similar ...“ (S. 35). Diese Beschreibung ist der obigen Charakterisierung von schlussfolgerndem Denken sehr ähnlich.
Im Gegensatz zu Spearman vertrat Thurstone (1938) die Auffassung, dass sich kognitive Leistungen nicht ausreichend durch einen g-Faktor und testspezifischen Faktoren erklären lassen. In seinen Faktorenanalysen extrahierte er neun sogenannte Primärfaktoren, die jeweils eine grundlegende Fähigkeit widerspiegeln. Eine dieser grundlegenden Fähigkeiten stellt bei Thurstone das schlussfolgernde Denken dar,[2] durch das in der vorliegenden Arbeit die kognitive Grundfertigkeit repräsentiert wird.
Horn und Cattell (1971) hingegen nahmen an, dass sich Spearmans g-Faktor aus zwei Komponenten zusammensetzt, der fluiden und der kristallinen Intelligenz, denen wiederum, ähnlich wie bei Thurstone, weitere Primärfaktoren untergeordnet sind. Die fluide Intelligenz wird dabei primär durch „Induction“, „Figural Relations“ und „Memory Span“ bestimmt (Horn, 1968, S. 249). Kognitive Grundfertigkeit, repräsentiert durch schlussfolgerndes Denken, ist damit im Modell von Cattell und Horn der fluiden Intelligenz zuzurechnen.
Die wohl umfangreichste faktorenanalytische Untersuchung zur Intelligenzmessung führte Carroll (1993) durch. Er reanalysierte 461 Datensätze aus der Intelligenzforschung und entwickelte auf dieser Grundlage seine „Three-Stratum-Theorie“. Auf der Ebene der höchsten Generalität, dem Stratum III, befindet sich die allgemeine Intelligenz im Sinne des g-Faktors nach Spearman. Es folgen auf Stratum II acht Intelligenzfähigkeiten mittlerer Generalität, die wiederum durch 69 relativ spezifische Fähigkeiten auf Stratum I beeinflusst werden. Auf Stratum II ist auch die fluide Intelligenz angesiedelt. Als dem Reasoning zuzurechnende Faktoren auf Stratum I nennt Carroll (1993) „Sequential Reasoning“, „Induction“ und „quantitative Reasoning“ (ebenda, S. 245). Kognitive Grund-

[1] Für eine ausführliche Darstellung der historischen Entwicklung faktorenanalytischer Methoden sei auf Jensen (1998, S. 39) verwiesen.

[2] In der Sekundärliteratur (vgl. Holling et al., 2004, S. 19; Süß, 2003, S. 218) werden meistens nur sieben Primärfaktoren aufgeführt. Hierbei werden die ursprünglich von Thurstone extrahierten Faktoren D (deductive), I (inductive) und R (restrictive thinking) zu einem Faktor „schlussfolgerndes Denken“ zusammengefasst.

fertigkeit, repräsentiert durch schlussfolgerndes Denken, ist in Carrolls Modell somit hier zu verorten.

In allen zuvor genannten Intelligenztheorien liefert kognitive Grundfertigkeit, repräsentiert durch schlussfolgerndes Denken, einen wesentlichen Beitrag zur allgemeinen Intelligenz. In der vorliegenden Arbeit wird daher untersucht, wie groß die Zusammenhänge zwischen kognitiver Grundfertigkeit und der Klausurleistung sind. Dazu ist kognitive Grundfertigkeit geeignet zu operationalisieren, worauf im nächsten Abschnitt eingegangen wird.

2.1.2 Operationalisierung

Einer der bekanntesten Tests zur Erfassung kognitiver Grundfertigkeit durch schlussfolgerndes Denken stellen die *Standard Progressive Matrices* (Raven & Court, 1998) dar, der als einer der besten Einzeltests zur Vorhersage fluider und allgemeiner Intelligenz gilt (Marshalek, Lohman und Snow, 1983, S. 122; Snow, Kyllonen und Marshalek, 1984, S. 91). Bei den *Standard Progressive Matrices* handelt es sich um einen figuralen Matrizentest, dessen Durchführung etwa 60 Minuten dauert.
Eine aus testökonomischer Sicht attraktive Alternative zur Messung kognitiver Grundfertigkeit stellt der Wiener Matrizen-Test 2 (Formann, Piswanger & Waldherr, 2011) dar. Hierbei handelt es sich ebenfalls um einen figuralen Matrizentest, dessen Durchführung allerdings weniger als 30 Minuten erfordert. Zudem gilt der Wiener Matrizen-Test 2 als Rasch-homogen. Dadurch stellt der übliche Verrechnungsmodus zu einem Testrohwert durch die Anzahl korrekt gelöster Aufgaben hier ein besonders faires Leistungsmaß dar (Gittler, 1999, S. 71), denn bei Rasch-skalierten Tests entscheidet über die Leistung lediglich, wie viele und nicht welche Aufgaben korrekt gelöst wurden.

Aus den vorgenannten Gründen wird auch im Rahmen der vorliegenden Arbeit mit dem Wiener Matrizen-Test 2 ein figuraler Matrizentest – und damit ein Testtyp, der als bester Einzelprädiktor für allgemeine Intelligenz gilt – zur Messung kognitiver Grundfertigkeit eingesetzt. Im folgenden Abschnitt wird dargelegt, mit welchen Korrelationen zur Klausurleistung zu rechnen ist.

2.1.3 Zusammenhänge mit Prüfungsleistung

Allgemeine Intelligenz gilt als der „beste Einzelprädiktor für den Erfolg in Schule, Hochschule, Ausbildung und Beruf“ (Rost & Sparfeldt, 2008, S. 309). Die Korrelation sinkt allerdings kontinuierlich in Abhängigkeit vom Alter bzw. der Schulstufe (vgl. Tabelle 2.1), was darauf hindeutet, dass spezifische Faktoren gegenüber der allgemeinen Intelligenz an Einfluss gewinnen (Schmidt-Atzert, Deter & Jaeckel, 2004). Gemäß Rost und Sparfeldt

Tabelle 2.1

Korrelationen der allgemeinen Intelligenz (IQ) mit verschiedenen Kriterien, unter anderem der Leistung im ersten Studienjahr

Untersuchtes Kriterium	Korrelation r
Schulleistungen (Grundschule, diverse Fächer)	$.56 - .71$
Leistungsrangplatz in der Abschlussklasse der high-school	.62
Leistung im ersten Studienjahr	**.44**
Durchschnittszensur (GPA) in diversen Colleges	$.30 - .70$
Abschlusszensur bei Jura-Studenten („law school“)	.30

Anmerkung. Survey zu Untersuchungsbefunden nach Jensen (1981, S. 31) (Tabelle gekürzt)

(2008, S. 310) hat sich die Befundlage zwischen 1981 und 2008 praktisch nicht verändert. Dennoch gelangen Richardson, Abraham und Bond (2012) in einer neueren Meta-Analyse, die Studien der vorangegangenen 15 Jahre berücksichtigt, für die Korrelation zwischen Intelligenz und Durchschnittsnoten im Studium zu einem anderen Ergebnis. Sie berichten eine deutlich geringere Korrelation von $r = .20$. Der relativ große Unterschied zu den Ergebnissen von Jensen (1981) ist möglicherweise auf unterschiedliche Erhebungsinstrumente und Berechnungen der Durchschnittsnote zurückzuführen.

Berücksichtigt man alle genannten Ergebnisse, so ist im ersten Studienjahr mit einer Korrelation zwischen allgemeiner Intelligenz und Prüfungsleistung zwischen $r = .20$ und $r = .44$ zu rechnen. Da kognitive Grundfertigkeit, operationalisiert durch figurale Matrizentests, als bester Einzelprädiktor für allgemeine Intelligenz gilt (Wilhelm, 2005, S. 385) und diese wiederum als bester Prädiktor für Leistung an der Hochschule angesehen wird (Rost & Sparfeldt, 2008, S. 309), kann im Durchschnitt mit einer ähnlichen Korrelation von $r = .20$ bis $r = .44$ zwischen Klausurleistung und kognitiver Grundfertigkeit gerechnet werden.

Fazit

Kognitive Grundfertigkeit im zuvor definierten Sinne und operationalisiert durch figurale Matrizentests gilt als bester Einzelprädiktor für allgemeine Intelligenz. Allgemeine Intelligenz wiederum wird als bester Prädiktor für Leistung in der Schule und Hochschule angesehen. In der vorliegenden Arbeit wird daher untersucht, wie groß die Zusammenhänge zwischen kognitiver Grundfertigkeit und der Leistung in Klausuren der Ingenieurmathematik sind. Dadurch ist es auch möglich, Korrelationen zwischen anderen Prädiktoren und der Klausurleistung unter Kontrolle der kognitiven Grundfertigkeit zu bestimmen.

Zwischen kognitiver Grundfertigkeit und Klausurleistung ist mit Korrelationen im Bereich von $r = .20$ bis $r = .44$ zu rechnen, was als ein Anhaltspunkt zur Überprüfung der

Repräsentativität der Stichprobe der vorliegenden Arbeit dient. Es werden ferner Vergleiche mit den Korrelationen der anderen in der Arbeit untersuchten Prädiktoren und der Klausurleistung angestellt. Darüber hinaus werden die Prognosegüten für Klausurerfolg und -leistung aller Prädiktoren verglichen.

2.2 Konzentration

Tests zur Messung von Konzentration weisen mit Mathematikklausuren für ingenieurwissenschaftliche Studiengänge einige Gemeinsamkeiten auf. In den Klausuren müssen Aufgaben in einer vorgegebenen Zeit bearbeitet werden. Die Klausuren enthalten damit ebenso wie Konzentrationstests eine Speed-Komponente. Ferner müssen in Mathematikklausuren neben den speziellen Anforderungen an prozedurales und konzeptuelles Wissen analog zu entsprechenden Konzentrationstests im Kern arithmetische Grundoperationen möglichst schnell und fehlerfrei durchgeführt werden. Folglich ist ein messbarer Zusammenhang zwischen der Leistung in Konzentrationstests und Klausuren der Ingenieurmathematik zu erwarten. Aus diesem Grund wurde Konzentration als weiterer Faktor zur Prognose von Klausurerfolg und -leistung in die Untersuchungen der vorliegenden Arbeit aufgenommen.

2.2.1 Konzeptualisierung

Konzentration wird allgemein als Aspekt des Arbeitens beschrieben, der die Selektion und Koordination von Informationen bzw. Reizen betrifft. Konzentration ist „die absichtsvolle nicht automatisierte Koordination von Handlungssteilen und deren kontrollierte Ausführung“ (Westhoff & Hagemeister, 2005, S. 20). Handlungsteile, die bei der Bearbeitung einer Aufgabe koordiniert werden, können beispielsweise Rechnungen, Abzählen und Merken sein. Konzentration ist damit abzugrenzen von Aufmerksamkeit: Während erstere primär die Weiterverarbeitung von Reizen erfordert, bezieht sich Aufmerksamkeit ausschließlich auf Wahrnehmungsprozesse und dient lediglich der Auswahl von relevanten Reizen oder Informationen (Leitner, 2005).

Konzentration gilt einerseits als Persönlichkeitsmerkmal, das über einen längeren Zeitraum transsituativ relativ stabil ist, andererseits unterliegt dieses stabile Konzentrationsniveau bei jedem Menschen Schwankungen, die von verschiedenen situativen, äußeren Bedingungen und individuellen Eigenschaften abhängen, wie beispielsweise Alter, körperlichen Faktoren, Motivation, Emotionen, Geübtheit, Strategien (Westhoff & Hagemeister, 2005, S. 22). Letzteres kommt auch in der Definition von Schmidt-Atzert, Büttner und Bühner (2004) zum Ausdruck, die unter Konzentration die „Fähigkeit, unter Bedingungen schnell und genau zu arbeiten, die das Erbringen einer kognitiven Leistung normalerweise erschweren“ (S. 9) verstehen. Erschwerende Bedingungen können Müdigkeit, Kopfschmerzen, Zeitdruck oder eine große Aufgabenmenge sein.

Es gibt allerdings keine einheitliche, allgemein anerkannte und gebräuchliche Definition von Konzentration (Schmidt-Atzert, Bühner & Enders, 2006, S. 33). Dennoch existieren zahlreiche Tests, die annehmen, Konzentration zu messen. Unter diesem Aspekt unterzogen Schmidt-Atzert et al. (2006) elf verschiedene Konzenzentrationstests einer gemeinsamen Faktorenanalyse. Dabei ließ sich ein Faktor extrahieren, der von numerischen und figuralen Fertigkeiten, von Intelligenz sowie von Reaktionsgeschwindigkeit abgegrenzt werden konnte. Schmidt-Atzert et al. (2006) bezeichnen diesen Faktor als „Konzentrationsfaktor“ und schlussfolgern, „dass die meisten hier verwendeten Konzentrationstests die gleiche Fähigkeit erfassen, nämlich Konzentration“ (S. 33). Einen Anhaltspunkt, was diesen Konzentrationsfaktor möglicherweise charakterisiert, liefert die im nachfolgenden Abschnitt vorgestellte Definition von Konzentrationstests nach Westhoff und Hagemeister (2005), denn dieser Definition genügen alle von Schmidt-Atzert et al. (2006) analysierten Tests, die eine hohe Ladung auf dem Konzentrationsfaktor aufweisen.

2.2.2 Operationalisierung

Konzentrationstests sind immer als Speed-Tests konzipiert, in denen es darum geht, so schnell wie möglich und richtig zu arbeiten. Die Anforderungen an Wahrnehmung, Gedächtnis, Lernen und Problemlösen werden so gering wie möglich gehalten, sodass die Aufgaben von jedem gesunden Menschen gelöst werden können. Damit sind Konzentrationstests abzugrenzen von Intelligenztests, die auch höhere mentale Operationen, wie zum Beispiel schlussfolgerndes, verbales, numerisches und visuell-räumliches Denken erfordern (Funsch, Martín & Halder-Sinn, 2011).

Zur Orientierung, ob es sich bei einem Test um einen Konzentrationstest handelt, kann die nachfolgende Definition für Konzentrationstests anhand qualitativer Merkmale von Westhoff und Hagemeister (2005) dienen.

> „Die Leistung eines hirnorganisch gesunden Probanden erzielt durch (mündliche oder manuelle) Reaktionen auf mehr oder weniger einfache (Bilder alltäglicher Gegenstände oder abstrakte Zeichen; Zahlen, Buchstaben oder andere) Reize, die er klar und eindeutig wahrnehmen kann und auf die er eine einfach zu erinnernde Regel anzuwenden hat, indem er absichtsvoll Teilhandlungen so schnell wie möglich bei (sehr niedriger bis sehr hoher) Geübtheit in der Ausführung dieses Tests korrekt koordiniert, kann abgebildet werden in die

> (sehr niedrige bis sehr hohe) Geschwindigkeit konzentrierten Handelns und den (sehr niedrigen bis sehr hohen) Anteil an Konzentrationsfehlern“ (Westhoff & Hagemeister, 2005, S. 40).

Konzentration wird in den meisten Tests über das Tempo und die Fehlerneigung beim konzentrierten Arbeiten operationalisiert (Schmidt-Atzert, Büttner & Bühner, 2004, S. 14). Beide Leistungsparameter gelten als weitgehend unabhängig voneinander (Steinborn, Flehmig, Westhoff & Langner, 2008; Westhoff & Graubner, 2003, S. 116). Das Tempo wird entweder über die Anzahl korrekt gelöster Aufgaben (Arbeitsleistung) oder über die Anzahl insgesamt bearbeiteter Aufgaben (Arbeitsgeschwindigkeit) erfasst.
Die Fehlerneigung wird meistens in Form des prozentualen Fehleranteils angegeben. Zum Teil werden Arbeitsleistung und Fehlerneigung miteinander verrechnet, um beispielsweise die Ratewahrscheinlichkeit oder das Überspringen von Aufgaben zu berücksichtigen (Goldhammer, Moosbrugger & Krawietz, 2009, S. 74).

Sowohl Tempo als auch Fehlerneigung gelten als retest-reliabel, sodass beide Leistungsparameter als Persönlichkeitsmerkmale angesehen werden (Westhoff & Hagemeister, 1992). Die Tempowerte sind in einer Population in der Regel normalverteilt, während die Fehlerneigung einer Poisson-Verteilung folgt (Westhoff & Hagemeister, 1992, S. 118).

Als Konzentrationstest wird in der vorliegenden Arbeit der Psychomeda-Konzentrationstest (Satow, 2011) eingesetzt. Zur Operationalisierung von Konzentration dienen hier die oben beschriebenen Kennzahlen Tempo und Fehlerneigung.

Schwierigkeiten bei der Messung von Konzentration

Bei der Durchführung des Konzentrationstests im Rahmen der vorliegenden Arbeit waren einige Aspekte zu beachten, die nachfolgend beschrieben werden. Auf weitere Details, insbesondere wie die diskutierten Aspekte bei der Testdurchführung berücksichtigt wurden, wird im Methodenteil eingegangen.

In Konzentrationstests sind über einen längeren Zeitraum kognitiv anspruchslose Aufgaben zu lösen. Die Bearbeitung ist anstrengend und monoton, da in der Regel immer Aufgaben vom selben Typ (z. B. Rechenaufgaben, Zahlen verbinden usw.) zu lösen sind. „Die Probanden stellen sich die Frage, zu welchem Zweck sie diese unangenehme Tätigkeit vollbringen. Geschieht das aus Gefälligkeit (man hat sich zu einer Testuntersuchung bereit erklärt) oder mit dem Ziel, eine Stelle zu bekommen?“ (Schmidt-Atzert, Büttner & Bühner, 2004, S. 20). Vor der Durchführung von Konzentrationstests ist also dafür zu sorgen, dass die Teilnehmerinnen und Teilnehmer motiviert sind, den gesamten Test schnell und sorgfältig zu bearbeiten. Westhoff und Graubner (2003) haben teilnehmenden

Schülerinnen und Schülern zur Motivation vor den Tests die Möglichkeit einer individuellen Rückmeldung angeboten und bemerken dazu: „Da fast alle diese Möglichkeit nutzten, gehen wir von einer hohen Motivation aus, in den Tests möglichst gute Leistungen zu zeigen“ (S. 113).

Eine weitere Schwierigkeit kann sich aus der weitgehenden Unabhängigkeit der Leistungsmaße Tempo und Fehlerneigung ergeben. Eine Testteilnehmerin oder ein Testteilnehmer könnte sich dazu entschließen, die Aufgaben mehrmals zu kontrollieren, um möglichst alle bearbeiteten Aufgaben korrekt zu bearbeiten. In diesem Fall würde die Fehlerneigung zu Lasten des Tempos gehen. Auch der umgekehrte Fall ist denkbar. Beide Fälle werden als „speed-accuracy-tradeoff“ (Schmidt-Atzert, Büttner & Bühner, 2004, S. 19) bezeichnet. Westhoff und Hagemeister (1992) bemerken hierzu:

> „Diese Strategie, jedes Item mehrmals zu bearbeiten, um ganz sicher keinen einzigen Fehler zu machen, ist zwar prinzipiell wählbar, führt aber zu einer äußerst geringen Arbeitsleistung und wird in der Regel von Gesunden unter der Standardinstruktion nicht benutzt“ (Westhoff & Hagemeister, 1992, S. 117).

Insofern sollte im Rahmen der Instruktion der Teilnehmenden deutlich darauf hingewiesen werden, so schnell wie möglich und richtig zu arbeiten. Ferner können bei der Analyse der Testergebnisse „Ausreißer“, die durch eine geringe Arbeitsgeschwindigkeit bei hoher Sorgfalt oder umgekehrt, durch eine hohe Arbeitsgeschwindigkeit bei geringer Sorgfalt auffallen, sorgfältiger inspiziert oder Teilnehmende ggf. nachträglich nach ihrem Arbeitsverhalten befragt werden.

2.2.3 Zusammenhänge mit Prüfungsleistung

Nach Rindermann und Neubauer (2000, S. 15) liegt die Korrelation zwischen dem Notendurchschnitt bei Schülerinnen und Schülern der Klassen 9 und 10 und der durch den Zahlen-Verbindungs-Test (Oswald & Roth, 1982) gemessenen Arbeitsleistung bei $r = .38$. Der Zahlen-Verbindungs-Test zählt zu den Konzentrationstest. Er lädt hoch auf dem von Schmidt-Atzert et al. (2006) gefundenen Konzentrationsfaktor.

In einer ähnlichen Untersuchung berichten Rindermann und Neubauer (2004, S. 580) Korrelationen von $r = .31$ bzw. $r = .25$ zwischen kombinierten Mathematik-Physik-Noten und dem Zahlen-Verbindungs-Test bzw. dem Kodierungstest (Sitzwohl, 1995). Am Test nahmen Schülerinnen und Schüler im Alter von 14 bis 17 Jahren teil.

Richardson et al. (2012) geben in einer Meta-Studie für die Korrelation zwischen Konzentrationsleistung und Durchschnittsnoten im Studium ein 95 %-Konfidenzintervall von [.14, .19] an.

Die unterschiedlich hohen Korrelationen können durch die verschiedenen Testinstrumente, Merkmale der Teilnehmenden und unterschiedliche Studiendesigns erklärt werden. Insgesamt ist von einem Zusammenhang zwischen Konzentrations- und Klausurleistung zwischen $r = .14$ und $r = .38$ auszugehen.

Fazit

Gemäß Westhoff und Hagemeister (2005, S. 30) erfasst ein Konzentrationstest die Geschwindigkeit und den Fehleranteil konzentrierten Handels. Beides ist wichtig zum Beispiel beim Anwenden arithmetischer Basisoperationen im Rahmen der Bearbeitung von Mathematikklausuren. Folglich ist ein Zusammenhang zwischen Konzentration und Klausurleistung naheliegend und in der Literatur bestätigt. Es ist mit Korrelationen zwischen $r = .14$ und $r = .38$ zu rechnen, was in der vorliegenden Arbeit auch zur Überprüfung dient, ob die vorliegende Stichprobe mit Stichproben aus anderen Studien vergleichbar ist. Dies würde ein Verallgemeinern der Resultate erleichtern, ohne auf Besonderheiten der Stichprobe eingehen zu müssen. Ferner wird ein Vergleich mit entsprechenden Korrelationen und Prognosegüten für Klausurerfolg und -leistung der in Kapitel 1 eingeführten fachspezifischen Faktoren, der kognitiven Grundfertigkeit, sowie der nachfolgend vorgestellten affektiven Faktoren durchgeführt. Zur Messung von Konzentration wird der Psychomeda-Konzentrationstest eingesetzt (Satow, 2011).

2.3 Prüfungsangst

Zum Thema Prüfungsangst existiert ein breites Spektrum an Ratgeberliteratur (z.B. Fehm & Fydrich, 2013; Walther, 2012; Knigge-Illner, 2010). An einigen Hochschulen finden zudem umfangreiche Interventionsmaßnahmen statt (Schaefer, Albus & Pfitzer, 2009). Beides lässt vermuten, dass Prüfungsangst ein ernst zu nehmendes, weitverbreitetes und behandlungsbedürftiges Phänomen bei Studierenden ist, das den Prüfungserfolg maßgeblich beeinflussen kann. Es könnte beispielsweise sein, dass Studierende durch ihre Prüfungsangst von der Bearbeitung der Klausur abgelenkt werden, was zu einem schlechteren Ergebnis oder gar dem Nichtbestehen der Klausur führt. Zusammenhänge zwischen Prüfungsangst und Prüfungserfolg werden durch Studien bestätigt. Aus diesen Gründen wurde Prüfungsangst in die Untersuchungen der vorliegenden Arbeiten einbezogen.

2.3.1 Konzeptualisierung

Es existiert keine einheitliche Definition von Prüfungsangst. Dies liegt unter anderem daran, dass Prüfungsangst in den wichtigsten internationalen Klassifikationssystemen für psychische Störungen, dem *Diagnostic and Statistical Manual of Mental Disorders* (DSM, siehe APA, 2013) sowie der *International Statistical Classification of Diseases and Related Health Problems* (kurz ICD, deutsche Übersetzung des psychische Störungen behandelnden Kapitels V siehe Dilling, Mombour, Schmidt et al., 1991) der Weltgesundheitsorgansisation (WHO) nicht unter den „Störungen mit Krankheitswert“ gelistet ist. Prüfungsangst wird in beiden Kompendien lediglich als Beispiel für eine soziale Phobie (DSM) bzw. für eine spezifische Phobie (ICD-10) erwähnt, allerdings ohne weitere Ausdifferenzierung bzw. Angabe von charakterisierenden Merkmalen.

In der psychologischen Literatur gibt es folglich verschiedene Definitionen von Prüfungsangst. Hong (1998) beschreibt Prüfungsangst wie folgt: „Test anxiety is a complex multidimensional construct involving cognitive, affective, physiological, and behavioral reactions to evaluative situations“ (S. 52). Zeidner (1998) formuliert konkreter: „The term ’test anxiety’, as a scientific construct, refers to the set of phenomenological, physiological, and behavioral responses that accompany concern about possible negative consequences or failure on an exam or similar evaluative situation ...“ (S. 17). Auch Schwarzer (2000) betont in seiner Definition die als bedrohlich empfundenen Leistungsanforderungen einer Prüfung: „*Leistungsangst ist die Besorgtheit und Aufgeregtheit angesichts von Leistungsanforderungen, die als selbstwertbedrohlich eingeschätzt werden*“ (S. 105).

Allen Definitionen gemeinsam ist die Charakterisierung von Prüfungsangst als mehrdimensionales Konstrukt, das aus kognitiven und emotionalen Komponenten besteht. Die Definitionen enthalten allerdings keinen Hinweis darauf, ob und wann Prüfungsängste unter klinisch-psychologischen Aspekten bedeutsam sind. Um diesen Aspekt besser zu berücksichtigen, definieren Fehm und Fydrich (2011) Prüfungsangst als

> „anhaltende und deutlich spürbare Angst in Prüfungssituationen und/oder während der Zeit der Prüfungsvorbereitung, die den Bedingungen der Prüfungsvorbereitung und der Prüfung selbst nicht angemessen ist. Die Angst äußert sich auf den Ebenen Verhalten, Emotion, Kognition und Physiologie. Klinisch relevante Prüfungsängste liegen vor, wenn die Ängste das alltägliche Leben und/oder den Ausbildungsverlauf bzw. das berufliche Weiterkommen deutlich beeinträchtigen“ (Fehm & Fydrich, 2011, S. 7).

Da ein Schwerpunkt der vorliegenden Arbeit auf der Untersuchung von Zusammenhängen zwischen verschiedenen Faktoren und Prüfungsleistung liegt, werden im Folgenden zwei prominente Funktionsmodelle vorgestellt, die sich primär mit einer Erklärung des Wirkzusammenhangs zwischen Prüfungsängsten und Prüfungsleistung befassen.

Das Interferenz-Modell

Das auf der Aufmerksamkeitshypothese von Wine (1971) gründende Interferenz-Modell geht davon aus, dass prüfungsängstliche Personen während einer Prüfung ihre Aufmerksamkeit auf prüfungsrelevante und prüfungsirrelevante kognitive Aktivitäten wie die gedankliche Beschäftigung mit Sorgen über den Prüfungsausgang, Selbstkritik und Auseinandersetzung mit somatischen Beschwerden, aufteilen. Diese Kognitionen stören ein effektives Zeitmanagement und ein erfolgreiches Absolvieren der Prüfung. Untersuchungen von Deffenbacher (1978) deuten darauf hin, dass prüfungsängstliche Studierende aus diesem Grund nur etwa 60 % der zur Verfügung stehenden Zeit auf die Bearbeitung der Prüfungsaufgaben verwenden.

Für das Interferenz-Modell gibt es zahlreiche empirische Bestätigungen (Deffenbacher & Deitz, 1978; Sarason, 1984; Blankstein, Toner & Flett, 1989). Allerdings wird es auch infrage gestellt, insbesondere durch die Beobachtung, dass durch geeignete Interventionen Prüfungsangst meistens signifikant reduziert wird, dies allerdings nur in 18 % der Fälle auch mit einer Verbesserung der Prüfungsleistung einhergeht (Hembree, 1988, S. 49), sodass Tryon (1980) anmerkt: „In summary, almost any type of treatment seems to reduce self-reported test anxiety. The same cannot be said when it comes to grades“ (S. 362). An diese Kritik knüpft das zu Beginn der 1980er-Jahre entwickelte Defizit-Modell an.

Das Defizit-Modell

Das Defizit-Modell postuliert, dass hochprüfungsängstliche Personen aus zwei Gründen schlechtere Prüfungsleistungen erbringen: Einerseits weil sie über eine mangelnde Studierfähigkeit verfügen und andererseits weil sie sich dessen mehr und mehr bewusst werden (Zeidner, 1998, S. 71). Im Gegensatz zum Interferenz-Modell sind schlechte Prüfungsleistungen damit primär eine Folge mangelnder Fähigkeiten. Das Bewusstsein darüber wiederum führt zu einer erhöhten Prüfungsangst, weshalb diese nur scheinbar mit schlechten Prüfungsleistungen korreliert. Auch für das Defizit-Modell finden sich zahlreiche empirische Bestätigungen (Naveh-Benjamin, McKeachie, Lin & Holinger, 1981; Culler & Holahan, 1980; Kirkland & Hollandsworth, 1980; Paulman & Kennelly, 1984). Bruch (1981, S. 49) berichtet eine Korrelation von $r = -.42$ zwischen „test-taking strategies“ und Prüfungsangst unter Studienanfängern, was bedeutet, dass eine Reduzierung von Defiziten mit einer Reduzierung von Prüfungsangst korreliert.

Gegen das Defizit-Modell ist einzuwenden, dass auch Personen mit guten Studierfertigkeiten über kognitive Interferenzen in Prüfungssituationen berichten (Naveh-Benjamin, McKeachie & Lin, 1987), was mit den Modellannahmen schwer vereinbar ist. Daher wurde der Versuch unternommen, Interferenz- und Defizit-Modell zu vereinen oder Modelle auf Grundlage anderer Annahmen zu entwickeln. Daraus resultierten beispielsweise das Selbstregulationsmodell, das Selbstwertmodell und das transaktionale Modell (vgl. Fehm & Fydrich, 2011; Zeidner, 1998).

2.3.2 Operationalisierung

Im letzten Abschnitt wurde deutlich, dass Prüfungsangst sowohl eine kognitive als auch eine emotionale Komponente zugeschrieben wird. Die Operationalisierung von Prüfungsangst war eng mit der Entwicklung und Anpassung entsprechender Testinstrumente verbunden. Dabei wurden unterschiedliche Ansätze verfolgt, welche und wie viele kognitive und emotionale Komponenten Prüfungsangst am besten repräsentieren. Operationalisierung, Testentwicklung und eine Untersuchung der Zusammenhänge mit der Prüfungsleistung waren daher eng miteinander verbunden und können folglich nicht getrennt voneinander beschrieben werden. Im Folgenden wird daher die historische Entwicklung kurz skizziert, um am Ende des Abschnitts die im deutschsprachigen Raum heute gebräuchliche Operationalisierung von Prüfungsangst gemeinsam mit einem passenden Erhebungsinstrument darzustellen.

Den Ausgangspunkt der Entwicklung psychometrischer Instrumente zur Messung von Prüfungsangst stellen die Arbeiten von Sarason und Mandler dar (Sarason & Mandler, 1952; Mandler & Sarason, 1952). Sarason und Mandler zählen zu den ersten Forschenden, die mit ihrem *Test Anxiety Questionaire* Studierende in hoch- und niedrigprüfungsängstlich einteilten und einen Zusammenhang zwischen dem *Test Anxiety Questionaire*-Score und der Prüfungsleistung aufzeigten.
Faktorenanalysen offenbarten eine zweifaktorielle Struktur des *Test Anxiety Questionaire*. Die beiden Faktoren interpretierten Liebert und Morris (1967) im Rahmen ihrer „Zwei-Komponenten-Theorie der Prüfungsangst“ als „Besorgtheit-Komponente“ („worry“) und „Aufgeregtheit-Komponente“ („emotionality“).
Liebert und Morris (1967) untersuchten mit ihrem *Worry-Emotionality Questionaire*, der auf dem *Test Anxiety Questionaire* aufbaut, den Einfluss der beiden Komponenten auf die Selbsteinschätzung von Studierenden hinsichtlich einer zu erbringenden Prüfungsleistung. Sie stellten eine negative Korrelation zur Besorgtheit fest, fanden aber keinen signifikanten Zusammenhang zur Aufgeregtheit. Obwohl dieses Resultat vielfach bestätigt wurde (Cassady & Johnson, 2002; Hodapp, 1982; Seipp, 1991), zählen bis heute Besorgtheit und Aufgeregtheit als Vertreter einer kognitiven und emotionalen Komponente zu integralen Facetten von Prüfungsangst.

Die teilweise unbefriedigende psychometrische Qualität des *Worry-Emotionality Questionaire* veranlasste Spielberger (1980) zu einer Weiterentwicklung dieses Fragebogens.

Auf diese Weise entstand das *Test Anxiety Inventory*, das auch heute noch im englischsprachigen Raum eingesetzt wird.
Etwa zeitgleich zur Etablierung des *Test Anxiety Inventory* erfolgte die Entwicklung der *Reactions to tests-Skala* durch Sarason (1984). Der *Reactions to tests-Skala* liegt der Gedanke zugrunde, dass Prüfungsangst neben den Faktoren Besorgtheit und Aufgeregtheit noch weitere Komponenten umfasst. Inspiriert durch die Aufmerksamkeitshypothese von Wine (1971), derzufolge Prüfungsangst Aufmerksamkeit von der Prüfungsaufgabe abzieht, enthält die *Reactions to tests-Skala* neben Items der Skalen Besorgtheit („worry") und Aufgeregtheit (von Sarason als „tension" statt „emotionality" bezeichnet) auch Items einer Skala Interferenz („test-irrelevant thinking"), die ein Maß für die Ablenkbarkeit durch Gedanken, die sich nicht auf die Prüfungsaufgaben beziehen, darstellt. Faktorenanalysen offenbarten jedoch eine vierfaktorielle Struktur der *Reactions to tests-Skala*, sodass Sarason nachträglich Items zu einer vierten Skala, „bodily symptoms", zusammenfasste.

Während beim *Test Anxiety Inventory* weiterhin die hohe Korrelation (und damit unzureichende Unabhängigkeit) der Skalen Besorgtheit und Aufgeregtheit kritisiert wurde (Hodapp, 1995), wurde bei der *Reactions to tests-Skala* unter anderem die geringe interne Konsistenz einiger Skalen bemängelt (ebenda). Daher unternahm Hodapp (1991) den Versuch einer Neukonstruktion eines Prüfungsangstfragebogens. Das Resultat war das *Test Anxiety Inventory-German*, das Wacker, Jaunzeme und Jaksztat (2008) als das „im deutschen Sprachraum ... gegenwärtig wohl am besten bewährte Verfahren" (S. 74) bezeichnen.
Die gute psychometrische Qualität des *Test Anxiety Inventory-German* wurde vielfach bestätigt (Musch & Bröder, 1999; Wacker et al., 2008; Hodapp, 1995). Seit 2011 wird das *Test Anxiety Inventory-German* in einer gekürzten und normierten Version unter der Bezeichnung *Prüfungsangstfragebogen* (Hodapp, Rohrmann & Ringeisen, 2011) von der Testzentrale (Hogrefe Verlag) vertrieben. Wie das *Test Anxiety Inventory-German* weist auch der *Prüfungsangstfragebogen* eine vierfaktorielle Struktur mit den Subskalen „Besorgheit", „Mangel an Zuversicht", „Aufgeregheit" und „Interferenz" (Analogie zur Skala „test-irrelevant thinking" aus der *Reactions to tests-Skala*) auf.
Das *Test Anxiety Inventory-German* und der *Prüfungsangstfragebogen* gelten damit als eine psychometrisch befriedigende Operationalisierung der angeführten Definitionen von Prüfungsangst, insbesondere jener nach Fehm und Fydrich (2011). Das *Test Anxiety Inventory-German* wird daher auch in der vorliegenden Arbeit eingesetzt.

Schwierigkeiten bei der Messung von Prüfungsangst

Auch für die vorliegende Arbeit ist wie bei allen Messungen von Prüfungsangst zu berücksichtigen, dass diese durch Fragebögen erhoben wird, deren Auswertung auf der Selbstauskunft der Teilnehmenden beruht. Selbstauskünfte können zu einer erheblichen Einschränkung der Validität führen (bezogen auf selbstberichtete Noten siehe beispielsweise Kuncel, Credé & Thomas, 2005; Bahrick, Hall & Berger, 1996). Es sollte daher stets die Verlässlichkeit der Selbstauskünfte reflektiert bzw. anhand psychometrischer Kenn- und Vergleichswerte die Zuverlässigkeit der Ergebnisse kontrolliert werden. Auf diesbezügliche Details der Erhebung für die vorliegende Arbeit wird im Methodenteil eingegangen.

2.3.3 Zusammenhänge mit Prüfungsleistung

Bezüglich Korrelationen zwischen Prüfungsangst und Prüfungsleistungen bemerkt Seipp (1991):

> „Hundreds of studies so far have investigated the complex relationship between anxiety and different kinds of performance. They have come up with an overwhelming amount of diverse results, ranging from positive to negative and from strong to weak relations, depending on different anxiety constructs, characteristics of subjects or the conceptualization of performance“ (Seipp, 1991, S. 27).

Anders ausgedrückt heißt dies, dass sich zu nahezu jedem „gewünschten“ Zusammenhang zwischen Prüfungsangst und Prüfungsleistung auch eine passende Studie findet, die diesen Zusammenhang belegt. Aus diesem Grund sollten ausschließlich Meta-Studien, die viele Einzelstudien zusammenfassen und mitteln, herangezogen werden, um einen möglichst objektiven Eindruck zu gewinnen.
Dabei und beim Vergleich von Meta-Studien sind stets die Heterogenität der Einzelstudien sowie die unterschiedlichen Auswahlkriterien zu bedenken. Die Werte aus Meta-Studien können daher nur als Anhaltspunkt für eigene Untersuchungen dienen. Tabelle 2.2 stellt die ermittelten Korrelationen zwischen der Besorgtheit-Komponente („worry“) und

Tabelle 2.2
Zusammenhang zwischen Komponenten von Prüfungsangst und Prüfungsleistung in zwei Meta-Studien

Variable und Studie	k	N	$\bar{r}$	r_{min}	r_{max}	KI95
„worry“ (Hembree)	13	1112	$-.31^{*}$	$-.41$	$-.25$	
„worry“ (Seipp)	38	6885	$-.22$	–	–	$[-.40, -.04]$
„emotionality“ (Hembree)	13	1112	$-.15^{*}$	$-.28$	$-.01$	
„emotionality“ (Seipp)	37	5182	$-.15$	–	–	–

Anmerkung. k=Anzahl Studien, N=Stichprobengröße, $\bar{r}$=mittlere Korrelation zur Leistung, r_{min}=Minimum, r_{max}=Maximum, KI95 = 95 %-Konfidenzintervall, $^{*}p < .01$, Quellen: Hembree (1988) und Seipp (1991) der Aufgeregtheit-Komponente („emotionality") mit der Prüfungsleistung aus den Meta-Analysen von Hembree (1988) und Seipp (1991) gegenüber. Die zu erwartende Korrelation zwischen Klausurleistung und Prüfungsangst bzw. den Skalen Besorgtheit und Aufgeregtheit liegt somit zwischen $r = -.31$ und $r = -.15$.

Fazit

Eine einheitliche Definition von Prüfungsangst existiert nicht. Den meisten Definitionen gemeinsam ist allerdings die Charakterisierung von Prüfungsangst als mehrdimensionales Konstrukt, das aus kognitiven und emotionalen Komponenten besteht. Im deutschsprachigen Raum gilt das *Test Anxiety Inventory-German* als eine bevorzugte Operationalisierung mit guten psychometrischen Eigenschaften. Aus diesem Grund wird es auch in der vorliegenden Arbeit eingesetzt. Die hierbei zu erwartende Korrelation zwischen Prüfungsangst und Klausurleistung liegt zwischen $r = -.31$ und $r = -.15$. Dieser aus Meta-Studien gewonnene Bereich dient als Anhaltspunkt für die Repräsentativität der Stichprobe der vorliegenden Arbeit, bezüglich der die Korrelationen mit der Klausurleistung sowie die Prognosegüte von Prüfungsangst für Klausurerfolg und -leistung mit den entsprechenden Werten der anderen fachspezifischen, allgemein kognitiven sowie der nachfolgend vorgestellten Prokrastination verglichen werden.

2.4 Prokrastination

Wie zur Prüfungsangst existieren auch zum Thema Prokrastination viele Ratgeber (z. B. Knaus, 2002; Rustemeyer & Callies, 2013; Kratz, 2011). An einigen Hochschulen finden zudem umfangreiche Interventionsmaßnahmen statt (Rist, Engberding, Patzelt & Beißner, 2006, S. 76). Beides lässt vermuten, dass auch Prokrastination ein ernst zu nehmendes, weitverbreitetes und behandlungsbedürftiges Phänomen bei Studierenden ist, das Prüfungserfolg maßgeblich beeinflussen kann, beispielsweise weil Lernen aufgeschoben wird und damit nicht im notwendigen Umfang oder gar nicht stattfindet. Aus diesen Gründen wurde neben Prüfungsangst auch Prokrastination als zweiter affektiver Faktor in die Untersuchungen der vorliegenden Arbeit einbezogen.

2.4.1 Konzeptualisierung

Prokrastination leitet sich aus dem lateinischen Wort *procrastino* (*pro* + *crastinus*) ab, was wörtlich „auf morgen verschieben" bedeutet (Georges, o. J.). In der Antike hatte Prokrastination eine neutrale bis positive Bedeutung. Es galt als weise, beispielsweise das Für und Wider eines Kriegsbeginns einen Tag länger abzuwägen, statt eine unüberlegte Entscheidung zu treffen. Seit der industriellen Revolution ist Prokrastination allerdings vor allem in westlichen Kulturkreisen eher negativ konnotiert (für einen historischen Überblick siehe Steel, 2007, S. 66).

Die aktuelle Forschung hat den Begriff Prokrastination auf viele Bereiche übertragen und damit verschiedene Forschungszweige eröffnet, die sich zum Teil unabhängig voneinander weiterentwickeln. Rustemeyer und Callies (2013, S. 22) z. B. unterscheiden in ihrem Überblick über die Taxonomie sechs Arten von Prokrastination, deren Untersuchung sich unter anderem in Abhängigkeit von der Interpretation der Prokrastinationsart als Persönlichkeitsmerkmal oder situationsspezifisches Merkmal weiterverzweigt.

Aufgrund dieser vielfältigen Betrachtungen existiert keine einheitliche Definition von Prokrastination: „Bislang liegen weder eine einheitliche Definition noch eine überzeugende Theorie vor, die in der Lage wären, die vielen atheoretischen empirischen Befunde und Erklärungsansätze zur Prokrastination zu bündeln“ (Rustemeyer & Callies, 2013, S. 23).

Dennoch gibt es Versuche einer Strukturierung. Ferrari, Johnson und McCown (1995, S. 3) ordnen die aus verschiedenen Forschungsrichtungen stammenden Definitionen in solche ein, die (1) einen zeitlichen Aspekt, (2) einen moralischen Aspekt, (3) einen irrationalen Aspekt, (4) einen operationalen Aspekt oder (5) einen mehrdimensionalen Aspekt betonen. Klingsieck (2013) versucht in einem Überblicksartikel, Gemeinsamkeiten der verschiedenen Definitionen zusammenzuführen. Sie extrahiert aus bestehenden Definitionen sieben Aspekte von Prokrastination:

(a) Eine offenkundige oder verborgene Handlung wird aufgeschoben.

(b) Der Start oder die Beendigung einer Handlung ist beabsichtigt.

(c) Die Handlung ist notwendig oder von persönlicher Wichtigkeit.

(d) Das Aufschieben ist freiwillig und wird nicht durch äußere Einflüsse erzwungen.

(e) Das Aufschieben ist unnötig oder irrational.

(f) Das Aufschieben findet trotz Bewusstsein über die negativen Konsequenzen statt.

(g) Das Aufschieben wird begleitet von subjektivem Unbehagen oder anderen negativen Konsequenzen.

Auf diese Weise gelingt eine Abgrenzung zu „stratetic delay“, das mit Prokrastination nur die Eigenschaften a)–d) teilt.
Ferner gelingt die Charakterisierung von Prokrastination als dysfunktionale Eigenschaft und damit eine Abgrenzung zu der von einigen Autorinnen und Autoren vertretenen positiven, aktiven und funktionalen Beschreibung von Prokrastination (Chu & Choi, 2005;

Schraw, Wadkins & Olafson, 2007). Die Charakterisierung von Prokrastination als dysfunktional wird durch Meta-Studien gestützt, die einen signifikanten negativen Zusammenhang zu Gewissenhaftigkeit und Prüfungsleistung sowie einen signifikant positiven Zusammenhang zu Neurotizismus, Selbstbehinderung und Ablenkbarkeit durch zur Lösung eines gegebenen Problems nicht relevante Gedanken nachweisen (Kim & Seo, 2015; Steel, 2007; Van Eerde, 2003).

Schließlich definiert Klingsieck (2013) Prokrastination wie folgt: „... procrastination can now be defined as *the voluntary delay of an intended and necessary and/or [personally] important activity, despite expecting potential negative consequences that outweigh the positive consequences of the delay*“ (S. 26).

Da es keine einheitliche Definition von Prokrastination gibt, existiert auch kein einheitlicher theoretischer Rahmen (Van Eerde, 2003, S. 1401). Klingsieck (2013, S. 26) schlägt daher eine differenzierte Betrachtung gemäß der psychologischen Forschungsrichtungen vor, die sich mit Prokrastination befassen. Sie unterscheidet zwischen der (1) motivations- und volitionspsychologischen, (2) der differentialpsychologischen, (3) der klinisch-psychologischen Perspektive.
(1) Aus motivations- und volitionspsychologischer Sicht wird Prokrastination als Selbstregulationsdefizit dargestellt, das den Prozess von der Intention bis zur Ausführung einer Handlung an mehreren Punkten stören und zu Prokrastination führen kann. Rist et al. (2006, S. 74) sowie Klingsieck und Fries (2012, S. 183) machen dies beispielsweise am „Rubikonmodells der Handlungsphasen“ (H. Heckhausen & Gollwitzer, 1987; J. Heckhausen & Heckhausen, 2006) deutlich.
(2) Prokrastination kann als eine zeit- und situationsstabile Eigenschaft angesehen werden (Steel, 2007, S. 67). In der differentiellen Psychologie wird Prokrastination daher als Persönlichkeitseigenschaft betrachtet und ihre Einordnung in das nomologische Netzwerk weiterer Persönlichkeitsmerkmale untersucht (Klingsieck & Fries, 2012, S. 183). In diesem Zusammenhang wird aufgrund einer hohen Korrelation zum Persönlichkeitsmerkmal Gewissenhaftigkeit („conscientiousness“) die Charakterisierung von Prokrastination als eigenständiges Merkmal kontrovers diskutiert (Schouwenburg, 2004; Steel, 2007). Aufgrund von faktorenanalytischen Betrachtungen verschiedener Forschungsarbeiten über die Facetten von Gewissenhaftigkeit gelangt Steel (2007) allerdings zu folgendem Schluss: „Consequently, procrastination may be considered to be the most central facet of conscientiousness, but it is not conscientiousness itself“ (S. 67).
(3) Der klinisch-psychologische Blick richtet sich auf die negativen Konsequenzen und Korrelate von Prokrastination wie Depression, Stress, Prüfungsangst etc. Es wird untersucht, wann Prokrastination, beispielsweise in Anlehnung an das *Diagnostic and Statistical Ma-*

nual of Mental Disorders (DSM, siehe APA, 2013), als behandlungsbedürftig eingestuft werden kann (Engberding, Frings, Höcker, Wolf & Rist, 2011) und welche Interventionsmaßnahmen geeignet erscheinen (Rist et al., 2006; Ferrari et al., 1995; Klingsieck, 2013, S. 187). Für eine Diskussion verschiedener Behandlungsansätze sei auf die Monografie von Rustemeyer und Callies (2013, S. 109) verwiesen.

2.4.2 Operationalisierung

Prokrastination wird meistens mithilfe von Fragebögen gemessen. In diesem Fall wird das Ergebnis durch die subjektive Selbsteinschätzung der Teilnehmerinnen und Teilnehmer sowie durch deren Motivation, die Fragen ehrlich zu beantworten, beeinflusst.
Um Prokrastination objektiv zu messen, kann beispielsweise registriert werden, wann Studierende sich zu einer Prüfung anmelden. Der zeitliche Abstand zwischen Anmeldung und letztmöglichem Anmeldezeitpunkt ist dann ein objektives Maß für Prokrastination (Rotenstein, Davis & Tatum, 2009).

Bei der Diagnose von Prokrastination durch Fragebögen ist bei der Auswahl eines geeigneten Messinstruments zu hinterfragen, ob dieses für den Untersuchungszweck tatsächlich geeignet ist. Denn existierende Fragebögen unterscheiden sich sowohl bezüglich ihrer Zielgruppe als auch bezüglich des zugrunde liegenden konzeptuellen Rahmens. Während beispielsweise die deutsche Version des Academic Procrastination State Inventory (APSI-d, deutsche Bearbeitung durch Helmke & Schrader, 2000) Prokrastination situativ bezogen auf die zurückliegende Woche erfasst, misst die Aitken Procrastination Scale (APS-d, deutsche Bearbeitung ebenfalls durch Helmke & Schrader, 2000) Verhaltensweisen und Gefühle ohne zeitliche Spezifizierung. Beide Fragebögen setzen sich aus jeweils drei Subskalen und 22 bzw. 19 Items zusammen.
Eine testökonomische Alternative bildet die eindimensionale Prokrastinationsskala von Schwarzer (1999) mit zehn Items und „guten psychometrischen Eigenschaften ...“ (Fydrich, 2009, S. 320), wobei die Autoren diese Eigenschaften nicht näher benennen. Die Prokrastinationsskala wird auch in der vorliegenden Arbeit eingesetzt. Die Messung von Prokrastination erfolgt also auf Grundlage eines Fragebogens und damit durch Selbstauskunft der Teilnehmerinnen und Teilnehmer. Eine eingehendere Beschreibung der Prokrastinationsskala erfolgt im Methodenteil.

Schwierigkeiten bei der Messung von Prokrastination

Steel, Brothen und Wambach (2001, S. 101) berichten in ihrer Studie über 152 Studierende eine nur geringe Korrelation von $r = .35$ zwischen selbstberichteter, d. h. durch Fragebögen erhobener, und anhand externer, objektiver Kriterien gemessener Prokrastination. Dadurch erklärt sich der Unterschied in den gefundenen Zusammenhängen zu wichtigen Korrelaten wie der Prüfunsleistung: Während die Korrelation zwischen Prüfungsleistung und objektiv gemessener Prokrastination bei $r = -.38$ lag, betrug die Korrelation zu selbstberichteter Prokrastination nur $r = -.23$ (ebenda). Kim und Seo (2015) greifen die Untersuchungen von Steel et al. (2001) auf und schließen, dass selbstberichtete Prokrastination die tatsächliche Prokrastination überschätzt. Sie konstatieren: „These results indicate that the using of self-report data may bias the results of investigations into the association between procrastination and academic performance“ (S. 27).

Der relativ schwache Zusammenhang zwischen selbstberichteter und objektiver Prokrastination wurde vielfach und mit verschiedenen Erhebungsinstrumenten bestätigt (Solomon & Rothblum, 1984; Tice & Baumeister, 1997). Dadurch kann die Aussagekraft der Erhebung durch Fragebögen geschwächt werden. Dies gilt auch für die vorliegende Arbeit, da eine objektive Erfassung von Prokrastination aus organisatorischen Gründen nicht möglich war.

2.4.3 Zusammenhänge mit Prüfungsleistung

Über den Zusammenhang zwischen Prokrastination und Prüfungsleistung finden sich verschiedene, zum Teil widersprüchliche Ergebnisse. Die gemessenen Korrelationen hängen unter anderem von den Erhebungsinstrumenten ab und auch davon, ob Daten objektiv oder durch Selbstauskunft der Studienteilnehmerinnen und -teilnehmer erhoben werden. Aus diesem Grund sollte bei der Beurteilung eines Zusammenhangs zwischen Prokrastination und Prüfungsleistung auf Ergebnisse aus Meta-Studien zurückgegriffen werden.

Kim und Seo (2015, S. 31) geben in ihrer Meta-Studie eine Korrelationen zwischen Prokrastination und Prüfungsleistung von $r = -.08$ an, falls die Datenerhebung sowohl für Prokrastination als auch für die Prüfungsleistung auf der Selbstauskunft der Teilnehmerinnen und Teilnehmer beruht. Für den Fall, dass in beiden Fällen die Daten objektiv erhoben werden, berichten sie eine Korrelation von $r = -.39$. In dem häufigen Fall, dass die Messung von Prokrastination durch einen Fragebogen, also durch Selbstauskunft, und die Erhebung der Prüfungsleistung objektiv erfolgt, berichten Kim und Seo (2015, S. 31) eine Korrelation von $r = -.15$. Das 95 %-Konfidenzintervall geben sie mit $[-.22, -.08]$ an. Insgesamt kann also nach Kim und Seo (2015), in deren Meta-Studie auch die Untersuchungen von Steel et al. (2001) enthalten sind, von einem eher schwachen Zusammenhang von $r = -.22$ bis $r = -.08$ zwischen selbstberichteter Prokrastination und objektiv gemessener Prüfungsleistung ausgegangen werden.

Fazit

Analog zu dem anderen in dieser Arbeit untersuchten affektiven Faktor Prüfungsangst existiert auch für Prokrastination keine einheitliche Definition. Weitestgehend übereinstimmend wird Prokrastination als dysfunktionale Eigenschaft angesehen, die das Aufschieben einer wichtigen Handlung trotz drohender negativer Konsequenzen beinhaltet. Zur Messung von Prokrastination wird in der vorliegenden Arbeit die Prokrastinationsskala von Schwarzer (1999) eingesetzt. Es ist mit einer eher schwachen Korrelation zwischen Prokrastination und Klausurleistung von $r = -.22$ bis $r = -.08$ zu rechnen. Dieser Bereich dient als Anhaltspunkt zur Überprüfung der Repräsentativität der Stichprobe, die den Untersuchungen der vorliegenden Arbeit zugrunde liegt und bezüglich der Zusammenhänge zu Klausurerfolg und -leistung mit den entsprechenden Werten der anderen Prädiktoren verglichen werden.

3 Übergreifende Fragestellung der Arbeit und Forschungsfragen

Ausgehend von der Identifikation prädiktiv isolierbarer (Grund-)Anforderungen in Klausuren der Ingenieurmathematik in Kapitel 1 werden nun die Forschungsfragen der vorliegenden Arbeit formuliert. Diese lassen sich drei Bereichen zuordnen: Im ersten Bereich wird die Stärke der Zusammenhänge zwischen Klausurergebnis und fachspezifischen, allgemein kognitiven sowie affektiven Faktoren untersucht (Forschungsfragen 1 und 2). Der zweiter Bereich umfasst die Analyse prozeduraler Performanz von Studierenden im Quer- und Längsschnitt sowie in Einzelfallanalysen (Forschungsfragen 3 bis 5). Der dritte Bereich behandelt die Vergleichbarkeit der den Untersuchungen zu Forschungsfrage 1 und 2 zugrunde liegenden Referenzklausur mit anderen Klausuren der Ingenieurmathematik (Forschungsfrage 6). Die drei Bereiche mit ihren sechs Forschungsfragen lassen sich der übergreifenden Frage *„Welche Hintergründe haben Erfolg und Misserfolg in Mathematikklausuren der ingenieurwissenschaftlichen Studiengänge?“* unterordnen.

Das Ziel des Kapitels ist damit eine ausführlichere Darstellung der übergreifenden Fragestellung sowie die Formulierung der Forschungsfragen und deren theoretische Einordnung. Um die Relevanz der Fragen aufzuzeigen, werden zu jeder Forschungsfrage mögliche Konsequenzen skizziert, die sich nach der Diskussion der Frage ergeben können. Eine vertiefte Diskussion der Konsequenzen bleibt dem Schlussteil der vorliegenden Arbeit vorbehalten.

Übergreifende Fragestellung: Welche Hintergründe haben Erfolg und Misserfolg in Mathematikklausuren der ingenieurwissenschaftlichen Studiengänge?

In der Einleitung wurde ausgeführt, dass Mathematikklausuren häufig als Hürde im Studium angesehen werden. Tolan (2013) führt in diesem Zusammenhang eine interne Untersuchung an, gemäß der es sich bei allen Klausuren eines Semesters mit einer Durchfallquote von über 50 % in 18 von 20 Fällen um Mathematikklausuren handelte. Im ersten Kapitel wurde daher betont, dass die Erstellung eines Anforderungsprofils für diese Klausuren relevant ist, beispielsweise um daraus mittelfristig geeignete Unterstützungsmaßnahmen abzuleiten. Im ersten Kapitel wurden anhand der Einschätzungen und Befunde aus vielen

Ländern festgestellt, dass im Anforderungsprofil von Klausuren der Ingenieurmathematik prozedurales Wissen dominiert. Um die Anforderungen genauer zu analysieren, wurden bestehende Definitionen für prozedurales Wissen und konzeptuelles Wissen so verfeinert, dass sich prädiktiv isolierbare Grundanforderungen in Klausuren der Ingenieurmathematik sowie mögliche Operationalisierungen prozeduralen Wissens ableiten lassen. Auf dieser Grundlage können prozedurales und konzeptuelles Wissen als mögliche fachspezifische Prädiktoren für Klausurerfolg untersucht werden. Die fachspezifischen Faktoren wurden in Kapitel 2 um jeweils zwei allgemein kognitive bzw. affektive Faktoren ergänzt, deren Einflüsse auf Klausurleistungen in der Literatur gut belegt sind.
Auf dieser Basis lässt sich die übergreifende Fragestellung der vorliegenden Arbeit spezifischer formulieren: „Was sind die besten Prädiktoren für den Erfolg in Mathematikklausuren der ingenieurwissenschaftlichen Studiengänge?“

Diese Frage wurde in der Literatur bereits unter verschiedenen Aspekten und für unterschiedliche Studiengänge diskutiert. Rach und Heinze (2013) erörtern beispielsweise, welche Lernstrategien in der Studieneingangsphase erfolgreich sind. Trapmann, Hell, Weigand und Schuler (2007) untersuchen in einer Meta-Studie die Prognosegüte von Schulnoten zur Vorhersage von Studienerfolg. Die Meta-Analyse von Hell, Trapmann und Schuler (2007) befasst sich mit der Studienerfolgsprognose durch fachspezifische Studierfähigkeitstests. Darüber hinaus existieren viele Untersuchungen, die sich mit Zusammenhängen zwischen Prüfungserfolg und weiteren Faktoren befassen, wie etwa psychologischen, soziodemographischen und institutionellen Merkmalen (z. B. Stankov & Lee, 2014; Parker, Marsh, Ciarrochi, Marshall & Abduljabbar, 2014). Es existiert jedoch keine Studie, die Zusammenhänge zwischen fachspezifischen Einzelprädiktoren in Form von prädiktiv isolierbaren Grundanforderungen und Erfolg in Klausuren der Ingenieurmathematik analysiert. Es existieren auch keine Untersuchungen zur Prognosegüte solcher fachspezifischen Einzelprädiktoren bezüglich des Klausurerfolgs. Diese Lücke schließt die vorliegende Arbeit im Rahmen der ersten beiden Forschungsfragen. In den Forschungsfragen 3 bis 5 folgen ausführliche Quer- und Längsschnittuntersuchungen sowie qualitative Analysen zum prozeduralen Wissen von Studierenden, da sich dieses hier als stärkster Einzelprädiktor für den Erfolg in Klausuren der Ingenieurmathematik erwiesen hat und prozedurales Wissen zudem Bestandteil des Anforderungsprofils von Klausuren der Ingenieurmathematik ist. Die letzte Forschungsfrage 6 befasst sich schließlich mit der Vergleichbarkeit der Referenzklausur mit anderen Klausuren der Ingenieurmathematik. Die Forschungsfragen werden im Folgenden detailliert vorgestellt.

Forschungsfrage 1: Wie groß sind die Zusammenhänge zwischen Klausurleistung und fachspezifischen, allgemein kognitiven sowie affektiven Faktoren?

In Kapitel 2 wurden die in die vorliegende Arbeit einbezogenen allgemein kognitiven und affektiven Prädiktoren vorgestellt. Deren Zusammenhänge mit Prüfungserfolg sind in der hochschuldidaktischen Literatur bereits gut erforscht. Daher konnten in Kapitel 2 die zu erwartenden Korrelationen zwischen den genannten Prädiktoren und der Klausurleistung angegeben werden. Diese Korrelationen werden im Rahmen der Untersuchungen von Forschungsfrage 1 zum Abgleich dienen, ob die hier untersuchte Stichprobe mit der aus anderen Studien vergleichbar ist. Eine Vergleichbarkeit würde die Generalisierbarkeit der Ergebnisse erleichtern, da nicht auf Besonderheiten der Stichprobe eingegangen werden muss.

Wie eingangs bereits erwähnt, fehlen dagegen Untersuchungen zu Korrelationen zwischen fachspezifischen Einzelprädiktoren und Erfolg in Klausuren der Ingenieurmathematik. Um diese Forschungslücke zu schließen, werden im Rahmen von Forschungsfrage 1 auch fachspezifische Prädiktoren untersucht.

Forschungsfrage 1 wird exemplarisch anhand einer Auswahl von Faktoren aus den genannten Bereichen (fachspezifisch, allgemein kognitiv, affektiv) untersucht. Dazu werden verschiedene Korrelationsanalysen zwischen diesen Faktoren und der Klausurleistung durchgeführt. Als fachspezifische Faktoren werden prozedurales und konzeptuelles Wissen (Kapitel 1), als allgemein kognitive Faktoren dagegen kognitive Grundfertigkeit (Abschnitt 2.1) und Konzentration (Abschnitt 2.2) und als affektive Faktoren Prüfungsangst (Abschnitt 2.3) und Prokrastination (Abschnitt 2.4) betrachtet. Nach der Bestimmung der Korrelationen mit der Klausurleistung können diese miteinander verglichen werden. Von besonderem Interesse ist dabei, ob und in welchem Ausmaß sich die Korrelationen unterscheiden und ob es Faktoren gibt, die herausragend hoch mit der Klausurleistung korrelieren und sich damit besonders gut als Prädiktoren eignen. Weiterhin ist von Interesse, wie groß die Zusammenhänge zwischen Klausurleistung und prozeduralem Wissen sind, da Letzteres durch prädiktiv isolierbare Grundanforderungen in Klausuren operationalisiert wird.

Die Untersuchung von Forschungsfrage 1 kann beispielsweise dabei helfen, (kontinuierliche) Leistungsüberprüfungen (Christoforou & Yigit, 2008; Cole & Spence, 2012; Shorter & Young, 2011) und andere diagnostische Maßnahmen als Teil eines Qualitätskreislaufs oder Modells zur Beschreibung von Erfolg und Fortschritt im Studium (Beekhoven, De Jong & Van Hout, 2002) zu entwickeln und zu bewerten. Zwei Beispiele sollen verdeutlichen, inwiefern das geschehen könnte: Sollte zum Beispiel Prüfungsangst besonders stark mit dem Erfolg in Klausuren der Ingenieurmathematik zusammenhängen, könnten weiterführende Überlegungen gerechtfertigt sein, wie betroffene Studierende frühzeitig unterstützt werden können. Falls hingegen prozedurales Wissen hoch mit der Klausurleistung korreliert, wäre dadurch dessen Berücksichtigung in Leistungsüberprüfungen und bei der Entwicklung entsprechender Testinstrumente zu begründen.

Forschungsfrage 2: In welchem Maße lassen sich Klausurerfolg und Klausurleistung vorhersagen?

Laut Formazin, Schroeders, Köller, Wilhelm und Westmeyer (2011, S. 221) sind an deutschen Hochschulen standardisierte Leistungstests wie etwa Einstufungs- und Studierfähigkeitstests zur Vorhersage des Studienerfolgs nach wie vor eher die Ausnahme als die Regel, da in den meisten Fächern gute Prädiktoren zur Vorhersage von Studienerfolg oder Klausurerfolg fehlen. In Ausnahmefällen wird lediglich die Schulabschlussnote herangezogen, obwohl die Berücksichtigung von Einstufungs- und Studierfähigkeitstests die Prognosegüte weiter verbessern kann (Trost et al., 1997; Kobrin et al., 2008). Es fehlt damit insbesondere die Berücksichtigung fachspezifischer Einzelprädiktoren zur Prognose von Erfolg in Klausuren der Ingenieurmathematik. Diese Forschungslücke wird mit der Untersuchung von Forschungsfrage 2 geschlossen. Während in Forschungsfrage 1 Korrelationen zwischen einzelnen Prädiktoren und der Klausurleistung untersucht werden, werden diese nun in ein gemeinsames (multivariates) lineares sowie in ein logistisches Regressionsmodell aufgenommen. Unter den Prädiktoren befinden sich insbesondere die in Kapitel 1 als Bestandteile des Anforderungsprofils von Klausuren der Ingenieurmathematik identifizierten fachspezifischen Faktoren prozedurales und konzeptuelles Wissen. Ziel ist es, durch Regressionsanalysen die Prognosegüte und erklärte Varianz des Klausurergebnisses zu bestimmen und daraus schließlich den unter den eingesetzten Instrumenten besten Prädiktor oder eine optimale Kombination von Prädiktoren für Klausurerfolg (bestanden versus nicht bestanden) und Klausurleistung (erreichte Klausurpunktzahl) abzuleiten. Zusätzlich zu den bereits in Forschungsfrage 1 berücksichtigten Prädiktoren wird in dem linearen Regressionsmodell der Zusammenhang zwischen der Klausurleistung und weiteren Hintergrundmerkmalen, wie Studiengang, Schuldauer, Wahl von Mathematik als Leistungs- oder Grundkurs, Geschlecht sowie Schultyp untersucht. Dadurch ist es möglich, den Einfluss dieser Hintergrundmerkmale mit dem Einfluss der genannten Prädiktoren auf die Zielvariable (Klausurleistung) zu vergleichen. Während Erhebungen für die Prädiktoren relativ aufwändig sind, lassen sich Hintergrundmerkmale vergleichsweise einfach abfragen. Sollten sich Hintergrundmerkmale ebenso gut zur Vorhersage von Klausurleistung eignen wie die untersuchten Prädiktoren, könnten aufwendige Erhebungen entfallen bzw. diese sich auf eine Abfrage von Hintergrundmerkmalen beschränken.

Im Rahmen der Untersuchungen von Forschungsfrage 2 wird auch die Aufspaltung von prozeduralem Wissen in Kalkülkenntnis und Kalkülfertigkeit empirisch begründet. For-

schungsfrage 2 liefert damit zusätzlich zu dem in Kapitel 1 aufgespannten theoretischen Rahmen einen empirisch fundierten Beitrag zur Konzeptualisierung prozeduralen Wissens in der Ingenieurmathematik.
Die Untersuchung von Forschungsfrage 2 sollte die Befunde zu Forschungsfrage 1 stützen. Zusätzlich ermöglicht sie eine differenzierte Sicht auf die Prognosegüte der Prädiktoren. Von besonderem Interesse ist hierbei analog zu Forschungsfrage 1, wie gut sich prozedurales Wissen, das durch die Fertigkeit, Basisprozeduren effizient und korrekt auszuführen operationalisiert wird und damit nur einen kleinen Teil des komplexen Anforderungsprofils von Klausuren der Ingenieurmathematik abdeckt, bereits als Prädiktor für Klausurleistung eignet. Ferner ist von Interesse, wie fachspezifische Prädiktoren im Vergleich zu den anderen Prädiktoren abschneiden, da erstere zur Prognose von Erfolg in Mathematikklausuren ingenieurwissenschaftlicher Studiengänge bisher nicht untersucht wurden. Die Untersuchung von Forschungsfrage 2 liefert damit erstmals Vergleichswerte in diesem Bereich.

Forschungsfrage 3: Welche Merkmalsverteilungen weisen Gruppen prozedural schwacher, prozedural durchschnittlicher und prozedural starker Studierender auf?

Aus der Literatur ist bekannt, dass Studierende mit einem Leistungskurs in Mathematik bessere Mathematikleistungen erzielen als Studierende, die einen Grundkurs Mathematik besuchten (Maag Merki, 2012; Büning, 2004). Dies gilt auch für Absolventinnen und Absolventen eines Gymnasiums, die im Durchschnitt leistungsstärker sind als Studierende, die eine Gesamtschule besucht haben (Trautwein, Köller, Lehmann & Lüdtke, 2006; Baumert, 1996). Analoge Vergleiche gibt es für zahlreiche weitere Hintergrundmerkmale. Es wurde allerdings noch nicht untersucht, ob sich entsprechende Unterschiede auch im prozeduralen Wissen erkennen lassen. Dies geschieht im Rahmen von Forschungsfrage 3. Dazu werden Studierende in Abhängigkeit von ihrem prozeduralen Wissens in drei Leistungsgruppen eingeteilt. Das Vorgehen entspricht etwa dem bei Einstufungstests an Universitäten in den USA („placement test“, siehe z. B. Denny, Nelson & Zhao, 2012; Madison et al., 2015; Melguizo, Kosiewicz, Prather & Bos, 2014). Anschließend wird untersucht, wie die in Forschungsfrage 1 aufgeführte Auswahl an fachspezifischen, allgemein kognitiven und affektiven Faktoren sowie Klausurergebnisse und die in Forschungsfrage 2 genannten Hintergrundmerkmale in den Leistungsgruppen verteilt sind. Ziel ist es, signifikante Unterschiede zwischen den Leistungsgruppen sichtbar zu machen. Die Unterschiede werden dann mithilfe inferenzstatistischer Methoden auf ihre Bedeutsamkeit untersucht.
Die Beantwortung von Forschungsfrage 3 kann versteckte Schwächen und Stärken innerhalb der Leistungsgruppen aufdecken und auf besonders heterogene Strukturen hinweisen.

Aus den USA ist beispielsweise bekannt, dass durch die Heterogenität der Studierenden ein unterschiedlicher Bedarf an Wiederholungskursen besteht (Meiner & Seiler, 2009, S. 3). Die aufgedeckten Unterschiede könnten in einem ähnlichen Rahmen bei einer zielgruppenspezifischen, adaptiven Gestaltung von Unterstützungsangeboten sowie ggf. bei einer heterogenitätssensitiven Gestaltung des Lehrbetriebs helfen. Falls beispielsweise auffällt, dass Studierende eines Studiengangs überproportional häufig zur Gruppe der Studierenden mit einem geringen prozeduralen Wissen gehören, ergeben sich daraus mehrere Konsequenzen. Einerseits sollte weitergehend untersucht werden, warum gerade dieser Studiengang betroffen ist, andererseits sollte überprüft werden, ob eine Unterstützung von Studierenden des betroffenen Studienganges auch in anderen Bereichen sinnvoll bzw. erforderlich ist.

Forschungsfrage 4: Wie entwickelt sich prozedurales Wissen bei ausgewählten Studierendengruppen während der ersten Studiensemester?

Unter den untersuchten Faktoren wird sich prozedurales Wissen als stärkster Einzelprädiktor für Klausurerfolg und -leistung erweisen. Damit bestätigt sich die in Kapitel 1 geäußerte These, dass prozeduralem Wissen in der ingenieurwissenschaftlichen Mathematikausbildung eine besondere Rolle zukommt. Aus diesem Grund erscheint es gerechtfertigt und erforderlich, prozedurales Wissen weiter zu erforschen. Bisher gibt es jedoch keine Untersuchungen, wie sich das prozedurale Wissen von Studierenden innerhalb der ersten Semester entwickelt. Insbesondere ist unbekannt, ob es wie zum Beispiel in der Schulleistungsstudie PALMA mit der Zeit zu einer Abnahme der Leistung bei schwächeren Studienteilnehmenden kommt (Pekrun et al., 2006, S. 34), was ggf. Auswirkungen auf eine frühzeitige Unterstützung dieser Studierenden haben könnte. Nachdem die Gruppe der prozedural schwachen, durchschnittlichen und starken Studierenden in Forschungsfrage 3 im Querschnitt analysiert worden ist, wird nun die Entwicklung des prozeduralen Wissens innerhalb dieser Leistungsgruppen über einen Zeitraum von zwei bzw. drei Semestern im Längsschnitt verfolgt. Von Interesse ist, ob sich prozedurales Wissen in den Gruppen unterschiedlich entwickelt und worin ggf. die Unterschiede bestehen.
Auf diese Weise können z. B. Unterstützungsangebote zum Aufbau prozeduralen Wissens differenziert bezüglich der Zielgruppen und der Art der Schwächen, d. h. passgenauer, gestaltet werden. Ferner lassen sich möglicherweise auch Aussagen über sinnvolle Zeitpunkte für Interventionen treffen. Die genannten Aspekte sind wichtig, weil sie sich positiv auf die Bestehensquoten in Klausuren auswirken können.

Forschungsfrage 5: Woran scheitern prozedural schwache Studierende bei der Durchführung von Basisprozeduren und wie entwickeln sich individuelle Ressourcen und Schwächen?

Mit dieser Frage wird die Analyseebene von den Studierenden als Gruppe auf individuelle Bearbeitungen prozeduraler Aufgaben bzw. deren Bestandteile, die Basisprozeduren und ihre Anforderungen an Kalkülkenntnis und -fertigkeit, verlagert. Aus der Literatur ist zwar bekannt, welche Schwächen Studierende in der Kalkülfertigkeit aufweisen (Kersten, 2015), ob diese Schwächen aber zeitlich stabil sind oder nach einigen Semestern verschwinden, ist bisher nicht erforscht. Das gilt auch für die Veränderung von Schwächen in der Kalkülkenntnis.

Im Rahmen von Forschungsfrage 5 wird daher für die besonders „gefährdete“ Gruppe der prozedural schwachen Studierenden exemplarisch untersucht, wie sich Ressourcen und Schwächen bei der Bearbeitung einer Basisprozedur innerhalb von drei Semestern verändern, woran diese Studierenden im Detail scheitern und ob Ressourcen und Schwächen individuell verschieden sind. Diese Analysen sind wichtig, da sie Hinweise für die Gestaltung von Unterstützungsangeboten liefern können, beispielsweise ob eine Förderung eher individuell gestaltet werden sollte, anstatt allen betroffenen Studierenden dieselbe Unterstützung zukommen zu lassen. Die Untersuchung von Forschungsfrage 5 kann auch Rückschlüsse darauf liefern, ob und welche Anforderungen von Basisprozeduren möglicherweise in Vergessenheit geraten und eventuell öfter aufgefrischt werden müssten (vgl. Engelbrecht, Harding & Du Preez, 2007). Das ist wichtig, weil der sichere und effiziente Umgang mit Basisprozeduren sowohl in allen Mathematikklausuren der ersten Semester als auch generell in einem ingenieurwissenschaftlichen Studium relevant ist (Carr, Bowe & Ní Fhloinn, 2013). Darüber hinaus weist Niss (2003) sehr deutlich darauf hin, dass mathematische Basisfertigkeiten und Faktenwissen eine notwendige Grundlage für den Aufbau mathematischer Kompetenz darstellen: „Necessary, but certainly not sufficient, prerequisites for mathematical competence are lots of factual knowledge and technical skills“ (Niss, 2003, S.6). Um diese Grundlage auch bei prozedural schwachen Studierenden zu schaffen, sind die Ergebnisse zu Forschungsfrage 5 wichtig.

Forschungsfrage 6: Inwiefern sind die in der Referenzklausur gestellten Anforderungen mit den entsprechenden Anforderungen analoger Klausuren an anderen Hochschulen vergleichbar?

Bei der Untersuchung von Forschungsfrage 6 wird sich herausstellen, dass in der Referenzklausur, die den Untersuchungen dieser Arbeit zugrunde liegt, die Anforderungen an prozedurales Wissen dominieren. Damit bestätigt sich die These einiger Autorinnen und Autoren aus Abschnitt 1.2, denen zufolge prozedurales Wissen eine dominante Rolle in Mathematikklausuren ingenieurwissenschaftlicher Studiengänge zukommt. Diese These wird weiter gestützt durch Untersuchungen von Bergqvist (2007), allerdings lediglich für schwedische Hochschulen. Ob sie auch in Deutschland auf Mathematikklausuren der ingenieurwissenschaftlichen Studiengänge zutrifft, ist bisher nicht erforscht. Mit der Untersuchung von Forschungsfrage 6 soll diese Forschungslücke geschlossen werden. Es wird daher untersucht, ob die der vorliegenden Arbeit zugrunde liegende Referenzklausur mit analogen Mathematikklausuren an anderen deutschen Hochschulen bezüglich ihrer Anforderungen insbesondere an prozedurales Wissen vergleichbar ist. Falls auch in diesen Klausuren ähnliche Anforderungen gestellt werden, ist dies eine Stütze für die begründete Hypothese, dass die Ergebnisse der ersten beiden Forschungsfragen auf andere Hochschulen übertragbar sein müssten. Die Untersuchungen von Forschungsfrage 6 sind damit relevant für die weitere Diskussion, inwiefern prozedurales Wissen in Klausuren der Ingenieurmathematik über den lokalen Rahmen hinaus eine wichtige Rolle spielt und wie aus der vorliegenden Arbeit resultierende Anknüpfpunkte für weitere Forschungsarbeiten hinsichtlich ihrer Generalisierbarkeit zu bewerten sind.

Fazit

Alle Forschungsfragen basieren auf der Analyse ausgewählter fachspezifischer, allgemein kognitiver und affektiver Faktoren. Während sich die ersten beiden Forschungsfragen mit den Zusammenhängen zwischen diesen Faktoren und dem Klausurerfolg bzw. der Klausurleistung beschäftigen, wenden sich die Forschungsfragen 3 und 4 in Quer- und Längsschnittanalysen den „betroffenenen“ Studierenden zu. Forschungsfrage 5 fokussiert indes durch die qualitative Analyse von Bearbeitungen einer Basisprozedur auf inhaltliche Anforderungen prozeduraler Aufgaben, die – wie sich durch Beantwortung von Forschungsfrage 6 zeigen wird – einen großen Teil in Mathematikklausuren für ingenieurwissenschaftliche Studiengänge in Deutschland und darüber hinaus (vgl. Bergqvist, 2007) ausmachen. Es geht also um (1) Zusammenhänge zwischen verschiedenen Faktoren und Klausurergebnissen von (2) Studierenden, die bestimmte (3) Anforderungen erfüllen müssen, was wiederum Auswirkungen auf die Relevanz der Faktoren (1) hat. Damit beleuchten die Forschungsfragen verschiedene Bereiche eines Wirkungskreislaufs im „student life cycle“ (Mooraj & Zervakis, 2014).

4 Forschungsdesign und Methode

Ausgehend von einer Beschreibung möglicher Prädiktoren für den Erfolg in Klausuren der Ingenieurmathematik in Kapitel 1 und 2 wurden in Kapitel 3 die Forschungsfragen der Arbeit formuliert. Im vorliegenden Kapitel werden nun die Methoden und Instrumente der Datenerhebung und -auswertung vorgestellt, die zur Untersuchung der Forschungsfragen dienen. Dabei steht die Darstellung der psychometrischen Qualität der Erhebungsinstrumente im Vordergrund. Die durchgeführten Untersuchungen waren eingebettet in den Veranstaltungsbetrieb der Mathematik für Ingenieure an der TU Dortmund. Dieser wurde flankiert durch Projekte zur Qualitätsverbesserung (siehe z. B. Fakultät für Mathematik der TU Dortmund, 2008), durch die die empirischen Betrachtungen inspiriert wurden.

Das Kapitel ist wie folgt gegliedert: Zunächst wird ein Überblick über die Messinstrumente gegeben. Neben den Erhebungszeitpunkten, einer exemplarischen Darstellung der Stichprobenzusammensetzung und einer Vorstellung der Referenzklausur, auf die sich die Korrelations- und Regressionsanalysen in der vorliegenden Arbeit beziehen, wird das Gütekriterium der Objektivität für alle Erhebungsinstrumente erläutert (4.1). Es folgt die Vorstellung der Methoden der Datenerhebung (4.2 bis 4.4) unter besonderer Berücksichtigung der Gütekriterien Reliabilität und Validität. Abschließend erfolgt die Beschreibung der Methoden zur Datenauswertung (4.5).

4.1 Überblick zur Datenerhebung

4.1.1 Überblick zu Instrumenten und Erhebungszeitpunkten

Zur Datenerhebung werden insgesamt sechs Erhebungsinstrumente verwendet. Tabelle 4.1 gibt einen Überblick über diese Instrumente, die in den nachfolgenden Abschnitten eingehend beschrieben werden. Die letzten fünf in Tabelle 4.1 aufgeführten Tests und Fragebögen werden für Querschnittsanalysen verwendet, während der Speed- und Leistungstest prozedurales Wissen zusätzlich Grundlage der Längsschnittuntersuchungen ist. Entsprechend wurde er zu drei verschiedenen Messzeitpunkten eingesetzt. Alle anderen Erhebungen wurden dagegen nur einmal durchgeführt. Die Messzeitpunkte sind in Abbildung 4.1 angegeben.

Tabelle 4.1
Überblick über die eingesetzten Erhebungsinstrumente

Messinstrument	Zuordnung	erfasstes Konstrukt	Operationalisierung
Speed- und Leistungstest prozedurales Wissen	fachspezifisch	prozedurales Wissen	Anzahl korrekter Antworten
		(Schwäche in) Kalkülfertigkeit	Anzahl mangels Kalkülfertigkeit nicht korrekt gelöster Items
		(Schwäche in) Kalkülkenntnis	Anzahl mangels Kalkülkenntnis nicht korrekt gelöster Items
Leistungstest konzeptuelles Wissen	fachspezifisch	konzeptuelles Wissen	Anzahl korrekter Antworten
Konzentrationstest	allgemein kognitiv	Konzentrationsfähigkeit	Arbeitsleistung (Anzahl korrekter Antworten)
			Arbeitsgeschwindigkeit (Anzahl bearbeiteter Items)
			Sorgfalt (Fehleranteil)
Kognitiver Leistungstest	allgemein kognitiv	schlussfolgerndes Denken	Anzahl korrekter Antworten
Prüfungsangst-Fragebogen	affektiv	Prüfungsangst	Summenscore
Prokrastination-Fragebogen	affektiv	Prokrastination	Summenscore

4.1.2 Stichprobenzusammensetzung

Testteilnehmende waren Studierende, die den Veranstaltungsbetrieb zu den Vorlesungen „Höhere Mathematik I“ (Wintersemester 2013/14), „Höhere Mathematik II“ (Sommersemester 2014) bzw. „Höhere Mathematik III“ (Wintersemester 2014/15) besuchten. Rund zwei Drittel davon gehörten der Kohorte der Studienanfängerinnen und -anfänger aus dem Wintersemester 2013/14 an. Alle Studierenden dieser Kohorte hatten einen mindestens zweisemestrigen Vorlesungszyklus in Mathematik zu absolvieren. Aus diesem Grund unterscheidet sich bei den Erhebungen in den ersten beiden Semestern die Zusammensetzung der Stichprobe kaum. Diese wird daher einmalig und stellvertretend für alle Erhebungen, die im Wintersemester 2013/14 und im Sommersemester 2014 stattfanden, in Tabelle 4.2 am Beispiel des Messzeitpunkts MZP0 berichtet.
Da nur drei der sechs einbezogenen Studiengänge einen dreisemestrigen Vorlesungszyklus in Mathematik absolvieren, war im Wintersemester 2014/15 mit einer Veränderung der Stichprobenzusammensetzung zu rechnen. Diese wird daher für den Speed- und Leistungstest prozedurales Wissen in Abschnitt 4.2.1 gesondert ausgewiesen.

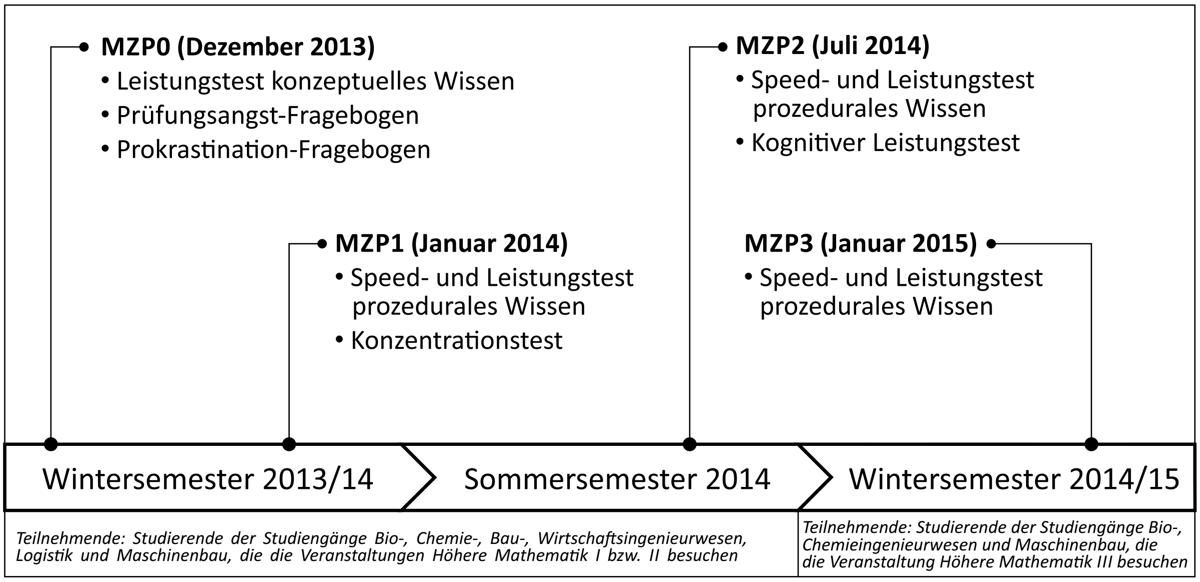


Abbildung 4.1 Zeitpunkte der Erhebungen und Beschreibung der Stichprobe

4.1.3 Objektivität der Erhebungsinstrumente

Alle Erhebungen wurden durch geschulte Testleitungen durchgeführt. Für alle Tests und Fragebögen liegen Testhandbücher vor, in denen die Instruktionen für die Testleitungen eindeutig und verständlich formuliert sind. Auf diese Weise war die Durchführungsobjektivität bei allen Erhebungen sichergestellt. Sofern in den nachfolgenden Abschnitten nicht anders erwähnt, erfolgte die Auswertung der Erhebungen vollautomatisch und EDV-basiert, nachdem die Bearbeitungsbögen eingescannt worden waren. Nicht eindeutig identifizierte Fälle wurden von der Auswertungssoftware zur manuellen Überprüfung vorgelegt. Somit war auch die Auswertungsobjektivität bei allen Erhebungen gewährleistet.

Tabelle 4.2

Stichprobenzusammensetzung zum Messzeitpunkt MZP0 ($N = 1616$)

Merkmal	abs.	%	Merkmal	abs.	%	Merkmal	abs.	%
Geschlecht			Schulform			Studiengang		
männlich	1207	75	Gymnasium	1317	81	Maschinenbau	406	25
weiblich	409	25	Gesamtschule	129	8	Wirtschaftsing.	382	24
Semesterzahl			keine Angabe	170	11	Logistik	228	14
regulär	1000	62	Schuldauer			Bioing.	169	11
höher	616	38	G8	357	22	Chemieing.	263	16
Kursform			G9	1141	71	Bauing.	165	10
GK Mathe	526	33	keine Angabe	118	7	sonstiges	3	0
LK Mathe	1004	62						
keine Angabe	86	5						

*A*nmerkung. abs.=absolute Anzahl; GK=Grundkurs, LK=Leistungskurs

Insgesamt kann damit gemäß Döring und Bortz (2006, S. 195) die Objektivität aller Erhebungsinstrumente als gegeben angesehen werden. Fälle, die durch die vorangegangene Beschreibung nicht abgedeckt sind, beispielsweise weil noch zusätzlich manuelle Auswertungen durchgeführt wurden, werden an entsprechender Stelle in den nachfolgenden Abschnitten gesondert behandelt.

4.1.4 Beschreibung der Referenzklausur „Höhere Mathematik I" zur Erfassung von Prüfungsleistung und Prüfungserfolg

Die Klausur, auf die sich die Fragestellungen zu Prüfungsleistung und -erfolg in der vorliegenden Arbeit beziehen, wurde im Wintersemester 2013/14 gestellt. Es handelt sich um die Klausur „Höhere Mathematik I", die von Studierenden der Studiengänge Maschinenbau, Logistik, Wirtschafts-, Bio-, Chemie- und Bauingenieurwesen im ersten Semester zu absolvieren ist.

Die Klausur bestand aus vier Aufgaben mit jeweils mehreren Teilaufgaben (vgl. Abbildung 4.3). In Aufgabe 1 wurden nur die angegebenen Endergebnisse gewertet, bei den übrigen Aufgaben wurde auch der Lösungsweg berücksichtigt. Insgesamt konnten 80 Punkte erreicht werden, wovon 32 Punkte (40 %) zum Bestehen ausreichten. Die Bearbeitungszeit betrug zwei Stunden.

An der Klausur nahmen 1552 Studierende teil. Die Verteilung der Hintergrundmerkmale entspricht in etwa der in Tabelle 4.2.

Aufgrund der reinen Klausurleistung, d. h. der in der Klausur erzielten Punkte, betrug die Bestehensquote rund 56 %. Tatsächlich lag sie aufgrund angerechneter Bonuspunkte aus vorher erbrachten Leistungen etwas höher. In den Analysen der vorliegenden Arbeit wird allerdings stets die reine Klausurleistung zugrunde gelegt. Die in der Klausur erreichten Punktzahlen (reine Klausurleistung) sind annäherend normalverteilt (vgl. Abbildung 4.2).

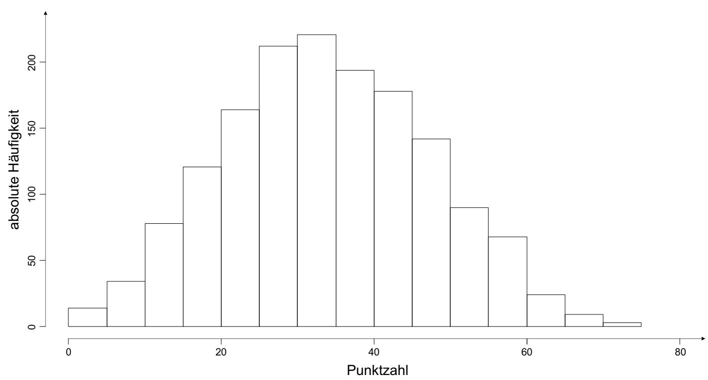


Abbildung 4.2 Häufigkeitsverteilung der erreichten Punkte in der Klausur „Höhere Mathematik I" aus dem Wintersemester 2013/14

Aufgabe 1: (20 Punkte)

a) Es sei $c \in \mathbb{R}$ und $z = \dfrac{1-2i}{2-ic}$. Für $c = \boxed{}$ ist iz reell.

Für $c = \frac{1}{2}$ gilt $\operatorname{Re} z = \boxed{}$ und $\operatorname{Im} z = \boxed{}$.

b) Es seien $A, B \in \mathbb{R}^{n\times n}$ und $\lambda \in \mathbb{C}$. Welche der folgenden Aussagen sind wahr? Nichtzutreffendes bitte deutlich streichen.

- $\det(A+B) = \det A + \det B$ wahr / falsch
- $\det A = \det\left(A^\top\right)$ wahr / falsch
- $\det\left(A - \lambda I\right) = \det\left(A^\top - \lambda I\right)$ wahr / falsch
- $\det(AB) = \det(A)\cdot\det(B)$ wahr / falsch

c) Es seien $v_1 = \frac{1}{\sqrt{2}}(1,0,1)^\top, v_2 = (0,1,0)^\top$. Bestimmen Sie $v_3 \in \mathbb{R}^3$ so, dass $\{v_1, v_2, v_3\}$ eine Orthonormalbasis des $\mathbb{R}^3$ ist. $v_3 = \boxed{\left(\quad , \quad , \quad \right)^\top}$

Es sei $y := (1,1,0)^T$ dargestellt als $y = \lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3$. Dann gilt $\lambda_1 = \boxed{}$, $\lambda_2 = \boxed{}$ und $\lambda_3 = \boxed{}$.

d) $\displaystyle\lim_{n\to\infty} \frac{2n^3 - 2n^2}{5n^3 + 5\sqrt{n}} = \boxed{}$, $\displaystyle\lim_{n\to\infty} \sqrt[n]{3^n + n^5 + 2^n + 2} = \boxed{}$.

e) Für $q(n) := \displaystyle\sum_{k=0}^{n} \left(\frac{1}{2}\right)^k$ gilt $q(n) = \dfrac{u}{v}$ mit $\dfrac{u}{v} = \boxed{}$ und $\displaystyle\lim_{n\to\infty} q(n) = \boxed{}$.

f) Es sei $U := \left\{x \in \mathbb{R}^3 \ : \ x = (1,0,1)^\top + \lambda(a,b,-1)^\top, \lambda \in \mathbb{R}\right\}$. Wie sind a und b zu wählen, so dass U ein Unterraum des $\mathbb{R}^3$ ist? $a = \boxed{}$ und $b = \boxed{}$.

Bonusaufgabe (2 Punkte)

Das Polynom $P(z) = z^3 - 39z + 70$ hat eine Nullstelle bei $z_1 = 2$ sowie zwei weitere Nullstellen $z_2, z_3 \in \mathbb{C}$. Es gilt $z_2 = \boxed{}$ und $z_3 = \boxed{}$.

Aufgabe 2: (20 Punkte)

a) Gegeben sind die folgenden Teilmengen des $\mathbb{R}^2$:

$$A = \{(x,y) \in \mathbb{R}^2 \,|\, x > 0 \text{ und } y > 0\}, \quad B = \{(x,y) \in \mathbb{R}^2 \,|\, 0 \le x \le 1 \text{ und } 0 < y \le 1\},$$
$$C = \{(x,y) \in \mathbb{R}^2 \,|\, (x-1)^2 + (y-1)^2 < 1\}.$$

Skizzieren Sie

i) A, B und C,

ii) $A \setminus B$ und $(A \setminus B) \cup C$.

Achten Sie insbesondere auf eine korrekte Beschriftung der Koordinatenachsen und der Achsenmarkierungen.

b) Zeigen Sie für alle $n \in \mathbb{N}$ die Identität $\displaystyle\sum_{k=1}^{n} \frac{1}{k(k+1)} = \frac{n}{n+1}$.

c) Sei das komplexe Polynom $P(z) = z^6 - z^5 + 4z^4 - 6z^3 + 2z^2 - 8z + 8$ gegeben.

i) Bestimmen Sie $b, c \in \mathbb{C}$ so, dass

$$P(z) = (z^3 - 2)(z^2 + b)(z - c)$$

gilt. *Hinweis: Verwenden Sie einen Koeffizientenvergleich.*

ii) Berechnen Sie drei komplexe Nullstellen von P.

d) Zeigen Sie für $n \in \mathbb{N}_0$ beliebig und alle $z = \cos\alpha + i\sin\alpha, \alpha \in \mathbb{R}$ die Identität

$$z^n + \frac{1}{z^n} = 2\cos(n\alpha).$$

Geben Sie bei Ihren Rechnungen an, welche Eigenschaften Sie benutzen.
Hinweis: Benutzen Sie die Formel von Moivre.

Abbildung 4.3 Klausur „Höhere Mathematik I" aus dem Wintersemester 2013/14 (Fortsetzung auf der nächsten Seite)

Aufgabe 3: (20 Punkte)

a) Vervollständigen Sie die Definitionen:

i) $v_1, \dots, v_s \in \mathbb{R}^n$ heißen **linear unabhängig** genau dann, wenn ...

ii) Seien $v_1, v_2, v_3 \in \mathbb{R}^n$ gegeben. Dann ist $\mathbf{span}(v_1, v_2, v_3) = \{\dots$

b) Sei $A = \begin{pmatrix} 1 & 0 & 4 \\ 0 & -3 & 3 \\ 4 & 3 & 12 \end{pmatrix} \in \mathbb{R}^{3\times 3}$ gegeben.

i) Berechnen Sie $\det(A)$.

ii) Ist A regulär? Begründung!

iii) Bestimmen Sie A^{-1}.

iv) Bestimmen Sie jeweils die Lösungsmenge von $Ax = b$

$$\text{mit} \qquad (1) \quad b = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \qquad\qquad (2) \quad b = \begin{pmatrix} 1 \\ 3 \\ 1 \end{pmatrix}.$$

Hinweis: Da es verschiedene Lösungswege gibt, muss klar erkennbar sein, wie Sie auf Ihr Ergebnis gekommen sind. Die alleinige Angabe einer Lösung ohne kurze Begründung/Rechnung reicht also nicht aus.

v) Bestimmen Sie $\mathrm{Bild}(A)$ sowie die Dimension von $\mathrm{Bild}(A)$ und geben Sie alle Elemente aus $\mathrm{Kern}(A)$ an.

vi) Geben Sie eine ON-Basis von $\mathrm{Bild}(A)$ an.

c) Zeigen Sie durch Nachprüfen der Unterraumkriterien bzw. widerlegen Sie durch ein Gegenbeispiel, dass es sich bei folgenden Mengen um Unterräume handelt:

i) $U_1 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3 \mid x_1 + x_2 = x_3 \right\} \subset \mathbb{R}^3$

ii) $U_2 = \big\{ p : \mathbb{R} \to \mathbb{R} \mid p(x) = ax^2 + bx + c \text{ mit } a, b, c \in \mathbb{R} \setminus \{0\} \big\} \subset \mathbb{P}_2$

(hierbei ist $\mathbb{P}_2$ der Vektorraum der reellen Polynome vom Grad kleiner oder gleich 2).

Aufgabe 4: (20 Punkte)

a) Betrachten Sie die Matrix

$$A = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix} \in \mathbb{R}^{3\times 3}.$$

i) Welche Eigenwerte besitzt A?

ii) Bestimmen Sie eine ON-Basis des $\mathbb{R}^3$ aus Eigenvektoren von A.

iii) Bestimmen Sie die Koordinaten des Vektors $\mathbb{R}^3 \ni x = (1, -2, 4)^\top$ bzgl. der in ii) berechneten ON-Basis.

iv) Geben Sie den Eigenraum zum Eigenwert 1, d.h. $E(1)$, in Hessescher Normalform an.

v) Begründen Sie, dass die Abbildung $\mathbb{R}^3 \ni x \mapsto Ax \in \mathbb{R}^3$ geometrisch eine Spiegelung darstellt.

b) Weisen Sie nach, dass für beliebige Matrizen $A \in \mathbb{R}^{n\times n}$ gilt: A und $A^\top$ besitzen identische Eigenwerte. Hinweis: charakteristisches Polynom.

c) Geben Sie eine Matrix $A \in \mathbb{R}^{2\times 2}$ an, mit der Sie belegen, dass die Eigenwerte von A und $A^\top$ übereinstimmen, die Eigenräume zu den Eigenwerten jedoch im Allgemeinen nicht.

4.2 Instrumente der Datenerhebung für fachspezifische Faktoren

4.2.1 Speed- und Leistungstest prozedurales Wissen

In Kapitel 1 wurde prozedurales Wissen als Teil des komplexen Anforderungsprofils von Klausuren der Ingenieurmathematik identifiziert. Exemplarisch wurde festgestellt, dass hier das Anforderungsspektrum an prozedurales Wissen von der Anwendung einfacher Basisprozeduren bis hin zur Durchführung komplexer Superprozeduren reicht. Im Rahmen des Speed- und Leistungstests prozedurales Wissen wird prozedurales Wissen durch Basisprozeduren operationalisiert. Diese stellen eine Grundanforderung in Klausuren (der Ingenieurmathematik) dar, deren Anforderungsprofil weit über diese Grundanforderungen hinausreicht. Im Folgenden wird der Test detailliert beschrieben.

Testbeschreibung

Der Speed- und Leistungstest prozedurales Wissen ist ein vom Autoren entwickeltes Messinstrument. Der Test dient zur Erfassung des prozeduralen Wissens von Studierenden und misst die Fertigkeit, Basisprozeduren (vgl. Abschnitt 1.1.1), die in der Höheren Mathematik für Ingenieure unterrichtet werden, in angemessener Zeit fehlerfrei durchführen zu können. Dabei wurde als angemessene Zeit das Doppelte der Arbeitszeit festgelegt, die bei Testrechnungen im Durchschnitt von drei Tutorinnen und Tutoren benötigt wurde, die mit der entsprechenden Basisprozedur vertraut waren. Diese Festlegung folgt damit der zeitlichen Auslegung von Klausuren in der Service-Mathematik für die in Tabelle 4.2 aufgeführten Studiengänge der TU Dortmund im Untersuchungszeitraum.
Neben der Messung der oben genannten Fertigkeit erlaubt die qualitative Auswertung der Testbearbeitungen Einblicke in die Ursachen für nicht korrekte Ergebnisse. Insbesondere können die Facetten prozeduralen Wissens, Kalkülkenntnis und Kalkülfertigkeit (vgl. Abschnitt 1.1.1), der Teilnehmenden differenziert analysiert werden.

Abbildung 4.4 zeigt drei typische Testitems. Allen Items ist gemeinsam, dass sie die Anwendung einer Basisprozedur der Höheren Mathematik unter Zeitbeschränkung erfordern. Das mathematische Modell ist somit vorgegeben und es erfolgt lediglich eine Verarbeitung durch das Ausführen einer Prozedur. Der kognitive Anspruch der Items ist folglich gering, da die Items durch Schema-Wissen lösbar sind. Konzeptuelles Wissen ist zur Lösung der Items nicht erforderlich. In den gezeigten Beispielen handelt es sich (von links nach rechts) um die Basisprozeduren Horner-Schema, Gauß-Algorithmus sowie konjugiert komplexe Erweiterung.

Gruppe A

LTK-13-01-19-FA010-2-Horner 1

Bearbeitungszeit: $\boxed{07{:}30}$ Minuten

Das Polynom

$$P(z) = z^4 + 7z^3 + \frac{73}{4}z^2 + \frac{17}{2}z - \frac{39}{4}$$

hat die Nullstellen $z_1 = \frac{1}{2}$, $z_2 = -\frac{3}{2}$ und $z_3, z_4 \in \mathbb{C}$. Bestimmen Sie z_3 und z_4 **mit Hilfe des Horner-Schemas** und der pq-Formel und berechnen Sie

$$\boldsymbol{z_1 - z_2 - z_3 - z_4}$$

Auftretende Brüche sind so weit wie möglich zusammenzufassen und zu kürzen.

Gruppe B

LTK-13-05-12-03-LGS 4

Bearbeitungszeit: $\boxed{05{:}00}$ Minuten

Das lineare Gleichungssystem

$$\begin{array}{rcrcrcr} x_1 & + & 3x_2 & & & = & -5 \\ 2x_1 & - & x_2 & + & 3x_3 & = & 4 \\ -x_1 & - & 2x_2 & + & x_3 & = & 3 \end{array}$$

hat eine eindeutige Lösung $x_1, x_2, x_3 \in \mathbb{R}$. Berechnen Sie x_1, x_2 und x_3 **mit Hilfe des Gauß-Algorithmus** und anschließend den gesuchten Wert

$$\boldsymbol{x = x_1 + x_2 + x_3}\,.$$

Gruppe B

LTK-13-05-30-1-Komplexe Zahlen 2

Bearbeitungszeit: $\boxed{1{:}30}$ Minuten

Gegeben sei die komplexe Zahl

$$z = \frac{1+5i}{2-i}\,.$$

Berechnen Sie

$$\mathbf{Im}(\boldsymbol{z}) - \mathbf{Re}(\boldsymbol{z})\,.$$

Vereinfachen Sie das Ergebnis durch Zusammenfassen und Kürzen so weit wie möglich.

Abbildung 4.4 Beispiele für Items aus dem Speed- und Leistungstest prozedurales Wissen

Insgesamt umfasst der Itempool 24 Items, die in verschiedenen Testvarianten eingesetzt werden. Jedes Item prüft genau eine der folgenden Basisprozeduren ab: Produktregel, Grenzwertberechnung mit den Regeln von de l'Hospital, Horner-Schema, Polynomdivision, bestimmtes Integral (Integration einer Polynomfunktion), (Bestimmung von Real- und Imaginärteil einer komplexwertigen Bruchzahl durch) konjugiert komplexe Erweiterung, (Lösung linearer Gleichungssysteme/Berechnung von Linearkombinationen mit dem) Gauß-Algorithmus, Matrixinvertierung, Matrizenmultiplikation, Rangbestimmung, Vektorrechnung (Kreuz- und Skalarprodukt), Bruchrechnung. Mit Ausnahme der Bruchrechnung werden alle Basisprozeduren mit expliziten Lerngelegenheiten im Rahmen des Vorlesungszyklus der Ingenieurmathematik behandelt. Alle Basisprozeduren sind üblicherweise ein Hilfsmittel bei der Bearbeitung komplexerer Probleme, beispielsweise im Rahmen von Superprozeduren (vgl. Abschnitt 1.1.1).

Der Test liegt in den Varianten *Semester 1-A*, *Semester 1-B*, *Semester 2-A*, *Semester 2-B* und *Semester 3* vor. Die Bezeichnungen weisen auf den Einsatzzeitpunkt hin (erstes, zweites bzw. drittes Studiensemester bzw. MZP1, MZP2, MZP3, vgl. Abschnitt 4.1.1). Tabelle 4.3 gibt einen Überblick über die Bearbeitungszeiten und die Itemanzahl der Testvarianten. Die Testvarianten sind über Anker-Items verbunden, um eine vertikale Verlinkung für Veränderungsmessungen zu ermöglichen, was ab Seite 81 detailliert beschrieben wird.

Tabelle 4.3

Dauer und Anzahl Items der Varianten des Speed- und Leistungstests prozedurales Wissen

	Testvariante Semester 1-A	Testvariante Semester 1-B	Testvariante Semester 2-A	Testvariante Semester 2-B	Testvariante Semester 3
Dauer in Minuten (gerundet)	81	81	52	54	79
Anzahl Items	12	12	9	9	12

Gruppe A	Gruppe A
LTK-13-01-19-FA010-2-Horner 1 Bearbeitungszeit: $\boxed{07{:}30}$ Minuten Das Polynom $P(z) = z^4 + 7z^3 + \frac{73}{4}z^2 + \frac{17}{2}z - \frac{39}{4}$ hat die Nullstellen $z_1 = \frac{1}{2}$, $z_2 = -\frac{3}{2}$ und $z_3, z_4 \in \mathbb{C}$. Bestimmen Sie z_3 und z_4 **mit Hilfe des Horner-Schemas** und der pq-Formel und berechnen Sie $\boldsymbol{z_1 - z_2 - z_3 - z_4}$ Auftretende Brüche sind so weit wie möglich zusammenzufassen und zu kürzen.	**Antwortoptionen** $z_1 - z_2 - z_3 - z_4 =$ 1 -8 2 -7 3 -6 4 -5 5 -4 6 8 7 5 8 6 9 7 10 4

Abbildung 4.5 Beispielitem „Horner 1“ mit Lösungsoptionen

Zur Kontrolle der Bearbeitungsgeschwindigkeit wird der Test beamerbasiert dargeboten. Dabei wird jede Aufgabe für eine vorgegebene Zeit auf die Projektionsfläche projiziert und die verbleibende Bearbeitungszeit parallel per Counter eingeblendet. 60 Sekunden vor Ablauf der Bearbeitungszeit ertönt ein akustisches Signal, um die teilnehmenden Studierenden darauf aufmerksam zu machen, dass noch eine Minute zur Bearbeitung der Aufgabe zur Verfügung steht. Nach Ablauf der Bearbeitungszeit ertönt wiederum ein akkustisches Signal und es werden für etwa 25 Sekunden zehn Lösungsoptionen mit einer Zahlenzuordnung eingeblendet (vgl. Abbildung 4.5). Unter den Antwortoptionen ist die korrekte Lösung auszuwählen (Single-Choice) und die korrekte Zuordnung auf dem Bearbeitungsbogen zu markieren. Nach Ausblenden der Lösungsoptionen ertönt erneut ein akkustisches Signal und die nächste Aufgabe erscheint. Durch diesen Ablauf ist sichergestellt, dass alle Teilnehmenden zur Bearbeitung jedes Items dieselbe Zeit zur Verfügung haben und Items nicht mehr nachträglich bearbeitet werden können. Die nachgelagerte Einblendung der Antwortoptionen verhindert die Umkehrung von Lösungsprozessen, z. B. durch Einsetzen der Lösungsoptionen in die Aufgabenstellung. Eine Demonstration des Tests kann bei Altieri (2014) aufgerufen werden.

Etwa vier Wochen vor jedem Test wurde eine Informationsbroschüre zum Download bereitgestellt. Die Broschüre enthielt die Testitems mit veränderten Zahlen und vollständigen Lösungswegen, sodass sich alle Studierenden relativ passgenau auf den Test vorbereiten konnten. Ferner wurde in der Broschüre angekündigt, dass falsch angekreuzte Lösungsoptionen mit negativen Punkten gewertet werden und dass Bonuspunkte für die Klausur gesammelt werden können. Ersteres diente zur Reduzierung der Ratewahrscheinlichkeit und Letzteres zur Erhöhung der Motivation, den Test gewissenhaft zu bearbeiten.

Zusammensetzung der Stichproben

Tabelle 4.4 ist die Zusammensetzung der Stichproben der einzelnen Testvarianten zu entnehmen.

Tabelle 4.4

Stichprobenzusammensetzungen beim Speed- und Leistungstest prozedurales Wissen

	Variante Semester 1-A		Variante Semester 1-B		Variante Semester 2-A		Variante Semester 2-B		Variante Semester 3	
Merkmal	abs.	%	abs.	%	abs.	%	abs.	%	abs.	%
N	669	100	617	100	479	100	718	100	495	100
Geschlecht										
männlich	493	74	475	77	356	74	538	75	394	80
weiblich	176	26	142	23	123	26	180	25	101	20
Semesterzahl										
regulär	439	66	367	59	339	71	317	45	271	55
höher	230	34	250	41	140	29	401	55	224	45
Kursform Mathematik										
Grundkurs	199	30	196	32	131	27	235	33	84	17
Leistungskurs	439	66	400	65	329	69	454	63	370	75
keine Angabe	31	4	21	3	19	4	29	4	41	8
Schulform										
Gymnasium	577	86	495	80	420	88	584	81	401	81
Gesamtschule	44	7	45	7	24	5	59	8	22	4
keine Angabe	48	7	77	13	35	7	75	11	72	15
Schuldauer										
G8	168	25	127	21	134	28	115	16	104	21
G9	466	70	447	72	313	65	553	77	338	68
keine Angabe	35	5	43	7	32	7	50	7	53	11
Studiengang										
Maschinenbau	166	25	153	25	121	25	182	25	256	52
Wirtschafts-ingenieurwesen	160	24	157	25	115	24	189	26	0	0
Logistik	92	14	86	14	69	15	127	18	0	0
Bioingenieur-wesen	87	13	50	8	72	15	49	7	85	17
Chemieingen-ieurwesen	91	14	108	18	72	15	104	15	146	29
Bauingenieur-wesen	68	10	62	10	29	6	57	8	5	1
sonstiges	5	1	1	0	1	0	10	1	3	1

Anmerkung. abs.=absolute Anzahl

Itemauswahl für Quer- und Längsschnittanalysen

In Tabelle 4.5 sind alle Items des Itempools und ihre Verteilung auf die einzelnen Testvarianten aufgeführt. Für jedes Item sind Trennschärfe, Lösungsquote sowie die Standardabweichung der Lösungsquote angegeben. Die Trennschärfe gilt als Maß dafür, wie gut ein Item zwischen Personen mit niedriger und hoher Merkmalsausprägung trennt. Sie sollte gemäß Bühner (2011, S. 81) nicht unter .30 liegen. Aus diesem Grund wurden bis auf zwei Ausnahmen alle Items mit einer geringeren Trennschärfe aus den Tests entfernt. Bei den Ausnahmen handelte es sich um Items mit einer grenzwertigen Trennschärfe von .29. Hier fand sich nach eingehender Analyse kein Grund, diese Items zu entfernen. Bei den ausgeschlossenen Items hingegen handelte es sich in 3 von 4 Fällen um Items zur Bruchrechnung. Diese erfordern als einzige Items im Test ausschließlich prozedurales Wissen auf dem Niveau der Sekundarstufe I und passen vermutlich aus diesem Grund nicht zu den übrigen Items. In Tabelle 4.5 sind die eliminierten Items durch „xx“ gekennzeichnet.

Tabelle 4.5

Trennschärfe (TS), Lösungsquote (LQ) und Standardabweichung (SD) der Items

Itembezeichnung	Variante Semester 1-A			Variante Semester 1-B			Variante Semester 2-A			Variante Semester 2-B			Variante Semester 3		
	TS	LQ	SD	TS	LQ	SD	TS	LQ	SD	TS	LQ	SD	TS	LQ	SD
Ableitungen5	.32	.28	.45				.40	.53	.50	.36	.57	.49	.38	.78	.42
Ableitungen6				.39	.13	.34									
Bruchrechnung1										xx	xx	xx			
Bruchrechnung1V1				xx	xx	xx									
Bruchrechnung2	xx	xx	xx				xx	xx	xx						
Grenzwerte1V1										.37	.48	.50			
Grenzwerte2							.36	.22	.42				.36	.57	.50
Horner1	.44	.37	.48	.44	.36	.48	.37	.50	.50	.37	.54	.50	.42	.61	.49
Integration1	.35	.65	.48				.31	.71	.45						
Integration2				.40	.22	.41							.31	.46	.50
KomplexeZahlen2				.40	.39	.49							.38	.59	.49
KomplexeZahlen5	.44	.21	.41				.41	.29	.45	.41	.27	.45			
LGS1	.33	.15	.36				.29	.18	.38	.33	.19	.39	.33	.27	.45
LGS4				.39	.54	.50									
Linearkombination3	.38	.34	.47												
Linearkombination3V1				.37	.29	.45							.32	.44	.50
Matrixinvertierung1	.38	.18	.38	.31	.17	.38							.30	.28	.45
Matrizenmultiplikation2				.36	.34	.47				.38	.45	.50			
Matrizenmultiplikation5	.42	.19	.39				.31	.22	.41				.33	.31	.46
Polynomdivision2V1	.47	.27	.44				.40	.38	.49				.40	.52	.50
Polynomdivision3				.42	.16	.37									
Rangbestimmung1	.39	.51	.50	.34	.50	.50				.35	.60	.49	xx	xx	xx
Vektorrechnung2										.33	.40	.49			
Vektorrechnung3	.40	.35	.48	.43	.38	.49							.29	.55	.50

Reliabilität

Tabelle 4.6 enthält die Reliabilitäten der gemäß Tabelle 4.5 modifizierten Tests. Zur Bestimmung der Split-Half-Reliabilität wird der Test zufällig in zwei Hälften geteilt und die Konsistenz der beiden Hälften zueinander beurteilt. Diese wird im Wesentlichen durch die Korrelation der beiden Testhälften bestimmt (Moosbrugger & Kelava, 2007, S. 130). Teilt man den Test weiter in so viele Teile, wie er Items enthält, resultiert Cronbachs Alpha als Maß für die Konsistenz aller Items der Skala. Diese Kennzahl wird auch als *interne Konsistenz* eines Tests bezeichnet, da sie durch die mittlere paarweise Korrelation der Items bestimmt wird (Steyer & Eid, 1993, S. 131).

Nach George (2007, S. 231) gilt eine Reliabilität ab 0.7 als akzeptabel, ab 0.8 als gut und ab 0.9 als exzellent. Die Reliabilitäten der im ersten und dritten Semester eingesetzten Testvarianten Semester 1-A, Semester 1-B und Semester 3 sind folglich akzeptabel. Die etwas niedrigeren Reliabilitäten der Varianten Semester 2-A und Semester 2-B können auf die geringere Itemzahl zurückgeführt werden, da die Reliabilitätskennzahl Cronbachs Alpha und die Split-Half-Reliabilität von dieser abhängen (Steyer & Eid, 1993, S. 131).

Zur Abklärung wird daher noch ein weiteres Reliabilitäts-Kriterium herangezogen: Die mittlere paarweise Korrelation der Items (*mean interitem correlation*, MIC) sollte zwischen .20 und .40 liegen (Briggs & Cheek, 1986, S. 115). In Tabelle 4.6 sind die MIC für jeden Test angegeben. Da es sich um dichotome Items handelt, wurden tetrachorische Korrelationen zugrunde gelegt (Bortz & Weber, 2005, S. 230). Die Forderung von Briggs und Cheek (1986) ist in allen Fällen erfüllt. Insgesamt gelten damit alle Tests als reliabel.

Tabelle 4.6
Reliabilitäten und mittlere Interitemkorrelation (MIC) der Testvarianten nach Ausschluss von Items mit ungenügender Trennschärfe

	Testvariante Semester 1-A	Testvariante Semester 1-B	Testvariante Semester 2-A	Testvariante Semester 2-B	Testvariante Semester 3
Anzahl Items	11	11	8	8	11
Cronbachs Alpha	.75	.74	.66	.67	.70
Split-Half-Reliabilität	.74	.73	.65	.65	.69
MIC	.37	.36	.34	.34	.29

Rasch-Homogenität

Als Nachweis, dass ein Test genau ein Persönlichkeitsmerkmal (*latent trait*) misst, wird die Gültigkeit des Rasch-Modells, d. h. die Rasch-Homogenität des Tests, angesehen (Strobl, 2012, S. 23). Rasch-homogene Tests werden daher auch als *eindimensionale* Messinstrumente bezeichnet.

Das Rasch-Modell ist ein Modell der probabilistischen Testtheorie. Es wird angenommen, dass das manifeste Lösungsverhalten einer Person bezüglich eines Testitems in probabilistischer Weise von genau einem zugrunde liegenden Persönlichkeitsmerkmal abhängt.
Gelingt es also, einen Rasch-homogenen Test zu konstruieren und das dem Lösungsverhalten zugrunde liegende Persönlichkeitsmerkmal zu identifizieren, lassen sich die im Folgenden beschriebenen Methoden der probabilistischen Testtheorie zur Beschreibung und Analyse des gemessenenen Merkmals anwenden.

Das Rasch-Modell für dichotome Daten wird durch folgende Modellgleichung beschrieben (Koller, Alexandrowicz & Hatzinger, 2012, S. 9):

$$P(X_{vi} = x_{vi} \,|\, \theta_v, \beta_i) = \frac{e^{x_{vi}(\theta_v - \beta_i)}}{1 + e^{\theta_v - \beta_i}} \tag{4.1}$$

Dabei stellen θ_v für $v = 1, \ldots, n$ die Personenfähigkeitsparameter und β_i für $i = 1, \ldots, k$ die Itemschwierigkeitsparameter oder kurz Itemparameter dar. Ferner bezeichnet X_{vi} die Variable für eine Antwort einer Person v auf ein Item i und $x_{vi} \in \{0, 1\}$ die konkrete Ausprägung der Variablen X_{vi}.
Gleichung (4.1) beschreibt damit die Wahrscheinlichkeit, dass die Zufallsvariable X_{vi} bei vorgegebener Personenfähigkeit θ_v und Itemschwierigkeit β_i die Ausprägung x_{vi} annimmt.

Die Wahrscheinlichkeit, dass ein Testergebnis $(x_{11}, \ldots, x_{nk})$ beobachtet wird, ergibt sich im Fall der stochastischen Unabhängigkeit der Einzelereignisse als Produkt dieser Ereignisse, d.h.:

$$P(x_{11}, \ldots, x_{nk} \,|\, \theta_1, \ldots, \theta_n, \beta_1, \ldots, \beta_k) = \prod_{v=1}^{n} \prod_{i=1}^{k} \frac{e^{x_{vi}(\theta_v - \beta_i)}}{1 + e^{\theta_v - \beta_i}} \tag{4.2}$$

Bei der Schätzung der Personen- und Itemparameter wird die Sichtweise umgekehrt und nicht mehr wie in Formel (4.2) auf die genannten Parameter bedingt, sondern auf das gemessene – und damit bekannte – Testergebnis:

$$L(\theta_1, \ldots, \theta_n, \beta_1, \ldots, \beta_k \,|\, x_{11}, \ldots, x_{nk}) = \prod_{v=1}^{n} \prod_{i=1}^{k} \frac{e^{x_{vi}(\theta_v - \beta_i)}}{1 + e^{\theta_v - \beta_i}} \tag{4.3}$$

L wird als Likelihood bezeichnet.

Sofern ein Test Rasch-homogen ist und folglich der Modellgleichung (4.1) genügt, werden die Personen- und Itemparameter bei einem vorgegebenen Testergebnis nun so geschätzt, dass die Likelihood maximal ist.

Nach Kubinger und Draxler (2007, S. 131) existiert kein standardisiertes Vorgehen zur Überprüfung, ob ein Test Rasch-homogen ist. Sie schlagen daher folgende Strategie vor (ebenda, S. 140): Durchführung von Andersons Likelihood-Quotienten-Test (LQT) mit vier verschiedenen Teilungskriterien. Darunter sollte insbesondere eine Stichprobenteilung in Personen mit hohem versus niedrigem Rohscore sein. Als fünften Signifikanztest empfehlen Kubinger und Draxler (2007) die Anwendung des Martin-Löf-Tests mit einer Teilung der Items nach hoher versus niedriger Schwierigkeit.
Das Alpha-Niveau sollte in allen Fällen auf $\alpha = 0.01$ gesetzt werden (Bonferroni-Korrektur, vgl. Koller et al., 2012, S. 164).

Die Größe der Stichproben aller Testvarianten (vgl. Tabelle 4.4) erlaubt die Bestimmung der in Tabelle 4.7 aufgeführten p-Werte mittels Bootstrapping (von Davier, 1997). Dabei wurden 500-mal aus der jeweiligen Stichprobe 75 % der teilnehmenden Studierenden zufällig ausgewählt und die oben genannten Signifikanztests mit dem R-Paket eRm (Mair, Hatzinger & Maier, 2015) durchgeführt.
Beim Likelihood-Quotienten-Test nach Andersonen dienten der Mittelwert der Rohscores, das Geschlecht (männlich versus weiblich), die Kursform (Leistungskurs versus Grundkurs Mathematik) sowie das Semester (Regel versus höheres Semester) als Teilungskriterium. In Tabelle 4.7 sind die mittleren p-Werte sowie die Standardabweichungen angegeben.

Die Signifikanztests wurden in keinem Fall signifikant, d. h. kein p-Wert unterschreitet das oben spzifizierte α-Niveau von 0.01. Damit gelten alle Testvarianten als Rasch-homogen und sind folglich als eindimensionale Messinstrumente anzusehen.

Tabelle 4.7
Rasch-Analysen aller Versionen des Speed- und Leistungstests prozedurales Wissen

	Variante Semester 1-A		Variante Semester 1-B		Variante Semester 2-A		Variante Semester 2-B		Variante Semester 3	
Siginifikanztest (Teilungskriterium)	$\bar{p}$	SD	$\bar{p}$	SD	$\bar{p}$	SD	$\bar{p}$	SD	$\bar{p}$	SD
LQT (Mittelwert)	.02	.03	.05	.06	.23	.18	.38	.21	.14	.13
LQT (Geschlecht)	.24	.18	.34	.23	.52	.22	.23	.17	.34	.18
LQT (Kursform)	.36	.21	.06	.08	.20	.17	.11	.13	.30	.20
LQT (Semester)	.26	.18	.16	.13	.64	.22	.05	.06	.37	.20
Loef (MW Itemschwierigkeit)	.65	.21	.23	.19	.45	.20	.36	.23	.65	.21

*A*nmerkung. $\bar{p}$: Durchschnittlicher p-Wert nach 500 Bootstrap-Runden, SD=Standardabweichung, MW=Mittelwert, LQT=Liklihood-Quotienten-Test nach Anderson

Hauptkomponentenanalyse zur Bestimmung der Anzahl an Hauptkomponenten

In Abbildung 4.6 sind die Screeplots (Cattell, 1966) aller Tests inklusive Parallelanalyse (Horn, 1965) nach einer Hauptkomponentenanalyse (Bortz & Weber, 2005, S. 516) dargestellt. Dabei gelten Eigenwerte oberhalb des simulierten Verlaufs als nicht mehr durch Zufall erklärbar (Horn, 1965). Dies gilt auch für alle tatsächlichen Eigenwerte, die größer als 1 sind (Kaiser-Kriterium nach Kaiser & Dickman, 1959).

Es bestätigt sich in allen Fällen die Eindimensionalität des Speed- und Leistungstests prozedurales Wissen in Gestalt einer dominanten Komponente, die je nach Test zwischen 30 % und 37 % der Varianz der Testergebnisse erklärt. Alle Testitems laden also hoch auf einem einzigen Faktor, der als prozedurales Wissen bezeichnet werden kann (vgl. Abschnitt über die Validität des Tests auf Seite 86).

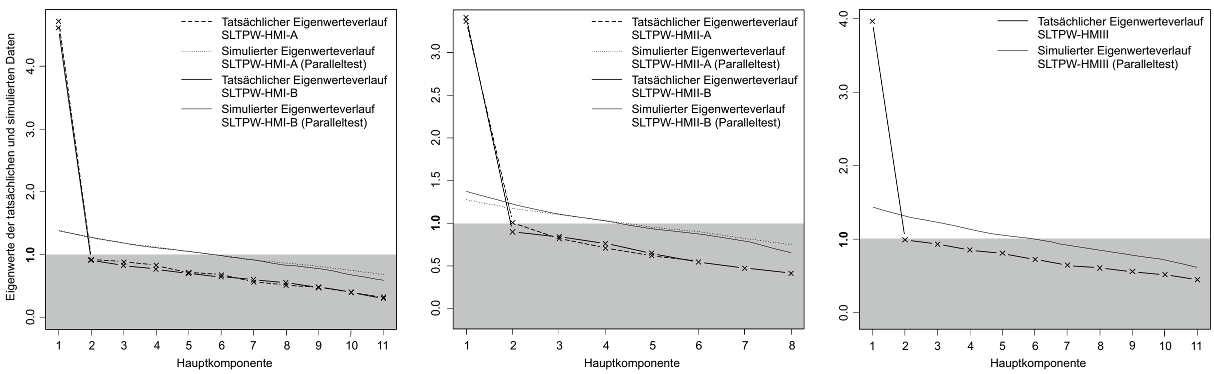


Abbildung 4.6 Parallelanalyse und Screeplots nach Hauptkomponentenanalyse der Tests. SLTPW=Speed- und Leistungstest prozedurales Wissen

Horizontale und vertikale Verlinkung der Testvarianten

Um Veränderungen des prozeduralen Wissens bei Studierenden zu messen, wurde der Speed- und Leistungstest prozedurales Wissen über einen Zeitraum von drei Semestern durchgeführt. Dabei wurden sowohl während der einzelnen Testzeitpunkte als auch zu verschiedenen Testzeitpunkten unterschiedliche Testvarianten eingesetzt (vgl. Tabelle 4.8).

Um die aus den verschiedenen Testvarianten resultierenden Personenparameter vergleichen zu können, ist eine Verortung dieser Parameter auf einer allen Tests gemeinsamen Skala erforderlich. Dies geschieht durch eine geeignete Skalierung der Parameter. Diese Skalierung wird bei Testvarianten, die zum selben Zeitpunkt durchgeführt werden, horizontale Verlinkung oder Equating genannt. Entsprechend wird sie bei Testvarianten, die zu verschiedenen Zeitpunkten dargeboten werden, als vertikale Verlinkung bezeichnet (Patz, 2007, S. 6).

Tabelle 4.8

Zeitpunkt der eingesetzten Varianten des Speed- und Leistungstests prozedurales Wissen

Zeitpunkt	Testvariante Semester 1-A	Testvariante Semester 1-B	Testvariante Semester 2-A	Testvariante Semester 2-B	Testvariante Semester 3
MZP1 (Wintersemester 13/14)	✓	✓			
MZP2 (Sommersemester 14)			✓	✓	
MZP3 (Wintersemester 14/15)					✓

Sowohl für die horizontale Verlinkung als auch für die vertikale Verlinkung kommen dieselben Methoden zum Einsatz. Abbildung 4.7 zeigt eine Auswahl der Entscheidungen, die bei der Methodenwahl zu treffen sind. Die breiter umrandeten Kästchen stellen die Entscheidungen in der vorliegenden Arbeit dar. Sie werden im Folgenden näher begründet.

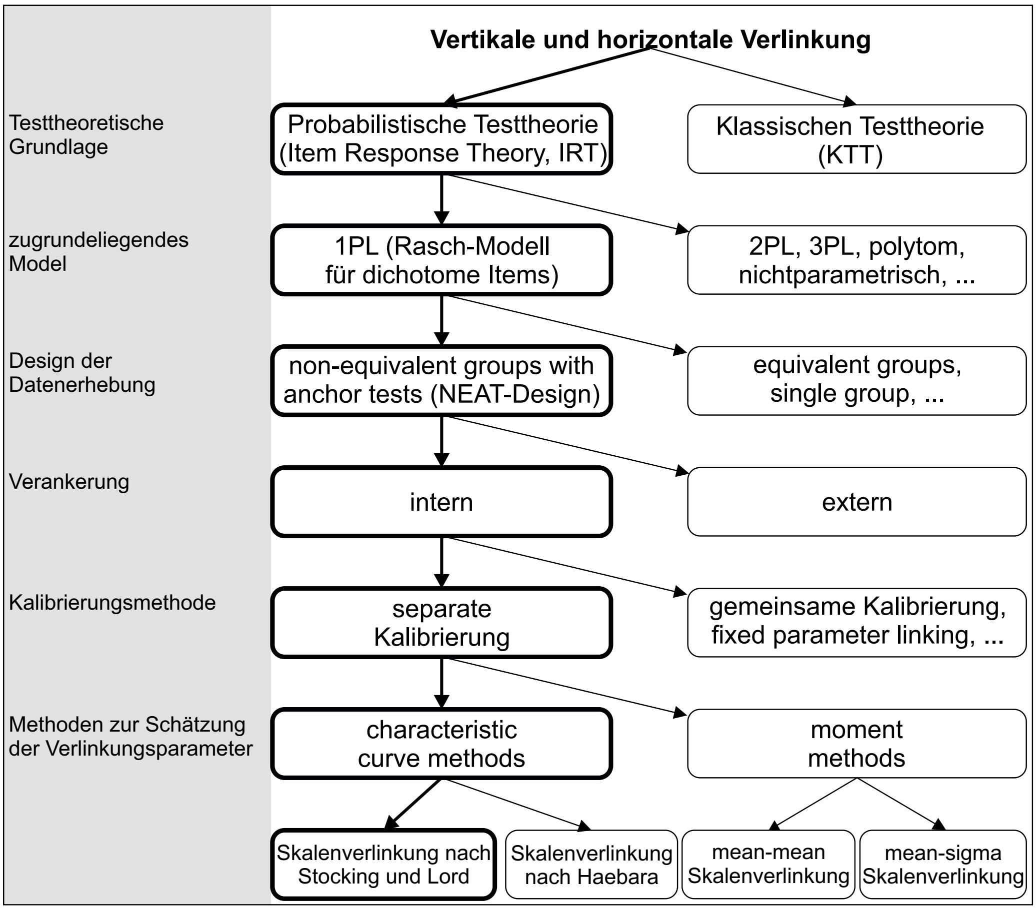


Abbildung 4.7 Entscheidungsweg zur Methodenauswahl bei der Verlinkung von Testvarianten. Quelle: Eigene Darstellung bzw. Zusammenstellung auf Grundlage der nachfolgend genannten Literatur

Da alle Varianten des Speed- und Leistungstests prozedurales Wissen Rasch-homogen sind, wird das Rasch-Modell für dichotome Items (1PL-Modell) zugrunde gelegt (Rasch, 1960). Testtheoretische Grundlage ist damit die probabilistische Testtheorie.

Aus organisatorischen Gründen und aufgrund der längsschnittlichen Anlage entfallen sowohl das *single group design* als auch das *equivalent group design* (von Davier, Carstensen & von Davier, 2008, S. 125). Daher wird in der Arbeit sowohl für die horizontale als auch für die vertikale Verlinkung das NEAT-Design (von Davier, Holland & Thayer, 2004, S. 13) verwendet. Hierbei werden zwei verschiedene Varianten eines Tests über einen Ankertest verlinkt, der im vorliegenden Fall aus einer Menge von Items desselben Tests besteht (interner Ankertest).

Es ist davon auszugehen, dass sich das prozedurale Wissen von Studierenden innerhalb der ersten drei Studiensemester verändert. Folglich wurde der Test zu den drei Messzeitpunkten von Personen mit einer unterschiedlichen Fähigkeit zur Bewältigung der Testitems bearbeitet. In diesem Fall gilt eine separate Kalibrierung den anderen Kalibrierungsmethoden als überlegen (von Davier et al., 2008, S. 132). Daher wird auch im vorliegenden Fall zur Verlinkung der Testvarianten eine separate Kalibrierung vorgenommen. Hierbei werden Schwierigkeits- und Personenparameter in den zu verlinkenden Testvarianten inklusive Ankeritems separat geschätzt.

Einen zentralen Aspekt bei der Verlinkung auf Grundlage des Rasch-Modells stellt die Invarianz der Likelihood (siehe S. 78) des Schätzmodells gegenüber einer Verschiebung der Schwierigkeits- und Personenparameter dar (Kolen & Brennan, 2004, S. 161). Dies bedeutet, dass sich an der Optimalität der geschätzen Schwierigkeits- und Personenparameter nichts ändert, wenn zu diesen eine reelle Konstante addiert wird. Formal lässt sich dieser Sachverhalt wie folgt beschreiben: Mit V_1 sei die erste Testvariante bezeichnet und mit V_2 die zweite. V_1 und V_2 sollen mithilfe eines internen Ankertests $A_{V_2 \to V_1}$ miteinander verlinkt werden. Der Index deutet an, dass die Skala von V_2 so skaliert wird, dass sie äquivalent zur Skala von V_1 ist und folglich die Items von V_2 auf der Skala von V_1 verortet werden können. Die Skala von V_1 wird daher auch als Basisskala bezeichnet (Kang & Petersen, 2012). Weiter sei β_{j,V_2} der Schwierigkeitsparameter eines Items j in V_2 und θ_{i,V_2} der Personenparameter einer Person i, die an V_2 teilgenommen hat. Ferner sei $B \in \mathbb{R}$. Dann ergibt sich nach Ersetzen der Parameter β_{j,V_2} und θ_{i,V_2} durch $\beta_j^* := \beta_{j,V_2} + B$ bzw. $\theta_i^* := \theta_{i,V_2} + B$ im Schätzmodell dieselbe Likelihood.

Ziel ist es nun, den Verlinkungsparameter B so zu bestimmen, dass die Items des Ankertests $A_{V_2 \to V_1}$ in V_1 und V_2 möglichst gut miteinander identifiziert werden. Zur Bestimmung des Verlinkungsparameters B gelten *characteristic curve methods* den *moment methods* als überlegen, da erstere den Parameter B im Rahmen eines Optimierungsproblems bestimmen, das die Verteilung der Personenparameter berücksichtigt (Carlson, 2011, S. 66). Unter den *characteristic curve methods* wird laut Wang (2013, S. 8) die Methode von Stocking und Lord (Stocking & Lord, 1983) am häufigsten verwendet, weshalb sie auch in der vorliegenden Arbeit eingesetzt wird.

Eine besondere Bedeutung kommt der Konzeption von Ankertests zu, an die mindestens die folgenden drei Forderungen gestellt werden (Tian, 2011, S. 33): Ein Ankertest sollte (1) aus einer ausreichenden Anzahl von Items bestehen. Seine Items sollten (2) gute Repräsentanten der zu verlinkenden Testvarianten sein. Ferner sollten sie (3) keinen *item parameter drift* (IPD) aufweisen. IPD stellt die Analogie zum *differential item functioning* (DIF) dar. Während DIF für ein unterschiedliches Funktionieren von Items bei verschiedenen Gruppen einer Population steht, bedeutet IPD ein unterschiedliches Funktionieren in unterschiedlichen Populationen, wie sie typischerweise in Längsschnittanalysen zu verschiedenen Testzeitpunkten vorliegen (Sukin, 2010, S. 3). Dabei gilt ein IPD der Ankeritems von 0.3 Logits als tolerierbar (Wright & Stone, 1979, S. 98).

Abbildung 4.8 veranschaulicht den IPD der Ankeritems nach der Verlinkung der angegebenen Testvarianten. Die Logits auf der horizontalen Achse beziehen sich auf die Basisskala. Auf der vertikalen Achse sind die Logits der skalierten Skala abgetragen. Falls sich die Ankeritems in den verlinkten Testvarianten identisch verhalten, liegen sie auf der skizzierten Winkelhalbierenden. Je weiter sie sich von dieser Geraden entfernen, desto höher ist ihr IPD. Die horizontalen Balken bei jedem Item stellen die tolerierbare Distanz von 0.3 Logits dar. Bei keinem Ankeritem liegt die Winkelhalbierende außerhalb dieser Entfernung. Folglich weist keines der Items einen problematischen IPD auf. Bedingung (3) ist damit erfüllt.

Die Repräsentativität eines Items für einen Test drückt sich in der Korrelation des Items zu den übrigen Testitems aus. In der vorliegenden Arbeit gilt ein Item als repräsentativ, wenn diese Korrelation der mittleren paarweisen Korrelation aller Testitems (MIC) ent-

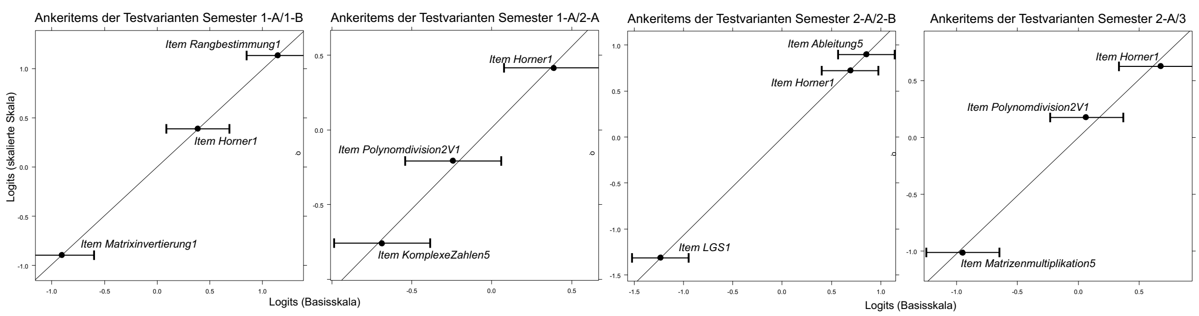


Abbildung 4.8 Parameter Item Drift (IPD) der Ankeritems bei den einzelnen Verlinkungen

Tabelle 4.9

Repräsentativität der Ankeritems, operationalisiert durch die durchschnittliche tetrachorische Korrelation der Ankeritems mit den restlichen Testitems

	Testvariante Semester 1-A	Testvariante Semester 1-B	Testvariante Semester 2-A	Testvariante Semester 2-B	Testvariante Semester 3
MIC	.37	.36	.34	.34	.29
Ankeritems					
Ableitungen5			.37 ↑	.32 ↓	
Horner1	.40 ↑	.38 ↑	.34 ↔	.33 ↓	.31 ↑
Komplexe Zahlen5	.40 ↑		.38 ↑		
LGS1			.31 ↓	.34 ↔	
Matrizenmultiplikation5			.30 ↓		.29 ↔
Matrixinvertierung1	.37 ↔	.31 ↓			
Polynomdivision2V1	.42 ↑		.37 ↑		.30 ↑
Rangbestimmung1	.35 ↓	.31 ↓			

*A*nmerkung. MIC=mean interitem correlation; ↑: Korrelation höher als MIC (hoch repräsentatives Item); ↓: Korrelation geringer als MIC (gering repräsenatives Item); ↔: Korrelation gleich MIC (repräsentatives Item)

spricht. Liegt die Korrelation oberhalb der MIC, wird das Item als hoch repräsentativ bezeichnet. Liegt sie unterhalb der MIC, wird es entsprechend als gering repräsentativ bezeichnet. In Tabelle 4.9 ist die Repräsentativität der Ankeritems dargestellt. Im Durchschnitt sind diese repräsentativ für die Testvarianten. Bei Items mit einer Korrelation unterhalb der MIC ist der Unterschied in allen Fällen gering. Insgesamt können daher alle Ankeritems als nahezu repräsentativ angesehen werden. Die an die Ankertests gestellte zweite Bedingung ist damit ebenfalls erfüllt.

Als eine angemessene Länge für einen Ankertest gibt Angoff (1984, S. 107) 20 % des Gesamttests an. Pibal und Cesnik (2011) beschreiben, wie sich bei einer Reduzierung der Ankeritemzahl von fünf auf drei unter Beibehaltung der Items mit dem geringsten IPD die Qualität der Verlinkung sogar verbessert. Die psychometrische Qualität der Ankeritems spielt somit eine wesentliche Rolle. In der vorliegenden Arbeit enthält jeder Ankertests daher drei besonders stabile Items (vgl. Abbildung 4.8). Der Anteil der Ankeritems am Gesamttest liegt damit in allen Fällen über 20 %. Somit ist auch die erste der eingangs an Ankertests gestellten Bedingungen erfüllt.

Als ein Resultat der Skalenverlinkung ist an dieser Stelle zu nennen, dass für den Verlinkungsparameter B zwischen den Testvarianten Semester 1-A und 1-B gilt $|B| \approx 0$, sodass die beiden Testvarianten direkt miteinander identifiziert werden können.

Transformation der Logit-Werte auf die PISA-Skala

Zur besseren Lesbarkeit werden die Logit-Werte durch die Transformation $\tau : z_{\text{Logit}} \mapsto z_{\text{PISA}}$ mit

$$\tau(z) = 500 + 100 \cdot \frac{z - \bar{z}}{\sigma} \tag{4.4}$$

in entsprechende Werte der PISA-Skala umgerechnet (Knoche et al., 2002, S. 170). Die PISA-Skala ist so normiert, dass der Mittelwert der Parameter gleich 500 ist und die Standardabweichung 100 beträgt. Dabei bezeichnet z einen geschätzten Parameter, $\bar{z}$ den Mittelwert und σ die Standardabweichung der geschätzten Parameter. Diese Transformation wird im Rahmen der vertikalen Verlinkung für die Personenparameter der einzelnen Testvarianten wie folgt durchgeführt: Zunächst werden $\bar{z}$ und σ für die Basisskala bestimmt. Anschließend werden die Logit-Werte der anderen Testvarianten in entsprechende Logit-Werte der Basisskala und anschließend mithilfe von Formel (4.4) in die entsprechenden Werte der PISA-Skala transformiert.

Validität

Als erstes Validitätskriterium kann die über zehnjährige Erfahrung des Autors in der Itementwicklung im Rahmen des Veranstaltungsbetriebs angeführt werden. Dadurch ist insbesondere sichergestellt, dass alle Testitems Basisprozeduren in dem in Abschnitt 1.1.1 definierten Sinne abprüfen, die in dieser Form in der Vorlesung und in den Übungen behandelt wurden.

Gemäß Nevo (1985, S. 292) sollte auch die Augenscheinvalidität routinemäßig in Validitätsüberprüfungen eines Tests einbezogen werden. Nevo (1985, S. 288) schlägt vor, Fachleute und Laien nach ihrer Einschätzung zu befragen, ob ein Test das misst, was er messen soll. Dementsprechend wurde im vorliegenden Fall eine Befragung einer Gruppe von Fachleuten und Laien durchgeführt. Die Gruppe der Fachleute bestand aus zwei Doktoranden und einer Doktorandin, die mindestens zwei Jahre in den Veranstaltungsbetrieb, unter anderem als Globalübungsleiterin bzw. Globalübungsleiter, involviert waren. Die Gruppe der Laien bestand aus drei neu eingestellten Tutorinnen und Tutoren des Veranstaltungsbetriebs zur Höheren Mathematik I im Wintersemester 2015/16. Fachleute und Laien kannten den Test nicht. Die Laien verfügten über eine ausreichende mathematische Vorbildung zur Beurteilung der Testinhalte. Nach der Präsentation des Tests sollte folgende Aussage reflektiert werden: „Der Test misst die Fähigkeit von Studierenden, Algorithmen/Kalküle, die in der Höheren Mathematik für Ingenieure unterrichtet werden, in angemessener Zeit fehlerfrei durchführen zu können.“ Zur Beurteilung wurde allen Ratern die von Nevo (1985, S. 292) vorgeschlagene Skala vorgelegt, die die Kategorien „hervorragend geeignet“, „sehr geeignet“, „geeignet“, „ungeeignet“ sowie „irrelevant“, umfasst. Zwei der Fachleute und einer der Laien stuften den Test als „hervorragend geeignet“ ein, während eine Fachfrau bzw. ein Fachmann und zwei Laien den Test als „sehr geeignet“ einschätzten. Demnach weist der Speed- und Leistungstest prozedurales Wissen sowohl für Fachleute als auch für Laien mit mathematischer Vorbildung eine hohe Augenscheinvalidität auf.

4.2.2 Leistungstest konzeptuelles Wissen

Konzeptuelles Wissen wurde in Kapitel 1 neben prozeduralem Wissen als Teil des Anforderungsprofils von Klausuren der Ingenieurmathematik identifiziert. In der vorliegenden Arbeit wird konzeptuelles Wissen daher gemeinsam mit prozeduralem Wissen als fachspezifischer Prädiktor für Klausurerfolg und -leistung untersucht. Im Folgenden wird die Operationalisierung konzeptuellen Wissens detailliert beschrieben.

Testbeschreibung

Das Ziel bei der Entwicklung des Leistungstests konzeptuelles Wissen war es, einen zum Speed- und Leistungstest prozedurales Wissen komplementären Test zu konstruieren, der konzeptuelles Wissen im Sinne der Definition aus Abschnitt 1.1.2 operationalisiert. Darunter ist Folgendes zu verstehen: Bei letztgenanntem Test sind Basisprozeduren der Ingenieurmathematik unter Zeitbeschränkung fehlerfrei anzuwenden (siehe Abschnitt 4.2.1). Das mathematische Modell ist vorgegeben und es erfolgt lediglich eine Verarbeitung durch Ausführen einer Basisprozedur. Der kognitive Anspruch der Items ist also gering, da die Items durch Schema-Wissen zu lösen sind. Der Leistungstest konzeptuelles Wissen dagegen besteht aus Items, die Mathematisieren erfordern. Hierbei sind verschiedene Basiskonzepte und zum Teil auch größere Einheiten konzeptuellen Wissens zu verknüpfen (vgl. Abschnitt 1.1.2). Der kognitive Anspruch an die Testteilnehmenden ist daher hoch, während die Komplexität der durchzuführenden Rechnung niedrig gehalten wird. Die Teilnehmenden dürfen die Items ohne Zeitbeschränkung und mit beliebigen (nicht elektronischen) Hilfsmitteln bearbeiten. Folglich lassen sich beide Tests – wie in Abbildung 4.9 zu sehen – in verschiedenen Bereichen des mathematischen Modellierungskreislaufs verorten.

Der Test wurde zu Beginn des Wintersemesters 2013/14 durchgeführt (siehe Abschnitt 4.1.1). Studienanfängerinnen und -anfänger schrieben den Test somit zu Beginn ihres Studiums. Daher sind alle Items so angelegt, dass sie mit Schulwissen lösbar sind. Die Items ähneln den Items zur Messung konzeptuellen Wissens von Engelbrecht et al. (2012, S. 146), Engelbrecht et al. (2005, S. 705) und insbesondere Mahir (2009, S. 203). Die psychometrische Qualität stand bei der Entwicklung nicht im Vordergrund, weshalb im

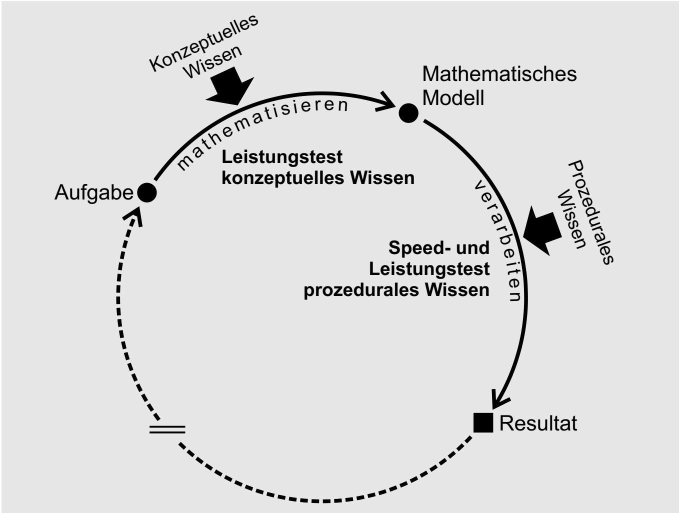


Abbildung 4.9 Verortung des Leistungstests im mathematischen Modellierungskreislauf (vereinfacht in Anlehnung an Blum & Leiß, 2005, S. 19)

weiteren Verlauf – wie bei den vorgenannten Arbeiten auch – keine Kennzahlen für Reliabilität und Validität angegeben werden. Stattdessen werden Test und Items primär aus qualitativer Sicht diskutiert. Über Eigenschaften der konstruierten Testhefte wird hier nur deskriptiv berichtet.

Aufgrund der hohen Teilnehmerzahl von 1616 Studierenden und einer begrenzten Anzahl an Aufsichtspersonen mussten besondere Vorkehrungen getroffen werden, um ein Abschreiben entsprechend nahe beieinandersitzender Studierender zu verhindern. Aus diesem Grund wurden sechs Testheftvarianten erstellt. Jedes Testheft enthält fünf Items, die sich inhaltlich oder in ihrer Anordnung innerhalb der Testhefte unterscheiden. In Tabelle 4.11 sind neben der Verteilung der Items auf die Testhefte auch die je Item zu leistenden Verknüpfungen von Basiskonzepten bzw. Einheiten konzeptuellen Wissens angegeben. Diese werden im nächsten Abschnitt exemplarisch an zwei Beispielen erläutert.

Ein weiteres Ziel der Testentwicklung war es, gleich schwierige Testhefte mit einer vergleichbaren Streuung der Ergebnisse zu konstruieren. Tabelle 4.10 zeigt, dass dies weitestgehend gelungen ist. Lediglich Testheft 4 verfügt über eine erhöhte durchschnittliche Lösungsquote. In der Nachanalyse stellte sich in der Tat heraus, dass Testheft 4 insgesamt über einfachere Items als die übrigen Testhefte verfügt, was beim Equating, d. h. der horizontalen Verlinkung der Testhefte, berücksichtigt wird (S. 91).

Tabelle 4.10

Teilnehmendenzahl, durchschnittliche Zahl korrekt gelöster Items und Standardabweichung

	Testheft 1	Testheft 2	Testheft 3	Testheft 4	Testheft 5	Testheft 6
Teilnehmerzahl	286	280	262	272	257	259
durchschnittliche Zahl korrekter Items	1.55	1.59	1.73	2.06	1.59	1.46
Standardabweichung	1.32	1.38	1.29	1.41	1.5	1.32

Tabelle 4.11

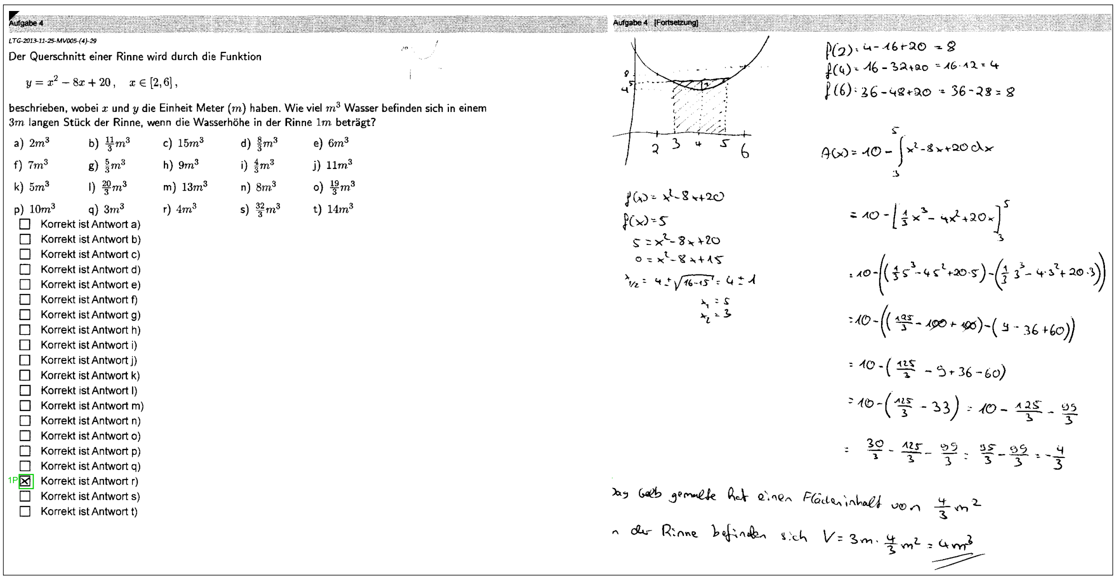
Verteilung der Items auf die sechs Testhefte (TH1 bis TH6) des Leistungstests konzeptuelles Wissen und zu verknüpfende Basiskonzepte/Einheiten konzeptuellen Wissens

Item	zu verknüpfende Basiskonzepte bzw. Einheiten konzeptuellen Wissens	TH1	TH2	TH3	TH4	TH5	TH6
Bohrinsel	geometrische Figuren (Kreis), nichtlineares Wachstum			X			
Böschung	geometrische Figuren (Kegel), Satzgruppe des Pythagoras		X				
Flächenvergleich	Parabeln, Flächenberechnung durch Integration, Proportionen und Ungleichungen	X					
Flächenvergleich V1	Parabeln, Flächenberechnung durch Integration, Proportionen und Ungleichungen			X			
Flächenvergleich V2	geometrische Figuren (Rechteck), Extremwertprobleme					X	
größtes Rechteck	geometrische Figuren (Rechteck), Extremwertprobleme		X				
größtes Rechteck V1	geometrische Figuren (Rechteck), Extremwertprobleme				X	X	
Konzertbühne	geometrische Figuren (Rechteck, Trapez), Satzgruppe des Pythagoras				X		
Konzertbühne V1	geometrische Figuren (Rechteck, Trapez), Satzgruppe des Pythagoras						X
Kugel	geometrische Figuren (Kugel, Zylinder), antiproportionale Zurodnung, Extremwertprobleme			X			
Lichtstrahler	geometrische Figuren (Dreieck, Kreis), Polynome, Extremwertberechnung, Prozentrechnung, Tangens		X				
Lichtstrahler V1	geometrische Figuren (Dreieck, Kreis), Polynome, Extremwertberechnung, Prozentrechnung, Tangens						X
Luftballon	geometrische Figuren (Kugel), nichtlineares Wachstum	X					
Luftballon V1	geometrische Figuren (Kugel), nichtlineares Wachstum				X		
Musikkonzert	Parameterdarstellung von Geraden aus Dreipunkteform, Schnittpunkte	X					
Polynom	Parameterdarstellung von Geraden, Schnittpunkte, Extremwertprobleme					X	
Rinne	Parabeln, Schnittpunkte, Flächenberechnung durch Integration, Volumenberechnung			X			
Rinne V1	Parabeln, Schnittpunkte, Flächenberechnung durch Integration, Volumenberechnung				X		
Stahlträger	Ableitung/Tangente, Extremwertprobleme						X
Tangenten	Parameterdarstellung von Geraden, Tangentengleichung, Skalarprodukt/Orthogonalität		X				
Trapez	geometrische Figuren (Trapez), Geradengleichung, Flächenberechnung durch Integration		X				
Trapez V1	geometrische Figuren (Trapez), Geradengleichung, Flächenberechnung durch Integration					X	
Umfang	geometrische Figuren (Dreieck), Satzgruppe des Pythagoras	X					
Umfang V1	geometrische Figuren (Dreieck), Satzgruppe des Pythagoras						X
Vergleiche	Ungleichungen, Umgang mit Nebenbedingungen	X					
Vergleiche V1	Ungleichungen, Umgang mit Nebenbedingungen			X			
Vergleiche V2	Ungleichungen, Umgang mit Nebenbedingungen						X
Wasserpfütze	geometrische Figuren (Kreis), nichtlineares Wachstum					X	
Zwei Pfützen	geometrische Figuren (Kreis), nichtlineares Wachstum				X		

Beispielitems

Bei der Konstruktion der Items wurde darauf geachtet, dass in der Regel mindestens zwei Basiskonzepte oder Einheiten konzeptuellen Wissens aktiviert und miteinander verknüpft werden müssen (vgl. Tabelle 4.11). Bei den meisten Items handelt es sich um eingekleidete Aufgaben, sodass mathematisches Modellieren erforderlich ist. Dies wird im Folgenden exemplarisch an zwei Studierendenlösungen demonstriert.

Abbildung 4.10 zeigt eine Bearbeitung des Items „Rinne“. Während der Anspruch an konzeptuelles Wissen hier hoch ist, stellt das Item nur geringe Anforderung an prozedurales Wissen. Zunächst erfolgt ein Zugriff auf das konzeptuelle Netzwerk (siehe Abschnitt 1.1.2). Hier wird das Basiskonzept „Parabel“ aktiviert. Um die Integrationsgrenzen zu bestimmen, ist ein Wechsel in das prozedurale Netzwerk (siehe Abschnitt 1.1.1) erforderlich und eine quadratische Gleichung zu lösen. Anschließend erfolgt die Rückkehr in das konzeptuelle Netzwerk, um auf das Basiskonzept der Flächenberechnung durch Integration zurückzugreifen. Dabei ist das bestimmte Integral wiederum mit Prozeduren aus dem prozeduralen Netzwerk zu berechnen. Abschließend wird noch auf das Basiskonzept der Volumenberechnung zurückgegriffen.



Aufgabe 4

LTG-2013-11-25-MV005-(4)-29

Der Querschnitt einer Rinne wird durch die Funktion

$$y = x^2 - 8x + 20\,, \quad x \in [2,6]\,,$$

beschrieben, wobei x und y die Einheit Meter (m) haben. Wie viel m^3 Wasser befinden sich in einem $3m$ langen Stück der Rinne, wenn die Wasserhöhe in der Rinne $1m$ beträgt?

a) $2m^3$	b) $\frac{11}{3}m^3$	c) $15m^3$	d) $\frac{8}{3}m^3$	e) $6m^3$
f) $7m^3$	g) $\frac{5}{3}m^3$	h) $9m^3$	i) $\frac{4}{3}m^3$	j) $11m^3$
k) $5m^3$	l) $\frac{20}{3}m^3$	m) $13m^3$	n) $8m^3$	o) $\frac{19}{3}m^3$
p) $10m^3$	q) $3m^3$	r) $4m^3$	s) $\frac{32}{3}m^3$	t) $14m^3$

- ☐ Korrekt ist Antwort a)
- ☐ Korrekt ist Antwort b)
- ☐ Korrekt ist Antwort c)
- ☐ Korrekt ist Antwort d)
- ☐ Korrekt ist Antwort e)
- ☐ Korrekt ist Antwort f)
- ☐ Korrekt ist Antwort g)
- ☐ Korrekt ist Antwort h)
- ☐ Korrekt ist Antwort i)
- ☐ Korrekt ist Antwort j)
- ☐ Korrekt ist Antwort k)
- ☐ Korrekt ist Antwort l)
- ☐ Korrekt ist Antwort m)
- ☐ Korrekt ist Antwort n)
- ☐ Korrekt ist Antwort o)
- ☐ Korrekt ist Antwort p)
- ☐ Korrekt ist Antwort q)
- 1P ☒ Korrekt ist Antwort r)
- ☐ Korrekt ist Antwort s)
- ☐ Korrekt ist Antwort t)

Aufgabe 4 [Fortsetzung]

$f(2) = 4 - 16 + 20 = 8$
$f(4) = 16 - 32 + 20 = 16 - 12 = 4$
$f(6) = 36 - 48 + 20 = 36 - 28 = 8$

$f(x) = x^2 - 8x + 20$
$f(x) = 5$
$5 = x^2 - 8x + 20$
$0 = x^2 - 8x + 15$
$x_{1/2} = 4 \pm \sqrt{16-15} = 4 \pm 1$
$x_1 = 5$
$x_2 = 3$

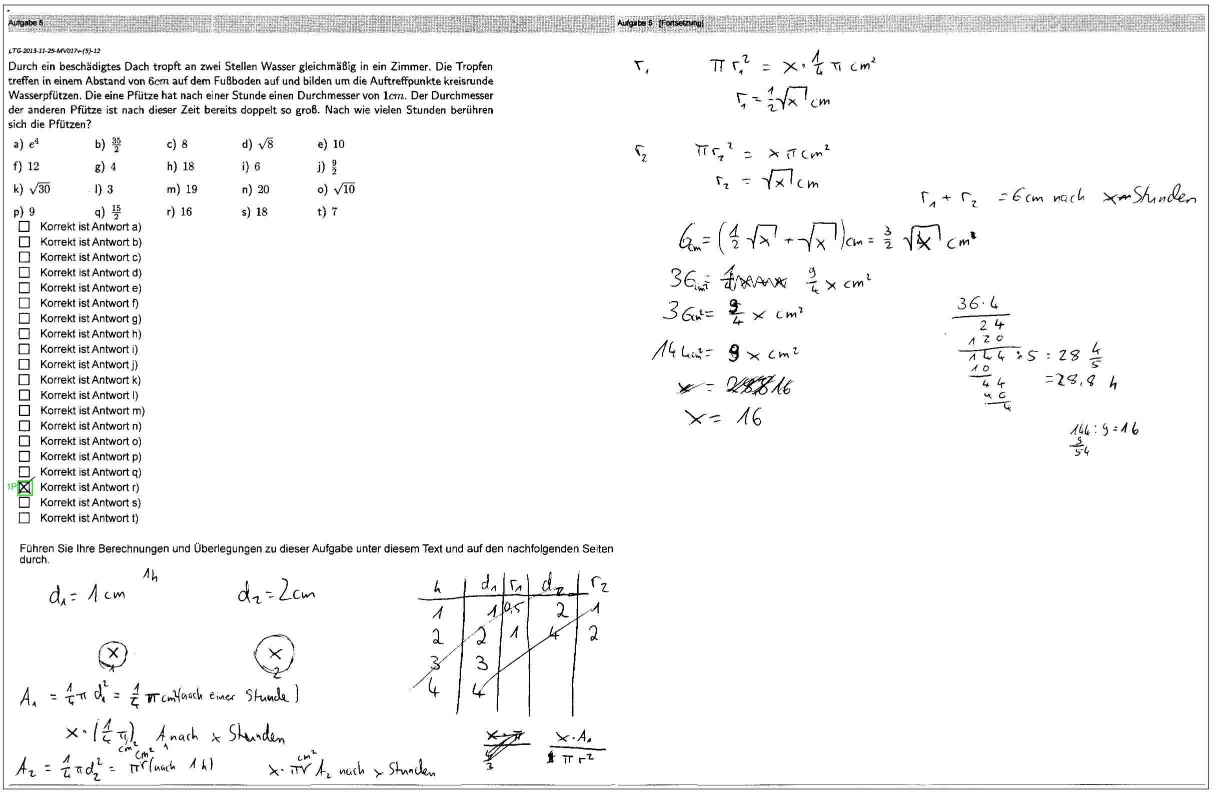
$$A(x) = 10 - \int_3^5 x^2 - 8x + 20\,dx$$
$$= 10 - \left[\tfrac{1}{3}x^3 - 4x^2 + 20x\right]_3^5$$
$$= 10 - \left(\left(\tfrac{1}{3}5^3 - 4\cdot5^2 + 20\cdot5\right) - \left(\tfrac{1}{3}3^3 - 4\cdot3^2 + 20\cdot3\right)\right)$$
$$= 10 - \left(\left(\tfrac{125}{3} - \cancel{100} + \cancel{100}\right) - (9 - 36 + 60)\right)$$
$$= 10 - \left(\tfrac{125}{3} - 9 + 36 - 60\right)$$
$$= 10 - \left(\tfrac{125}{3} - 33\right) = 10 - \tfrac{125}{3} - \tfrac{99}{3}$$
$$= \tfrac{30}{3} - \tfrac{125}{3} - \tfrac{99}{3} = \tfrac{95}{3} - \tfrac{99}{3} = -\tfrac{4}{3}$$

Das Gelb gemalte hat einen Flächeninhalt von $\frac{4}{3}m^2$

In der Rinne befinden sich $V = 3m \cdot \frac{4}{3}m^2 = \underline{\underline{4m^3}}$

Abbildung 4.10 Bearbeitung des Items „Rinne“: Der Sprung in das konzeptuelle Netzwerk gelingt hier ebenso wie das Navigieren und Wechseln zwischen konzeptuellem und prozeduralem Netzwerk.

In Abbildung 4.11 ist eine Bearbeitung des Items „Zwei Pfützen“ zu sehen. Nach einem zunächst fehlerhaften Sprung in das konzeptuelle Netzwerk, bei dem die Person fälschlicherweise ein lineares Wachstum annimmt (siehe angefertigte Tabelle am unteren Rand), gelingt die Aktivierung des Basiskonzepts „nichtlineares Wachstum“ und dessen Verknüpfung mit dem Basiskonzept „geometrische Figuren (Kreis)“. Das anschließende Navigieren durch das prozedurale Netzwerk ist erfolgreich und führt zur korrekten Lösung.



Aufgabe 5

LTG-2013-11-25-MV017v-(5)-12

Durch ein beschädigtes Dach tropft an zwei Stellen Wasser gleichmäßig in ein Zimmer. Die Tropfen treffen in einem Abstand von $6cm$ auf dem Fußboden auf und bilden um die Auftreffpunkte kreisrunde Wasserpfützen. Die eine Pfütze hat nach einer Stunde einen Durchmesser von $1cm$. Der Durchmesser der anderen Pfütze ist nach dieser Zeit bereits doppelt so groß. Nach wie vielen Stunden berühren sich die Pfützen?

a) e^4 b) $\frac{35}{2}$ c) 8 d) $\sqrt{8}$ e) 10
f) 12 g) 4 h) 18 i) 6 j) $\frac{9}{2}$
k) $\sqrt{30}$ l) 3 m) 19 n) 20 o) $\sqrt{10}$
p) 9 q) $\frac{15}{2}$ r) 16 s) 18 t) 7

- [ ] Korrekt ist Antwort a)
- [ ] Korrekt ist Antwort b)
- [ ] Korrekt ist Antwort c)
- [ ] Korrekt ist Antwort d)
- [ ] Korrekt ist Antwort e)
- [ ] Korrekt ist Antwort f)
- [ ] Korrekt ist Antwort g)
- [ ] Korrekt ist Antwort h)
- [ ] Korrekt ist Antwort i)
- [ ] Korrekt ist Antwort j)
- [ ] Korrekt ist Antwort k)
- [ ] Korrekt ist Antwort l)
- [ ] Korrekt ist Antwort m)
- [ ] Korrekt ist Antwort n)
- [ ] Korrekt ist Antwort o)
- [ ] Korrekt ist Antwort p)
- [ ] Korrekt ist Antwort q)
- [x] Korrekt ist Antwort r) 1P
- [ ] Korrekt ist Antwort s)
- [ ] Korrekt ist Antwort t)

Führen Sie Ihre Berechnungen und Überlegungen zu dieser Aufgabe unter diesem Text und auf den nachfolgenden Seiten durch.

Aufgabe 5 [Fortsetzung]

Abbildung 4.11 Bearbeitungsbeispiel des Items „Zwei Pfützen"

Testdurchführung und -auswertung

Der Test wurde als Paper & Pencil-Test dargeboten. Um die Ratewahrscheinlichkeit zu reduzieren, standen bei jedem Item 20 Antwortoptionen zur Verfügung, von denen die korrekte markiert werden musste (Single-Choice-Format). Ferner wurden für falsch markierte Optionen negative Punkte vergeben.
Um die Motivation zu erhöhen, den Test gewissenhaft zu bearbeiten, konnten Bonuspunkte für die nachfolgende Klausur gesammelt werden. Alle Regelungen wurden zuvor per Aushang und vor Testbeginn durch die Testleitung angekündigt.
Die Testauswertung erfolgte nach dem in Abschnitt 4.1.3 geschilderten standardisierten Verfahren.

Horizontale Verlinkung der Testhefte

Um die Ergebnisse der Testhefte vergleichen zu können, ist eine geeignete Verlinkung der Testhefte erforderlich. Da die Tests nicht auf Grundlage der probabilistischen Testtheorie entwickelt wurden, ist bei der Suche einer geeigneten Verlinkungsstrategie im ersten Knoten des Entscheidungsbaums in Abbildung 4.7 die rechte Abzweigung zu wählen. Hier existieren zwei klassische Methoden: das lineare Equating und das equiperzentile Equating (Livingston, 2004, S. 14).

Da der hier untersuchte Leistungstest konzeptuelles Wissen lediglich fünf Items umfasst, ist eine Differenzierung der erzielten Scores nach Prozenträngen nur sehr eingeschränkt möglich. Aus diesem Grund wird in der vorliegenden Arbeit das lineare Equating dem equiperzentilen Equating vorgezogen.
Nach dem linearen Equating gelten zwei Summenscores verschiedener Testhefte T_1 und T_2 als äquivalent, falls ihr Abstand zum Mittelwert, gemessen in Standardabweichungen, identisch ist. Formal besteht dann zwischen einem Summenscore z_{T_1} in T_1 und einem äquivalenten Summenscore z_{T_2} in T_2 die Beziehung

$$z_{T_1} = \bar{z}_{T_1} + \frac{\sigma_{T_1}}{\sigma_{T_2}} \cdot \left(z_{T_2} - \bar{z}_{T_2}\right) \tag{4.5}$$

Dabei bezeichnen $\bar{z}_{T_1}$ und $\bar{z}_{T_2}$ die Mittelwerte der Summenscores in T_1 bzw. T_2 und σ_{T_1} bzw. σ_{T_2} die entsprechenden Standardabweichungen. Mithilfe von Formel (4.5) werden die erzielten Summenscores aller Testhefte auf die Skala des ersten Testheftes transformiert.

4.3 Instrumente der Datenerhebung für allgemein kognitive Faktoren

4.3.1 Psychomeda-Konzentrationstest (KONT-P)

Konzentration ist der Aspekt des Arbeitens, der die Selektion und Koordination von Informationen bzw. Reizen betrifft. Sie wird über das Tempo und die Fehlerneigung beim konzentrierten Arbeiten (siehe Abschnitt 2.2) operationalisiert. Tempo und Fehlerneigung werden auch beim Psychomeda-Konzentrationstest (KONT-P) erfasst, der im Rahmen der vorliegenden Arbeit zur Messung von Konzentation verwendet wurde. Im Folgenden werden die Details zur Operationalisierung beschrieben.

Testbeschreibung

Beim KONT-P handelt es sich um ein von Satow (2011) entwickeltes Testinstrument zur Erfassung der Konzentrationsfähigkeit. Der Test steht als Online-Test zur Verfügung und unterliegt der „Creative Commons Namensnennung-Nicht-kommerziell-Weitergabe unter gleichen Bedingungen 3.0 Unported“-Lizenz, d. h. der Test darf kostenlos verwendet und zu nicht kommerziellen Zwecken adaptiert werden.

Der Konzentrationstest besteht aus insgesamt 28 Items. Jedes Item umfasst 7 Zeilen. In den Items $1-7$ (Aufgabenblock 1) und $15-21$ (Aufgabenblock 3) sind in jeder der 7 Zeilen einstellige Zahlen zu addieren, in den Items $8-14$ (Aufgabenblock 2) und $22-28$ (Aufgabenblock 4) sind in jeder der 7 Zeilen Einsen zu zählen. Diese stehen in den Zeilen neben ähnlichen Zeichen, wie I und i (vgl. Abbildung 4.12). Die Aufgabenblöcke 1 und

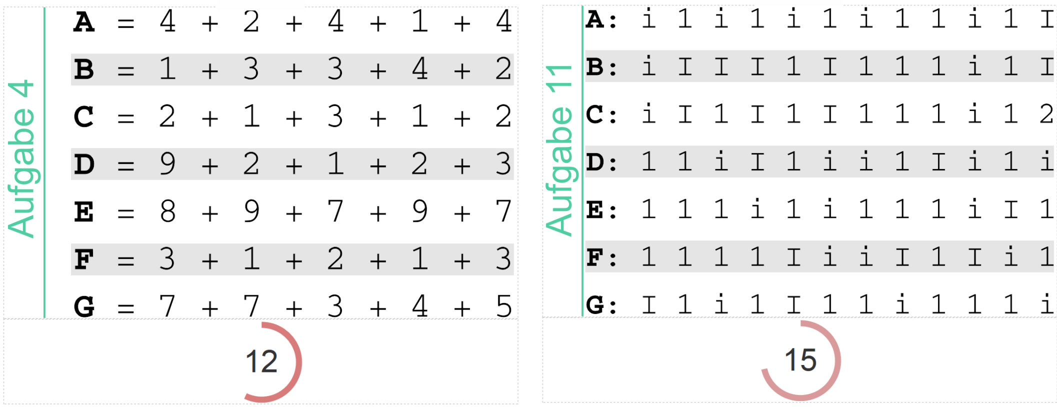


Abbildung 4.12 Screenshots von Beispielitems aus der Beamer-Version des Tests für Hörsaaltestungen zu den Aufgabenblöcken 1 und 3 (elementares Addieren) bzw. 2 und 4 (elementares Abzählen)

3 werden zur Subskala *Rechnen*, die Blöcke 2 und 4 zur Subskala *Zählen* und alle Aufgabenblöcke zusammen zur Gesamtskala *Gesamt* zusammengefasst. Für die Bearbeitung eines Items stehen 20 Sekunden zur Verfügung. In der Regel ist es aus Zeitgründen nicht möglich, alle Zeilen eines Items zu bearbeiten. Die Ergebnisse sind auf einem Bearbeitungsbogen zu notieren.
Um den Test für Hörsaaltestungen nutzbar zu machen, wurde die Online-Version in eine Java-Anwendung portiert und beamerbasiert dargeboten.

Durchführung und Auswertung

Die Testdurchführung inklusive Instruktion war vollständig automatisiert. Die Instruktionen wurden vor jedem Aufgabenblock für 25 Sekunden per Beamer eingeblendet. Es wurde jeweils darauf hingewiesen, die Aufgaben von oben nach unten und so schnell und sorgfältig wie möglich zu bearbeiten. Um die Motivation für eine gewissenhafte Bearbeitung zu erhöhen wurde darauf hingewiesen, dass alle Teilnehmenden eine individuelle Rückmeldung zu ihren Ergebnissen erhalten. Damit wurde den typischen Schwierigkeiten bei der Durchführung von Konzentrationstests begegnet (vgl. Abschnitt 2.2.2).

Nach Ablauf der Bearbeitungszeit eines Items ertönte ein akustisches Signal und es erschien für fünf Sekunden der Hinweis „nächste Aufgabe folgt". Sobald ein Umblättern des Bearbeitungsbogens erforderlich war, wurde nach dem letzten Item für sieben Sekunden der Hinweis „bitte umblättern" eingeblendet.

Die Auswertung der Bearbeitungsbögen erfolgte mithilfe einer Schablone. Es wurden auf dem Bearbeitungsbogen in dem dafür vorgesehenen Feld pro Item die Anzahl bearbeiteter sowie die Anzahl korrekt bearbeiteter Zeilen vermerkt. Anschließend wurden die Bearbeitungsbögen eingescannt und die Auswertung vollautomatisch elektronisch erfasst.

Gleichzeitig wurden automatisch für jeden Teilnehmer die Arbeitsleistung, die Arbeitsgeschwindigkeit und die Fehlerneigung, jeweils getrennt für die Items der Blöcke 1 und 3 (Subskala Rechnen), der Blöcke 2 und 4 (Subskala Zählen) sowie aller Items (Gesamtskala), bestimmt. Die Arbeitsleistung wurde durch die Anzahl korrekt gelöster Zeilen und die Arbeitsgeschwindigkeit durch die Anzahl insgesamt bearbeiteter Zeilen operationalisiert. Die Fehlerneigung stellt den prozentualen Anteil falsch bearbeiteter Zeilen an der Anzahl insgesamt bearbeiteter Zeilen dar (vgl. Abschnitt 2.2.2).

Reliabilität

In Tabelle 4.12 sind die Reliabilitäten (Cronbachs Alpha und Split-Half-Reliabilität, vgl. Seite 78) für die einzelnen Skalen aufgeführt. Diese stimmen sehr gut mit den von Satow (2011) genannten Reliabilitäten überein. Für die Gesamtskala Arbeitsleistung berichtet Satow ein Cronbachs Alpha von $r = .94$ und für die Gesamtskala Arbeitsgeschwindigkeit nennt er eine Reliabilität von $r = .96$. Weitere Reliabilitäten gibt Satow nicht an.
Insgesamt konnte somit die von Satow (2011) berichtete hohe Reliabilität des Konzentrationstests bestätigt werden.

Tabelle 4.12
Cronbachs Alpha und Split-Half-Reliabilität der einzelnen Skalen

	Arbeitsleistung Gesamtskala	Arbeitsleistung Subskala Rechnen	Arbeitsleistung Subskala Zählen	Arbeitsgeschwindigkeit Gesamtskala	Arbeitsgeschwindigkeit Subskala Rechnen	Arbeitsgeschwindigkeit Subskala Zählen	Fehlerneigung Gesamtskala	Fehlerneigung Subskala Rechnen	Fehlerneigung Subskala Zählen
Cronbachs Alpha	.93	.91	.90	.96	.94	.94	.74	.56	.73
Split-Half-Reliabilität	.93	.91	.90	.96	.94	.94	.74	.56	.73

Validität

Satow (2011) berichtet über eine bei Konzentrationstests zu erwartende Diskriminationsfähigkeit des vorliegenden Tests bezüglich Alter, Schulbildung und Einkommensgruppe. Ferner variieren die Ergebnisse in Abhängigkeit vom Bildungsstand (vgl. hierzu auch Westhoff, 1995, S. 394–395). Satow führt als weiteres Validitätskriterium an, dass erwartungsgemäß Personen mit einer ADHS-Diagnose oder Personen, die bei sich ADHS vermuten, im Test unterdurchschnittliche Ergebnisse erzielen.
Bei dem eingesetzten Konzentrationstest handelt es sich um eine nicht begutachtete Online-Publikation. Die Angaben zur Validität können dennoch als vertrauenswürdig angesehen werden, da sie auf einer Stichprobe von 2183 Teilnehmern beruhen (Satow, 2011, S. 7).
Die Validität wird zusätzlich dadurch gestärkt, dass im Rahmen der Untersuchungen der vorliegenden Arbeit die Arbeitsgeschwindigkeit und -leistung normalverteilt sind, während

die Fehlerneigung einer Poisson-Verteilung folgt (Abschnitt 2.2.2).

Zuverlässigkeit der Ergebnisse

Die Testergebnisse können nur in weitere Analysen einbezogen werden, wenn der zugrunde liegende Test von den teilnehmenden Studierenden gewissenhaft bearbeitet wurde. Dies kann dann angenommen werden, wenn die vorliegenden Testergebnisse ähnliche Merkmale aufweisen wie die Ergebnisse aus der Vergleichsstichprobe von Satow (2011). Im Folgenden werden einige Merkmale kurz diskutiert und auf Übereinstimmung geprüft.

Die Kennzahlen für Reliabilität sowie die Verteilung der Testergebnisse sind erwartungskonform. Ferner zeigt sich in Übereinstimmung mit den Ergebnissen von Satow (2011) kein signifikanter Unterschied zwischen den Tempoleistungen weiblicher und männlicher Teilnehmenden. Darüber hinaus liegen hohe Übereinstimmung mit den Angaben von Satow (2011) bezüglich der Itemmittelwerte, Standardabweichungen und Trennschärfen sowie der Interskalenkorrelationen vor. Insbesondere bestätigt sich der hohe Zusammenhang zwischen den Gesamtskalen Arbeitsleistung und Arbeitsgeschwindigkeit. Erwartungskonform korrelieren in der vorliegenden Stichprobe Fehlerneigung und Tempoleistungen (Arbeitsleistung und Arbeitsgeschwindigkeit) nicht bzw. gering.

Cluster- und exploratorische Faktorenanalysen bestätigen die zweifaktorielle Struktur des Konzentrationstests (vgl. Abschnitt 4.1.2) auch bei der dieser Arbeit zugrunde liegenden Stichprobe (vgl. Abbildung 4.13). Die Faktoren lassen sich eindeutig den Subskalen Rechnen und Zählen zuordnen, die erklärte Varianz beträgt kumuliert 41 % gegenüber 48 % bei Satow (2011).

Aufgrund der hohen Übereinstimmung in zahlreichen Merkmalen der Ergebnisse des vorliegenden Tests mit den Ergebnissen der Vergleichsstichprobe von Satow (2011) ist davon auszugehen, dass der Test von den teilnehmenden Studierenden gewissenhaft bearbeitet wurde und die Ergebnisse für weitergehende Analysen geeignet sind.

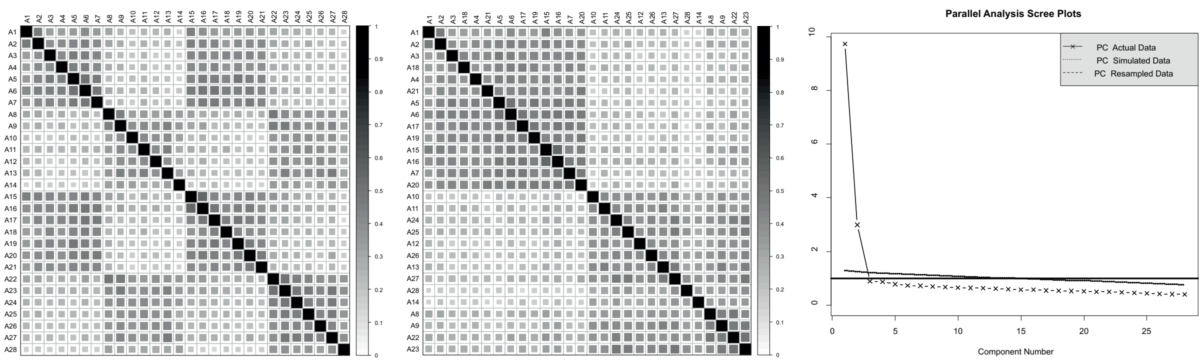


Abbildung 4.13 Links: Levelplot der Korrelationsmatrix (Items in originaler Reihenfolge); Mitte: Levelplot der Korrelationsmatrix nach Clusteranalyse (Complete-Linkage, vgl. Backhaus, Erichson, Plinke & Weiber, 2011, S. 425); Rechts: Screeplot nach Hauptkomponentenanalyse

4.3.2 Wiener Matrizen-Test 2

In der vorliegenden Arbeit wird kognitive Grundfertigkeit durch schlussfolgerndes Denken repräsentiert (siehe Abschnitt 2.1.1). Im Folgenden wird dessen Operationalisierung durch den Wiener Matrizen-Test 2 detailliert beschrieben.

Testbeschreibung

Bei dem Wiener Matrizen-Test 2 (WMT2) handelt es sich um eine gekürzte und neu normierte Fassung des Wiener Matrizen-Tests (Formann & Piswanger, 1979). Er wird von der Testzentrale (Hogrefe Verlag) vertrieben. Der WMT2 ist den figuralen Matrizentests zuzuordnen. Laut Testhandbuch ist er ein „sprachfreies, kurzes und daher sehr ökonomisches Verfahren zur Erfassung des intellektuellen Leistungsniveaus. Entsprechend dem Aufgabentyp erfasst er die Fähigkeit zum schlussfolgernden Denken im Umgang mit abstrakten Symbolen“ (Formann et al., 2011).

Der Test besteht aus insgesamt 18 Items. Jedes Item enthält eine 3×3-Matrix, in der ein Eintrag fehlt. Die vorhandenen acht Matrixeinträge sind nach Regeln aufgebaut, auf deren Grundlage der fehlende Eintrag aus einer Auswahl von jeweils acht Möglichkeiten durch induktives Schließen zu ergänzen ist (vgl. Abbildung 4.14). Die korrekte Lösung ist auf dem Bearbeitungsbogen zu markieren.
Die Bearbeitungsdauer beträgt gemäß Testhandbuch 20 bis maximal 30 Minuten. Der Test wurde aufgrund dieser kurzen Bearbeitungszeit und der guten psychometrischen Qualität für die vorliegende Arbeit ausgewählt.

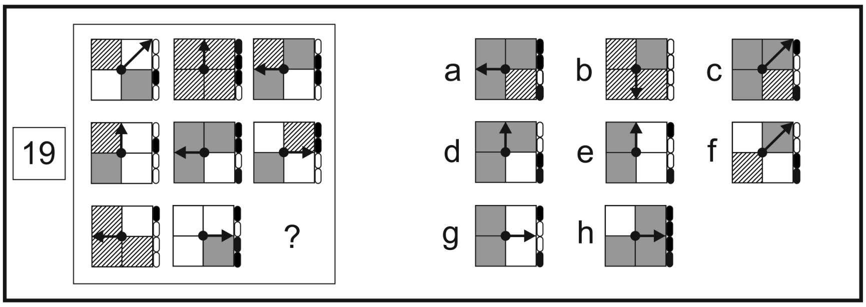


Abbildung 4.14 Beispielitem (Eigenentwicklung), das den Items des WMT2 ähnelt

Reliabilität

Gütekriterien der klassischen Testtheorie

In Tabelle 4.13 sind die Reliabilitäten für die vorliegende Stichprobe, die Normstichprobe aus dem Testhandbuch, sowie einer weiteren Vergleichsstichprobe aufgeführt. Die geringere interne Konsistenz bei der vorliegenden Stichprobe deutet an, dass der WMT2 nicht

Tabelle 4.13
Reliabilität des Wiener Matrizen-Tests 2 bezüglich der aktuellen Stichprobe im Vergleich mit anderen Untersuchungen

Untersuchung	*N*	Cronbachs Alpha	Split-Half-Reliabilität
aktuelle Stichprobe (Sommersemester 2014=MZP2)	993	.64	.62
Normstichprobe (Gesamtstichprobe Österreich), Formann et al. (2011)	2494	.82	.82
Vergleichsstichprobe (Schülerinnen und Schüler, 14–25 Jahre) von Marakovits (2008)	660	.76	.71

optimal zu den Teilnehmenden passte. Dies wird auch in der linken Grafik von Abbildung 4.15 deutlich: Von den 993 Teilnehmenden haben 152 die maximale Punktzahl erzielt. Daraus resultiert ein Deckeneffekt, der zu einer Varianzeinschränkung führt. Die rechte Grafik in Abbildung 4.15 zeigt zum Vergleich die Punkteverteilung in der Normstichprobe, deren Zusammensetzung der vorliegenden Stichprobe ähnelt. Auch hier zeigt sich ein deutlicher Deckeneffekt und ein ähnlicher Leistungsverlauf. Insgesamt kann die Reliabilität bezüglich der vorliegenden Stichprobe trotz Deckeneffekt als ausreichend angesehen werden, da die Leistung korrekt abgebildet wird.

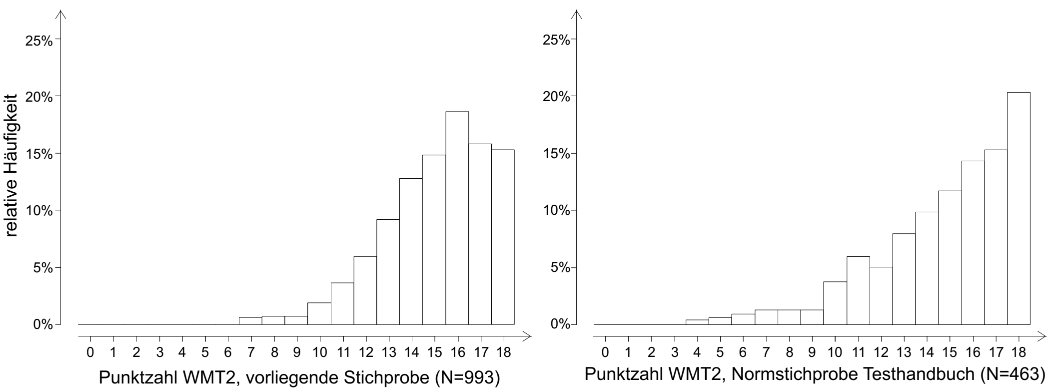


Abbildung 4.15 Punkteverteilung in der vorliegenden Stichprobe (links) und einer Normstichprobe mit ähnlicher Teilnehmendenzusammensetzung aus dem Testhandbuch (österreichische Erwachsene, 19–25 Jahre, Hochschulreife)

Gütekriterien der probabilistischen Testtheorie

Laut Testhandbuch genügt der WMT2 den Voraussetzungen des Rasch-Modells (siehe Seite 78). Von den dort durchgeführten grafischen Modelltests können an der vorliegenden Stichprobe jene mit Mittelwert und Geschlecht als Teilungskriterium nachvollzogen werden.

Von den 18 Aufgaben erweisen sich dabei in der vorliegenden Stichprobe lediglich zwei Items als nicht Rasch-konform. Anhand dieses Kriteriums bestätigt sich die Eindimensionalität des WMT2 weitestgehend auch bei der vorliegenden Stichprobe.

Validität

Gemäß Testhandbuch misst der WMT2 den „Grad der individuellen Fähigkeit schlussfolgernden Denkens im Umgang mit abstrakten Symbolen“ (Formann et al., 2011, S. 21). Eine Korrelation in Höhe von $r = .74$ zu den *Standard Progressive Matrices* (SPM) von Raven und Court (1998) sowie $r = .69$ zu dem *Dreidimensionalen Würfeltest* (3DW) von Gittler (1990) weisen auf konvergente Validität hin. Die geringere Korrelation von $r = .36$ mit dem *Zahlenverbindungstests* (ZVT) von Oswald und Roth (1982), der als Konzentrationstest eingestuft werden kann, spricht für diskriminante Validität.

Als weiteres Indiz für konvergente Validität wird die Diskriminationsfähigkeit des WMT2 zwischen Schülerinnen und Schülern mit sehr guten Leistungen im Fach Mathematik gegenüber jenen mit schlechteren Mathematikleistungen angeführt. Schülerinnen und Schüler mit der Note „sehr gut“ in Mathematik haben gemäß Testhandbuch signifikant bessere Testleistungen im WMT2 (Median= 12) als die Schülerinnen und Schüler mit schlechteren Noten (Median = 11).

4.4 Instrumente der Datenerhebung für affektive Faktoren

4.4.1 Prüfungsangstfragebogen TAI-G

Gemäß Fehm und Fydrich (2011) ist Prüfungsangst eine „anhaltende und deutlich spürbare Angst in Prüfungssituationen und/oder während der Zeit der Prüfungsvorbereitung, die den Bedingungen der Prüfungsvorbereitung und der Prüfung selbst nicht angemessen ist“ (S. 7). Prüfungsangst äußert sich auf emotionaler und kognitver Ebene (vgl. Abschnitt 2.3.1) und wird daher meistens durch mehrdimensionale Testinstrumente operationalisiert, wie z. B. durch das *Test Anxiety Inventory-German* (TAI-G). Das TAI-G wurde auch im Rahmen der vorliegenden Arbeit eingesetzt. Es wird daher im Folgenden eingehend beschrieben.

Testbeschreibung

Das TAI-G basiert auf einer vierfaktoriellen Struktur mit 30 Items, die den Subskalen *Besorgtheit*, *Aufgeregtheit*, *Interferenz* und *Mangel an Zuversicht* zugeordnet sind. Bei-

spielitems der einzelnen Skalen lauten: „Ich denke darüber nach, wie wichtig mir die Klausur oder Prüfung ist.“, „Ich spüre ein komisches Gefühl im Magen.“, „Mir schießen plötzlich Gedanken durch den Kopf, die mich blockieren.“ bzw. „Ich bin zuversichtlich, was meine Leistung betrifft.“. Das Antwortformat ist vierstufig: „stimmt nicht“, „stimmt kaum“, „stimmt eher“, „stimmt genau“.
Das TAI-G wurde für die vorliegende Arbeit ausgewählt, da es gut erforscht ist und mit einer Bearbeitungszeit von ca. 15 Minuten eine hohe Testökonomie aufweist. Der Fragebogen wurde mit Zustimmung seines Autors, Prof. Dr. Volker Hodapp, eingesetzt. Die Teilnahme war freiwillig und auf Wunsch erfolgte eine differenzierte Rückmeldung der Ergebnisse.

Reliabilität und Zuverlässigkeit der Ergebnisse

Alle Skalen weisen eine sehr gute und mit anderen Untersuchungen vergleichbare Reliabilität zwischen $\alpha = .81$ und $\alpha = .92$ auf. Es ist daher davon auszugehen, dass der Fragebogen von den teilnehmenden Studierenden ehrlich und gewissenhaft bearbeitet wurde und folglich die Ergebnisse für weitere Analysen geeignet sind.
Die Zuverlässigkeit der Ergebnisse wird ferner dadurch bestärkt, dass die vierfaktorielle Struktur des TAI-G eindeutig repliziert werden konnte. Darüber hinaus stimmen auch weitere wichtige Kennzahlen (Mittelwerte, Standardabweichungen, Trennschärfe der Items, Interskalenkorrelation) sehr gut mit entsprechenden Werten aus der Literatur überein.

Validität

Für die Validität des TAI-G lassen sich folgende Gründe anführen: Einerseits wurden entsprechende Analysen während der Entwicklung durchgeführt (Hodapp, 1991), andererseits wurde die Validität des Prüfungsangstfragebogens PAF (Hodapp et al., 2011) bestätigt, dessen Items eine echte Teilmenge der Items des TAI-G darstellen. Ferner steht das TAI-G am Ende einer rund vierzigjährigen Entwicklung psychometrischer Instrumente zur Messung von Prüfungsangst. Darüber hinaus kann auch die häufige Verwendung des TAI-G als Bestätigung seiner Validität gelten.

4.4.2 Prokrastinationsskala von Schwarzer

Prokrastination wird von den meisten Autorinnen und Autoren als dysfunktionale Eigenschaft angesehen, die im Wesentlichen das Aufschieben einer wichtigen Handlung trotz drohender negativer Konsequenzen beinhaltet (vgl. Abschnitt 2.4.1). Im Rahmen der vorliegenden Arbeit erfolgte die Operationalisierung durch die Prokrastinationsskala von Schwarzer (1999), die im Folgenden näher beschrieben wird.

Testbeschreibung

Mit der Prokrastinationsskala von Schwarzer (1999) wird die „Tendenz von Personen, Handlungen aufzuschieben, selbst wenn dieser Aufschub unter Umständen vorhersehbare streßreiche Konsequenzen hat“ (Schwarzer, 1999) gemessen. Es wird die allgemeine Prokrastination als Persönlichkeitsmerkmal erfasst (vgl. Abschnitt 2.4.1). Die Messung erfolgt mithilfe eines Fragebogens durch Selbstauskunft der Teilnehmenden.
Die Skala besteht aus zehn Items, von denen die Hälfte positiv und die andere Hälfte negativ gepolt ist. Beispielitems lauten „Ich komme oft erst nach Tagen dazu, Dinge zu tun, die ich eigentlich sofort erledigen wollte.“, „Ich beginne jeden Tag mit einer klaren Vorstellung davon, was ich schaffen will“. Das Antwortformat ist vierstufig: „stimmt nicht“, „stimmt kaum“, „stimmt eher“, „stimmt genau“.

Für die vorliegende Arbeit wurde die genannte Skala zur Untersuchung von Prokrastination ausgewählt, weil sie nur wenige Items umfasst und damit gut in Kombination mit umfangreicheren Tests einsetzbar ist. Die Teilnahme war freiwillig und auf Wunsch erfolgte eine differenzierte Rückmeldung der Ergebnisse.

Reliabilität und Zuverlässigkeit der Ergebnisse

Die Skala weist eine gute und mit anderen Untersuchungen vergleichbare Reliabilität auf (vgl. Tabelle 4.14). Dies deutet auf eine ehrliche Bearbeitung durch die Teilnehmerinnen und Teilnehmer hin, sodass die Ergebnisse für weitere Analysen geeignet sind.

Tabelle 4.14
Reliabilität der Prokrastinationsskala bezüglich der aktuellen Stichprobe im Vergleich mit anderen Untersuchungen

Untersuchung	*N*	Cronbachs Alpha	Split-Half-Reliabilität
aktuelle Stichprobe (WS13/14)	857	.80	.80
Schwarzer (1999)	293	.84	nicht berichtet
Eckstein (2001)	63	.77	nicht berichtet

Validität

Bezüglich der Validität bemerkt Schwarzer (1999), dass sich erwartungsgemäß „vor allem Zusammenhänge zu Selbstwirksamkeitserwartungen, Selbstregulation sowie zur Einschätzung stressreicher Ereignisse als Bedrohung“ zeigen. Darüber hinaus liefern Studien von Kröner und Fritzsche (2015), König und Kleinmann (2004) sowie Eckstein (2001) im Vergleich mit Korrelationsanalysen aus Metastudien (Van Eerde, 2003; Steel, 2007) deutliche Hinweise sowohl auf konvergente als auch auf divergente Validität. Die Prokrastinationsskala von Schwarzer kann somit als valide angenommen werden.

4.5 Methoden der Datenauswertung

Im Folgenden werden die Methoden der Datenauswertung vorgestellt. Die Reihenfolge orientiert sich dabei an der Reihenfolge der Forschungsfragen.

Bei den Untersuchungen zu Forschungsfrage 1 werden drei verschiedene Typen von Korrelationen bestimmt. Neben Produkt-Moment-Korrelationen nach Pearson werden Rangkorrelationen nach Spearman sowie entsprechende Partialkorrelationen berechnet. Hierbei handelt es sich um Standardmethoden, die beispielsweise in Bortz und Weber (2005, S. 204, S. 232, S. 443) beschrieben werden.
Falls zwischen den untersuchten Merkmalen kein linearer Zusammenhang angenommen werden kann, sollte alternativ zur Produkt-Moment-Korrelation die Rangkorrelation zur Beurteilung eines Zusammenhangs herangezogen werden. Bei der Angabe der entsprechenden Kennzahlen werden in der vorliegenden Arbeit Rangbindungen stets berücksichtigt. Dabei gelten Korrelationen bis 0.5 als gering, zwischen 0.5 und 0.7 als mittel und ab 0.7 als hoch (Nachtigall & Wirtz, 2009, S. 91).
Der Signifikanztest für die Produkt-Moment-Korrelation setzt eine bivariate Normalverteilung der Merkmale voraus. Der Test erweist sich allerdings als äußerst robust gegenüber Verletzungen dieser Annahme (Bortz & Weber, 2005, S. 214). Im Zweifel oder zur Absicherung kann auch hier alternativ der Signifikanztest zur Rangkorrelation herangezogen werden, der keine Verteilungsanforderungen stellt.
Die Daten wurden in allen Fällen auf Ausreißer geprüft, da diese insbesondere die Produkt-Moment-Korrelation verzerren können. In der vorliegenden Arbeit ergaben sich diesbezüglich keine Auffälligkeiten.
Für die Analysen werden die Methoden aus den R-Paketen „ppcor“ und „Hmisc“ verwendet.

Für die Untersuchung von Forschungsfrage 2 werden multivariate lineare und logistische Regressionsanalysen durchgeführt (Backhaus et al., 2011, S. 55, S. 509). Um zu klären, welche Tests bzw. welche Kombination von Tests die Klausurleistung am besten vorhersagen, wird die sogenannte schrittweise lineare Regression mit Rückwärts-Elimination (Sachs & Hedderich, 2006, S. 573) verwendet. Dabei werden nach und nach Tests und Variablen ausgeschlossen, deren Anteil an der erklärten Varianz der Klausurpunktzahl gering ist. Um die berechneten Regressionsgewichte der unabhängigen Variablen vergleichen zu können,

werden diese standardisiert (standardisierte β-Koeffizienten, vgl. Schroeder, Sjoquist & Stephan, 1986, S. 31). Für das Ausgangsmodell wurde die Normalverteilung der Residuen mithilfe eines q-q-Plots inspiziert und die Homoskedastizität der Residuen visuell durch ein Streudiagramm überprüft (Backhaus et al., 2011, S. 103). In beiden Fällen ergaben sich keine Anhaltspunkte für die Vermutung von Prämissenverletzungen.
Die Prognosegüte der verschiedenen Prädiktoren zur Vorhersage von Klausurerfolg wird mithilfe logistischer Regression und Bootstraping wie folgt bestimmt: Zunächst wird aus der Stichprobe eine 400 Teilnehmer umfassende Lernstichprobe zufällig ausgewählt. Anhand dieser Lernstichprobe erfolgt die Modellbildung. Der verbleibende Teil der Stichprobe dient als Teststichprobe zur Modellüberprüfung. Eine Prognose des Klausurerfolgs wird ausschließlich für Studierende aus der Teststichprobe vorgenommen, um ein Overfitting zu vermeiden (*out of sample*-Prognose). Dieser Prozess wird 2000-mal wiederholt. Durch dieses Bootstrapping ergibt sich eine Verteilungsfunktion der Prognosegüte (in der Regel eine Normalverteilung), auf deren Grundlage sich der Mittelwert der Prognosegüte inklusive Streuung und Konfidenzintervallen schätzen lässt (von Davier, 1997). Dieses Vorgehen ist weitaus robuster als die einmalige Modellbildung und erlaubt darüber hinaus die Angabe von Konfidenzintervallen für die Prognosegüte.
Alle Analysen werden mithilfe der Methoden des R-Pakets „stats“ durchgeführt.

Im Rahmen der Untersuchungen der Forschungsfragen 3 und 4 werden Mittelwertunterschiede auf Signifikanz überprüft. In der Regel unterscheiden sich die Varianzen in den untersuchten Gruppen, sodass in diesem Fall als Signifikanztest der (zweiseitige) Welch-Test (Hatzinger, Hornik & Nagel, 2011, S. 338) anstatt des (zweiseitigen) t-Tests (Bortz & Weber, 2005, S. 136) bei unverbundenen Stichproben eingesetzt wird. Da der Stichprobenumfang bei allen Untersuchungen größer als $N = 30$ ist, kann die Normalverteilungsannahme an die Merkmale vernachlässigt werden (Hatzinger et al., 2011, S. 338). Aus den Teststatistiken der eben genannten Signifikanztests wird als Maß für die Effektstärke Cohens d berechnet (Borenstein, 2009, S. 228). Im Fall von verbundenen Stichproben wird die Effektstärke gemäß Dunlap, Cortina, Vaslow und Burke (1996, S. 171) bestimmt.
Unterschiede in den Leistungsgruppen bezüglich dichotomer Hintergrundmerkmale werden mithilfe eines χ^2-Tests (Nachtigall & Wirtz, 2009, S. 165) auf Signifikanz untersucht. Die Berechnung der Effektstärke erfolgt in diesem Fall mithilfe der zugrunde liegenden Kontingenztafel gemäß Hasselblad und Hedges (1995, S. 170). Die Interpretation kann gemäß Nachtigall und Wirtz (2008, S. 160) nach dem Klassifikationsschema von Cohen erfolgen (Cohen, 1988). Demnach liegt bei $|d| \approx 0.2$ ein schwacher, bei $|d| \approx 0.5$ ein mittlerer und ab $|d| \approx 0.8$ ein starker Effekt vor.
Bei den Untersuchungen der Forschungsfrage 3 werden auch die Scheiterungsgründe bei der Bearbeitung prozeduraler Aufgaben, d. h. Kalkülfertigkeit und Kalkülkenntnis (vgl.

Abschnitt 1.1.1), analysiert. Dazu wurde in einer Voranalyse bei jedem Item entschieden, ob dieses korrekt gelöst oder aufgrund eines Mangels an Kalkülkenntnis falsch oder nicht gelöst wurde. Nicht bearbeitete Items wurden hierbei ebenfalls einem Mangel an Kalkülkenntnis zugerechnet, da dies als der naheliegendste Grund für eine Nichtbearbeitung erscheint und in diesem Fall teilweise auch so auf den Bearbeitungsbögen von den Studierenden zurückgemeldet wurde. Falls das Item nicht aufgrund von Kalkülkenntnis falsch gelöst wurde, kam als Scheiterungsgrund nur Kalkülfertigkeit in Betracht, was in diesem Fall automatisch kodiert wurde. Die Rater hatten also zu entscheiden, ob ein Item mangels Kalkülfertigkeit nicht korrekt gelöst wurde oder ob es korrekt gelöst wurde. Die Beurteilung erfolgte nach detailliert festgelegten Richtlinien eines Kodierhandbuchs. Exemplarisch seien hier die Kodierrichtlinien für das Item Horner1 (vgl. Abbildung 4.5) ungekürzt wiedergegeben:

- Kalkülkenntnis gilt als vorhanden, sofern das Horner-Schema korrekt angewendet wird (natürlich nur von der Form her, nicht bzgl. Rechenfehlern).
  Wenn am Ende mit dem $\pm\sqrt{-4}$-Ausdruck falsch umgegangen wird (z. B. weil $\sqrt{-4} = -2$ gerechnet wird oder die Rechnung abgebrochen wird), ist wie folgt zu verfahren:
  - Wird $\sqrt{-4}$ falsch berechnet und damit weitergerechnet und ergibt sich am Ende das korrekte Ergebnis (weil sich dieser Fehler aufhebt bzw. nicht auswirkt), dann lassen wir das durchgehen und werten es als Flüchtigkeitsfehler, der sich nicht weiter auswirkt.
  - Wird $\sqrt{-4}$ falsch berechnet und damit weitergerechnet und ergibt sich am Ende eine falsches Ergebnis oder wird die Rechnung abgebrochen, weil der/die Studierende mit dem Ausdruck $\sqrt{-4}$ nicht klarkommt, werten wir das nicht als Schwäche in der Kalkülkenntnis. Denn zu wissen, was $\sqrt{-4}$ ist, zählt zu den Elementarkompetenzen, die eine Facette der Kalkülfertigkeit darstellen.
- Manchmal verrechnen sich die Studierenden, erhalten am Ende des Horner-Schemas keine „0“ und rechnen einfach weiter. Das werten wir nicht als Schwäche in der Kalkülkenntnis, da der oder die Studierende offenbar weiß, wie das Horner-Schema zur Nullstellenbestimmung anzuwenden ist. Ohne den Rechenfehler bekäme sie/er ja das korrekte Ergebnis heraus.
- Einige spalten auch $-\frac{1}{2}$ statt $\frac{1}{2}$ bzw. $\frac{3}{2}$ statt $-\frac{3}{2}$ ab. Geschieht das immer mit dem falschen Vorzeichen, werten wir das als Schwäche in der Kalkülkenntnis, da die Person zu keinem korrekten Ergebnis gelangen kann. Wird einmal ein falsches, einmal ein richtiges Vorzeichen verwendet, werten wir das dagegen nicht als Schwäche in der Kalkülkenntnis, sondern als Flüchtigkeitsfehler. Falls (z. B. aus Zeitgründen) nur ein Horner-Schema durchgeführt wird und $-\frac{1}{2}$ statt $\frac{1}{2}$ bzw. $\frac{3}{2}$ statt $-\frac{3}{2}$ verwendet wird, werten wir das als Schwäche in der Kalkülkenntnis.
- Im Horner-Schema sind stets die Zahlen in den ersten beiden Zeilen zu addieren, um die Einträge in der letzten Zeile und damit die Koeffizienten des Restpolynoms zu erhalten. Einige Studierende subtrahieren die Terme aber, anstatt sie zu addieren. Falls

> die Studierenden bei beiden Horner-Schemata Zeilenelemente subtrahieren, werten wir das als Schwäche in der Kalkülkenntnis (weil das Kalkül nicht richtig verstanden ist). Falls die Studierenden bei einem Horner-Schema subtrahieren, beim anderen addieren, werten wir das als Flüchtigkeitsfehler (und damit nicht als Schwäche in der Kalkülkenntnis). Falls (z. B. aus Zeitgründen) nur ein Horner-Schema durchgeführt wird und die Zeilenelemente subtrahiert werden, dann werten wir das als Schwäche in der Kalkülkenntnis.

Die Bearbeitung jedes Items wurde von zwei unabhängigen Bewertungsteams analysiert und wie oben erläutert kodiert. In Tabelle 4.15 sind die Interraterreliabilität sowie die prozentuale Beurteilerübereinstimmung (Wirtz & Caspar, 2002; Fleiss, Levin & Paik, 2003, S. 599) angegeben. Zur Bestimmung der Interraterreliabilität wurde der Koeffizient κ_n von Brennan und Prediger (1981, S. 693) verwendet, da im vorliegenden Fall die Randhäufigkeiten nicht festgelegt sind.

Legt man die Bewertung von Landis und Koch (1977, S. 165) zur Beurteilung des verwandten Koeffizienten Cohens Kappa zugrunde, so ist die Interraterreliabilität als „fast perfekt" zu bewerten, was der besten von fünf Kategorien entspricht. Unabhängig davon wurden bei allen abweichenden Beurteilungen der beiden Bewertungsteams die betroffenen Items durch den Autor untersucht und in Form eines Expertenratings abschließend bewertet. Ferner lagen von 836 Items aus technischen Gründen keine Doppelsichtungen vor. Auch hier erfolgte die abschließende Bewertung durch den Autor.

Tabelle 4.15

Interraterreliabilität κ_n nach Brennan und Prediger (1981, S. 693) und prozentuale Bewerterübereinstimmung (PÜ) bei der Beurteilung der Korrektheit bzw. einer Schwäche in der Kalkülkenntnis bei der Bearbeitung der Items

	Testvarianten Semester 1-A und 1-B	Testvariante Semester 2-A	Testvariante Semester 2-B	Testvariante Semester 3	$\sum$ / ∅
Inspizierte Aufgaben	18263	4400	6479	5507	34649
κ_n	0.90	0.93	0.90	0.92	0.91
PÜ (in %)	91.63	96.41	95.23	95.79	93.57

5 Ergebnisse

In dem vorliegenden Kapitel werden die Untersuchungsergebnisse zu den in Kapitel 3 formulierten Forschungsfragen in chronologischer Reihenfolge vorgestellt, diskutiert und interpretiert.

5.1 Zusammenhänge zwischen Klausurleistung und fachspezifischen, allgemein kognitiven sowie affektiven Faktoren

Frage und Vorgehensweise

Im Rahmen von Forschungsfrage 1 (*Wie groß sind die Zusammenhänge zwischen Klausurleistung und fachspezifischen, allgemein kognitiven sowie affektiven Faktoren?*) wird untersucht, inwiefern die Klausurleistung mit verschiedenen Faktoren zusammenhängt. Dazu werden die Korrelationen zwischen den einzelnen Faktoren (Prädiktorvariablen) und der Klausurleistung (Zielvariable) berechnet. Die Klausurleistung wird dabei durch die in der Klausur „Höhere Mathematik I“ im Wintersemester 2013/14 erreichte Punktzahl (vgl. Abschnitt 4.1.4) operationalisiert. Die Operationalisierung der Prädiktorvariablen wird in Tabelle 4.1 beschrieben.

Neben den Produkt-Moment-Korrelationen nach Pearson (Bortz & Weber, 2005, S. 204) werden zusätzlich die Rangkorrelationen nach Spearman (Bortz & Weber, 2005, S. 232) ausgewiesen. Falls beide Korrelationen stark differieren, müsste eine vertiefte Inspektion der Daten erfolgen, um die Ursache dafür zu finden. Eine mögliche Ursache wurde hier vorab ausgeschlossen, indem alle Daten auf Ausreißer geprüft wurden (vgl. Abschnitt 4.5). Da auch die Prädiktorvariablen korrelieren können, ist in jeder Korrelation zwischen einer Prädiktorvariablen und der Zielvariablen in der Regel ein Teil der Korrelation auf Zusammenhänge der übrigen Prädiktorvariablen mit der untersuchten Variablen und der Zielvariablen zurückzuführen. Um solche Scheinkorrelationen zu berücksichtigen, müssen aus der Korrelation zwischen einer Prädiktorvariablen und der Zielvariablen die Einflüsse der anderen Prädiktorvariablen herausgerechnet werden. Dies geschieht üblicherweise durch die Bestimmung von Partialkorrelationen (Bortz & Weber, 2005, S. 443). Im vorliegenden Fall wurde auf diese Weise für den jeweiligen Prädiktor der gemeinsame Zusammenhang zwischen allen anderen untersuchten Prädiktorvariablen und der Zielvariablen herauspartialisiert. Da hierzu bei jedem Studierenden der Stichprobe Informationen über alle berücksichtigten Prädiktorvariablen vorliegen müssen, ist der Stichprobenumfang in diesem Fall entsprechend geringer als die Teilnehmendenzahl der einzelnen Tests.

Ergebnisse

In Tabelle 5.1 sind oberhalb der Hauptdiagonalen die Produkt-Moment-Korrelation nach Pearson und unterhalb der Hauptdiagonalen die Rangkorrelationen nach Spearman angegeben. Erwartungskonform unterscheiden sich die jeweiligen Korrelationen kaum, sodass keine weitere Inspektion der Daten erfolgen muss.
Tabelle 5.2 enthält oben rechts bzw. unten links die entsprechenden Partialkorrelationen der Prädiktorvariablen mit der Klausurleistung. Auf die Berücksichtigung der Arbeitsgeschwindigkeit beim Konzentrationstest wurde hier verzichtet, da diese fast perfekt mit der Arbeitsleistung korreliert und letztere als Repräsentant der Tempoleistung im Konzentrationstest ausreicht (vgl. Abschnitt 2.2.2). Die Partialkorrelationen geben den um Scheinkorrelationen bereinigten Zusammenhang zwischen Prädiktor- und Zielvariable an. Die Zusammenhänge zwischen Prädiktorvariablen und Klausurleistung sind jeweils fett gedruckt.

Tabelle 5.1

Produkt-Moment-Korrelation (oberes Dreieck) und Rangkorrelation (unteres Dreieck) zwischen Prädiktorvariablen und Klausurpunktzahl

	Klausur-punktzahl	Leistungs-test proze-durales Wissen	Leistungs-test konzept-uelles Wissen	Wiener Matrizen-Test 2	Konzen-tration Arbeits-leistung	Konzen-tration Arbeits-geschwin-digkeit	Konzen-tration Fehler-neigung	Prüfungs-angst	Prokras-tination
Klausurpunktzahl		$\mathbf{.63^{**}}$	$\mathbf{.44^{**}}$	$\mathbf{.20^{**}}$	$\mathbf{.22^{**}}$	$\mathbf{.16^{**}}$	$\mathbf{.21^{**}}$	$\mathbf{-.12^{**}}$	$\mathbf{-.13^{**}}$
Leistungstest prozedurales Wissen	$\mathbf{.64^{**}}$		$.43^{**}$	$.25^{**}$	$.29^{**}$	$.22^{**}$	$.25^{**}$	$-.13^{**}$	$-.06$
Leistungstest konzeptuelles Wissen	$\mathbf{.41^{**}}$	$.45^{**}$		$.24^{**}$	$.12^{**}$	$.07^{*}$	$.16^{**}$	$-.18^{**}$	$.02$
Wiener Matrizen-Test 2	$\mathbf{.19^{**}}$	$.25^{**}$	$.24^{**}$		$.24^{**}$	$.17^{**}$	$.24^{**}$	$-.09^{*}$	$.07$
Konzentration Arbeitsleistung	$\mathbf{.21^{**}}$	$.28^{**}$	$.11^{**}$	$.24^{**}$		$.94^{**}$	$.34^{**}$	$-.13^{**}$	$-.04$
Konzentration Arbeitsgeschwindigkeit	$\mathbf{.16^{**}}$	$.21^{**}$	$.05^{*}$	$.18^{**}$	$.93^{**}$		$-.01$	$-.12^{**}$	$-.04$
Konzentration Fehlerneigung	$\mathbf{.18^{**}}$	$.24^{**}$	$.15^{**}$	$.23^{**}$	$.28^{**}$	$-.03$		$-.07$	$-.01$
Prüfungsangst	$\mathbf{-.12^{**}}$	$-.15^{**}$	$-.19^{**}$	$-.09^{*}$	$-.12^{**}$	$-.12^{**}$	$-.05$		$.31^{**}$
Prokrastination	$\mathbf{-.13^{**}}$	$-.08^{*}$	$.02$	$.06$	$-.04$	$-.04$	$-.02$	$.30^{**}$	

Anmerkung. $^{*}p < .05$, $^{**}p < .01$; minimale Stichprobengröße: $N = 616$, maximale Stichprobengröße: $N = 1310$

Tabelle 5.2

Produkt-Moment-Partialkorrelation (oberes Dreieck) und Partialkorrelation nach Spearman (unteres Dreieck) zwischen Prädiktorvariablen und Klausurpunktzahl

	Klausur-punktzahl	Leistungs-test proze-durales Wissen	Leistungs-test konzept-uelles Wissen	Wiener Matrizen-Test 2	Konzen-tration Arbeits-leistung	Konzen-tration Fehler-neigung	Prüfungs-angst	Prokras-tination
Klausurpunktzahl		**.49****	**.22****	**.08**	**−.02**	**.10***	**.12***	**−.15****
Leistungstest prozedurales Wissen	**.51****		.23**	.07	.22**	.00	−.01	.01
Leistungstest konzeptuelles Wissen	**.21****	.22**		.17*	.02	−.01	−.22**	.14**
Wiener Matrizen-Test 2	**.07**	.08	.18**		.11*	.15**	.05	.11*
Konzentration Arbeitsleistung	**.00**	.19**	−.00	.11*		.23**	−.03	−.07
Konzentration Fehlerneigung	**.07**	.02	.01	.12*	.18**		−.05	−.01
Prüfungsangst	**.11****	−.01	-.21**	−.06	−.02	−.02		.33**
Prokrastination	**-.13****	.01	.12**	.10*	.08	.00	.31**	

Anmerkung. $^{*}p < .05$, $^{**}p < .01$; Stichprobengröße: $N = 454$

Interpretation der Ergebnisse im Hinblick auf Forschungsfrage 1

Im Folgenden werden die Ergebnisse zu den untersuchten fachspezifischen, allgemein kognitiven und affektiven Faktoren zusammengefasst und interpretiert.

Fachspezifische Faktoren: Die fachspezifischen Faktoren prozedurales und konzeptuelles Wissen weisen einen deutlichen Zusammenhang mit der Klausurleistung auf. Dabei ist prozedurales Wissen mit einer Korrelation von $r = .63$ (Pearson) bzw. $r = .64$ (Spearman) der dominierende Faktor. Auch nach einem Herauspartialisieren der anderen in Tabelle 5.2 aufgeführten Faktoren weisen Partialkorrelationen von $r = .49$ (Pearson) bzw. $r = .51$ (Spearman) auf einen deutlichen Zusammenhang zwischen prozeduralem Wissen und der Klausurleistung hin. Eine Einordnung der Höhe dieser Korrelationen erfolgt weiter unten.

Allgemein kognitive Faktoren: Die allgemein kognitiven Faktoren weisen mit Korrelationen zwischen $r = .22$ für Konzentration (Arbeitsleistung) und $r = .20$ für kognitive Grundfertigkeit (Wiener Matrizen-Test 2) einen schwachen und erwartungskonformen (vgl. Abschnitt 2.1.3 und 2.2.3) Zusammenhang mit der Klausurleistung auf. Letzteres ist gleichzeitig ein Indiz für die Repräsentativität der Stichprobe (vgl. Fazit in den Abschnitten 2.1 und 2.2).

Affektive Faktoren: In der Gesamtbetrachtung spielen die affektiven Faktoren Prüfungsangst und Prokrastination die geringste Rolle. In Einzelfällen kann das zwar anders sein

(Schaefer et al., 2009), bei der Betrachtung aller Studierenden weisen diese Faktoren jedoch keinen relevanten Zusammenhang mit der Klausurleistung auf. Die Korrelationen liegen betragsmäßig weit unterhalb von $r = .20$ und die Partialkorrelationen sind nicht signifikant (von Null verschieden). Dieses Resultat wird durch Untersuchungen von Fischer, Schult und Hell (2015) bestätigt, die bei einem Vergleich der Korrelationen von fachspezifischen und affektiven Faktoren mit Studiennoten zu folgendem Ergebnis gelangen: Während die höchste Korrelation zwischen einem fachspezifischen Studierfähigkeitstest und Studiennoten bei $r = -.49$ liegt, erreichen affektive Faktoren höchstens eine Korrelation von $r = -.24$. Affektive Faktoren weisen demnach auch hier im Vergleich zu fachspezifischen Faktoren einen viel schwächeren Zusammenhang mit Studiennoten bzw. der Klausurleistung auf.

Einordnung der Stärke der Zusammenhänge: Die vorangehenden Betrachtungen haben gezeigt, dass unter den untersuchten Faktoren das prozedurale Wissen den stärksten Zusammenhang mit der Klausurleistung aufweist. Die Produkt-Moment-Korrelation nach Pearson beträgt hier $r = .63$. Dies ist insofern bemerkenswert, da prozedurales Wissen durch die Fertigkeit, Basisprozeduren der Ingenieurmathematik (vgl. Abschnitt 1.1.1) in angemessener Zeit korrekt durchzuführen, operationalisiert wird (vgl. Abschnitt 4.2.1) und diese Fertigkeit nur einen kleinen Teil des komplexen Anforderungsspektrums von Klausuren der Ingenieurmathematik darstellt (vgl. Abschnitt 1.2.2).
Zur Einordnung der Stärke der Zusammenhänge werden im Folgenden einige Korrelationen zwischen Studienerfolg und Schulabschlussnoten angeführt. Letztere dienen in vielen Fächer der Auswahl von Studienbewerberinnen und -bewerbern und gelten als guter Prädiktor für Studienerfolg. In einer Meta-Studie untersuchten Schuler, Funke und Baron-Boldt (1990) sowie Trapmann et al. (2007) Zusammenhänge zwischen durchschnittlichen Schulabschlussnoten und dem Studienerfolg, wobei letzterer über Studiennoten operationalisiert wurde (Trapmann et al., 2007, S. 13). In den beiden Studien werden Korrelationen von $r = .46$ bzw. $r = .52$ berichtet. Damit sind durchschnittliche Schulabschlussnoten vermutlich bessere Prädiktoren für Klausurleistung und -erfolg als die in der vorliegenden Arbeit untersuchten allgemein kognitiven und affektiven Faktoren, deren Korrelationen mit der Klausurleistung maximal $r = .22$ betragen. Die Korrelation von $r = .63$ zwischen prozeduralem Wissen und der Klausurleistung erreichen Schulabschlussnoten dagegen deutlich nicht. Aus diesem Grund und da Schulabschlussnoten häufig nicht für eine Erfolgsprognose verfügbar sind, stellt der Leistungstest prozedurales Wissen möglicherweise eine adäquate Alternative zur Prognose von Klausurerfolg und -leistung dar. Dies wird im nächsten Abschnitt noch eingehender untersucht.
Einige Autorinnen und Autoren zeigen weiter, dass auch zwischen Studierfähigkeitstests und akademischer Leistung ein vergleichsweise hoher Zusammenhang bestehen kann. Hell,

Linsner und Kurz (2008, S. 177) berichten über eine Korrelation von $r = .53$ zwischen einem fachspezifischen Studierfähigkeitstest und der Klausur „Mathematik 1“ für Maschinenbau-Studierende. Gemäß Madison et al. (2015, S. 148) liegen die Korrelationen zwischen College-Einstufungstests („placement tests“) und den späteren Mathematikkursen in der Regel zwischen $r = .20$ und $r = .44$. Auch diese Korrelationen liegen deutlich unterhalb der Korrelation zwischen Klausurleistung und prozeduralem Wissen, was dessen bereits in Abschnitt 1.2 angedeutete Relevanz in der Ingenieurmathematik nochmals unterstreichen könnte.

Fazit

Im Hinblick auf Forschungsfrage 1 ist festzuhalten, dass fachspezifische Faktoren einen deutlichen Zusammenhang mit der Klausurleistung aufweisen, während allgemein kognitive Faktoren und Klausurleistung nur schwach zusammenhängen und affektive Faktoren diesbezüglich kaum relevant sind. Unter allen untersuchten Prädiktoren weist prozedurales Wissen den stärksten Zusammenhang mit der Klausurleistung auf. Die gemessene Korrelation liegt sogar deutlich über den in der Literatur berichteten Korrelationen zwischen Klausur- bzw. Studienleistung und anderen Faktoren, für die ein Zusammenhang zur akademischen Leistung nachgewiesen ist. Zu diesen Faktoren zählen beispielsweise die durchschnittliche Schulabschlussnote sowie das Abschneiden in Studierfähigkeitstests.
Der starke Zusammenhang zwischen der Klausurleistung und prozeduralem Wissen ist auch deshalb bemerkenswert, weil letzteres durch das korrekte Anwenden von Basisprozeduren in angemessener Zeit operationalisiert wird. Diese Fertigkeit stellt lediglich eine Grundanforderung im weitaus komplexeren Anforderungsprofil von Klausuren der Ingenieurmathematik dar und deckt damit nur einen kleinen Teil der Anforderungen ab.
An dieser Stelle ließe sich anmerken, dass prozedurales Wissen dennoch das im Hinblick auf die Klausur handlungsnächste der untersuchten Konstrukte darstellt, sowohl inhaltlich als auch mit Blick auf die Lehr/Lern- bzw. Prüfungssituation. Auch die Zeitbeschränkung ist in ähnlicher Weise gegeben. Folglich ließe sich einwenden, dass der im Vergleich zu den anderen Faktoren stärkste Zusammenhang zwischen Klausurleistung und prozeduralem Wissen nicht überrascht. Dagegen ist einzuwenden, dass die Kenntnis aller Basisprozeduren alleine nicht ausreicht, um die hier zugrunde liegende Klausur (vgl. Abschnitt 4.1.4) zu bestehen. Die korrekte und effiziente Anwendung von Basisprozeduren ist im vorliegenden Fall lediglich eine notwendige, aber keinesfalls eine hinreichende Bedingung für eine gute Klausurleistung. Darüber hinaus ist nicht allein der Vergleich der Korrelationen bemerkenswert, sondern die absolute Höhe, die prozedurales Wissen im vorliegenden Fall als einen Prädiktor ausweist, der auch vielen anderen aus der Literatur bekannten Faktoren überlegen zu sein scheint.

Abschließend ist aufgrund der Ergebnisse festzuhalten, dass die Fertigkeit, Basisprozeduren korrekt und effizient anwenden zu können, als prädiktiv isolierbare Grundanforderung in Klausuren angesehen werden kann (vgl. Kapitel 2), die daher auch aus fachdidaktischer Sicht als Grundlage von Förderinstrumenten dienen könnte.

5.2 Zur Prognose von Klausurerfolg und Klausurleistung

Im vorangegangenen Abschnitt wurden die Untersuchungen zu Forschungsfrage 1 vorgestellt. Dabei ging es darum herauszufinden, welche Faktoren besonders stark mit der Klausurleistung zusammenhängen. Ergänzend dazu wird im vorliegenden Abschnitt analysiert, wie gut sich Klausurergebnisse durch verschiedene Faktoren vorhersagen lassen.

Frage und Vorgehensweise

Im Rahmen von Forschungsfrage 2 (*In welchem Maße lassen sich Klausurerfolg und Klausurleistung vorhersagen?*) wird die Vorhersagbarkeit von Klausurerfolg und Klausurleistung behandelt. Der Klausurerfolg wird durch die beiden Kategorien „Klausur bestanden“ und „Klausur nicht bestanden“ operationalisiert. Die Klausurleistung drückt sich in der insgesamt erreichten Punktzahl in der Klausur aus (vgl. Abschnitt 4.1.4).
Mithilfe eines multivariaten logistischen Regressionsmodells (vgl. Abschnitt 4.5) wird untersucht, in welchem Umfang der Klausurerfolg durch die eingesetzten Tests prognostiziert werden kann. Anschließend wird durch ein lineares Regressionsmodell (ebendort) analysiert, in welchem Maße die Klausurleistung durch die einbezogenen Prädiktorvariablen erklärbar ist. In beiden Fällen soll unter den einbezogenen fachspezifischen, allgemein kognitiven und affektiven Faktoren der beste Einzelprädiktor ermittelt werden.

Während die Prognose des Klausurerfolgs ein grobes Erfolgsmaß ist, bei dem eine Aussage über lediglich zwei – wenngleich sehr wichtige – Kategorien („bestanden“ oder „nicht bestanden“) getroffen wird, ist die Prognose der Klausurleistung, die im vorliegenden Fall zwischen 0 und 80 Punkten liegt, wesentlich differenzierter.

Es soll der Erfolg bzw. die Leistung in der Referenzklausur „Höhere Mathematik I“ aus dem Wintersemester 2013/14 (vgl. Abschnitt 4.1.4) prognostiziert werden. Als Maß für die Prognosegüte bezüglich des Klausurerfolgs gilt der Anteil korrekter Prognosen und bezüglich der Klausurleistung die Höhe der erklärten Varianz der Klausurleistung.

Ergebnisse zur Prognose des Klausurerfolgs

An der genannten Referenzklausur nahmen 1552 Studierende teil, von denen 676 die Klausur nicht bestanden (vgl. Abschnitt 4.1.4). Daraus resultiert eine proportionale Zufallswahrscheinlichkeit von 52.0 % (Backhaus et al., 2011, S. 273). Würde also die Entscheidung „Klausur bestanden“ bzw. „Klausur nicht bestanden“ bei jeder teilnehmenden Person zufällig getroffen, so ergäbe sich in 52 % der Fälle eine korrekte Prognose. Das bedeutet, dass die Prognose all jener Tests als „nicht zufällig“ eingestuft werden kann, bei denen die proportionale Zufallswahrscheinlichkeit kleiner als die untere Grenze des entsprechenden Konfidenzintervalls ist. D. h. diese Tests haben eine dem Zufall überlegende Prognosegüte. In Tabelle 5.3 sind die Prognosegüten zur Vorhersage des Klausurerfolgs („Klausur bestanden“ versus „Klausur nicht bestanden“) dargestellt. Die Trennlinien dienen zur besseren Unterscheidung zwischen den fachspezifischen, allgemein kognitiven und affektiven Prädiktoren. Unter „Auswahl“ ist diejenige Testkombination angegeben, die zu der höchsten Prognosegüte führt. Damit ist gemeint, dass ein Verändern dieser Testkombination durch Weglassen oder Hinzunahme von Tests zu einer Reduzierung der Prognosegüte führt. Folglich ist die angegebene Prognosegüte von 75.6 % die mithilfe der in die vorliegenden Arbeit einbezogenen Tests maximal erreichbare Prognosegüte.

Tabelle 5.3

Prognose des Klausurerfolgs (bestanden/nicht bestanden) durch Einzeltests und Testkombinationen

Nr.	Test	Simulationen	*N*	*N* Lernstichprobe	*N* Teststichprobe	**Korrekte Prognose (in %)**	*SD* (in %)	95%-KI
(1)	Auswahl	2000	670	400	270	**75.6**	2.1	[71.5, 79.7]
(2)	Leistungstest prozedurales Wissen	2000	1286	400	886	**74.5**	0.9	[72.7, 76.3]
(3)	Leistungstest konzeptuelles Wissen	2000	1359	400	959	**65.9**	0.8	[64.3, 67.5]
(4)	Konzentrationstest Arbeitsleistung	2000	1158	400	758	**61.3**	1.5	[58.4, 64.2]
(5)	Konzentrationstest Arbeitsgeschwindigkeit	2000	1158	400	758	**59.6**	3.9	[52.0, 67.2]
(6)	Konzentrationstest Sorgfalt	2000	1158	400	758	**61.8**	1.7	[58.5, 65.1]
(7)	Wiener Matrizen-Test 2	2000	943	400	543	**62.2**	2.9	[56.5, 67.9]
(8)	Prüfungsangst	2000	857	400	457	**55.3**	5.4	[44.7, 65.9]
(9)	Prokrastination	2000	857	400	457	**54.8**	5.5	[44.0, 65.6]
	proportionale Zufallswahrscheinlichkeit	-	1552	-	-	**52.0**		

Anmerkung. Auswahl=(2)+(3)+(4)+(6)+(7); *SD*=Standardabweichung; KI=Konfidenzintervall

Als Einzeltest liefert der Speed- und Leistungstest prozedurales Wissen die mit Abstand beste Vorhersage. Mit im Schnitt 74.5 % korrekter Prognosen besitzt er nahezu dieselbe Prognosekraft wie die unter „Auswahl“ angegebene optimale Testkombination.

Ergebnisse zur Prognose der Klausurleistung

Um zu klären, welche Tests bzw. welche Kombination von Tests die Klausurleistung am besten vorhersagen, wird die sogenannte schrittweise lineare Regression mit Rückwärtselimination (Sachs & Hedderich, 2006, S. 573) verwendet. Dabei werden nach und nach Tests und Variablen ausgeschlossen, deren Anteil an der erklärten Varianz der Klausurpunktzahl gering ist. Die Ergebnisse sind in Tabelle 5.4 angegeben.

Tabelle 5.4

Schrittweise lineare Regression mit Rückwärtselimination

Modell	N	Prädiktor	β	p	$\tilde{R}^2$	$\Delta\tilde{R}^2$
saturiert	454	prozedurales Wissen	0.47	.000	0.444	0.444
		konzeptuelles Wissen	0.20	.000		
		Wiener Matrizen-Test 2	0.06	.108		
		Konzentration Arbeitsleistung	-0.01	.754		
		Konzentration Sorgfalt	0.09	.022		
		Prüfungsangst	0.10	.015		
		Prokrastination	-0.12	.001		
		Semester	-0.11	.004		
		Schulform	-0.03	.361		
		Schuldauer	0.03	.377		
		Kursform	-0.07	.056		
		Geschlecht	0.02	.576		
Modell 7	454	prozedurales Wissen	0.47	.000	0.443	0.001
		konzeptuelles Wissen	0.21	.000		
		Konzentration Sorgfalt	0.10	.009		
		Prüfungsangst	0.10	.011		
		Prokrastination	-0.12	.001		
		Semester	-0.10	.011		
		Kursform	-0.09	.030		
Modell 8	454	prozedurales Wissen	0.52	.000	0.429	0.014
		konzeptuelles Wissen	0.19	.000		
		Konzentration Sorgfalt	0.09	.018		
		Prokrastination	-0.08	.019		
Modell 9	454	prozedurales Wissen	0.54	.000	0.416	0.013
		konzeptuelles Wissen	0.19	.000		
Modell 10a	454	prozedurales Wissen	0.63	.000	0.390	0.026
Modell 10b	1286	prozedurales Wissen	0.63	.000	0.397	-

Anmerkung. β=standardisierter Beta-Koeffizient; $\tilde{R}^2$=korrigiertes R^2

In das Ausgangsmodell (saturiertes Modell=Modell 0) wurden neben den Testergebnissen noch zusätzliche zur Verfügung stehende Hintergrundmerkmale wie Semesterzahl (Regelstudienzeit versus höheres Semester=Wiederholer der Veranstaltung), Schulform (Gymnasium versus andere Schulform), Schuldauer (G8 versus G9), Kursform (Leistungskurs Mathematik versus Grundkurs Mathematik) sowie das Geschlecht aufgenommen. Das Ziel bei der Berücksichtigung dieser Hintergrundmerkmale ist es, ihren Einfluss auf die Zielvariable im Vergleich zu den anderen Faktoren zu untersuchen. Das ist wichtig, da Hintergrundmerkmale wesentlich einfacher zu erheben sind als Ergebnisse aus Tests, deren Organisation und Auswertung in der Regel einen erheblichen Aufwand darstellen. Sollte sich herausstellen, dass Hintergrundmerkmale eine ähnliche Prognosegüte bezüglich der Klausurleistung wie die anderen untersuchten Faktoren aufweisen, würden Hintergrundmerkmale möglicherweise für eine Vorhersage von Klausurleistung ausreichen und ein aufwendiges Testen wäre vermeidbar.

Die Modelle 1 bis 6 sind in der Tabelle nicht aufgeführt. Hier wurden sukzessive folgende Faktoren ausgeschlossen, die kaum einen Beitrag zur erklärten Varianz leisten: Kognitive Grundfertigkeit (operationalisiert durch den Wiener Matrizen-Test 2, vgl. Abschnitt 4.3.2), Konzentration (Arbeitsleistung) (vgl. Abschnitt 2.2 und 4.3.1), Schulform, Schuldauer und Geschlecht.
Somit gehen in Modell 7 neben den beiden fachspezifischen Faktoren prozedurales und konzeptuelles Wissen (vgl. Kapitel 1 und Abschnitt 4.2.1 bzw. 4.2.2) die beiden affektiven Faktoren Prüfungsangst (vgl. Abschnitt 2.3 und 4.4.1) und Prokrastination (vgl. Abschnitt 2.4 und 4.4.2) sowie Konzentration (Sorgfalt) (vgl. Abschnitt 2.2 und 4.3.1), die Semesterzahl und die Kursform mit einem auf dem 5 %-Niveau signifikant von 0 verschiedenen Einfluss ein.
Eine merkliche Reduktion der erklärten Varianz von $44,3$ % auf $42,9$ % bzw. $41,6$ % ergibt sich allerdings erst beim Übergang von Modell 7 zu Modell 8 bzw. 9, d. h. nach Ausschluss von Konzentration Sorgfalt und Prokrastination. Wird anschließend noch konzeptuelles Wissen ausgeschlossen, reduziert sich die erklärte Varianz in Modell 10a um weitere 2.6 %. Als einziger Prädiktor bleibt dann prozedurales Wissen übrig, das alleine 39 % der Varianz der Klausurleistung erklärt.
Die Modelle 10a und 10b zeigen damit, dass von der maximal erklärten Varianz je nach Stichprobengröße bis zu $39,7$ % allein durch prozedurales Wissen erklärt wird. Der Unterschied der beiden Modelle besteht darin, dass bei Modell 10a aus Gründen der Vergleichbarkeit dieselbe Stichprobe wie bei den vorangegangenen Modellen zugrunde liegt, während bei Modell 10b die maximal mögliche Stichprobengröße zugelassen wurde.
Im nächsten Abschnitt wird erläutert, wie sich die durch prozedurales Wissen erklärte Varianz noch steigern lässt.

Ergebnisse zur Aufspaltung prozeduralen Wissens in Kalkülkenntnis und Kalkülfertigkeit

In Abschnitt 1.1.1 wurde prozedurales Wissen als Verknüpfung von Kalkülkenntnis und Kalkülfertigkeit definiert. Hier wurde an Beispielen deutlich, dass Schwächen in der Kalkülkenntnis oder der Kalkülfertigkeit verschiedene Fehlerquellen bei der Bearbeitung prozeduraler Aufgaben darstellen. Berücksichtigt man diese Aufspaltung prozeduralen Wissens in der linearen Regression, bestätigt sich auch hier, dass Kalkülkenntnis und Kalkülfertigkeit zumindest teilweise unabhängig voneinander sind, da sich die erklärte Varianz nach der Aufspaltung um fast 7 % erhöht (vgl. Tabelle 5.5). In die Berechnungen geht jeweils ein, wie viele Aufgaben wegen einer Schwäche in Kalkülkenntnis bzw. Kalkülfertigkeit nicht oder nicht korrekt gelöst wurden. Aus diesem Grund sind die standardisierten β-Koeffizienten nun negativ und weisen erwartungsgemäß auf eine negative Korrelation zwischen der Klausurleistung und der Anzahl aufgrund von Schwächen in der Kalkülkenntnis bzw. Kalkülfertigkeit nicht oder nicht korrekt gelöster Aufgaben hin. Durch die Unterscheidung zwischen Kalkülfertigkeit und Kalkülkenntnis als Fehlerquelle bei der Durchführung von Basisprozeduren lässt sich somit die durch prozedurales Wissen erklärte Varianz der Klausurleistung auf rund 47 % erhöhen. Dieser Wert liegt deutlich oberhalb der erklärten Varianz des ursprünglichen saturierten Modells.
Dabei ist zu beachten, dass der standardisierte β-Koeffizient von Kalkülkenntnis fast doppelt so groß wie der entsprechende β-Koeffizient von Kalkülfertigkeit ist. Dieser Unterschied bestätigt sich auch in den in Tabelle 5.6 aufgeführten Korrelationen. Kalkülkenntnis weist mit $|r| = .57$ eine etwa doppelt so hohe Produkt-Moment-Korrelation mit der Klausurleistung auf wie Kalkülfertigkeit mit $|r| = .30$. Ein ähnliches Verhältnis ergibt sich auch für die entsprechenden Rangkorrelationen nach Spearman in Höhe von $|\rho| = .55$ zwischen Kalkülkenntnis und Klausurleistung bzw. von $|\rho| = .32$ zwischen Kalkülfertigkeit und Klausurleistung.

Tabelle 5.5
Lineare Regression nach Aufspaltung des prozeduralen Wissens bzw. Unterscheidung der Scheiterungsgründe

Modell	N	Prädiktor	β	p	$\tilde{R}^2$	$\Delta\tilde{R}^2$
Modell 10b	1286	prozedurales Wissen	0.63	.000	0.397	0.397
Modell 11	1286	Schwäche in Kalkülkenntnis	-0.62	.000	0.465	0.068
	1286	Schwäche in Kalkülfertigkeit	-0.37	.000		

Tabelle 5.6

Produkt-Moment-Korrelation (oberes Dreieck) und Rangkorrelation (unteres Dreieck) zwischen prozeduralem Wissen sowie seiner Facetten und der Klausurpunktzahl

	Klausur-leistung	Leistungstest prozedurales Wissen Anzahl korrekter Lösungen	Leistungstest prozedurales Wissen Anzahl mangels Kalkülkenntnis falscher Ergebnisse	Leistungstest prozedurales Wissen Anzahl mangels Kalkülfertigkeit falscher Ergebnisse
Klausurleistung		$.63^{**}$	$-.57^{**}$	$-.30^{**}$
Leistungstest prozedurales Wissen Anzahl korrekter Lösungen	$.64^{**}$		$-.57^{**}$	$-.68^{**}$
Leistungstest prozedurales Wissen Anzahl mangels Kalkülkenntnis falscher Ergebnisse	$-.55^{**}$	$-.64^{**}$		$-.22^{**}$
Leistungstest prozedurales Wissen Anzahl mangels Kalkülfertigkeit falscher Ergebnisse	$-.32^{**}$	$-.65^{**}$	$-.02$	

Anmerkung. $^{*}p < .05$, $^{**}p < .01$; Stichprobengröße: $N = 1286$

Interpretation der Ergebnisse im Hinblick auf Forschungsfrage 2

Der Klausurerfolg („Klausur bestanden“ versus „Klausur nicht bestanden“) konnte bei rund 76 % der teilnehmenden Studierenden korrekt prognostiziert werden. Als mit Abstand bester Einzelprädiktor erwies sich prozedurales Wissen, operationalisiert durch die Anzahl korrekt gelöster Aufgaben im Speed- und Leistungstest prozedurales Wissen, mit einer Prognosegüte von rund 75 % korrekter Vorhersagen.

Die Resultate der nachfolgend genannten Studien ermöglichen eine Einordnung dieser Prognosegüte: Faulkner, Hannigan und Fitzmaurice (2014, S. 648) erreichen durch eine Diskriminanzanalyse auf Grundlage einer schulischen Vorbenotung und eines Eingangstests bei 71,3 % der Studierenden eine korrekte Vorhersage über deren Klausurerfolg in der Ingenieurmathematik. Kolb et al. (2006) sowie Christensen und Meier (2014) gelingt die Frühidentifikation von Studienabbrechern durch eine Kombination geeigneter Inputvariablen und deskriptiv-statistischer Überlegungen bzw. multivariater logistischer Regression mit einer Prognosegüte zwischen 70 % und 85 %.

Auf Grundlage dieser Resultate kann die durch den Speed- und Leistungstest prozedurales Wissen erzielte Prognosegüte in Höhe von 75 % als vergleichsweise hoch eingestuft werden. Bemerkenswert ist dabei, dass es sich bei prozeduralem Wissen um einen Einzelprädiktor handelt, während die Prognosegüte in den zitierten Studien aus einer Kombination mehrerer Erhebungsinstrumente resultiert. Im letzteren Fall werden also mehr Informationen berücksichtigt, was dennoch teilweise nicht ausreicht, um die Prognosegüte von prozeduralem Wissen als Einzelprädiktor zu erreichen.

Die in Tabelle 5.3 angegebenen Prognosegüten bestätigen darüber hinaus die Ergebnisse zu Forschungsfrage 1: Auch im Rahmen der Untersuchungen zu Forschungsfrage 2 erweisen sich die fachspezifischen Faktoren bezüglich der Prognosegüte allen anderen ein-

bezogenen Faktoren als überlegen. Ferner liegt die Prognosegüte affektiver Faktoren mit maximal 55.3 % korrekter Prognosen nur knapp oberhalb der Zufallswahrscheinlichkeit, sodass sich die insgesamt geringe Relevanz dieser Faktoren bestätigt.
Ein Vorteil der eingesetzten Methode (logistische Regression) besteht darin, dass für jeden Studierenden eine Erfolgswahrscheinlichkeit ausgewiesen werden kann, was im Hinblick auf Interventionsmaßnahmen Differenzierungsmöglichkeiten eröffnet. Auf diese Weise könnten beispielsweise homogene Fördergruppen mit einer ähnlichen Bestehenswahrscheinlichkeit gebildet werden. Eine Diskussion, ob und ggf. in welchem Lehr-/Lernarrangement dieses Vorgehen sinnvoll ist, muss allerdings an anderer Stelle geführt werden.

Schließlich erweist sich prozedurales Wissen mit einer erklärten Varianz von rund 40 % auch als bester Einzelprädiktor für die Klausurleistung (in der Klausur erreichte Punktzahl). Alle anderen Faktoren und Hintergrundmerkmale sind aufgrund einer insgesamt zusätzlich erklärten Varianz von 4.4 % kaum relevant. Daraus kann gefolgert werden, dass zur Prognose von Klausurleistung auf Tests wie den Speed- und Leistungstest prozedurales Wissen vermutlich nicht verzichtet werden kann, da beispielsweise die Hintergrundmerkmale kaum Varianz in der Klausurleistung erklären. Am wichtigsten erscheinen unter diesen die in der Schule besuchte Kursform (Leistungskurs bzw. Grundkurs Mathematik) und die Semesterzahl, worauf im nachfolgenden Abschnitt im Rahmen der Diskussion von Forschungsfrage 3 noch einmal näher eingegangen wird.
Die durch prozedurales Wissen erklärte Varianz der Klausurleistung von rund 40 % kann nach einer Aufspaltung prozeduralen Wissens in Kalkülkenntnis und Kalkülfertigkeit nochmals um 7 % gesteigert werden. Die folgenden Ausführungen sollen die Einordnung dieses Wertes ermöglichen: Im linearen Regressionsmodell (vgl. Tabelle 5.4) liegt die Differenz in der erklärten Varianz beim Vergleich des saturierten Modells mit Modell 10a, das nur noch prozedurales Wissen als Prädiktor enthält, bei nur 4.4 %. Aus der umgekehrten Perspektive betrachtet bedeutet dies, dass nach Hinzunahme aller Prädiktoren lediglich eine zusätzliche Varianzaufklärung von 4.4 % gegenüber der durch prozedurales Wissen erklärten Varianz erfolgt. Für einen weiteren Vergleich dient der „Test für medizinische Studiengänge TMS“, der von 1986 bis 1997 bei der Zulassung zu medizinischen Studiengängen deutschlandweit eingesetzt wurde. Dieser Test konnte über die Abiturnote hinaus 7 % zusätzliche Varianz in der Gesamtnote der ärztlichen Vorprüfung erklären (Trost et al., 1997, S. 69). Die in den USA eingesetzten Studierfähigkeitstests SAT I und SAT II erklären gegenüber der Schulabschlussnote zusätzlich zwischen 8 % und 10 % der Varianz in der Studienleistung (Bridgeman, McCamley-Jenkins & Ervin, 2000; Ramist, Lewis & Jenkins, 2001). Auf Grundlage dieser Ergebnisse ist eine durch Aufspalten prozeduralen Wissens in Kalkülkenntnis und Kalkülfertigkeit zusätzlich erklärbare Varianz von rund 7 % als bedeutsam einzustufen. Dieser relativ hohe Wert kann zusammen mit den

Korrelationsanalysen in Tabelle 5.6 die Interpretation rechtfertigen, dass Schwächen in Kalkülkenntnis und Schwächen in Kalkülfertigkeit zwei zumindest teilweise voneinander unabhängige Fehlerquellen sind.

Nach Aufspaltung prozeduralen Wissens in die beiden Facetten Kalkülkenntnis und Kalkülfertigkeit erhöht sich die erklärte Varianz der Klausurleistung auf 47 %. Zur Einordnung dieses Resultats werden die folgenden Untersuchungen angeführt: Denny et al. (2012, S. 179) berichten für die Prognose der Noten in „intermediate algebra“, „precalculus“, und „calculus“ durch eine Kombination von Notendurchschnitt an der Highschool, „placement tests“ und Studierfähigkeitstest Mathematik (SATM) eine erklärte Varianz von 43 %, 36 % bzw. 32 %. Im Vergleich hierzu stellt prozedurales Wissen zumindest im vorliegenden Fall mit einer erklärten Varianz der Klausurleistung von bis zu 47 % auch im internationalen Vergleich einen starken Prädiktor der Klausurleistung dar.
Das Aufspalten prozeduralen Wissens in seine Facetten Kalkülkenntnis und Kalkülfertigkeit liefert noch ein weiteres Resultat: Den Tabellen 5.5 und 5.6 ist zu entnehmen, dass Kalkülkenntnis einen etwa doppelt so hohen Zusammenhang mit der Klausurleistung aufweist wie Kalkülfertigkeit ($\beta = -0.62$ gegenüber $\beta = -0.37$ bzw. $r = -.57$ gegenüber $r = -.30$). Eine mögliche Erklärung für diesen deutlichen Unterschied ergibt sich aus der Tatsache, dass die Kenntnis von Basisprozeduren, durch deren effiziente und fehlerfreie Ausführung prozedurales Wissen im vorliegenden Fall operationalisiert wird (vgl. Abschnitt 4.2.1), eine Grundanforderung im komplexen Anforderungsprofil von Klausuren der Ingenieurmathematik darstellt (vgl. Abschnitt 1.2.2). Dies bedeutet insbesondere, dass ein Beherrschen von Basisprozeduren eine Voraussetzung darstellt, um komplexe mathematische Verfahren, die in Klausuren abgeprüft werden, erfolgreich durchführen zu können. Beispielsweise ist das Durchführen von Superprozeduren ohne die Kenntnis der involvierten Basisprozeduren nicht möglich (vgl. Abbildung 1.5 für einige Beispiele solcher Abhängigkeiten). Eine unzureichende Kenntnis von Basisprozeduren führt dann dazu, dass Superprozeduren und andere komplexe mathematische Verfahren nicht durchgeführt werden können und in der Folge viele Aufgaben nicht lösbar sind. Schwächen in der Kalkülfertigkeit führen dagegen in der Regel lediglich dazu, dass vereinzelte Aufgaben nicht korrekt gelöst werden, beispielsweise aufgrund eines Rechenfehlers. Die Auswirkungen von Schwächen in der Kalkülkenntnis auf die Klausurleistung können deshalb wesentlich größer sein als die Auswirkungen von Schwächen in der Kalkülfertigkeit. Für die Praxis kann dies bedeuten, dass gegebenenfalls eine Förderung von Kalkülkenntnis gegenüber einer Verbesserung von Kalkülfertigkeit zu priorisieren ist. In einer kurzen Formel lässt sich das wie folgt formulieren: „Kalkülkenntnis vor Kalkülfertigkeit“. Eine Diskussion, ob und wie dieses Resultat bei Unterstützungsmaßnahmen im Detail zu berücksichtigen ist, kann die vorliegende Arbeit nicht leisten und muss an anderer Stelle erfolgen.

Fazit

Aufgrund der Ergebnisse kann Forschungsfrage 2 für den vorliegenden Fall wie folgt beantwortet werden: Mithilfe aller in der vorliegenden Arbeit eingesetzten Test- und Erhebungsinstrumente lässt sich in rund 76 % aller Fälle der Klausurerfolg („Klausur bestanden“ versus „Klausur nicht bestanden“) korrekt vorhersagen. Dieser Wert liegt nur knapp über der Prognosegüte des stärksten Einzelprädiktors. Hierbei handelt es sich um prozedurales Wissen, das eine Prognosegüte von rund 75 % erzielt. Im Vergleich mit entsprechenden Werten aus der Literatur ist diese Prognosegüte als hoch zu bewerten. Dies gilt auch für die prognostizierte Klausurleistung: Nach Aufspalten von prozeduralem Wissen in seine beiden Facetten Kalkülkenntnis und Kalkülfertigkeit (vgl. Abschnitt 1.1.1) beträgt die hierdurch erklärte Varianz der Klausurleistung rund 47 %. Auch dieses Ergebnis ist im Vergleich mit internationalen Studien als hoch zu werten. Es bestätigt sich damit die bereits in Abschnitt 1.2 angedeutete und im Rahmen von Forschungsfrage 1 bestätigte hohe Relevanz prozeduralen Wissens für Klausuren der Ingenieurmathematik. Dabei ist zu beachten, dass die hohe Relevanz zunächst nur für die untersuchte Referenzklausur (vgl. Abschnitt 4.2) gilt. Die Verallgemeinerbarkeit dieses Resultats wird im Rahmen der letzten Forschungsfrage diskutiert.

Als weiteres Resultat lässt sich festhalten, dass der Zusammenhang zwischen Kalkülkenntnis und Klausurleistung etwa doppelt so groß wie der Zusammenhang zwischen Kalkülfertigkeit und Klausurleistung ist, was zu der Formel „Kalkülkenntnis vor Kalkülfertigkeit“ führt. Für die Planung von Interventionen ist daher zu überlegen, ob ein Schwerpunkt zunächst auf die Festigung von Kalkülkenntnissen gelegt werden sollte.

5.3 Querschnittsanalyse der untersuchten Studierendenkohorte

Während in den letzten beiden Abschnitten der Zusammenhang zwischen Klausurergebnissen und verschiedenen Faktoren sowie Hintergrundmerkmalen untersucht worden ist, wird im vorliegenden Abschnitt die Verteilung dieser Faktoren und Hintergrundmerkmale in verschiedenen Leistungsgruppen analysiert. Dies geschieht auf Grundlage der Fragestellung von Forschungsfrage 3 durch eine Querschnittsanalyse der untersuchten Studierendenkohorte (vgl. Abschnitt 4.1.2).

Frage und Vorgehensweise

Im Rahmen von Forschungsfrage 3 (*Welche Merkmalsverteilungen weisen Gruppen prozedural schwacher, prozedural durchschnittlicher und prozedural starker Studierender auf?*) wird untersucht, ob es innerhalb verschiedener Studierendengruppen charakteristische Merkmale gibt, die mit Blick auf Interventionsmaßnahmen relevant sein könnten. Dazu werden die Studierenden zunächst bezüglich ihres prozeduralen Wissens in drei Leistungsgruppen eingeteilt: Prozedural schwache Studierende bilden Gruppe 1, prozedural durchschnittliche Studierende Gruppe 2 und prozedural starke Studierende Gruppe 3. Im Folgenden wird erklärt, wie die Studierenden auf Grundlage ihres prozeduralen Wissens in diese drei Gruppen eingeteilt werden und warum gerade prozedurales Wissen als Einstufungskriterium dient.

In den letzten Abschnitten hat sich herausgestellt, dass unter den untersuchten Prädiktorvariablen prozedurales Wissen den Erfolg in der Referenzklausur (vgl. Abschnitt 4.1.4) am besten vorhersagt. Prozedurales Wissen wurde hierbei durch die Fertigkeit operationalisiert, Basisprozeduren der Ingenieurmathematik erfolgreich durchführen zu können (Abschnitt 4.2.1). Es konnte gezeigt werden, dass der auf Grundlage dieser Operationalisierung konstruierte Speed- und Leistungstest prozedurales Wissen ein reliables und valides Testinstrument darstellt, das zudem dem Rasch-Modell genügt (ebenda). Letzteres bringt insbesondere den Vorteil mit sich, dass die Anzahl korrekt gelöster Items eine suffiziente Statistik bildet (Rost, 2004, S. 124). Das bedeutet, dass zur Messung eines Merkmals mit diesem Test nicht mehr entscheidend ist, *welche* Items Teilnehmende lösen, sondern lediglich, *wie viele*. Im konkreten Fall des Speed- und Leistungstests prozedurales Wissen heißt dies, dass bereits die Anzahl der im Test korrekt durchgeführten Basisprozeduren das prozedurale Wissen der Teilnehmenden beschreibt. Das wiederum erlaubt eine Definition der Kategorien *prozedural stark*, *prozedural durchschnittlich* bzw. *prozedural schwach* allein auf Grundlage der Anzahl der im Test korrekt gelösten Items. Legt man nun wie geschildert prozedurales Wissen mit der beschriebenen Operationalisierung dieser Kategorienbildung zugrunde, müssen noch die Schwellenwerte hinsichtlich der Anzahl korrekt gelöster Aufgaben festgelegt werden, die die Kategorien voneinander abgrenzen. Diese Schwellenwerte werden hier nach einem pragmatischen Kriterium festgelegt, nach dem sich in jeder Leistungsgruppe etwa gleich viele Studierende befinden sollten. Das ist sinnvoll, da dann aufgrund der vergleichbaren Stichprobengröße statistische Aussagen über die einzelnen Leistungsgruppen eine ähnliche Qualität aufweisen. In Abhängigkeit von der Anzahl der im Speed- und Leistungstest prozedurales Wissen korrekt gelösten Aufgaben werden also drei etwa gleich große Gruppen wie folgt gebildet: Gruppe 1 besteht aus knapp 37 % der Studierenden mit dem geringsten prozeduralen Wissen. Diese Studierenden haben zum ersten Erhebungszeitpunkt (MZP1, vgl. Abschnitt 4.1) im Test maximal zwei von elf Aufgaben korrekt gelöst. Gruppe 3 bilden rund 24 % der Studierenden, die mindestens sechs Aufgaben korrekt gelöst haben und damit über das größte prozedurale Wissen verfügen. Die restlichen 39 % der Studierenden bilden Gruppe 2. Diese Studierenden haben zwischen drei und fünf Aufgaben korrekt gelöst. Studierende in Gruppe 1 werden im Folgenden als „prozedural schwach“, jene in der Mittelgruppe 2 als „prozedural durchschnittlich“ und Studierende in Gruppe 3 als „prozedural stark“ bezeichnet. Die genauen Zahlen sind in Tabelle 5.7 zusammengestellt.

In den letzten beiden Zeilen in Tabelle 5.7 findet sich noch eine weitere Rechtfertigung für die Wahl des Schwellenwertes zur Unterscheidung der prozedural schwachen von den prozedural durchschnittlichen Studierenden: Aus anekdotischer Evidenz gilt üblicherweise eine Bestehenswahrscheinlichkeit von 50 % als „kritische Marke“. Im vorliegenden Fall

Tabelle 5.7

Verteilung der Testteilnehmer auf die drei Leistungsgruppen

	prozedural schwache Studierende			prozedural durchschnittliche Studierende			prozedural starke Studierende					
Anzahl korrekter Lösungen:	0	1	2	3	4	5	6	7	8	9	10	11
absolute Häufigkeit	146	148	178	199	160	149	106	82	62	31	17	8
relative Häufigkeit (in %)	11.4	11.5	13.8	15.5	12.4	11.6	8.2	6.4	4.8	2.4	1.3	0.9
kummulierte Häufigkeit (in %)	11.4	22.9	36.6	15.5	27.9	39.4	8.2	14.6	19.4	21.8	23.1	24.0
pBQ Klausur (in %)	18.4	28.5	41.4	55.6	69.0	79.8	87.4	92.4	95.5	97.4	98.5	99.1
tBQ Klausur (in %)	8.9	33.8	42.7	58.8	71.3	81.2	84.9	85.4	95.2	96.8	100	100

Anmerkung. pBQ Klausur=durch das logistische Regressionsmodell prognostizierte Bestehensquote in der Klausur; tBQ Klausur=tatsächliche Bestehensquote in der Klausur trennt diese Bestehenswahrscheinlichkeit die prozedural schwachen von den prozedural durchschnittlichen Studierenden, denn prozedural schwache Studierende haben in der Referenzklausur bei dem festgelegten Schwellenwert von zwei korrekt gelösten Aufgaben eine prognostizierte Erfolgswahrscheinlichkeit von maximal 41 %, während die Bestehenswahrscheinlichkeit durchschnittlicher Studierender bei über 57 % liegt. Eine vergleichbare Begründung für die Wahl des oberen Schwellenwertes zur Unterscheidung prozedural durchschnittlicher von prozedural starken Studierenden lässt sich nicht angeben, sodass hierfür nur das oben genannte Kriterium der etwa gleich großen Leistungsgruppen angeführt werden kann.

Im Folgenden wird noch der eingangs erwähnte Aspekt näher erläutert, warum gerade prozedurales Wissen als Kriterium ausgewählt wurde, um Studierende in die genannten Kategorien *prozedural schwach, durchschnittlich* und *stark* einzuordnen. Dazu ist zunächst festzuhalten, dass für Studierende vor allem die Klausurleistung in Mathematik eines der wichtigsten Leistungsmaße darstellt, denn die Leistung in Mathematikklausuren geht in der Regel mit einem verhältnismäßig hohen Anteil in die Bachelor-Abschlussnote ein. Aus diesem Grund liegt es nahe, die Klausurleistung oder die Leistung in einem Test, der hoch mit der Klausurleistung korreliert, der Kategorienbildung zugrunde zu legen. In Abschnitt 5.1 wurde diesbezüglich gezeigt, dass Klausurleistung und prozedurales Wissen deutlich korrelieren. Dadurch besitzt prozedurales Wissen eine hohe Prognosegüte zur Vorhersage von Klausurerfolg und -leistung (vgl. Abschnitt 5.2). Dies wird auch nochmals in Tabelle 5.7 und Abbildung 5.1 deutlich, denn die Bestehensquote in der Klausur wird hier nahezu perfekt durch prozedurales Wissen vorhergesagt. Aus diesen Gründen

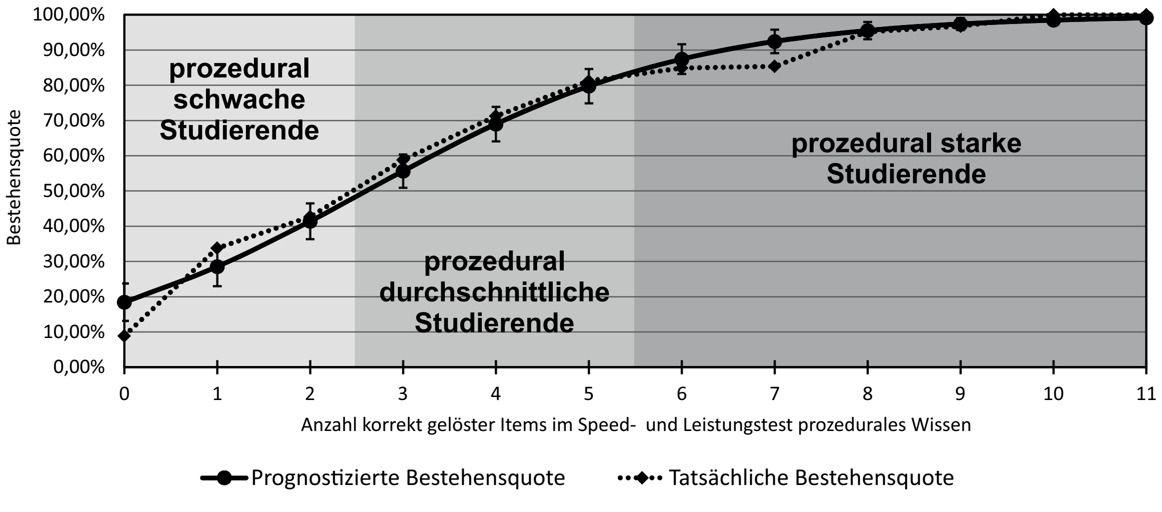


Abbildung 5.1 Prognostizierte und tatsächliche Bestehensquote in der Referenzklausur in Abhängigkeit von der Anzahl korrekt gelöster Items im Speed- und Leistungstest prozedurales Wissen. Die Fehlerbalken beschreiben das 95 %-Konfidenzintervall der durch Bootstraping bestimmten Bestehenswahrscheinlichkeiten (vgl. Abschnitt 4.5).

erscheint der Speed- und Leistungstest prozedurales Wissen zur Kategorienbildung nicht nur einen adäquaten Ersatz für die spätere Klausurleistung zu sein, es lassen sich sogar einige Vorteile hierfür nennen: Dazu zählen eine effiziente Operationalisierbarkeit, eine frühzeitige Messbarkeit sowie die guten psychometrischen Eigenschaften des Messinstruments (vgl. Abschnitt 4.2.1).

Ergebnisse bezüglich der Verteilung der Hintergrundmerkmale in den Leistungsgruppen

Tabelle 5.8 enthält deskriptiv statistische Informationen über die Verteilung der Hintergrundmerkmale in den drei Leistungsgruppen. Es wurden insbesondere diese Merkmale erhoben, da einerseits aus der Literatur Zusammenhänge zur Prüfungs- oder Studienleistung bekannt sind, andererseits sind sie auch im Hinblick auf einen möglichen Förderbedarf von Relevanz. Hierauf wird im nachfolgenden Abschnitt im Rahmen der Interpretation der Ergebnisse näher eingegangen. Zur besseren Übersicht ist die Merkmalsverteilung aus Tabelle 5.8 in Abbildung 5.2 grafisch aufbereitet dargestellt. Tabelle 5.9 hingegen enthält Informationen zur Signifikanz der Unterschiede und zu den Effektstärken. Diese werden

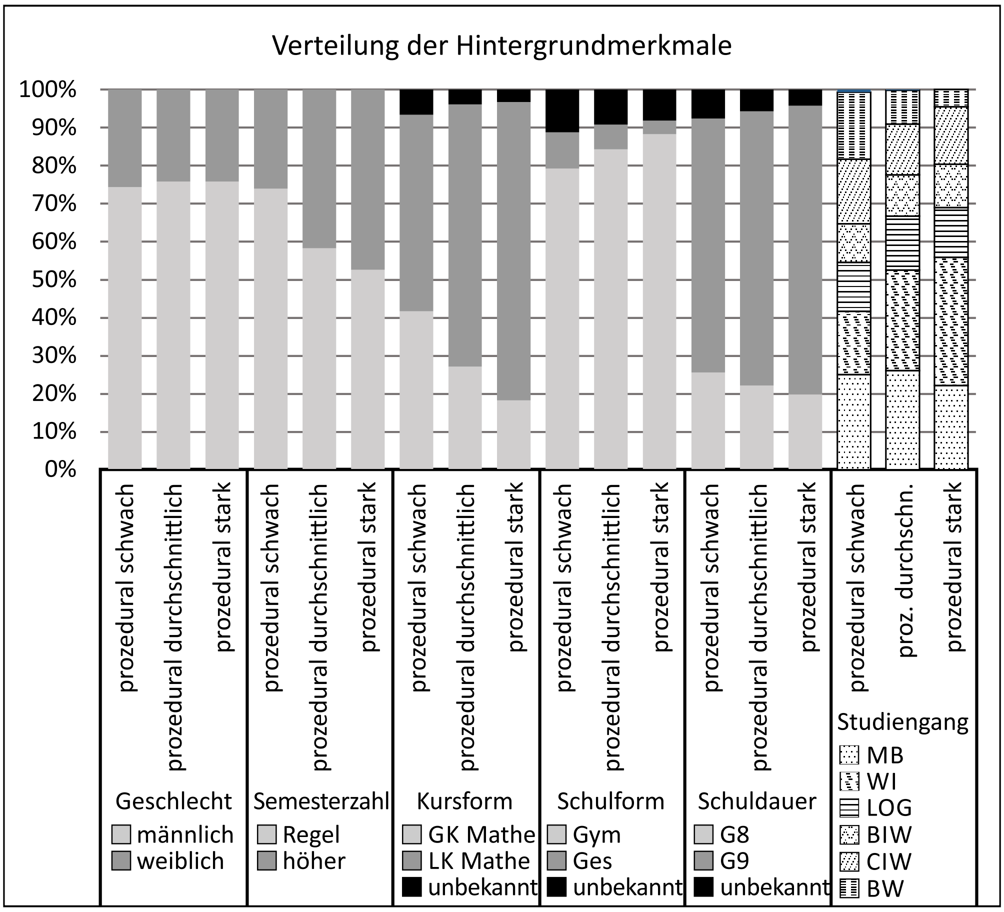


Abbildung 5.2 Verteilung der Hintergrundmerkmale in den Leistungsgruppen der prozedural schwachen, durchschnittlichen und starken Studierenden

Tabelle 5.8

Verteilung der Hintergrundmerkmale in den Leistungsgruppen

	prozedural schwache Studierende		prozedural durchschnittlich Stud.		prozedural starke Studierende	
Merkmal	absolut	%	absolut	%	absolut	%
N	472	100	508	100	306	100
Geschlecht						
männlich	351	74	385	76	232	76
weiblich	121	26	123	24	74	24
Semesterzahl						
Regelstudienzeit	349	74	296	58	161	53
höher	123	26	212	42	145	47
Kursform Mathematik						
Grundkurs	201	43	138	27	56	18
Leistungskurs	249	53	350	69	240	79
keine Angabe	22	4	20	4	10	3
Schulform						
Gymnasium	374	79	428	84	270	88
Gesamtschule	45	10	33	6	11	4
keine Angabe	53	11	47	10	25	8
Schuldauer						
G8	121	26	113	22	61	20
G9	315	67	366	72	232	76
keine Angabe	36	7	29	6	13	4
Studiengang						
Maschinenbau	118	25	133	26	68	22
Wirtschaftsing.	80	17	134	26	103	34
Logistik	66	14	72	14	40	13
Bioingenieurwesen	47	10	55	11	35	11
Chemieingenieurwesen	85	18	68	14	46	15
Bauingenieurwesen	71	15	45	9	14	5
sonstiges	5	1	1	0	0	0

Tabelle 5.9

Effektstärken beim paarweisen Vergleich der Hintergrundmerkmale zwischen den Leistungsgruppen

	prozedural schwach ↔ prozedural durchschnittlich			prozedural durchschnittlich ↔ prozedural stark			prozedural schwach ↔ prozedural stark		
Merkmal	χ^2	p	$\|d\|$	χ^2	p	$\|d\|$	χ^2	p	$\|d\|$
Geschlecht	0.2	.660		0.538	1		0.1	.709	
Semesterzahl	26.0	.000	0.4	4.1	.042	0.1	36.5	.000	**0.5**
Kursform	26.5	.000	0.4	8.2	.004	0.3	51.3	.000	**0.7**
Schulform	3.1	.080		2.7	.100		9.7	.002	**0.6**
Schuldauer	1.9	.172		1.2	.277		4.1	.042	0.2

Anmerkung. $|d|$=Absolutbetrag der Effektstärke; Interpretation gemäß Cohen (1988, S. 355): $|d| \approx 0.2$: schwacher Effekt, $|d| \approx 0.5$: mittlerer Effekt, $|d| \approx 0.8$: starker Effekt. Die Anzahl an Freiheitgraden ist in allen Fällen gleich 1 und daher nicht gesondert ausgewiesen ($df = 1$).

ab dem Niveau $p < .05$ berichtet. Studierende mit *missing data* wurden bei dem betroffenen Merkmal ausgeschlossen. Da es sich mit Ausnahme der Variable Studiengang bei allen Hintergrundvariablen um dichotome Merkmale handelt, werden Unterschiede in den Leistungsgruppen bezüglich dieser Merkmale mithilfe eines χ^2-Tests (Nachtigall & Wirtz, 2009, S. 165) auf Signifikanz untersucht. Die Berechnung der Effektstärke erfolgt in diesem Fall ggf. mithilfe der zugrunde liegenden Kontingenztafel gemäß Hasselblad und Hedges (1995, S. 170). Starke Effekte sind fett gedruckt. Für weitere Details wird auf den Methodenteil verwiesen (vgl. Abschnitt 4.5).

Diskussion der Ergebnisse

Im Folgenden werden die Ergebnisse aus Tabelle 5.9 sowie die Verteilung der in Tabelle 5.8 aufgeführten Hintergrundmerkmale in der dort aufgeführten Reihenfolge diskutiert.

Geschlecht: Das Geschlechterverhältnis ist in allen Leistungsgruppen nahezu identisch. Männliche und weibliche Studierende als Gruppe unterscheiden sich nicht in ihrem Leistungsvermögen. Das Ergebnis entspricht den Beobachtungen des Autors, wonach sich männliche und weibliche Studierende auch in der Klausurleistung in den letzten Jahren nicht signifikant unterscheiden. Die aus den Untersuchungen im Rahmen von PISA bekannte, signifikante Differenz in der Mathematikleistung zwischen männlichen und weiblichen Schülern bzw. Schülerinnen (Prenzel et al., 2007; Prenzel, Sälzer, Klieme & Köller, 2013) setzt sich damit im vorliegenden Fall nicht im tertiären Bereich fort. Ein Grund dafür könnte die Selbstselektion der Studierenden durch die Studiengangwahl sein, denn vermutlich entscheiden sich Schülerinnen und Schüler mit schwachen Mathematikleistungen eher selten dafür, ein ingenieurwissenschaftliches und damit mathematikbetontes Studium aufzunehmen.

Semesterzahl: Der Unterschied bezüglich der Semesterzahl ist dagegen bei jedem Gruppenvergleich mindestens auf dem 5 %-Niveau signifikant: Je leistungsstärker eine Gruppe ist, desto mehr Studierende aus höheren Semestern, also Wiederholer der Veranstaltung, enthält die Gruppe. Die Effektstärke beim paarweisen Vergleich der Leistungsgruppen bezüglich dieses Merkmals ist klein bis mittel ($|d| = 0.1$ bis $|d| = 0.5$). Eine Interpretation dieses Ergebnisses erfolgt weiter unten zusammen mit den nachfolgenden Resultaten zur Kursform.

Kursform: Unter den prozedural starken Studierenden haben in der Schule 78 % einen Leistungskurs in Mathematik besucht, während dieser Anteil unter den prozedural schwachen Studierenden nur 53 % beträgt. Dieser Unterschied stellt einen mittleren bis starken Effekt dar ($|d| = 0.7$). Auch beim Vergleich prozedural schwacher mit prozedural durchschnittlichen sowie prozedural durchschnittlicher mit prozedural starken Studierenden befinden sich in der jeweils prozedural stärkeren Gruppe signifikant mehr Studierende, die zuvor einen Leistungskurs in Mathematik besuchten. Dies drückt sich in schwachen bis mittleren Effektstärken aus ($|d| = 0.4$ bzw. $|d| = 0.3$). Daraus kann gefolgert werden, dass Studierende mit einem Leistungskurs in Mathematik auch bezüglich ihres an der Hochschule erworbenen prozeduralen Wissens in diesem Bereich leistungsstärker sind als Studierende, die in der Schule einen Grundkurs in Mathematik besuchten. Da die Fertigkeit, Basisprozeduren erfolgreich anwenden zu können, eine Grundanforderung in Klausuren der Ingenieurmathematik ist (vgl. Abschnitt 1.2.2), haben auch aus diesem Grund Studierende mit einem Leistungskurs in Mathematik vermutlich eine höhere Erfolgsprognose, die anstehende Klausur zu bestehen, als Studierende, die einen Grundkurs in Mathematik absolvierten. Dieses Resultat wird in ähnlicher Form durch Untersuchungen von Heublein et al. (2010, S. 73) gestützt, denen zufolge von den Studienabbrecherinnen und Studienabbrechern in den ingenieurwissenschaftlichen Studiengängen im Wintersemester 2007/2008 64 % einen Grundkurs in Mathematik besuchten, während es unter den Hochschulabsolventinnen und -absolventen nur 35 % waren. Auch Büning (2004) stellt fest, dass Studienanfängerinnen und -anfänger mit einem Leistungskurs in Mathematik in Eingangstests besser als ihre Mitstudierenden abschneiden, die einen Grundkurs besuchten. Damit lässt sich festhalten, dass insbesondere Studierende mit einem Grundkurs in Mathematik gefährdet erscheinen, zu den leistungsschwachen Studierenden zu gehören. Dieses Resultat wird auch dadurch weiter gestärkt, dass im Vergleich mit den anderen Hintergrundmerkmalen die in der Schule getroffene Wahl der Kursform den stärksten Effekt bezüglich der Leistung im Speed- und Leistungstest prozedurales Wissen aufweist. Zudem sind auch in den Regressionsanalysen Kursform und Semesterzahl bezüglich der Klausurleistung unter den Hintergrundmerkmalen die bedeutendsten Merkmale (vgl. Abschnitt 5.2).

Kursform und Semesterzahl haben gemeinsam, dass beide mit dem zeitlichen Umfang zusammenhängen, wie lange sich Studierende insgesamt mit Mathematik beschäftigt haben. Studierende mit einem Leistungskurs in Mathematik und Wiederholer einer Lehrveranstaltung haben mehr Zeit mit Mathematik verbracht als Studierende, die einen Grundkurs in Mathematik besuchten, sowie Studierende mit einem regulären Studienverlauf ohne Wiederholungen. Die Ergebnisse zeigen, dass es einen signifikanten Zusammenhang zwischen dieser Beschäftigungszeit mit Mathematik und der Leistungsfähigkeit im Test oder sogar Studium geben könnte. Diese Schlussfolgerung wird von zahlreichen Untersuchungen bestätigt (Perdigones, Benedicto, Sánchez-Espinosa, Gallego & García, 2014; Büning, 2004, S. 619). Eine Konsequenz für leistungsschwache Studierende könnte damit sein, dass diese Studierenden während ihres Studiums Lernzeit aufholen müssen. Wie die Zahlen nahelegen, geschieht dies häufig durch eine oder mehrere Wiederholungen einer Veranstaltung. Dies ist vermutlich für betroffene Studierende eine eher frustrierende Art, Lernzeit aufzuholen, zumal Studien zeigen, dass ein „Mehr“ an investierter Lernzeit noch kein Garant für mehr Erfolg im Studium ist, da auch die Art und Weise bzw. die Effizienz des Lernens hierfür eine wesentliche Rolle spielen (Barnett, Sonnert & Sadler, 2014; Kolari, Savander-Ranne & Viskari, 2008; Trautwein, 2007). Dieser Aspekt sollte bei der Implementierung von Unterstützungsmaßnahmen berücksichtigt werden.

Schulform: Beim Vergleich von prozedural schwachen mit prozedural starken Studierenden hinsichtlich der zuvor besuchten Schulform (Gymnasium versus Gesamtschule) liegt ein mittlerer bis starker Effekt vor ($|d| = 0.6$). Demzufolge ist unter den prozedural schwachen Studierenden der Anteil der Absolventinnen und Absolventen einer Gesamtschule signifikant höher als in der Gruppe der leistungsstarken Studierenden. Auch dieses Ergebnis, das vermutlich auf die stärkere Selektivität der Gymnasien zurückzuführen ist, wird durch Resultate mehrerer Studien bestätigt (Trautwein et al., 2006; Baumert, 1996).

Schuldauer: Bezüglich der Schuldauer (G8 versus G9) zeigen sich mit Ausnahme eines vernachlässigbar schwach signifikanten Zusammenhangs und einem schwachen Effekt beim Vergleich der prozedural schwachen mit den prozedural starken Studierenden keine signifikanten Unterschiede. Zu einem vergleichbaren Ergebnis kommen Trautwein, Hübner, Wagner und Kramer (2015, S. 3) in ihrer Untersuchung von rund 5000 Schülerinnen und Schülern an 48 Gymnasien.

Studiengang: Studierende der Studiengänge Maschinenbau (MB), Bio- und Chemieingenieurwesen (BIW, CIW) sowie Logistik (LOG) verteilen sich etwa gleichmäßig auf alle drei Leistungsgruppen. Anders verhält sich das bei der Gruppe der Studierenden des Studiengangs Wirtschaftsingenieurwesen (WI) sowie der Gruppe der Studierenden des Bauingenieurwesens (BW). Erstere sind in der Gruppe der prozedural starken Studierenden deutlich überrepräsentiert, während letztere in der Gruppe der prozedural schwachen Studierenden überproportional häufig vertreten sind. Dieses Ergebnis wird gestützt durch die Resultate von Heublein et al. (2014, S. 4), denen zufolge unter den ingenieurwissenschaftlichen Studiengängen das Bauingenieurwesen mit 51 % die höchste Abbruchquote aufweist. Dabei gibt ein großer Teil der Abbrecherinnen und Abbrecher Leistungsprobleme als Grund für den Studienabbruch an. Ein Grund dafür, dass Studierende des Studiengangs Wirtschaftsingenieurwesen überproportional häufig unter den prozedural starken Studierenden zu finden sind, könnte der höhere NC für die Zulassung zu diesem Studium sein. Dies steht allerdings im Widerspruch dazu, dass sich Studierende der anderen Studiengänge mit einem ähnlichen NC eher gleichmäßig auf die Leistungsgruppen verteilen. Die Erforschung der Gründe, warum Studierende des Bauingenieurwesens häufiger in der Gruppe der prozedural schwachen Studierenden vertreten sind bzw. Studierende des Wirtschaftsingenieurwesens öfter in der Gruppe der prozedural starken Studierenden zu finden sind, bleibt damit nachfolgenden Forschungsarbeiten vorbehalten.

Ergebnisse bezüglich der Verteilung der fachspezifischen, allgemein kognitiven und affektiven Faktoren sowie der Klausurleistung in den Leistungsgruppen

Tabelle 5.10 gibt Aufschluss über die Verteilung der fachspezifischen, allgemein kogitiven und affektiven Faktoren sowie der Klausurleistung in den Leistungsgruppen. Im Folgenden werden kurz einige der angegebenen Faktoren näher erläutert: Zum Bestehen der Referenzklausur (vgl. Abschnitt 4.1.4) sind 32 Punkte erforderlich (40 % der erreichbaren Punktzahl). Die Klausurleistung wird hier durch die in der Klausur erreichten Punkte operationalisiert. Aus den Bearbeitungen des Speed- und Leistungstests prozedurales Wissen, der Basisprozeduren der Ingenieurmathematik abprüft (vgl. Abschnitt 4.2.1), lassen sich drei Faktoren extrahieren: Die Anzahl korrekter Lösungen stellt die Operationalisierung des prozeduralen Wissens der Teilnehmenden dar (ebenda). Ferner wurden alle fehlerhaften Bearbeitungen hinsichtlich der Gründe untersucht, warum das

Tabelle 5.10

Verteilung der fachspezifischen, allgemein kognitiven und affektiven Merkmale sowie der Klausurleistung in den Leistungsgruppen

		prozedural schwache Studierende			prozedural durchschnittliche Studierende			prozedural starke Studierende		
Faktor	erreichbare Punktzahl	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
Klausurleistung	80	472	26.2	10.6	508	37.8	10.9	306	45.6	10.7
Speed- und Leistungstest prozedurales Wissen		472			508			306		
korrekte Lösungen	11		1.1	0.8		3.9	0.8		7.3	1.3
Kalkülkenntnis als Scheiterungsgrund	11		2.4	2.2		0.8	0.9		0.3	0.5
Kalkülfertigkeit als Scheiterungsgrund	11		7.5	2.0		6.3	1.0		3.4	1.3
Leistungstest konzeptuelles Wissen	5	432	1.0	1.1	465	1.9	1.2	292	2.4	1.4
Wiener Matrizen-Test 2	18	284	14.5	2.6	343	15.5	2.7	210	16.8	3.1
Konzentrationstest		377			419			274		
Arbeitsleistung	196		117.2	16.8		122.4	16.8		129.6	16.2
Arbeitsgeschwindigkeit	196		126.8	17.2		131.0	17.1		136.7	17.0
Sorgfalt	1.00		0.92	0.05		0.93	0.04		0.95	0.03
Prüfungsangst	120	279	67.2	13.4	293	64.6	12.0	187	64.4	14.0
Prokrastination	40	279	23.9	5.0	293	23.1	4.9	187	23.6	5.0

Anmerkung. M=Mittelwert, SD=Standardabweichung

jeweilige Item nicht korrekt gelöst wurde. Hierbei wurde zwischen Kalkülkenntnis und Kalkülfertigeit als Fehlerursache unterschieden (ebenda). Da im Test maximal elf Punkte erreicht werden können, ist es entsprechend möglich, maximal in elf Aufgaben aus den genannten Gründen zu scheitern. Weitere Details zu allen aufgeführten Faktoren werden im Theoriteil (Kapitel 1 und 2) sowie im Methodenteil (Kapitel 4) beschrieben.
Während Tabelle 5.10 einen deskriptiv statistischen Überblick über die Daten darstellt, finden sich in Tabelle 5.11 die inferenzstatistischen Auswertungen dieser Daten. Hier werden insbesondere die zugehörigen Effektstärken berichtet, sofern die Mittelwerte mindestens auf dem 5 %-Niveau signifikant voneinander abweichen. Starke Effekte sind fett gedruckt.

Tabelle 5.11
Effektstärken beim paarweisen Vergleich der Faktoren in den Leistungsgruppen

	prozedural schwach ↔ prozedural durchschnittlich				prozedural durchschnittlich ↔ prozedural stark				prozedural schwach ↔ prozedural stark			
Faktor	$\|t\|$	*df*	*p*	$\|d\|$	$\|t\|$	*df*	*p*	$\|d\|$	$\|t\|$	*df*	*p*	$\|d\|$
Klausurleistung	17.0	976	.00	**1.1**	10.0	650	.00	0.7	24.9	646	.00	**1.8**
Speed- und Leistungstest prozedurales Wissen												
korrekte Lösungen	53.7	972	.00	**3.4**	40.6	447	.00	**3.3**	73.6	458	.00	**5.9**
Kalkülkenntnis als Scheiterungsgrund	15.3	607	.00	**1.0**	10.6	811	.00	0.7	20.6	547	.00	**1.2**
Kalkülfertigkeit als Scheiterungsgrund	11.5	705	.00	0.7	33.2	540	.00	**2.5**	34.8	776	.00	**2.4**
Leistungstest konzeptuelles Wissen	10.5	894	.00	0.7	5.8	558	.00	0.4	14.3	525	.00	**1.1**
Wiener Matrizen-Test 2	3.9	585	.00	0.3	3.7	511	.00	0.3	7.3	492	.00	0.6
Konzentrationstest												
Arbeitsleistung	4.4	419	.00	0.3	5.7	599	.00	0.4	9.6	599	.00	0.7
Arbeitsgeschwindigkeit	3.4	784	.00	0.2	4.4	587	.00	0.3	7.3	593	.00	0.6
Sorgfalt	3.1	756	.00	0.2	4.8	676	.00	0.4	7.5	645	.00	0.6
Prüfungsangst	2.4	556	.02	0.2	0.18	352	.85		2.2	387	.03	0.2
Prokrastination	2.1	566	.04	0.2	1.08	388	.28		0.8	401	.45	

*A*nmerkung. $|d|$=Absolutbetrag der Effektstärke (Cohens d); Interpretation gemäß Cohen (1988, S. 355): $|d| \approx 0.2$: schwacher Effekt, $|d| \approx 0.5$: mittlerer Effekt, $|d| \approx 0.8$: starker Effekt; $|t|$ Absolutbetrag des t-Wertes (zweiseitiger t- bzw. Welch-Test auf Gleichheit der Mittelwerte); *df*=Anzahl Freiheitsgrade

Diskussion der Ergebnisse

Im Folgenden wird die Verteilung der genannten Faktoren innerhalb der drei Leistungsgruppen gemäß der in Tabelle 5.11 aufgeführten Reihenfolge diskutiert.

Klausurleistung: Die bezüglich ihres prozeduralen Wissens als schwach eingestuften Studierenden erreichen in der späteren Klausur im Durchschnitt nicht die erforderliche Punktzahl, um die Klausur zu bestehen: Ihre Klausurleistung liegt durchschnittlich bei 26.2 Punkten, während zum Bestehen 32 Punkte erforderlich sind. Die Klausurleistung der prozedural durchschnittlichen und prozedural starken Studierenden liegt hingegen mit im Schnitt 37.8 bzw. 45.6 Punkten deutlich oberhalb der Bestehensgrenze. Die Unterschiede in der Klausurleistung zwischen prozedural schwachen und durchschnittlichen sowie zwischen prozedural schwachen und starken Studierenden sind jeweils als stark zu bewerten ($|d| = 1.1$ bzw. $|d| = 1.8$). Eine Konsequenz dieser Resultate könnte eine gezielte Förderung leistungsschwacher Studierender in klausurrelevanten Themen sein.

Prozedurales Wissen, Anzahl korrekt gelöster Items: In diesem Bereich zeigen sich die größten Effektstärken bei der Bewertung der Unterschiede zwischen den Leistungsgruppen. Dies liegt allerdings auch daran, dass die Kategorienbildung auf Grundlage prozeduralen Wissens bzw. der Anzahl der im Speed- und Leistungstest prozedurales Wissen korrekt gelöster Items erfolgte.

Prozedurales Wissen, Kalkülkenntnis als Scheiterungsgrund: In allen paarweisen Vergleichen liegen signifikante Unterschiede zwischen den Leistungsgruppen vor. Diese sind zwischen prozedural schwachen und durchschnittlichen sowie zwischen prozedural schwachen und prozedural starken Studierenden als stark zu bewerten ($|d| = 1.0$ bzw. $|d| = 1.2$). Wie bei der Klausurleistung sind die Unterschiede zwischen prozedural durchschnittlichen und starken Studierenden weniger stark ausgeprägt. Folglich könnte der zielgruppenspezifische Fokus einer möglichen Förderung von Kalkülkenntnis erneut auf die prozedural schwachen Studierenden gerichtet werden.

Prozedurales Wissen, Kalkülfertigkeit als Scheiterungsgrund: Bezüglich ihrer Kalkülfertigkeit sind die Unterschiede zwischen prozedural schwachen und durchschnittlichen Studierenden weniger ausgeprägt als zwischen prozedural schwachen und starken sowie zwischen prozedural durchschnittlichen und starken Studierenden. Prozedural schwache Studierende scheitern durchschnittlich in 7.5 von elf Items wegen Schwächen in der Kalkülfertigkeit, durchschnittliche Studierende in 6.3 Aufgaben und prozedural starke Studierende in nur 3.4 Aufgaben. Eine Förderung von Kalkülfertigkeit könnte sich daher sowohl an prozedural schwache als auch an durchschnittliche Studierende richten.

Konzeptuelles Wissen: Bezüglich Unterschieden im konzeptuellen Wissen sind lediglich die Unterschiede zwischen prozedural schwachen und prozedural starken Studierenden als stark zu bewerten ($|d| = 1.1$). Ähnlich wie bei der Klausurleistung und der Kalkülkenntnis sollte sich eine Unterstützung in diesem Bereich primär auf prozedural schwache Studierende konzentrieren.

Wiener Matrizen-Test 2 (kognitive Grundfertigkeit) und Konzentration (allgemein kognitive Faktoren): Insgesamt zeigen sich beim paarweisen Vergleich der Leistungsgruppen zwar signifikante Unterschiede, die Effektstärken sind allerdings klein bis mittel ($|d| = 0.2$ bis $|d| = 0.7$). Analog zu den Resultaten zu den ersten beiden Forschungsfragen scheinen damit auch bezogen auf das prozedurale Wissen von Studierenden allgemein kognitive Faktoren eine geringere Bedeutung zu spielen als die fachspezifischen Faktoren.

Prüfungsangst und Prokrastination (affektive Faktoren): Prozedural schwache Studierende weisen gegenüber durchschnittlichen und prozedural starken Studierenden eine auf dem 5 %-Niveau signifikant höhere Prüfungsangst auf. Der Effekt ist allerdings in beiden Fällen schwach ($|d| = 0.2$). Damit spielen in der Gesamtbetrachtung affektive Faktoren im Gruppenvergleich keine Rolle. Im Einzelfall kann dies allerdings anders sein (Schaefer et al., 2009). Aus diesem Grund sollten im Hinblick auf ein ganzheitliches Modell zur Beschreibung des „student life cycle“ affektive Faktoren dennoch Beachtung finden. Neuere Untersuchungen rücken beispielsweise das Vertrauen in die eigene Leistung als möglichen „best non-cognitive predictor of academic success“ (Stankov & Lee, 2014, S. 1) in den Fokus der Forschung (Stankov, Morony & Lee, 2014; Parsons, Croft & Harrison, 2009). Auch im Bereich der Heterogenitäts- und Dropout-Diskussion an den Hochschulen könnten vermutlich noch weitere affektive Faktoren relevant sein, wie beispielsweise motivationale Aspekte, die zu den primären Ursachen eines Studienabbruches zählen (Heublein et al., 2010, S. 28).

Zusammenfassend ist festzuhalten, dass die Unterschiede zwischen den Leistungsgruppen bezüglich fachspezifischer Faktoren häufig groß, bezüglich allgemein kognitiver Faktoren schwach bis mittel und bezüglich affektiver Faktoren nicht relevant sind.
Es fällt auf, dass prozedural schwache Studierende gegenüber prozedural starken bezüglich aller Leistungsindikatoren mindestens mittlere, meistens aber große Unterschiede aufweisen. Deutliche Leistungsunterschiede zwischen prozedural schwachen und starken Studierenden zeigen sich also nicht nur in isolierten Bereichen, sondern im ganzen Spektrum der untersuchten leistungsbezogenen Faktoren. Dieser Befund wird von aktuellen Untersuchungen gestützt, die weitergehend aufzeigen, dass die Schwächen sich noch weit über die in der vorliegenden Arbeit untersuchten Bereiche hinaus erstrecken (Yip, 2013). Eine Konsequenz für die Lehre könnte daher eine Verlängerung der Orientierungsphase

(Empfehlung des Wissenschaftsrat, 2004, S. 60) oder die Einführung eines Vorstudienjahres für leistungsschwache Studienanfängerinnen und Studienanfänger sein, wie es aus ähnlichen Gründen bereits an mehreren Stellen gefordert wird (Gensch & Kliegl, 2011, S. 121) und wofür international und inzwischen auch national Beispiele existieren (Parnell & Statham, 2007; Haase, 2014). Durch diese intensiven Maßnahmen ist es vermutlich möglich, ein breiteres Spektrum von Schwächen aufzuarbeiten und vorhandene Ressourcen zu stärken. In den USA wird dies beispielsweise durch „pre-calculus"-Kurse realisiert, die im Umfang von etwa 10 Wochen leistungsschwache Studierende in die Lage versetzen sollen, anschließend die regulären Veranstaltungen der Hochschulmathematik erfolgreich zu absolvieren. Allerdings kommen Sonnert und Sadler (2014) diesbezüglich in einer repräsentativen Untersuchung von 5507 Studierenden an 132 Institutionen zu folgendem Ergebnis: „Thus, our investigation suggests a simple answer to the principal question posed in this article (Does taking a college pre-calculus course improve students' subsequent performance in college calculus?): No" (Sonnert & Sadler, 2014, S. 1203). Zu einem ähnlichen Resultat kommt Horstkötter (2006) bei seiner Untersuchung, ob Studierende mit geringerer Vorbildung Schwächen durch spezielle Unterstützungsmaßnahmen während der Studieneingangsphase ausgleichen können: „Dies lässt den Schluss zu, dass eine Substitution von mangelnder Inputqualität durch die universitären Mehrleistungen im vorliegenden Fall der Universität Paderborn nicht ausgeglichen werden konnten" (S. 206). Ein möglicher Grund für den geringen Erfolg derartiger Unterstützungsmaßnahmen könnte eventuell in einer zu niedrigen Passgenauigkeit dieser Angebote liegen, die durch eine stärkere fachdidaktische Fundierung bzw. eine höhere Adaptivität hinsichtlich der dignostizierten Schwächen der Studierenden verbessert werden könnte. Um diesbezüglich aber abgesicherte Aussagen treffen zu können, ist eine genauere Analyse der genannten Angebote insbesondere aus fachdidaktischer Perspektive erforderlich, was nachfolgenden Forschungsarbeiten vorbehalten bleiben muss. Eine weitere Ursache für den zum Teil eingeschränkten Erfolg von Unterstützungsangeboten in der Vor- oder Studieneingangsphase könnte in deren zeitlichem Umfang liegen. In der vorangegangenen Diskussion bezüglich des Zusammenhangs zwischen Hintergrundmerkmalen und prozeduraler Leistung hatte sich gezeigt, dass prozedural schwache Studierende häufig weniger Lernzeit in Mathematik investierten, beispielsweise weil sie in der Schule einen Grundkurs in Mathematik besuchten. Möglicherweise kann diese Lernzeit durch die oben genannten Maßnahmen nicht in dem erforderlichen Maße aufgeholt werden, weil die Angebote zu kurz oder noch nicht effektiv genug sind.

Vor dem Hintergrund dieser Resultate scheint es ratsam, vor der Einführung unterstützender Maßnahmen diese grundsätzlich auf ein solides fachdidaktisches Fundament zu stellen. Hierfür finden sich in der Literatur zahlreiche Ausgangspunkte (Hieb, Lyle, Ralston & Chariker, 2015; Parnell & Statham, 2007). Dabei sollte bei der Bewertung von Maßnah-

men, die als Vorbild dienen könnten, berücksichtigt werden, dass Erfolgsmeldungen über eine Verbesserung von Studierenden nach der Teilnahme an einem Unterstützungsangebot häufig auf der Grundlage eines Vortests und identischen Nachtests zur Erfolgskontrolle erfolgen. Das kann darüber hinwegtäuschen, dass lediglich ein „teaching to the test" stattgefunden hat, das die Studierenden nicht zwingend befähigt, in Anschlussveranstaltungen mit neuen Kontexten erfolgreich zu sein.
Darüber hinaus sollte bei der Durchführung von Unterstützungsmaßnahmen gewissenhaft eruiert werden, welche Studierenden diese Maßnahmen in welchem Umfang benötigen. Bebermeier und Nußbeck (2014) konstatieren diesbezüglich mit Verweis auf weiterführende Studien: „Allerdings hat sich gezeigt, dass für den Erfolg von Unterstützungsmaßnahmen vor allem die Passung zwischen den Angeboten und den Merkmalen der Studierenden zentral ist ..." (S. 86). Untersuchungen belegen beispielsweise, dass es auch innerhalb der Gruppe der leistungsschwachen Studierenden Unterschiede gibt: Während einige nach einem dreiwöchigen Auffrischungskurs auf dem gewünschten Stand sind, benötigen andere dafür ein ganzes Semester (Rodgers, Posler & Trible, 2011).

Ergebnisse bezüglich der Verteilung prozeduralen Wissens bei Studierenden sowie seiner beiden Facetten Kalkülfertigkeit und Kalkülkenntnis

Da prozedurales Wissen einen Schwerpunkt der vorliegenden Arbeit bildet, wird dieses hier nochmals eingehender und gesondert untersucht.
In Abbildung 5.3 ist dazu oben links die Verteilung der Scheiterungsgründe innerhalb der Leistungsgruppen beim Lösen der Items des Speed- und Leistungstests prozedurales Wissen dargestellt. Auf der horizontalen Achse ist die Anzahl an Items abgetragen, bei denen Teilnehmerinnen und Teilnehmer mangels Kalkülfertigkeit scheiterten. Die vertikale Achse gibt an, bei wie vielen der elf Items Teilnehmende mangels Kalkülkenntnis scheiterten. Die horizontalen und vertikalen Linien im Schaubild oben links markieren die gruppenspezifischen Mittelwerte.
Die übrigen drei Schaubilder in Abbildung 5.3 stellen einen Perspektivenwechsel dar. Hier wurden ebenfalls drei Leistungsgruppen gebildet, die Klassifizierung erfolgte allerdings bezüglich der Klausurpunktzahl. Die Verteilung der Studierenden entspricht dem Verteilungsverhältnis der bisher betrachteten, d. h. bezüglich der Anzahl im Speed- und Leistungstest prozedurales Wissen korrekt gelöster Aufgaben gebildeten, Leistungsgruppen. Das heißt, die leistungsschwachen Studierenden bilden in den übrigen drei Schaubildern die rund 37 % der Studierenden mit der geringsten Klausurpunktzahl, die Gruppe der leistungsstarken Studierenden umfasst die 24 % Studierenden mit der höchsten Klausurpunktzahl und den Rest bilden die bezüglich ihrer Klausurpunktzahl durchschnittlichen Studierenden.

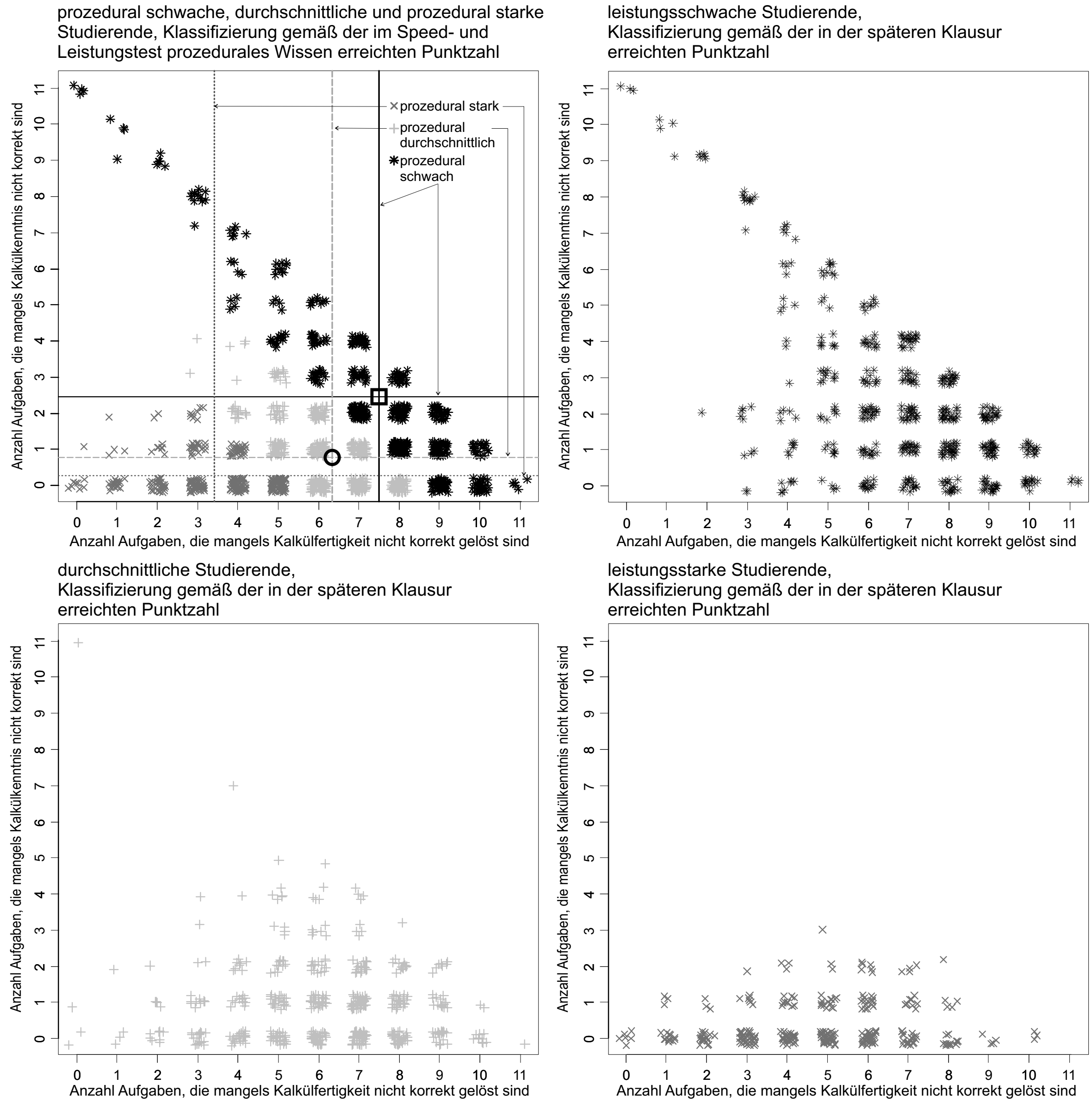


Abbildung 5.3 Oben links: Verteilung der prozedural schwachen, durchschnittlichen und starken Studierenden bzgl. der Anzahl wegen einer Schwäche in Kalkülkenntnis bzw. Kalkülfertigkeit im Speed- und Leistungstest prozedurales Wissen nicht oder falsch gelöster Aufgaben. Oben rechts und unten: Analoge Verteilung für bezüglich der Klausurleistung schwache, durchschnittliche bzw. starke Studierende

Diskussion der Ergebnisse

Zunächst wird das Schaubild oben links in Abbildung 5.3 diskutiert. Hier erfolgte die Kategorienbildung (*prozedural schwach, prozedural durchschnittlich, prozedural stark*) auf Grundlage des prozeduralen Wissens der Studierenden, d. h. in Abhängigkeit von der Anzahl der im Speed- und Leistungstest prozedurales Wissen korrekt gelösten Aufgaben. Es fällt auf, dass sich die drei Leistungsgruppen hinsichtlich der Anzahl an Items, die mangels Kalkülkenntnis bzw. Kalkülfertigkeit nicht oder falsch gelöst wurden, deutlich voneinander abgrenzen. Darauf deuteten auch bereits die Effektstärken in Tabelle 5.11 hin. Pro-

zedural schwache Studierende unterscheiden sich sowohl bezüglich ihrer Kalkülkenntnis als auch ihrer Kalkülfertigkeit von Studierenden der anderen beiden Leistungsgruppen, wobei der Unterschied zu den prozedural durchschnittlich Studierenden in der Kalkülkenntnis größer ist als in der Kalkülfertigkeit. Prozedural durchschnittliche und prozedural starke Studierende unterscheiden sich dagegen bezüglich ihrer Kalkülkenntnis weniger stark, was sich auch in einer geringeren, aber immer noch vergleichsweise hohen Effektstärke ausdrückt ($|d| = 0.7$, vgl. Tabelle 5.11). Bezüglich Kalkülfertigkeit sind prozedural starke Studierende sowohl den durchschnittlichen als auch prozedural schwachen Studierenden deutlich überlegen, während sich durchschnittliche und prozedural schwache Studierende diesbezüglich weniger unterscheiden. Auch dies wird durch eine entsprechend geringere Effektstärke von $|d| = 0.7$ in Tabelle 5.11 bestätigt.
Eine Konsequenz für die Lehre könnten daher Angebote insbesondere für prozedural schwache Studierende zum Auf- und Ausbau beider Facetten prozeduralen Wissens sein. Die Grundlage hierfür könnten die Items des Speed- und Leistungstests prozedurales Wissen bilden. Diese prüfen Basisprozeduren der Ingenieurmathematik ab (vgl. Abschnitt 4.2.1). Ein Abprüfen von Basisprozeduren erscheint auch mit Hinblick auf den Klausurerfolg sinnvoll, denn einerseits stellen Basisprozeduren eine Grundanforderung in Klausuren der Ingenieurmathematik dar (vgl. Abschnitt 1.2.2), andererseits konnte in den Untersuchungen zu den ersten beiden Forschungsfragen ein deutlicher Zusammenhang zwischen prozeduralem Wissen und seinen beiden Facetten, Kalkülkenntnis und -fertigkeit, sowohl mit dem Klausurerfolg als auch mit der Klausurleistung gezeigt werden.
Abbildung 5.3 liefert auch erste Anhaltspunkte für die Definition von Mindest- und Regelstandards, die regelmäßig überprüft werden könnten (Wilson & MacGillivray, 2007; Fuller, Deshler, Kuhn & Squire, 2014, S. 50): So erscheint es plausibel, dass alle Studierenden mindestens die durschnittliche Kalkülkenntnis und -fertigkeit prozedural schwacher Studierenden erreichen sollten. In Abbildung 5.3 entspricht dies dem Schnittpunkt der horizontalen und vertikalen Linien, welche die jeweiligen Mittelwerte kennzeichnen. Der Schnittpunkt ist in diesem Fall mit einem Quadrat markiert. Entsprechend könnten die Mittelwerte der prozedural durchschnittlichen Studierenden als Regelstandard festgelegt werden, der in Abbildung 5.3 mit einem Kreis gekennzeichnet ist. Die vorangegangenen Überlegungen können nur Ansatzpunkte für zielgruppenspezifische und adaptive Unterstützungsangebote zum Auf- und Ausbau prozeduralen Wissens sowie zum Abbau von Schwächen in der Kalkülfertigkeit und -kenntnis sein. Eine detaillierte Ausarbeitung solcher Maßnahmen, insbesondere auf einer fachdidaktischen Grundlage, bleibt damit nachfolgenden Forschungsvorhaben vorbehalten.

Während in dem zuvor diskutierten Schaubild oben links in Abbildung 5.3 die Einteilung der Studierenden in *prozedural schwach, durchschnittlich* und *stark* auf Grundlage

der Ergebnisse im Speed- und Leistungstest prozedurales Wissen vorgenommen wurde, erfolgte in den übrigen drei Schaubildern die Einteilung derselben Studierenden in die Kategorien leistungsschwach, durchschnittlich und leistungsstark in Abhängigkeit von den später erzielten Klausurergebnissen. Diese Einteilung wurde so vorgenommen, dass korrespondierende Leistungsgruppen gleich groß sind. Das heißt, die Gruppe der prozedural schwachen Studierenden umfasst genauso viele Studierende wie die Gruppe der (bezüglich der Klausurleistung) leistungsschwachen Studierenden usw.
Es folgt nun die Interpretation der übrigen drei Schaubilder oben rechts und unten in Abbildung 5.3: Zunächst fällt auf, dass bezüglich der Klausurleistung als leistungsschwach eingestufte Studierende bei der Bearbeitung der Testitems wesentlich häufiger sowohl mangels Kalkülkenntnis als auch mangels Kalkülfertigkeit scheitern als Studierende in den anderen Leistungsgruppen. Die bezüglich der Klausurleistung durchschnittlichen und leistungsstarken Studierenden scheitern dagegen deutlich seltener mangels Kalkülkenntnis. Während bezüglich ihrer Klausurleistung leistungsschwache Studierende in bis zu elf Aufgaben (von elf Aufgaben im Speed- und Leistungstest prozedurales Wissen) mangels Kalkülkenntnis scheitern, ist dies bei den bezüglich der Klausurleistung durchschnittlichen und leistungsstarken Studierenden, von wenigen Ausreißern abgesehen, nur bei maximal vier bzw. zwei Aufgaben der Fall. Im Gegensatz zur Kalkülkenntnis lassen sich die drei Leistungsgruppen durch mangelnde Kalkülfertigkeit als Scheiterungsgrund nicht so klar gegeneinander abgrenzen. Dies unterstützt den Befund von Forschungsfrage 2, dem zufolge Kalkülkenntnis einen höheren Anteil der Varianz der Klausurleistung erklärt als Kalkülfertigkeit. Folglich ist eine Förderung von Kalkülkenntnis gegebenenfalls zu priorisieren („Kalkülkenntnis vor Kalkülfertigkeit“).

Fazit

Die Einteilung der Studierenden in die drei Leistungsgruppen *prozedural schwach, durchschnittlich* und *stark* in Abhängigkeit von ihren Ergebnissen im Speed- und Leistungstest prozedurales Wissen stellt ein geeignetes Mittel dar, um differenzierte Unterschiede in den Hintergrund- und Leistungsmerkmalen aufzudecken. Die Resultate hierzu werden weiter unten nochmals zusammengefasst. Die beschriebene Einteilung in die Leistungsgruppen gemäß der Testergebnisse hat gegenüber der Einteilung nach späteren Klausurergebnissen einige Vorteile. Dazu zählen insbesondere die frühzeitige Verfügbarkeit der Ergebnisse, die effiziente Operationalisierbarkeit prozeduralen Wissens sowie die Rasch-Skalierbarkeit des Erhebungsinstruments.

Bezüglich der Hintergrundmerkmale sind unter den leistungsstarken Studierenden überproportional viele Studierende mit einem Leistungskurs Mathematik, überproportional viele Wiederholer der Veranstaltung sowie überproportional viele Studierende des Studiengangs Wirtschaftsingenieurwesen vertreten. Dabei repräsentieren die ersten beiden genannten Hintergrundmerkmale denselben Aspekt: Sie stehen für den zeitlichen Umfang, der über eine längere Zeit in die Beschäftigung mit Mathematik investiert wurde. Aus diesen Ergebnissen könnte gefolgert werden, dass Studierende umso leistungsstärker sind, desto mehr Zeit sie für Mathematik aufgewendet haben oder aufwenden mussten, beispielsweise aufgrund einer nicht bestandenen Klausur durch wiederholtes Lernen desselben Vorlesungsstoffs.

In der Gruppe der prozedural schwachen Studierenden ist dagegen der Anteil der Studierenden, die in der Schule einen Grundkurs in Mathematik besucht haben, sowie der Absolventen einer Gesamtschule signifikant höher als in der Gruppe der prozedural starken Studierenden. Dies gilt unter den untersuchten Studiengängen auch für Studierende des Studiengangs Bauingenieurwesen. Aus diesem Grund sollte diesen drei Gruppen im Hinblick auf einen späteren Klausurerfolg eine besondere Beachtung zukommen, da prozedural schwache Studierende in der Referenzklausur (vgl. Abschnitt 4.1.4) im Durchschnitt lediglich rund 26 Punkte erreichen, was zum Bestehen dieser Klausur nicht ausreicht, d. h. prozedural schwache Studierende bestehen im Schnitt die spätere Klausur nicht.

Aus den Ergebnissen kann daher ein Unterstützungsbedarf für prozedural schwache Studierende in den Bereichen prozedurales Wissen inklusive seiner beiden Facetten, Kalkülkenntnis und Kalkülfertigkeit, sowie in den Bereichen konzeptuelles Wissen und klausurspezifische Themen abgeleitet werden. Bezüglich der anderen beiden Leistungsgruppen empfiehlt sich dagegen lediglich eine Förderung von Kalkülfertigkeit bei den durchschnittlichen Studierenden.

5.4 Längsschnittanalyse zur Entwicklung prozeduralen Wissens bei Studierenden

Im Rahmen der ersten beiden Forschungsfragen stellte sich prozedurales Wissen unter den untersuchten Konstrukten und Hintergrundmerkmalen als bester Prädiktor für Erfolg und Leistung in der Referenzklausur heraus. Das Ergebnis passt zu der Meinung vieler Autorinnen und Autoren, denen zufolge Ingenieurmathematik eher prozedural gelehrt, gelernt und geprüft wird (vgl. Abschnitt 1.2). Aus diesen Gründen konzentrierten sich die Untersuchungen zu Forschungsfrage 3 auf die Analyse prozeduralen Wissens bei Studierenden, was nun in Forschungsfrage 4 fortgesetzt wird. Während es sich bei den Untersuchungen zu Forschungsfrage 3 um Querschnittsanalysen handelte, geht es in Forschungsfrage 4 um die längsschnittliche Entwicklung prozeduralen Wissens bei Studierenden. Dementsprechend lautet Forschungsfrage 4: *Wie entwickelt sich prozedurales Wissen bei ausgewählten Studierendengruppen während der ersten Studiensemester?* Im Folgenden wird diese Frage im Detail untersucht.

Vorgehensweise

Bei der Diskussion von Forschungsfrage 3 fiel auf, dass Studierende, die in der Schule einen Grundkurs in Mathematik besuchten, sowie Studierende des Studienganges Bauingenieurwesen überproportional häufig in der Gruppe der prozedural schwachen Studierenden vertreten sind. Studierende mit einem Leistungskurs in Mathematik und Studierende des Studiengangs Wirtschaftsingenieurwesen sind hingegen häufiger in der Gruppe der prozedural starken Studierenden zu finden. Neben den prozedural schwachen, prozedural durchschnittlichen und prozedural starken Studierenden (zur Definition dieser Kategorien siehe Abschnitt 5.3) wird hier daher zusätzlich die Entwicklung des prozeduralen Wissens bei Studierenden mit einem Grundkurs bzw. Leistungskurs Mathematik sowie von Studierenden der Studiengänge Bau- bzw. Wirtschaftsingenieurwesen untersucht.
Die Untersuchungen werden dabei einmal bei allen Studierenden der genannten Gruppen über einen Zeitraum von zwei Semestern (2 Messzeitpunkte) und einmal über einen Zeitraum von drei Semestern durchgeführt (3 Messzeitpunkte). Die exakten Messzeitpunkte sind Abschnitt 4.1.1 zu entnehmen. Berücksichtigt werden dabei alle Studierenden, die am Speed- und Leistungstest prozedurales Wissen zu beiden bzw. zu allen drei Messzeitpunkten teilgenommen haben und einer der genannten Gruppen angehören.

Da sich für viele Studierende die Mathematikausbildung nur über zwei Semester erstreckt, können im zwei Semester umfassenden Zeitraum entsprechend mehr Studierende einbezogen werden als im drei Semester umfassenden Zeitraum. Im letztgenannten Zeitraum scheidet beispielsweise die Untersuchung der Studiengänge Logistik, Bau- und Wirtschaftsingenieurwesen aus, da für diese nur eine zweisemestrige Mathematikvorlesung vorgesehen ist.

Prozedurales Wissen wird in der vorliegenden Arbeit operationalisiert durch die Fertigkeit, Basisprozeduren der Ingenieurmathematik in angemessener Zeit durchführen zu können. Die entsprechenden Erhebungen erfolgen mit den verschiedenen Varianten des Speed- und Leistungstests prozedurales Wissen (vgl. Abschnitt 4.2.1). Die Varianten sind Rasch-skaliert und über Anker-Tests miteinander verlinkt. Die Testrohwerte (Logit-Werte) werden zur besseren Lesbarkeit in entsprechende Werte der PISA-Skala umgerechnet (ebenda). Diese ist so normiert, dass der Mittelwert der Parameter gleich 500 ist und die Standardabweichung 100 beträgt. Die Normierung wurde auf der Grundlage aller Testergebnisse zum ersten Messzeitpunkt (MZP1) vorgenommen, damit von allen anderen Messzeitpunkten aus ein Bezug zur Leistung zum ersten Messzeitpunkt hergestellt werden kann.
Um die nachfolgenden Ergebnisse bessere einordnen zu können, seien hier einige Vergleichszahlen aus PISA 2012, wo dieselbe Skalierung zugrunde liegt, genannt: Die Differenz in der mathematischen Kompetenz zwischen Mädchen und Jungen in Deutschland liegt bei 13 Punkten auf der PISA-Skala (Prenzel et al., 2013, S. 76). Der Unterschied in mathematischer Kompetenz zwischen Schülerinnen und Schülern in Deutschland (514 Punkte) und dem erstplatzierten Land (Korea, 554 Punkte) beträgt 40 Punkte. Der Abstand zum letztplatzierten Land (Mexiko, 413 Punkte) liegt bei 99 Punkten. Deutschland liegt 20 Punkte über dem OECD-Durchschnitt von 494 Punkten (Prenzel et al., 2013, S. 71).

Ergebnisse zur Veränderung prozeduralen Wissens innerhalb von zwei Semestern

Abbildung 5.4 gibt einen Überblick über das prozedurale Wissen in verschiedenen Gruppen von Studierenden zum ersten und zweiten Messzeitpunkt (MZP1 bzw. MZP2). Die

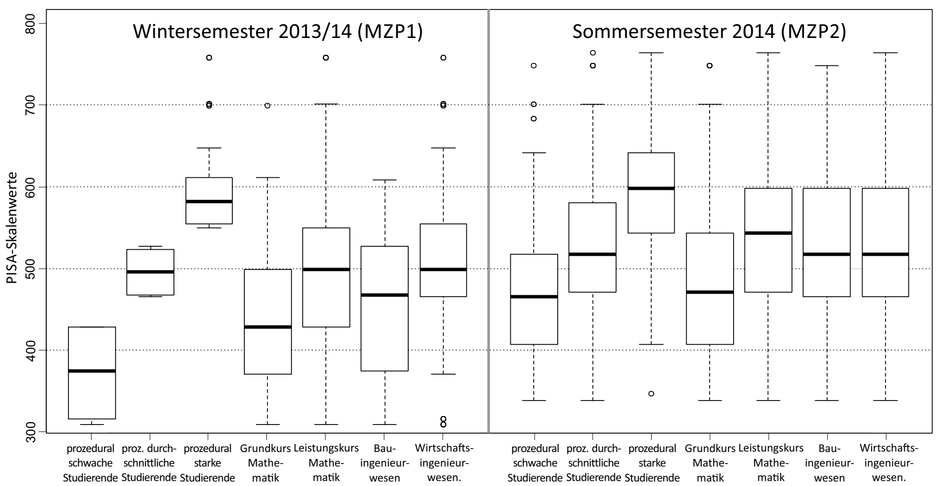


Abbildung 5.4 Leistungsvergleich aller untersuchten Gruppen in den ersten beiden Semestern

Tabelle 5.12

Veränderung des prozeduralen Wissens vom 1. zum 2. Semester

		1. Sem.			2. Sem.			Veränderung				
	N	M	SD	SE	M	SD	SE	ΔM	$\|t\|$	df	p	$\|d\|$
prozedural schwache Studierende	285	379	47	2.8	458	85	5.0	79	16.0	284	.00	**<u>1.1</u>**
proz. durchschnittliche Studierende	338	494	24	1.3	530	82	4.5	36	8.5	337	.00	**0.6**
prozedural starke Studierende	200	593	47	3.3	595	81	5.7	2	0.4	199	.73	
Grundkurs Mathematik	258	436	86	5.4	480	93	5.8	44	8.1	257	.00	**0.5**
Leistungskurs Mathematik	565	497	86	3.6	539	94	4.0	42	11.9	564	.00	**0.5**
Bauingenieurwesen	45	463	83	12.4	519	101	15.1	56	5.0	44	.00	**0.6**
Wirtschaftsingenieurwesen	204	500	89	6.2	530	96	6.7	30	5.0	203	.00	0.3

Anmerkung. N=Stichprobengröße, M=Mittelwert, SD=Standardabweichung, SE=Standardfehler, $|d|$=Absolutbetrag der Effektstärke (Cohens d); Interpretation gemäß Cohen (1988, S. 355): $|d| \approx 0.2$: schwacher Effekt, $|d| \approx 0.5$: mittlerer Effekt, $|d| \approx 0.8$: starker Effekt; $|t|$ Absolutbetrag des t-Wertes (zweiseitiger t-Test auf Gleichheit der Mittelwerte, gepaarte Stichproben); df=Anzahl Freiheitsgrade

horizontalen Balken in den Boxplots geben den Median an, die darauffolgenden Begrenzungen das obere bzw. untere Quartil, zwischen denen 50 % der Daten liegen. Die restlichen Daten verteilen sich, von Ausreißern abgesehen, zwischen der Box und den oberen bzw. unteren Begrenzungen.

In Tabelle 5.12 sind die entsprechenden deskriptiv- und inferenzstatistischen Detailinformationen über die mittlere Leistungsentwicklung angegeben. Die Tests sind miteinander verlinkt, sodass die Ergebnisse in den einzelnen Semestern miteinander vergleichbar sind und daraus die entsprechende Leistungsveränderung bestimmt werden kann (vgl. Abschnitt 4.2.1). Diese wird in der Spalte ΔM angegeben. In den nachfolgenden Spalten ist die Signifikanz der Abweichung und ggf. die Effektstärke (vgl. Abschnitt 4.5) ausgewiesen. Diese werden ab $p < .05$ berichtet. Mittlere bis starke Effekte sind fett gedruckt, starke Effekte sind zusätzlich unterstrichen. Die Daten werden in Abbildung 5.5 grafisch aufbereitet präsentiert.

Interpretation der Ergebnisse in Hinblick auf Forschungsfrage 4

Vergleich der Leistungsgruppen (prozedural schwach, prozedural durchschnittlich und prozedural stark): In Abbildung 5.4 fällt auf, dass die Streuung in den drei Leistungsgruppen (*prozedural schwach, durchschnittlich, stark*) zum MZP2 deutlich zunimmt. Dies ist durch

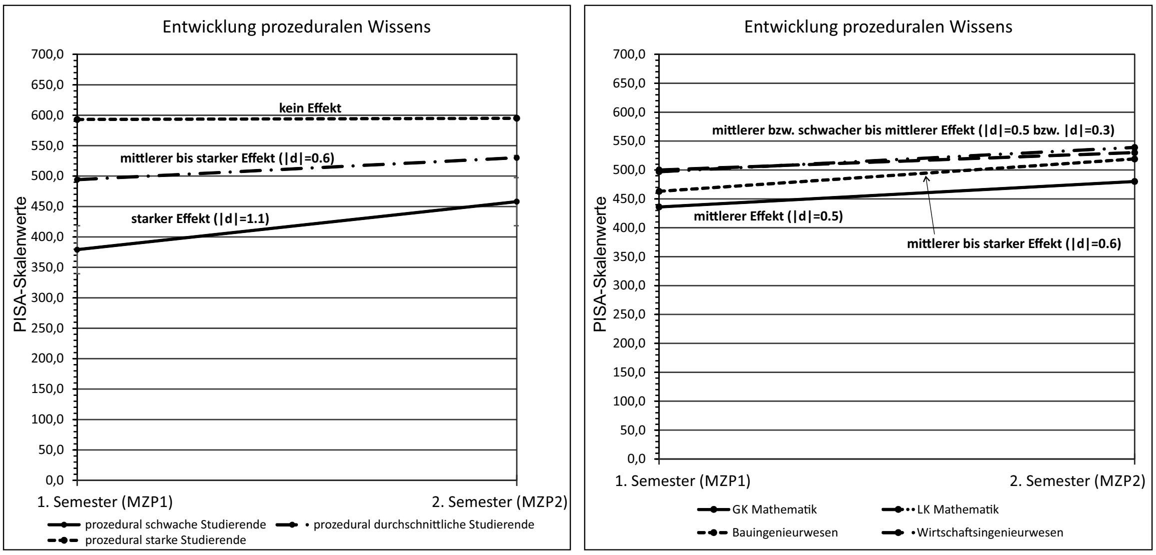


Abbildung 5.5 Entwicklung des prozeduralen Wissens vom ersten zum zweiten Semester innerhalb verschiedener Studierendengruppen

die Definition der Leistungsgruppen zu erklären: Zum MZP1 werden alle Studierenden in Abhängigkeit von ihrer im Speed- und Leistungstest prozeduralen Wissen erreichten Punktzahl in die genannten drei Leistungsgruppen eingeteilt. Die Streuung zum MZP1 ist daher entsprechend gering. Zum MZP2 nimmt dann die Streuung zu, da sich in allen Leistungsgruppen Studierende unterschiedlich stark verbessern, aber auch verschlechtern. Prozedural schwache Studierende steigern sich im Durchschnitt um 79 Punkte, prozedural durchschnittliche Studierende um 36 Punkte. Die Zuwächse sind jeweils auf dem 1 %-Niveau signifikant und entsprechen einem starken ($|d| = 1.1$) bzw. mittleren bis starken Effekt ($|d| = 0.6$). Keine signifikante Veränderung liegt hingegen bei den prozedural starken Studierenden vor. Möglicherweise befinden sich diese bereits zu Studienbeginn an einer testbedingten und damit von den teilnehmenden Studierenden unabhängigen Leistungsgrenze. Diese Vermutung wird nachfolgend erneut bei der Analyse des dreisemestrigen Zeitraums mit drei Messzeitpunkten aufgegriffen und diskutiert.

Aufgrund der vorliegenden Beobachtungen lässt sich also feststellen, dass der Leistungszuwachs im prozeduralen Wissen und damit bezüglich der Fertigkeit, Basisprozeduren der Ingenieurmathematik in angemessener Zeit durchführen zu können, umso größer ist, je schwächer eine Gruppe anfangs ist. Bei der späteren Betrachtung des dreisemestrigen Zeitraums ist daher von besonderem Interesse, ob diese Aussage auch bezüglich der Leistungsentwicklung vom zweiten zum dritten Messzeitpunkt zutrifft.

Obwohl prozedural schwache Studierende vom ersten zum zweiten Messzeitpunkt den stärksten Leistungszuwachs aufweisen, erreichen sie zum MZP2 mit durchschnittlich 458 Punkten nicht das Ausgangsniveau der prozedural durchschnittlichen Studierenden zum MZP1 von im Schnitt 494 Punkten. Eine mögliche Konsequenz daraus könnte sein, leistungsschwache Studierende frühzeitig, idealerweise vor dem ersten Messzeitpunkt, zu fördern und damit ihr Ausgangsniveau bezüglich prozeduralen Wissens anzuheben.

Vergleich der Kurstypen (Grundkurs Mathematik, Leistungskurs Mathematik): Studierende, die einen Grundkurs in Mathematik besuchten, erreichen zum ersten Messzeitpunkt durchschnittlich 436 Punkte. Studierende, die einen Leistungskurs in Mathematik absolvierten, erzielen dagegen 497 Punkte. Die Zuwächse sind jeweils auf dem 1 %-Niveau signifikant. Beide Gruppen steigern ihre Leistung um eine halbe Standardabweichung ($|d| = 0.5$). Der Leistungszuwachs ist also vergleichbar hoch. Studierende mit einem Leistungskurs in Mathematik sind hinsichtlich ihres Ausgangsniveaus und ihrer Leistungsentwicklung mit den prozedural durchschnittlichen Studierenden vergleichbar. Studierende mit einem Grundkurs in Mathematik, die in der Gruppe der prozedural schwachen Studierenden überproportional häufig vertreten sind (vgl. Abschnitt 5.3), sind dagegen mit dieser Gruppe nicht vergleichbar, denn innerhalb der Gruppe der prozedural schwachen Studierenden ist der Leistungszuwachs deutlich höher und das durchschnittliche Ausgangsniveau mit 379 Punkten geringer. Dies könnte darauf hindeuten, dass der stärkere Leistungszuwachs in dieser Gruppe insbesondere auf die Studierenden mit einem Leistungskurs in Mathematik zurückzuführen ist, die zu Beginn zur prozedural schwachen Gruppe gehören. Möglicherweise können diese Studierende aufgrund ihrer Kurswahl auf mehr Ressourcen zurückgreifen, was dann zu einem stärkeren Leistungszuwachs führt. Die Hintergrundmerkmale „Grundkurs Mathematik“ und „Leistungskurs Mathematik“ sind folglich alleine nicht geeignet, um prozedural schwache von prozedural starken Studierenden zu trennen, was auch bereits aus Tabelle 5.8 und Abbildung 5.2 hervorgeht. Als Kriterium für die Teilnahme an möglichen Unterstützungsangeboten sollten daher leistungsbezogene Kriterien verwendet werden, wie etwa die erreichte Punktzahl im Speed- und Leistungstest prozedurales Wissen.

Vergleich der Studiengänge Bauingenieurwesen und Wirtschaftsingenieurwesen: Studierende des Studienganges Bauingenieurwesen erzielen zum MZP1 durchschnittlich 463 Punkte, während Studierende des Wirtschaftsingenieurwesens im Schnitt 500 Punkte erreichen. Zum MZP2 holen Studierende des Studiengangs Bauingenieurwesen auf und erreichen durchschnittlich 519 Punkte, während Studierende des Studiengangs Wirtschaftsingenieurwesen im Schnitt 530 Punkte erzielen. Die Leistungszuwächse von 56 bzw. 30 Punkten sind jeweils signifikant auf dem 1 %-Niveau. Es handelt sich im ersten Fall um einen mittleren bis starken Effekt ($|d| = 0.6$) und im zweiten Fall um einen schwachen bis mittleren Effekt ($|d| = 0.3$). Der Abstand zwischen Studierenden der beiden Studiengänge liegt somit zu Beginn bei 37 Punkten und reduziert sich zum MZP2 auf 11 Punkte, was auf einen größeren Leistungszuwachs der Bauingenieurwesen-Studierenden zurückzuführen ist. Möglicherweise hängt dies damit zusammen, dass der Studiengang Bauingenieurwesen in anderen Vorlesungen mehr Übungsgelegenheiten bietet, um prozedurales Wissen bzw. Fertigkeit in der Durchführung von Basisprozeduren aufzubauen. Dieser Frage könnte in nachfolgenden Forschungsvorhaben nachgegangen werden.

Studierende des Studienganges Wirtschaftsingenieurwesen sind mit prozedural durchschnittlichen Studierenden vergleichbar, Studierende des Bauingenieurwesens liegen dagegen etwa zwischen prozedural schwachen und prozedural durchschnittlichen Studierenden. Von den untersuchten Gruppen ähnelt keine der Gruppe der prozedural starken Studierenden und auch prozedural schwache Studierende sind keiner anderen Gruppe sehr ähnlich. Dies bestätigt eine Interpretation der Ergebnisse aus Tabelle 5.8 und Abbildung 5.2, die zeigen, dass die Gruppen der leistungsstarken und leistungsschwachen Studierenden eher heterogen bezüglich der Hintergrundmerkmale zusammengesetzt sind. Dies spricht erneut dafür, bei der Implementierung möglicher Fördermaßnahmen eher leistungsbezogene Merkmale, wie beispielsweise das prozedurale Wissen, zu berücksichtigen.

Ergebnisse zur Veränderung prozeduralen Wissens innerhalb von drei Semestern

Analog zur vorangegangenen Darstellung der Entwicklung prozeduralen Wissens in dem zwei Semester umfassenden Zeitraum gibt Abbildung 5.6 einen groben Überblick über die Leistungsverteilung und -entwicklung in den untersuchten Gruppen innerhalb von drei Semestern. Tabelle 5.13 enthält analog die deskriptiv- und inferenzstatistischen Detailinformationen, während in Abbildung 5.7 die Daten nochmals grafisch aufbereitet dargestellt sind.

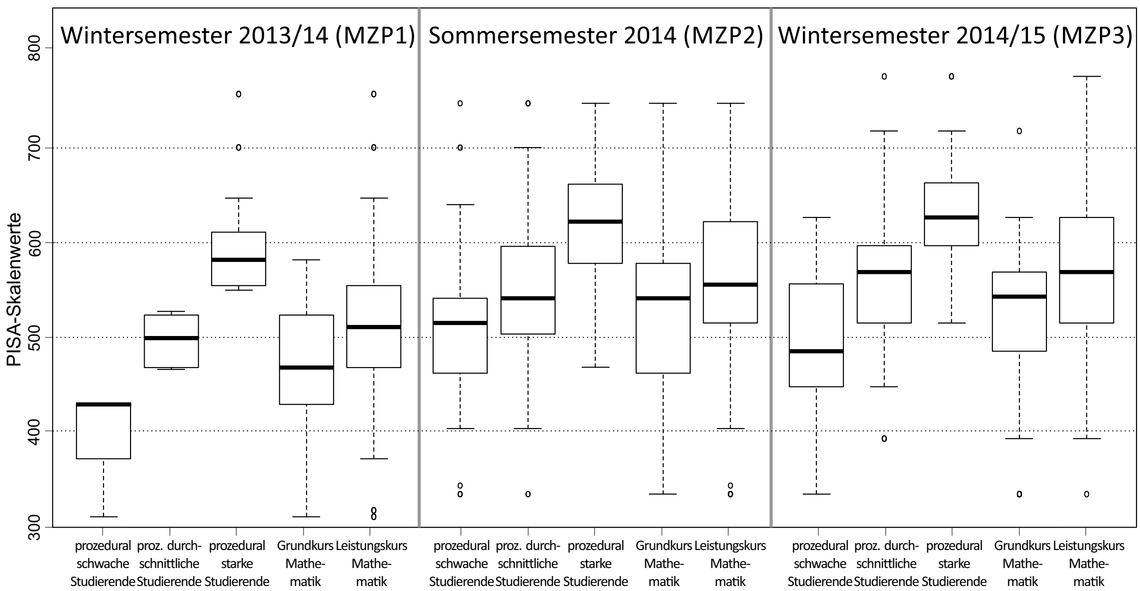


Abbildung 5.6 Leistungsvergleich aller untersuchten Gruppen

Tabelle 5.13

Veränderung des prozeduralen Wissens zwischen den Messzeitpunkten

		1. Semester (MZP1)			3. Semester (MZP3)							
	N	*M*	*SD*	*SE*	*M*	*SD*	*SE*	ΔM	$\lvert t \rvert$	*df*	*p*	$\lvert d \rvert$
prozedural schwache Studierende	59	389	46	6.0	486	82	10.7	98	9.3	58	.00	**1.4**
proz. durchschnittliche Studierende	123	499	24	2.2	561	69	6.2	62	10.1	122	.00	**1.2**
prozedural starke Studierende	71	598	47	5.6	621	60	7.1	23	3.0	70	.00	0.4
Grundkurs Mathematik	45	458	81	12.1	519	90	13.5	61	5.0	44	.00	**0.7**
Leistungskurs Mathematik	208	510	81	5.6	569	81	5.6	59	11.6	207	.00	**0.7**
		1. Semester (MZP1)			2. Semester (MZP2)			Veränderung				
	N	*M*	*SD*	*SE*	*M*	*SD*	*SE*	ΔM	$\lvert t \rvert$	*df*	*p*	$\lvert d \rvert$
prozedural schwache Studierende	59	389	46	6.0	505	87	11.3	116	10.5	58	.00	**1.6**
proz. durchschnittliche Studierende	123	499	24	2.2	547	76	6.8	48	7.2	122	.00	**0.8**
prozedural starke Studierende	71	598	47	5.6	614	66	7.9	16	1.9	70	.06	
Grundkurs Mathematik	45	458	81	12.1	529	90	13.4	71	4.9	44	.00	**0.8**
Leistungskurs Mathematik	208	510	81	5.6	562	84	5.8	51	9.2	207	.00	**0.6**
		2. Semester (MZP2)			3. Semester (MZP3)							
	N	*M*	*SD*	*SE*	*M*	*SD*	*SE*	ΔM	$\lvert t \rvert$	*df*	*p*	$\lvert d \rvert$
prozedural schwache Studierende	59	505	87	11.3	486	82	10.7	-19	1.8	58	.08	
proz. durchschnittliche Studierende	123	547	76	6.8	561	69	6.2	14	1.8	122	.08	
prozedural starke Studierende	71	614	66	7.9	621	60	7.1	7	0.7	70	.47	
Grundkurs Mathematik	45	529	90	13.5	519	90	13.5	-10	0.8	44	.46	
Leistungskurs Mathematik	208	562	84	5.8	569	81	5.6	7	1.3	207	.18	

Anmerkung. Erläuterungen siehe Tabelle 5.12

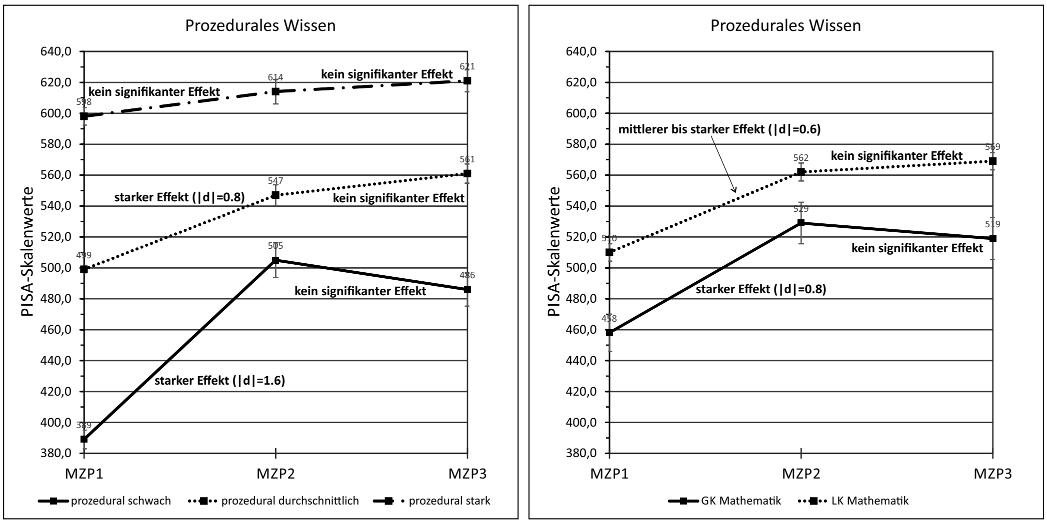


Abbildung 5.7 Entwicklung des prozeduralen Wissens vom ersten bis zum dritten Semester (MZP1 bis MZP3) innerhalb verschiedener Studierendengruppen

Interpretation der Ergebnisse im Hinblick auf Forschungsfrage 4

Vergleich der Leistungsgruppen (prozedural schwach, prozedural durchschnittlich und prozedural stark): Bei den Studierenden, die eine dreisemestrige Mathematikveranstaltung absolvieren, verläuft die Leistungsentwicklung vom ersten zum zweiten Messzeitpunkt ähnlich wie bei der Gesamtgruppe derjenigen Studierenden, die eine drei oder nur eine zweisemestrige Veranstaltung besuchen (vgl. Tabelle 5.4). Die zunehmende Streuung zum MZP2 erklärt sich wie zuvor dadurch, dass die Leistungsgruppen zum MZP1 auf Grundlage ihrer Testergebnisse im Speed- und Leistungstest prozedurales Wissen gebildet wurden. Hier ist die Streuung entsprechend gering. Da sich die Studierenden innerhalb der Leistungsgruppen unterschiedlich entwickeln, nimmt die Streuung zum MZP2 zu.

Den größten Leistungszuwachs zwischen MZP1 und MZP2 verzeichnen mit 1.6 Standardabweichungen ($|d| = 1.6$) die prozedural schwachen Studierenden, gefolgt von den prozedural durchschnittlichen Studierenden mit $|d| = 0.8$. Während diese Veränderungen auf dem 1 %-Niveau signifikant sind, erfolgt bei den prozedural starken Studierenden erneut keine signifikante Leistungszunahme zwischen MZP1 und MZP2, weshalb auch keine Effektstärke auszuweisen ist.

Im Gegensatz zu der zuvor untersuchten Gesamtgruppe aller Studierenden, die eine zwei oder drei Semester umfassende Mathematikvorlesung besuchen (vgl. Tabelle 5.4), erreichen in der hier untersuchten Gruppe der Studierenden, die eine drei Semester umfassende Veranstaltung besuchen, prozedural schwache Studierende nach einem Semester knapp das Ausgangsniveau der prozedural durchschnittlichen Studierenden. Dies kann als Indiz dafür gedeutet werden, dass ein „Aufstieg“ nach einem Semester in eine höhere Leistungsklasse möglich ist. Durch eine geeignete Unterstützung könnte daher versucht werden, dieses „Aufsteigen“ vorzeitig zu erreichen und bestenfalls bis zum ersten Messzeitpunkt prozedural schwache Studierende so weit zu fördern, dass sie das durchschnittliche Ausgangsniveau der prozedural durchschnittlichen Studierenden (ca. 500 Punkte) erreichen.

Zwischen dem zweiten und dritten Messzeitpunkt gelingt keiner Gruppe eine signifikante Leistungssteigerung. Damit bestätigt sich auch die bereits zuvor geäußerte Vermutung, dass prozedural starke Studierende bereits zum ersten Messzeitpunkt eine „natürliche“, vermutlich testbedingte, Leistungsgrenze erreicht haben, die sich nur noch in geringem Maße überschreiten lässt. Zwischen dem ersten und dritten Messzeitpunkt nimmt die Leistung der prozedural starken Studierenden lediglich um durchschnittlich 23 Punkte zu (vgl. Tabelle 5.13), was eine signifikante Steigerung mit einer schwachen bis mittleren Effektstärke ($|d| = 0.4$) darstellt.

Auffällig ist, dass die Leistung der leistungsschwachen Studierenden zwischen MZP2 und MZP3 – wenngleich auch nicht signifikant – *sinkt*. Dies steht in einem starken Gegensatz dazu, dass deren Leistung zwischen MZP1 und MZP2 unter den drei Leistungsgruppen am stärksten *wächst*. Dieser Trend kann also nicht fortgesetzt werden, er wird sogar umgekehrt. Ein ähnliches Ergebnis findet sich in der Längsschnittstudie PALMA (Pekrun et al., 2006): Hier nimmt die Leistung der Schülerinnen und Schüler in der schwächsten Leistungsgruppe ebenfalls ab, allerdings kontinuierlich und bereits ab dem ersten Messzeitpunkt. Hinweise auf eine Leistungsabnahme zu späteren Messzeitpunkten bei leistungsschwachen Studierenden finden sich auch bei Henn und Polaczek (2007) in ihrer Untersuchung von 971 Studierenden der Ingenieurwissenschaften über einen Zeitraum von 5 Semestern: „Defizite, die im ersten Studienfachsemester auftraten, vergrößern sich sogar noch in den folgenden Studienfachsemestern“ (S. 146). Über die genauen Gründe für die stark differente Leistungsentwicklung vom ersten zum zweiten und vom zweiten zum dritten Semester können hier nur Vermutungen angestellt werden: In den ersten drei Forschungsfragen wurde gezeigt, dass prozedural schwache Studierende auch schwach in der späteren Klausur abschneiden. Ferner bestehen Schwächen nicht nur im prozeduralen Wissen, sondern auch in anderen leistungsrelevanten Bereichen, was ein Hinweis darauf sein könnte, dass prozedural schwache Studierende auch in anderen Erstsemesterklausuren eher schwach abschneiden. Folglich könnte es sein, dass prozedural schwache Studierende innerhalb des ersten Semesters oder Studienjahres vergleichsweise erfolglos sind. Erfolg und Misserfolg während des ersten Studienjahres haben einen wesentlichen Einfluss auf den weiteren Studienverlauf (Ortiz & Dehon, 2013; Henn & Polaczek, 2007; Kolb et al., 2006). Aus diesen Gründen könnte es nach dem zweiten Semester bei prozedural schwachen Studierenden möglicherweise zu motivationalen Problemen kommen oder auch dazu, dass weniger Übungsgelegenheiten genutzt werden. Genauere Erkenntnisse hierüber und zu den tatsächlichen Ursachen für die Umkehrung der Leistungsentwicklung nach dem zweiten Semester müssen nachfolgende Forschungsvorhaben erbringen. Bezogen auf die vorliegenden Resultate erscheint es insgesamt empfehlenswert, leistungsschwache Studierende so früh wie möglich und insbesondere erneut nach dem zweiten Messzeitpunkt zu unterstützen, um die Dynamik in der Leistungsentwicklung aufrechtzuhalten.

Vergleich der Kurstypen (Grundkurs Mathematik, Leistungskurs Mathematik): Studierende, die in der Schule einen Leistungskurs in Mathematik absolvierten, sind prozedural etwas stärker als prozedural durchschnittliche Studierende (510 gegenüber 499 Punkten), Studierende mit einem Grundkurs Mathematik mit 458 Punkten dagegen etwas schwächer. Es fällt auf, dass die Leistung von Studierenden mit Grundkurs Mathematik analog zur Leistung der prozedural schwachen Studierenden zwischen MZP1 und MZP2 zunächst stark ($|d| = 0.8$) und auf dem 1 %-Niveau signifikant zunimmt und anschließend von MZP2 zu MZP3 um 10 Punkte *sinkt*. Das heißt auch bei Studierenden, die einen Grundkurs in Mathematik besuchten, findet nach dem zweiten Messzeitpunkt eine Trendumkehr in der Leistungsentwicklung statt. Aus diesem Grund könnte möglicherweise eine Empfehlung für die Teilnahme an einer Intervention nach MZP2 auch allein auf Grundlage des Hintergrundmerkmals „Grundkurs in Mathematik besucht“ getroffen werden. Wie oben schon geschildert, sollten allerdings zunächst in anderen Forschungsvorhaben die genauen Gründe für die Trendumkehr in der Leistungsentwicklung nach MZP2 aufgeklärt werden.

Fazit

Je prozedural schwächer die untersuchte Studierendengruppe zu Beginn ist, desto stärker ist der Leistungszuwachs innerhalb der ersten beiden Semester. Dieser Trend setzt sich allerdings nicht im dritten Semester fort, d. h. in den prozedural schwachen Gruppen stagniert die Entwicklung oder die Leistung nimmt sogar ab.
Prozedural schwache Studierende erreichen je nach Betrachtungszeitraum nach einem Semester nicht bzw. gerade eben das Anfangsniveau der prozedural durchschnittlichen Studierenden. Gleichzeitig erreicht die Leistungsdifferenz zwischen prozedural schwachen und prozedural durchschnittlichen Studierenden zum MZP2 ein Minimum. Danach vergrößert sich die Leistungsdifferenz wieder (Schereneffekt).

Aus diesen Beobachtungen könnten sich zwei Konsequenzen für mögliche Unterstützungsmaßnahmen ergeben: Erstens sollte versucht werden, durch eine frühzeitige Förderung das Anfangsniveau prozedural schwacher Studierender anzuheben. Zweitens sollte versucht werden, die positive Leistungsentwicklung prozedural schwacher Studierender zwischen MZP1 und MZP2 zum dritten Messzeitpunkt fortzusetzen. Hierfür ist vermutlich eine geeignete Intervention im zweiten Studiensemester erforderlich. Im Rahmen weiterer Forschungsarbeiten sollte allerdings geklärt werden, warum es zu der beobachteten

Trendumkehr in der Leistungsentwicklung der prozedural schwachen Studierenden nach dem zweiten Semester kommt. Im Rahmen der PALMA-Studie wurde beispielsweise festgestellt, dass leistungsschwache Schülerinnen und Schüler deutlich ungünstigere Fähigkeitsselbstkonzepte und höhere Angstwerte in Mathematik haben als der Durchschnitt (Pekrun et al., 2006, S. 34). Letzteres wird durch die Werte in Tabelle 5.11 bestätigt. Denkbar ist ferner, dass prozedural schwache Studierende – eventuell motivationsbedingt nach einigen Misserfolgen während der Studieneingangsphase – insgesamt weniger Lern- und Übungsgelegenheiten nutzen als ihre prozedural durchschnittlichen und prozedural starken Kommilitoninnen und Kommilitonen und sich der starke Leistungszuwachs auch deshalb nicht über das zweite Semester hinaus fortsetzen lässt.
In weiterführenden Forschungsarbeiten sollten auch situative Einflüsse untersucht werden, wie beispielsweise der Umgang der Lehrkräfte mit schwächeren Studierenden und deren Berücksichtigung bei der Konzeption von Lehrveranstaltungen. Hieran knüpft unmittelbar eine Reihe Anschlussfragen an, beispielsweise wie gut Lehrkräfte im Umgang mit leistungsschwächeren Studierenden und deren speziellen Bedürfnissen geschult sind und ob Lehrveranstaltungen einseitig starke Studierende fördern, während schwache Studierende möglicherweise eher verunsichert werden. Darüber hinaus ist zu fragen, ob schwächere Studierende tatsächlich weniger Lerngelegenheiten nutzen, beispielsweise indem sie Lehr- und Übungsveranstaltungen häufiger fernbleiben als ihre durchschnittlichen Kommilitoninnen und Kommilitonen. Möglicherweise nutzen stärkere Studierende auch andere Lerngelegenheiten, die sie in ihren Mathematikleistungen stärker fördern als schwächere Studierende. Um also schwächere Studierende bei einem positiven Studienverlauf zu unterstützen, erscheint es erforderlich, Studierende und ihre Interaktionen mit studiumrelevanten Personen, Institutionen und Situationen sowie die allgemeinen Rahmenbedingungen möglichst gut in der Breite zu erforschen.

5.5 Analyse von Ressourcen und Schwächen prozedural schwacher Studierender bei der Durchführung einer Basisprozedur

Im Rahmen der ersten beiden Forschungsfragen wurden Zusammenhänge zwischen dem Klausurergebnis und verschiedenen Faktoren sowie Hintergrundmerkmalen untersucht (vgl. Abschnitt 5.1 und 5.2). In diesen Untersuchungen stellte sich prozedurales Wissen als stärkster Prädiktor für Klausurerfolg und -leistung heraus. Dabei wurde prozedurales Wissen operationalisiert durch die Anzahl der im Speed- und Leistungstest prozedurales Wissen korrekt gelösten Aufgaben (vgl. Abschnitt 4.2.1). Dieser Test prüft Basisprozeduren der Ingenieurmathematik (vgl. Abschnitt 1.1.1) ab, die korrekt und in angemessener Zeit durchgeführt werden müssen. Auf Grundlage der Testergebnisse wurden dann in den Forschungsfragen 3 und 4 Quer- und Längsschnittanalysen durchgeführt, um das prozedurale Wissen von Studierenden noch detaillierter zu analysieren. Dabei stellte sich in der längsschnittlichen Betrachtung heraus, dass prozedural schwache Studierende im Durchschnitt zwar besser werden, aber auch zu späteren Messzeitpunkten kaum das Ausgangsniveau der prozedural durchschnittlichen Studierenden erreichen. Oder anders formuliert: Die meisten prozedural schwachen Studierenden bleiben trotz gewisser Verbesserungen prozedural schwach. Anknüpfend an dieses Resultat gilt es nun in einer qualitativen Analyse zu untersuchen, woran prozedural schwache Studierende zu den einzelnen Messzeitpunkten scheitern. Dabei kann auch untersucht werden, wie sich individuelle Ressourcen und Schwächen entwickeln. Forschungsfrage 5 lautet dementsprechend: *Woran scheitern prozedural schwache Studierende bei der Durchführung von Basisprozeduren und wie entwickeln sich individuelle Ressourcen und Schwächen?* Während die vorangegangenen Forschungsfragen quantitativer Natur waren, verlagert sich nun also das Erkenntnisinteresse auf die inhaltliche Analyse konkreter Studierendenbearbeitungen. Dieses ist mit qualitativen Methoden tiefer zu bearbeiten.

Die Untersuchung von Forschungsfrage 5 ist relevant für die Gestaltung von Fördermaßnahmen, die zum Ziel haben, prozedural schwache Studierende bei der korrekten und effizienten Verwendung von Basisprozeduren zu unterstützen, denn die fehlerfreie und zeitgerechte Bearbeitung der Basisprozeduren stellt eine wichtige Grundanforderung für den Klausurerfolg in der Ingenieurmathematik dar (vgl. Abschnitt 1.2.2).

Vorgehensweise

Definition von Ressourcen und Schwächen: Zur Untersuchung von Forschungsfrage 5 wird analysiert, wie sich Ressourcen und Schwächen prozedural schwacher Studierender bei der Durchführung von Basisprozeduren über mehrere Semester ohne gezielte Fördermaßnahmen verändern. Als (mathematische) Ressourcen werden diejenigen Aspekte der Kalkülfertigkeit und -kenntnis betrachtet, die den jeweiligen Lernenden zur Lösung des betrachteten Items zur Verfügung stehen. Sie werden mit dem Anforderungsprofil des Items verglichen. Als Schwächen werden die nicht gezeigten Aspekte des itemspezifischen Anforderungsprofils untersucht.

Auswahl des Items: Die notwendigen Ressourcen und mögliche Schwächen bei Studierenden werden exemplarisch an den Bearbeitungen eines Items des Speed- und Leistungstests prozedurales Wissen rekonstruiert. Hierfür wurde das Item „Ableitungen 5“ (vgl. Abbildung 5.8) aus folgenden Gründen ausgewählt:

1. Das Item umfasst zwölf Anforderungen, die fünf verschiedene Aspekte der Kalkülfertigkeit bzw. -kenntnis abdecken (Ableiten, Funktion auswerten, Grundrechenarten verwenden, Bearbeitungszeit, Regelkenntnis). Dies macht die Analyse überschaubar und dennoch können unterschiedliche Aspekte prozeduralen Wissens auf einer differenzierten Ebene untersucht werden.

2. Der wesentliche technische Aspekt des Items (Produktregel) ist aus der Schule bekannt und wird im zweiten Semester vor MZP2 im Veranstaltungsbetrieb noch einmal mit entsprechenden Lerngelegenheiten eingeführt. Dies kommt in der besonders deutlichen Verbesserung der Lösungsquoten zum Ausdruck: Die Lösungsquote der prozedural durchschnittlichen Studierenden steigt beispielsweise von 43 % (MZP1) auf 77 % (MZP2) bzw. 86 % (MZP3). Eine positive Entwicklung ist zwar auch bei den prozedural schwachen Studierenden zu verzeichnen, dennoch liegt die Lösungsquote zum MZP3 erst bei 66 %. Dies macht das Item besonders interessant für die Frage, an welchen Anforderungen viele prozedural schwache Studierenden auch nach erneuten Lerngelegenheiten, also nach dem MZP1, noch scheitern.

Sachanalyse zur Erstellung des Anforderungsprofils: Die Analyse wird wie folgt durchgeführt: Zunächst wird das Item im Rahmen einer Sachanalyse betrachtet und der Lösungsweg vorgestellt. Da es sich um eine Basisprozedur handelt, bauen die einzelnen Lösungsschritte aufeinander auf und müssen folglich in der dargestellten Reihenfolge ausgeführt werden (vgl. hierzu auch die Definition von Prozedur in Abschnitt 1.1.1). Anschließend werden die Anforderungen, die zum erfolgreichen Lösen des Items erforderlich sind, expliziert und in einem Anforderungsprofil zusammengefasst und nummeriert. Das Anforderungsprofil dient als Vergleichsmaßstab für die Inhaltsanalyse von neun Bearbeitungsbeispielen von drei Studierenden zu drei Messzeitpunkten.

Auswahl der Fälle: Mit Blick auf die Gestaltung von möglichen Förderinstrumenten zur Unterstützung der Gruppe der prozedural schwachen Studierenden wurden drei Fälle ausgewählt, die innerhalb der Gruppe der prozedural schwachen Studierenden zu den schwächsten Studierenden zählen und zu keinem der drei Messzeitpunkte das betrachtete Item korrekt lösen konnten. Es sind alle drei durch die erste Mathematikklausur gefallen. Bezüglich der anderen Hintergrundmerkmale wurden der Schultyp, die G9-Zugehörigkeit sowie das Studiensemester konstant gehalten, die Grund- und Leistungskurszugehörigkeit gezielt kontrastiert. Präzisere Informationen zu den drei Fällen Lena, Max und Tim (Namen geändert) sind der Tabelle 5.14 zu entnehmen.

Tabelle 5.14
Überblick über die Hintergrundmerkmale der Studierenden

Name	Lena	Max	Tim
Studiengang	Chemieing.	Maschinenbau	Maschinenbau
Kurswahl Mathematik	Grundkurs	Leistungskurs	Leistungskurs
Schultyp	Gymnasium	Gymnasium	Gymnasium
G8/G9	G9	G9	G9
Studiensemester zum MZP1	1	1	1
Punktzahl Test* zum MZP1	0	2	0
Punktzahl Test* zum MZP2	1	1	0
Punktzahl Test* zum MZP3	1	1	2
Punktzahl Klausur (von 80)	25	30	23

Anmerkung. *Speed- und Leistungstest prozedurales Wissen

Die grobe Durchsicht zahlreicher Bearbeitungen bestätigt, dass die Bearbeitungen dieser drei Fälle für die Gruppe der prozedural schwachen Studierenden insofern typisch sind, als dass die aufzuzeigenden Phänomene auch bei anderen Studierenden zu beobachten sind (für die quantitative Übersicht vgl. Abschnitt 5.4).

Längsschnittliche Inhaltsanalyse: Bei der Inhaltsanalyse dieser Bearbeitungen wird mit Bezug auf das Anforderungsprofil des Items auf Ressourcen und Schwächen in der Kalkülkenntnis und -fertigkeit sowie auf deren Entwicklung innerhalb des Beobachtungszeitraums von drei Semestern eingegangen. Bezogen auf die Anforderungen werden dabei Ressourcen in der Kalkülkenntnis bzw. -fertigkeit mit einem „+“ gekennzeichnet, Schwächen entsprechend mit einem „−“. Die Bearbeitungsschritte sind in der Form [Xm-n] kodiert. Dabei steht X für den Anfangsbuchstaben der oder des Studierenden, $m \in \{1, 2, 3\}$ bezeichnet den Messzeitpunkt (MZP1, MZP2 oder MZP3) und $n \in \mathbb{N}$ den Bearbeitungsschritt.

Sachanalyse

Beschreibung des Items und der Testbedingungen

In Abbildung 5.8 ist im linken Teil die Aufgabenstellung wiedergegeben. Diese wurde im Rahmen des Speed- und Leistungstests prozedurales Wissen (vgl. Abschnitt 4.2.1) für

90 Sekunden über einen Beamer eingeblendet. Die vorgegebene Bearbeitungszeit von 90 Sekunden entspricht dem Doppelten der Zeit, die drei Tutorinnen und Tutoren, die mit dem Thema vertraut waren, im Durchschnitt zur Lösung des Items inklusive Aufschreiben der Lösung benötigten. Nach Ablauf dieser Zeit wurde die Aufgabenstellung ausgeblendet und es erschienen für etwa 15 Sekunden die im rechten Teil von Abbildung 5.8 dargestellten Lösungsoptionen. Die richtige Option musste dann von den Teilnehmenden auf dem Bearbeitungsbogen markiert werden. Durch dieses Testdesign war sichergestellt, dass alle Teilnehmenden dieselbe Zeit zur Bearbeitung des Items zur Verfügung hatten. Insbesondere konnte das Item später nicht erneut bearbeitet werden.
Zur Vorbereitung auf den Test wurde etwa vier Wochen vor der Testdurchführung Übungsmaterial mit vollständigen Lösungswegen online gestellt. Das Material enthielt zur Übung auch ein zum Item „Ableitung 5“ strukturgleiches Item.

LTK-2013-05-24-10-Ableitungen 5

Bearbeitungszeit: $\boxed{1{:}30}$ Minuten **Aufgabe 5**

Gegeben sei die Funktion

$$f(x) = \ln(x) \cdot (x^3 + x^{3/2}) \,.$$

Berechnen Sie den Funktionswert der ersten Ableitung an der Stelle $x = 1$, d.h. berechnen Sie $\boldsymbol{f'(1)}$. Vereinfachen und kürzen Sie Ihr Ergebnis so weit wie möglich.

Antwortoptionen **Aufgabe 5**

$f'(1) =$

1 9	2 1	3 4	4 6
5 0	6 5	7 3	8 8
9 7	10 2		

Abbildung 5.8 Item „Ableitungen 5“ mit Lösungsoptionen. Diese werden nach Ablauf der Bearbeitungszeit für 15 Sekunden eingeblendet.

Lösungsweg

- Um die Ableitungen von f zu bestimmen, ist die Produktregel anzuwenden. Definiere dazu Funktionen g und h wie folgt:
  $f(x) = \underbrace{\ln(x)}_{=:g(x)} \cdot \big(\underbrace{x^3 + x^{3/2}}_{=:h(x)}\big)$
- Es gilt $g'(x) = 1/x$ (Ableitung der Logarithmusfunktion).
- $h'(x) = 3x^2 + \frac{3}{2}x^{1/2}$ (Ableitung der Potenzfunktion)
  Vorausschauend gedacht ist die Berechnung von h' nicht erforderlich, da nach Anwenden der Produktregel der Vorfaktor $\ln(x)$ an der Stelle $x = 1$ gleich Null ist.
- Anwenden der Produktregel ($f' = g'h + gh'$) ergibt:
  $f'(x) = \frac{1}{x}\big(x^3 + x^{3/2}\big) + \ln(x)\big(3x^2 + \frac{3}{2}x^{1/2}\big)$
- Funktionsauswertung(en) an der Stelle $x = 1$ und Zusammenfassen der Terme:
  $f'(1) = \underbrace{\frac{1}{1}\big(1^3 + 1^{3/2}\big)}_{=2} + \underbrace{\ln(1)}_{=0}\big(3 \cdot 1^2 + \frac{3}{2} \cdot 1^{1/2}\big) = \underline{\underline{2}}$

Anforderungen zur Lösung des Items

Das Item erfordert die Anwendung der Produktregel, die aus der Schule bekannt sein sollte (Ministerium für Schule und Weiterbildung NRW, 2014, S. 28). Bei der Produktregel handelt es sich um eine Basisprozedur. Die Kenntnis der Formel für die Produktregel ist daher der Kalkülkenntnis zuzurechnen. Die benötigten Ableitungen von Grundfunktionen (Logarithmusfunktion und Potenzfunktionen) stellen mathematische Grundoperationen zur erfolgreichen Durchführung einer Basisprozedur dar und werden daher der Kalkülfertigkeit zugerechnet. Die Auswertung der Logarithmusfunktion und Potenzfunktionen an der Stelle $x = 1$ sind ebenfalls mathematische Grundoperationen und damit der Kalkülfertigkeit zuzurechnen. Bei vorausschauender Rechnung könnte auf die Ableitung der Potenzfunktion verzichtet werden, da der Vorfaktor gleich Null wird. Aus diesem Grund werden diese Anforderungen in Klammern gesetzt. Folgt man allerdings der Prozedur, so ist zur vollständigen Angabe von f' die Berechnung der Ableitung doch erforderlich. Zudem hat dieser Aspekt auf die Analyse der Scheiterungsgründe keinen Einfluss, vielmehr ergeben sich nützliche Zusatzinformationen über die Kalkülfertigkeit prozedural schwacher Studierender. Aus diesem Grund werden alle aufgeführten Anforderungen in die Analyse einbezogen.
Eine weitere Anforderung an die Kalkülfertigkeit besteht darin, eine Lösung innerhalb der vorgegebenen Zeit von 90 Sekunden zu bestimmen. Im Detail ergeben sich damit für das vorliegende Item die in Tabelle 5.15 dargestellten Anforderungen an die Kalkülfertigkeit und -kenntnis.

Tabelle 5.15
Anforderungsprofil für das Item „Ableitungen 5“

Nr.	**Anforderungen an Kalkülfertigkeit**
A1	Ableitung von $\ln x$ bestimmen
A2	(Ableitung von x^3: Exponent berechnen)
A3	(Ableitung von x^3: Koeffizient bestimmen)
A4	(Ableitung von $x^{3/2}$: Exponent berechnen)
A5	(Ableitung von $x^{3/2}$ Koeffizient bestimmen)
A6	Auswertung von $\ln x$ an der Stelle $x = 1$
A7	Auswertung von x^3 an der Stelle $x = 1$
A8	Auswertung von $x^{3/2}$ an der Stelle $x = 1$
A9	Addition/Multiplikation von Zahlen
A10	Resultat innerhalb von 90 Sekunden erzielen
	Anforderungen an Kalkülkenntnis
A11	Faktoren g und h identifizieren
A12	Produktregel formal korrekt anwenden

Inhaltsanalyse

Entwicklung der Bearbeitungen im Fall Lena

Die Bearbeitungen des Items „Ableitungen 5“ des ersten Falls Lena zum MZP1 (Ende der Vorlesungszeit des ersten Semesters), MZP2 (Ende der Vorlesungszeit des zweiten Semesters) und MZP3 (Ende der Vorlesungszeit des dritten Semesters, vgl. Abschnitt 4.1.1) sind in Abbildung 5.9 abgedruckt.

Die Analyseergebnisse werden in Tabelle 5.16 zunächst zusammenfassend dargestellt.

Lena - MZP1

[L1-01] $f'(x) = \ln \cdot \left(\frac{1}{2} x^2 + 3 x^{\frac{1}{2}}\right)$

[L1-02] $= 0,1 \cdot \left(\frac{1}{2} + \frac{3}{2}\right)$

[L1-03] $= 0,1 \cdot 2$

[L1-04] $= 0,2$

Lena - MZP2

[L2-01] $f(x) = \ln(x) \cdot (x^3 + x^{\frac{3}{2}})$

[L2-02] $f'(x) = 3x + \frac{3}{2} x^{\frac{1}{2}} \cdot \ln(x) + \ln(x) \cdot x^3 + x^{\frac{3}{2}}$

[L2-03] $f(1) = 0$

Lena - MZP3

Ich habe keine Lösung bestimmt, weil (diese Angaben gehen nicht in die Bewertung ein):

- [x] verrechnet
- [ ] zu wenig Zeit
- [ ] Kalkül noch nicht so richtig gelernt bzw. nicht mehr so richtig bekannt
- [ ] sonstiges (bitte hier vermerken):

[L3-01] $f(x) = \ln(x) \cdot (x^3 + x^{\frac{3}{2}})$

[L3-02] $= \ln(x) \cdot \left(3x + \frac{3}{2} x^{\frac{1}{2}}\right)$

[L3-03] $= \ln(1) \cdot \left(3 + \frac{3}{2} \cdot 1^{\frac{1}{2}}\right)$

[L3-04] $= 1 \cdot \left(3 + \frac{3}{2}\right)$

[L3-05] $= 1 \cdot \left(\frac{9}{2}\right)$

[L3-06] $= \frac{9}{2}$

Abbildung 5.9 Bearbeitung des Items „Ableitung 5“ durch Lena zu den drei Messzeitpunkten

Tabelle 5.16
Zusammenfassung der Ressourcen und Schwächen von Lena in der Kalkülkenntnis und -fertigkeit bei der Lösung des Items „Ableitungen 5“

Nr.	**Anforderungen an Kalkülfertigkeit**	Lösungsschritt in Abb. 5.9	MZP1 +/ − /?	MZP2 +/ − /?	MZP3 +/ − /?
A1	Ableitung von $\ln x$ bestimmen	[L1-01] [L2-02] [L3-02]	-	-	-
A2	(Ableitung von x^3: Exponent berechnen)	[L1-01] [L2-02] [L3-02]	+	-	-
A3	(Ableitung von x^3: Koeffizient bestimmen)	[L1-01] [L2-02] [L3-02]	-	+	+
A4	(Ableitung von $x^{3/2}$: Exponent berechnen)	[L1-01] [L2-02] [L3-02]	+	+	+
A5	(Ableitung von $x^{3/2}$ Koeffizient bestimmen)	[L1-01] [L2-02] [L3-02]	+	+	+
A6	Auswertung von $\ln x$ an der Stelle $x = 1$	[L1-02] [L2-03] [L3-04]	-	+	-
A7	Auswertung von x^3 an der Stelle $x = 1$	[L1-02] [L2-03] [L3-03]	+	+	+
A8	Auswertung von $x^{3/2}$ an der Stelle $x = 1$	[L1-02] [L2-03] [L3-03]	+	+	+
A9	Addition/Multiplikation von Zahlen	[L1-04] [L2-03] [L3-05/06]	+	+	+
A10	Resultat innerhalb von 90 Sekunden erzielen		+	+	+
	Anforderungen an Kalkülkenntnis				
A11	Faktoren g und h identifizieren	[L1-01] [L2-01] [L3-02]	+	+	+
A12	Produktregel formal korrekt anwenden	[L1-01] [L2-02] [L3-02]	-	+	-

*A*nmerkung. „+“=Anforderung erfüllt, „−“=Anforderung nicht erfüllt; MZP=Messzeitpunkt

Analyse von Lenas Bearbeitungen zum MZP1: Lena schreibt in [L1-01] als erstes die Ableitung von f auf. Dabei wendet sie nicht die korrekte Formel für die Produktregel an (A12−) und leitet die Logarithmusfunktion nicht (korrekt) ab. Dies kann an einer fehlerhaften Vorstellung der Produktregel liegen, der zufolge z. B. der erste Faktor nicht abzuleiten ist, möglicherweise glaubt Lena aber auch, dass sich die Ableitung wie bei der Exponentialfunktion reproduziert. Mit Blick auf Lenas Bearbeitungen zum MZP3 und insbesondere zum MZP2, wo sie die Produktregel korrekt anzuwenden scheint, ist eher davon auszugehen, dass sie die Ableitung der Logarithmusfunktion nicht kennt (A1−). Bei der Ableitung der Potenzfunktion in [L1-01] reduziert Lena den Grad der Exponenten korrekt (A2+, A4+), wählt aber die Koeffizienten falsch (A3−), was insgesamt auf falsche Ableitungen der Potenzfunktionen führt. In [L1-02] setzt Lena dann den Wert $x = 1$ in die von ihr berechnete Ableitung ein. Hierbei wertet sie die Logarithmusfunktion nicht korrekt aus. Für Lena gilt $\ln(1) = 0,1$ (A6−). Wie sie zu dieser Vorstellung gelangt, kann der Bearbeitung nicht entnommen werden. Die Auswertung der Potenzfunktionen gelingt dagegen (A7+, A8+). Ferner deutet der Bruch $\frac{3}{2}$ in [L1-02] an, dass Lena die zweite Potenzfunktion ($x^{3/2}$) korrekt abgeleitet und den Koeffizienten korrekt bestimmt hat (A4+, A5+). Den anderen Koeffizienten ($\frac{1}{2}$) korrigiert sie allerdings nicht, sodass dieser fehlerhaft bleibt. In [L1-03] fasst Lena die Brüche korrekt zusammen und auch die Multiplikation mit $0,1$ wird in [L1-04] korrekt ausgeführt (A9+). Ihr Ergebnis erzielt Lena innerhalb der vorgegebenen Zeit (A10+).

Insgesamt ist damit festzuhalten, dass Lena zum MZP1 vermutlich die Addition gleichnamiger Brüche und das Kürzen sowie die arithmetischen Grundoperationen beherrscht. Die Grundfunktionen (Logarithmus- und Potenzfunktionen) und deren Ableitungen scheint sie jedoch nicht im hier erforderlichen Maße zu kennen. Die Produktregel wendet Lena nicht korrekt an. Ihre Ressourcen reichen damit für ein erfolgreiches Lösen des vorliegenden Items nicht aus.

Analyse von Lenas Bearbeitungen zum MZP2: Die Markierung von $\ln(x)$ als g und $(x^3 + x^{3/2})$ als f in [L2-01] deutet an, dass Lena erkennt, dass sie die Produktregel anwenden muss. Die Produktregel scheint sie dann in [L2-02] auch korrekt anzuwenden (A11+, A12+), wenn man annimmt, dass Lena wie zum MZP1 $\ln(x)$ für die Ableitung von $\ln(x)$ hält (A1−). Die Ableitung der Potenzfunktionen in [L2-02] gelingt fast: Die Koeffizienten sind nun korrekt (A3+, A5+), allerdings ist ein Exponent falsch (A2−). Auch die Funktionsauswertung in [L2-03] gelingt (A6+, A7+, A8+), sofern man voraussetzt, dass Lena „im Kopf" in [L2-02] Klammern korrekt setzt. Insbesondere scheint Lena nun zu wissen, dass $\ln(1) = 0$ gilt. Lena schafft es erneut, ihr Ergebnis im Rahmen der vorgegebenen Zeit zu erzielen (A10+).

Insgesamt ist davon auszugehen, dass Lena zu einem richtigen Ergebnis kommen würde, wenn sie die Ableitung der Logarithmusfunktion korrekt angeben könnte. Lenas Bearbeitung ist daher ein wichtiges Beispiel dafür, wie „kleine" basismathematische Schwächen, wie ein fehlerhaftes Verständnis von der Ableitung der Logarithmusfunktion, zum Scheitern bei der Durchführung einer Basisprozedur führen. Dies wiederum führt vermutlich dazu, dass auch komplexere Klausuraufgaben nicht korrekt gelöst werden können, wofür Lenas Klausurergebnis (25 Punkte von 32 Punkten, die zum Bestehen nötig sind) ein Indiz ist.

Analyse von Lenas Bearbeitungen zum MZP3: Zum MZP3 wendet Lena die Formel für die Produktregel in [L3-02] nicht korrekt an (A12−). Im Gegensatz zu ihrer Bearbeitung zum MZP2 ist hier nicht klar, ob Lena erkennt, dass sie die Produktregel anwenden muss. Vermutlich hält sie $\ln(x)$ nach wie vor für die Ableitung von $\ln(x)$ (A1−). Die Potenzfunktionen werden fast korrekt abgeleitet, allerdings reduziert Lena wieder den Exponenten von x^3 nicht korrekt (A2−). In [L3-04] wertet Lena die einzelnen Funktionen aus. Diesmal gelingt die Auswertung der Logarithmusfunktion nicht (A6−), ihre Auswertung der Potenzfunktion ist hingegen korrekt (A7+, A8+) und auch die Addition der Brüche und die korrekte Erweiterung gelingen (A9+). Lena bestimmt ihr Ergebnis erneut innerhalb der vorgegebenen Zeit (A10+).

Damit lässt sich festhalten, dass Lena auf wichtige Ressourcen (Bruchrechnen, arithmetische Grundoperationen) nach wie vor erfolgreich zurückgreifen kann, während zuvor

erworbene Ressourcen (Kenntnis der Produktregel, $\ln(1) = 0$) nicht mehr Teil ihres prozeduralen Netzwerkes bzw. zum MZP3 nicht abrufbar sind.
Vergleich von Lenas Bearbeitungen zu den drei Messzeitpunkten: Der zeilenweise Vergleich in Tabelle 5.16 ist ausgesprochen aufschlussreich für die Analyse der längsschnittlichen Entwicklung. Lena wendet zum MZP2 die Produktregel korrekt an, was als positive Entwicklung gegenüber MZP1 zu deuten ist. Zum MZP3 kann sie dieses Wissen allerdings nicht mehr adäquat abrufen. Eine analoge Entwicklung liegt bei der Auswertung der Logarithmusfunktion an der Stelle $x = 1$ vor: Hier gelingt die Auswertung zum MZP1 nicht, zum MZP2 hingegen schon, zum MZP3 allerdings wieder nicht. Die Ableitung der Logarithmusfunktion beherrscht Lena zu keinem Zeitpunkt. Während sich bei der Ableitung von x^3 die Bestimmung des Koeffizienten zum MZP2 und MZP3 gegenüber MZP1 verbessert, liegt für die Berechnung des Exponenten der umgekehrte Fall vor: Hier weist Lena zum MZP2 und MZP3 Schwächen auf, zum MZP1 gibt sie hingegen den korrekten Exponenten an. Insgesamt scheint Lena zum MZP3 über ein geringeres prozedurales Wissen zu verfügen als zum MZP2. Dies entspricht den quantitativen Ergebnissen zu Forschungsfrage 4, denen zufolge im Durchschnitt das prozedurale Wissen der prozedural schwachen Studierenden nach dem MZP2 abnimmt (von 505 Einheiten auf der PISA-Skala zum MZP2 auf 486 Punkte zum MZP3, vgl. Tabelle 5.13).
Bemerkenswert ist ferner, dass Lena zu den Messzeitpunkten jeweils mindestens zwei Schwächen aufweist. Eine Reduzierung um nur eine einzelne Schwäche reicht bei Lena meistens nicht aus, um die Basisprozedur erfolgreich durchzuführen.
Zusammenfassend fällt auf, dass einige der eben genannten Aspekte der Kalkülfertigkeit und -kenntnis in Lenas prozeduralem Netzwerk eher flüchtig vorhanden zu sein scheinen und nicht verankert werden konnten (vgl. auch die Bemerkungen zur Stabilität und Nachhaltigkeit prozeduralen Wissens in Abschnitt 1.1.3).

Entwicklung der Bearbeitungen im Fall Max

Die Bearbeitungen des Items „Ableitungen 5“ des zweiten Falls Max zu allen drei Messzeitpunkten sind in Abbildung 5.10 abgedruckt.

Max - MZP1	[M1-01]	$f(x) = \ln(x) \cdot (x^3 + x^{3/2})$
	[M1-02]	$f'(x) = \frac{1}{x} \cdot (x^3 + x^{3/2}) + \ln(x) \cdot 3x^2 + \frac{3}{2} x^{-\frac{2}{3}}$
	[M1-03]	$\frac{1}{1} \cdot (1^3 + 1^{3/2}) + \ln(1) \cdot 3(1)^2 + \frac{3}{2} \cdot 1^{-\frac{2}{3}}$
	[M1-04]	$f(x)$ $1 + (\ln(1)) \cdot 3 + \frac{3}{2}$

Abbildung 5.10 Bearbeitung des Items „Ableitung 5“ durch Max zu den drei Messzeitpunkten

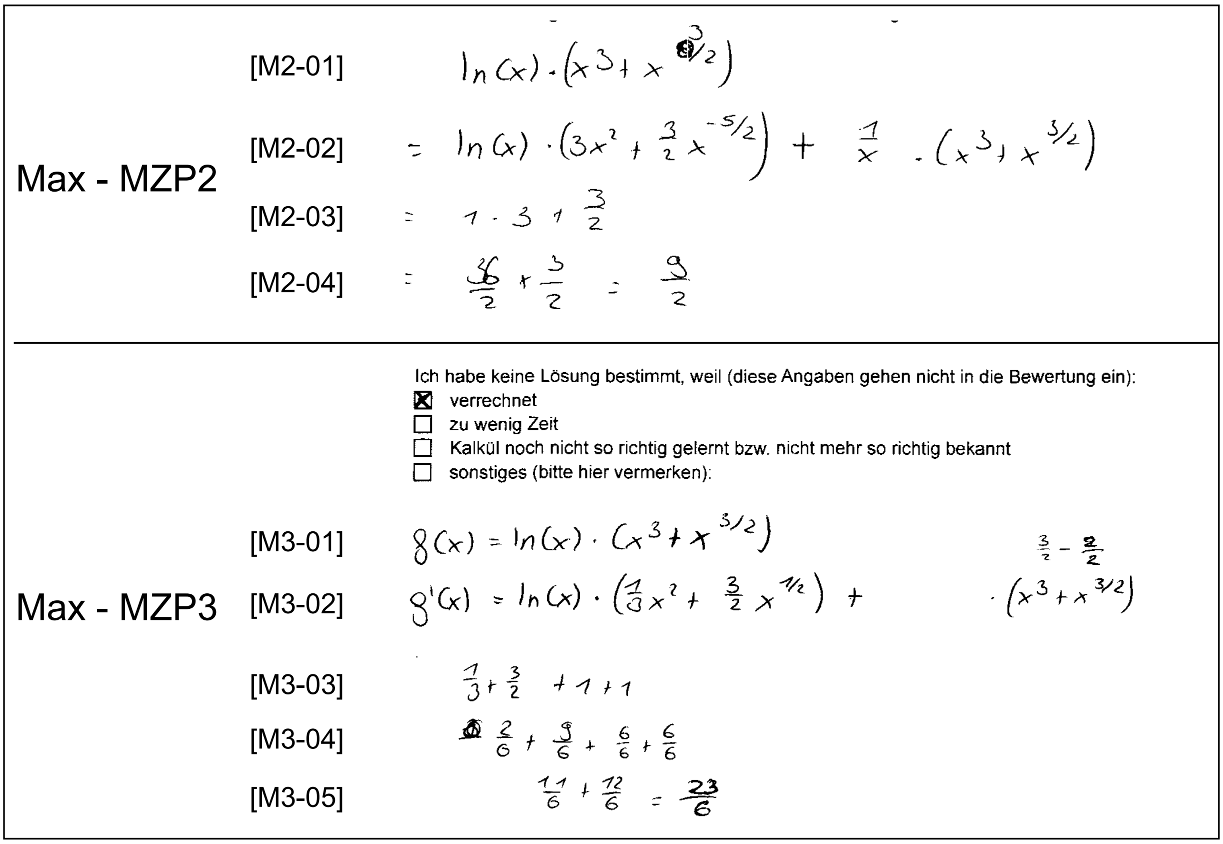


Abbildung 5.10 – Fortsetzung

Die Analyseergebnisse werden in Tabelle 5.17 zunächst zusammenfassend dargestellt.

Tabelle 5.17

Zusammenfassung der Ressourcen und Schwächen von Max in der Kalkülkenntnis und -fertigkeit zur Lösung des Items „Ableitungen 5“

Nr.	**Anforderungen an Kalkülfertigkeit**	Lösungsschritt in Abb. 5.10	MZP1 +/ − /?	MZP2 +/ − /?	MZP3 +/ − /?
A1	Ableitung von $\ln x$ bestimmen	[M1-02] [M2-02] [M3-02]	+	+	+
A2	(Ableitung von x^3: Exponent berechnen)	[M1-02] [M2-02] [M3-02]	+	+	+
A3	(Ableitung von x^3: Koeffizient bestimmen)	[M1-02] [M2-02] [M3-02]	+	+	-
A4	(Ableitung von $x^{3/2}$: Exponent berechnen)	[M1-02] [M2-02] [M3-02]	-	-	+
A5	(Ableitung von $x^{3/2}$ Koeffizient bestimmen)	[M1-02] [M2-02] [M3-02]	+	+	+
A6	Auswertung von $\ln x$ an der Stelle $x = 1$	[M1-04] [M2-03] [M3-03]	-	-	-
A7	Auswertung von x^3 an der Stelle $x = 1$	[M1-04] [M2-03] [M3-03]	+	+	+
A8	Auswertung von $x^{3/2}$ an der Stelle $x = 1$	[M1-04] [M2-03] [M3-03]	+	+	+
A9	Addition/Multiplikation von Zahlen	[M1-04] [M2-04] [M3-04/05]	-	+	+
A10	Resultat innerhalb von 90 Sekunden erzielen		+	+	+
	Anforderungen an Kalkülkenntnis				
A11	Faktoren g und h identifizieren	[M1-02] [M2-02] [M3-02]	+	+	+
A12	Produktregel formal korrekt anwenden	[M1-02] [M2-02] [M3-02]	+	+	+

Anmerkung. „+“=Anforderung erfüllt, „−“=Anforderung nicht erfüllt; MZP=Messzeitpunkt

Analyse von Max Bearbeitungen zum MZP1: Max wendet in [M1-02] die Produktregel korrekt an (A11+, A12+). Die Ableitung der Logarithmusfunktion kennt er ebenfalls (A1+). Die Ableitung der Potenzfunktionen gelingt ihm fast. Max gibt hierbei einen Exponenten fehlerhaft an: Statt $\frac{1}{2}$ bestimmt er $-\frac{2}{3}$ (A4−). Er scheint den Zusammenhang $\ln(1) = 0$ nicht zu kennen (A6−). Bemerkenswert hierbei ist, dass Max zwar die Ableitung der Logarithmusfunktion kennt, aber vermutlich nicht den Verlauf des Graphen dieser Funktion, da er den Ausdruck $\ln(1)$ nicht auswertet. Aufgrund einer fehlenden Klammer in [M1-02] um den Term $3x^3 + \frac{3}{2}x^{-\frac{2}{3}}$ fasst Max schließlich in [M1-04] die Ausdrücke falsch zusammen, ferner rechnet er $1^3 + 1^{\frac{3}{2}} = 1$ (A9−).
Analyse von Max Bearbeitungen zum MZP2: Zum zweiten Messzeitpunkt ist die Produktregel nach wie vor Bestandteil seines prozeduralen Netzwerkes. In [M2-02] wendet er die Regel korrekt an (A11+, A12+). Die Ableitung der Logarithmusfunktion ist ebenfalls korrekt (A1+). Bei der Ableitung der Potenzfunktion bereitet ihm wie zum MZP1 die Ableitung von Potenzfunktionen mit einem rationalen Exponenten Schwierigkeiten. Hier bestimmt Max erneut den Exponenten falsch (A4−). In [M2-03] nimmt Max $\ln(1) = 1$ an (A6−). Ferner scheint er zu vergessen, den Term $\frac{1}{x} \cdot (x^3 + x^{\frac{3}{2}})$ aus [M2-02] auszuwerten. Die abschließende Erweiterung und Addition der Brüche gelingt (A9+).
Analyse von Max Bearbeitungen zum MZP3: Wie zu den vorangegangenen Messzeitpunkten wendet Max auch zum MZP3 die Produktregel in [M3-02] korrekt an (A11+, A12+). Ferner gelingt ihm erstmals die Ableitung der Potenzfunktion mit rationalem Exponenten (A4+, A5+). Dafür unterläuft Max ein Fehler bei der Ableitung von x^3. Hier gibt Max $\frac{1}{3}x^2$ an, was darauf hindeuten könnte, dass er Differentiation und Integration vermischt (A3−). Bei der anschließenden Funktionsauswertung in [M3-03] nimmt Max wieder $\ln(1) = 1$ an (A6−). Die anschließende Erweiterung und Addition von Brüchen gelingen (A9+).
Vergleich von Max Bearbeitungen zu den drei Messzeitpunkten: Max verfügt bereits zum MZP1 über wichtige Ressourcen zur Lösung des Items (Produktregel, Ableitung der Logarithmsfunktion), die auch Bestandteil seines prozeduralen Netzwerkes bleiben bzw. auf die er bis zum MZP3 zugreifen kann. Ferner gelangt er zu allen Messzeitpunkten innerhalb der vorgegebenen Zeit zu einem Ergebnis. Allerdings gelingt es ihm nicht, Potenzfunktionen fehlerfrei abzuleiten sowie Funktionswerte der Logarithmusfunktion korrekt zu bestimmen. Hierin zeigt sich, wie kleine basismathematische Schwächen, die zum Teil bis zum dritten Semester nicht behoben werden, zum Scheitern bei der Durchführung einer Basisprozedur führen können. Dies kann sich negativ auf die Durchführung komplexerer mathematischer Verfahren auswirken und zu einem schlechteren Abschneiden in Klausuren beitragen, in denen die korrekte und effiziente Durchführung von Basisprozeduren zu den Grundanforderungen gehört. Bemerkenswert ist ferner, dass Max zu jedem Messzeitpunkt mindestens zwei Schwächen aufweist.

Entwicklung der Bearbeitungen im Fall Tim

Die Bearbeitungen des Items „Ableitungen 5“ des dritten Falls Tim zu allen drei Messzeitpunkten sind in Abbildung 5.11 abgedruckt.

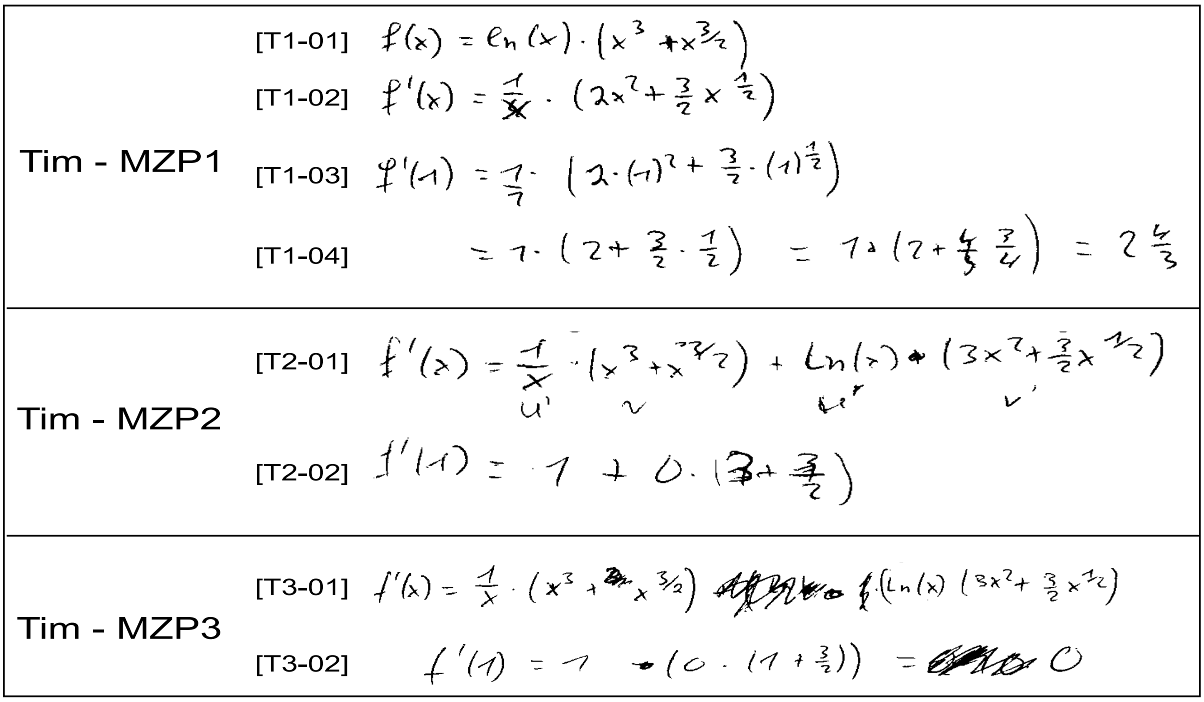


Abbildung 5.11 Bearbeitung des Items „Ableitung 5“ durch Tim zu den drei Messzeitpunkten

Die Analyseergebnisse werden in Tabelle 5.18 zunächst zusammenfassend dargestellt.

Tabelle 5.18

Zusammenfassung der Ressourcen und Schwächen von Tim in der Kalkülkenntnis und -fertigkeit zur Lösung des Items „Ableitungen 5“

Nr.	**Anforderungen an Kalkülfertigkeit**	Lösungsschritt in Abb. 5.11	MZP1 +/ − /?	MZP2 +/ − /?	MZP3 +/ − /?
A1	Ableitung von $\ln x$ bestimmen	[T1-02] [T2-01] [T3-01]	+	+	+
A2	(Ableitung von x^3: Exponent berechnen)	[T1-02] [T2-01] [T3-01]	+	+	+
A3	(Ableitung von x^3: Koeffizient bestimmen)	[T1-02] [T2-01] [T3-01]	-	+	+
A4	(Ableitung von $x^{3/2}$): Exponent berechnen)	[T1-02] [T2-01] [T3-01]	+	+	+
A5	(Ableitung von $x^{3/2}$ Koeffizient bestimmen)	[T1-02] [T2-01] [T3-01]	+	+	+
A6	Auswertung von $\ln x$ an der Stelle $x = 1$	[T2-02] [T3-02]	?	+	+
A7	Auswertung von x^3 an der Stelle $x = 1$	[T1-04] [T2-02] [T3-02]	+	+	+
A8	Auswertung von $x^{3/2}$ an der Stelle $x = 1$	[T1-04] [T2-02] [T3-02]	+	+	+
A9	Addition/Multiplikation von Zahlen	[T1-04] [T2-02] [T3-02]	-	-	-
A10	Resultat innerhalb von 90 Sekunden erzielen		+	+	+
	Anforderungen an Kalkülkenntnis				
A11	Faktoren g und h identifizieren	[T1-02] [T2-01] [T3-01]	+	+	+
A12	Produktregel formal korrekt anwenden	[T1-02] [T2-01] [T3-01]	-	+	-

Anmerkung. „+“=Anforderung erfüllt, „−“=Anforderung nicht erfüllt, „?“=nicht erkennbar, ob Anforderung erfüllt ist; MZP=Messzeitpunkt

Analyse von Tims Bearbeitungen zum MZP1: Zum MZP1 wendet Tim in [T1-02] die Produktregel nicht korrekt an (A12−). Er geht von dem nicht korrekten Zusammenhang $(f \cdot g)' = f' \cdot g'$ aus. Die Ableitung der Logarithmusfunktion ist Tim bekannt (A1+). Die Ableitung der Potenzfunktionen gelingt fast: Tim reduziert die Exponenten korrekt um 1 (A2+, A4+), bei der Ableitung von x^3 gibt er allerdings den Koeffizienten falsch an (2 statt 3) (A3−). Die anschließende Funktionsauswertung in [T1-03] und [T1-04] gelingt (A7+, A8+). Bei der Zusammenfassung der Brüche unterläuft Tim ein Fehler (A9−).
Insgesamt lässt sich festhalten, dass Tim als wichtige Ressource die Kenntnis der Ableitung der Logarithmusfunktion mitbringt. Die Produktregel hat er hingegen nicht verinnerlicht.
Analyse von Tims Bearbeitungen zum MZP2: Zum zweiten Messzeitpunkt verwendet Tim die Produktregel nun in [T2-01] korrekt (A11+, A12+). Ferner gelingt ihm nun auch die Ableitung der Potenzfunktionen (A2+ bis A5+). Die Ableitung der Logarithmusfunktion gehört ebenfalls weiterhin zu seinem Repertoir (A1+). Tim verwendet ferner die Beziehung $\ln(1) = 0$ (A6+). Am Ende scheitert er jedoch, da er in [T2-02] $\frac{1}{1} \cdot (1^3 + 1^{\frac{3}{2}}) = 1$ berechnet (A9−).
Damit lässt sich festhalten, dass Tim zum MZP2 über fast alle Ressourcen verfügt, um das Item korrekt zu lösen. Er hat sowohl seine Kalkülfertigkeit als auch -kenntnis verbessert und scheitert am Ende aufgrund eines Rechenfehlers.
Analyse von Tims Bearbeitungen zum MZP3: In [T3-01] berechnet Tim die Ableitungen der Logarithmus- und Potenzfunktionen erneut korrekt (A1+ bis A5+). Bei der Produktregel verwendet er allerdings die nicht korrekte Beziehung $(g \cdot h)' = g'h \cdot gh'$. Die Produktregel scheint damit nicht mehr in adäquater Form Bestandteil von Tims prozeduralem Netzwerk zu sein (A12−). Die Funktionsauswertungen in [T3-02] sind korrekt, wobei Tim erneut den Fehler $\frac{1}{1} \cdot (1^3 + 1^{\frac{3}{2}}) = 1$ macht (A9−).
Vergleich von Tims Bearbeitungen zu den drei Messzeitpunkten: Im Vergleich der einzelnen Messzeitpunkte fällt auf, dass die Produktregel möglicherweise nicht gut genug verankert wurde. Während Tim diese zum MZP2 korrekt anwendet, gelingt ihm das ein Semester zuvor und danach nicht. Auffällig ist ferner, dass Tim zu zwei Messzeitpunkten denselbe elementaren Fehler ($\frac{1}{1} \cdot (1^3 + 1^{\frac{3}{2}}) = 1$) macht und darüber hinaus zu jedem Messzeitpunkt Schwächen bei der Durchführung arithmetischer Grundoperationen hat bzw. hier zu keinem korrekten Ergebnis gelangt.

Vergleich der drei Fälle Lena, Max und Tim

Beim Vergleich der Bearbeitung des Items durch Lena, Max und Tim zu den verschiedenen Messzeitpunkten fallen folgende Punkte auf.

- Die drei Studierenden weisen unterschiedliche durchgängige Schwächen auf: Lena leitet zu keinem Messzeitpunkt die Logarithmusfunktion korrekt ab, während dies Max und Tim zu jedem Messzeitpunkt gelingt. Max wertet $\ln(1)$ zu keinem Zeitpunkt korrekt aus, während Lena und Tim hiermit seltener Schwierigkeiten haben. Tim hingegen weist zu allen Messzeitpunkten Schwächen in der Ausführung arithmetischer Grundoperationen auf, was Lena und Max kaum Probleme bereitet.

- Wie aufgrund der Lerngelegenheiten zur Produktregel im zweiten Semester und damit vor MZP2 zu erwarten war, weist keiner der drei Studierenden zum MZP2 Schwächen in der Kalkülkenntnis der Produktregel auf. Bei Lena und Tim, die beide die Produktregel zum MZP1 nicht beherrschten, liegt zum MZP3 hierbei allerdings wieder eine Schwäche vor. Dies ist ein Hinweis darauf, dass die Kenntnis der Produktregel im prozeduralen Netzwerk dieser beiden Studierenden nicht ausreichend verankert werden konnte.

- Bei der Betrachtung der durchgängigen Schwächen fällt auf, dass diese zum Teil sehr schnell beseitigt werden könnten: Lena müsste lernen, dass $(\ln x)' = \frac{1}{x}$ gilt und Max müsste den Zusammenhang $\ln(1) = 0$ erfassen. Hierbei handelt es sich um Faktenwissen. Tim dagegen müsste mehr Sicherheit im Umgang mit arithmetischen Grundoperationen unter Zeitdruck erlangen.

- Bei 8 der insgesamt 9 Bearbeitungen des Items (jeweils drei Bearbeitungen pro Studierenden) weisen alle drei Studierenden mindestens zwei Schwächen auf.

Fazit

Eingangs wurde bereits erwähnt, dass der Nachweis einer Übertragbarkeit der Ergebnisse der durchgeführten qualitativen Analysen in nachfolgenden Forschungsarbeiten erbracht werden muss. Dennoch erscheint es im Hinblick auf Forschungsfrage 5 plausibel, an dieser Stelle anhand der analysierten Bearbeitungen erste mögliche Antworten zusammenzutragen, die in späteren Forschungsvorhaben überprüft werden können.

Alle beobachteten Schwächen der prozedural schwächsten Studierenden sind ausschließlich im basismathematischen Bereich zu verorten. Ohne diese basismathematischen Schwächen wären alle Studierenden in der Lage, das analysierte Item innerhalb der vorgegebenen Zeit korrekt zu lösen. Die (selbstständige) Aufarbeitung der Schwächen scheint jedoch nicht

immer zu gelingen, da einige Schwächen nicht verschwinden, sondern zu allen Messzeitpunkten auftreten: Lena bereitet die Ableitung der Logarithmusfunktion zu allen drei Messzeitpunkten Schwierigkeiten, Max die Auswertung der Logarithmusfunktion und Tim die korrekte Anwendung arithmetischer Grundoperationen. Dabei fällt auf, dass diese „dauerhaften“ Schwächen in der vorliegenden Untersuchung individuell verschieden sind.
Die Analysen enthalten darüber hinaus Hinweise, dass sich Lerngelegenheiten im Veranstaltungsbetrieb punktuell und kurzfristig positiv auf das prozedurale Netzwerk auswirken können. Dies wird zum MZP2 deutlich, zu dem alle Studierenden nach entsprechenden Lerngelegenheiten in der Vorlesung Kalkülkenntnis der Produktregel aufweisen. Allerdings scheint es zum Teil nicht zu gelingen, dieses aufgenommene prozedurale Wissen anschließend nachhaltig zu verankern, da zwei der drei Studierenden im darauffolgenden Semester hier wieder Schwächen zeigen.
Die Analysen enthalten schließlich noch eine Beobachtung, die zu einer Erklärung beitragen könnte, warum prozedural schwache Studierende ihre prozedurale Leistung nicht in dem erhofften Umfang steigern: Mit Ausnahme von Tim zum Messzeitpunkt 2 weisen alle drei Studierenden zu jedem Messzeitpunkt mindestens zwei Schwächen auf, d. h. mindestens zwei der zwölf Anforderungen werden jeweils nicht erfüllt. Dies könnte auf eine erhöhte Fehlerquote prozedural schwacher Studierender bei der Durchführung von Basisprozeduren hindeuten, die sich auch mit der Zeit nicht zu verändern scheint. Auch dieser Aspekt sollte in weiterführenden Forschungsarbeiten genauer untersucht werden.

Im Hinblick auf die Gestaltung von Fördermaßnahmen lassen sich aus den obigen Darstellungen folgende Konsequenzen ableiten:

- Die Schwächen im basismathematischen Bereich könnten und sollten zum Teil schnell behoben werden, da es sich in einigen Fällen lediglich um fehlendes Faktenwissen handelt. Durch die Zusammenstellung eines Katalogs, welches Faktenwissen abrufbar sein sollte (in Analogie zu Dürrschnabel et al., 2012), könnten der Wissensstand regelmäßig überprüft und Lücken leicht geschlossen werden.

- Eine mögliche Folgerung aus der beobachteten Heterogenität der Schwächen könnte sein, Fördermaßnahmen individuell zu gestalten, anstatt allen Studierenden dieselbe Förderung zukommen zu lassen. Dies wiederum setzt eine geeignete Diagnose voraus, die rechtzeitig erfolgen sollte, damit Fehlermuster frühzeitig beseitigt werden. Ziel der genannten Unterstützung sollte es darüber hinaus sein, prozedurales Wissen möglichst nachhaltig anzubinden, was durch eine stärkere konzeptuelle Verankerung möglich sein könnte (vgl. Abschnitt 1.1.3).

Die vorangehenden Betrachtungen fokussieren auf eine Verbesserung prozeduraler Fertigkeiten der prozedural schwachen Studierenden. Natürlich ist auch eine Überprüfung, Entwicklung und gegebenenfalls Unterstützung konzeptuellen Wissens und der mathematischen Kompetenz (Niss, 2003) insgesamt wünschenswert. Allerdings weist Niss (2003, S. 6) sehr deutlich darauf hin, dass mathematische Basisfertigkeiten und Faktenwissen eine notwendige Grundlage für den Aufbau mathematischer Kompetenz darstellen. Um diese Grundlage auch bei prozedural schwachen Studierenden zu schaffen, sind die Ergebnisse des vorliegenden Abschnitts und darauf aufbauende Forschungsvorhaben in den Bereichen der Diagnose und adaptiven Förderung mathematischer Basiskompetenzen von Studierenden wichtig sowie Untersuchungen zur Wirksamkeit und Nachhaltigkeit entsprechender Maßnahmen erforderlich.

5.6 Zur Vergleichbarkeit der Referenzklausur

Bereits im Theorieteil wurde ausgeführt, dass nach Auffassung vieler Autorinnen und Autoren in Klausuren der Ingenieurmathematik die Anforderungen an prozedurales Wissen dominieren (vgl. Abschnitt 1.2). Diese These konnte in den Untersuchungen zu den ersten beiden Forschungsfragen (vgl. Abschnitt 5.1 und 5.2) insofern bestätigt werden, als dass prozedurales Wissen, operationalisiert durch den Speed- und Leistungstest prozedurales Wissen (vgl. Abschnitt 4.2.1), einen starken Zusammenhang zu Klausurleistung und -erfolg aufweist. Allerdings bezogen sich hier alle Analysen auf eine einzige Referenzklausur (vgl. Abschnitt 4.1.4). Folglich stellt sich die Frage, inwiefern die Referenzklausur mit anderen Klausuren der Ingenieurmathematik vergleichbar ist. Forschungsfrage 6 lautet entsprechend: *Inwiefern sind die in der Referenzklausur gestellten Anforderungen mit den entsprechenden Anforderungen analoger Klausuren an anderen Hochschulen vergleichbar?* Zur Untersuchung dieser Frage werden sieben weitere Klausuren der Ingenieurmathematik analysiert und das Verhältnis von prozeduralem Wissen zu konzeptuellem Wissen und Faktenwissen untersucht. Falls sich diesbezüglich eine Vergleichbarkeit mit der Referenzklausur nachweisen lässt, ist dies eine Stütze für die begründete Hypothese, dass die Ergebnisse auf andere Hochschulen übertragbar sein müssten. Diese Hypothese wird im Interpretationsteil zu Forschungsfrage 6 aufgegriffen und detailliert diskutiert.

Vorgehensweise

Zur Untersuchung von Forschungsfrage 6 wurden die weiteren Klausuren der Ingenieurmathematik zufällig ausgewählt und für einen Vergleich wie folgt zusammengestellt: Zwei Klausuren stammen aus den Service-Bereichen[1] der Fakultät für Mathematik der TU Dortmund, weitere fünf Klausuren sind externe Klausuren von drei repräsentativen deutschen Universitäten mit 15000 bis 45000 Studierenden, deren Dozentinnen und Dozenten ihre Klausuren für anonymisierte Analysen zur Verfügung stellten. Zusammen mit der Referenzklausur liegen damit acht Klausuren zur Analyse vor. Bei allen Klausuren handelt es sich um Mathematikklausuren für Studierende ingenieurwissenschaftlicher Studiengänge im ersten Studienjahr.

In einem ersten Schritt wurde jede Klausur in Anlehnung an das Vorgehen von Jordan (2006, S. 18) in Analyseeinheiten zerlegt, wobei es sich bei den Analyseeinheiten in der Regel um die Teilaufgaben einer Hauptaufgabe handelt. Anschließend wurde für jede Analyseeinheit entschieden, ob es sich um eine sogenannte technische Aufgabe handelt

[1] Mathematik für Maschinenbau, Logistik, Wirtschafts-, Bio-, Chemie-, und Bauingenieurwesen; Mathematik für Physik, Mediziniphysik, Ingenieurinformatik, Informations- und Kommunikationstechnik; Mathematik für Chemiestudierende.

oder nicht. Unter technischen Aufgaben werden im PISA- und COACTIV-Kontext Aufgaben verstanden, „die nur technisches Wissen außerhalb jeglicher Kontextanbindungen erfordern“ (Jordan, 2006, S. 45). Dabei werden unter technischem Wissen Faktenwissen und Fertigkeiten verstanden. Fertigkeiten wiederum beinhalten „eine prozedurale Abfolge von einzelnen Schritten, die hierarchisch oder linear aufgebaut ist und deren Ablauf, wie bei einem ‚Algorithmus‘, im Großen und Ganzen festgelegt ist und gegebenenfalls automatisiert werden kann“ (ebendort). Dies bedeutet, dass technische Aufgaben entweder Faktenwissen, wie beispielsweise eine Definition aus der Vorlesung, oder prozedurales Wissen abprüfen, wobei jeweils kein konzeptuelles Wissen erforderlich ist. Die Anforderungen an prozedurales Wissen reichen dabei von einfachen Operationen auf dem Niveau der Sekundarstufe I über Basisprozeduren bis hin zu Superprozeduren der

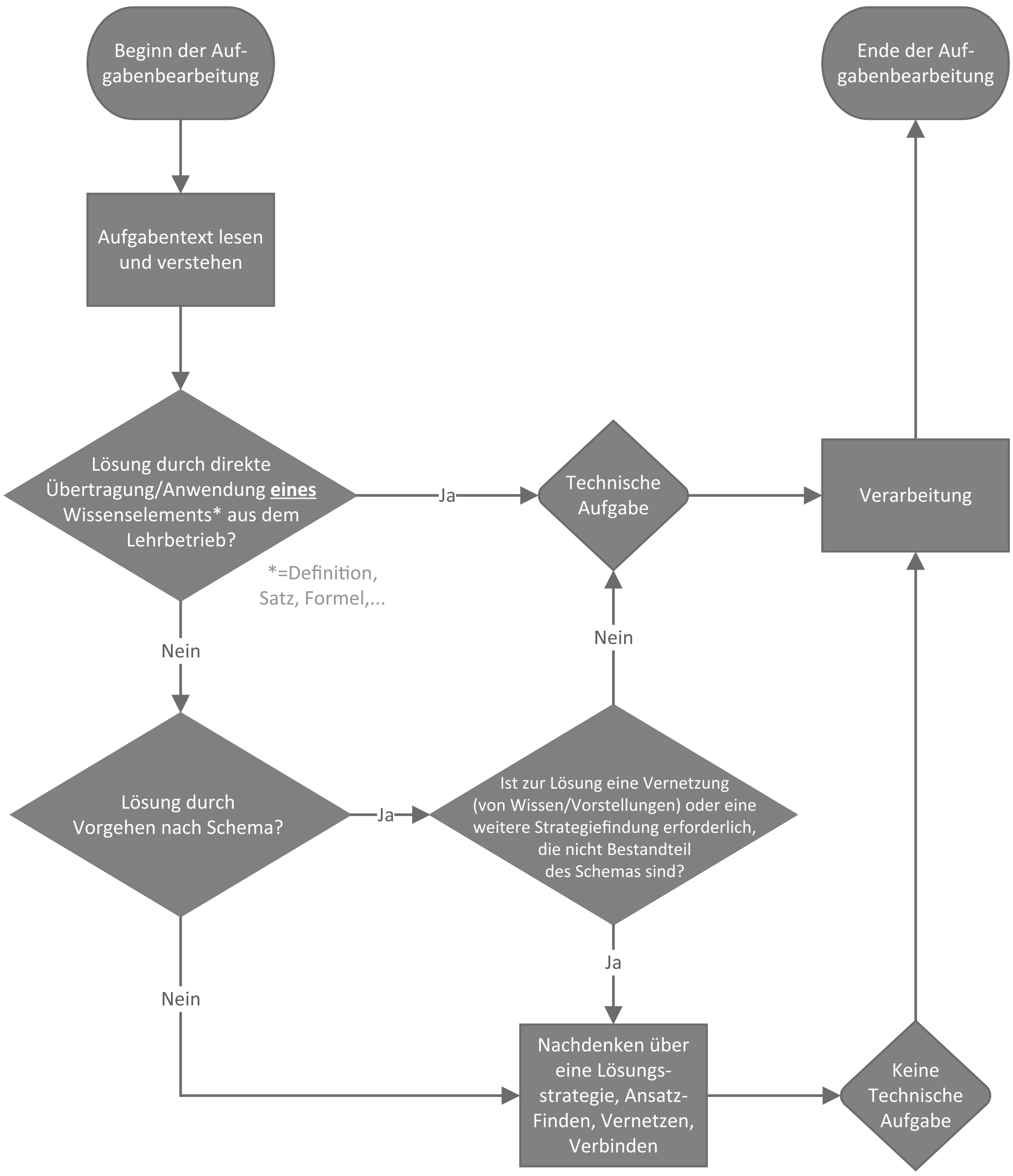


Abbildung 5.12 Ablaufschema zur Typisierung von Mathematikaufgaben als technische Aufgabe

Ingenieurmathematik (vgl. Abschnitt 1.1.1). Zur Kategorisierung wurde zunächst für jede Analyseeinheit mithilfe des in Abbildung 5.12 dargestellten, an Jordan (2006) angelehnten Ablaufschema untersucht, ob es sich um eine technische Aufgabe handelt oder nicht. Die entsprechende Kategorisierung wurde von zwei unabhängigen Kodierern vorgenommen. Die mittlere Interraterreliabilität (vgl. Abschnitt 4.5) beträgt $\kappa_n = 0.86$ und ist nach Landis und Koch (1977, S. 165) als „almost perfect“ einzustufen. Die prozentuale Beurteilerübereinstimmung (Fleiss et al., 2003, S. 599) liegt im Schnitt bei 94 %. Zur Einordnung dieses Wertes kann folgender Vergleich dienen: Jordan et al. (2008, S. 97) berichten bei der Kategorisierung von Aufgaben in die Kategorien „technische Aufgabe“, „rechnerische Aufgabe“ und „begriffliche Aufgabe“ zu drei verschiedenen Zeitpunkten und jeweils zwölf Kodiererinnen und Kodierern von einer prozentualen Übereinstimmung von 66 % bis 81 %. Die vorliegende Übereinstimmung von 94 % kann damit als vergleichsweise hoch bewertet werden.
Im Folgenden wird das Ablaufschema exemplarisch auf die fünf Klausuraufgaben in Abbildung 1.9 angewendet: Aufgabe 1 erfordert lediglich die Übertragung eines Wissenselements, nämlich einer Definition, aus der Vorlesung. Es handelt sich damit um eine technische Aufgabe. Aufgabe 2 ist mithilfe einer Basisprozedur zur Determinantenberechnung zu lösen. Somit ist auch Aufgabe 2 eine technische Aufgabe. In den Aufgaben 3 und 5 hingegen muss zunächst über eine geeignete Lösungsstrategie nachgedacht werden. Dabei sind einige Themen der Veranstaltung zu vernetzen, sodass diese beiden Aufgaben keine technischen Aufgaben darstellen. Aufgabe 4 erfordert das Anwenden einer Superprozedur, wobei keine zusätzliche Strategiefindung erforderlich ist. Aufgabe 4 ist damit eine technische Aufgabe.

Nachdem alle Analyseeinheiten entsprechend klassifiziert worden waren, wurde bei allen technischen Aufgaben anschließend unterschieden, ob diese Faktenwissen abfragen oder ob zur Lösung prozedurales Wissen erforderlich ist. Auf diese Weise konnte für jede Klausur ein Anforderungsprofil erstellt werden, aus dem der Anteil Analyseeinheiten hervorgeht, der ausschließlich prozedurales Wissen bzw. ausschließlich Faktenwissen bzw. konzeptuellen Wissen erfordert. Der Auswahl gerade dieser Merkmale für die Erstellung eines Anforderungsprofils liegt die These zugrunde, dass insbesondere der Anteil prozeduralen Wissens in Klausuren die korrekte und effiziente Anwendung von Basisprozeduren und damit die Resultate des Speed- und Leistungstests prozedurales Wissen zu einem starken Prädiktor für Klausurerfolg und Klausurleistung macht. Diese These wird im Interpretationsteil aufgegriffen und dort im Hinblick auf Forschungsfrage 6 und eine mögliche Übertragbarkeit der Resultate der ersten beiden Forschungsfragen diskutiert.

Ergebnisse

In Abbildung 5.13 sind die Ergebnisse der Klausuranalysen dargestellt. Demnach handelt es sich bei mehr als 70 % der Analyseeinheiten in allen untersuchten Klausuren um technische Aufgaben. Entsprechend stellen in allen Klausuren weniger als 30 % der Analyseeinheiten Anforderungen an konzeptuelles Wissen. Der Anteil an Analyseeinheiten, der ausschließlich prozedurales Wissen erfordert, liegt zwischen 50 % und 88 %. Ausschließlich Faktenwissen wird in 4 % bis 25 % der Analyseeinheiten abgeprüft. Nachfolgend werden diese Resultate diskutiert und interpretiert.

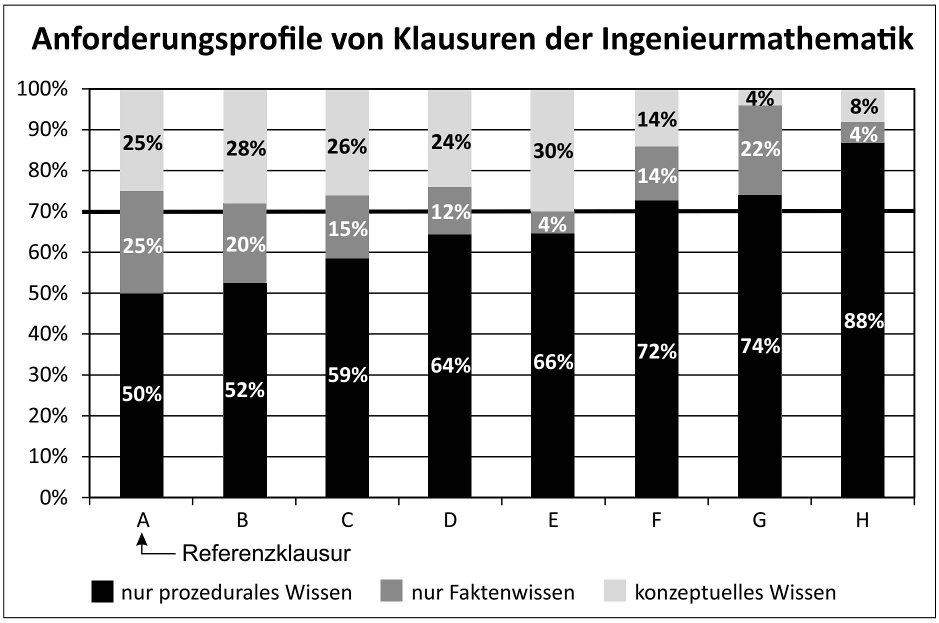


Abbildung 5.13 Anteile prozeduralen Wissens, konzeptuellen Wissens und Faktenwissens als Anforderungen in Klausuren der Ingenieurmathematik. Die Klausuren sind mit den Buchstaben A bis H kodiert, wobei A die Referenzklausur darstellt.

Interpretation der Ergebnisse in Hinblick auf Forschungsfrage 6

Die Anforderungen der Klausuren sind insofern vergleichbar, als dass in allen Klausuren mindestens 50 % der Analyseeinheiten ausschließlich Anforderungen an prozedurales Wissen und höchstens 30 % Anforderungen an konzeptuelles Wissen stellen. Damit kann bestätigt werden, dass zumindest in der vorliegenden Auswahl von Klausuren der Ingenieurmathematik Anforderungen an prozedurales Wissen dominieren. Dabei weist die den Analysen der vorliegenden Arbeit zugrunde liegende Referenzklausur den geringsten Anteil an Analyseeinheiten auf, die ausschließlich prozedurales Wissens abprüfen.
Damit wäre unter der Voraussetzung, dass allein der Anteil prozeduralen Wissens in den Klausuren über die Stärke des Zusammenhangs zwischen den Ergebnissen des Speed- und Leistungstests prozedurales Wissen und dem Klausurerfolg bzw. der Klausurleistung entscheidet, eine Übertragbarkeit der Ergebnisse zu den ersten beiden Forschungsfragen auf die analysierten sieben Klausuren naheliegend.

Um die Hypothese zu verifizieren, dass die Stärke des Zusammenhangs zwischen den Ergebnissen des Speed- und Leistungstests prozedurales Wissen und Klausurerfolg bzw. Klausurleistung durch den Anteil prozeduraler Aufgaben in der Klausur bestimmt wird, wären aufwendige Erhebungen erforderlich, die im Rahmen der vorliegenden Arbeit nicht zu leisten sind. Aus diesem Grund erfolgt nun eine kritische Diskussion der Einwände gegen diese Hypothese. Mögliche Einsprüche könnten sein:

- Wenn zwei Klausuren denselben Anteil Analyseeinheiten enthalten, die ausschließlich prozedurales Wissens einfordern, die Klausuren aber an das prozedurale Wissen verschiedene Anforderungen stellen, dann werden die Zusammenhänge des Speed- und Leistungstests prozedurales Wissen mit dem Klausurerfolg bzw. der Klausurleistung unterschiedlich sein. Im Extremfall könnte die eine Klausur nur aus einfachen Basis- und die andere Klausur nur aus komplexen Superprozeduren bestehen.
  Diesem Einwand ist entgegenzusetzen, dass Klausuren auch im prozeduralen Bereich Vorlesungswissen abprüfen, das in der Ingenieurmathematik relativ kanonisch ist und hier weder nur aus Basisprozeduren noch ausschließlich aus Superprozeduren besteht. Insofern werden Klausuren in der Regel sowohl Basis- als auch Superprozeduren abprüfen. Das Verhältnis kann hingegen durchaus variieren, weshalb der Einwand in abgeschwächter Form berechtigt bleibt.

- Zu den Zusammenhängen zwischen Klausurerfolg bzw. -leistung und den Ergebnissen des Speed- und Leistungstests prozedurales Wissen tragen andere Klausuranforderungen mehr bei als prozedurales Wissen.
  Hiergegen ist mit Blick auf die Anforderungsprofile in Abbildung 5.13 einzuwenden, dass dafür nur der Anteil an konzeptuellem Wissen in Betracht kommt, da der Anteil an Faktenwissen in den meisten Fällen geringer ist und zudem ein Zusammenhang zwischen der korrekten und effizienten Anwendung von Basisprozeduren und Faktenwissen nicht offensichtlich ist. Folglich ist das Argument in dem Fall relevant, wenn die Ergebnisse des Speed- und Leistungstests prozedurales Wissen stärker mit der Bearbeitung konzeptueller als prozeduraler Aufgaben zusammenhängen. An dieser Stelle wird davon ausgegangen, dass aufgrund der vergleichbaren Anforderungen der Zusammenhang mit prozeduralen Aufgaben stärker ist, womit das Argument insgesamt abgeschwächt werden kann.

- Die Zusammenhänge könnten von den Studierenden, deren Vorwissen und deren Geübtheit sowie von institutionellen Rahmenbedingungen abhängen.
  Dieser Einwand erscheint grundsätzlich berechtigt, beispielsweise könnte eine spezielle Klausurvorbereitung die Zusammenhänge insofern verändern, als dass prozedural schwache Studierende in der Klausur durch einen „teaching to the test“-Effekt

besser abschneiden. Allerdings würde das vermutlich auch für alle anderen Studierenden gelten. Zudem könnten Vorwissen und Geübtheit sich ebenso auf den Speed- und Leistungstest prozedurales Wissen auswirken und nicht nur auf die Klausur. Für eine Untersuchung der obigen Hypothese sollte dennoch vorausgesetzt werden, dass die Gestaltung der zu vergleichenden Veranstaltungsbetriebe nicht zu verschieden ist.

Schließlich sei noch ein Argument genannt, das für die Gültigkeit der obigen Hypothese angeführt werden kann: Der Speed- und Leistungstest prozedurales Wissen erfordert die korrekte und effiziente Anwendung von Basisprozeduren der Ingenieurmathematik. Diese werden in prozeduralen Klausuraufgaben entweder direkt abgeprüft oder sind Grundvoraussetzung für die Durchführung von Superprozeduren innerhalb dieser Aufgaben. Je höher also der Anteil prozeduraler Aufgaben ist, desto stärker kann der Einfluss von Basisprozeduren sein, was sich entsprechend auf Zusammenhänge zwischen Testergebnissen und Klausurerfolg bzw. Klausurleistung überträgt.

Im Hinblick auf Forschungsfrage 6 ist festzuhalten, dass in allen sieben Vergleichsklausuren der Anteil an Analyseeinheiten, die ausschließlich prozedurales Wissen abprüfen, höher ist als in der Referenzklausur. Aus der vorangegangenen Diskussion könnte daher geschlossen werden, dass die Ergebnisse der ersten beiden Forschungsfragen sowie die Relevanz prozeduralen Wissens zumindest tendenziell auf diese Klausuren der Ingenieurmathematik übertragbar sind.
Ob die Resultate des Klausurvergleichs sowie die diskutierten möglichen Konsequenzen auf weitere Klausuren der Ingenieurmathematik übertragbar sind, müssten mehr Klausuranalysen sowie genauere Untersuchungen der Rahmenbedingungen zeigen, was nachfolgenden Forschungsvorhaben vorbehalten bleibt. Ein Indiz für eine weitergehende, internationale Vergleichbarkeit der Klausuren liefern die Resultate von Bergqvist (2007), die eine ähnliche Untersuchung an schwedischen Hochschulen durchführte. Auch hier erwiesen sich mindestens 70 % der Klausuraufgaben als technische Aufgaben, wovon 52 % ausschließlich durch prozedurales Wissen und 17 % allein durch Faktenwissen zu lösen waren.

Fazit

In allen untersuchten Klausuren beträgt der Anteil der Analyseeinheiten, die ausschließlich prozedurales Wissen einfordern, mindestens 50 %. Damit bestätigt sich die von vielen Autorinnen und Autoren vertretene These, dass Klausuren der Ingenieurmathematik eher prozedural ausgerichtet sind (vgl. Abschnitt 1.2). Der Anteil technischer Aufgaben, die ausschließlich prozedurales Wissen oder Faktenwissen abprüfen, liegt in allen Klausuren bei über 70 %. Dies bedeutet, dass ein hoher Anteil der Klausuraufgaben mit Schema-und Faktenwissen zu lösen ist. Dieser Aspekt könnte in aktuelle Diskussionen über eine outcome-orientierte Neuausrichtung der Hochschullehre (Heinisch, Romeike & Eichler, 2016; Kultusminsterkonferenz, 2010) einfließen und auf die Frage fokussieren, ob bzw. wie sich eine stärkere Kompetenzorientierung der Lehre in Anforderungsprofilen von Klausuren ausdrücken sollte.

Die Dominanz prozeduralen Wissens in allen analysierten Klausuren stützt die Hypothese, dass tendenziell auch an anderen Hochschulen mit ähnlichen Rahmenbedingungen

- die Fertigkeit, Basisprozeduren der Ingenieurmathematik korrekt und effizient ausführen zu können, einen bedeutsamen Zusammenhang mit der Klausurleistung aufweist,
- prozedurales Wissen, operationalisiert durch obige Fertigkeit, eine hohe Prognosegüte zur Vorhersage von Klausurerfolg hat und
- Kalkülkenntnis mehr Varianz der Klausurleistung erklärt als Kalkülfertigkeit.

Diese Punkte sind zentrale Ergebnisse der ersten beiden Forschungsfragen. Ob und in welchem Maße diese tatsächlich auf andere Hochschulen übertragbar sind, sollte in nachfolgenden Forschungsvorhaben überprüft werden.

6 Fazit und Ausblick

In diesem Kapitel werden zunächst die Ergebnisse der Arbeit resümierend in die aktuelle Forschung eingeordnet sowie (methodische) Grenzen und mögliche Konsequenzen aufgezeigt (6.1). Abschließend erfolgt aufbauend auf den Resultaten ein Ausblick auf offene Fragen und resultierende Forschungsperspektiven (6.2).

6.1 Fazit

Bereits in der Einleitung wurde die übergreifende Fragestellung der vorliegenden Arbeit wie folgt formuliert:

Welche Hintergründe haben Erfolg und Misserfolg in Mathematikklausuren der ingenieurwissenschaftlichen Studiengänge?

An dieser Stelle wird nun ein Teil dieser Hintergründe durch die Ergebnisse zu den einzelnen Forschungsfragen ausgeleuchtet. Dabei ist zu betonen, dass alle Resultate *systembezogen* sind. Mit *System* ist hier insbesondere der *derzeitige* Prüfungsbetrieb in der Mathematikausbildung für Studierende ingenieurwissenschaftlicher Fächer gemeint. Im Rahmen der Untersuchungen zu Forschungsfrage 6 konnte gezeigt werden, dass dieser Prüfungsbetrieb vermutlich an den meisten deutschen Hochschulen überwiegend prozedural ausgerichtet ist. Darauf deutet die in der vorliegenden Arbeit durchgeführte Analyse von acht zufällig ausgewählten Klausuren der Ingenieurmathematik hin: In den Klausuren lassen sich zwischen 50 % und 88 % der Aufgaben allein durch prozedurales Wissen, d. h. durch die Kenntnis von Basis- und Superprozeduren, lösen, während höchstens 30 % der Aufgaben konzeptuelles Wissen erfordern. Ein ähnliches Resultat liegt für schwedische Hochschulen vor (Bergqvist, 2007). Gestützt wird dieses Ergebnis auch durch die Aussagen vieler Autorinnen und Autoren, denen zufolge Mathematik in den Ingenieurwissenschaften prozedural gelehrt, gelernt und geprüft wird (Engelbrecht et al., 2012; Binti et al., 2014; Zerr, 2009; El Gaidi & Ekholm, 2015; Aspinwall & Miller, 1997; Carpenter et al., 1983; Schoenfeld, 1989). Um dieses Resultat noch weiter abzusichern, sollten im Rahmen nachfolgender Forschungsvorhaben die Anforderungsprofile weiterer Klausuren der Ingenieurmathematik analysiert werden. Hier könnten auch, wie etwa bei Lithner (2006, 2004), weitere wichtige prozessbezogene Anforderungen wie Problemlösen oder Beweisen

berücksichtigt werden, was außerhalb des Fokus der vorliegenden Arbeit lag und daher hier noch nicht berücksichtigt wurde.
Die zuvor geschilderten Resultate sind wichtig, da sie auf ein relativ starkes Ungleichgewicht hinsichtlich der Anforderungen an prozedurales und konzeptuelles Wissen im aktuellen Hochschulsystem der Ingenieurmathematik hinweisen. Hier scheinen Anforderungen an prozedurales Wissen zu dominieren. Besonders bemerkenswert ist daher, dass viele Innovationsprojekte ihre hochschuldidaktische Aufmerksamkeit auf konzeptuelles Wissen legen, wodurch ein weiteres Ungleichgewicht, allerdings diesmal zugunsten (der Erforschung) konzeptuellen Wissens, entstanden ist (vgl. Abschnitt 1.2.1). Durch die Schwerpunktsetzung auf prozedurales Wissen trägt die vorliegende Arbeit daher zu einer Verbesserung der Balance in der Forschung über prozedurales und konzeptuelles Wissen in der Ingenieurmathematik bei. Da die Ergebnisse der vorliegenden Arbeit innerhalb des aktuellen Hochschulsystems der Ingenieurmathematik generiert wurden, erstreckt sich ihr Gültigkeitsanspruch zunächst auch nur auf den Status quo dieses Systems.

Ein Hauptergebnis der vorliegenden Arbeit ist weiter, dass im aktuellen Hochschulsystem der Ingenieurmathematik ein solides prozedurales Wissen nahezu ein Garant für das erfolgreiche Absolvieren einer Klausur zu sein scheint. Darauf weisen folgende Daten hin: In rund 75 % der Fälle sagt das prozedurale Wissen eines bzw. einer Studierenden den Klausurerfolg korrekt voraus. Dieses Resultat gilt zwar zunächst nur im Mikrokosmos der analysierten Referenzklausur, im Rahmen der Untersuchungen von Forschungsfrage 6 konnte die Vergleichbarkeit dieser Klausur mit anderen Klausuren der Ingenieurmathematik aber bereits angedeutet werden. Ferner wurde die Prognosegüte durch *out of sample*-Prognosen bestimmt, sodass mit keinem zu starken Overfit zu rechnen ist. Unabhängig davon müsste dieses Ergebnis noch an Hand anderer Klausuren der Ingenieurmathematik bestätigt werden, was in der vorliegenden Arbeit nicht geleistet werden konnte. Letzteres gilt auch für eine Bestätigung der erklärten Varianz der Klausurleistung. Prozedurales Wissen alleine klärt hier bereits bis zu 47 % der Varianz auf und erzielt unter allen untersuchten Prädiktoren die höchste Produkt-Moment-Korrelation mit der Klausurleistung ($r = .63$). Die folgenden Resultate aus vergleichbaren Studien dienen zur Einordnung dieser Werte: Faulkner et al. (2014) gelingt es, auf Grundlage der schulischen Vorbenotung und eines Eingangstests in 71.3 % der Fälle den Klausurerfolg korrekt vorherzusagen. Denny et al. (2012) erklären mit einer Kombination aus Einstufungstest und weiteren Instrumenten 32 % der Varianz in den Ergebnissen einer „Calculus"-Klausur. Hell et al. (2008) berichtet eine Korrelation von $r = .53$ zwischen einem fachspezifischen Studierfähigkeitstest und der Klausur Mathematik 1 für Maschinenbau-Studierende. In Hinblick auf die übergreifende Fragestellung der vorliegenden Arbeit beleuchten diese Zusammenhänge die Prognosekraft prozeduralen Wissens für die Vorhersage von Klausurerfolg und -leistung im aktuellen Hochschulsystem der Ingenieurmathematik. Es kann festgehalten werden, dass prozedurales Wissen hier ein vergleichsweise starker Prädiktor ist.

An dieser Stelle könnte jedoch auch kritisch angemerkt werden, dass der starke Zusammenhang zwischen dem Klausurergebnis und prozeduralem Wissen zu erwarten ist, da dieses aktuell schwerpunktmäßig abgeprüft wird. Dagegen ist einzuwenden, dass in der vorliegenden Arbeit prozedurales Wissen durch das Anwenden von *Basisprozeduren* operationalisiert wurde. Diese bilden nachweislich lediglich einen kleinen Teil des weitaus komplexeren Anforderungsspektrums von Klausuren der Ingenieurmathematik. Insofern ist es nicht selbstverständlich, dass prozedurales Wissen in der hier operationalisierten Weise einen so guten Prädiktor darstellt. Daran anknüpfend zeigte sich als weiteres wichtiges Resultat in den Untersuchungen zu Forschungsfrage 2, dass Kalkül*kenntnis* (operationalisiert durch eine Kenntnis von Basisprozeduren) etwa einen doppelt so großen Einfluss auf die Klausurleistung hat wie Kalkül*fertigkeit* (operationalisiert durch Fertigkeit bei der Durchführung einer Basisprozedur). Eine mögliche Erklärung hierfür besteht darin, dass unzureichende Kalkülkenntnis dazu führt, dass komplexe mathematische Verfahren und Superprozeduren, die in Klausuren abgeprüft werden, nicht erfolgreich durchgeführt werden können, da die Kenntnis von Basisprozeduren hierfür eine wichtige Voraussetzung ist. Durch mangelhafte Kalkülkenntnis können also möglicherweise gleich mehrere Aufgabenteile, welche die Kenntnis von Basisprozeduren als Grundanforderung beinhalten, nicht gelöst werden. Eine unzureichende Kalkülfertigkeit wie beispielsweise ein Rechenfehler wirkt sich hingegen nur isoliert auf die betroffene Aufgabe aus. Mit Blick auf die übergreifende Fragestellung der Arbeit nach den Hintergründen für Erfolg und Misserfolg in Klausuren lässt sich damit zusammenfassend sagen, dass für einen erfolgreichen Klausurabschluss in der Ingenieurmathematik beide Facetten prozeduralen Wissens wesentlich sind, wobei Kalkülkenntnis wichtiger als Kalkülfertigkeit zu sein scheint. Ferner konnte das Anwenden von Basisprozeduren als prädiktiv isolierte Grundanforderung im Anforderungsprofil von Klausuren der Ingenieurmathematik identifiziert werden. Diese Resultate sind wichtig, da sie Ansatzpunkte für adaptive Lernangebote darstellen, worauf im nächsten Abschnitt eingegangen wird.

Wie bereits die vorangehenden Ausführungen zeigen, ist die Unterscheidung von Basisprozeduren und Superprozeduren sowie Kalkülkenntnis und Kalkülfertigkeit wichtig, um Fehlerquellen bei der Bearbeitung prozeduraler Aufgaben sowie unterschiedliche Anforderungen an prozedurales Wissen präziser beschreiben zu können. Um diese Unterscheidungen vornehmen zu können, war es erforderlich, die Theoriebildung zu prozeduralem Wissen zu erweitern. Dazu wurden im Theorieteil (vgl. Kapitel 1) bestehende, aus der Literatur bekannte Definitionen von prozeduralem Wissen ausgeschärft und ergänzt. So wurde prozedurales Wissen in der vorliegenden Arbeit als Verknüpfung von Kalkülkenntnis und

Kalkülfertigkeit definiert, wobei Basisprozeduren und Superprozeduren die beiden Pole des Anforderungsspektrums prozeduralen Wissens repräsentieren. Durch diese Unterscheidungen war es in Forschungsfrage 5 möglich, die zeitliche Veränderung von Ressourcen und Schwächen in der Kalkülkenntnis und -fertigkeit exemplarisch an einigen Studierendenbearbeitungen aufzuzeigen. Ein Resultat der Untersuchung war hier, dass Schwächen individuell verschieden sind und sowohl im Bereich der Kalkülkenntnis als auch in der Kalkülfertigkeit über mehrere Semester hinweg bestehen bleiben können. Dies ist eine wichtige Beobachtung, weil dadurch möglicherweise ein erfolgreiches Verwenden der entsprechenden Basisprozeduren in Klausuren dauerhaft verhindert wird. Das wiederum führt zu einer schlechteren Erfolgsprognose für die anstehende Klausur, was in den Forschungsfragen 1 und 2 gezeigt wurde. Dieser Zusammenhang gilt auch für eine weitere Beobachtung, dass prozedurales Wissen mit der Zeit nicht mehr in dem erforderlichen Maße abrufbar sein kann und damit nicht mehr Teil des prozeduralen Netzwerks der oder des Studierenden ist (vgl. hierzu auch Engelbrecht et al., 2007).
Für die übergreifende Fragestellung nach den Hintergründen für Klausurerfolg und -misserfolg ist damit festzuhalten, dass die in der vorliegenden Arbeit eingeführt Verfeinerung der Definition prozeduralen Wissens eine bessere Gewichtung und Ausleuchtung möglicher Ursachen von Fehlern bei der Bearbeitung prozeduraler Aufgaben in (Klausuren) der Ingenieurmathematik ermöglicht. Für das aktuelle Hochschulsystem der Ingenieurmathematik sind diese Erkenntnisse wichtig, da hier häufig weit über 50 % der Klausuraufgaben ausschließlich Anforderungen an prozedurales Wissen stellen. An diese Stelle knüpft daher unmittelbar die Forderung nach einer noch differenzierteren Analyse der Fehlerquellen an, beispielsweise im Bereich der Kalkülfertigkeit (vgl. Kersten, 2015). Dies konnte im Rahmen der vorliegenden Arbeit allerdings nur exemplarisch und für den Fall prozedural schwacher Studierender geleistet werden und wird daher im nachfolgenden Ausblick als Forschungsperspektive aufgegriffen.

Während in Forschungsfrage 5 schwerpunktmäßig die Entwicklung von Ressourcen und Schwächen im prozeduralen Wissen analysiert wurde, beleuchtete Forschungsfrage 4 die zeitliche Entwicklung des prozeduralen Wissens bei Studierenden im Studium. Dieses wurde hier durch die erfolgreiche Anwendung von Basisprozeduren im Speed- und Leistungstest prozedurales Wissen operationalisiert. Zunächst wurden die teilnehmenden Studierenden in Abhängigkeit von ihrem prozeduralen Wissen im ersten Studiensemester in drei Leistungsgruppen eingeteilt. Anschließend wurde die Entwicklung ihres prozeduralen Wissens über einen Zeitraum von bis zu drei Semestern dokumentiert. Dabei stellte sich heraus, dass die Leistung der gemäß den Kriterien der vorliegenden Arbeit als prozedural stark eingestuften Studierenden in diesem Zeitraum kaum zunimmt. Dies bedeutet, dass sich die prozedural starken Studierenden bereits auf einem sehr hohen Ausgangsniveau

befinden. Die Leistung der prozedural durchschnittlichen Studierenden steigt hingegen im Verlauf von drei Semestern moderat, aber kontinuierlich. Am auffälligsten erscheint dagegen die Leistungsentwicklung der prozedural schwachen Studierenden: Deren Leistung nimmt zunächst vom ersten zum zweiten Semester stark zu und *sinkt* anschließend vom zweiten zum dritten Semester. Dieses paradoxe Resultat stimmt zum Teil mit Ergebnissen von Pekrun et al. (2006) sowie Henn und Polaczek (2007) überein, wonach die Leistungsfähigkeit schwächerer Schülerinnen und Schüler bzw. Studierender im Verlauf der Zeit *abnimmt*. Hier erfolgt die Leistungsabnahme allerdings kontinuierlich und nicht – wie im vorliegenden Fall – nach einem Leistungsanstieg. Damit bleibt die Frage offen, warum der positive Trend im Leistungszuwachs nicht im dritten Semester fortgesetzt werden kann. Dabei ist zu beachten, dass prozedural schwache Studierende trotz eines starken Leistungsanstiegs vom ersten zum zweiten Semester dennoch kaum das Ausgangsniveau der prozedural durchschnittlichen Studierenden erreichen. Die meisten prozedural schwachen Studierenden bleiben damit trotz gewisser Verbesserungen prozedural schwach. Dieses Resultat ist auch relevant für die Erforschung des „student life cycle", denn es betont die Relevanz des ersten Studienjahres für die weitere Leistungsentwicklung von Studierenden, was auch durch andere Studien gestützt wird (Ortiz & Dehon, 2013; Henn & Polaczek, 2007; Kolb et al., 2006).
Die genannten Resultate beleuchten den Leistungsverlauf verschieden leistungsstarker Studierendengruppen. Für die übergreifende Forschungsfrage nach den Hintergründen für Erfolg und Misserfolg in Klausuren der Ingenieurmathematik ist insbesondere festzuhalten, dass das prozedurale Wissen prozedural schwacher Studierender auf einem vergleichweise niedrigen Niveau bleibt und im zweiten Studienjahr sogar abnimmt. Diese Entwicklung der prozedural schwachen Studierenden wirft weiterführende Fragen auf, die allerdings in der vorliegenden Arbeit nicht geklärt werden konnten. Dazu zählt beispielsweise die für die Diskussion von Dropouts und Durchfallquoten wichtige Frage, ob sich ein ähnlicher Leistungsverlauf auch in anderen Fächern beobachten lässt und was genau die Ursachen für die Trendumkehr in der Leistungsentwicklung nach dem zweiten Semester sind.

In der vorangegangenen Diskussion wurde festgestellt, dass die Gruppe der prozedural schwachen Studierenden einige bemerkenswerte Eigenschaften aufweist, die weiteren Diskussionsbedarf einfordern. Im Rahmen der Untersuchungen von Forschungsfrage 3 konnten bezüglich dieser Studierendengruppe noch weitere Details aufgedeckt werden. Zum einen stellte sich heraus, dass prozedural schwache Studierende auch in anderen Bereichen als prozeduralem Wissen wie etwa bei konzeptuellem Wissen, Klausurleistung, Konzentration usw. schwächer als Studierende der anderen Leistungsgruppen sind. Dieses Resultat ist wichtig, denn es bedeutet, dass leistungsschwache Studierende vermutlich in *mehreren* Bereichen einen erhöhten Unterstützungsbedarf aufweisen. Zum anderen gibt es Gruppen

von Studierenden, die überproportional häufig unter den prozedural schwachen Studierenden zu finden sind. Hierzu zählen Studierende des Studiengangs Bauingenieurwesen, Studierende mit einem Grundkurs in Mathematik in der Schule und Absolventinnen und Absolventen einer Gesamtschule. Diese Ergebnisse stimmen überein mit Untersuchungen von Heublein et al. (2014), Heublein et al. (2010), Trautwein et al. (2006) und Baumert (1996). Umgekehrt gibt es auch Studierendengruppen, die unter den prozedural starken Studierenden überproportional häufig anzutreffen sind. Hierzu zählen Studierende des Studiengangs Wirtschaftsingenieurwesen, Studierende, die in der Schule einen Leistungskurs in Mathematik besuchten, sowie Studierende, die die Veranstaltung wiederholen. Diese Resultate deuten teilweise darauf hin, dass Studierende umso leistungsstärker sind, je länger sie sich mit Mathematik beschäftigt haben. Diese Folgerung wird durch andere Studien gestützt (z. B. Perdigones et al., 2014; Büning, 2004). Dabei handelt sich um eine sehr wichtige Schlussfolgerung, da sie die *Lernzeit* als möglichen Beitrag zu Klausurerfolg thematisiert. Offenbar wird Lernzeit teilweise dadurch aufgeholt, indem Semester und Klausuren einmal oder öfter wiederholt werden. Dies stellt einen für Studierende vermutlich sehr frustrierenden Weg dar, Lernzeit aufzuholen.
Von den obigen Resultaten konnte im Rahmen der vorliegenden Arbeit nicht aufgeklärt werden, warum Studierende des Studiengangs Bauingenieurwesen besonders häufig zu den prozedural schwachen Studierenden gehören. Dieser Frage müssen daher andere Forschungsvorhaben nachgehen. Des weiteren fand sich im Rahmen der Untersuchungen von Forschungsfrage 3 noch eine Abweichung gegenüber Ergebnissen aus dem Schulbereich: Während Schülerinnen und Schüler sich bezüglich ihrer Mathematikleistung signifikant unterscheiden (Prenzel et al., 2007, 2013), weisen männliche und weibliche Studierende nach Selbstselektion durch die Studienfachwahl keinen signifikanten Unterschied in ihrer Leistung bezüglich prozeduralen Wissens auf.
Im Hinblick auf die übergreifende Forschungsfrage der vorliegenden Arbeit beleuchten die vorgenannten Resultate insbesondere die Konturen der Gruppe der prozedural schwachen Studierenden. Daraus können im nachfolgenden Abschnitt wichtige Konsequenzen für selektive Lernangebote und -formate (vgl. z. B. Altieri et al., 2013; Fakultät für Mathematik der TU Dortmund, 2013) und auch für eine weitergehende Theoriebildung für prozedurales Wissens abgeleitet werden.

An dieser Stelle soll nochmals auf die ersten beiden Forschungsfragen zurückgekommen werden. Wie zu Beginn erwähnt, stellte sich hier prozedurales Wissen als stärkster Prädiktor für Klausurerfolg und -leistung dar, gefolgt von konzeptuellem Wissen und den beiden allgemein kognitiven Faktoren Konzentration und kognitive Grundfertigkeit. Den geringsten Zusammenhang mit der Klausurleistung weisen die affektiven Faktoren Prüfungsangst und Prokrastination auf. Die berichteten Zusammenhänge stimmten, sofern Vergleichswerte existierten, erwartungskonform mit Ergebnissen aus der Literatur überein. Dennoch ist nicht auszuschließen, dass es weitere Prädiktoren gibt, die in der vorliegenden Arbeit nicht untersucht wurden und die einen noch besseren Prädiktor für Klausurerfolg und -leistung im aktuellen Hochschulsystem der Ingenieurmathematik darstellen. Auch könnte eine Verbesserung der Tests den deutlichen Vorsprung prozeduralen Wissens hinsichtlich der Prognosegüte verkleinern. Dies deutete sich beispielsweise bei der Messung der kognitiven Grundfertigkeit an. Hier passte das zugrunde liegende Erhebungsinstrument, der Wiener Matrizen-Test 2, nicht optimal zu den Teilnehmenden, da der Test im oberen Leistungsbereich zu wenig differenzierte. Umgekehrt könnte eine weitere Verbesserung des Speed- und Leistungstests prozedurales Wissen, beispielsweise auf Grundlage einer weitergehenden Theoriebildung für prozedurales Wissen, zu einer noch höheren Prognosegüte prozeduralen Wissens führen.

Im Hinblick auf die übergreifende Fragestellung können damit die zentralen Ergebnisse der vorliegenden Arbeit wie folgt zusammengefasst werden: Im aktuellen Hochschulsystem der Ingenieurmathematik gilt prozedurales Wissen als stärkster Prädiktor für Klausurerfolg und -leistung. Prozedurales Wissen wurde operationalisiert durch das Anwenden von Basisprozeduren, die allerdings nur einen kleinen Teil des Anforderungsprofils von Klausuren der Ingenieurmathematik abdecken. Die Kenntnis von Basisprozeduren erwies sich als wichtiger als die Fertigkeit in der Ausführung von Basisprozeduren („Kenntnis vor Fertigkeit“). Schwächen in der Kalkülkenntnis und Kalkülfertigkeit können zeitlich stabil sein, was Klausurerfolg dauerhaft verhindern kann. Das prozedurale Wissen prozedural schwacher Studierender nimmt im Gegensatz zu den anderen Leistungsgruppen nach dem zweiten Semester ab. Zu der Gruppe der prozedural schwachen Studierenden zählten überproportional häufig Studierende des Studiengangs Bauingenieurwesen und Studierende, die einen Grundkurs in Mathematik besuchten, sowie Absolventinnen und Absolventen einer Gesamtschule. In der Gruppe der prozedural starken Studierenden waren überproportional häufig Studierende des Studiengangs Wirtschaftsingenieurwesen, Studierende mit einem Leistungskurs Mathematik sowie Studierende, welche die Veranstaltung wiederholen, vertreten. Die Resultate weisen darauf hin, dass Studierende umso erfolgreicher sein können, je mehr Lernzeit sie effektiv in Mathematik investiert haben. Die höchste Prognosegüte für Klausurerfolg und -leistung wiesen fachspezifische Faktoren auf, gefolgt von den allgemein kognitiven Faktoren. Affektive Faktoren spielten hier kaum eine Rolle.

Aus diesen Ergebnissen lassen sich folgende Punkte für eine spätere Weiterentwicklung der Lehre im aktuellen Hochschulsystem der Ingenieurmathematik festhalten:

- Leistungsschwache Studierende mit einer entsprechend geringen Erfolgswahrscheinlichkeit in der anstehenden Klausur müssen frühzeitig identifiziert werden, um Unterstützung rechtzeitig anbieten zu können.

- Unterstützungsangebote müssen zielgruppenspezifisch so früh wie möglich erfolgen.
- Kalkülkenntnis und Kalkülfertigkeit sind aufgrund der Heterogenität des prozeduralen Wissens adaptiv zu fördern, wobei gilt: „Kalkülkenntnis vor Kalkülfertigkeit".

Diese Ergebnisse sollten jetzt allerdings nicht leichtfertig auf die Lehre übertragen werden, da – um nur einen Grund zu nennen – beispielsweise nach wie vor unklar ist, *wie* Kalkülkenntnis und Kalkülfertigkeit sinnvoll zu fördern sind. Vielmehr sind anknüpfend an diese Resultate noch weitere Fragen zu klären, bevor eine theoriegeleitete und didaktisch fundierte Veränderung der Lehre zum Nutzen der Studierenden stattfinden kann. Einige Aspekte, die hierzu thematisiert werden müssten, werden im folgenden Abschnitt beschrieben.

6.2 Ausblick

Am Ende dieser Arbeit bleiben mehr Fragen als Antworten. Das ist allerdings positiv zu bewerten, denn jede Frage eröffnet Forschungsperspektiven und deutet darauf hin, dass sich die Lehre in der Ingenieurmathematik in einem dynamischen Prozess befindet und weiter verbessern lässt. Offene Fragen lauten beispielsweise:

- Weist prozedurales Wissen in anderen Klausuren der Ingenieurmathematik einen ebenso starken Zusammenhang mit der Klausurleistung auf wie bei der Referenzklausur?
- Wie erfolgreich sind Studierende mit zeitlich stabilen Fehlermustern im prozeduralen Wissen in Mathematikklausuren und Prüfungen anderer Fächer?
- Warum erfolgt eine Trendumkehr in der Leistungsentwicklung prozedural schwacher Studierender nach dem zweiten Semester?
- Warum gehören Studierende des Studiengangs Bauingenieurwesen überproportional häufig zu der Gruppe der prozedural schwachen Studierenden?

Während all diese Punkte *lokale* Forschungsbereiche eröffnen, erscheint es sinnvoll, sich zunächst auf *globale* Anknüpfpunkte an die vorliegende Arbeit zu konzentrieren. Im Folgenden werden davon exemplarisch zwei aufgezeigt, die jeweils einen anderen Schwerpunkt bilden.

Der erste Anknüpfpunkt betrifft die Weiterführung der Theoriebildung im Bereich der Kalkülkenntnis. Letztere stellte sich im Hinblick auf die Klausurleistung als eine noch wichtigere Facette von prozeduralem Wissen heraus als Kalkülfertigkeit. Insofern erscheint

es im Hinblick auf ein erfolgreiches Absolvieren einer Klausur im aktuellen Hochschulsystem der Ingenieurmathematik relevant, zu erforschen, wie sich Kalkülkenntnis festigen und verstetigen lässt. Gemäß der in der vorliegenden Arbeit getroffenen Definition von Kalkülkenntnis beinhaltet diese insbesondere die Kenntnis von Prozeduren. Zu den Prozeduren zählen in der Ingenieurmathematik sowohl einfache Basisprozeduren als auch die weitaus komplexeren Superprozeduren (vgl. Abschnitt 1.1.1). Darauf bezogen lautet obige Frage, wie sich die Kenntnis von Basis- und Superprozeduren nachhaltig unterstützen lässt. Die Diskussion in Abschnitt 1.1.3 deutete an, dass dies insbesondere durch eine stärkere Verknüpfung von prozeduralem und konzeptuellem Wissen möglich ist. Wie das Beispiel einer Superprozedur in Abbildung 1.8 zeigte, sind diese Verknüpfungen in der Ingenieurmathematik ebenso zahlreich wie Basis- und Superprozeduren selbst. Gemäß der in Abschnitt 1.1.1 und 1.1.2 eingeführten Terminologie geht es also um eine möglichst reichhaltige Verknüpfung des prozeduralen mit dem konzeptuellen Netzwerk von Studierenden und damit auch um die fachdidaktisch relevante Frage, wie diese Verknüpfungen am besten gelehrt und implementiert werden können. Bevor diese Fragestellung allerdings in der Breite bearbeitet werden kann, muss zunächst eine tiefgehende Exploration von Prozeduren der Ingenieurmathematik sowie deren konzeptuellen Anker erfolgen. Abbildung 1.8 stellt hierfür lediglich ein erstes, noch rudimentäres Beispiel dar. Erst nach dieser Exploration von Prozeduren und ihrer konzeptuellen Anker kann theoriegeleitet die Entwicklung eines geeigneten Interventionsdesigns erfolgen und dieses mit einem fachdidaktischen Fundament versehen werden.
Der hier vorgestellte erste Ansatzpunkt eröffnet somit die Perspektive für ein Entwicklungsforschungsprojekt in der Hochschuldidaktik an der Schnittstelle zwischen prozeduralem und konzeptuellem Wissen auf Grundlage der Erkenntnisse und theoretischen Ausarbeitungen der vorliegenden Arbeit.

Der zweite Anknüpfpunkt zielt auf die Entwicklung und Verfeinerung von Diagnoseinstrumenten für prozedurales Wissen bei Studierenden ab. Prozedurales Wissen ist in der vorliegenden Arbeit als Verknüpfung von Kalkülkenntnis und Kalkülfertigkeit definiert. Kalkülfertigkeit ist in der Schulmathematik insbesondere im Hinblick auf Diagnose und Förderung bereits umfangreich erforscht worden. In der Hochschulmathematik hingegen existieren hierzu wenige Untersuchungen aus fachdidaktischer Perspektive, obwohl Studien zeigen, dass Studierende die gleichen Probleme in der Kalkülfertigkeit aufweisen können wie Schülerinnen und Schüler (Kersten, 2015). In der vorliegenden Arbeit konnte zudem gezeigt werden, dass Kalkülfertigkeit und Klausurleistung in der Ingenieurmathematik zusammenhängen. Aus diesen Gründen erscheint es sinnvoll, Kalkülfertigkeit noch genauer zu untersuchen. Dazu ist es zunächst erforderlich, Kalkülfertigkeit präziser zu beschreiben. Ansatzpunkte hierfür liefern die Arbeiten von Star et al. (2015), Hatano und Inagaki

(1984), Kilpatrick, Swafford und Findell (2001) sowie Kersten (2015). Sobald die verschiedenen Facetten von Kalkülfertigkeit benannt und beschrieben sind, kann spezifiziert werden, was im Detail aus fachdidaktischer Sicht und im Hinblick auf die hochschuldidaktische oder institutionelle Zielsetzung ein entsprechendes Diagnoseelement leisten soll. Aufgrund der Ergebnisse der vorliegenden Arbeit ist bei Studierenden mit einer heterogenen Verteilung der Facetten von Kalkülfertigkeit zu rechnen. Eine Diagnose von Kalkülfertigkeit nach dem oben beschriebenen Ansatz bildet daher gleichzeitig den Ausgangspunkt für die Entwicklung heterogenitätssensitiver, d. h. adaptiver Förderinstrumente, die auf der reichhaltigen Erfahrung der Fachdidaktik in dem entsprechenden Bereich der Schulmathematik aufsetzen könnten.
Auf eine ähnliche Art und Weise könnte auch eine Diagnose von Kalkülkenntnis erfolgen. Kalkülkenntnis steht wie bereits geschildert einerseits im Zusammenhang mit der Klausurleistung, andererseits kann die Kenntnis von Basis- und Superprozeduren mit der Zeit in Vergessenheit geraten (Engelbrecht et al., 2007), obwohl diese in nachfolgenden Lehrveranstaltungen noch bzw. wieder benötigt werden. Hier ist eine Art adaptives Kontinuitätstraining denkbar, bei dem die Gestaltung der Förderkomponenten auf den oben genannten ersten Anknüpfpunkt zur Theoriebildung von Kalkülkenntnis zugreift. Damit schließt sich insgesamt ein möglicher Entwicklungskreislauf von Diagnose- und Förderinstrumenten für prozedurales und konzeptuelles Wissen.

Bereits eingangs wurde erwähnt, dass die Resultate der vorliegenden Arbeit an das aktuelle Hochschulsystem der Ingenieurmathematik gebunden sind. Die vorangehenden Ansatzpunkte für Veränderungen der Lehre beschreiben in der Terminologie der Systemtheorie allesamt einen sogenannten *Wandel erster Ordnung* (Levy & Merry, 1986, S. 6). Hierdurch werden inkrementelle Veränderungen innerhalb eines bestehenden Systems beschrieben, die weder eine Veränderung der grundsätzlichen Ausrichtung noch einen Paradigmenwechsel nach sich ziehen. Eine grundsätzliche Systemveränderung, beispielsweise durch eine deutlich veränderte Gewichtung von prozeduralem und konzeptuellem Wissen, wie sie einige Autorinnen und Autoren vorschlagen oder sogar fordern (Dreyfus & Eisenberg, 1990; Drüke-Noe, 2014; Chappell & Killpatrick, 2003), wird hingegen als *Wandel zweiter Ordnung* bezeichnet. Aufgrund der in der Einleitung aufgezeigten hohen Hürde, die die Ingenieurmathematik aktuell für viele Studierende darstellt, wird daher am Ende dieser Arbeit zunächst für einen fachdidaktisch abgesicherten Wandel erster Ordnung innerhalb des aktuellen Hochschulsystems der Ingenieurmathematik plädiert, der von einer weiteren Öffnung für Diskussionen über Veränderungen zweiter Ordnung begleitet werden sollte. So erscheint eine langfristige und nachhaltige Verbesserung der Studienleistung möglich, die sowohl von der Hochschuldidaktik als auch von den Studierenden angestrebt werden sollte.

Literatur

Altieri, M. (2014). *Testdemonstration SLTPW (frühere Bezeichnung: LTKR).* Zugriff am 18.7.2015 auf http://youtu.be/zeHWB21ccpA/

Altieri, M., Oecking, C., Paluch, D., Pomwap, C., Squiccimarro, F. & Wunde, N. (2013, Juli). *Analyse der „Lernumgebung Mathematik" für Studierende der INT–Fächer in der Studieneingangsphase* (Bericht). Fakultät für Mathematik, TU Dortmund. (Ergebnisberichte des Instituts für Angewandte Mathematik, Nummer 475)

Altieri, M. & Prediger, S. (voraussichtlich 2016). Unpacking procedural knowledge in mathematics exams for first-year engineering students. In *KHDM-Report.* Kompetenzzentrum Hochschuldidaktik Mathematik.

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Angoff, W. H. (1984). *Scales, norms, and equivalent scores.* Educational testing service.

APA. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5).* American Psychiatric Association.

Arslan, S. (2010). Traditional instruction of differential equations and conceptual learning. *Teaching Mathematics and its Applications*, *29* (2), 94–107.

Aspinwall, L. & Miller, L. (1997). Students' positive reliance on writing as a process to learn first semester calculus. *Journal of Instructional Psychology*, *24*, 253–261.

Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2011). *Multivariate Analysemethoden.* Berlin: Springer.

Bahrick, H. P., Hall, L. K. & Berger, S. A. (1996). Accuracy and distortion in memory for high school grades. *Psychological Science*, *7* (5), 265–271.

Barnett, M. D., Sonnert, G. & Sadler, P. M. (2014). Productive and ineffective efforts: how student effort in high school mathematics relates to college calculus success. *International Journal of Mathematical Education in Science and Technology*, *45* (7), 996–1020.

Baroody, A. J., Feil, Y. & Johnson, A. R. (2007). An alternative reconceptualization of procedural and conceptual knowledge. *Journal for Research in Mathematics Education*, 115–131.

Baumert, J. (1996). *Bildungsverläufe und psychosoziale Entwicklung im Jugendalter (BIJU): 2. Bericht für die Schulen.* Berlin: Max-Planck-Institut für Bildungsforschung.

Bebermeier, S. & Nußbeck, F. W. (2014). Heterogenität der Studienanfänger/innen und Nutzung von Unterstützungsmaßnahmen. *Zeitschrift für Hochschulentwicklung* (5), 83–100.

Beekhoven, S., De Jong, U. & Van Hout, H. (2002). Explaining academic progress via combining concepts of integration theory and rational choice theory. *Research in*

Higher Education, *43* (5), 577–600.

Bergqvist, E. (2007). Types of reasoning required in university exams in mathematics. *The Journal of Mathematical Behavior*, *26* (4), 348–370.

Bergsten, C., Engelbrecht, J. & Kågesten, O. (2015). Conceptual or procedural mathematics for engineering students – views of two qualified engineers from two countries. *International Journal of Mathematical Education in Science and Technology*, *46* (7), 979–990.

Binti, I. Z., Zeynivandnezhad, F., Binti, M. Y. & David, E. (2014). Computing in Differential Equations with Mathematical Thinking Approach among Engineering Students. In *Teaching and Learning in Computing and Engineering (LaTiCE), 2014 International Conference* (S. 163–170).

Blankstein, K. R., Toner, B. B. & Flett, G. L. (1989). Test anxiety and the contents of consciousness: Thought-listing and endorsement measures. *Journal of Research in Personality*, *23* (3), 269–286.

Blum, W. & Leiß, D. (2005). Modellieren im Unterricht mit der Tanken-Aufgabe. *mathematik lehren*, *128*, 18–21.

Boesen, J. (2006). *Assessing mathematical creativity: comparing national and teacher-made tests, explaining differences and examining impact* (Unveröffentlichte Dissertation). Matematik och matematisk statistik, Umeå universitet.

Borenstein, M. (2009). Effect sizes for continuous data. In H. Cooper, L. Hedges & J. Valentine (Hrsg.), *The handbook of research synthesis and meta-analysis* (S. 221–237). New York: Russell Sage Foundation.

Bortz, J. & Weber, R. (2005). *Statistik für Human-und Sozialwissenschaftler* (6. Aufl.). Springer Medizin.

Bosse, M. J. & Bahr, D. L. (2008, November). The state of balance between procedural knowledge and conceptual understanding in mathematics teacher education. *International Journal for Mathematics Teaching and Learning*, 1–28.

Brennan, R. L. & Prediger, D. J. (1981). Coefficient kappa: Some uses, misuses, and alternatives. *Educational and psychological measurement*, *41* (3), 687–699.

Bridgeman, B., McCamley-Jenkins, L. & Ervin, N. (2000). Predictions of Freshman Grade-Point Average From the Revised and Recentered SAT I: Reasoning Test. *ETS Research Report Series* (1), 1–16.

Briggs, S. R. & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality*, *54* (1), 106–148.

Brown, J. S., Collins, A. & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, *18* (1), 32–42.

Bruch, M. A. (1981). Relationship of test-taking strategies to test anxiety and performance: Toward a task analysis of examination behavior. *Cognitive Therapy and*

Research, *5* (1), 41–56.

Bühner, M. (2011). *Einführung in die Test-und Fragebogenkonstruktion.* Pearson Deutschland.

Büning, H. (2004). Breites Angebot an falschen Lösungen. Mathematikkenntnisse von Studienanfängern im Test. *Forschung & Lehre*, *11*, 618–620.

Carlson, J. E. (2011). Statistical models for vertical linking. In A. A. von Davier (Hrsg.), *Statistical models for test equating, scaling, and linking* (S. 59–70). New York: Springer.

Carpenter, T. P., Lindquist, M. M., Matthews, W. & Silver, E. A. (1983). Results of the third NAEP mathematics assessment: Secondary school. *The Mathematics Teacher*, 652–659.

Carr, M., Bowe, B. & Ní Fhloinn, E. (2013). Core skills assessment to improve mathematical competency. *European Journal of Engineering Education*, *38* (6), 608–619.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York: Cambridge University Press.

Cassady, J. C. & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology*, *27* (2), 270–295.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1* (2), 245–276.

Cattell, R. B. (1971). *Abilities: Their Structure, Growth, and Action.* Boston: Houghton Mifflin.

Chappell, K. K. & Killpatrick, K. (2003). Effects of concept-based instruction on students'conceptual understanding and procedural knowledge of calculus. *Problems, Resources, and Issues in Mathematics Undergraduate Studies*, *13* (1), 17–37.

Christensen, B. & Meier, J.-H. (2014). Zur Frühidentifikation von Studienabbrüchen. *Das Hochschulwesen*, *62* (6), 183–185.

Christoforou, A. & Yigit, A. (2008). Improving teaching and learning in engineering education through a continuous assessment process. *European Journal of Engineering Education*, *33* (1), 105–116.

Chu, A. & Choi, J. (2005). Rethinking procrastination: Positive effects of „active“ procrastination behavior on attitudes and performance. *The Journal of Social Psychology*, *145*, 245–264.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. Aufl.). Hillsdale, New Jersey: Lawrence Earlbaum Associates.

Cole, J. S. & Spence, S. W. (2012). Using continuous assessment to promote student engagement in a large class. *European Journal of Engineering Education*, *37* (5), 508–525.

Culler, R. E. & Holahan, C. J. (1980). Test anxiety and academic performance: the effects of study-related behaviors. *Journal of Educational Psychology*, *72* (1), 16.

von Davier, A. A., Carstensen, C. H. & von Davier, M. (2008). Linking competencies in horizontal, vertical, and longitudinal settings and measuring growth. In J. Hartig, E. Klieme & D. Leutner (Hrsg.), *Assessment of competencies in educational contexts* (S. 53–80). Göttingen: Hogrefe & Huber.

von Davier, A. A., Holland, P. W. & Thayer, D. T. (2004). *The kernel method of test equating.* New York: Springer.

Deffenbacher, J. L. (1978). Worry, emotionality, and task-generated interference in test anxiety: An empirical test of attentional theory. *Journal of Educational Psychology*, *70* (2), 248–254.

Deffenbacher, J. L. & Deitz, S. R. (1978). Effects of test anxiety on performance, worry, and emotionality in naturally occurring exams. *Psychology in the Schools*, *15* (3), 446–450.

Denny, J. K., Nelson, D. G. & Zhao, M. Q. (2012). Creating and analyzing the effectiveness of a mathematics placement policy for new freshmen. *PRIMUS*, *22* (3), 177–185.

Dilling, H., Mombour, W., Schmidt, M. H. et al. (1991). *Internationale Klassifikation psychischer Störungen: ICD-10, Kapitel V.* Bern: Huber.

Döring, N. & Bortz, J. (2006). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (4. Aufl.). Berlin: Springer.

Doyé, V. (2013). *Die hohe Hürde Mathematik.* Zugriff am 18.7.2015 auf http://www.donaukurier.de/themen/wissen/hochschule/art133692,2717549

DPA. (2012). *Abbruch im Ingenieurstudium vorbeugen.* Zugriff am 18.7.2015 auf http://www.focus.de/wissen/mensch/campus/bildung-abbruch-im-ingenieurstudium-vorbeugen_aid_748891.html

Dreyfus, T. & Eisenberg, T. (1990). Conceptual calculus: Fact or fiction?. *Teaching Mathematics and Its Applications*, *9* (2), 63–67.

Drüke-Noe, C. (2014). *Aufgabenkultur in Klassenarbeiten im Fach Mathematik: Empirische Untersuchungen in neunten und zehnten Klassen.* Wiesbaden: Springer Spektrum.

Dunlap, W. P., Cortina, J. M., Vaslow, J. B. & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods*, *1* (2), 170–177.

Dürrschnabel, K., Klein, H., Niederdrenk-Felgner, C., Dürr, R., Weber, B. & Wurth, R. (2012). *Mindestanforderungskatalog Mathematik der Hochschulen Baden-Württembergs für ein Studium von MINT oder Wirtschaftsfächern (WiMINT).* Zugriff am 7.6.2016 auf https://www.hs-karlsruhe.de/hochschule/lehre/didaktische-weiterbildung-und-austausch/cosh-cooperation-schule-hochschule.html

Eckstein, K. (2001). *Evaluation-Report (BO-SE) Burnout and Self-Efficacy.* Zugriff am 27.6.2015 auf http://www.ebop.salzburg.at/TExte/evaluation%20reports_4/EVALUATION_BO-SE_REPORT.pdf

El Gaidi, K. & Ekholm, T. (2015). Contextualizing calculus with everyday examples to enhance conceptual learning. In *2015 122nd ASEE Annual Conference and Exposition; Washington Convention CenterSeattle; United States* (Bd. 122).

Engberding, M., Frings, E., Höcker, A., Wolf, J. & Rist, F. (2011). *Is procrastination a symptom or a disorder like other Axis-1-disorders in the DSM? Steps towards delineating a case definition.* Zugriff am 14.7.2015 auf http://www.fmg.uva.nl/procrastinationconference2011/programme.cfm

Engelbrecht, J., Bergsten, C. & Kågesten, O. (2009). Undergraduate students' preference for procedural to conceptual solutions to mathematical problems. *International journal of mathematical education in science and technology*, *40* (7), 927–940.

Engelbrecht, J., Bergsten, C. & Kågesten, O. (2012). Conceptual and procedural approaches to mathematics in the engineering curriculum: Student conceptions and performance. *Journal of Engineering Education*, *101* (1), 138–162.

Engelbrecht, J., Harding, A. & Du Preez, J. (2007). Long-term retention of basic mathematical knowledge and skills with engineering students. *European Journal of Engineering Education*, *32* (6), 735–744.

Engelbrecht, J., Harding, A. & Potgieter, M. (2005). Undergraduate students' performance and confidence in procedural and conceptual mathematics. *International Journal of Mathematical Education in Science and Technology*, *36* (7), 701–712.

Fakultät für Mathematik der TU Dortmund. (2008). *Qualitätsbericht der Fakultät für Mathematik über das Projekt HMplus (Wintersemester 2008/2009 und Sommersemester 2009).* Zugriff am 3.9.2016 auf http://www.mathematik.tu-dortmund.de/media/studinfo/studienbeitraege/Bericht-HMplus-2008-1.pdf

Fakultät für Mathematik der TU Dortmund. (2013). *Qualitätsbericht 2013 der Fakultät für Mathematik – Berichtszeitraum: 01.08.2011-31.12.2013.* Zugriff am 3.9.2016 auf http://www.mathematik.tu-dortmund.de/media/studinfo/studienbeitraege/Qualitaetsbericht-2013.pdf

Faulkner, F., Hannigan, A. & Fitzmaurice, O. (2014). The role of prior mathematical experience in predicting mathematics performance in higher education. *International Journal of Mathematical Education in Science and Technology*, *45* (5), 648–667.

Fehm, L. & Fydrich, T. (2011). *Prüfungsangst* (D. Schulte, K. Hahlweg, J. Margraf & D. Vaitl, Hrsg.). Göttingen: Hogrefe.

Fehm, L. & Fydrich, T. (2013). *Ratgeber Prüfungsangst: Informationen für Betroffene und Angehörige*. Göttingen: Hogrefe.

Ferrari, J. R., Johnson, J. L. & McCown, W. G. (1995). *Procrastination and task avoidance: Theory, research, and treatment.* New York: Springer.

Ferrini-Mundy, J. & Gaudard, M. (1992). Secondary school calculus: Preparation or pitfall in the study of college calculus? *Journal for Research in Mathematics Education*, 56–71.

Fischer, F., Schult, J. & Hell, B. (2015). Unterschätzung der Studienleistungen von Frauen durch Studierfähigkeitstests: Erklärbar durch Persönlichkeitseigenschaften? *Diagnostica*, 34–46.

Fleiss, J. L., Levin, B. & Paik, M. C. (2003). *Statistical methods for rates and proportions* (3. Aufl.). Hoboken, New Jersey: Wiley & Sons.

Formann, A. & Piswanger, K. (1979). *WMT. Wiener Matrizen-Test.* Weinheim: Beltz.

Formann, A., Piswanger, K. & Waldherr, K. (2011). *Wiener Matrizen-Test 2: Ein Rasch-skalierter sprachfreier Kurztest zur Erfassung der Intelligenz.* Göttingen: Hogrefe.

Formazin, M., Schroeders, U., Köller, O., Wilhelm, O. & Westmeyer, H. (2011). Studierendenauswahl im Fach Psychologie. *Psychologische Rundschau*, 221–236.

Foster, C. (2014). 'Can't you just tell us the rule?' Teaching procedures relationally. In *Proceedings of the 8th British Congress of Mathematics Education 2014.*

Freesemann, O. (2014). Inhaltliche Schwierigkeiten rechenschwacher Schülerinnen und Schüler. In *Schwache Rechnerinnen und Rechner fördern* (S. 31–51). Wiesbaden: Springer.

Frerich, T. & Pitsch, A. (2016). *Flexibilitätsförderung bei routinebasiertem Entscheidungsverhalten im Falle eines komplexen hochschulmathematischen Verfahrens* (unveröffentlichte Masterarbeit). TU Dortmund.

Fuller, E., Deshler, J. M., Kuhn, B. & Squire, D. (2014). Tracking the success of pre-college algebra workshop students in subsequent college mathematics classes. *PRIMUS*, *24* (1), 46–60.

Funsch, K., Martín, B. A. & Halder-Sinn, P. (2011). Differentieller Konzentrationstest für Kinder (DKT-K). *Diagnostica*, *57* (1), 2–16.

Fydrich, T. (2009). Arbeitsstörungen und Prokrastination. *Psychotherapeut*, *54* (5), 318–325.

Gensch, K. & Kliegl, C. (2011). *Studienabbruch – was können Hochschulen dagegen tun?* München: IHF.

George, D. (2007). *SPSS for windows step by step: A simple study guide and reference, 17.0 update, 10/e* (7. Aufl.). New York: Pearson.

Georges, K. E. (o. J.). *Ausführliches lateinisch-deutsches Handwörterbuch.* Zugriff am 15.7.2015 auf http://www.zeno.org/Zeno/0/Suche?q=procrastino&k=Georges-1913

Gittler, G. (1990). *Dreidimensionaler Würfeltest: 3 DW: Ein rasch-skalierter Test zur Messung des räumlichen Vorstellungsvermögens.* Weinheim: Beltz.

Gittler, G. (1999). Sind Raumvorstellung und Reasoning separierbare Fähigkeitsdimensionen? Dimensionalitätsanalysen zweier Rasch-skalierter Tests: 3 DW und WMT. *Diagnostica*, *45* (2), 69–81.

Goldhammer, F., Moosbrugger, H. & Krawietz, S. A. (2009). Fact-2-The Frankfurt Adaptive Concentration Test: Convergent validity with self-reported cognitive failures. *European Journal of Psychological Assessment*, *25* (2), 73–82.

Grundmeier, T. A., Hansen, J. & Sousa, E. (2006). An exploration of definition and procedural fluency in integral calculus. *Problems, Resources, and Issues in Mathematics Undergraduate Studies*, *16* (2), 178–191.

Haapasalo, L. & Kadijevich, D. (2000). Two types of mathematical knowledge and their relation. *Journal für Mathematik-Didaktik*, *21* (2), 139–157.

Haase, D. (2014). Studieren im MINT-Kolleg Baden-Württemberg. In I. Bausch et al. (Hrsg.), *Mathematische Vor-und Brückenkurse* (S. 123–136). Wiesbaden: Springer Spektrum.

Hallett, D. H. (2006). What have we learned from calculus reform? The road to conceptual understanding. *MAA NOTES*, *69*, 43.

Hasselblad, V. & Hedges, L. V. (1995). Meta-analysis of screening and diagnostic tests. *Psychological bulletin*, *117* (1), 167–178.

Hatano, G. & Inagaki, K. (1984). Two courses of expertise. In *Research and clinical center for child development Annual Report* (S. 27–36). Hokkaido: HUSCAP.

Hatzinger, R., Hornik, K. & Nagel, H. (2011). *R – Einführung durch angewandte Statistik*. München: Pearson.

Heckhausen, H. & Gollwitzer, P. M. (1987). Thought contents and cognitive functioning in motivational versus volitional states of mind. *Motivation and emotion*, *11* (2), 101–120.

Heckhausen, J. & Heckhausen, H. (2006). *Motivation und Handeln*. Berlin: Springer.

Heinisch, I., Romeike, R. & Eichler, K.-P. (2016). Outcome-orientierte Neuausrichtung der Hochschullehre für das Fach Mathematik. In A. Hoppenbrock, R. Biehler, R. Hochmuth & H.-G. Rück (Hrsg.), *Lehren und Lernen von Mathematik in der Studieneingangsphase* (S. 261–275). Wiesbaden: Springer.

Hell, B., Linsner, M. & Kurz, G. (2008). Studieneignung und Studierendenauswahl – Untersuchungen und Erfahrungsberichte. In M. Rentschler & H.-P. Voss (Hrsg.), (S. 132–177). Aachen: Shaker.

Hell, B., Trapmann, S. & Schuler, H. (2007). Eine Metaanalyse der Validität von fachspezifischen Studierfähigkeitstests im deutschsprachigen Raum. *Empirische Pädagogik*, *21* (3), 251–270.

Helmke, A. & Schrader, W. (2000). Prokrastination im Studium. In D. Danner & A. Glöckner-Rist (Hrsg.), *Interesse und Lernmotivation. Untersuchungen zu Entwicklung, Förderung und Wirkung* (S. 207–225). Münster: Waxmann.

Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research*, *58* (1), 47–77.

Henn, G. & Polaczek, C. (2007). Studienerfolg in den Ingenieurwissenschaften. *Das Hochschulwesen* (5), 144–147.

Heublein, U., Hutzsch, C., Schreiber, J., Sommer, D. & Besuch, G. (2010). *Ursachen des Studienabbruchs in Bachelor- und in herkömmlichen Studiengängen.* Hannover: HIS, Forum Hochschule.

Heublein, U., Richter, J., Schmelzer, R. & Sommer, D. (2014). *Die Entwicklung der Studienabbruchquoten an den deutschen Hochschulen.* Hannover: HIS, Forum Hochschule.

Hieb, J. L., Lyle, K. B., Ralston, P. A. & Chariker, J. (2015). Predicting performance in a first engineering calculus course: implications for interventions. *International Journal of Mathematical Education in Science and Technology*, *46* (1), 40–55.

Hiebert, J. (1986). *Conceptual and procedural knowledge: The case of mathematics.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Hiebert, J. & Lefevre, P. (1986). Conceptual and Procedural Knowledge in Mathematics: An Introductory Analysis. In J. Hiebert (Hrsg.), *Conceptual and procedural knowledge: The case of mathematics* (S. 1–28). Lawrence Erlbaum Associates, Inc.

Hodapp, V. (1982). Causal Inference From Nonexperimental Research on Anxiety and Educational Achievement. In H. H. Krone & L. Laux (Hrsg.), *Achievement, Stress, and Anxiety.* Washington, D.C.: Hemisphere.

Hodapp, V. (1991). Das Prüfungsängstlichkeitsinventar TAI-G: Eine erweiterte und modifizierte Version mit vier Komponenten. *Zeitschrift für Pädagogische Psychologie*, 121–130.

Hodapp, V. (1995). The TAI-G: A multidimensional approach to the assessment of test anxiety. *Stress, anxiety, and coping in academic settings*, 95–130.

Hodapp, V., Rohrmann, S. & Ringeisen, T. (2011). *Prüfungsangstfragebogen: PAF.* Göttingen: Hogrefe.

Holling, H., Preckel, F. & Vock, M. (2004). *Intelligenzdiagnostik.* Hogrefe Verlag.

Hong, E. (1998). Differential stability of individual differences in state and trait test anxiety. *Learning and Individual Differences*, *10* (1), 51–69.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30* (2), 179–185.

Horn, J. L. (1968). Organization of abilities and the development of intelligence. *Psychological Review*, *75* (3), 242–259.

Horstkötter, A. (2006). Studentischer Input und universitäre Lehre: Inwieweit kann eine Universität mangelnde studentische Vorbildung ausgleichen? *Das Hochschulwesen* (6), 202–206.

Huang, C.-H. (2010). Conceptual and procedural abilities of engineering students to perform mathematical integration. In *Proceedings of the Conference of the Joint International IGIP-SEFI Annual Conference.* Trnava, Slovakei, 19.–22. Sept. 2010.

Jensen, A. R. (1981). *Straight talk about mental tests.* New York: Free Press.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport: Praeger.

Jordan, A. (2006). *Klassifikationsschema für Mathematikaufgaben: Dokumentation der Aufgabenkategorisierung im COACTIV-Projekt.* Berlin: Max-Planck-Institut für Bildungsforschung.

Jordan, A., Krauss, S., Löwen, K., Blum, W., Neubrand, M., Brunner, M., … Baumert, J. (2008). Aufgaben im COACTIV-Projekt: Zeugnisse des kognitiven Aktivierungspotentials im deutschen Mathematikunterricht. *Journal für Mathematik-Didaktik*, *29* (2), 83–107.

Kaiser, H. F. & Dickman, K. (1959). Analytic determination of common factors. *American Psychologist*, *14* (7), 425–439.

Kane, M., Crooks, T. & Cohen, A. (1999). Validating measures of performance. *Educational measurement: issues and practice*, *18* (2), 5–17.

Kang, T. & Petersen, N. S. (2012). Linking item parameters to a base scale. *Asia Pacific Education Review*, *13* (2), 311–321.

Kersten, I. (2015). Kalkülfertigkeiten an der Universität: Mängel erkennen und Konzepte für die Förderung entwickeln. In *Übergänge konstruktiv gestalten* (S. 33–49). Wiesbaden: Springer.

Kieran, C. (2013). The false dichotomy in mathematics education between conceptual understanding and procedural skills: an example from Algebra. In *Vital directions for mathematics education research* (S. 153–171). New York: Springer.

Kilpatrick, J., Swafford, J. & Findell, B. (2001). *Adding it up: Helping children learn mathematics.* Washington, D.C.: National Academies Press.

Kim, K. R. & Seo, E. H. (2015). The relationship between procrastination and academic performance: A meta-analysis. *Personality and Individual Differences*, *82*, 26–33.

Kirkland, K. & Hollandsworth, J. G. (1980). Effective test taking: skills-acquisition versus anxiety-reduction techniques. *Journal of Consulting and Clinical Psychology*, *48* (4), 431.

Klingsieck, K. B. (2013). Procrastination. *European Psychologist*, *18* (1), 24–34.

Klingsieck, K. B. & Fries, S. (2012). Allgemeine Prokrastination: Entwicklung und Validierung einer deutschsprachigen Kurzskala der General Procrastination Scale (Lay, 1986). *Diagnostica*, *58* (4), 182–193.

Knaus, W. J. (2002). *The procrastination workbook: Your personalized program for breaking free from the patterns that hold you back.* Oakland, Kalifornien: New Harbinger.

Knigge-Illner, H. (2010). *Prüfungsangst besiegen: Wie Sie Herausforderungen souverän* *meistern.* Campus Verlag: Frankfurt.

Knoche, N., Lind, D., Blum, W., Cohors-Fresenborg, E., Flade, L., Löding, W., ... Wynands, A. (2002). Die PISA-2000-Studie, einige Ergebnisse und Analysen. *Journal für Mathematik-Didaktik*, *23* (3–4), 159–202.

Kobrin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D. & Barbuti, S. M. (2008). *Validity of the SAT for Predicting First-Year College Grade Point Average (College Board Research Report No. 2008-5).* New York: The College Board.

Kolari, S., Savander-Ranne, C. & Viskari, E.-L. (2008). Learning needs time and effort: a time-use study of engineering students. *European Journal of Engineering Education*, *33* (5-6), 483–498.

Kolb, M., Kraus, M., Pixner, J. & Schüpbach, H. (2006). Analyse von Studienverlaufsdaten zur Identifikation von studienabbruchgefährdeten Studierenden. *Das Hochschulwesen*, *54* (6), 196–201.

Kolen, M. J. & Brennan, R. L. (2004). *Test equating, scaling, and linking.* Springer.

Koller, I., Alexandrowicz, R. & Hatzinger, R. (2012). *Das Rasch Modell in der Praxis: Eine Einführung in eRm* (Bd. 3786). Wien: UTB.

König, C. J. & Kleinmann, M. (2004). Business before pleasure: no strategy for procrastinators? *Personality and Individual Differences*, *37* (5), 1045–1057.

Kratz, H.-J. (2011). *Aufschieben - nein danke!* Regensburg: Walhalla Fachverlag.

Kröner, S. & Fritzsche, E. S. (2015). Wer riskiert eine Mahnung? *Diagnostica*, 169–181.

Kubinger, K. D. (2009). *Psychologische Diagnostik: Theorie und Praxis psychologischen Diagnostizierens.* Göttingen: Hogrefe.

Kubinger, K. D. & Draxler, C. (2007). Probleme bei der Testkonstruktion nach dem Rasch-Modell. *Diagnostica*, *53* (3), 131–143.

Kultusminsterkonferenz. (2010). *Ländergemeinsame Strukturvorgaben für die Akkreditierung von Bachelor- und Masterstudiengänge (Beschluss der Kultusministerkonferenz vom 10.10.2003 i.d.F. vom 04.02.2010).* Zugriff am 10.6.2016 auf http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2003/2003_10_10-Laendergemeinsame-Strukturvorgaben.pdf

Kuncel, N. R., Credé, M. & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of educational research*, *75* (1), 63–82.

Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159–174.

Leitner, W. (2005). *Konzentrationsleistung und Aufmerksamkeitsverhalten: Begriff, Einflussfaktoren, Entwicklung, Diagnostik, Prävention und Intervention.* Regensburg: Roderer.

Levy, A. & Merry, U. (1986). *Organizational transformation: Approaches, strategies,* *theories.* New York: Praeger.

Liebert, R. M. & Morris, L. W. (1967). Cognitive and emotional components of test anxiety: A distinction and some initial data. *Psychological reports*, *20* (3), 975–978.

Lithner, J. (2004). Mathematical reasoning in calculus textbook exercises. *The Journal of Mathematical Behavior*, *23* (4), 405–427.

Lithner, J. (2006). *A framework for analysing creative and imitative mathematical reasoning. research reports in mathematics education.* Department of Mathematics and Mathemical Statistics, Umeå Universitet.

Livingston, S. A. (2004). *Equating test scores.* Princeton, New Jersey: Educational Testing Service.

Maag Merki, K. (2012). Die Leistungen der Gymnasiastinnen und Gymnasiasten in Mathematik und Englisch. In *Zentralabitur. Die längsschnittliche Analyse der Prozesse und Wirkungen der Einführung zentraler Abiturprüfungen in zwei Bundesländern* (S. 263–292). Wiesbaden: VS Verlag für Sozialwissenschaften.

Madison, B. L., Linde, C. S., Decker, B. R., Rigsby, E. M., Dingman, S. W. & Stegman, C. E. (2015). A study of placement and grade prediction in first college mathematics courses. *PRIMUS*, *25* (2), 131–157.

Mahir, N. (2009). Conceptual and procedural performance of undergraduate students in integration. *International Journal of Mathematical Education in Science and Technology*, *40* (2), 201–211.

Mair, P., Hatzinger, R. & Maier, M. (2015). *eRm: Extended Rasch modeling.* Zugriff am 10.6.2016 auf https://cran.r-project.org/web/packages/eRm/index.html

Mandler, G. & Sarason, S. B. (1952). A study of anxiety and learning. *The Journal of Abnormal and Social Psychology*, *47* (2).

Marakovits, N. (2008). *Vergleich einer revidierten Form des Wiener Matrizen-Tests (WMT) in Berlin und Österreich* (Diplomarbeit). Universität Wien.

Marshalek, B., Lohman, D. F. & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence*, *7* (2), 107–127.

Marshall, S. (1980). Procedural networks and production systems in adaptive diagnosis. *Instructional Science*, *9* (2), 129–143.

Mason, J. & Spence, M. (1999). Beyond mere knowledge of mathematics: The importance of knowing-to act in the moment. *Educational Studies in Mathematics*, *38* (1–3), 135–161.

Matic, L. J. (2014). Mathematical knowledge of non-mathematics students and their beliefs about mathematics. *Mathematics Education*, *9* (1), 13–24.

Mauermeister, S., Zylla, B. & Wagner, L. (2015). Wie gut sind die Konzepte zum Studieneingang? Das StuFo-Verbundprojekt zur Wirksamkeit der Studieneingangsphase. *Qualität in der Wissenschaft* (2), 50–55.

Meiner, S. & Seiler, R. (2009, January). *Abschlussbericht Expertentreffen Brückenkurs Mathematik.* Zugriff am 9.8.2015 auf http://page.math.tu-berlin.de/~seiler/publications/Abschlussbericht.pdf

Melguizo, T., Kosiewicz, H., Prather, G. & Bos, J. (2014). How Are Community College Students Assessed and Placed in Developmental Math?: Grounding Our Understanding in Reality. *The Journal of Higher Education*, *85* (5), 691–722.

Ministerium für Schule und Weiterbildung NRW. (2014). *Kernlehrplan für die Sekundarstufe II Gymnasium / Gesamtschule in Nordrhein-Westfalen.* Zugriff am 24.5.2016 auf http://www.schulentwicklung.nrw.de/lehrplaene/lehrplannavigator-s-ii/\\gymnasiale-oberstufe/

Mooraj, M. & Zervakis, P. (2014). *Der Umgang mit studentischer Heterogenität in Studium und Lehre. Chancen, Herausforderungen, Strategien und gelungene Praxisansätze aus den Hochschulen.* Zugriff am 10.6.2016 auf http://www.inklusion-online.net/index.php/inklusion-online/article/\linebreakview/222/224

Moosbrugger, H. & Kelava, A. (2007). *Testtheorie und Fragebogenkonstruktion.* Berlin: Springer.

Moser Opitz, E. (2007). *Rechenschwäche, Dyskalkulie: Theoretische Klärungen und empirische Studien an betroffenen Schülerinnen und Schülern.* Bern: Haupt.

Moser Opitz, E. & Nührenbörger, M. (2015). Diagnostik und Leistungsbeurteilung. In *Handbuch der Mathematikdidaktik* (S. 491–512). Berlin: Springer.

Musch, J. & Bröder, A. (1999). Psychometrische Eigenschaften und Validität des multidimensionalen Prüfungsängstlichkeitsinventars TAI-G. *Zeitschrift für Pädagogische Psychologie*, *13* (1/2), 100–105.

Nachtigall, C. & Wirtz, M. (2008). *Wahrscheinlichkeitsrechnung und Inferenzstatistik* (5. Aufl., Bd. 1). Weinheim: Juventa.

Nachtigall, C. & Wirtz, M. (2009). *Wahrscheinlichkeitsrechnung und Inferenzstatistik* (5. Aufl., Bd. 2). Weinheim: Juventa.

Naveh-Benjamin, M., McKeachie, W. J. & Lin, Y.-G. (1987). Two types of test-anxious students: Support for an information processing model. *Journal of Educational Psychology*, *79* (2), 131–136.

Naveh-Benjamin, M., McKeachie, W. J., Lin, Y.-G. & Holinger, D. P. (1981). Test anxiety: Deficits in information processing. *Journal of Educational Psychology*, *73* (6), 816–824.

Nevo, B. (1985). Face validity revisited. *Journal of Educational Measurement*, *22* (4), 287–293.

Niss, M. (2003). Mathematical competencies and the learning of mathematics: The Danish KOM project. In *3rd mediterranean conference on mathematical education*

(S. 115–124).

Ortiz, E. A. & Dehon, C. (2013). Roads to success in the Belgian French community's higher education system: predictors of dropout and degree completion at the Université libre de Bruxelles. *Research in Higher Education*, *54* (6), 693–723.

Oswald, W. & Roth, E. (1982). *Der Zahlenverbindungstest*. Göttingen: Hogrefe.

Parker, P. D., Marsh, H. W., Ciarrochi, J., Marshall, S. & Abduljabbar, A. S. (2014). Juxtaposing math self-efficacy and self-concept as predictors of long-term achievement outcomes. *Educational Psychology*, *34* (1), 29–48.

Parnell, S. & Statham, M. (2007). An effective preparation for tertiary mathematics. *International Journal of Mathematical Education in Science and Technology*, *38* (7), 869–879.

Parsons, S., Croft, T. & Harrison, M. (2009). Does students' confidence in their ability in mathematics matter? *Teaching Mathematics and its Applications*, *28* (2), 53–68.

Patz, R. J. (2007). *Vertical scaling in standards-based educational assessment and accountability systems*. Washington, DC: The Council of Chief State School Officers.

Paulman, R. G. & Kennelly, K. J. (1984). Test anxiety and ineffective test taking: Different names, same construct? *Journal of Educational Psychology*, *76* (2), 279–288.

Pekrun, R., Vom Hofe, R., Blum, W., Götz, T., Wartha, S., Frenzel, A. & Jullien, S. (2006). Projekt zur Analyse der Leistungsentwicklung in Mathematik (PALMA). *Untersuchungen zur Bildungsqualität von Schule*, 21–53.

Perdigones, A., Benedicto, S., Sánchez-Espinosa, E., Gallego, E. & García, J. L. (2014). How many hours of instruction are needed for students to become competent in engineering subjects? *European Journal of Engineering Education*, *39* (3), 300–308.

Pibal, F. & Cesnik, H. S. (2011). Evaluating the quantity-quality trade-off in the selection of anchor items: A vertical scaling approach. *Practical Assessment, Research & Evaluation*, *16* (6), 1–12.

Prediger, S. & Leuders, T. (2016). *Flexibel differenzieren und fokussiert fördern im Mathematikunterricht*. Berlin: Cornelsen.

Prediger, S. & Wittmann, G. (2009). Aus Fehlern lernen – (wie) ist das möglich? *Praxis der Mathematik in der Schule*, *51* (3), 1–8.

Prenzel, M., Artelt, C., Baumert, J., Blum, W., Hammann, M. & Klieme. (2007). *PISA 2006: Die Ergebnisse der dritten internationalen Vergleichsstudie* (R. Pekrun, Hrsg.). Münster: Waxmann.

Prenzel, M., Sälzer, C., Klieme, E. & Köller, O. (2013). *PISA 2012: Fortschritte und Herausforderungen in Deutschland*. Münster: Waxmann.

Rach, S. & Heinze, A. (2013). Welche Studierenden sind im ersten Semester erfolgreich? *Journal für Mathematik-Didaktik*, *34* (1), 121–147.

Ramist, L., Lewis, C. & Jenkins, L. M. (2001). *Using Achievement Tests/SAT II: Subject Tests to Demonstrate Achievement and Predict College Grades: Sex, Language, Ethnic, and Parental Education Groups* (Bericht Nr. 2001-5). New York: College Entrance Examination Board.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Kopenhagen: Danish Institute for Educational Research.

Raven, J. C. & Court, J. H. (1998). *Raven's progressive matrices and vocabulary scales.* Oxford: Oxford Psychologists Press.

Richardson, M., Abraham, C. & Bond, R. (2012). Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychological bulletin*, *138* (2), 353–387.

Rindermann, H. & Neubauer, A. (2004). Processing speed, intelligence, creativity, and school performance: Testing of causal hypotheses using structural equation models. *Intelligence*, *32* (6), 573–589.

Rindermann, H. & Neubauer, A. C. (2000). Informationsverarbeitungsgeschwindigkeit und Schulerfolg: Weisen basale Maße der Intelligenz prädiktive Validität auf? *Diagnostica*, *46* (1), 8–17.

Rist, F., Engberding, M., Patzelt, J. & Beißner, J. (2006). „Aber morgen fange ich richtig an" – Prokrastination als verbreitete Arbeitsstörung. *Personalführung*, *6*, 64–78.

Rittle-Johnson, B. & Schneider, M. (2014). Developing conceptual and procedural knowledge of mathematics. In R. Kadosh & A. Dowker (Hrsg.), *Oxford handbook of numerical cognition.* New York: Oxford University Press.

Rodgers, K., Posler, B. & Trible, L. (2011). A faster track: Decreasing the number of semesters students spend in developmental mathematics without lowering academic expectations. *PRIMUS*, *21* (3), 252–262.

Rost, D. & Sparfeldt, J. (2008). Intelligenz und Hochbegabung. In M. K. Schweer (Hrsg.), *Lehrer-Schüler-Interaktion* (2. Aufl., S. 303–325). Wiesbaden: VS Verlag für Sozialwissenschaften.

Rost, J. (2004). *Lehrbuch Testtheorie, Testkonstruktion* (2. Aufl.). Bern: Huber.

Rotenstein, A., Davis, H. Z. & Tatum, L. (2009). Early birds versus just-in-timers: the effect of procrastination on academic performance of accounting students. *Journal of Accounting Education*, *27* (4), 223–232.

Rustemeyer, R. & Callies, C. (2013). *Aufschieben, Verzögern, Vermeiden: Einführung in die Prokrastination.* Darmstadt: WBG.

Sacerdoti, E. D. (1975). *A Structure for Plans and Behavior* (Dissertation). Stanford University.

Sachs, L. & Hedderich, J. (2006). *Angewandte Statistik. Methodensammlung mit R.* (12. Aufl.). Berlin: Springer.

Sarason, I. G. (1984). Stress, anxiety, and cognitive interference: reactions to tests. *Journal of personality and social psychology*, *46* (4), 929–938.

Sarason, S. B. & Mandler, G. (1952). Some correlates of test anxiety. *The Journal of Abnormal and Social Psychology*, *47* (4), 561–565.

Satow, L. (2011). *Psychomeda-Konzentrationstest (KONT-P): Skalendokumentation und Normen.* Zugriff am 15.7.2015 auf www.psychomeda.de/online-tests/Psychomeda-Konzentrationstest.pdf

Schaefer, A., Albus, C. & Pfitzer, G. (2009). *Gruppentherapeutisches Angebot für hoch prüfungsängstliche Studierende der Human- und Zahnmedizin.* Zugriff am 10.6.2016 auf www.annschaefer.de/gruppenprojekt.pdf

Scheffler, I. (1965). *Conditions of knowledge: An introduction to epistemology and education.* Chicago: Scott, Foresman.

Scherer, P. & Moser Opitz, E. (2010). *Fördern im Mathematikunterricht der Primarstufe.* Heidelberg: Springer.

Schmidt-Atzert, L., Bühner, M. & Enders, P. (2006). Messen Konzentrationstests Konzentration? *Diagnostica*, *52* (1), 33–44.

Schmidt-Atzert, L., Büttner, G. & Bühner, M. (2004). Theoretische Aspekte von Aufmerksamkeits-/Konzentrationsdiagnostik. *Diagnostik von Konzentration und Aufmerksamkeit*, 3–22.

Schmidt-Atzert, L., Deter, B. & Jaeckel, S. (2004). Prädiktion von Ausbildungserfolg: Allgemeine Intelligenz (g) oder spezifische kognitive Fähigkeiten? *Zeitschrift für Personalpsychologie*, *3* (4), 147–158.

Schoenfeld, A. H. (1989). Explorations of students' mathematical beliefs and behavior. *Journal for Research in Mathematics Education*, 338–355.

Schoenfeld, A. H. (2004). The math wars. *Educational policy*, *18* (1), 253–286.

Schouwenburg, H. C. (2004). Academic procrastination: Theoretical notions, measurement and research. In H. C. Schouwenburg, C. H. Lay, T. A. Pychyl & J. R. Ferrari (Hrsg.), *Counseling the procrastinator in academic settings.* (S. 3–17). Washington, DC: American Psychological Association.

Schraw, G., Wadkins, T. & Olafson, L. (2007). Doing the things we do: A grounded theory of academic procrastination. *Journal of Educational Psychology*, *99* (1), 12–25.

Schroeder, L. D., Sjoquist, D. L. & Stephan, P. E. (1986). *Understanding regression analysis: An introductory guide.* Newbury Park, Kalifornien: Sage.

Schuler, H., Funke, U. & Baron-Boldt, J. (1990). Predictive validity of school grades-a meta-analysis. *Applied Psychology*, *39* (1), 89–103.

Schwarzer, R. (1999). *Prokrastination.* Zugriff am 27.6.2015 auf http://userpage.fu-berlin.de/~gesund/skalen/Prokrastination/prokrastination.htm

Schwarzer, R. (2000). *Stress, Angst und Handlungsregulation.* Stuttgart: Kohlhammer.

Seipp, B. (1991). Anxiety and academic performance: A meta-analysis of findings. *Anxiety Research*, *4* (1), 27–41.

Shorter, N. A. & Young, C. Y. (2011). Comparing assessment methods as predictors of student learning in an undergraduate mathematics course. *International Journal of Mathematical Education in Science and Technology*, *42* (8), 1061–1067.

Sitzwohl, E. M. (1995). *Konstruktion und empirische Überprüfung eines Papier-Bleistift-Tests zur Erfassung von Informationsverarbeitungsgeschwindigkeit: Der Coding-Test* (Graz: unveröffentlichte Diplomarbeit).

Skemp, R. R. (1979). *Intelligence, learning and action.* London: Wiley & Sons.

Snow, R., Kyllonen, P. & Marshalek, B. (1984). The Topography of Ability and Learning Correlations. In R. Sternberg (Hrsg.), *Advances in the psychology of human intelligence.* Hillsdale, New York: Erlbaum.

Solomon, L. J. & Rothblum, E. D. (1984). Academic procrastination: Frequency and cognitive-behavioral correlates. *Journal of Counseling Psychology*, *31* (4), 503–509.

Sonnert, G. & Sadler, P. M. (2014). The impact of taking a college pre-calculus course on students' college calculus performance. *International Journal of Mathematical Education in Science and Technology*, *45* (8), 1188–1207.

Spearman, C. (1904). "general intelligence" objectively determined and measured. *The American Journal of Psychology*, *15* (2), 201–292.

Spearman, C. (1927). *The abilities of man.* New York: Macmillan.

Spielberger, C. D. (1980). *Test anxiety inventory. Preliminary professional manual.* Palo Alto: Consulting Psychologists Press.

Stankov, L. & Lee, J. (2014). Quest for the best non-cognitive predictor of academic achievement. *Educational Psychology*, *34* (1), 1–8.

Stankov, L., Morony, S. & Lee, Y. P. (2014). Confidence: the best non-cognitive predictor of academic achievement? *Educational Psychology*, *34* (1), 9–28.

Star, J. R. (2005). Reconceptualizing procedural knowledge. *Journal for research in mathematics education*, 404–411.

Star, J. R., Pollack, C., Durkin, K., Rittle-Johnson, B., Lynch, K., Newton, K. & Gogolen, C. (2015). Learning from comparison in algebra. *Contemporary Educational Psychology*, *40*, 41–54.

Steel, P. (2007). The nature of procrastination: a meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychological Bulletin*, *133* (1), 65–94.

Steel, P., Brothen, T. & Wambach, C. (2001). Procrastination and personality, performance, and mood. *Personality and individual differences*, *30* (1), 95–106.

Steinborn, M. B., Flehmig, H. C., Westhoff, K. & Langner, R. (2008). Predicting school achievement from self-paced continuous performance: Examining the contributions of response speed, accuracy, and response speed variability. *Psychology Science*

Quarterly, *50* (4), 613–634.

Sternberg, R. (2011). *Cognitive psychology*. Belmont, Kalifornien: Wadsworth.

Steyer, R. & Eid, M. (1993). *Messen und Testen*. Berlin: Springer-Verlag.

Stocking, M. L. & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied psychological measurement*, *7* (2), 201–210.

Strobl, C. (2012). *Das Rasch-Modell*. München: Hampp.

Sukin, T. M. (2010). *Item parameter drift as an indication of differential opportunity to learn: An exploration of item flagging methods & accurate classification of examinees* (Dissertation). University of Massachusetts.

Süß, H. (2003). Intelligenztheorien. In K. D. Kubinger & R. Jäger (Hrsg.), *Schlüsselbegriffe der psychologischen Diagnostik* (S. 217–224). Weinheim: Beltz.

Tall, D. (1996). Functions and calculus. In *International Handbook of Mathematics Education* (S. 289–325). Dordrecht: Springer.

Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.

Tian, F. (2011). *A comparison of equating/linking using the stocking-lord method and concurrent calibration with mixed-format tests in the non-equivalent groups common-item design under irt* (Dissertation). Boston College.

Tice, D. M. & Baumeister, R. F. (1997). Longitudinal study of procrastination, performance, stress, and health: The costs and benefits of dawdling. *Psychological science*, 454–458.

Tolan, M. (2013). *Mathematik ist die größte Hürde*. Zugriff am 9.3.2016 auf https://www.youtube.com/watch?v=wtpoAR6vYBU

Trapmann, S., Hell, B., Weigand, S. & Schuler, H. (2007). Die Validität von Schulnoten zur Vorhersage des Studienerfolgs – eine Metaanalyse. *Zeitschrift für pädagogische Psychologie*, *21* (1), 11–27.

Trautwein, U. (2007). The homework-achievement relation reconsidered: Differentiating homework time, homework frequency, and homework effort. *Learning and Instruction*, *17* (3), 372–388.

Trautwein, U., Hübner, N., Wagner, W. & Kramer, J. (2015). *Die G8-Reform in Baden-Württemberg – Zusammenfassung zentraler Befunde*. Zugriff am 10.6.2016 auf https://www.uni-tuebingen.de/uploads/media/2015-04-20_Studie_Konsequenzen_der_G8-Reform.pdf

Trautwein, U., Köller, O., Lehmann, R. & Lüdtke, O. (2006). *LAU Aspekte der Lernausgangslage und der Lernentwicklung – Klassenstufe 13*. Hamburg: Behörde für Bildung und Sport.

Trost, G., Klieme, E. & Nauels, H. (1997). Prognostische Validität des Tests für medizinische Studiengänge (TMS). In T. Herrmann (Hrsg.), *Hochschulentwicklung - Aufgaben und Chancen* (S. 57–78). Heidelberg: Asanger.

Tryon, G. S. (1980). The measurement and treatment of test anxiety. *Review of Educational research*, *50* (2), 343–372.

Tynjälä, P., Salminen, R. T., Sutela, T., Nuutinen, A. & Pitkänen, S. (2005). Factors related to study success in engineering education. *European Journal of Engineering Education*, *30* (2), 221–231.

Van Eerde, W. (2003). A meta-analytically derived nomological network of procrastination. *Personality and Individual Differences*, *35* (6), 1401–1418.

Vollmers, F. (2011). *Keine Angst vor großen Zahlen.* Zugriff am 10.6.2016 auf http://www.faz.net/aktuell/beruf-chance/campus/mathematik-keine-angst-vor-grossen-zahlen-11340157-p2.html?printPagedArticle=true#pageIndex_2

vom Hofe, R. (1992). Grundvorstellungen mathematischer Inhalte als didaktisches Modell. *Journal für Mathematik-Didaktik*, *13* (4), 345–364.

von Davier, M. (1997). *Methoden zur Prüfung probabilistischer Testmodelle* (Nr. 158). Kiel: IPN.

Wacker, A., Jaunzeme, J. & Jaksztat, S. (2008). Eine Kurzform des Prüfungsängstlichkeitsinventars TAI-G. *Zeitschrift für Pädagogische Psychologie*, *22* (1), 73–81.

Walther, H. (2012). *Ohne Prüfungsangst studieren.* Konstanz: UVK.

Wang, X. (2013). *Linking across forms in vertical scaling under the common-item nonequvalent groups design* (Dissertation). University of Iowa.

Westhoff, K. (1995). Aufmerksamkeit und Konzentration. *Verhaltens- und Leistungsunterschiede*, *2*, 375–402.

Westhoff, K. & Graubner, J. (2003). Konstruktion eines komplexen Konzentrationstests. *Diagnostica*, *49* (3), 110–119.

Westhoff, K. & Hagemeister, C. (1992). Reliabilität von Fehlern in Konzentrationstests. *Diagnostica*, 116–129.

Westhoff, K. & Hagemeister, C. (2005). *Konzentrationsdiagnostik.* Lengerich: Pabst.

White, P. & Mitchelmore, M. (1996). Conceptual knowledge in introductory calculus. *Journal for Research in Mathematics Education*, 79–95.

Wilhelm, O. (2005). *Measuring reasoning ability.* London: Sage.

Wilhelm, O. & Engle, R. (2004). *Handbook of understanding and measuring intelligence.* London: Sage.

Wilson, T. M. & MacGillivray, H. L. (2007). Counting on the basics: mathematical skills among tertiary entrants. *International Journal of mathematical education in science and technology*, *38* (1), 19–41.

Wine, J. (1971). Test anxiety and direction of attention. *Psychological Bulletin*, *76* (2).

Wirtz, M. A. & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen.* Göttingen: Hogrefe.

Wissenschaftsrat. (2004). *Empfehlungen zur Reform des Hochschulzuganges.* Zugriff am 17.8.2015 auf www.wissenschaftsrat.de/download/archiv/5920-04.pdf

Wright, B. D. & Stone, M. H. (1979). *Best Test Design. Rasch Measurement.* Chicago: University of Chicago.

Wu, H. (1999). Basic skills versus conceptual understanding. *American Educator*, *23* (3), 14–19.

Yip, M. C. (2013). Learning strategies and their relationships to academic performance of high school students in hong kong. *Educational Psychology*, *33* (7), 817–827.

Zeidner, M. (1998). *Test anxiety: The state of the art.* New York: Springer.

Zerr, R. J. (2009). Promoting Students' Ability to Think Conceptually in Calculus. *Primus*, *20* (1), 1–20.

Abbildungsverzeichnis

Tabellenverzeichnis