Schnelle Fehlererkennung als Facette von Lehrerprofessionswissen: Analysen und Validierung

– Eine Detailanalyse im Rahmen der Studie TEDS-Follow-up –

Kumulative Dissertation
zur Erlangung des akademischen Grades einer Doktorin
der Philosophie (Dr. phil.)
an der Fakultät für Erziehungswissenschaft
der Universität Hamburg

vorgelegt von
Lena Pankow
bei Prof. Dr. Gabriele Kaiser
Hamburg, Januar 2020

Für meine Familie

Gutachterinnen und Gutachter:

1. Prof. Dr. Gabriele Kaiser, Universität Hamburg (Betreuerin)

2. Prof. Dr. Jörg Doll, Universität Hamburg

3. Prof. Dr. Marianne Nolte, Universität Hamburg

Datum der mündlichen Disputation:
21. April 2020

Inhaltsverzeichnis

Auflistung der Einzelarbeiten in dieser Dissertation

Pankow, L., Kaiser, G., Busse, A., König, J., Hoth, J., Döhrmann, M., & Blömeke, S. (2016). Early career teachers' ability to focus on typical students errors in relation to the complexity of a mathematical topic. *ZDM Mathematics Education, 48*(1), 55-67.

Pankow, L., Kaiser, G., König, J. & Blömeke, S. (2018). Perception of student errors under time limitation: are teachers faster than mathematicians or students? *ZDM Mathematics Education, 50*(4), 631–642.

Pankow, L. & Kaiser, G. (2018). Ein zeitbeschränkter Test zur schnellen Erkennung von Schülerfehlern durch Junglehrkräfte – Qualitative Merkmale zur Schwierigkeitsbestimmung von Items. *mathematica didactica, 41*(2), 147-162.

1. Einleitung

Die Konzeptualisierung der professionellen Kompetenz von Lehrkräften und deren Messung haben in den letzten zwei Dekaden der empirischen Bildungsforschung deutlich an Bedeutung gewonnen. Die Einbeziehung der Wahrnehmung von Schülerfehlern als Teil dieser Kompetenz wie auch die Art und Weise des Umgangs mit Schülerfehlern wurde verstärkt erst in den letzten Jahren untersucht, insbesondere im Kontext des adaptiven Lehrerverhaltens (vgl. Südkamp & Praetorius 2017) sowie im Rahmen der Förderung von Diagnosekompetenz in der Lehrerausbildung[1] (vgl. KMK 2015).

Schülerfehler gehören in den Alltag eines jeden Unterrichts. Angesichts der zunehmenden Heterogenität der Schülerschaft, die nach Terhart (2011) aktuell zu den größten Herausforderungen gehört, sind die Fehler-Diagnose sowie der Umgang mit auftretenden Fehlern eine wichtige Kompetenz von Lehrkräften. Die produktive Nutzung der Schülerfehler bedarf seitens der Lehrkraft neben der Fähigkeit, diese schnell zu erkennen, auch eines breiten mathematischen Wissens, um erklären zu können, was dem Fehler zugrunde liegt. Dabei müssen die Lehrkräfte den Schülerfehler erkennen, diesen einordnen und entscheiden, ob der Fehler für die ganze Lerngruppe von Bedeutung ist, ob es sich um einen Flüchtigkeitsfehler handelt, bei dem keine Fehlvorstellung vorliegt, oder um einen systematischen Fehler. Beutelspacher (2008) betont, dass anhand der auftretenden Fehler zu erkennen sei, wie das Denken funktioniere, im Englischen wird dies häufig als Mathematical Thinking bezeichnet (Sherin & van Es 2002).

Für das Erkennen der Schülerfehler, die in 75% der Fälle im Mathematikunterricht echte Lerngelegenheiten bieten (vgl. Schoy-Lutz 2005, S. 333), verbleibt der Lehrkraft nur wenig Zeit. Dabei ermöglicht erst das schnelle Identifizieren der Fehler auch das Erkennen und die Nutzung dieser Lerngelegenheiten. Herppich, Altmann, Wittwer und Nückles (2017) heben hervor, dass das Verständnis, das einzelne Schüler(innen) zu haben scheinen, ständig aktualisiert wird und der Unterricht an die Lernvoraussetzungen laufend anzupassen ist, damit günstige Lernbedingungen hergestellt werden (vgl. Brühwiler 2017, S. 124). Eine wichtige Facette des Wissens der Mathematiklehrkräfte stellt das Wissen über Fehlkonzepte bzw. Fehlvorstellungen der Schüler(innen) dar (vgl. Altmann & Nückles 2017). In dem Test zur schnellen Fehlererkennung, der im Rahmen der Studie „**T**eacher **E**ducation and **D**evelopment **S**tudy in Mathematics – **F**ollow **U**p“ TEDS-FU (Details in Blömeke et al., 2014) eingesetzt wurde, um eine Facette der Diagnosekompetenz zu erheben, wurde die Kompetenz zur schnellen Fehlererkennung von angehenden Mathematiklehrkräften (im Folgenden auch: Junglehrkräfte) erfasst.

2. Ziele der Studie und Gliederung der Arbeit

Übergeordnetes Ziel der dieser publikationsbasierten Dissertation zugrundeliegenden Studie ist die detaillierte Untersuchung einer Messung der Kompetenz zur schnellen Fehlererkennung. Dazu wurde im Rahmen der Studie TEDS-Follow Up (TEDS-FU) – einer

[1] Anstelle von Schülerinnen und Schüler wird im Folgenden die Kurzform Schüler(innen) verwendet. Gemeint sind alle Geschlechter.

Nachfolgestudie der internationalen Vergleichsstudie TEDS-M zur Lehrerausbildung – ein Test zur schnellen Fehlererkennung entwickelt, der diese Facette professioneller Kompetenz erfassen soll. Grundlage dieser Dissertation sind drei im Peer-Review-Verfahren veröffentlichte Zeitschriftenbeiträge, die sich mit den oben dargelegten Forschungsfragen auseinandersetzen und diese beantworten.

Der erste Artikel „Early Career Teachers' ability to focus on typical students‘ errors in relation to the complexity of a mathematical topic" erschien 2016 in der Zeitschrift *ZDM Mathematics Education* und untersucht die Fragestellung nach dem Zusammenhang von Itemkomplexität und Antizipationszeit in Relation zu möglichen Fehlern, bevor das Item gezeigt wird. Hieraus resultiert folgende erste Fragestellung der Dissertation: Erkennen Junglehrkräfte ein inhaltlich komplexes Item und welche Zusammenhänge gibt es zu ihrer Antizipationszeit?

Der zweite Artikel „Ein zeitbeschränkter Test zur schnellen Erkennung von Schülerfehlern durch Junglehrkräfte – Qualitative Merkmale zur Schwierigkeitsbestimmung von Items ", der 2018 in der Zeitschrift *mathematica didactica* veröffentlicht wurde, thematisiert die zweite und die dritte Fragestellung: Welche schwierigkeitsgenerierenden Merkmale für die Items können in dem Test zur schnellen Fehlererkennung identifiziert werden? Sind diese Merkmale im Test stabil? Die Bearbeitung beider Fragestellungen kann als eine erste Validierung angesehen werden.

Der dritte Beitrag „Perception of student errors under time limitation: are teachers faster than mathematicians or students?", der ebenfalls 2018 in der Zeitschrift *ZDM Mathematics Education* erschienen ist, validiert den Test zur schnellen Fehlererkennung mittels Kontrastgruppen und beantwortet somit die vierte Forschungsfrage: Kann mittels Kontrastgruppenvergleich die Validität des Tests bestimmt werden?

Im Folgenden wird das Ziel der kumulativen, publikationsbasierten Dissertation dargelegt, die somit in vier Fragestellungen differenziert wird. In dem anschließend explizierten theoretischen Rahmen werden die elementaren Begriffe und Zusammenhänge dargestellt, die den Publikationen zugrunde liegen. Dem schließt sich der Abdruck der Publikationen an, die den Kern der Arbeit bilden, sowie eine Spezifizierung des diesbezüglichen eigenen Anteils. Nach der Beantwortung der genannten Fragestellungen werden die Grenzen des Instruments aufgezeigt und es wird ein Ausblick auf weitere Forschungsoptionen gegeben.

3. Theoretischer Rahmen

3.1. Professionelle Kompetenz von Lehrkräften

Die Erforschung der professionellen Kompetenz von Lehrkräften basiert auf der Annahme, dass eine hohe Kompetenz von Lehrkräften einen positiven Einfluss auf die Unterrichtsqualität hat, was beispielsweise in der COACTIV-Studie, aber auch im Rahmen des Learning Mathematics for Teaching Project (LMT) (Baumert & Kunter 2006; Ball & Bass 2002) gezeigt werden konnte. Ein hoher Ausbildungsgrad professioneller Kompetenz erscheint demzufolge unabdingbar. Hierbei stellt sich unmittelbar die Frage nach der konkreten Definition des Begriffs, die für die Verwendung im Rahmen dieser Arbeit nachstehend kurz festgelegt wird.

Der Begriff der Kompetenz soll nach Weinert als „die bei Individuen verfügbaren oder durch sie erlernten kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können“ (Weinert 2001, S. 27f) verstanden werden. Damit ist hier nicht von „trägem Wissen“ (vgl. Gruber et al. 1999) die Rede, vielmehr soll zielgerichtet Wissen aufgenommen werden, das geeignet ist, um Probleme zu lösen. Dieses zielgerichtet aufzunehmende Wissen soll im Folgenden näher betrachtet werden.

Die Facetten des professionellen Wissens, die der hier verwendeten Studie TEDS-FU, einer Folgestudie der internationalen Vergleichsstudie TEDS-M zu Kompetenzen von zukünftigen Mathematiklehrkräften am Ende ihrer Ausbildung, zugrunde liegen (Blömeke, Kaiser & Lehmann 2010), basieren auf der Definition von Shulman (1986) zu Professionswissen von Lehrkräften. Shulman unterscheidet zwischen dem fachlichen Wissen, dem pädagogischen Wissen und dem fachdidaktischen Wissen. In der vorliegenden Arbeit liegt der nachgeordnete Fokus auf dem pädagogischen Wissen, im Vordergrund steht die Funktion der übrigen Wissensfacetten, die unter Rekurs auf Shulman detailliert erörtert werden. Zu dem fachlichen Wissen, dem Content Knowledge (CK), gehört nach Shulman (1986) das Wissen, das den Inhalt umschließt und sich zugleich nicht auf reines Faktenwissen reduzieren lässt. Im gegebenen Kontext wird diese Facette bzw. Wissenskomponente als Mathematical Content Knowledge (MCK) bezeichnet. Weiterhin unterscheidet Shulman (1986) das fachspezifisch-pädagogische Wissen, Pedagogical Content Knowledge (PCK), das hier als Mathematical Pedagogical Content Knowledge (MPCK) gefasst wird. Dieses Wissen betreffe „ […] the most regularly taught topics in one’s subject area, the most useful forms of representation of those ideas, the most powerful analogies, illustrations, examples, explanations, and demonstrations – in a word, the ways of representing and formulating the subject that make it comprehensible to others“ (Shulman 1986, S. 9). Ergänzend ist mit Shulman hervorzuheben, dass die Lehrkraft in Situationen, in denen sie mit Fehlern von Schüler(innen) konfrontiert ist, über Strategien zur Behebung der Fehlvorstellungen verfügen muss, um das Wissen der Schüler(innen) reorganisieren zu können (Shulman 1986, S. 9f).

Eine weitere Ausdifferenzierung der Facetten fachliches Wissen sowie fachdidaktisches Wissen wurde im Rahmen des Projekts Mathematics Teaching and Learning to Teach (MTLT) von Loewenberg Ball, Thames und Phelps (2008) vorgenommen. Ihnen zufolge ist insbesondere das Erkennen von Fehlern, das dem Specialized Content Knowledge (SCK) zugeordnet wird, bedeutsam, in dem das Wissen über die Ursachen typischer Fehler zusammengefasst wird, sowie das Knowledge of Content and Students (KCS), in dem das Wissen über das Lehren von Mathematik und seine speziellen Fehler subsumiert ist (vgl. Loewenberg Ball et al. 2008, S. 401). In Übereinstimmung mit den Ausführungen von Loewenberg Ball et al. (2008) ist bei der Definition professionellen Wissens bzw. professioneller Kompetenz von Lehrkräften zu berücksichtigen, dass Fehleranalysen unter Zeitdruck zu ihren charakteristischen Aufgaben gehören, in denen sich ihre Tätigkeit deutlich von anderen Professionen, insbesondere von Mathematiker(innen), abgrenzt (vgl. Loewenberg Ball et al. 2008, S. 403).

3.2. Die Studie TEDS-FU als Rahmen der eigenen Studie zur schnellen Fehlererkennung

Die vorliegende Dissertation entstand im Rahmen der Studie TEDS-FU, einer Nachfolgestudie der internationalen Lehrerbildungsstudie TEDS-M (Teacher Education and Delevopment Study in Mathematics), die 2008 in 17 Ländern durchgeführt wurde (für Details siehe Blömeke, Kaiser & Lehmann 2010). TEDS-M verfolgte das Ziel, die professionelle Kompetenz von angehenden Lehrkräften zu erfassen. Hierfür wurden zwei inhaltlich unterschiedliche, aber strukturgleiche Tests sowohl für die Primar- als auch für die Sekundarstufe entwickelt. In Deutschland nahmen an TEDS-M ca. 2000 Referendar(innen) teil, die sich am Ende ihrer Lehrerausbildung befanden.

In der Nachfolgestudie TEDS-FU wurden ehemalige Teilnehmende der Studie TEDS-M erneut zur Mitwirkung eingeladen. Im Testzeitraum 2012 nahmen insgesamt 304 Primar- und Sekundarstufen-Lehrkräfte teil. Die Teilnehmer(innen) verfügten zum Testzeitpunkt über ca. vier Jahre Berufserfahrung. Die Stichprobe der Primarstufen-Lehrkräfte besteht aus 133 Lehrkräften, während 171 Lehrkräfte an der Sekundarstufentestung teilnahmen. Im Fokus der Arbeit steht das Sample der Sekundarstufe, bestehend zu 55,8% aus Gymnasiallehrkräften und zu 43,6% aus Haupt- bzw. Realschullehrkräften. Die Mathematiklehrer(innen) sind durchschnittlich 32 Jahre alt (SD=5,9; min=26; max=53) und zu 58,9% weiblich. 73,3% der Teilnehmer(innen) gaben an, einen Mathematikleistungskurs besucht und das Abitur mit der Durchschnittsnote 2,1 (SD=0,6; min=1,0; max=3,9) absolviert zu haben (für Details siehe Blömeke et al. 2014).

Besonders interessant war die Betrachtung der Entwicklung der Kompetenz der Junglehrkräfte während der ersten Jahre der Lehrtätigkeit. Die Studie TEDS-FU wurde aufgrund der Verteilung der Teilnehmer(innen) über die gesamte Bundesrepublik online mithilfe des CBA ItemBuilder (Rölke 2012) durchgeführt. Um die Kompetenz der Junglehrkräfte zu spezifizieren, wurde mit teilweise gekürzten, aber auch erweiterten Testteilen der TEDS-M-Studie erhoben (für Details siehe Blömeke et al. 2014). Darüber hinaus wurden neue Testinstrumente eingesetzt, um die situationsbezogenen Fähigkeiten von Mathematiklehrkräften zu erfassen. So wurden drei kurze Videovignetten gezeigt und bezüglich der gezeigten Unterrichtssequenzen spezifische Fragen gestellt, die es zu beantworten galt (für Details siehe Blömeke et al. 2014). Der Einsatz dieser Unterrichtsvideos diente dazu, die Wahrnehmung und die Interpretation von mathematikbezogenen

Unterrichtssituationen sowie die Reaktion und die Fortführung des Unterrichts durch die Mathematiklehrkräfte präzise zu erheben und hinsichtlich der Zielangemessenheit auszuwerten (vgl. Blömeke 2015). Die Studie folgt somit dem Kompetenzmodell von Blömeke et al. und fasst Kompetenz als Kontinuum (2015). Das Konzept von TEDS-FU bietet die Möglichkeit einer Evaluation der professionellen Kompetenz, ergänzt durch „the original purely cognitive and affective facets of the professional competence of teachers by perceptual, interpretative and decision-making skills" (Kaiser et al. 2015, S. 373). Einen weiteren Aspekt liefert der Beitrag von Kaiser et al. (2015) mit dem theoretischen Bezug zum Teacher Noticing Concept, der von Blömeke et al. (2015) konkretisiert wird, indem sie die situationsbezogenen Fähigkeiten der Wahrnehmung von Unterricht mittels dreier Facetten beschreiben: perception, interpretation und decision making – kurz gefasst als PID-Modell. Anhand des PID-Modells, das den TEDS-FU-Studien zugrunde liegt, lassen sich die situationsspezifischen Kompetenzen wie folgt beschreiben: „(a) Perceiving particular events in an instructional setting, (b) Interpreting the perceived activities in the classroom and (c) Decision-making, either as anticipating a response to students' activities or as proposing alternative instructional strategies" (Kaiser et al. 2015, S. 373). Zur Frage der Wahrnehmung von Schülerfehlern, vgl. Abschnitt 3.3., wird auf diese Konzeptualisierung genauer eingegangen. Der Test zur schnellen Fehlererkennung wird im Rahmen des Kapitels 3.4. (Methodisches Vorgehen) erläutert.

3.3. Fehler im Mathematikunterricht und ihre Wahrnehmung

Oser et al. (1999) definieren einen Fehler als Abweichung von der Norm und fügen hinzu, dass es ohne dieses normative Bezugssystem nicht möglich wäre, zwischen richtig und falsch zu unterscheiden (vgl. Oser et al. 1999, S. 11). Heinze (2004) spezifiziert die Definition wie folgt: „Ein Fehler ist eine Äußerung, die gegen die allgemeingültigen Aussagen und Definitionen der Mathematik sowie gegen allgemein akzeptiertes mathematisch-methodisches Vorgehen verstößt" (Heinze 2004, S. 223). Schülerfehler können mehrere Gründe haben, die der Flüchtigkeit oder auch der Nichtbeachtung der Regeln zugeordnet werden können. Viele Schülerfehler sind systematischer Art, indem sie in strukturgleichen Aufgaben wiederholt auftreten. In diesem Fall ist eine adäquate Intervention seitens der Lehrkraft nötig (Swan 2004), um der Gefahr der regelhaften Wiederholung zu begegnen (Türling 2014).

Des Weiteren muss die Lehrkraft einschätzen, inwieweit der wahrgenommene Fehler im Unterrichtsgespräch mit der ganzen Lerngruppe oder in einem eins zu eins mit dem oder der betreffenden Schüler(in) zu klären ist. Schoy-Lutz (2005) konnte zeigen, dass 75% der Fehlersituationen Lerngelegenheiten bieten. Gleichzeitig wird hervorgehoben, dass erst ein fehlerfreundliches Unterrichtsklima Voraussetzungen schafft, um die eine Fehlersituation als Lernsituation nutzen zu können (Wuttke & Seifried 2017, S. 4) und dass zugleich zwischen Lern- und Leistungssituationen zu trennen ist (Oser & Spychiger 2005).

Für ein genaues Begriffsverständnis ist ein Sprachvergleich aufschlussreich: Während dt. Fehler ein weites Bedeutungsspektrum umfasst, wird im englischen Sprachgebrauch nach spezifischen Ursachen differenziert und error (regelhafter Fehler), slip (Flüchtigkeitsfehler) und mistake (Verwechslungsfehler) unterschieden (für Details siehe Steuer 2014). Mangels dieser Differenzierung ist der Fehler per se stark negativ konnotiert, grundsätzlich gelten

Fehler im Unterricht als unangenehm und potenziell bedrohlich für den Lernprozess (vgl. Blanck 2006). Andererseits jedoch können Schülerfehler als Indikatoren des Status quo betrachtet werden, die kenntlich machen, wo sich die Schüler(innen) gerade „in ihrem Denken“ befinden (vgl. Beutelspacher 2008). Wichtig ist in diesem Zusammenhang, dass nicht die Perspektive der Defizitorientierung eingenommen wird, sondern dass Fehler per se positiv konnotiert als Lerngelegenheit betrachtet werden.

Dies setzt seitens der Lehrkraft die Fähigkeit zum *schnellen* Erkennen von Schülerfehlern voraus, die daher eine wichtige Facette der professionellen Kompetenz von Lehrer(innen) bildet. Seidel und Prenzel (2003) stellen das schnelle Erkennen als notwendige Voraussetzung dar, um auf den Fehler adäquat eingehen zu können. Die zeitnahe Reaktion der Lehrkräfte ist unabdingbar, um den unmittelbar konsekutiven Lernprozess der Schüler(innen) optimal zu unterstützen. Des Weiteren ist das schnelle Erkennen von Schülerfehlern elementar, um Schülerdokumente zu evaluieren und mit stringent nachfolgenden Aufgaben Impulse setzen zu können (Brühwiler 2017). Zusammenfassend kann die Fähigkeit, Schülerfehler schnell zu erkennen, als Teil der diagnostischen Kompetenz von Lehrkräften bezeichnet werden (vgl. Leuders, 2001; Schumacher, 2008).

Basierend auf der aktuellen Diskussion um Lehrerkompetenz und die Notwendigkeit, auch die Performanz der Lehrkräfte genau zu erfassen, konzeptualisieren Blömeke et al. (2015) Kompetenz als ein Kontinuum (vgl. 3.2.). Mit dem Performanzbegriff wird die situative Komponente von Lehrerkompetenzen und insbesondere die Komponente der Wahrnehmung – hier die Phase der Wahrnehmung von Schülerfehlern – explizit einbezogen. Die Komponente der Wahrnehmung kann nicht zuletzt angesichts der hohen Bedeutung von Schülerfehlern als wichtige Komponente der Situationskompetenzen von Lehrer(innen) und im engeren Sinne auch als Bestandteil der diagnostischen Kompetenz angesehen werden.

Die diagnostische Kompetenz ist verbunden mit MCK und MPCK, da dieses Wissen für die Wahrnehmung und Interpretation von Schülerfehlern sowie für die Entwicklung adäquater Lehrmaßnahmen (für einen Überblick über die aktuelle Diskussion siehe Südkamp & Praetorius 2017; Leuders et al. 2018) benötigt wird. Ein entscheidendes Element der Wahrnehmung von Schülerfehlern ist die Schnelligkeit der Fehlerwahrnehmung unter Zeitbeschränkungen im Kontext der Unterrichtspraxis, in der Lehrer(innen) die Aussagen bzw. Lösungsvorschläge der Schüler(innen) in kürzester Zeit analysieren müssen (Lindmeier et al. 2013, S. 106). Auch in der Evaluation der Schülerdokumente ist das Umgehen mit diesen zeitlichen Einschränkungen Teil des Berufslebens der Lehrkräfte und unverzichtbarer Bestandteil von professionellen Lehrerkompetenzen (Wahl et al. 1984).

Bromme (1997) bezeichnet die diagnostische Kompetenz als eine relevante Facette für erfolgreiches Lehrerhandeln und betont die besondere Rolle, die der Lehrer(innen)wahrnehmung zukommt. McElvany et al. (2009) stützen diese Aussage und akzentuieren die Bedeutung der diagnostischen Kompetenz mit Blick darauf, dass die Ausrichtung des Unterrichts auf die Lernvoraussetzungen der Schüler(innen) als zentraler Aspekt der Lehrerkompetenz gelten muss.

Im Sinne des MCK muss die Lehrkraft zudem auch über das mathematische Wissen sicher verfügen, um den Fehler nachvollziehen zu können. Die Bedeutung des mathematischen Wissens bringen Reiss und Hammer (2013) auf den Punkt:

> „Fachwissen ist zwingend notwendig, um die manchmal kompliziert erscheinenden Ideen und Argumentationen von Kindern nachzuvollziehen, ihren sachlichen Gehalt zu erkennen und angemessene Unterstützung daraus ableiten zu können. [...] Man darf sicher verallgemeinern: Fachliches Wissen ist unverzichtbar, wenn es darum geht, einen Fehler zu erkennen“ (Reiss & Hammer 2013, S. 116).

Im Sinne des MPCK ist ebenso mathematikdidaktisches Wissen (vgl. Abb. 1) notwendig, um im Unterricht generell und im Besonderen nach der Fehlerwahrnehmung angemessen reagieren zu können. Hier formulieren Reiss und Hammer (2013) prägnant, dass „diagnostische Fähigkeiten [...] also genauso auf fachdidaktischem wie auf fachlichem Wissen“ (Reiss & Hammer 2013, S. 117) basieren. Kontrastierend betonen Prediger, Wessel, Tschierschky, Seipp und Özdil (2013), dass neben mathematikdidaktischem Wissen die Diagnosefähigkeit notwendig sei, um die Anpassung der inhaltlichen Handlungsfähigkeit sicherzustellen (Prediger et al. 2013, S. 172).

Dies führt zu dem Anliegen zu untersuchen, in welche Phasen die Wahrnehmung von Schülerfehlern im Einzelnen strukturiert werden kann. Ausgehend von der allgemeinen Diskussion diagnostischer Kompetenzen als Teil der Lehrerkompetenz und der Frage, auf welche Weise Lehrerkompetenz im Rahmen der Lehrerausbildung ganzheitlich gefördert werden kann, entwickelten Heinrichs (2015) und Hoth (2016) gemeinschaftlich ein Modell zur Wahrnehmung und zu dem Umgang mit Fehlern im Mathematikunterricht. Sie beschrieben den Diagnoseprozess in Unterrichtssituationen, in denen Fehler von der Lehrkraft wahrgenommen werden, als zyklischen Prozess. Der Kreislauf beginnt demnach mit der Wahrnehmung eines Schülerfehlers und setzt sich fort mit der Entwicklung von Hypothesen über seine Ursache(n). Mögliche Interventionsansätze bilden den nächsten Schritt für den Umgang mit dem Fehler, wobei abschließend durch eine strukturell äquivalente Aufgabe geprüft wird, ob der/die Schüler(in) die Fehlvorstellung überwunden hat (siehe auch Hoth et al. 2016). Ist dies noch nicht der Fall, muss der zyklische Vorgang ein weiteres Mal durchgeführt werden. In dieser Arbeit steht die Wahrnehmung des Fehlers im Fokus, hier verstanden als die erste Phase des Diagnoseprozesses.

Die erste Phase der Diagnose in einer Fehlersituation ist zunächst das Wahrnehmen einer Abweichung von der erwarteten Norm. Bei Reisman (1976) – auf die sich sowohl Heinrichs (2015) als auch Hoth (2016) beziehen – wird diese Phase als Identification bezeichnet, da sie sich auf das Wahrnehmen und Analysieren von Verhaltensergebnissen bezieht (vgl. Reisman 1976, S. 6). Die Notwendigkeit der Schnelligkeit der Identifizierung von Fehlern, um die damit zusammenhängenden Fehlvorstellungen im Lernprozess produktiv nutzen zu können, wurde zu Beginn des Abschnitts bereits ausführlich dargelegt. Nachstehende Abbildung führt die bis hierhin erläuterten theoretischen Konstrukte zur Wahrnehmung von Schülerfehlern zusammen:

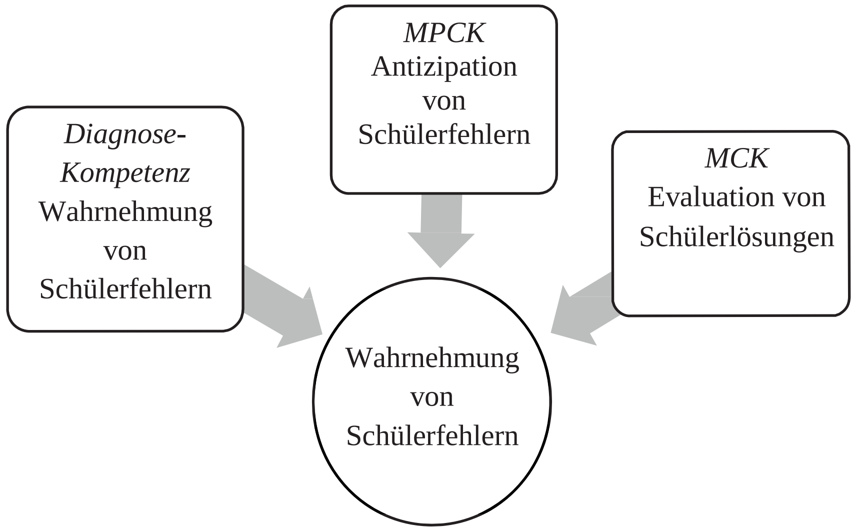


Abbildung 1: Theoretische Konstrukte zur Wahrnehmung von Schülerfehlern (eigene Darstellung).

Zusammenfassend beruht die Wahrnehmung von Schülerfehlern auf drei Kompetenzen bzw. Wissensfacetten (vgl. Abbildung 1): Auf Basis der Diagnose-Kompetenz wird der Fehler erkannt; MPCK ermöglicht die Antizipation des Fehlers auf Basis von mathematikdidaktischem Wissen über den Fehler und seinen (möglichen) Entstehungskontext. MCK ermöglicht die Evaluation der Schülerlösung als „richtig“ oder „falsch“ und optional als Lerngelegenheit.

3.4. Methodisches Vorgehen bei der Testentwicklung und Abgrenzung verwandter Studien

Im Folgenden wird das Design des Tests zur schnellen Erkennung von Schülerfehlern im Detail dargestellt und von dem Design verwandter Studien abgegrenzt. Wie bereits erläutert, ist die Kompetenz zur schnellen Fehlererkennung als eine Facette der Lehrerkompetenz zuzuordnen. Im Folgenden wird die Konzeption des Tests kurz skizziert. Jedes der 16 Items pro Test für die Primarstufe bzw. Sekundarstufe folgt dem gleichen Ablauf und bietet so eine gleichmäßige Erhebung der Fehlererkennungskompetenz. Durch die Verteilung der Teilnehmer(innen) über die Bundesrepublik wurde, wie in Abschnitt 3.2. dargestellt, internetbasiert getestet.

Das Ziel der Testfacette „schnelles Erkennen von Schülerfehlern“ bestand darin, die maximale Leistungsfähigkeit der Testteilnehmer(innen) im Erkennen von typischen Schülerfehlern herauszufinden. Im Rahmen des Designs des zeitbegrenzten Tests wurde die Bearbeitungszeit bewusst knapp angesetzt (Bortz & Döring 2006, S. 453), um die Ausprägung der Fähigkeit zur Erkennung von Schülerfehlern analog zur realen Unterrichtspraxis messen zu können. Das Zeitlimit bis zur Antwort, das aus jedem Item hervorgeht, wurde mittels eines Expertenratings bestehend aus 15 Mathematikdidaktiker(inne)n und erfahrenen Lehrkräften auf vier Sekunden pro Item beschränkt. Mittels der Zeitbeschränkung sollte auch verhindert werden, dass die Items, die von allen Teilnehmer(innen) bei ausreichender Zeit hätten gelöst werden können, tatsächlich nachgerechnet werden konnten (Bühner 2011, S. 21).

Als Leistungsindikator wurde die Anzahl der richtigen Antworten herangezogen. Items, die nicht innerhalb des Zeitlimits beantwortet wurden, wurden als falsch gewertet. Da die Zeitspanne bis zur Antwort aus technischen Gründen nicht gemessen werden konnte, wird nur zwischen richtig/falsch unterschieden. Alle Items sind neben der identischen Zeitbeschränkung durch eine unterschiedliche Schwierigkeit aufgrund ihrer Komplexität gekennzeichnet (Pankow et al., 2016). Nach Moosbrugger und Kelava (2012) handelt es sich aufgrund der unterschiedlichen Schwierigkeit deshalb nicht nur um einen Speedtest, sondern ebenso um einen Power-/ bzw. Niveautest (vgl. Moosbrugger & Kelava 2012, S. 30).

Die Beantwortung der Items wird nachfolgend anhand der Beschreibung eines Items exemplarisch dargestellt (vgl. Abb. 2; weiteren Items siehe Abschnitt 4, 5 & 6).

Beispielitem	1. Phase	Ankündigung des Themas mit der Aufforderung, typische Schülerfehler zu antizipieren (≤ 5 min)	Bruchrechnung: Addition zweier Brüche		
	2. Phase	Identifizierung der falschen Schülerlösung (≤ 4 s)	*s* $\frac{1}{2}+\frac{1}{4}=\frac{3}{4}$	*d* $\frac{2}{7}+\frac{3}{7}=\frac{5}{14}$	*f* $\frac{1}{3}+\frac{2}{3}=1$

Abbildung 2: Design der Itemstruktur (kursiv: diese Angaben haben die Lehrkräfte in der Testung gesehen) [„d“ ist die falsche Antwort]

In der ersten Phase wird das Themengebiet mitgeteilt, dem die in der Folgephase präsentierte Aufgabe/falsche Schülerlösung entstammt. Thematisch sind die Items in dem Gebiet der Arithmetik und der Algebra sowie in kleinerem Umfang in der Geometrie und der Stochastik verortet, die in der Sekundarstufe 1 klassischerweise curricular verankert sind. Während der Antizipation, die bis zu fünf Minuten dauern konnte, waren die Teilnehmer(innen) explizit gefordert, auf ihre Erfahrungen mit typischen Schülerfehlern zurückzugreifen. Die Zeit, die die Teilnehmer(innen) auf der Seite des eingeblendeten Themenfeldes verblieben, wurde gemessen und wird im Folgenden als Antizipationszeit bezeichnet. Dieses Testdesign wurde bewusst gewählt, um sich stärker als sonstige Assessments der Realität in Klassenräumen anzunähern.

In der zweiten Phase werden den Teilnehmer(innen) drei Schülerlösungen gezeigt, wobei zwei Schülerlösungen richtig sind und eine Lösung als typischer Schülerfehler zu identifizieren ist. Durch das eingesetzte Onlinetool war es möglich, die zeitliche Limitierung in das Instrument zu integrieren. Nach Ablauf der vier Sekunden und ohne eine Eingabe über die Tastatur wurde das Item als falsch gewertet. Um einen sonstigen Zeitverlust zu verhindern, wurden die Teilnehmer(innen) zu Beginn der Testung gebeten, die Finger ihrer linken Hand auf die Tastatur in der Zehn-Finger-Tippposition zu legen, insbesondere auf die Buchstaben „s“, „d“ und „f“.

In der Mathematikdidaktik liegen einige Studien vor, die ähnliche Ziele verfolgten und einen ähnlichen Ansatz gewählt haben. Von diesen Studien soll nun die eigene Studie abgegrenzt werden.

Auch Krauss und Brunner (2011) konnten in ihrem Reaktionszeittest zur Fehlererkennung Erkenntnisse hinsichtlich des schnellen Erkennens von Schülerfehlern generieren. Nach ihren Messungen lag die in dem Reaktionstest festgestellte Lesezeit der Items in etwa bei 3 bis 6 Sekunden (Krauss & Brunner 2011, S. 241). Hinsichtlich der möglichen Korrelation des Testerfolgs und des Alters der Lehrkräfte konnte keine Signifikanz festgestellt werden, lediglich eine negative Tendenz, die sich durch die abnehmende Reaktionsgeschwindigkeit mit zunehmendem Alter erklären lässt (Krauss & Brunner 2011, S. 245). Weiterhin ließ sich ein signifikanter Unterschied zugunsten der Gymnasiallehrkräfte erkennen, die schneller reagierten und mehr richtige Lösungen produzierten (Krauss & Brunner 2011, S. 245). Die Autoren weisen darauf hin, dass die Gründe hierfür vielfältig sein können. Außerdem konnten sie einen signifikanten Zusammenhang zwischen dem Fachwissen, das in der COACTIV-Studie erfasst wurde, und dem Abschneiden im Fehlererkennungstest herausarbeiten (Krauss & Brunner, 2011, S. 246).

Um die Ergebnisse der COACTIV-Studie und der TEDS-FU-Studie genauer einordnen zu können, ist es aufschlussreich, die jeweiligen Tests unter einigen Aspekten zu vergleichen.

An der COACTIV-Studie nahmen 185 Personen mit einem Altersdurchschnitt von 47 Jahren teil (min. 28 J., max. 65 J.), bei der TEDS-FU-Studie waren es 137 Personen mit einem Altersdurchschnitt von 29 Jahren (min. 21 J., max. 50 J.). Während es sich bei den COACTIV-Teilnehmer(innen) um erfahrene Lehrkräfte handelte, verfügten die Teilnehmer(innen) bei TEDS-FU als ehemalige Teilnehmer(innen) von TEDS-M über drei bis vier Jahre Lehrerfahrung nach dem Referendariat.

Der COACTIV-Fehlererkennungstest besteht aus 12 Items, der TEDS-FU-Test zur schnellen Fehlererkennung aus 16 Items. Beide Tests beginnen mit einem einführenden Beispiel. Bei dem COACTIV-Test werden die Aufgabenstellung, die Schülerlösung und auch die Bewertung, ob eine Schülerlösung richtig oder falsch ist, auf einer Seite präsentiert. Bei dem TEDS-FU-Test hingegen werden auf einer ersten Seite das Thema des Schülerfehlers und auf einer zweiten Seite die drei Antwortmöglichkeiten gezeigt. Während bei COACTIV entschieden werden muss, ob die präsentierte Antwort richtig oder falsch ist, muss bei dem TEDS-FU die falsche Schülerlösung innerhalb von drei Lösungen identifiziert werden. Weitere Unterschiede beziehen sich auf das technische Verfahren der Bearbeitung: Während bei COACTIV mit dem Cursor der Maus das gewählte Feld angeklickt werden muss, ist bei TEDS-FU die zu der Schülerlösung zugehörige Buchstabentaste per Tastatur einzugeben. Die Ratewahrscheinlichkeit liegt bei 50% für COACTIV und bei 33% für TEDS-FU. Diese Wahrscheinlichkeit verändert sich während der beiden Tests nicht, sondern bleibt aufgrund der Beschaffenheit der Tests gleich (vgl. Tabelle 1), worauf das Problem des Anordnungsbias zurückzuführen ist.

Tabelle 1: Vergleich des Testaufbaus von COACTIV und TEDS-FU

COACTIV	TEDS-FU
Aufgabe: $\frac{a^4 + a^3}{a^2} =$ Schülerlösung: $a^2 + 1$ richtig falsch Die Schülerlösung ist falsch! Die richtige Lösung lautet: $a^2 + a$ Sie haben 2.2 Sek. für Ihre Antwort benötigt. Bitte klicken Sie auf weiter, um zur nächsten Aufgabe zu gelangen. >>weiter	*Addition von Brüchen* weiter s d f $\frac{1}{2} + \frac{1}{4} = \frac{3}{4}$ $\frac{2}{7} + \frac{3}{7} = \frac{5}{14}$ $\frac{1}{3} + \frac{2}{3} = 1$ Die FALSCHE Antwort ist

Inhaltlich behandelt der COACTIV-Test eine Mischung verschiedener Themengebiete aus der Arithmetik und der Algebra (vgl. Krauss & Brunner 2011, S. 239), die ein Mindestmaß an Rechenaufwand erfordern und somit eine „Neuartigkeit“ (Krauss & Brunner 2011, S. 238) enthalten. Die Items des TEDS-FU-Tests umfassen typische Schülerfehler, die durch die vorherige Einblendung des Themas eine Auseinandersetzung mit dem Themengebiet initiieren.

Nach der Eingabe wird bei COACTIV sofort angezeigt, ob das Item richtig oder falsch gelöst wurde und welche Bearbeitungszeit benötigt wurde. Um mit diesem Feedback verbundene (Lern)effekte auszuschließen, war bei dem TEDS-FU-Test nach der Eingabe der Lösung lediglich das Feld „weiter“ anzuklicken. War die Zeit zur Eingabe der Antwort abgelaufen und es war keine Antwort erfolgt, erschien ein Feld mit der Mitteilung „Die Zeit ist abgelaufen“.

Beide Tests wurden computerbasiert durchgeführt. Während bei COACTIV eine Testleitung zur Verfügung stand, fanden die Tests im Rahmen der TEDS-FU-Studie an einem selbstgewählten Ort statt, wobei nur ein Internetanschluss bestehen musste.

Zusammenfassend unterscheiden sich die Ansätze von COACTIV und TEDS-FU zentral darin, dass bei TEDS-FU das Themengebiet bzw. die Aufgabe mit der Aufforderung, typische Fehler zu antizipieren, vorher eingeblendet wurde. Auf diese Weise wurde intentional gesteuert eine stärkere Nähe zur Unterrichtswirklichkeit geschaffen, da sich auch eine Lehrkraft auf das Thema der geplanten Unterrichtsstunde eingestellt hat.

3.5. Weitere Studien zur schnellen Wahrnehmung von Schülerfehlern

Die Kompetenz von Lehrkräften im Hinblick auf die Wahrnehmung von Schülerfehlern – oder allgemeiner auf die Bewertung der Richtigkeit der Schülerlösungen – ist in verschiedenen empirischen Projekten und Studien gemessen worden.

Im Rahmen der Studien Mathematical Knowledge for Teaching (MKT) und Mathematics Teaching and Learning to Teaching (MTLT) konnten Hill et al. (2008) die Wahrnehmung und Analyse häufiger Schülerfehler und die Bereitstellung von Erklärungen dieser Fehler als einen Hauptbestandteil des mathematischen Wissens von Lehrkräften und als wichtige Komponente des Mathematikunterricht identifizieren. Loewenberg Ball et al. (2008) fassten die Wahrnehmung von Fehlern unter den Wissensfacetten Specialized Content Knowledge (SCK) und Knowledge of Content and Students (KCS) zusammen und definierten für die Wahrnehmung von Fehlern zwei Grundlagen: das mathematische Wissen der Lehrkräfte und ihr Wissen über das Wissen der Schüler(innen) und deren Nähe zur Schulmathematik.

KCS wurde im Unterschied zum mathematischen Wissen der Profession der professionellen Mathematiker(innen) als die spezifische Profession von Lehrer(innen) determiniert. Von professionellen Mathematiker(innen) wurde entsprechend erwartet, dass sie mit häufigen Schülerfehlern weniger vertraut sind. Weitere Analysen konnten bestätigen, dass Mathematiker(innen) mit geringerer Wahrscheinlichkeit Items richtig beantworten konnten, die auf „Knowledge of Students“ statt auf „Content Knowledge“ beruhten (Hill et al. 2007). Wie oben bereits angesprochen, betonten Loewenberg Ball et al. (2008) in ihrem Modell, dass zeitlich begrenzte Fehleranalysen zu den charakteristischen Aufgaben von Lehrer(innen) gehören, durch die sich ihre Tätigkeit deutlich von anderen Professionen, speziell von Mathematiker(innen), abgrenzt.

Die COACTIV-Studie analysierte die Bewertung der Lehrer(innen) hinsichtlich der Lösungen der Schüler(innen) für mathematische Aufgaben als „richtig“ oder „falsch“ (Binder et al. 2018), indem die Reaktionszeit der Lehrer(innen) gemessen und mit der Anzahl der korrekt identifizierten Schülerlösungen korreliert wurde. Diese Testkonstruktion ist als schnelle Auswertung elementarer mathematischer Aussagen zu bezeichnen, die sich bei COACTIV im Unterschied zu dem oben diskutierten MKT-Projekt, das sich auf Grundschullehrkräften konzentrierte, auf das Fachwissen von Sekundarstufen-Lehrkräfte bezog. Da bei COACTIV zusätzlich mathematisches Inhaltswissen und mathematisch-pädagogisches Inhaltswissen gemessen wurde, konnte die Autor(inn)engruppe die Korrelationen zwischen den Ergebnissen zu den Reaktionszeiten und den Punktzahlen der Teilnehmer(innen) im Kontext dieser Wissensfacetten herstellen und untersuchen. Die Studie zeigte eine starke Korrelation zwischen der schnellen Bewertung von mathematischen Aussagen und mathematischen inhaltlichen Kenntnissen der Lehrkräfte sowie deren mathematikpädagogischen Kenntnissen (Krauss & Brunner 2011, S. 246f).

Wie das MKT-Projekt analysierte auch COACTIV die Validität seiner Instrumente und stellte im Gegensatz zum MKT-Projekt fest, dass professionelle Mathematiker(innen) bessere

Ergebnisse in den Aufgabenlösungen erzielten, was den starken Zusammenhang zwischen mathematischem Inhaltswissen und der Fähigkeit zur Wahrnehmung von Schülerfehlern verdeutlicht (Krauss et al. 2008).

3.6. Ansätze zur Validierung im Rahmen der Dissertation

Die Erfüllung von Gütekriterien ist als zentrale Anforderung jeder Disziplin obligatorisch zu überprüfen (vgl. Bortz & Döring, 2016). Jenßen, Dunekacke, und Blömeke (2015) bezeichnen die Validität als fundamentales und zugleich komplexestes Gütekriterium, indem es angibt, ob ein Test das misst, was er zu messen vorgibt. Im Vergleich zu den Gütekriterien der Objektivität und Reliabilität ist die Überprüfung der Validität „sehr viel aufwendiger" (Bortz & Döring, 2006, S. 200) und es gibt hierfür kein einheitliches Routineverfahren (vgl. Hartig, Frey & Jude, 2012, S. 145). Nach Bortz und Döring (2006) werden drei Hauptarten von Validität unterschieden: Inhaltsvalidität, Kriteriumsvalidität und Konstruktvalidität. Mit Hartig et al. (2012) ist auf den Wandel im Verständnis der Konstruktvalidität hinzuweisen: Während diese in den Anfängen als Ergänzung der Kriteriums- und Inhaltsvalidität angesehen wurde, wird die Konstruktvalidität seit Ende der 1970er Jahre als übergeordnete Validität verstanden, die die anderen Formen mit einschließt (vgl. Hartig et al. 2012, S. 153, für einen Überblick siehe Newton & Shaw 2016). Im Rahmen der Studie von Lange, Kleickmann und Möller (2012) wurde eine besondere Variante zur Überprüfung der Kriteriumsvalidität angewendet, die „Technik der bekannten Gruppen" (Bortz & Döring 2006, S. 201). Maßgebliches Kriterium ist die Zugehörigkeit zu Gruppen, die mit der Durchführung des Tests herausgearbeitet wird und für die Unterschiede in der Ausprägung des zu messenden Konstrukts erwartet werden (vgl. Schnell, Hill & Esser 2011, S. 148).

Die „Technik der bekannten Gruppen" wurde bereits in einigen Studien (ein Überblick ist im dritten Beitrag dieser Arbeit zu finden) angewendet und wird im Folgenden als Kontrastgruppenvalidierung vorgestellt.

COACTIV hat als erste deutsche Studie die Kontrastgruppenvalidierung (vgl. Krauss et al., 2008) gewählt, indem als Teilnehmer(innen) Mathematiklehrkräfte, Mathematikstudierende, Lehrkräfte der Biologie und Chemie sowie Oberstufenschüler(innen) gewählt wurden. Im Ergebnis erwies sich das vorhandene mathematische Wissen aller Gruppen als vergleichbar hoch, das Wissen der Mathematikstudierenden im Bereich PCK überstieg die Erwartungen. Um diese Ergebnisse erklären zu können, wurde das mathematikdidaktische Wissen differenziert in das Wissen über das Erklären und Repräsentieren, das Wissen über typische Schülerfehler und Schülerschwierigkeiten sowie in das Wissen über das multiple Lösungspotenzial von Mathematikaufgaben (Krauss et al. 2011, S. 138f). Die Analysen ergaben bezüglich des Wissens über das multiple Lösungspotenzial signifikante Unterschiede zwischen den Mathematiklehrkräften und den weitaus kompetenteren Studierenden, die eine weitere Ausdifferenzierung in der weiteren Forschung notwendig macht.

Auch in der US-amerikanischen Studie von Buschang, Chung, Delacruz und Baker (2012) wurde mit dem Fokus auf Sekundarstufenlehrkräfte eine Validierung durchgeführt. Getestet wurden drei Gruppen sogenannter Mathematics Content Experts: Mathematikexpert(inn)en mit einer besonderen Zertifizierung, erfahrene und am Beginn ihrer Laufbahn stehende Lehrkräfte. Entgegen den Erwartungen konnten hinsichtlich MKT keine signifikanten Unterschiede im Vergleich der ersten beiden Gruppen mit der Gruppe der Noviz(inn)en gewonnen werden. Demzufolge wurde deutlich, dass eine ganzheitliche Betrachtung von MKT die Gruppenunterschiede nicht ausreichend abzubilden vermag.

Weiteren Aufschluss zur Frage der Validierung bietet die Studie von Charalambous (2016). In seiner Auswertung weist Charalambous darauf hin, dass weitere Studien erforderlich seien, um zu ermitteln, wie die Lehrkräfte ihre speziellen Kenntnisse einsetzen. Nach seiner Auffassung spielt dabei die Lehrerfahrung die entscheidende Rolle, mit der die unerwarteten Ausnahmen erklärt werden könnten (vgl. Charalambous 2016, S. 231). In dieser Studie konnten keine Unterschiede zwischen Lehramtsstudierenden und praktizierenden Lehrkräften sowie Mathematiker(inne)n in der Analyse von Schülerfehlern und deren Denken festgestellt werden. Mögliche Gründe hierfür wurden anhand von im Rahmen der Studie durchgeführten Interviews identifiziert.
Wichtig erscheint der Hinweis von Charalambous (2016), das Multiple Choice-Verfahren für die Erhebung eher ungeeignet seien. Er schlägt den Einsatz von Videos vor und betont die Notwendigkeit der Entwicklung innovativer Ansätze, um das MKT von Lehrkräften zu verbessern und zugleich mathematisches und kohärentes pädagogische Wissen zu fördern.
Als weitere Gründe für die Studienergebnisse nennt er die Möglichkeit, dass die Teilnehmer(innen) zu richtigen Ergebnissen auch dann gekommen sind, wenn sie die Wege der Schülerlösungen nicht nachvollziehen konnten und folglich nur zu einem mathematisch richtigen Ergebnis kamen. Charalambous (2016) fordert daher andere Ansätze, die das Wissen der Lehrkräfte genauer erfassen können (Charalambous 2016, S. 232). Eine Option wäre die Erfassung über externe Variablen, "[which] may include measures of some criteria that the test is expected to predict, as well as relationships to other tests hypothesized to measure the same constructs, and tests measuring related or different constructs" (AERA, APA & NCME 2014, S. 16). So sollte es möglich sein, die Kriteriumsvalidität abzuschätzen, d.h. zu messen "how accurately the test scores predict criterion performance" (AERA, APA & NCME 2014, S. 17). Die Bedeutsamkeit dieser externen Variablen konnten u.a. Blömeke et al. (2014) zeigen, indem sie einen Zusammenhang zwischen der Abiturnote und dem schnellen Erkennen von Schülerfehlern bzw. zwischen dem mathematischen Wissen und dem mathematikdidaktischen Wissen identifizierten.
Die Ergebnisse der Validierungsstudie von Krauss et al. (2008) konnten durch die Studie von Charalambous (2016) aktuell bestätigt werden. Diese Studie sollte daher als explorative Validierungsstudie verstanden und ihre Hinweise auf das Wissen von Lehrkräften, das für die erforderliche Unterrichtskompetenz grundlegend ist, sollte berücksichtigt werden (vgl. Charalambous 2016, S. 230).

Neben den gezeigten Unterschieden zu dem Testinstrument von COACTIV zeigt der Beitrag von Charalambous (2016) die Forschungslücke, die mit der vorliegenden publikationsbasierten Dissertation geschlossen werden konnte: Gefordert erschien ein Test, der unterrichtsnah (mittels Antizipation) und unter Einbezug von (fehlerhaften) Schülerlösungen das Wissen von Lehrkräften erhebt und – wie von Charalambous angeregt – externe Variablen einbezieht.

4. Publikation I

Die erste Publikation dieser kumulativen, publikationsbasierten Dissertation mit dem Titel „Early Career Teachers' ability to focus on typical students errors in relation to the complexity of a mathematical topic" ist in der Zeitschrift *ZDM Mathematics Education* im Jahr 2016 erschienen, genauer im Themenheft mit dem Titel „Perception, interpretation and decision making: understanding the missing link between competence and performance" herausgegeben von Sigrid Blömeke und Jon R. Star.

Die Publikation fokussiert die Antizipationsphase, die im Rahmen des zeitbegrenzten Tests zur schnellen Fehlererkennung enthalten ist. In der ersten Phase der Beantwortung eines jeden Items wurden die Teilnehmer(innen) aufgefordert, zu einem gegebenen mathematischen Themengebiet typische Schülerfehler zu antizipieren und hierbei auf ihr Wissen zurückzugreifen.

In der theoretischen Fundierung des Beitrags werden der aktuelle Stand der Expertiseforschung und der Unterschied zwischen Expert(inn)en und Noviz(inn)en dargelegt. Expert(inn)en verfügen demzufolge über die Fähigkeit, wahrgenommene Informationen nach relevant und irrelevant zu filtern und bei Bedarf zu fokussieren. Weiterhin wird das Konstrukt der schnellen Fehlererkennung entwickelt und es werden die Zusammenhänge zwischen der Fehlererkennung und dem mathematischen bzw. dem mathematikdidaktischen Wissen verdeutlicht, das Lehrkräfte besitzen müssen, um Schülerfehler identifizieren zu können. Die Forschungsfrage gilt der Korrelation zwischen der gemessenen Antizipationszeit und der Kompetenz, Schülerfehler zu identifizieren. Dabei wird die Hypothese aufgestellt, dass die richtige Antworthäufigkeit mit der Komplexität der Itemankündigung korreliert.

Im methodischen Teil werden die Studie TEDS-FU und die Stichprobe vorgestellt sowie die Testkonstruktion erläutert. Anhand eines Beispielitems wird der Ablauf der Bearbeitung eines Items beispielhaft gezeigt und die unterschiedlichen Phasen der Beantwortung werden verdeutlicht. Mittels eines Expert(inn)enratings werden drei weitere Beispielitems didaktisch begründet und Antizipationsbeispiele dargelegt, ebenso wird die Komplexität der Items bestätigt. In der sich anschließenden Datenanalyse wird die Auswertungsmethode expliziert und es werden die Begrifflichkeiten wie CA (correct answer) und FA (false answer) eingeführt. Die Mittelwerte der Gruppen wurden jeweils in einem Wilcoxon-Mann-Whitney auf Signifikanz geprüft. Weiterhin wurde ein Mediansplit vorgenommen, um einen Gruppenvergleich vornehmen zu können. Dem schloss sich ein itembezogener Mediansplit an, um die CAs mit den FAs zu vergleichen.

Die Ergebnisse des Beitrages lassen sich wie folgt zusammenfassen: Es konnte ein signifikanter Unterschied hinsichtlich der Komplexität der Items festgestellt werden. Die Lehrkräfte, die richtig geantwortet haben, waren offensichtlich in der Lage, die Komplexität des Items zu identifizieren und haben entsprechend für eine längere Zeitspanne antizipiert. Bei den falsch Antwortenden waren die Antizipationszeiten bei komplexen und bei nicht komplexen Items gleich lang, sodass keine Fokussierung ermittelt werden konnte. Insgesamt konnte im Vergleich dieser Messresultate sogar bei den Junglehrkräften ein Unterschied bezüglich der Fähigkeit zum Fokussieren nachgewiesen werden.

4.1. Darlegung des eigenen Anteils

Die Beitragsidee wurde von mir in enger Zusammenarbeit mit meiner Betreuerin der Promotion Gabriele Kaiser entwickelt. Inhaltlich ist der Beitrag in enger Zusammenarbeit mit meinem damaligen Kollegen Andreas Busse, der den Test zur schnellen Fehlererkennung für die Sekundarstufe entwickelt hat, entstanden, wobei die inhaltlichen Ansätze von mir entwickelt wurden. Andreas Busse hat im Zuge der Entwicklung des Testinstruments einen Teil des theoretischen Hintergrundes aus der Expertiseforschung zusammengetragen. Den weiteren theoretischen Rahmen habe ich verfasst, ebenso wie die Testdarstellung und den methodischen Abschnitt. Die Darstellung der erzielten Ergebnisse wurde von der Autorengruppe gemeinschaftlich überarbeitet, basierend auf Textentwürfen von mir. Die finalen Überarbeitungen habe ich im Wesentlichen vorgenommen.

4.2. Abdruck der Publikation I

Pankow, L., Kaiser, G., Busse, A., König, J., Hoth, J., Döhrmann, M., & Blömeke, S. (2016). Early career teachers’ ability to focus on typical students errors in relation to the complexity of a mathematical topic. *ZDM Mathematics Education, 48*(1), 55-67.

ZDM Mathematics Education (2016) 48:55–67
DOI 10.1007/s11858-016-0763-2

ORIGINAL ARTICLE

Early Career Teachers' ability to focus on typical students errors in relation to the complexity of a mathematical topic

Lena Pankow[1] · Gabriele Kaiser[1] · Andreas Busse[1] · Johannes König[2] · Sigrid Blömeke[3] · Jessica Hoth[4] · Martina Döhrmann[4]

Accepted: 7 February 2016 / Published online: 13 February 2016

Abstract The paper presents results from a computer-based assessment in which 171 early career mathematics teachers from Germany were asked to anticipate typical student errors on a given mathematical topic and identify them under time constraints. Fast and accurate perception and knowledge-based judgments are widely accepted characteristics of teacher competence. The item-wise length of anticipation time, the complexity of mathematical topics and the frequency of right or wrong given answers were used as indicators for teacher competence. The data revealed that anticipation time and the complexity of mathematical topics were related with each other. The groups of test persons with correct and incorrect answers behaved contrarily to the length of the anticipation time. Whereas test persons with correct answers needed more time to anticipate typical errors with an increasing complexity of the errors, the test persons giving false answers responded very quickly, even with an increasing error complexity. This finding confirms results of the expertise research, which emphasize that expert teachers focus more intensively if this is required by a complex task.

Keywords Professional competences · Typical errors · Error anticipation · Early career teacher · Timed test

✉ Lena Pankow
lena.pankow@uni-hamburg.de

1 Faculty of Education, University of Hamburg, Hamburg, Germany

2 University of Cologne, Cologne, Germany

3 Centre for Educational Measurement at University of Oslo, Oslo, Norway

4 University of Vechta, Vechta, Germany

1 Introduction

Mathematics teachers' daily routine work comprises dealing with student errors and misconceptions, often connected with fundamental mathematical concepts. Fast identification of these errors is an important professional requirement, because an immediate feedback by the teachers is expected by the students. In addition, the identification of student errors is needed in order to evaluate students' work, to plan the next instructional steps and to foster students' learning.

It is well-known that expert teachers are more likely to identify errors faster than novice teachers, because experienced teachers know or even anticipate typical errors students may make in advance (Groves et al. 2013). The results from expertise research lead to the assumption that experienced teachers focus in a special way when they observe a student's written answer. Livingston and Borko (1989) state: "Experts' knowledge systems provide a framework for determining what information is relevant to their planning and interactive decisions and what information can be ignored." (p. 39). That means expert teachers anticipate what kind of solution can be expected, which leads an experienced teacher not to read a students' written answer carefully from the beginning to the end, but to skip the less mistaken-bound parts in order to reach the problematic part quickly.

The fast identification of errors is part of teachers' overall professional expertise. On the one hand, it roots in mathematical knowledge and in knowledge about teaching and learning mathematics (Krauss and Brunner 2011). On the other hand, it features a high degree of accuracy and speed in combination with a clear idea of distinction between relevant and less relevant aspects, which can be named as information processing capacity (Livingston and Borko 1989).

In this paper, results of a national follow-up study to the international comparative Teacher Education and Development Study in Mathematics (TEDS-M) are presented. Teachers' competence to identify typical student errors was conceptualized and examined as part of early career mathematics teachers' competencies in the field of mathematics, mathematics education and general pedagogy (König et al. 2014; Blömeke et al. 2014b; Kaiser et al. 2015). In this follow-up study of TEDS-M (so-called TEDS-FU study) a computer-based assessment was developed that provided the participating early career teachers with an opportunity to anticipate typical errors to a given mathematical topic before they were asked to identify such typical errors under time constraint. The evaluation of the data revealed a strong relation between the competence to identify errors quickly with both mathematical content knowledge and mathematics pedagogical content knowledge (Blömeke et al. 2014b). Building on this, the present paper provides an in-depth analysis of the relation between the anticipation time of possible students' errors and the success in identifying students' errors. We also examine how the complexity of mathematical topics influences this relationship. As part of the teacher pre-service and in-service education, it is essential for the learning progress of students that teachers recognize situations, in which typical errors may appear, quickly and reliably in a correct way.

2 Theoretical framework

2.1 Expertise research

In accordance with Carter (1990) expert teachers act in a context-bound way and focus more intensively if necessary.

According to Mulder and Gruber (2011), expertise research refers to successful people who have achieved excellent performance over a long period of time. A person is attributed expertise when he or she achieves an expert status in a complex field. In particular professionalization in terms of professional competencies seems to play an important role in the difference between experts and novices, which holds for the teaching profession too, as Chi (2011) points out. A solid knowledge base, a portfolio of experiences, and an efficient mode of operation are characteristics of expert teachers (Mulder and Gruber 2011). They are able to perceive situations faster, in an overall and more holistic way due to their strong experience base. In this respect, they outperform novices as they are able to quickly interpret a given situation (Blömeke et al. 2014a).

When regarding the identification of students' errors, at a more general level the concept of noticing (Sherin et al. 2011) is applicable and important. According to this approach, experts are able to perceive more selectively due to their "professional eye" (Gobet 2005) and are able better to interpret situations than novice teachers. In addition, expert teachers analyze the learning processes of their students in a more focused way and they better anticipate possible consequences, developing more adequate and more flexible strategies for teaching. Carter et al. (1988) point out that teachers' precise perception of classroom events plays an essential role for the quality of instruction. Concerning the acquisition of professional knowledge during the professional life of a teacher, a restructuring process takes place in the professional years of practice due to their increasing experience (Berliner 1986; Livingston and Borko 1989; Chi 2011).

A timed test represents a suitable methodological approach to assess teachers' expertise to identify typical students' errors, because the relation to typical classroom situations and their complexity is strengthened by the time restriction. As previous research has pointed out the competence to notice classroom situations is considered to be an important premise for the successful mastering of teaching (van Es and Sherin 2002). In the literature, various frameworks classifying different cognitive processes related to noticing exist for the analysis of teachers' activities concerning classroom activities (e.g., van Es and Sherin 2002; Seidel et al. 2010).

According to Weinert et al. (1990), teachers with practical experience have mental models about the knowledge of their students and ways to reach a higher level with them. Compared with novice teachers, i.e., teachers without practical experience, experienced teachers are better aware of the methods and typical difficulties that might occur, based on their richer knowledge of possible errors that they may anticipate during class. Practical experiences support the development of this knowledge (Berliner 1986). Already existing knowledge and the developed planning strategy admits the detention and the access of the data. By building chunks between information a fast and accurate access is available. Based on schemes with typical scripts experienced teachers can anticipate and realize a possible problem and can react faster than novices (Calderhead 1981). Because of this cognitive connection experienced teachers are able to perceive a substantial amount of relevant information accurately (Carter et al. 1988) and can afterwards make decisions for different alternatives, and are able to act effectively (Weinert 1996).

Expertise in a specific domain is stored in long-term memory (Kirschner et al. 2006). Knowledge in long-term memory is organized into meaningful structures and can be used for the interpretation of novel information (Baumert and Kunter 2006). Concerning the structures an activation of knowledge for categorizing new information is necessary when salient cues are present. Since connectivity and complexity of schemata required for identifying and categorizing information evolve with practice (e.g., Dehoney 1995), perceptual accuracy is an indicator of expertise. Consequently, it can be reasonably assumed that expert teachers identify students' errors more precisely and correctly than novices (Sabers et al. 1991).

The depth of information processing serves as an explanatory model for these differences since experts draw conclusions about meaning and reasons behind teaching sequences; for example, "In such a community, students might help one another solve problems by building on each other's knowledge, asking questions to clarify explanations, and suggesting avenues that would move the group toward its goal" (Brown and Campione 1994, quoted in Bransford et al. 2000, p. 25).

So the quantitative amount and the qualitative structure of prior knowledge gives an advantage to expert teachers who are able quickly to interpret a given classroom situation holistically and draw conclusions even about the non-visible features such as the background and instructional rationale behind the perceived scenario. Based on the interpretation of students' errors, teachers have to develop measures to cope with the underlying reason for this error. Based on the kind of the interpretation teachers will act differently, influenced by their professional competence.

In the following, we concentrate on the identification of students' errors and the relation between the identification of errors and teachers' anticipation.

The term anticipation was originally used in research on sport education concerning perceptual anticipation and was made by Jones and Miles (1978). They examined the ability of amateur tennis players and experienced tennis coaches to anticipate the trajectory of the tennis ball within the tennis match. Therefore film clips of a tennis play were shown to the participants. The anticipation in this field is described as the ability of tennis players to predict upcoming events based on partial information and the anticipation predicts where, when, and how something might happen (McMorris 2004): anticipation can help us to make fast responses and helps us to determine what is going to happen before it actually does happen. We adopt this approach and transfer it to the professional competencies of teachers, i.e., we describe in our paper anticipation as the thought process based on experiences within specific thematic areas in order to predict typical students' errors.

2.2 Identification of student errors as specific component of the perception of classroom incidents

The fast identification of students' errors as part of the perception of important classroom incidents is a daily challenge for mathematics teachers under various perspectives: Fast identification of errors can be conceptualized as a teachers' sub-competence of diagnostic competence. The first type of sub-competence focuses on the entire class and becomes relevant, when a teacher chooses specific tasks regarding the class' abilities. Another type is a diagnostic situation in which the teacher chooses different tasks for different students and, finally, the third type is during one-to-one situations between teacher and student [for more details on diagnostic competence see Hoth et al. (2015) in this issue].

Within the diagnosis of classroom situations or other learning situations, the identification of students' errors plays an important role. It is an essential requirement in the classroom where many actions take place simultaneously and a teacher ought to be able to give a fast adaptive feedback to students concerning their written or oral work. In a mathematics classroom, usually there is little time to read a students' work thoroughly. So generally a quick glance should already be sufficient for the teacher to identify a possible error. Furthermore, the fast identification of student errors is an activity highly important at all levels of students' learning from Kindergarten to university learning and a special characteristic of mathematical teaching processes.

For teaching mathematics, student errors inform the teacher about the learning process of his or her students (Beutelspacher 2008). It is the teacher's task to perceive those errors quickly in order constructively and sustainably to make use of them during teaching. As already mentioned, the competence of the fast perception of a students' error can be described as sub-component of the diagnostic competence of a teacher, which comes into action as early as when the teacher plans the lesson anticipating typical students' errors ("planning decisions", Peterson et al. 1978, p. 418). This process (perception and anticipation) happens at the same time as the teaching and learning process goes ahead ("interactive decisions"). Results in expertise research emphasize that expert teachers know which kind of problems are likely to appear and they include this knowledge in their lesson planning (Leinhardt 1989; Clermont et al. 1994; König et al. 2015).

Empirical research points out that the fast judgement of the right or wrong students' solution is a fundamental competence of expert teachers (Krauss and Brunner 2011). Fast perception and assessment of students' utterances—in oral and written form—qualifies teacher performance in the specific classroom situation. Even under time pressure and under observation an expert teacher has to cope with these situations.

How to combine the competence facets, accuracy and speed of error recognition, is an often discussed issue in the literature about speed and timed tests that can be answered differently. Some researchers prefer to measure the time length until an answer is given (Phillips and Rabbitt 1995; Arthur et al. 2003), as it was done for example in the COACTIV study (Krauss and Brunner 2011). Others are in favor of 'experimenter-paced testing' (Davison et al. 2011) providing a certain time limit to answer a question. The latter puts an external pressure on the test person and can therefore be seen as a simulation of a real classroom situation.

Empirically, the ability quickly to identify student errors is grounded in mathematical content knowledge (MCK) as the COACTIV study, TEDS-M, and TEDS-FU have pointed out (Krauss & Brunner 2011; Blömeke et al. 2014a) and in mathematics pedagogical content knowledge (MPCK) shown by TEDS-M and TEDS-FU (Blömeke et al. 2014a). According to these studies, the ability of fast error recognition was more highly correlated with MCK than with MPCK. The high correlation (ß = 0,42*) (see Blömeke et al. 2014a) might be influenced by the high relevance of errors and their identification at all levels of mathematical teaching and learning processes, so the competence facet of identification of student errors is even supported by mathematical teaching-and-learning processes without specific relations to classroom situations.

Since experts excel in the ability to focus on the essential parts of a situation (Borko and Livingston 1989), it is reasonable to assume that expert teachers focus on those aspects of a students' solution where errors are most likely to appear, i.e., they anticipate the possible error. By introducing anticipation into the test component of error recognition, teachers' preparation for a lesson was simulated (Leinhardt 1989).

We can summarize the key features of the construct *Competence quickly to identify typical students' errors*, on which the current study will focus, as follows: Overall this competence is a sub-competence of diagnostic competence. It refers to the first phase of noticing, namely the perception and anticipation of important classroom incidents that are subject-specific, i.e., related to the teaching of mathematics. The competence is based on the broad knowledge of expert teachers in a wide variety of school mathematical topics; the knowledge about errors is on the one hand strongly connected to mathematics content knowledge (MCK). On the other hand, it also refers to mathematics pedagogical content knowledge (MPCK) due to its relation to school mathematics. This competence quickly to identify typical students' errors contains the efficient anticipation of possible students' errors, which will speed up their correct identification of errors.

2.3 Research question and hypotheses

We hypothesize that potential expert teachers in contrast to potential novice teachers know where typical errors are likely to appear and look for these errors first. This speed aspect is operationalized by limiting the available time for an answer to a few seconds. The limited time encourages the test person to focus on expected typical errors. Based on a comparison of the anticipation time and the solution frequency per item, we assume that the duration of anticipation corresponds to the intensity of anticipation. In other words, we assume that potential expert teachers will not only have the ability to identify the error very fast, but also anticipate the error faster than potential novices. We therefore study in this article the question: What is the *relation* between the *measured anticipation time* and the *competence of identifying students' errors*?

We hypothesize that participants with higher expertise are better in recognizing the complexity of an item based on an adequate analysis of the announcement and thus take more time to anticipate possible students' errors within complex topics, whereas they take less anticipation time, when their analysis of the announcement leads them to assume that the student error is less complex.

3 Method

3.1 Study and sample

TEDS-FU was a follow-up of the Teacher Education and Development Study-Learning to Teach Mathematics (TEDS-M; Blömeke et al. 2014b) carried out in 2008 under the auspices of the International Association for the Evaluation of Educational Achievement (IEA). TEDS-FU was carried out in Germany in 2012; the study tested web-based former TEDS-M participants now being early career teachers. Departing from the seminal work by Shulman (1987) TEDS-FU like TEDS-M tested various facets of the professional competencies of teachers, namely mathematical content knowledge (MCK), mathematics pedagogical content knowledge (MPCK) and general pedagogical knowledge (GPK). TEDS-FU enriched the theoretical framework of TEDS-M by including situated, performance-oriented competence facets such as perception of classroom incidents, interpretation, and decision-making (for details see Kaiser et al. 2015). In addition the competence facet of fast error recognition of students' errors was evaluated with a time-limited computer-based assessment. The paper will concentrate on this competence facet.

In the following, the lower-secondary TEDS-FU sample will be analyzed, out of which 137 study participants completed the timed assessment. The participants had worked as in-service teachers by this time for about 3 years and were on average 32 years old (SD = 5.9, min = 26, max = 53) (Blömeke et al. 2014a). 58.9 % of the TEDS-FU participants were female, the sample consisted of 55.8 % teachers for higher track secondary schools (so-called Gymnasium), and 43.6 % lower track secondary schools (so-called Hauptschule and Realschule) (Blömeke et al. 2014a). The majority of the participants (53.7 %) had high cultural capital which was indicated by the possession of more than 200 books at home (Blömeke et al. 2014a). It has to be pointed out that the participants of TEDS-FU represented a positively selected convenience sample, displayed amongst others by the high mean score in TEDS-M, because participation in the test took place on a voluntary basis with a small amount of money as incentive.

3.2 Test instrument

In order to measure the fast recognition of students' errors as important facet of teachers' professional competencies a new test-instrument was developed. The computer-based test on the identification of students' errors consists of 16 items.

The online implementation of TEDS-FU was supported by the tool CBA Itembuilder (Rölke 2012). This online environment made it possible to implement time restrictions in order to control the answering time. The test persons participated in the test via their computers at home, i.e. that the test conditions were not controlled. However, due to the short testing time, it can be assumed that no external help was used by the test persons.

As a resume the following variables of interest will be investigated: the anticipation time, the complexity of each item, the correctness of the answer and its coding. Concerning the short time of the answer the experts, mentioned above, expressed consensus that 4 s are too short, even for experienced teachers, to recalculate the items. Therefore, the time limitation of 4 s was selected for the timed tests. The same group was asked to evaluate the item complexity based on the announcement. Each item was rated with a three-point-Likert scale concerning its complexity based on the three factors described above.

The items cover the following range of topics from mathematics taught at lower secondary level: fractions, percentage, manipulations of simple number expressions, calculation of values of functions, computations with zero, and calculation of probabilities. Overall, typical students' errors occurring often in mathematical lessons in German lower secondary schools were covered by these items. Since the items contain typical school mathematics issues, all errors could have been identified easily by mathematics teachers, if enough time was provided. Therefore, the given time was limited in order to prevent test-persons from re-calculating the three students' answers. In order to validate the influence of the time limitation on the evaluated competence facet, a workshop was held with educators from the second phase of teacher education in Germany, experts (n = 7) from university, as well as experienced mathematics teachers.

Based on an analysis of the structure of the announcement and the underlying mathematical field each item is rated as non-complex or complex. A non-complex item will not have any disturbing words in the announcement, no precise task will be given and only a small variety and diversity of possible errors exist in the according mathematical field.

We assumed that through the opportunity to anticipate student errors in advance, the test would come closer to the classroom reality, where teachers are acquainted with the topic they are teaching and therefore know which errors to expect. Thus, by announcing the mathematical topic, we were aiming at the early career teachers' expertise in anticipating problems. The young teachers may either recall their teaching experience or academically acquired knowledge about student errors. The anticipation time per item—i.e., how much time an early career teacher took to think about possible errors—was measured due to technical reasons in whole seconds per person.

Once the early career teacher had pushed the "forward"-button, he or she was confronted with three different students' expressions related to the announced mathematical topic. The expressions were marked by the letters "a", "s", and "d" corresponding to the respective keys on the keyboard that had to be pressed to identify the wrong student solution quickly. Within the given time limit of 4 s, the test

Table 1 Design of the testing process (italic: these aspects were displayed to the teachers) [the corresponding answer to "d" is the wrong answer]

Item	1. segment	Announcement of the topic in order to anticipate typical errors (≤ 5 min)	*Solution of equations*		
	2. segment	Identification of the incorrect answer (≤ 4 s)	*a*	*s*	*d*
			$3x=3 \Rightarrow x=1$	$5x=10 \Rightarrow x=2$	$\frac{1}{2}x=6 \Rightarrow x=3$

Table 2 Expert rating on the complexity of the items within the timed test [the italic item numbers are described in detail in chapter 3.3]

Item no.	Item topic	Ratings as non-complex item/missing answers/ratings as complex item	Ranking
2	Transforming simple number terms with brackets	(17/2/2)	1
3	Transformation of terms I	(16/4/1)	2
7	Calculation of fractions III	(16/5/0)	3
12	Solution of equation I	(16/5/0)	4
16	Probability calculation	(15/2/4)	5
6	Calculation of fractions II	(15/6/0)	6
13	Solution of equation II	(14/2/5)	7
5	Calculation of fractions I	(14/5/2)	8
1	Extract a root	(10/2/9)	9
8	Calculation of fractions IV	(10/7/4)	10
9	Calculation of percentage	(7/5/9)	11
4	Transformation of terms II	(7/5/9)	12
10	Theorem of Pythagoras	(7/10/4)	13
15	Linear functions	(6/5/10)	14
11	Typical trigonometric values	(5/4/12)	15
14	Application of the formula for quadratic equations	(4/1/16)	16

person had to decide, which of the three options includes an error. Using the method of experimenter-paced testing (Davison et al. 2011), it was no longer possible to give an answer after the time had elapsed.

In Table 1, the procedure of the time limited test to identify typical students' errors is illustrated by an example.

To make participants familiar with the form of the test, they were asked to put three of their fingers on the keys a, s, and d of their computer keyboard before the tests starts. Furthermore, they were provided with an example item in advance to get used to the technique. In addition, the first two items had an extended time limit (5 s instead of 4 s).

Each item started with the announcement of a mathematical topic (e.g. addition of fractions). While reading the announcement, the test persons were asked to think simultaneously about typical students' errors in this field. The test persons had up to 5 min to anticipate these errors. They were free to decide when to proceed.

The announcement of the error, which has to be anticipated, is influenced by three factors, which were discussed on their influence of the test persons by an expert rating. The *first factor* focuses on the aspiration level within the announcement, for example the words "easy" or "mental arithmetic" may lead the participants to reflect longer or shorter on possible student errors. The *second factor* refers to additional aspects within the announcement, which may stimulate an intense reflection such as the indication of specific numbers or task contexts. For example if a quadratic equation with numbers is given and typical errors should be identified, this could lead the test person to solve the task given. The *third factor* is the variety and diversity of the field, from which the typical student error is taken. The announcement can include a high diversity of typical student errors: for example concerning the addition of fractions a high number of possible error patterns exists, which may influence the anticipation time (Table 2).

Each item was rated by each expert as followed: The announcement was shown to the expert and has to be rated within the three categories: aspiration level, indication of specific numbers and the variety and diversity of the field, from which the typical student error is taken. For each of these aspects the item should be examined by typing (−1/0/+1) meant as (not existing/missing answers/existing). The experts rated the item "Addition of fraction" (Fig. 1) as follows: (7/0/0), (5/2/0) and (4/3/0). As an overall result these points can be calculated to (16/5/0) these numbers can be ranked towards the complexity of all items.

3.3 Item examples

Based on typical errors in the following we describe the announcement of the mathematical topics "Transforming simple number terms with brackets", "Addition of fraction", "Trigonometric values to specific angles" and "Application of the formula for quadratic equations", give some short background information, discuss reasons why the item might be easy or difficult to anticipate and give the result of the expert rating of the complexity. These items were selected in order to clarify the different complexity of the items.

Transforming simple number terms with brackets		
a	s	d
$3\cdot(5+2)=3\cdot5+3\cdot2$	$3+(5+2)=3+5+2$	$3-(5+2)=3-5+2$

Fig. 1 Item no. 2 of TEDS-FU "*Transforming simple number terms with brackets*" [the corresponding answer to "d" contains the wrong answer]

Addition of fractions		
a	s	d
$\frac{1}{2}+\frac{1}{4}=\frac{3}{4}$	$\frac{2}{7}+\frac{3}{7}=\frac{5}{14}$	$\frac{1}{3}+\frac{2}{3}=1$

Fig. 2 Item no. 5 of TEDS-FU "*Addition of fractions*" [the corresponding answer to "s" contains the wrong answer]

The first item refers to "Transforming simple number terms with brackets" (Fig. 1). The mathematical operation is not clearly specified, various combinations and thus errors (distributive law etc.) can occur. The word "simple" within the announcement labels the topic as easy. The word "brackets" expresses a multitude of brackets. In case the students' expressions contain more than two brackets the following rule "calculate from outside to the inside" has to be known and applied by the students. On the one hand the mentioned aspects point out the complexity of the topic. On the other hand this topic is likely to be anticipated, because a minus sign in combination with a bracket leads to one typical student error which can be expected and is often made by all age grades of students.

This item was rated as non-complex with an increasing aspiration level, because the word simple let the participants anticipate longer.

Although answer "a" and "s" have potential for typical students' errors too, the last answer "d" displays the typical students' error. In this case, there is a minus sign in front of the brackets and the minus sign has to be applied to the whole algebraic expression in the brackets. The application of the minus sign only to the first number in the bracket is a common student error.

The second item is about adding fractions (see Fig. 2). According to Padberg (2009), the extension of the set of numbers of natural to rational numbers represents central learning topics at the beginning of secondary mathematics education in German schools. Many students have problems with the understanding of fractions, which create difficulties for their further school career. Wu (2008) points out that by the introduction of fractions an intuitive calculation with fingers is no longer possible. Students not only have to acquire conceptual knowledge of fractions such as the importance of the different function of denominator and numerator (Eichelmann et al. 2011), but also associated operations and procedural knowledge of applicable rules. Eichelmann et al. (2011) emphasize that learners often rely on memorized rules and, as a result, fail in more complex situations.

Trigonometry: trigonometric values to specific angles		
A	s	d
sin 30° = 0,5	cos 0° = 0	tan 45° = 1

Fig. 3 Item no. 11 of TEDS-FU "Trigonometry: trigonometric values to specific angles [the corresponding answer to "s" contains the wrong answer] trigonometric values to specific angles"

The announcement "Addition of fraction" (Fig. 2) which is item no. 5 in TEDS-FU describes the number range in action, namely fractions. Furthermore, the type of calculation is clearly mentioned.

The group of researchers rated this item as non-complex. All agreed to this decision.

The error used in the test on the addition of two fractions (see Fig. 2) is based on the erroneous approach of the separate addition of the numerator and the denominator, which has been analyzed in several empirical studies so far (Eichelmann et al. 2011). According to a study by Wartha (2007) in German mathematics education, 45 % of all errors on fractions result from this approach. Some of the students even develop their own addition rule (Padberg 2009). Confusion with rules concerning the multiplication and addition of fractions is frequently observed (Wartha 2007). According to Padberg (1986, 1995), this often occurs after the introduction of the—at first—easier multiplication method. In this context, unrealistic results are accepted without reflection and without going back to the basic concept of fractions.

The item "Trigonometry: trigonometric values to specific angles", item number 11 (see Fig. 3) in the test, indicates within the wording the specific area of mathematics. The focus could be the periodicity, the miscalculation of radians and degree, shifting phases as well as vertical asymptotes. The term "specific angles" indicates a

Application of the formula for quadratic equations $x^2 + 8x - 4 = 0$		
a	s	d
$x_{1,2} = -4 \pm \sqrt{16+4}$	$x_{1,2} = -4 \pm \sqrt{16-(-4)}$	$x_{1,2} = -4 \pm \sqrt{16-4}$

Fig. 4 Item no. 14 of TEDS-FU "Application of the formula for quadratic equations $x^2 + 8x - 4 = 0$" [The corresponding answer to "d" contains the wrong answer] trigonometric values to specific angles"

geometrical aspect of the task. In the case of specific angles only angles between 0° and 90° can be assumed. All these possible and diverse interpretations of the announced topic make this item to a complex one.

The group of raters reached consensus on the complexity of this item. They agreed that an expert may start writing down the characteristic values of these functions or may visualize these functions in order to spot the erroneous answer quickly. In contrast, an unexperienced teacher may give less thought to this subject matter due to their lower experience.

Concerning the problem of the calculation of trigonometric values, the teachers need to recall the relevant features of trigonometric functions and they need to anticipate possible errors. All three answers have erroneous potential, however, the wrong students' result "s" contains a special feature. In this wrong students' answer, the students may associate wrongly the multiplication with zero leading to zero and may therefore calculate the cosine value wrongly.

Another topic, which creates typical errors in German secondary school is the "Application of the formula for quadratic equations of the type $x^2 + px + q = 0$" (see Fig. 4).

In Germany, the solution of quadratic equations is usually done by applying the formula $x_{1,2} = -\frac{p}{2} \pm \sqrt{\left(\frac{p}{2}\right)^2 - q}$, so-called pq-formula. The complexity of the item No. 14 might arise on the one hand from the possible misunderstanding of the announcement. The participants could think that the displayed number expression is only one possible example for this formula and will not think about the typical errors of this expression. On the other hand the topic itself is complex, because of the high number of possible errors. Firstly, the participants may expect an error with p or q being inserted in the formula. Secondly, an error might arise within the calculation of the roots. Furthermore, the announcement has a high variety of other possible solution approaches like completing the square, graphing or using Vieta's formula. The expert group therefore agrees on the high complexity on this item.

According to Hensel (2012), many German students have problems with the usage of this pq-formula in order to solve quadratic equations. The first possible error may appear, if the students do not realize that this quadratic polynomial is already normalized. A second possible error might appear, if the negative algebraic sign in the original form is ignored or overlooked while transforming the numbers into the pq-formula, as is done in the wrong students' answer d. Other students may have difficulties with the formula itself. A common error is the negative algebraic sign at the beginning of the pq-formula, which is quite often not taken into account. The teacher can also anticipate that students may not think about the fact that a root has maximally two solutions, which leads to the omission of the plus or the minus sign in front of the root and the calculation of only one solution.

3.4 Data analysis

The third table contains descriptive statistics for all 16 items, providing an overview of the anticipation time per mathematical topic and item. As each item can be solved correctly (CA) or falsely (FA), the average anticipation time of early career teachers providing correct answers or false answers respectively was computed (see Table 3), then compared item-wise by calculating the ratio CA/FA (see Table 3 first column providing statistical findings). This ratio is below one if the CA group needs less time for anticipation than the FA group. In contrast, the ratio is above one, when the CA group takes more time than the FA group. By applying a Wilcoxon–Mann–Whitney test, mean differences between the CAs and the FAs were tested for statistical significance (see Table 3) using the significance level of 5 % (Bortz & Schuster 2010). The items are displayed in the order of their appearance in the timed test.

In a second step of the analysis the sample was split into high and low achievers by using a median split (Mitchell & Jolley 2013) according to the achievements based on the whole test. This was done, amongst other reasons, in order to distinguish homogeneous groups of participants in their anticipation behavior regardless of their response in single test items. Afterwards an item-wise comparison of the anticipation time between the two median-split groups based on the person's competence to identify student errors (based on the results in the timed test) was carried out.

4 Results

The group of correct answers and false answers will be further compared by statistic characteristics (average, standard deviation, min and max), the effect size (Cohen's d) and the frequency of solution (see Table 3). For an overview on the groups of correct and false answers based on the results of

Table 3 Descriptive statistics for anticipation of all items and item-specific average time for groups with correct and false answers (CA, FA) [the italic numbers are for the item examples shown in Sect. 3.3]

Item no.	Item topic	Ratio (CA antic. time/FA antic. time)	CA [M(SD)]	FA[M(SD)]	Total [M(SD)]	Max (s) anticipation	U test (*p* value)	Effect size
12	Solution of equation I	0.51	3.75 (2.26)	7.4 (22.01)	5.35 (14.71)	173	0.112	0.3
16	Probability calculation	0.66	6.35 (4.86)	9.68 (23.79)	8.07 (17.48)	191	0.551	0.2
2	Transforming simple number terms with brackets	0.71	5.77 (4.1)	8.09 (10.3)	7.21 (8.56)	71	*0.039**	0.3
13	Solution of equation II	0.78	4.03 (2.21)	5.14 (6.88)	4.88 (6.1)	66	0.682	0.2
3	Transformation of terms I	0.80	4.88 (3.17)	6.15 (6.55)	5.76 (5.74)	49	0.465	0.2
6	Calculation of fractions II	0.83	4.87 (3.31)	5.89 (4.56)	5.2 (3.77)	25	0.365	0.3
5	Calculation of fractions I	0.92	4.59 (2.46)	4.98 (5.7)	4.78 (4.32)	43	0.235	0.1
8	Calculation of fractions IV	0.92	5.07 (3.75)	5.51 (6.58)	5.32 (5.52)	49	*0.945*	0.1
10	Theorem of Pythagoras	0.92	6.02 (4.07)	6.57 (5.17)	6.39 (4.8)	35	0.687	0.1
1	Extract a root	0.99	5.75 (4.41)	5.82 (7.61)	5.79 (6.43)	58	0.360	0.0
9	Calculation of percentage	0.99	6.22 (4.64)	6.3 (10.31)	6.27 (8.4)	87	0.151	0.0
4	Transformation of terms II	1.28	6.81 (5.21)	5.33 (3.01)	6.26 (4.56)	26	0.289	−0.3
7	Calculation of fractions III	1.40	5.7 (5.96)	4.06 (2.87)	5.31 (5.41)	47	0.129	−0.3
15	Linear functions	1.61	19.28 (10.28)	12.08 (10.87)	13.97 (11.15)	45	0.000*	−0.7
11	Typical trigonometric values	1.96	11.29 (17.72)	6.85 (12.33)	8.76 (15)	87	*0.031**	−0.4
14	Application of the formula for quadratic equations	2.32	35.57 (39.68)	15.36 (18.27)	20.53 (26.83)	140	*0.011**	−0.8

* Significant at 5 % level

the whole test, the participants were arranged according to their total score of correct answers.

The analysis yields the followings results. The items are ordered by the ratio of the anticipation time of the group of correct and false answerers' results, using the Wilcoxon–Mann–Whitney (U test) for significance testing item per item:

As shown in Table 3, the results of the Wilcoxon–Mann–Whitney test are significant at the 5 % level for the items "Transforming simple number terms with brackets" (item 2), "Linear functions" (item 15), "Calculation of typical trigonometric values" (item 11), and Application of the formula for quadratic equations (item 14). The analysis of all items relating the anticipation time with the outcome of the item (right or false answer) points out that there is a relation between the ratio CA/FA's length of anticipation time and the complexity of the item as evaluated by the raters. When dealing with complex items, the group of correct answerers needs more time to anticipate possible errors compared to the group of false answerers (see Table 3). Concerning elementary items the situation is reversed. Here, the group of correct answerers take a shorter anticipation time than the group of false answerers. As exemplification we give an example for the second case, namely item 2 on 'Transforming simple number terms with brackets". The item, which is rated with high consensus as non-complex, has an anticipation ratio of 0.71, which means that the group of correct answerers uses a shorter time for anticipation. The discrepancy is significant concerning the group of false answerers, who use a longer time for anticipation. As described in 3.4, this typical error is quite simple and well-known and can be recalled in short time.

A different situation emerges with items such as the items on "Typical trigonometric values" (item 11) or the "Application of the solution formula for quadratic equations" (item 14), which are both classified as complex by the raters. The group of correct answerers uses significantly longer time for anticipation than the false answerers.

Cohen's d concerning item 14 "Application of the formula for quadratic equations" which was presented in 3.3 in detail, has an high effect concerning the anticipation time and the skill to solve the item based on the differences between the two groups of correct and false answerers. Item 15 "Linear functions" has a middle to high effect, item 11 a small effect on the anticipation time. Item 2 displays a small effect, representing the reverse response pattern of the skill to solve the item and the anticipation time taken.

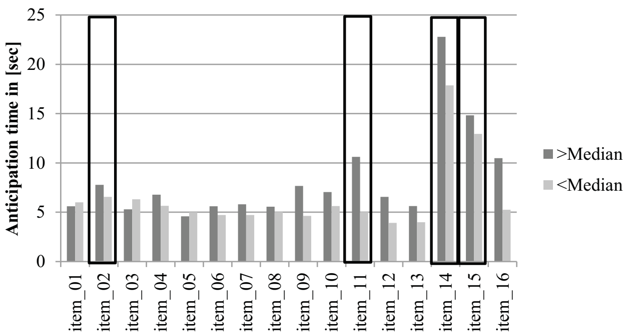


Fig. 5 The median split into high and low achievers per item in (s)

The other items' mean differences are not statistically significant, but show a similar tendency. The results of our further analyses support these findings described above. The scale reliability of the dichotomous test item implemented in TEDS-FU showed a sufficient scale reliability, namely WLE = 0.73 (for details see Blömeke et al. 2014a).

The median split consisting of the high and low achievers (see Fig. 5) over the whole timed test is able to show that the high achievers need a longer time to anticipate the typical student errors than the low achievers. Those differences are most of the time very small, but in cases of higher item complexity the group of high achievers needs significantly longer time whereas the group of low achievers does not change their behavior on anticipation.

As shown in Table 4, the results of the test are in line with the expert ranking explained in chapter 3.2. The distance between the rated item in the test and the rating by the experts is 0–4 places in the ranking. The item number 7 "Calculation of fractions III" was rated as non-complex by the experts, but the ranking within the test emphasis a complex item. This item will be focused on in further detailed analyses of all items.

5 Summary and discussion

5.1 Summary

The fast and accurate identification of problematic classroom incidents as well as knowledge-based judgments are widely accepted characteristics of professional expertise.

Table 4 Expert rating complexity contrasting the test results

Item no.	Item topic	Ratings as non-complex item/ missing answers/ratings as complex item	Ranking (experts)	Ranking (test)	Distance (ranking experts − ranking test)
2	Transforming simple number terms with brackets	(17/2/2)	1	3	2
3	Transformation of terms I	(16/4/1)	2	5	3
7	Calculation of fractions III	(16/5/0)	3	13	10
12	Solution of equation I	(16/5/0)	4	1	3
16	Probability calculation	(15/2/4)	5	2	3
6	Calculation of fractions II	(15/6/0)	6	6	0
13	Solution of equation II	(14/2/5)	7	4	3
5	Calculation of fractions I	(14/5/2)	8	7	1
1	Extract a root	(10/2/9)	9	10	1
8	Calculation of fractions IV	(10/7/4)	10	8	2
9	Calculation of percentage	(7/5/9)	11	11	0
4	Transformation of terms II	(7/5/9)	12	12	0
10	Theorem of Pythagoras	(7/10/4)	13	9	4
15	Linear functions	(6/5/10)	14	14	0
11	Typical trigonometric values	(5/4/12)	15	15	0
14	Application of the formula for quadratic equations	(4/1/16)	16	16	0

Students 'learning processes and classroom situations, i.e. amongst others, recognizing students' errors under time pressure, interpreting and acting accordingly are well-known indicators of teachers' expertise (Sherin et al. 2011).

In the field of mathematics education, this can be operationalized by the fast recognition of mathematical errors made by students. One part of the Follow-up study of the international Teacher Education Development Study in Mathematics (TEDS-FU) evaluating the professional competencies of early career teachers focuses on this aspect of teachers' expertise.

The results of the study give a hint that the time early career teachers use to anticipate possible students' errors strongly depends on the complexity of the item. The group of correct answerers use a longer anticipation time due to a stronger knowledge base on specific items, which needs to be activated. In contrast, the group of false answerers take a shorter anticipation time, because they possibly have a smaller knowledge base to activate.

These findings confirm results from the expertise research that experts spot more adequately the complexity of a task and react accordingly, which means they take more time to anticipate.

To sum up our conclusions, we describe our findings concerning the relation of the complexity of the items and the anticipation time as follows (Table 5).

Thus, our study emphasizes that the group of correct answerers are faster in anticipating errors within non-complex items while the group of wrong answerers use the same period as for the other items. In contrast, the group of right answerers take a longer time for anticipating possible students' errors than the group of false answerers when they are confronted with complex or less familiar topics. These findings confirm results of the expertise research, which emphasize that expert teachers act context-bound and focus more intensively if necessary (Carter 1990).

6 Discussion

Based on results of the timed test we were able to show differences concerning the relation between the measured anticipation time and the competence of identifying students' errors. Although only four items showed significant results concerning the relation of the competence to solve complex items and the anticipation time, the other items confirmed the results.

Table 5 Relation of anticipation time and group of answerers

	Complex items	Non-complex items
CAs need	Long anticipation time	Short anticipation time
FAs need	Short anticipation time	Short anticipation time

We want to close with a discussion of the limitations of the study. The sample of TEDS-FU is a positively selected sub-sample of the original TEDS-M sample, which can be seen by the higher average results of the new sample compared to the original sample and the strong competence growth in general pedagogy (GPK) between the two measurement points in TEDS-M and TEDS-FU (König et al. 2014). The decrease of the sample size may have various reasons, amongst others the voluntary participation with only small incentives. Furthermore, it is unclear, how many future teachers tested at the end of the second phase of teacher education in TEDS-M really entered the teaching profession. This might lead to the open problem, whether the achieved result is generalizable across all performance levels. Furthermore, the tested group of early career teachers was homogeneous with nearly the same amount of teaching experience (i.e. 4 years teaching experience or less) (König et al. 2015), which leads to missing variance in the results.

Other aspects such as reading speed, eye-sight and form on the day as influencing factors may play an additional role, but were not assessed in this study due to the unsolved problem of how to evaluate them. The anticipation time was measured in whole seconds, due to a technical reason, which turned out not to be an accurate measure. Furthermore, it might be questioned whether early career teachers really benefit from the opportunity to think about typical errors due to their short teaching experience. Further studies with more experienced teachers are currently being carried out (TEDS-Instruct) and will allow further insight into this unsolved problem.

The design of the study differed in several respects from another study—the COACTIV study—which also included a test component on the identification of students' errors (Krauss & Brunner 2011). In contrast to COACTIV, the topic from which the error would stem was given in TEDS-FU in order to enable the participants to anticipate the topic and typical students' errors in advance. Furthermore, our study uses time limitations in contrast to COACTIV, in which the time that was needed for developing answers by the participants was measured. The errors used in our study were typical students' errors in contrast to more complex errors used in COACTIV. In COACTIV, each item consists of a task and one students' solution is given that can be either right or wrong. After answering the item, the solution is shown to the participants as well as their time needed. It is an open question so far, whether both studies measure the same construct, namely fast error recognition as part of diagnosis competence, which is a facet of teachers' professional competencies. In addition, it is not clear yet, whether the construct measured is more linked to mathematical content knowledge (MCK) or mathematics pedagogical content knowledge (MCPK). Both studies confirm the strong influence of MCK and the lower influence of MPCK (see Blömeke et al. 2014a; Krauss & Brunner 2011). However, it cannot be decided on the basis of the data evaluated so far and the design chosen, whether these influences reflect a general relation, namely that the competence of fast error recognition belongs to MCK as subfacet or whether the kind of measurement used, yields this strong relation. Further tests with specialized groups of test persons such as highly trained mathematicians or students from upper secondary level are needed in order to develop answers to these questions. Further research is needed in order to examine how this opportunity to anticipate student errors is used in detail and how the anticipated errors relate to the competence of rightly identifying students' errors.

References

Arthur, W., Tubré, T., Paul, D. S., & Edens, P. S. (2003). Teaching effectiveness: the relationship between reactions and learning evaluation criteria. *Educational Psychology, 23*, 275–285.

Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft, 9*, 469–520.

Berliner, D. C. (1986). In pursuit of the expert pedagogue. *Educational Researcher, 15*, 5–13.

Beutelspacher, A. (2008). Horizonterweiternde Stolpersteine: über die Unmöglichkeit und die Notwendigkeit von Fehlern in der Mathematik. In R. Caspary (Ed.), *Nur wer Fehler macht, kommt weiter* (pp. 86–96). Freiburg: Herder.

Blömeke, S., Hsieh, F.-J., Kaiser, G., & Schmidt, W. H. (Eds.). (2014a). *International perspectives on teacher knowledge, beliefs and opportunities to learn: TEDS-M results*. Dordrecht: Springer.

Blömeke, S., König, J., Busse, A., Suhl, U., Benthien, J., Döhrmann, M., & Kaiser, G. (2014b). Von der Lehrerausbildung in den Beruf-Fachbezogenes Wissen als Voraussetzung für Wahrnehmung, Interpretation und Handeln im Unterricht. *Zeitschrift für Erziehungswissenschaft, 17*, 509–542.

Borko, H., & Livingston, C. (1989). Cognition and improvisation: differences in mathematics instructed by expert and novice teachers. *American Educational Research Journal, 26*, 413–498.

Bortz, J., & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler*. Berlin: Springer.

Bransford, J., Brown, A., Cocking, R. (2000). *How people learn: Brain, mind, experience, and school. National Research Council Commission on Behavioral & Social Sciences & Education, Committee on Developments in the Science of Learning*. Washington, DC: US National Academy Press.

Brown, A. L., & Campione, J. C. (1994). Guided discovery in a community of learners. In K. McGilly (Eds.), *Classroom Lessons:Integrating Cognitive Theory and Classroom Practice* (pp. 229–270). Cambridge, MA: MIT Press.

Calderhead, J. (1981). A psychological approach to research on teachers' classroom decision-making. *British Education Research Journal, 7*, 51–57.

Carter, K. (1990). Teachers' knowledge and learning to teach. In W. R. Houston (Ed.), *Handbook of research on teacher education* (pp. 291–310). New York: Macmillan.

Carter, K., Cushing, K., Sabers, D., Stein, P., & Berliner, D. C. (1988). Expert-novice differences in perceiving and processing visual information. *Journal of Teacher Education, 39*, 25–31.

Chi, M. T. H. (2011). Theoretical perspectives, methodological approaches, and trends in the study of expertise. In Y. Li & G. Kaiser (Eds.), *Expertise in mathematics instruction: An international perspective* (pp. 17–39). New York: Springer.

Clermont, C. P., Borko, H., & Krajcik, J. S. (1994). Comparative study of the pedagogical content knowledge of experienced and novice chemical demonstrators. *Journal of Research in Science Teaching, 31*(4), 419–441.

Davison, M. L., Semmes, R., Huang, L., & Close, C. N. (2011). On the reliability and validity of a numerical reasoning speed dimension derived from response times collected in computerized testing. *Educational and Psychological Measurement, 72*, 245–263.

Dehoney, J. (1995). Cognitive task analysis: implications for the theory and practice of instructional design. In *Proceedings of the annual national convention of the Association for Educational Communications and Technology* (pp. 113–123) (ERIC document reproduction service no. ED 383 294).

Eichelmann, A., Narciss, S., Schnaubert, L., & Melis, E. (2011). Typische Fehler bei der Addition und Subtraktion von Brüchen-Ein Review zu empirischen Fehleranalysen. *Journal für Mathematik-Didaktik, 33*(1), 29–57.

Gobet, F. (2005). Chunking models of expertise: implications for education. *Applied Cognitive Psychology, 19*(2), 183–204.

Groves, M., O'Rourke, P., & Alexander, H. (2013). The clinical reasoning characteristics of diagnostic experts. *Medical Teachers, 25*(3), 308–313.

Hensel, M. (2012). *Kurvendiskussion. Lern- und Übungsbuch für die Abiturprüfung Mathematik*. Norderstedt: Books on Demand.

Hoth, J., Döhrmann, M., Kaiser, G., Busse, A., König, J., & Blömeke, S. (2015). Diagnostic competence of primary school mathematics teachers during classroom situations. *ZDM Mathematics Education*, 47, (this issue).

Jones, C. M., & Miles, T. R. (1978). Use of advance cues in predicting the flight of a lawn tennis ball. *Journal of Human Move Studies, 4*, 231–235.

Kaiser, G., Busse, A., Hoth, J., König, J., & Blömeke, S. (2015). About the complexities of video-based assessments: theoretical and methodological approaches to overcoming shortcomings of research on teachers' competence. *International Journal of Science and Mathematics Education, 13*(2), 369–387.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75–86.

König, J., Blömeke, S., & Kaiser, G. (2015). Early career mathematics teachers' general pedagogical knowledge and skills: do teacher education, teaching experience, and working conditions make a difference? *International Journal of Science and Mathematics Education, 13*(2), 331–350.

König, J., Blömeke, S., Klein, P., Suhl, U., Busse, A., & Kaiser, G. (2014). Is teachers' general pedagogical knowledge a premise for noticing and interpreting classroom situations? A video-based assessment approach. *Teaching and Teacher Education, 38*, 76–88.

Krauss, S., & Brunner, M. (2011). Schnelles Beurteilen von Schülerantworten: Ein Reaktionszeittest für Mathematiklehrer/innen. *Journal für Mathematikdidaktik, 32*(2), 233–251.

Leinhardt, G. (1989). Math Lessons: A contrast of novice and expert competence. *Journal for Research in Mathematics Education, 20*(1), 52–75.

Livingston, C., & Borko, H. (1989). Expert-novice differences in teaching: a cognitive analysis and implications for teacher education. *Journal of Teacher Education, 40*, 36–42.

McMorris, T. (2004). *Acquisition and performance of sports skills*. Chichester: Wiley.

Mitchell, M. L., & Jolley, J. M. (2013). *Research design explained*. Belmont: Wadsworth.

Mulder, R., & Gruber, H. (2011). Die Lehrperson im Lichte von Professions-, Kompetenz- und Expertiseforschung: Die drei Seiten einer Medaille. In O. Zlatkin-Troitschanskaia (Ed.), *Stationen empirischer Bildungsforschung* (pp. 427–438). Wiesbaden: Verlag für Sozialwissenschaften.

Padberg, F. (1986). Über typische Schülerschwierigkeiten in der Bruchrechnung-Bestandsaufnahme und Konsequenzen. *Der Mathematikunterricht, 32*(3), 58–77.

Padberg, F. (1995). *Didaktik der Bruchrechnung: gemeine Brüche-Dezimalbrüche*. Heidelberg: Spektrum.

Padberg, F. (2009). *Didaktik der Bruchrechnung für Lehrerausbildung und Lehrerfortbildung*. Heidelberg: Spektrum.

Peterson, P. L., Marx, R. W., & Clark, C. M. (1978). Teacher planning, teacher behavior, and student achievement. *American Educational Research Journal, 15*(3), 417–432.

Phillips, L. H., & Rabbitt, P. M. A. (1995). Impulsivity and speed-accuracy strategies in intelligence test performance. *Intelligence, 21*, 13–29.

Rölke, H. (2012). The Item Builder: A graphical authoring system for complex item development. In T. Bastiaens & G. Marks (Eds.), *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education* (pp. 344–353). Chesapeake: AACE.

Sabers, D. S., Cushing, K. S., & Berliner, D. C. (1991). Differences among teachers in a task characterized by simultaneity, multidimensionality, and immediacy. *American Educational Research Journal, 28*, 63–88.

Seidel, T., Blomberg, G., & Stürmer, K. (2010). „Observer"-validation of a video-based instrument to assess the professional view on classrooms. *Zeitschrift für Pädagogik*, 56. Beiheft, 296–300.

Sherin, M. G., Jacobs, V. R., & Philipp, R. A. (Eds.). (2011). *Mathematics teacher noticing. Seeing through teachers' eyes*. New York: Routledge.

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*, 1–22.

Van Es, E. A., & Sherin, M. G. (2002). Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *Journal of Technology and Teacher Education, 10*(4), 571–596.

Wartha, S. (2007). *Längsschnittliche Untersuchungen zur Entwicklung des Bruchzahlbegriffs*. Hildesheim: Franzbecker.

Weinert, F. E. (1996). „Der gute Lehrer", „die gute Lehrerin"im Spiegel der Wissenschaft: Was macht Lehrende wirksam und was führt zu ihrer Wirksamkeit? *Beiträge zur Lehrerbildung, 14*, 141–151.

Weinert, F. E., Schrader, F.-W., & Helmke, A. (1990). Unterrichtsexpertise: Ein Konzept zur Verringerung der Kluft zwischen zwei theoretischen Paradigmen. In L.-M. Alisch, J. Baumert, & K. Beck (Eds.), *Professionswissen und Professionalisierung* (pp. 173–206). Braunschweig: Copy-Center Colmsee.

Wu, H. (2008). Fractions, decimals, and rational numbers. University of California, Berkeley, Department of Mathematics. https://math.berkeley.edu/~wu. Accessed 12 Dec 2014.

5. Publikation II

Der zweite Artikel „Ein zeitbeschränkter Test zur schnellen Erkennung von Schülerfehlern durch Junglehrkräfte – Qualitative Merkmale zur Schwierigkeitsbestimmung von Items“, ist 2018 in der Zeitschrift *mathematica didactica* erschienen.

Der Beitrag fokussiert die zweite Phase der Beantwortung der Items, in der die Teilnehmer(innen) aus drei Schülerlösungen die falsche Lösung identifizieren sollen. Dies ist neben der Herausforderung durch die zeitliche Beschränkung auch inhaltlich anspruchsvoll.

In der theoretischen Fundierung des Beitrags wird neben dem Professionswissen von Lehrkräften ihr Wissen über Fehler und deren Erkennung dargestellt. Weiterhin wird die Itemschwierigkeit im Kontext der Itembeantwortung detailliert erläutert. Zwei Merkmale, welche die Itemschwierigkeit inhaltlich aufklären, werden genauer ausgeführt: die „curriculare Relevanz“ sowie die „strukturelle Komplexität“.

Im Rahmen der Erläuterung des methodischen Vorgehens werden die Studie TEDS-FU sowie die Stichprobe vorgestellt und das verwendete Testinstrument wird detailliert dargelegt. Anschließend erfolgt die Analyse der Items bezogen auf die beiden genannten Merkmale beispielhaft anhand von fünf Items. Neben der inhaltlichen Analyse, in der die Items von Ratern bewertet werden, erfolgt die Berechnung der jeweiligen Itemschwierigkeit.

Die Ergebnisse zeigen, dass die leicht und schwer zu lösenden Items mittels der Merkmale zur Itemschwierigkeit erklärt werden können, sowohl bezogen auf die gewählten Items der Sekundarstufe als auch auf die Items der Primarstufe.

Im Rahmen der Diskussion wird darauf hingewiesen, dass das Merkmal der curricularen Relevanz am robustesten zu sein scheint: Items aus Themengebieten, die im Spiralcurriculum häufig vorkommen und in einer früheren Klassenstufe bereits unterrichtet wurden, waren leicht zu lösen. Weiterhin kann ein Item, das viele Zeichen beinhaltet und zu dessen Lösung verschiedene Regeln angewandt werden müssen, aufgrund dieser strukturellen Komplexität als schwer gelten.

Die Merkmale sollten in weiteren Tests eingesetzt werden, um ihre Robustheit zu untersuchen und bei Bedarf weitere Tests zu entwickeln.

5.1. Darlegung des eigenen Anteils

Die Beitragsidee wurde nach Rücksprache mit meiner Betreuerin der Promotion Gabriele Kaiser eigenständig von mir entwickelt. Die Darstellung des theoretischen Rahmens, der Methodik und der Ergebnisse sowie die Überprüfung der merkmalsbezogenen Analysen habe im Wesentlichen ich verfasst. Die Ergebnisse und die Schlussfolgerungen wurden von Gabriele Kaiser und mir gemeinsam verfasst. Die Hinweise, die im Review angemerkt wurden, sind zunächst von mir und anschließend von Gabriele Kaiser eingearbeitet worden.

5.2. Abdruck der Publikation II

Pankow, L. & Kaiser, G. (2018). Ein zeitbeschränkter Test zur schnellen Erkennung von Schülerfehlern durch Junglehrkräfte – Qualitative Merkmale zur Schwierigkeitsbestimmung von Items. *mathematica didactica, 41*(2), 147-162.

Ein zeitbeschränkter Test zur schnellen Erkennung von Schülerfehlern durch Junglehrkräfte – Qualitative Merkmale zur Schwierigkeitsbestimmung von Items

LENA PANKOW, HAMBURG; GABRIELE KAISER, HAMBURG & BRISBANE

Zusammenfassung*: Zur Förderung von Lernchancen ist ein konstruktiver Umgang mit Schülerfehlern nötig, der das schnelle Erkennen von Schülerfehlern voraussetzt. Im Test zur schnellen Fehlererkennung, der im Rahmen der TEDS-FU-Studie durchgeführt wurde, waren Junglehrkräfte gefordert, typische Schülerfehler unter Zeitdruck zu erkennen. Zwei Merkmale – aus mathematikdidaktischer und aus wahrnehmungspsychologischer Perspektive – können in den beiden Tests für die Primar- und Sekundarstufe als schwierigkeitsgenerierend identifiziert und validiert werden. Die identifizierten Merkmale können als Analyseinstrument bei der Weiter- oder Neuentwicklung von Tests zur schnellen Fehlererkennung eingesetzt werden und bieten eine inhaltliche Grundlage zur Erklärung der Itemschwierigkeit.*

Abstract: *In order to promote learning opportunities, a fast recognition of students' errors is a necessary prerequisite for a constructive approach to students' errors. In the time-limited test for identifying students' errors from the TEDS-FU study, early career teachers were requested to recognize typical students' errors under time constraints. In the two tests for primary and secondary teachers, two characteristics could be identified as difficulty generating aspects, focusing the perspective of mathematics didactics and the psychology of perception. The identified characteristics can be used as an analysis tool for a new test for fast recognition of students' errors in the design process as well as the basis for the explanation of item difficulty.*

1. Einleitung und Forschungsfrage

Lehrkräfte haben tagtäglich mit Schülerfehlern aller Art zu tun, die sich im Mathematikunterricht z. B. in systematische und unsystematische Fehler unterscheiden lassen (Radatz, 1980). Die schnelle Wahrnehmung dieser Fehler im Unterricht und der adäquate Umgang mit den Fehlern sind für Lehrkräfte bedeutsam. Die Fehler weisen auf den Lernprozess der Schüler*innen hin und „lassen erkennen, wie das Denken funktioniert" (Beutelspacher, 2008, S. 87). Schoy-Lutz (2005) konnte zeigen, dass in 75 % der Schülerfehler im Mathematikunterricht echte Lerngelegenheiten verborgen sind. Voraussetzung zum Erkennen dieser Lerngelegenheiten ist das schnelle Identifizieren der Schülerfehler. Die Lehrkräfte müssen darüber hinaus, um diese Fehler als Lerngelegenheiten nutzen zu können, entscheiden, ob ein Fehler für alle Lernenden von Bedeutung ist oder ob dieser individuell zu klären ist. Neben der benötigten Kompetenz der Klassenführung muss die Lehrkraft daher flexibel mit der Fehlersituation umgehen (vgl. Klug et al., 2013) und zudem über das mathematische und mathematikdidaktische Wissen verfügen, um die Fehler nachzuvollziehen und einordnen zu können. Im Test zur schnellen Fehlererkennung, der im Rahmen der nationalen Folgestudie der internationalen Vergleichsstudie „Teacher Education and Development Study in Mathematics – TEDS-M 2008" (Blömeke et al., 2010) erhoben wurde, wurde die Kompetenz zur schnellen Fehlererkennung von Junglehrkräften im Fach Mathematik als ein Indikator für Lehrerexpertise erfasst, die im Folgenden vorgestellt wird (für Details zur TEDS-FU Studie siehe Blömeke et al., 2014). TEDS-FU setzte bei Lehrkräften der Primar- und der unteren Sekundarstufe zwei mathematisch unterschiedliche, aber strukturgleiche Tests zur schnellen Fehlererkennung ein. Es stellt sich a posteriori die Frage nach den schwierigkeitsgenerierenden Merkmalen der einzelnen Items, d. h. welche Merkmale eines Items machen dieses schwer bzw. leicht. Des Weiteren ist zu klären, ob diese Merkmale über die verschiedenen Lehramtstypen stabil sind, d. h. ob sich die gleichen Merkmale sowohl für Primarstufen- als auch für Sekundarstufenlehrkräfte zeigen.

Auf Basis einer mathematikdidaktischen Perspektive, aber auch einer wahrnehmungspsychologischen Perspektive aus der testbezogenen Diskussion über Fehlererkennungstests werden im Folgenden schwierigkeitsgenerierende Merkmale von Items theoriegeleitet entwickelt, die anhand der empirischen Ergebnisse überprüft werden. Die Analysen erlauben, zukünftig a priori die Itemschwierigkeit ähnlich konzipierter Tests zu bestimmen und bieten so die Möglichkeit, diese theoriegeleitet zu entwickeln. Zudem soll mit dem Artikel ein Beitrag zur Überprüfung der Güte des im Rahmen der TEDS-FU Studie entwickelten Testinstruments geleistet werden.

2. Theoretischer Rahmen

2.1 Lehrerprofessionswissen, Fehler und deren Erkennung

Die kognitiven Voraussetzungen, die Lehrkräfte für den Lehrberuf mitbringen müssen, werden in vielen Studien – wie in der COACTIV Studie (Kunter et al., 2011) oder in Studien der Michigan-Group (Hill, Ball & Schilling, 2008) – untersucht. In der TEDS-FU-Studie werden in Anlehnung an Shulman (1986) drei Wissensdimensionen des Lehrerprofessionswissens ausdifferenziert. Aufgrund ihrer empirisch gezeigten Relevanz für schnelle Fehlererkennung (Blömeke et al., 2014) werden im Folgenden zwei der drei Facetten genauer dargestellt. Die dritte Facette, General Pedagogical Knowledge (GPK), ist unter der Perspektive der Erkennung mathematischer Fehler kaum bedeutsam und wird deshalb im Folgenden nicht berücksichtigt.

Zum fachlichen Wissen, dem Content Knowledge (CK), gehört nach Shulman (1986) das Wissen, das den Inhalt umschließt und sich nicht auf reines Faktenwissen reduzieren lässt. Darüber hinaus soll die Lehrkraft auch die innere Struktur der Mathematik, die Grammatik und das Warum verstehen können (vgl. Shulman, 1986, S. 9). Für die Mathematik wird diese Wissenskomponente als *Mathematical Content Knowledge* (MCK) bezeichnet. Weiterhin unterscheidet Shulman (1986) das fachspezifisch-pädagogische Wissen, *Pedagogical Content Knowledge* (PCK), für die Mathematik *Mathematical Pedagogical Content Knowledge* (MPCK). Dieses Wissen beinhaltet

> for the most regularly taught topics in one's subject area, the most useful forms of representation of those ideas, the most powerful analogies, illustrations, examples, explanations, and demonstrations – in a word, the ways of representing and formulating the subject that make it comprehensible to others (Shulman, 1986, S. 9).

Des Weiteren hebt Shulman (1986, S. 9 f.) hervor, dass die Lehrkraft in Situationen, in denen sie mit Schülerfehlern konfrontiert wird, Strategien zur Behebung der Fehlvorstellungen haben muss, um das Wissen der Schüler*innen reorganisieren zu können.

Auch im Kompetenzmodell der COACTIV-Studie spielt das Erkennen von Schülerfehlern und der Umgang damit eine bedeutende Rolle, und zwar im fachdidaktischen Wissen, welches im Rahmen der Dimension Wissen über Schülervorstellungen ausdifferenziert wird in Wissen über „Fehlkonzeptionen, typische Fehler, Strategien" (Kunter et al., 2011, S. 38).

Fehler und der Umgang mit diesen spielen insbesondere im Kontext des Lernens von Mathematik eine große Rolle. Nach Oser, Hascher und Spychiger (1999) ist ein Fehler ein „von der Norm abweichender Sachverhalt oder ein von der Norm abweichender Prozess [...]. Normen stellen also das Bezugsystem dar und ohne Normen bzw. Regeln wäre es nicht möglich, fehlerhafte und fehlerfreie Leistungen, das Richtige vom Falschen zu unterscheiden" (S. 11). Heinze (2004) konkretisiert für den Mathematikunterricht Fehler als eine „Äußerung, die gegen die allgemeingültigen Aussagen und Definitionen der Mathematik sowie gegen allgemein akzeptiertes mathematisch-methodisches Vorgehen verstößt" (S. 223). Aufgrund der hohen Relevanz für mathematisches Lernen wird im Folgenden auf den Fehler des regelhaften Missverständnisses fokussiert; sog. Flüchtigkeitsfehler, die ggf. aufgrund mangelnder Aufmerksamkeit oder einer andersartigen Störung der Ausführung der Prozedur auftreten (Malle, 1993), werden aufgrund ihrer fehlenden Regelhaftigkeit nicht berücksichtigt.

Eine Fehlersituation wird von Seifried und Wuttke (2010) sowie von Heinze (2004) durch das Erkennen eines Fehlers definiert. Somit bildet die Identifikation des Fehlers bereits den ersten Schritt vor der vertieften Analyse des Fehlers (ähnlich Seidel & Prenzel, 2003). Die Wahrnehmung von Schülerfehlern ist darüber hinaus auch von dem erlernten Wissen über Schülerfehler beeinflusst (vgl. König et al., 2014). Aufgrund der Notwendigkeit, im Unterricht schnell zu reagieren, fokussieren wir im Folgenden auf schnelle Fehlererkennung als Voraussetzung für ein kompetentes Unterrichtshandeln und damit als eine zentrale Facette professioneller Kompetenz von Lehrkräften. Inhaltlich kann schnelle Fehlererkennung als eine Subkompetenz des umfassenderen Konstrukts der Diagnosekompetenz von Lehrkräften konzeptualisiert werden (Südkamp & Praetorius, 2017), die nach dem Ansatz von Blömeke, Gustafsson und Shavelson (2015) zur Kompetenz als Kontinuum als Teil der professionellen Unterrichtswahrnehmung, dem sog. Noticing, gezählt werden kann.

2.2 Itemschwierigkeit

Um die Kompetenzfacette ‚Schnelle Fehlererkennung' bei Lehrkräften empirisch erfassen zu können, ist es entweder möglich, die potentiell zu erkennenden Fehler induktiv aus vorliegendem Datenmaterial zu identifizieren oder deduktiv-systematisch aus der Literatur abzuleiten. Wir haben uns aufgrund fehlender Daten für die zweite Vorgehensweise entschieden. Eine wesentliche Komponente der Klassifikation der Itemschwierigkeit liegt in der Unterscheidung in eine formale und inhaltliche Komponente. Nach Kauertz (2008) bewirkt die formale Struktur in

Kompetenztests eine Schwierigkeit, die gezielt variiert oder konstant (wie in diesem Test: bei allen Items muss die falsche Schülerlösung identifiziert werden) gehalten werden solle. Besonders soll auf die inhaltlichen Aufgabenmerkmale der Items fokussiert werden, denen Schumann und Eberle (2011) ein bedeutendes Potenzial zur Aufklärung zuschreiben. Schumann und Eberle (2011) heben hervor, dass es möglich sein sollte, in einer nachträglichen Analyse der Items (post-Identifikation) eines Tests die beobachteten Itemschwierigkeiten in (fach-) didaktischer Hinsicht sinnvoll zu interpretieren. Auch Hartig (2007) betont dies als Ansatz zur Vorhersage schwierigkeitsbestimmender Aufgabenmerkmale, um etwas über die Validität eines Kompetenztests aussagen zu können. Angelehnt an Arbeiten von Cohors-Fresenborg et al. (2004) zu schwierigkeitsgenerierenden Merkmalen zur Einschätzung der Aufgabenschwierigkeit bei der PISA-Studie werden im Folgenden Merkmale entwickelt, die auf die Items des Tests zur schnellen Fehlererkennung von TEDS-FU angewendet werden sollen.

Das erste Merkmal aus einer mathematikdidaktisch geprägten Perspektive untersucht die curriculare Einführung des Themas und Häufigkeit des curricularen Auftretens im nachfolgenden Mathematikunterricht (im Folgenden als *curriculare Relevanz* bezeichnet). Die Inhalte der Schulmathematik werden im deutschen Mathematikunterricht im Sinne des Spiralprinzips nach Bruner (1973) zu verschiedenen Zeitpunkten und verschiedenen Kontexten wieder aufgegriffen (Brinkmann, 2002). Fehler zu Themenfeldern, die in den unteren Klassenstufen unterrichtet werden, können auf diese Weise häufiger von Lehrkräften beobachtet werden, als jene, die erst in höheren Schulstufen eingeführt werden. Items, die typische Fehler zu Themenfeldern darstellen, die in unteren Klassenstufen unterrichtet werden, sollten somit zu leichteren Items gehören, als jene, die in höheren Klassenstufen zum ersten Mal unterrichtet werden. Im Sinne des Spiralprinzips sollten ebenfalls die Items leichter lösbar sein, die nach dessen Einführung häufig unterrichtet werden, da sie in anderen Themenfeldern involviert sind.

Um die Schwierigkeit hinsichtlich des ersten Auftretens des Themas bis hin zur Wiederholung des Themas im weiteren Verlauf des Mathematikunterrichts einordnen zu können, wurden die Bildungspläne der einzelnen Bundesländer analysiert. Die Themen, die in den Bildungsplänen der Bundesländer der Primarstufe sowie der unteren Sekundarstufe für das Fach Mathematik aufgeführt werden, unterscheiden sich nur wenig in der Reihenfolge der Themengebiete, ggf. aufgrund der starken Orientierung an den von der KMK erlassenen Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss bzw. die Primarstufe, weshalb im Folgenden beispielhaft auf die Lehrpläne von Hamburg zurückgegriffen wird (vgl. Freie und Hansestadt Hamburg Behörde für Schule und Berufsbildung, 2012 a, b).

Das zweite Merkmal aus einer wahrnehmungspsychologischen Perspektive spiegelt den strukturellen Aufbau der Items wider. In dem genannten Merkmal (im Folgenden als *strukturelle Komplexität* bezeichnet) werden die Zeichen als Objekte – ihre Definition und ihre Grammatik (Rechenregeln/Gesetze) – näher betrachtet. Unter Bezug auf die in den Bildungsstandards formulierte allgemeine Kompetenz des Umgangs mit den symbolischen, formalen und technischen Elementen der Mathematik, gehen wir davon aus, dass die Zahlen und Symbole, aus denen die Items bestehen, seitens der Teilnehmenden wahrgenommen, interpretiert und im Gesamtzusammenhang beantwortet werden müssen. Dabei ist zu bemerken, dass die mathematischen Ausdrücke häufig in Gruppen oder Clustern zusammengefügt werden, die einer anderen Logik folgen als die der geschriebenen Sprache (vgl. Dyrvold, 2016, S. 11). Die Cluster, aus denen die Darstellungen der Items bestehen, sollen für die Analyse herangezogen werden, wobei hierarchische Beziehungen berücksichtigen werden (vgl. Drollinger-Vetter, 2011). Bei dem Gebrauch der Zeichen muss jede*r Lernende die Regeln zum Gebrauch der Zeichen – in diesem Fall nach den Konventionen der Mathematik – erwerben, um sie entsprechend verwenden zu können, gleiches gilt für die Verwendung von Regeln (vgl. Brunner, 2013, S. 54). Um die Itemschwierigkeit mittels des zweiten Merkmals zu identifizieren, wird im ersten Schritt die Anzahl der verwendeten Zeichen, aus denen die Items bestehen, berücksichtigt. Die Zeichenlänge wird als komplexitätsgenerierend angesehen, da innerhalb einer kurzen Testsituation (4 Sekunden) alle Zeichen wahrzunehmen sind, was eine hohe Anforderung an die Reaktionsfähigkeit, die kognitive Aufnahme- und Verarbeitungsgeschwindigkeit darstellt. Dabei ist zu bemerken, dass Bedeutungseinheiten wie die Buchstabenfolge ‚sin' als ein Zeichen gewertet werden, während die Zahl 0,5 als zwei gesonderte Zeichen gezählt werden, da die Ziffern gesondert in das Stellenwertsystem eingeordnet werden müssen und damit zusätzliche kognitive Verarbeitungsschritte nötig sind. Im zweiten Schritt werden die zu verwendeten Rechenregeln, die zur Lösung des Items benötigt werden, festgestellt.

3. Methode

3.1 Studie und Stichprobe

TEDS-FU ist eine deutsche Folgestudie zu der 2008 durchgeführten internationalen Lehrerbildungsstudie TEDS-M (Teacher Education and Development Study in Mathematics). Die 2012 durchgeführte Studie TEDS-FU (vgl. Blömeke et al., 2014) erhob die Kompetenz von Junglehrkräften nach vier Jahren Berufserfahrung. TEDS-FU wurde in Form einer Onlinestudie mithilfe des vom DIPF (Deutsches Institut für Internationale Pädagogische Forschung) entwickelten CBA ItemBuilder (Rölke, 2012) durchgeführt. An zwei Tests, die für die Primarstufen- und Sekundarstufenlehrkräfte strukturgleich gestaltet, aber inhaltlich an das studierte Lehramt angepasst sind, nahmen 171 Sekundar- und 131 Primarstufenlehrkräfte teil. Neben den Instrumenten zur Erhebung des mathematischen Fachwissens, des fachdidaktischen Wissens, des allgemeinpädagogischen Wissens und ausgewählter Belief-Facetten aus TEDS-M, wurden Tests zur Erfassung der situierten Facetten von Lehrerprofessionalität neu entwickelt und eingesetzt (vgl. Blömeke et al., 2014; Kaiser et al., 2017). Neben diesen videobasierten Instrumenten (vgl. Kaiser et al., 2015) wurden wie in der COACTIV-Studie Tests zur Identifikation von Schülerfehlern entwickelt (Krauss et al., 2008; Krauss & Brunner, 2011), die im Gegensatz zur COACTIV-Studie die schnelle Erkennung von Schülerfehlern (mit Zeitbegrenzung) fokussierten.

3.2 Testinstrument und Analyse der Items

Der neu entwickelte Test zum schnellen Erkennen von Schülerfehlern intendierte eine möglichst unterrichtsnahe Erfassung von Schülerfehlern. Da die Lehrperson üblicherweise das von ihr im Unterricht behandelte Themengebiet kennt, wurde vorweg das Themengebiet, aus dem die gezeigten Fehler kommen würden, eingeblendet mit der Aufforderung, typische Fehler zu antizipieren (vgl. Pankow et al., 2016).

Der Sekundarstufentest besteht aus 16 und der Primarstufentest aus 15 typischen Schülerfehlern, die auf einschlägig bekannte Fehlkonzepte zurückzuführen und für mathematisches Lehren und Lernen bedeutsam sind.

Im Anschluss an die Antizipationsphase wurden den Probanden drei Schülerlösungen präsentiert, von der eine Lösung einen typischen Schülerfehler im Kontext des angekündigten Themengebietes enthält, während die anderen beiden Lösungen korrekt sind.

Die Buchstaben, zu den dargestellten Schülerlösungen werden über die Tastatur der Probanden eingegeben, und zwar über die Buchstaben S, D, F, auf denen die Finger der linken Hand während des Tests ruhen sollen. Im ersten Schritt wurde der typische Schülerfehler angekündigt, z. B. „Anwenden des Satzes des Pythagoras“ mit einem abgebildeten Dreieck. Es wurde maximal 5 Minuten Zeit gegeben, um sich mögliche Schülerfehler in diesem Bereich zu vergegenwärtigen.

Im zweiten Schritt, der Identifikation der falschen Schülerlösung, wurden drei Lösungen gezeigt und es musste die Taste mit der falschen Schülerlösung gedrückt werden. Um zu verhindern, dass die Aufgaben nachgerechnet werden konnten, wurde die Zeit zur Antwort auf vier Sekunden beschränkt. Die Zeit, die einer Expertenbefragung zufolge benötigt wird, um die Ergebnisse nachzurechnen, liegt bei ca. fünf bis sechs Sekunden. Es wurde bei der Testkonstruktion davon ausgegangen, dass erfahrene Lehrkräfte wissen, welche typischen Fehler in dem betreffenden Themengebiet zu erwarten sind, d. h. ihre Fehlersuchstrategie ist strukturiert, sodass sie den Fehler ohne Rechnung schnell, auf einen Blick erkennen. Innerhalb von vier Sekunden musste daher der zu dem Fehler zugehörige Buchstabe gedrückt werden, um die Antwort in das System einzugeben. Nach dieser Zeit war es nicht länger möglich, eine Antwort einzugeben, und das Item wurde als falsch bewertet. Als Leistungsindikator wurde die Anzahl der richtigen Antworten herangezogen.

Im Folgenden wird eine exemplarische Auswahl der Sekundarstufenitems bzgl. ihrer Schwierigkeit analysiert (Abb. 1). Das erste Item (Nr. 15) befasst sich mit der Addition zweier Brüche, da empirisch aufgezeigt wurde, dass 45 % aller Schülerfehler auf dieses Vorgehen zurückgehen, d. h. es werden Zähler und Nenner getrennt addiert (vgl. Wartha, 2007).

Weiterhin konnte Padberg (2009) zeigen, dass einige Schüler*innen ein eigenes Regelwerk (vgl. Allmendinger et al., 2013, 81 f.) zur Addition zweier Brüche entwickeln. Eichelmann, Narciss, Schnaubert und Melis (2012) weisen drauf hin, dass Schüler*innen konzeptuelles Wissen über Brüche besitzen müssen, unter anderem zu der Bedeutung des Zählers und des Nenners, aber auch zur adäquaten Bruchvorstellung selbst (vgl. Eichelmann et al. 2012, S. 30).

Das zweite Item (Nr. 14) zur Umformung einfacher Zahlterme mit Klammern beinhaltet zwei primäre Schwierigkeiten: Zum einen ist das Ausmultiplizieren eines Terms, in dem Klammern enthalten sind, komplexer als bei einem Term ohne Klammern, da zwei Rechenschritte sowie eine Verknüpfung (Distributivgesetz) zu beachten sind.

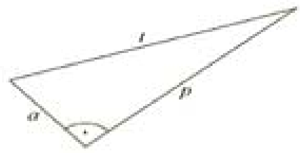
	Nr.	Ankündigung	„S“	„D“	„F“
Sekundarstufentest	15	Bruchrechnung: Addition zweier Brüche	$\frac{1}{2}+\frac{1}{4}=\frac{3}{4}$	$\frac{2}{7}+\frac{3}{7}=\frac{5}{14}$	$\frac{1}{3}+\frac{2}{3}=1$
	14	Umformung einfacher Zahlterme mit Klammern	$3 \cdot (5 + 2) = 3 \cdot 5 + 3 \cdot 2$	$3 + (5 + 2) = 3 + 5 + 2$	$3 - (5 + 2) = 3 - 5 + 2$
	7	Trigonometrie: Werte spezieller Winkel	$\sin 30° = 0{,}5$	$\cos 0° = 0$	$\tan 45° = 1$
	8	Anwenden der *p*-*q*-Formel bei der quadratischen Gleichung $x^2 + 8x - 4 = 0$	$x^2 + 8x - 4 = 0$		
			$x_{1,2} = -4 \pm \sqrt{16+4}$	$x_{1,2} = -4 \pm \sqrt{16-(-4)}$	$x_{1,2} = -4 \pm \sqrt{16-4}$
	5	Anwenden des Satzes des Pythagoras			
			$p^2 + a^2 = t^2$	$t^2 = a^2 + p^2$	$a^2 + t^2 = p^2$

Abb. 1: Auswahl (5 von 16) der Items zur schnellen Fehlererkennung aus dem TEDS-FU Sekundarstufentest

Zum anderen müssen Vorzeichen, wie in dem hier vorliegenden Fall, berücksichtigt werden, die nach Lüddecke (2015) auch unter anderem bei Problembearbeitungsprozessen Lösungsbarrieren beinhalten.

Das dritte Item (Nr. 7) aus der Trigonometrie kann unter zwei Sichtweisen interpretiert werden, was in der Literatur bisher kaum diskutierte Schwierigkeiten hervorrufen kann. So kann das Item einerseits funktional interpretiert werden als Frage nach den Funktionswerten von trigonometrischen Funktion zu speziellen Winkeln gegeben im Gradmaß. Andererseits ist auch eine geometrische, nichtfunktionale Auffassung möglich als Frage nach dem Verhältnis von Kathete und Hypothese im rechtwinkligen Dreieck zu speziellen Winkeln. Ein typischer Schülerfehler ist eine Verwechslung von cos 0° mit sin 0°, sodass cos 0° = 0 angegeben wird.

Das vierte Item (Nr. 8) enthält einen typischen Schülerfehler, der bei der Lösung einer quadratischen Gleichung mittels der sog. *p*-*q*-Formel $x_{1,2} = -\frac{p}{2} \pm \sqrt{\left(\frac{p}{2}\right)^2 - q}$ entstehen kann. Neben einer möglichen fehlenden Normierung (vgl. Ritter & Voß 2015, 82 f.) spielen die Vorzeichen der Parameter „*p*“ und „*q*“ eine zentrale Rolle (Goebbels & Ritter 2013, S. 55), die häufig nicht beachtet werden.

Das fünfte Item (Nr. 5) zum Anwenden des Satzes des Pythagoras beinhaltet den wohlbekannten Fehler der Vertauschung von Katheten und Hypotenuse im rechtwinkligen Dreieck, die durch die Nichtbeachtung der Definition der geometrischen Begriffe hervorgerufen sein kann oder durch die Fixierung auf übliche geometrische Darstellungen rechtwinkliger Dreiecke mit einer Kathete parallel zum unteren Rand der Darstellungsfläche (vgl. Draschoff, 2000).

3.3 Methodisches Vorgehen

Die Items beider Tests zur schnellen Fehlererkennung wurden basierend auf dem einparametrischen Raschmodell skaliert. Die EAP/PV-Reliabilität erreicht im Falle des Sekundarstufentests einen Wert von 0,62 und im Falle der Primarstufe von 0,64. Beide können als ausreichend angesehen werden (vgl. Magenheim et al., 2015; Robinson et al., 1991).

Zur Beantwortung der Forschungsfrage wird eine Analyse der Items durchgeführt, die bzgl. der Itemschwierigkeit extreme Ausprägungen aufweisen, da sie einerseits besonders einfach bzw. besonders schwer sind. Ein besonderer Fokus auf extreme Ausprägungen eines Merkmals oder auf Extremgruppen ist ein in der qualitativen Forschung weitverbreitetes Vorgehen. So werden z. B. im Rahmen der Typenbildung Teilgruppen gebildet, die hinsichtlich gewisser Merkmale besonders ähnlich sind, um Einsicht in allgemeinere Strukturen und Muster zu erhalten (vgl. Kelle & Kluge, 2010, S. 83 ff.). Auch in der quantitativen Forschung werden Items mit extremer Schwierigkeit bei der finalen Analyse ausgeschlossen, da sie aufgrund schwacher Diskrimination nicht substantiell zur Messung des Konstrukts beitragen (vgl. Bond & Fox, 2007).

Um die Merkmale zur Itemschwierigkeit zu bestimmen, bewerteten drei Rater die Merkmale curriculare Einführung des Themas und Häufigkeit des curricularen Auftretens, Zeichen als Objekte – ihre Definition und ihre Grammatik (wie bspw. Rechengesetze). Diese Kodierung fand nach einer Schulung der drei Rater im Rahmen eines niedrig inferenten Verfahrens statt. Die Rater erhielten Angaben zum Rating der curricularen Einführung der Themen der aufgeführten Items sowie eine Anleitung, wie das spätere Auftreten zu raten ist. Die Angaben, wie Zeichen und deren Gesetzmäßigkeiten zu zählen seien, wurden anhand von Beispielen verdeutlicht (für Details siehe Kodierhinweise am Ende des Beitrages).

Die Interraterreliabilität für die Kodierung der Merkmale ergab einen substantiellen Übereinstimmungswert (vgl. Landis & Koch, 1977) des Fleiss‘ Kappas ($\kappa = 0{,}61$). Der dargestellte Wert erhöhte sich bezogen auf die Interraterreliabilität bei den leichten und schwierigen Items noch deutlich und erreichte einen Wert von $\kappa = 0{,}80$.

4. Ergebnisse

Die Analyse der Itemschwierigkeit im Test zur schnellen Fehlererkennung geschieht im Folgenden in zwei Schritten. Zunächst werden in Abschnitt 4.1. die Itemschwierigkeiten zum Primar- und Sekundarstufentest zur schnellen Fehlererkennung dargestellt, die mithilfe der Software ConQuest (vgl. Wu et al., 1998) geschätzt wurden. In Abschnitt 4.2. werden zunächst die qualitativen Merkmale, die zur Erklärung der Itemschwierigkeiten herangezogen werden, beschrieben und beim Sekundarstufentest exemplarisch an fünf Items dargestellt. Abschließend werden die Merkmale auf den Test zur schnellen Fehlererkennung der Primarstufe übertragen. Für dieses Vorhaben werden vier Items der Primarstufe zunächst kurz vorgestellt und im Anschluss ebenfalls hinsichtlich der schwierigkeitsgenerierenden Merkmale analysiert.

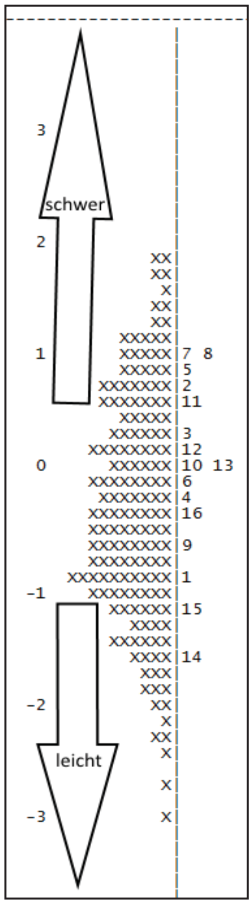
4.1 Schätzung der Itemschwierigkeiten

Die Schwierigkeit der einzelnen Items des Testinstruments wird mittels des Output des Programms ConQuest näher vorgestellt (Abb. 2 und 3).

Auf der linken Seite der Skala sind jeweils die teilnehmenden Personen repräsentiert, wobei jedes Kreuz einen Probanden repräsentiert.

Auf der linken Seite der Skala sind jeweils die teilnehmenden Personen repräsentiert, wobei jedes Kreuz einen Probanden repräsentiert.

Auf der rechten Seite kann anhand der Lage der Itemnummern die Schwierigkeit abgelesen werden. Die Schwierigkeitsgrade der Skala reichen von -3 bis +3. Items, die oberhalb von Null liegen, sind eher schwer, während Items, die unterhalb von Null liegen, eher leicht sind. Je weiter die Items von Null entfernt sind, desto seltener bzw. häufiger wurden sie richtig gelöst.



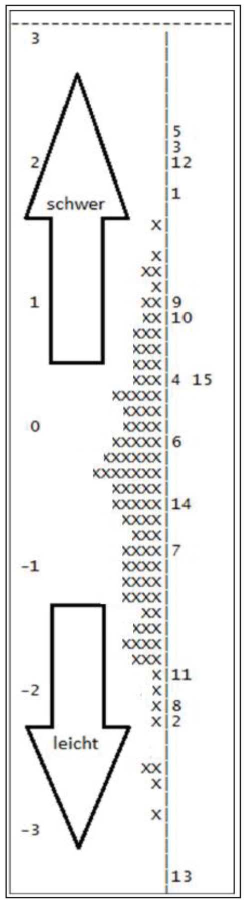


Abb. 2 (links): Itemschwierigkeiten des zeitbegrenzten Tests zur schnellen Fehlererkennung der Sekundarstufe

Abb. 3 (rechts): Itemschwierigkeiten des zeitbegrenzten Tests zur schnellen Fehlererkennung der Primarstufe

Die Outputs zeigen, dass die Itemstreuung für die Primarstufe (Abb. 3) gleichmäßiger über die Personenfähigkeiten verteilt ist als für die Sekundarstufe (Abb. 2). Dies bedeutet, dass für die Primarstufe Schätzungen in den Randbereichen besser möglich sind, während für die Sekundarstufe die Schätzungen im mittleren Schwierigkeitsbereich genauer sind.

Die Items Nr. 7 und Nr. 8 gehören im Sekundarstufentest (Abb. 2) zu den schwer zu lösenden Items, die Items Nr. 14 und Nr. 15 zu den leicht zu lösenden. Zur Beantwortung der Forschungsfrage soll in diesem Zusammenhang geklärt werden, wie die Schwierigkeit an beiden Enden der Skalen inhaltlich erklärt werden kann.

Im Vergleich zu dem Sekundarstufentest stellen die Items Nr. 2 und Nr. 8 im Primarstufentest (vgl. Abb. 3) Repräsentanten für leicht zu lösende Items dar, während die Items Nr. 1 und Nr. 12 als schwer zu lösende Items anzusehen sind. Hervorzuheben ist das Item Nr. 13, welches von allen Teilnehmer*innen richtig gelöst und daher als zu niedrig differenzierend ausgeschlossen wurde.

4.2 Qualitative Analysen zur Itemschwierigkeit

Im Folgenden werden die Merkmale zunächst auf die Items der Sekundarstufe bezogen, gefolgt von den Primarstufenitems.

Das Merkmal *curriculare Relevanz*, das die Einführung des Themas und Häufigkeit des curricularen Auftretens im weiteren Mathematikunterricht fokussiert, wird – wie in der Tab. 1 dargestellt – hinsichtlich zweier Aspekte untersucht: Zum einen wird die Klassenstufe des ersten Auftretens im Curriculum analysiert, zum anderen wird die curriculare Häufigkeit der Verwendung in anderen Themengebieten nach der Einführung des Begriffs betrachtet. Dies wird im Folgenden für das Themengebiet „Addition zweier Brüche“ (Item Nr. 15 aus Abb. 1) exemplarisch ausgeführt. Das Thema ‚Addition zweier Brüche‘ tritt curricular in der Sekundarstufe insgesamt in 48 der 114 Themen, die unterrichtet werden, auf (ca. 40 %).

Um dies zu analysieren, wurde das Beispiel in einem schulinternen Curriculum für die Stadtteilschule (ein Zusammenschluss der ehemaligen Haupt- und Realschulen) Mathematik“ (vgl. Freie und Hansestadt Hamburg Behörde für Schule und Berufsbildung, 2012b) analysiert.

Die Unterrichtsvorhaben für die Jahrgänge 5-10 sind in diesem Beispielcurriculum in den Doppeljahrgängen 5/6 sowie 7/8 und 9/10 aufgeführt. Die Themen werden weiter in Inhalte differenziert. So wird das fünfte Thema in Jahrgang 5/6 „Einteilen – Verteilen“ in sieben Unterthemen eingeteilt wie bspw. Grundvorstellungen, Darstellungen, Größenvergleiche, Umwandlung von Bruch in Dezimalzahl und Grundrechenarten in Verbindung mit Brüchen und Dezimalzahlen.

Insgesamt entstehen so 114 Themen. Bei jedem der Fehler, die in beiden Tests aufgeführt wurden, war es so möglich, durch drei Rater zu bewerten, ob der Fehler in den Themengebieten des Unterrichts enthalten war oder nicht. Ähnliche curriculare Auswertungen wurden für die anderen Items ebenfalls vorgenommen, die Ergebnisse sind in Tab. 1 dargestellt.

	Merkmal / Item	empirische Schwierigkeit (Rasch-Wert -3 (leicht) bis +3 (schwer) aus Abb. 2)	Einführung des Themas [in Klassenstufe(n)]	Häufigkeit curricularen Auftretens in anderen Themen ab Einführung
Sekundarstufen Items	Addition zweier Brüche (Item 15)	-1	5/6	ca. 40 %
	Umformen mit Klammern (Item 14)	-1,5	5	ca. 25 %
	Trigonometrie (Item 7)	1	9/10	ca. 3 %
	Quadratische Gleichung (*p*-*q*-Formel) (Item 8)	1	9/10	ca. 6 %
	Anwendung des Satzes des Pythagoras (Item 5)	1	9/10	ca. 4 %

Tab. 1: Analysen zum Merkmal *curriculare Relevanz*

Item \ Merkmal		empirische Schwierigkeit (Rasch-Wert -3 (leicht) bis +3 (schwer) aus Abb. 2)	Merkmal: Zeichen je Antwort	Merkmal: Zeichen je Item	Merkmal: Regeln
Sekundarstufen Items	Addition zweier Brüche (Item 15)	-1	10 bis 11	32	3
	Umformen mit Klammern (Item 14)	-1,5	13 bis 15	41	4
	Trigonometrie (Item 7)	1	5 bis 6	16	9
	Quadratische Gleichung (*p*-*q*-Formel) (Item 8)	1	9 bis 15	48	8
	Anwendung des Satzes des Pythagoras (Item 5)	1	14	29	7

Tab. 2: Analysen zum Merkmal *strukturelle Komplexität*

Zusammenfassend ist erkennbar, dass Items in dem zeitbeschränkten Test zur schnellen Fehlererkennung, die sich auf curriculare Inhalte in niedrigeren Klassenstufen beziehen (wie die Items Nr. 14 und Nr. 15), häufiger richtig beantwortet wurden als jene, die in höheren Klassenstufen unterrichtet wurden (Items Nr. 7, Nr. 8 und Nr. 5). Weiterhin zeigt sich, dass Items zu curricularen Themen, die häufig in nachfolgenden Klassenstufen wiederaufgenommen werden, häufiger richtig beantwortet wurden. Insgesamt tragen diese beiden Merkmale dazu bei, dass Items als schwer oder leicht angesehen werden können.

Die Auswertung des Merkmals *strukturelle Komplexität*, in dem Zeichen als Objekte – d. h. ihre Definition und ihre Grammatik (wie bspw. Rechengesetze) – analysiert werden, ist in der Tab. 2 wiedergegeben. Bezüglich des genannten Merkmals sind zwei unterschiedliche Aspekte zu analysieren. Zunächst werden die wahrzunehmenden Zeichen der Items beschrieben, während im Anschluss die zu verwendenden Regeln näher betrachtet werden, die zu einer Lösung als bekannt vorausgesetzt und korrekt angewandt werden müssen. Aus Platzgründen werden nur exemplarisch die Items ‚Addition zweier Brüche', Quadratische Gleichungen (*p*-*q*-Formel)' sowie ‚Anwendung des Satzes des Pythagoras' detailliert dargestellt.

Im Item zur „Bruchrechnung" müssen die Brüche zunächst als solche identifiziert werden und im Anschluss hinsichtlich der Operation betrachtet werden. So bestehen die Bruchzahl ½ wie auch die Bruchzahl ¼ und das Ergebnis jeweils aus insgesamt drei Zeichen. Es verbleiben der Operator und das Gleichheitszeichen. Das Gleichheitszeichen wird als ein Zeichen interpretiert. Insgesamt werden so 10–11 Zeichen je Antwortmöglichkeit erreicht. Wenn man die drei Antwortmöglichkeiten addiert, handelt es sich auf diese Weise um ca. 32 Zeichen, die im Rahmen dieses Items wahrzunehmen sind. Die verwendeten Regeln, die in diesem Item Anwendung finden, sind die folgenden: Bildung und Umgang mit dem Hauptnenner sowie die Addition der Zähler und anschließendes Kürzen. Eine Analyse des Items zur ‚Lösung der quadratischen Gleichung mittels der *p*-*q*-Formel' zeigt, dass bei diesem Item die meisten Zeichen auftreten. Das Item besteht aus einer zusammengesetzten Antwort: Zunächst ist der übergeordnete Ausdruck, der aus 9 Zeichen besteht, wahrzunehmen, die zu den Antwortmöglichkeiten hinzuaddiert werden müssen, da der Ausdruck die Voraussetzung zur Lösung des Items darstellt. Zwei der Antwortmöglichkeiten umfassen 12 Zeichen, da das Symbol ‚$\pm$' als ein Zeichen gezählt wird. Die dritte Antwortmöglichkeit enthält 15 Zeichen, da neben zwei Klammern noch ein weiteres Vorzeichen hinzukommt. Neben der Zeichenanzahl werden ebenfalls die Rechengesetze überprüft: Die Indizes in dem Ausdruck ‚$x_{1,2}$' benötigen Wissen über die Verwendung dieses Ausdrucks, nämlich, dass zwei Ergebnisse erwartet werden müssen. Ebenfalls sollte die Verwendung des ‚$\pm$'-Zeichens bekannt sein. Die Verwendung des Rechenzeichens vor Klammern gehört ebenfalls in die Kategorie der als bekannt vorauszusetzenden Rechenregeln. Des Weiteren ist der Umgang mit Wurzeln von Bedeutung, da die Ergebnisse nicht einzeln berechnet werden dürfen, sondern aus dem gesamten Ausdruck die Wurzel zu ziehen ist. Bei dem Ziehen der Wurzel muss die Kenntnis, dass es im Bereich der reellen Zahlen nicht möglich ist, aus negativen Zahlen die Wurzel zu ziehen, vorhanden sein, um ein falsches Ergebnis einordnen zu können. Weiterhin ist Wissen zum Umgang mit Brüchen bzw. deren Umwandlung in Dezimalzahlen umgehen zu können.

Das Item zur ‚Anwendung des Satzes des Pythagoras' ist zusammengesetzt: Zunächst muss das Bild eines rechtwinkligen Dreiecks samt den Bezeichnungen wahrgenommen werden. Dieses besteht aus fünf Zeichen. Im Anschluss müssen die drei Antwortmöglichkeiten mit je acht Zeichen erfasst werden. Die Buchstaben je Antwortmöglichkeit sowie die Quadrate gelten als einzelne Zeichen, die wahrgenommen werden müssen. Darüber hinaus müssen Regeln zum Umgang mit den Zeichen inkorporiert sein, da neben dem Quadrieren ebenfalls die Überlegungen zu den Eigenschaften eines Dreiecks abgerufen werden müssen. Außerdem müssen die Voraussetzungen für die Anwendung und die Definition der Hypotenuse und der Katheten des Satzes des Pythagoras bekannt sein und angewendet werden. Abschließend müssen auch Kenntnisse zum Wurzelziehen und zur Gleichungsumformung angewendet werden.

Zusammenfassend ist erkennbar, dass das Merkmal *strukturelle Komplexität* geeignet ist, aus einer wahrnehmungspsychologischen Perspektive die Komplexität von Items vorherzusagen. Die Anzahl der Zeichen in einem Item beeinflusst die Schwierigkeit des Items, da die Zeit, in der die vorhandenen Zeichen wahrgenommen werden müssen, begrenzt ist. Allerdings ist nicht die reine Anzahl von Zeichen entscheidend, die die Schwierigkeit eines Items beeinflusst, sondern auch die Anzahl der anzuwendenden Regeln und somit die Art der Verknüpfungen zwischen den Zeichen. Die Subkategorien des Merkmals *strukturelle Komplexität* beeinflussen einander, so dass zu erkennen ist, dass die Items mit einer hohen Zeichenanzahl oder Anzahl von Regeln schwieriger sind als Items mit einer niedrigen Zeichenanzahl oder Anzahl von Regeln.

Abschließend lässt sich feststellen, dass sich bei den Analysen über alle Items der Sekundarstufe ein ähnliches, aber nicht so starkes Bild ergibt wie bei den Items aus der Extremgruppe der schwierigen bzw. einfachen Items. Ersichtlich spielen bei Items mittleren Schwierigkeitsgrads unterschiedliche Merkmale eine Rolle, so dass die zwei ausgewählten Merkmale an Bedeutung verlieren.

Allerdings weisen die Ergebnisse auch bei den mittelschweren Items in dieselbe Richtung. Insgesamt erscheinen damit weitere Analysen mit einem größeren Itempool nötig, um die oben vorgestellten Analysen zu erhärten. Aufgrund dieser Einschränkung wurden dieselben Analysen am Primarstufentest durchgeführt, um damit insgesamt zu einer größeren Breite in den Analysen zu gelangen und zusätzlich die Stabilität der Ergebnisse unabhängig von der unterrichteten Schulstufe der Testbeteiligten zu überprüfen.

5. Überprüfung der merkmalsbezogenen Analysen mithilfe der Items des Primarstufen-Tests

Die Merkmale zur Erklärung der Itemschwierigkeit wurden am Test zur schnellen Fehlererkennung aus der Sekundarstufe entwickelt und angewendet. Um zu abgesicherten Aussagen zu kommen, auch gerade bzgl. der Abhängigkeit von der unterrichteten Schulstufe, wird die Angemessenheit der schwierigkeitsgenerierenden Merkmale für Items anhand des Primarstufentests zur schnellen Fehlererkennung überprüft.

5.1 Vorstellung der Primarstufenitems

Die Items, die in dem Test zur schnellen Fehlererkennung verwendet wurden, beziehen sich auf Fehler, die im Grundschulunterricht typischerweise vorkommen. Zunächst sollen die Items mit niedriger Schwierigkeit dargestellt werden (Abb. 4).

Der Fehler mit der Null (Item Nr. 2) tritt zunächst bei der Bedeutungszuweisung der Null durch Grundschüler*innen auf. Das neutrale Element bei der Addition und der Subtraktion wird als ein solches auf die Multiplikation übertragen (Padberg & Benz, 2011). Der Fehler mit der Null ist einer der „häufigsten Einmaleins-Fehler“ (ebd., S. 147 f.). Eine weitere mögliche Ursache ist, dass die Schüler*innen eine falsche Vorstellung von der Null haben, sodass diese als „Nichts“ interpretiert wird und die 12 als Ergebnis angegeben wird, obwohl eine Multiplikation angegeben ist. Des Weiteren können Lernende die Vorstellung haben, dass das Ergebnis einer Multiplikation mindestens so groß sein muss wie der größte Faktor.

Beim Erkennen von Symmetrieachsen (Item Nr. 8) ist zu berücksichtigen, dass nach Franke und Reinhold (2016) die Lösungshäufigkeit sowohl von der Figur als auch der Lage der Achse abhängig ist. Sowohl bei Schulanfänger*innen als auch bei älteren Schüler*innen werden Schwierigkeiten sichtbar, wenn eine Spiegelachse schräg zur Bildkante verläuft. Eine Entwicklung des Symmetrieverständnisses ist daher auf vielfältige Weise durch eigene Erfahrungen zu vermitteln (vgl. Franke & Reinhold, 2016). In dem Test zur schnellen Fehlererkennung stellen die Lösungen „D“ und „F“ die Parallelität zu den Seitenkanten dar, sodass keine Schwierigkeit bei der Achsenerkennung zu erwarten ist.

Im Gegensatz dazu steht die Lösung „S“, in der die Spiegelachse die Diagonale darstellt. Die als schwierig identifizierten Items fokussieren andere Aspekte, z. B. Rechnen mit der Null.

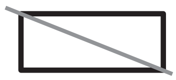
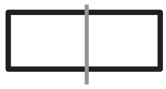
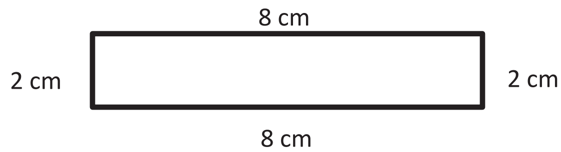
	Nr.	Ankündigung	„S“	„D“	„F“
Primarstufenitems	1	Schriftliche Multiplikation	31·24 62 124 …..	31·24 62 124 …..	31·24 620 124 …..
	2	Rechnen mit der Null	12 – 0 = 12	12 · 0 = 12	12 + 0 = 12
	8	Eine Symmetrieachse einzeichnen			
	12	Umfang eines Rechtecks	 U = 2 · 8 cm	U = 2 · 10 cm	U = 2cm+8cm+2cm+8cm

Abb. 4: Auswahl (4 von 15) der Items zur schnellen Fehlererkennung aus TEDS-FU Primarstufentest

Sowohl die Arbeiten von Gerster (1982), wie später auch die von Padberg und Benz (2011), analysieren die Schwierigkeiten der schriftlichen Multiplikation (Item Nr. 1). Sie weisen neben Schwierigkeiten bei der Multiplikation mit der Null auf die Schwierigkeit des Umgangs mit dem Stellenwert hin, also auf die Möglichkeit der falschen Anordnung der errechneten Teilprodukte, sog. Stellenwertfehler (Padberg & Benz, 2011, S. 280). Der aufgeführte Fehler kann, so Padberg und Benz (2011), als systematischer und somit als typischer Fehler angesehen werden, da die ggf. nicht verstandene Bedeutung des Ausrückens der Ziffernreihe die Stellenwerte der einzelnen Rechnungen außer Acht lässt.

Weiterhin kann das frühe Weglassen der Endnull in der ersten Rechenzeile für die Fehlvorstellung verantwortlich sein.

Am Ende der Grundschulzeit werden Flächen und Umfänge von geometrischen Figuren berechnet, auf die in Item 12 Bezug genommen wird. Im Rahmen dieser Berechnungen spielt die Verwendung der korrekten Begriffe des Umfangs und des Flächeninhaltes und deren Unterscheidung eine Rolle. Franke und Reinhold (2016) weisen darauf hin, dass die Begriffe aufgrund der fehlenden Begriffsvorstellung häufig verwechselt werden (Franke & Reinhold, 2016). Auch wenn eine algebraische Betrachtung für die Berechnung des Flächeninhalts und des Umfangs in der Grundschule noch nicht betrachtet werden, kann damit das Verständnis für Formeln zu den einschlägigen Berechnungen gefördert werden (Franke & Reinhold, 2016).

5.2 Überprüfung der merkmalsbezogenen Analyse der Aufgabenschwierigkeit mit Hilfe der Items des Primarstufen-Tests

Im Folgenden werden die beiden Merkmale zur Analyse der Aufgabenschwierigkeit, die für die Items des Sekundarstufentest entwickelt wurden, auf die Items der Primarstufe angewendet, um die Validität des Analyseinstruments zu überprüfen. Zunächst wird das Merkmal *curriculare Relevanz*, d. h. Einführung des Themas und Häufigkeit des curricularen Auftretens im weiteren Mathematikunterricht analysiert (Tab. 3).

Die schriftliche Multiplikation wird in Grundschulen in der 3./4. Klasse unterrichtet (vgl. Freie und Hansestadt Hamburg Behörde für Schule und Berufsbildung, 2012a, S. 3) und wird dann in weiteren acht von 42 möglichen Unterrichtsinhalten (vgl. Tab. 3) curricular aufgegriffen. Ebenso verhält es sich mit dem Themengebiet ‚Umfang eines Rechtecks‘ (Item 12).

Dieses Thema wird in der 4. Klasse unterrichtet und wird nach der Einführung in vier weiteren Themengebieten wiederaufgegriffen. Das Rechnen mit der Null, welches in Item Nr. 2 fokussiert wird, wird thematisch bei der Einführung der Addition in der 1. Klasse im Mathematikunterricht thematisiert und wird im Anschluss mit dem Bezug auf die weiteren Rechenarten sukzessive weitergeführt, sodass es insgesamt in 25 der 42 Unterrichtsthemen wiederaufgegriffen wird. Umgang mit Symmetrie (Item Nr. 8) wird zwar spielerisch in der 1.–2. Klassenstufe mit Hilfe von einfachen Spiegelungen an Flächen durchgeführt, lässt sich allerdings im weiteren Verlauf der Schullaufbahn nur in neun der 42 möglichen Unterrichtsinhalte finden.

Item \ Merkmal		empirische Schwierigkeit (Rasch-Wert -3 (leicht) bis +3 (schwer) aus Abb. 3)	Merkmal Zeichen je Antwort	Zeichen je Item	Regeln
Primarstufen Items	Schriftliche Multiplikation (Item 1)	1	7	21	6
	Fehler mit der Null (Item 2)	-2	5	15	2
	Symmetrie (Item 8)	-2	2	6	2
	Umfang eines Rechtecks (Item 12)	2	9 (Rechteck) +6-13 (abhängig von der Antwortmöglichkeit)	34	3

Tab. 3: Analysen zum Merkmal *curriculare Relevanz*

Item \ Merkmal		empirische Schwierigkeit (Rasch-Wert -3 (leicht) bis +3 (schwer) aus Abb. 3)	Merkmal Klassenstufe	Häufigkeit
Primarstufen Items	Schriftliche Multiplikation (Item 1)	1	3/4	ca. 25 %
	Fehler mit der Null (Item 2)	-2	1–3	ca. 60 %
	Einzeichnen der Symmetrieachse (Item 8)	-2	1/2	ca. 25 %
	Umfang eines Rechtecks (Item 12)	2	4	ca. 10 %

Tab. 4: Analysen zum Merkmal *strukturelle Komplexität*

Die Ergebnisse der Analysen zum Merkmal *strukturelle Komplexität*, d. h. die Auffassung von Zeichen als Objekte – ihre Definition und ihre Grammatik (wie bspw. Rechengesetze), werden zunächst übersichtsartig (Tab. 4) dargestellt und anschließend beispielhaft an Item 1 und 12 inhaltsbezogen analysiert.

Das Item zur „schriftlichen Multiplikation" (Item Nr. 1) enthält fünf Zeichen bestehend aus zwei- bis dreistelligen Zahlen, die mit einer Multiplikation verbunden sind und im Anschluss untereinander angeordnet sind. Zusammengefasst handelt es damit auf diese Weise um 21 Zeichen, die im Rahmen dieses Items wahrzunehmen sind. Weiterhin müssen die Regeln zur schriftlichen Multiplikation von zweistelligen Zahlen berücksichtigt werden, ebenso wie folgende Regeln: Rechne von rechts nach links, verwende die Kenntnisse aus dem 1x1, berücksichtige eventuelle Überträge, beachte die Kommutativität der Multiplikation, beachte die stellenwertberücksichtigende Notation.

Das vierte Item zum ‚Umfang eines Rechtecks' (Item Nr. 12) besteht aus zwei Teilen, der Abbildung eines Rechtecks (ein Zeichen) sowie der jeweiligen Maßeinheiten. Das Rechteck kann als ein Zeichen gezählt werden, da dies als Gesamtfigur identifiziert werden muss.

Insgesamt, inkl. der zu identifizierenden Maßeinheiten an den Seiten des Rechtecks, handelt es sich um 9 Zeichen. Je Antwortmöglichkeit erhöhen sich die wahrzunehmenden Zeichen auf 15 bis 22 Zeichen. Bei allen drei Antworten handelt es sich auf diese Weise um 34 Zeichen, die im Rahmen dieses Items zu erkennen sind. Für die Lösung des Items sind weiterhin Definitionen und Rechenregeln einzuhalten. So ist es notwendig, dass die Regeln des Einmaleins verinnerlicht sind, wie auch die Formel zur Berechnung des Umfangs eines Rechtecks, weiterhin ist die Verwendung der Einheiten zur korrekten Darstellung notwendig.

6. Diskussion der Ergebnisse, Grenzen der Studie und Schlussfolgerungen

Die quantitative Auswertung der Itemschwierigkeit (Abb. 2 und 3) konnte die unterschiedlichen Schwierigkeiten der Items, die in beiden Tests zur schnellen Fehlererkennung verwendet wurden, zeigen. Die Items, die einen hohen sowie einen niedrigen Schwierigkeitsgrad aufweisen, wurden genauer analysiert, da davon ausgegangen wurde, dass gerade bei stark oder schwach ausgeprägtem Schwierigkeitsgrad theoretisch entwickelte Erklärungsansätze greifen. Es zeigt sich, dass die literaturgeleiteten Merkmale zur Erklärung der qualitativen Schwierigkeit geeignet sind, die Itemschwierigkeit inhaltlich zu erklären (Tab. 5).

Das aus der mathematikdidaktischen Perspektive entwickelte Merkmal der *curricularen Relevanz* (d. h. curriculare Einführung des Themas und Häufigkeit des curricularen Auftretens im weiteren Mathematikunterricht) kann die Itemschwierigkeit der analysierten Items des Sekundarstufentests inhaltlich aufklären. Es weisen solche Items eine geringe Itemschwierigkeit auf, die curricular früh eingeführt werden und anschließend im Rahmen eines spiralförmig geprägten Schulcurriculums häufig wieder aufgegriffen werden. Die Themen der Items, die später im Curriculum vorkommen und damit wohl auch später und damit seltener unterrichtet werden, konnten als schwieriger identifiziert werden. Dieses Merkmal konnte auf den strukturgleichen Primarstufentest übertragen werden, wo sich dieselben Ergebnisse zeigten. Dabei ist zu berücksichtigen, dass die Häufigkeit des Aufgreifens eines Themas in anderen curricularen Themenfeldern keinen exakten Wert darstellt, da die Werte auf Grundlage eines Ratings erfasst wurden. Die Daten ermöglichen aber eine recht gesicherte Einsicht in die Häufigkeit des curricularen Auftretens eines Themengebiets.

Das aus einer wahrnehmungspsychologischen Ebene entwickelte Merkmal *strukturelle Komplexität* (d. h. Auffassung von Zeichen als Objekte – ihre Definition und ihre Grammatik) ist ebenfalls gut geeignet, die Itemschwierigkeit aufzuklären. So ist die Anzahl der Zeichen in Kombination mit ihren Definitionen und Rechengesetzen zu sehen. Die Anzahl der Zeichen kann generell wie folgt ausgewertet werden: Je weniger Zeichen in der begrenzten Zeit wahrgenommen werden müssen, desto häufiger wird das Item richtig gelöst. Allerdings muss darüber hinaus auch die zweite Komponente des Merkmals beachtet werden: Je mehr Definitionen und Rechengesetze das Item zur Lösung benötigt werden, desto schwieriger ist das Item trotz weniger Zeichen, wie sich beispielsweise in dem Item zu ‚Trigonometrischen Funktionen bzw. Trigonometrie im rechtwinkligen Dreieck' zeigt. Bei diesem Item müssen insgesamt lediglich fünf Zeichen je Antwortkategorie wahrgenommen werden, allerdings ist es notwendig, die Werte der trigonometrischen Funktionen bzw. die entsprechenden Definitionen der Verhältnisse von Katheten und Hypotenuse im rechtwinkligen Dreieck zu kennen und die damit in Verbindung stehenden Voraussetzungen.

Die Ergebnisse zur Aufklärung der Itemschwierigkeit mittels der identifizierten Merkmale konnten von dem Test der schnellen Fehlererkennung der Sekundarstufe auf den Test der Primarstufe übertragen werden (Tab. 5).

Das Merkmal der *curricularen Relevanz* und damit der curricularen Verankerung und Vertrautheit mit dem Thema scheint am robustesten zu sein, da weder im Primar- noch im Sekundarstufentest Ausnahmen zu erkennen sind. Die Items, die positive Werte in der ersten Spalte besitzen, gehören zu den schweren Items, während die Items mit negativen Ziffern zu den leichten Items gehören. Bei den Primarstufen-Items haben die Items ‚Schriftliche Multiplikation' und ‚Einzeichnen der Symmetrieachse' den gleichen Wert zum curricularen Auftreten in der weiteren Schullaufbahn, aber aus unterschiedlichen Gründen. Dies ist vermutlich dadurch bedingt, dass die ‚Schriftliche Multiplikation' erst am Ende der dritten bzw. zu Beginn der vierten Klasse erfolgen kann, da sie auf den anderen Rechenarten beruht, wohingegen Symmetrieachsen aufgrund ihrer relativ niedrigen Bedeutung im Grundschulcurriculum erst spät eingeführt werden. Des Weiteren können die zwei Teilkomponenten des Merkmals Zeichenanzahl und Anzahl der anzuwendenden Regeln nicht losgelöst voneinander betrachtet werden. So besteht das Item zum Umformen mit Klammern aus 41 Zeichen, benötigt allerdings vier Regeln zum Lösen.

Bezüglich der Zeichen würde es sich um ein schweres Item handeln, bzgl. der Regeln um ein leichtes Item. Umgekehrt würde es sich bei dem Item zur Trigonometrie um ein leichtes Item bzgl. der Zeichenanzahl handeln, da neun Regeln anzuwenden sind, um es zu lösen. Zusammenfassend können die Subfacetten des Merkmals die Itemschwierigkeit qualitativ erklären.

Die Streuung der Itemschwierigkeiten wird in den quantitativen Analysen deutlich, allerdings nicht inhaltlich erklärt. Die Analysemerkmale für Schüleraufgaben im Unterricht, die von Draxler (2006) entwickelt und durch Schumann und Eberle (2011) ergänzt wurden, konnten an die Gegebenheiten des vorliegenden Tests angepasst und analysiert werden.

Mithilfe der vorgestellten Merkmale, d. h. das aus einer mathematikdidaktisch geprägten Perspektive der *curricularen Relevanz* und das aus einer wahrnehmungspsychologisch geprägten Perspektive der *strukturellen Komplexität*, zur Aufklärung der Itemschwierigkeit in dem zeitbeschränkten Test war es möglich, die Unterschiede in der Itemschwierigkeit qualitativ aufzuklären. Obwohl die schwierigkeitsgenerierenden Merkmale für Items der Sekundarstufe entwickelt wurden, konnten damit auch die Itemschwierigkeiten des Primarstufentests erklärt werden, was als eine erste Validierung der Analysen anzusehen ist. Damit können die zwei aus unterschiedlichen Perspektiven identifizierten Merkmale als Basis zur Analyse der Schwierigkeit von Items dienen und einen Beitrag für die Weiterentwicklung von Tests zur schnellen Fehlererkennung leisten.

<table>
<tr><th colspan="2" rowspan="3">Merkmale / Items</th><th rowspan="3">empirische Schwierigkeit (Rasch-Wert -3 (leicht) bis +3 (schwer) aus Abb. 2 und Abb. 3)</th><th colspan="2">1. Merkmal
curriculare Relevanz</th><th colspan="3">2. Merkmal
strukturelle Komplexität</th></tr>
<tr><th rowspan="2">Einführung des Themas [in Klassenstufe]</th><th rowspan="2">Häufigkeit des Auftretens</th><th colspan="2">Zeichen</th><th rowspan="2">Regeln</th></tr>
<tr><th>je Antwort</th><th>je Item</th></tr>
<tr><td rowspan="5">Sekundarstufen Items</td><td>Trigonometrie (Item 7)</td><td>1</td><td>9/10</td><td>ca. 3 %</td><td>5 bis 6</td><td>16</td><td>9</td></tr>
<tr><td>Quadratische Gleichung (*p*-*q*-Formel) (Item 8)</td><td>1</td><td>9/10</td><td>ca. 6 %</td><td>9+
12 bis 15</td><td>48</td><td>8</td></tr>
<tr><td>Anwendung Satz des Pythagoras (Item 5)</td><td>1</td><td>9/10</td><td>ca. 4 %</td><td>14</td><td>29</td><td>7</td></tr>
<tr><td>Addition zweier Brüche (Item 15)</td><td>-1</td><td>5/6</td><td>ca. 40 %</td><td>10 bis 11</td><td>32</td><td>3</td></tr>
<tr><td>Umformen mit Klammern (Item 14)</td><td>-1,5</td><td>5</td><td>ca. 25 %</td><td>13 bis15</td><td>41</td><td>4</td></tr>
<tr><td rowspan="4">Primarstufen Items</td><td>Umfang eines Rechtecks (Item 12)</td><td>2</td><td>3/4</td><td>ca. 10 %</td><td>9 bis 13</td><td>34</td><td>3</td></tr>
<tr><td>Schriftliche Multiplikation (Item 1)</td><td>1</td><td>3/4</td><td>ca. 25 %</td><td>7</td><td>21</td><td>6</td></tr>
<tr><td>Einzeichnen der Symmetrieachse (Item 8)</td><td>-2</td><td>1/2</td><td>ca. 25 %</td><td>2</td><td>6</td><td>2</td></tr>
<tr><td>Fehler mit der Null (Item 2)</td><td>-2</td><td>1–3</td><td>ca. 60 %</td><td>5</td><td>15</td><td>2</td></tr>
</table>

Tab. 5: Übersicht der Merkmale zur Schwierigkeitsgenerierung bei beiden zeitbeschränkten Tests

Dabei sollten in Weiterentwicklung solcher Tests aufgrund ihrer Einbindung in Studien zur Professionalität von Mathematiklehrkräften stärker mathematikdidaktisch-orientierte schwierigkeitsgenerierende Merkmale in den Vordergrund gerückt werden. Aber die Analysen machen deutlich, dass auch wahrnehmungspsychologische Aspekte eine nicht unwichtige Rolle spielen und berücksichtigt werden müssen.

Als eine Grenze der Studie muss festgestellt werden, dass die Ergebnisse nur für die einfachen und die schwierigen Items stabil sind, bei mittlerer Itemschwierigkeit scheinen noch andere Faktoren eine Rolle zu spielen, die über diese Analysen hinausgehen. Weitere Untersuchungen mit einem größeren Itempool erscheinen daher nötig, um einerseits die Relevanz der zwei rekonstruierten schwierigkeitsgenerierenden Aufgabenmerkmale zu erhärten und andererseits weitere Einflussfaktoren zu identifizieren. Dass die Ergebnisse sowohl für Primarstufenitems als auch für Sekundarstufenitems stabil sind, kann allerdings als ein Hinweis auf die Robustheit der Ergebnisse angesehen werden.

Abschließend kann festgestellt werden, dass das schnelle Erkennen von Schülerfehlern eine grundlegende Kompetenz von Lehrkräften darstellt und daher bereits in der Lehrerausbildung gefördert werden sollte durch einschlägige Seminarsequenzen, wie von Heinrichs (2015) entwickelt. Als Konsequenz aus den Ergebnissen kann gefolgert werden, dass solche Seminarsequenzen zwar mit Fehlern aus Themengebieten, die im jeweiligen Curriculum der unteren Klassenstufen auftreten, ansetzen sollten, da diese leichter zugänglich sind und die Lehramtsstudierenden damit weniger überfordern. Allerdings sollten aufgrund der hohen Relevanz für die Entwicklung der Kompetenz zur Fehlererkennung auch die schwerer zu identifizierenden Schülerfehler, deren Themen erst in den höheren Klassenstufen im Curriculum eingeführt werden, in den Seminarsequenzen behandelt werden, um damit Lehramtsstudierende für alle Schulstufen ihres Lehramts adäquat auszubilden.

Insgesamt weisen die Analysen auf die Bedeutung von Fehlern beim Umgang mit den in den Bildungsstandards geforderten Kompetenzen zur Verwendung mathematischer Darstellungen und zum Umgang mit symbolischen, formalen und technischen Elementen der Mathematik hin (Leiß & Blum, 2007), mit denen nicht nur Schülerinnen und Schüler, sondern auch die sie unterrichtenden Lehrkräfte vertraut sein sollten.

Literatur

Allmendinger, H., Lengnink, K., Vohns, A. & Wickel, G. (2013). *Mathematik verständlich unterrichten. Perspektiven für Unterricht und Lehrerbildung*. Wiesbaden: Springer Fachmedien.

Beutelspacher, A. (2008). Über die Unmöglichkeit und die Notwendigkeit von Fehlern in der Mathematik. In R. Caspary (Hrsg.), *Nur wer Fehler macht, kommt weiter. Wege zu einer neuen Lernkultur* (S. 86–96). Freiburg: Herder.

Blömeke, S., Gustafsson, J.-E. & Shavelson, R. J. (2015). Beyond Dichotomies. Competence Viewed as a Continuum. *Zeitschrift für Psychologie, 223* (1), 3–13.

Blömeke, S., Kaiser, G. & Lehmann, R. (Hrsg.). (2010). *TEDS-M 2008. Professionelle Kompetenz und Lerngelegenheiten angehender Mathematiklehrkräfte für die Sekundarstufe I im internationalen Vergleich*. Münster: Waxmann.

Blömeke, S., König, J., Busse, A., Suhl, U., Benthien, J., Döhrmann, M. et al. (2014). Von der Lehrerausbildung in den Beruf – Fachbezogenes Wissen als Voraussetzung für Wahrnehmung, Interpretation und Handeln im Unterricht. *Zeitschrift für Erziehungswissenschaft, 17* (3), 509–542.

Bond, T. & Fox, C. (2007). *Applying the Rasch model: Fundamental measurement in the human science*s. Mahwah, NJ: LEA.

Brinkmann, A. (2002). *Über Vernetzungen im Mathematikunterricht - eine Untersuchung zu linearen Gleichungssystemen in der Sekundarstufe I.* Unveröffentlichte Dissertation. Duisburger Elektronische Texte.

Bruner, J. S. (1973). *Der Prozeß der Erziehung*. Berlin: Berlin Verlag.

Brunner, M. (2013). Didaktikrelevante Aspekte im Umfeld der Konzepte token und type. *Journal für Mathematik-Didaktik, 34* (1), 53–72.

Cohors-Fresenborg, E., Sjuts, J. & Sommer, N. (2004): Komplexität von Denkvorgängen und Formalisierung von Wissen. In M. Neubrand (Hrsg.), *Mathematische Kompetenzen von Schülerinnen und Schülern in Deutschland: Vertiefende Analysen im Rahmen von PISA-2000*, 109-144. Wiesbaden: VS-Verlag für Sozialwissenschaften.

Draschoff, S. (2000). *Lernen am Computer durch Konfliktinduzierung. Gestaltungsempfehlungen und Evaluationsstudie zum interaktiven computerunterstützten Lernen*. Münster: Waxmann.

Draxler, D. (2006). *Aufgabendesign und basismodellorientierter Physikunterricht*. Unveröffentlichte Dissertation Duisburg-Essen (https://d-nb.info/983890943/34).

Drollinger-Vetter, B. (2011). *Verstehenselemente und strukturelle Klarheit. Fachdidaktische Qualität der Anleitung von mathematischen Verstehensprozessen im Unterricht.* Münster: Waxmann.

Dyrvold, A. (2016). *Difficult to read or difficult to solve? The role of natural language and other semiotic resources in mathematics tasks*. Umeå: Print & Media.

Eichelmann, A., Narciss, S., Schnaubert, L. & Melis, E. (2012). Typische Fehler bei der Addition und Subtraktion von Brüchen – Ein Review zu empirischen Fehleranalysen. *Journal für Mathematik-Didaktik*, *33* (1), 29–57.

Franke, M. & Reinhold, S. (2016). *Didaktik der Geometrie in der Grundschule* (3. Aufl.). Berlin: Springer Spektrum.

Freie und Hansestadt Hamburg Behörde für Schule und Berufsbildung. (2012a). *Beispiel für ein schulinternes Fachcurriculum. Mathematik Grundschule*. Hamburg.

Freie und Hansestadt Hamburg Behörde für Schule und Berufsbildung. (2012b). *Beispiel für ein schulinternes Fachcurriculum. Mathematik. Stadtteilschule Sekundarstufe I*. Hamburg.

Gerster, H.-D. (1982). *Schülerfehler bei schriftlichen Rechenverfahren. Diagnose und Therapie*. Freiburg: Herder.

Goebbels, S. & Ritter, S. (2013). *Mathematik verstehen und anwenden – von den Grundlagen bis zu Fourier-Reihen und Laplace-Transformation* (2. Aufl). Berlin: Springer.

Hartig, J. (2007). Skalierung und Definition von Kompetenzniveaus. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch Englisch Schülerleistungen International)* (S. 83–99). Weinheim: Beltz Verlag.

Heinrichs, H. (2015). *Diagnostische Kompetenz von Mathematik-Lehramtsstudierenden. Messung und Förderung*. Wiesbaden: Springer Fachmedien.

Heinze, A. (2004). Zum Umgang mit Fehlern im Unterrichtsgespräch der Sekundarstufe I. *Journal für Mathematik-Didaktik, 25* (3-4), 221–244.

Hill, H.C., Ball, D.L. & Schilling, S.G. (2008) Unpacking Pedagogical Content Knowledge. *Journal for Research in Mathematics Education*, 39, 372-400.

Kaiser, G., Blömeke, S., König, J., Busse, A., Döhrmann, M. & Hoth, J. (2017). Professional competencies of (prospective) mathematics teachers—cognitive versus situated approaches. *Educational Studies in Mathematics, 94* (2), 161–182.

Kaiser, G., Busse, A., Hoth, J., König, J., & Blömeke, S. (2015). About the Complexities of Video-Based Assessments. Theoretical and Methodological Approaches to Overcoming Shortcomings of Research on Teachers' Competence. *International Journal of Science and Mathematics Education, 13* (2), 369–387.

Kauertz, A. (2008). *Schwierigkeitserzeugende Merkmale physikalischer Leistungstestaufgaben*. Berlin: Logos Verlag.

Kelle, U. & Kluge, S. (2010). *Vom Einzelfall zum Typus*. Wiesbaden: VS Verlag.

Klug, J., Bruder, S., Kelava, A., Spiel, C. & Schmitz, B. (2013). Diagnostic competence of teachers. A process model that accounts for diagnosing learning behavior tested by means of a case scenario. *Teaching and Teacher Education, 30,* 38–46.

König, J., Blömeke, S., Klein, P., Suhl, U., Busse, A. & Kaiser, G. (2014). Is teachers' general pedagogical knowledge a premise for noticing and interpreting classroom situations? A video-based assessment approach. *Teaching and Teacher Education*, *38*, 76–88.

Krauss, S. & Brunner, M. (2011). Schnelles Beurteilen von Schülerantworten: Ein Reaktionszeittest für Mathematiklehrer/innen. *Journal für Mathematik-Didaktik, 32*(2), 233–251.

Krauss, S., Neubrand, M., Blum, W., Baumert, J., Brunner, M., Kunter, M. et al. (2008). Die Untersuchung des professionellen Wissens deutscher Mathematik-Lehrerinnen und -Lehrer im Rahmen der COACTIV-Studie. *Journal für Mathematik-Didaktik, 29* (3-4), 233–258.

Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S. & Neubrand, M. (2011). *Professionelle Kompetenz von Lehrkräften. Ergebnisse der Forschungsprogramms COACTIV*. Münster: Waxmann.

Landis, J. R. & Koch G. G. (1977). The measurement of observer agreement for categorical data. In: *Biometrics*. 33, 1977, 159–174

Leiß, D., & Blum, W. (2007). Beschreibung zentraler mathematischer Kompetenzen. In W. Blum, C. Drüke-Noe, R. Hartung & O. Köller. Bildungsstandards Mathematik: konkret. Sekundarstufe I: Aufgabenbeispiele, Unterrichtsanregungen, Fortbildungsideen (3. Aufl.) (S. 33-50). Berlin: Cornelsen Scriptor.

Lüddecke, J. (2015). *Fehler beim Problemlösen. Empirische Erkundungen zu Fehlern beim Bearbeiten mathematischer Probleme*. Hamburg: disserta Verlag.

Magenheim, J., Schubert, S. & Schaper, N. (2015). Competencies in Computer Science Education. In T. Brinda, N. Reynolds, R. Romeike & A. Schwill (Hrsg.), *KEYCIT 2014: key competencies in informatics and ICT* (S. 33–56). Universitäts-Verlag Potsdam.

Malle, G. (1993). Didaktische Probleme der elementaren Algebra. Braunschweig: Vieweg.

Oser, F., Hascher, T. & Spychiger, M. (1999). Lernen aus Fehlern. Zur Psychologie des "negativen" Wissens. In W. Althof (Hrsg.), *Fehlerwelten. Vom Fehlermachen und Lernen aus Fehlern* (S. 11–41). Wiesbaden: Verlag für Sozialwissenschaften.

Padberg, F. & Benz, C. (2011). *Didaktik der Arithmetik. Für Lehrerausbildung und Lehrerfortbildung.* Heidelberg: Spektrum.

Padberg, F. (2009). *Didaktik der Bruchrechnung. Für Lehrerausbildung und Lehrerfortbildung* (4., erw., Aufl.). Heidelberg: Springer Spektrum.

Pankow, L., Kaiser, G., Busse, A., König, J., Blömeke, S., Hoth, J. et al. (2016). Early career teachers' ability to focus on typical students errors in relation to the complexity of a mathematical topic. *ZDM Mathematics Education, 48* (1-2), 55–67.

Radatz, H. (1980). *Fehleranalysen im Mathematikunterricht*. Wiesbaden: Vieweg+Teubner Verlag.

Ritter, S. & Voß, U. (2015). *Erfolgreich Starten ins Ingenieurstudium. Grundlagen der Mathematik anwendungsorientiert erklärt*. Berlin: Springer Vieweg.

Robinson, J. P., Shaver, P. R. & Wrightsman, L. S. (1991). Criteria for Scale Selection and Evaluation. In J. P. Robinson, P. R. Shaver & L. S. Wrightsman (Hrsg.), *Measures of Personality and Social Psychological Attitudes* (S. 1-15). San Diego: Academic Press.

Rölke, H. (2012). The ItemBuilder: A Graphical Authoring System for Complex Item Development. In T. Bastiaens & G. Marks (Hrsg.), *Proceedings of E-Learn. World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2012* (S. 344–353). Chesapeake: Aace. (https://www.learntechlib.org/p/41614; letzter Zugriff am 15.02.2017).

Schoy-Lutz, M. (2005). *Fehlerkultur im Mathematikunterricht. Theoretische Grundlegung und evaluierte unterrichtspraktische Erprobung anhand der Unterrichtseinheit "Einführung in die Satzgruppe des Pythagoras"*. Hildesheim: Franzbecker.

Schumann, S. & Eberle, F. (2011). Bedeutung und Verwendung schwierigkeitsbestimmender Aufgabenmerkmale für die Erfassung ökonomischer und beruflicher Kompetenzen. In U. Faßhauer (Hrsg.), *Grundlagenforschung zum Dualen System und Kompetenzentwicklung in der Lehrerbildung* (S. 77–89). Leverkusen: Barbara Budrich.

Seidel, T. & Prenzel, M. (2003). Mit Fehlern umgehen - Zum Lernen motivieren. *Praxis der Naturwissenschaften - Physik, 51* (1), 30–34.

Seifried, J. & Wuttke, E. (2010). Professionelle Fehlerkompetenz – Operationalisierung einer vernachlässigten Kompetenzfacette von (angehenden) Lehrkräften. *Wirtschaftspsychologie, 12* (4), 17–28.

Shulman, L. S. (1986). Those Who Understand. Knowledge Growth in Teaching. *Educational Researcher, 15* (2), 4–14.

Südkamp, A. & Praetorius, A.-K. (Hrsg.) (2017). *Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen*. Münster: Waxmann.

Wartha, S. (2007). *Längsschnittliche Untersuchungen zur Entwicklung des Bruchzahlbegriffs*. Hildesheim: Franzbecker.

Wu, M. L., Adams, R. J. & Wilson, M. (1998). *ACER ConQuest. Generalised item reponse modelling software manual*. Melbourne: ACER Press.

Anschrift der Verfasserinnen

Lena Pankow
Universität Hamburg,
Fakultät für Erziehungswissenschaft
Von-Melle-Park 8
20146 Hamburg, Deutschland
lena.pankow@uni-hamburg.de
ORCID ID: 0000-0001-8941-7133

Prof. Dr. Gabriele Kaiser
Universität Hamburg
Fakultät für Erziehungswissenschaft
Von-Melle-Park 8
20146 Hamburg, Deutschland
gabriele.kaiser@uni-hamburg.de
Australian Catholic University
Institute for Learning Sciences and Teacher Education
ORCID ID: 0000-0002-6239-0169

Anhang

Kodierhinweise (Auszug aus dem Manual):

Rating des Merkmals: *curricularen Relevanz*

In einer Kreuztabelle wurden alle Themengebiete, die in dem Beispielcurriculum der Freien und Hansestadt Hamburg enthalten sind, auf der x-Achse aufgeführt. Die Items, aus denen der Test besteht, sind auf der y-Achse aufgeführt. In der so entstandenen Kreuztabelle wurde je Item geratet, ob das Thema des typischen Schülerfehlers in dem Themengebiet des Beispielcurriculums vorkommt (1) oder nicht (0).

Hinweis: Ein Thema des typischen Schülerfehlers kann erst dann mit 1 (enthalten) aufgeführt werden, wenn es zuvor eingeführt wurde. Bsp. Die Bruchrechnung wird erst in der 5/6 Klasse eingeführt, könnte allerdings schon in der Grundschule bei der Berechnung von Wahrscheinlichkeiten ihre Anwendung finden. Dies galt es beim Rating zu beachten.

Rating des Merkmals *strukturelle Komplexität*

Die *Anzahl der Zeichen*, aus denen die Items bestehen, soll bestimmt werden. Dafür gelten nach Literaturgrundlage die folgenden Regeln:

- Ein Bruch wird als 3 Zeichen gezählt, da zunächst erfasst werden muss, dass es sich um einen Bruch handelt. Des Weiteren muss der Wert des Bruchs bestimmt werden, hier müssen Zähler und Nenner in Beziehung gesetzt werden.
- Eine Zahl wie bspw. 30 wird hingegen als ein Zeichen gezählt, da die 3 und die 0 im Stellenwertsystem miteinander in Verbindung stehend gesehen werden muss.
- Zeichen, die eine Operation beschreiben, wie „+, –, ·, %, √ oder auch ±, =, sin “, werden als ein Zeichen betrachtet.
- Enthaltene Einheiten wie „cm, ° oder m“ werden als ein Zeichen gezählt.
- Bei einigen Items ist über den Antwortmöglichkeiten eine Grafik zu sehen, die darin enthaltenen Zeichen müssen ebenfalls gezählt werden. Die Zeichen müssen allerdings nicht zu jeder Antwortmöglichkeit einzeln addiert werden, sondern nur einmal gesondert vermerkt werden.

In den Items enthaltene Regeln:

Die verwendeten Regeln, die zu einer Lösung als bekannt vorausgesetzt und korrekt angewandt werden müssen, sollen ebenfalls gezählt **und** aufgeführt werden.

Während bei einer einfachen Addition von zwei einstelligen Ziffern unter 5 nur eine Regel verwendet werden muss, ist es notwendig, bei einer Addition bei Ziffern über 5 neben der zuvor genannten Regel ebenfalls die des Stellenwerts einzuhalten. Wichtig beim Raten der Items ist das Aufführen der verwendeten Regeln, hier gilt es möglichst elementar vorzugehen. Man beginne mit den Regeln des Stellenwerts und „untersuche“ die einzelnen Operationen, die weiterhin notwendig sind.

6. Publikation III

Der dritte Artikel im Rahmen dieser kumulativen Dissertation mit dem Titel „Perception of student errors under time limitation: are teachers faster than mathematicians or students?" ist ebenfalls 2018 in der Zeitschrift *ZDM Mathematics Education* erschienen. Das Heft ist dem Thema „Assessment in mathematics education: issues regarding methodology, policy and equity" gewidmet und wurde von Guri A. Nortvedt und Nils F. Buchholtz herausgegeben.

Diese Publikation widmete sich dem Forschungsziel, den Test zur schnellen Fehlererkennung mittels Kontrastgruppen zu validieren. Der theoretische Rahmen des Beitrags bietet einen Forschungsüberblick bezüglich des Umgangs mit Schülerfehlern, zu deren Wahrnehmung und insbesondere zu empirischen Studien hinsichtlich der schnellen Wahrnehmung von Schülerfehlern. Dem schließt sich eine Darstellung der Ansätze zur Validierung an. Insbesondere wird auf die Technik der bekannten Gruppen hingewiesen, bei der mit sogenannten Kontrastgruppen gearbeitet wird. Außerdem werden externe Variablen als Indikatoren für die Validierung herangezogen.

In der Folge werden vier Hypothesen vorgestellt, in denen anhand der Kontrastgruppen ein Ranking des Abschneidens in dem zeitbegrenzten Test zur schnellen Fehlererkennung abgeleitet wird. Das Studiendesign mit dem Testinstrument und die Stichproben werden detailliert vorgestellt. Die Methodologie der angewandten Testung sowie das Vorgehen bei der Berechnung der Zusammenhänge zwischen den Stichproben werden ebenso erläutert wie die Messinvarianz als Voraussetzung für den adäquaten Stichprobenvergleich.

Es kann festgestellt werden, dass die Stichproben miteinander verglichen werden dürfen, da dasselbe Konstrukt gemessen wird. Weiterhin konnte ein Ranking der Stichproben in folgender Reihenfolge ermittelt werden: die grundständig Mathematikstudierenden, die zukünftigen Mathematiklehrkräfte, die Teilnehmer(innen) von TEDS-FU, die Teilnehmer(innen) von TEDS-Unterricht und die Oberstufenschüler(innen). Weiterhin ist festzustellen, dass sich die Ergebnisse im Test zur schnellen Fehlererkennung in der Gruppe der Lehrkräfte nicht signifikant unterscheiden.

Der Beitrag schließt mit der kritischen Erläuterung der Grenzen des Testinstruments, die insbesondere darin liegen, dass mittels des Instruments nur ein Element eines Diagnosekreislaufs analysiert wird, während die anderen Elemente nicht betrachtet werden. Weiterhin wird aufgezeigt, dass einerseits immer noch weitere Informationen erhoben werden könnten, was andererseits nicht immer umsetzbar sein wird. Das Fazit schließt mit der Erkenntnis, dass das schnelle Erkennen von Schülerfehlern eine Kernkompetenz von Lehrkräften darstellt, die neben dem mathematikdidaktischen Wissen auch das mathematische Wissen für deren Wahrnehmung voraussetzt.

6.1. Darlegung des eigenen Anteils

Die dritte Publikation ist ebenfalls in enger Kooperation mit der Betreuerin meiner Promotion Gabriele Kaiser entstanden, wobei die Idee des Vorgehens von mir entwickelt und die Analysen von mir durchgeführt wurden.

Die Darstellung der erzielten Ergebnisse und deren Diskussion erfolgte zunächst von mir und wurden von der Autorengruppe gemeinschaftlich überarbeitet. Im fortlaufenden Begutachtungsprozess wurde die vorliegende Publikation von mir unter Mitarbeit der Projektleitung von TEDS-FU (Gabriele Kaiser, Johannes König und Sigrid Blömeke) überarbeitet.

6.2. Abdruck der Publikation III

Pankow, L., Kaiser, G., König, J. & Blömeke, S. (2018). Perception of student errors under time limitation: are teachers faster than mathematicians or students? *ZDM Mathematics Education, 50*(4), 631–642.

ZDM (2018) 50:631–642
https://doi.org/10.1007/s11858-018-0945-1

ORIGINAL ARTICLE

Perception of student errors under time limitation: are teachers faster than mathematicians or students?

Lena Pankow[1] · Gabriele Kaiser[1,2] · Johannes König[3] · Sigrid Blömeke[4]

Accepted: 12 May 2018 / Published online: 19 May 2018

Abstract
The ability to offer constructive feedback to students concerning their errors is an indispensable requirement for mathematics teachers, for the purpose of providing cognitively challenging learning opportunities. However, if they are to react adequately, teachers need to identify student errors immediately. The fast perception of student errors can therefore be described as an indispensable part of mathematics teachers' professional competence. Data on this facet of teacher competence were gathered as part of a national follow-up-study of the IEA's international TEDS-M (Teacher Education and Development Study in Mathematics) that used a time-limited test to measure teachers' perception of student errors. This paper aims to provide evidence for the validity of the test interpretation of fast student-error perception as an indicator of professional competence by drawing on contrast groups already used in other studies. Overall, the study could support the validity of the test interpretation because the chosen contrast groups were found to perform either better than the tested teachers—as is the case for the contrast group of mathematicians—or more poorly, as is the case for the group of students. Furthermore, the present study shows that the competence facet of fast student error perception is closer to the domain of teachers' mathematical content knowledge than it is to the domain of teachers' mathematics pedagogical content knowledge.

Keywords Validation study · Contrast groups · Timed test · Knowledge facets · Student error recognition

1 Introduction

Perceiving and adequately responding to student errors is part of mathematics teachers' daily activities in the classroom. Teachers not only have to be aware of student errors; they must also identify, interpret and classify them. Further, they must decide if the error is relevant enough to be addressed extensively in front of the whole group or individually, and whether the error is caused by an underlying misconception. The time they have for noticing and addressing student errors is limited, but doing so, according to a study by Schoy-Lutz (2005), creates genuine learning opportunities in most cases. A necessary condition for recognizing these learning opportunities is the fast identification of student errors.

Knowledge of students' misconceptions is an important facet of teacher competence in mathematics (cf. Altmann and Nückles 2017). This competence was assessed in TEDS-FU, a German follow-up study to the international Teacher Education and Development Study in Mathematics (TEDS-M; for details see Blömeke et al. 2014). TEDS-FU evaluated the competence of a subsample of German TEDS-M participants who were practicing mathematics teachers in their fourth year of teaching. The results presented in this paper are related to lower secondary teachers. In addition to the original TEDS-M study, TEDS-FU contained a test component on fast perception of student errors by teachers (see Pankow et al. 2016). The fast perception of students' mathematical errors can be described as a sub-facet of *diagnostic competence* (Shulman 1986).

To validate the interpretation of the test as an indicator of mathematics teachers' competence and to further examine whether the instrument is testing the targeted construct, the present study includes a contrasting group analysis based on five groups with varying expertise:

✉ Lena Pankow
Lena.Pankow@uni-hamburg.de

[1] University of Hamburg, Hamburg, Germany

[2] Australian Catholic University, Brisbane, Australia

[3] University of Cologne, Cologne, Germany

[4] Centre for Educational Measurement at University of Oslo, Oslo, Norway

future mathematics teachers in their master's study (Sample 1), early career teachers (about 3 to 4 years of teaching experience) (Sample 2), experienced teachers (> 4 years of teaching experience) (Sample 3), university students enrolled in a mathematics degree program (Sample 4), and high school students (Sample 5). This approach was adopted from the COACTIV (Professional Competence of Teachers, Cognitively Activating Instruction, and Development of Students' Mathematical Literacy) study (see Krauss et al. 2008), in which the tests on mathematical content knowledge and mathematics pedagogical content knowledge—including a test of fast perception of student errors—were validated using contrast groups. Like the ones in the COACTIV study, the groups in this study were chosen according to their different levels and qualities of expertise, i.e., mathematical content knowledge and mathematics pedagogical content knowledge, Therefore, this study aims to examine the relations between a key dependent variable—fast perception of student errors—and a set of external variables, an approach that can be considered as part of criterion-based *validation*, at least when accounting for a broader understanding of the classical criterion validity concept. Furthermore, the relation between the content of the test and the construct it is intended to measure is analyzed, providing content-oriented evidence that, in a classical sense, could be labelled as *construct validity* (AERA, APA, and NCME 2014).

After a description of the theoretical background of the study, which includes an overview of the current discussion on student errors in mathematics education, the construct of fast perception of student errors under time pressure is developed. Furthermore, the validation process using the contrast group method is described along with other sources of evidence for validity. Finally, the roles and measurement of both mathematical content knowledge and mathematics pedagogical content knowledge within the TEDS-FU study are discussed. After developing the research aim and the hypotheses, the design of the validation study is presented, followed by a description of the results. The paper closes with a summary and discussion of those results.

2 Theoretical framework of the study

This section describes the theoretical framework of student errors and teachers' familiarity with those errors as part of teachers' diagnostic competence. That discussion is followed by a description of the empirical studies that operationalize and measure student-error perception. The chapter closes with a description of different approaches to the validation of the test instrument.

2.1 Dealing with student errors

Oser et al. (1999) defined an error as a "deviation from the standard" and continued, "Standards represent the reference system; without standards or rules, it would not be possible to distinguish between correct and incorrect" (p. 11). Heinze (2004) specified this general definition of an error tailored to mathematics education: "An error is an expression which is against the general statements and definitions of mathematics as well as against a generally accepted mathematical-methodical approach" (p. 223). Student errors in mathematics can happen for a variety of reasons, including carelessness, ignoring the given rules, or feeling uncertain. However, many student errors are of a systematic nature—i.e., they will occur again in another mathematical task of equivalent structure. If these types of errors are not tackled adequately by a teacher's intervention (Swan 2004), they will probably be repeated over subsequent school years (Radatz 1980a, b, p. 16; Türling 2014). The ability to perceive student errors quickly represents an important part of professional competence, then, because immediate feedback from a teacher supports student learning processes and enhances active learning. In addition, the perception of student errors is needed in order to evaluate student work and plan the next instructional steps (Brühwiler 2017).

2.2 Student-error perception

The conceptualization of teacher competence and its measurement has been the focus of much empirical research in the last two decades. However, the inclusion of student-error perception as part of that measurement, and the ways of dealing with that component, have been introduced into that discussion only recently, especially in the context of adaptive teacher behavior (Südkamp and Praetorius 2017).

Departing from the general discussion of diagnostic competencies as part of teacher competence and how teacher competence can be fostered holistically as part of teacher education, Heinrichs and Kaiser (2018) developed a model for the perception and handling of errors in mathematics instruction. They described the diagnostic process in instructional situations where errors occur, by way of a cyclic process starting from the perception of a student error, leading to the development of hypotheses concerning the causes of such errors, and ending with possible instructional approaches for dealing with the error in class (see also Hoth et al. 2016).

If the student has not understood the error and therefore overcome the problem, the cyclic process has to be carried out another time (see Fig. 1). The first step

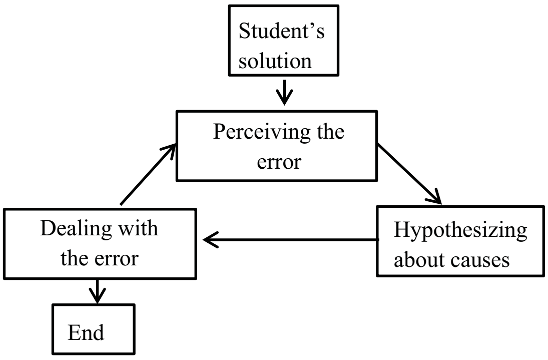


Fig. 1 Model of the diagnostic process in error situations (Heinrichs and Kaiser 2018, p. 84)

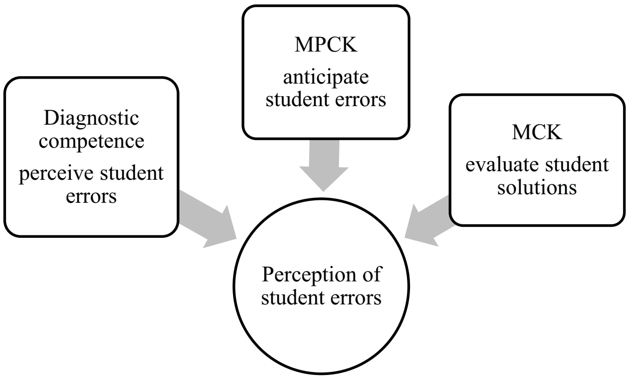


Fig. 2 Contribution of different theoretical constructs to the perception of students' errors

of the diagnostic process, the perception of the error, was the focus of the time-limited test presented in this paper. Neither of the other steps of the diagnostic process—'hypothesizing about causes' and 'dealing with the error'—were part of the test developed in TEDS-FU.

Based on the current discussion of teacher competence, Blömeke et al. (2015) described competence as a continuum, departing from teachers' knowledge as their dispositions, including situational components such as perception, interpretation and decision-making, leading to the observable behavior of teachers in class (p. 7). As error perception may be considered to be embedded within the perception component seen here as part of the situational competences of teachers, it can, in a narrower sense, also be seen as a component of diagnostic competence. Diagnostic competence is connected to mathematical content knowledge and mathematics pedagogical content knowledge, as this kind of knowledge is needed for the perception and interpretation of student errors as well as for the development of adequate teaching measures (for an overview of the current discussion on diagnostic competence see Südkamp and Praetorius 2017, Leuders et al. 2018). A crucial element of student-error perception is the speed of error perception under time constraints, as teachers have to analyze students' statements and responses within a limited time frame (Lindmeier et al. 2013, p. 106); dealing with those time constraints is part of teachers' professional lives (Wahl et al. 1984).

The knowledge dimensions of teachers' professional competence are conceptualized in TEDS-FU based on the work of Shulman (1986), who distinguished, amongst others, content knowledge, pedagogical content knowledge and pedagogical curricular knowledge. In the context of the test of fast perception of student errors, two of Shulman's three facets are relevant, namely *content knowledge* as the part of knowledge that encloses the content and is not reduced, or pure factual knowledge. As this facet relates to mathematics, it is called *Mathematical Content Knowledge* (MCK). Furthermore, Shulman (1986) hones in on specific pedagogical knowledge, the Pedagogical Content Knowledge (PCK), which includes the following:

> .. . for the most regularly taught topics in one's subject area, the most useful forms of representation of those ideas, the most powerful analogies, illustrations, examples, explanations, and demonstrations—in a word, the ways of representing and formulating the subject that make it comprehensible to others. (Shulman 1986, p. 9).

This form of knowledge translates to mathematics as Mathematical Pedagogical Content Knowledge (MPCK). Additionally, Shulman points out that the teacher who is confronted with student errors has to have strategies to correct misconceptions in order to reorganize students' knowledge (Shulman 1986, pp. 9).

Overall, the perception of student errors has different theoretical starting points as described in Fig. 2. The initial step in diagnosing a student error is the perception of a deviation from the expected norm. Reisman (1976) describes this phase as *identification* as it refers to noticing and analyzing student behavior. To use errors and the misconceptions connected to them productively in the learning process, it is indispensable to identify them quickly (cf. Leuders 2001; Radatz 1980a, b).

2.3 Empirical studies on fast student-error perception

Teacher competence with regard to perceiving student errors—or, more generally, to evaluating the correctness of students' answers—has already been measured in various empirical projects and studies.

Within the framework of the Mathematical Knowledge for Teaching (MKT) and Mathematics Teaching and Learning to Teach (MTLT) projects, Hill et al. (2008) described the identification of common student errors and the provision of explanations as one of the major parts of teachers' mathematical knowledge for teaching. Ball et al. (2008) summarized the perception of errors under the knowledge facets *Specialized Content Knowledge* (SCK) and *Knowledge of Content and Students* (KCS), pointing out that error perception has two bases—mathematical knowledge and the knowledge of students and their relation to school mathematics. The latter kind of knowledge was described as the province of teachers, in contrast to mathematical knowledge, which resided in the province of professional mathematicians. Professional mathematicians were expected to be less familiar with common student errors. Further analyses could confirm that mathematicians were less likely to answer items correctly based on knowledge of students than content knowledge items (Hill et al. 2007). Ball et al. (2008) emphasized in their model that error analyses under time restriction are part of the characteristic tasks of teachers and clearly delineate their work differently from other professions, especially from the work of mathematicians.

The COACTIV study analyzed teachers' evaluations of students' solutions to mathematical tasks as either "correct" or "incorrect" (Binder et al. 2018) by measuring the teachers' reaction time and setting these results in relation to the number of student responses they assessed correctly. This construct is described as the *fast evaluation of elementary mathematical statements*. In contrast to the MKT project previously discussed, which focused on primary teachers, COACTIV evaluated the professional knowledge of secondary teachers. As COACTIV had in addition measured mathematical content knowledge and mathematics pedagogical content knowledge, the authors were able to estimate and examine the relations between the results from the test of reaction times and subjects' scores on these knowledge facets. The study showed a strong correlation between the fast evaluation of mathematical statements and mathematical content knowledge by the teachers as well as their mathematics pedagogical content knowledge (Krauss and Brunner 2011).

Like the MKT project, COACTIV analyzed the validity of its instruments and found—in contrast to the MKT project—that mathematicians solved the items significantly better, which supports the strong relation of the perception of student errors to mathematical content knowledge (Krauss et al. 2008).

The test of student-error perception that is presented in this paper, carried out as a separate test component within the TEDS-FU study, limited the time it took study participants to perceive student errors or assess the quality of students' statements. To bring the test closer to actual teaching environments, common student errors—ones that are often encountered in everyday teaching—were used as test items. Furthermore, the test was time limited in such a way that it would make any recalculation by the participants impossible (Bühner 2011, p. 21). The number of correct answers was used as a performance indicator, and items not answered within the given time period were classified as wrong. The items used in the test for fast perception of student errors are of varying complexity (Pankow et al. 2016).

The present study examines the relations among mathematical content knowledge, mathematics pedagogical content knowledge, and fast perception of student errors using data from the TEDS-FU study. TEDS-FU measured the achievement of a subsample of early career (fourth-year) lower-secondary mathematics teachers in Germany who had participated in the TEDS-M study at the end of their teacher education. The data reveal a significant correlation of medium effect size between MCK and the fast perception of student errors ($r = .46$, $p < .001$; Blömeke et al. 2014, p. 528) and a significant correlation of weak effect size between MPCK and the error perception ($r = .28$).

Further studies of TEDS-FU data confirm the strong relation of mathematics pedagogical content knowledge and mathematical content knowledge with fast perception of student errors (Blömeke et al. 2016).

The dependence of error-perception speed on age has not yet been discussed in these studies, although many psychological studies point out that increasing age may bring with it a decrease in the speed with which many processing operations are executed, and that this reduction in speed may lead to disturbances in cognitive functions (Salthouse 1996). The decrease in sensory abilities in connection with the decrease in the ability for selective attention and multi-tasking are especially well-documented, particularly in research on aging and driving (Madden 2007; Schlag 1993). This age dependency may play a role in the measurement of fast perception of student errors in TEDS-FU, where the speed of visual perception and physical reaction were emphasized.

Although the construct of fast perception of student errors was already described (Pankow et al. 2016), until now no external validation study concerning the test and its interpretation has been provided. Currently, only analyses of the internal structure of the whole set of tests applied within TEDS-FU exist, in which the relationships between the test components and its conformity to the construct are evaluated (Blömeke et al. 2016). In the following, we describe the results of external validation studies.

2.4 Approaches for validation

In educational research it is necessary to provide evidence for the quality of the instruments used in studies. Validity as a fundamental quality criterion "refers to the degree to which evidence and theory support the interpretation of tests scores for proposed uses of tests. Validity is, therefore, the most fundamental consideration in developing tests and evaluating tests" (AERA, APA, and NCME 2014, p. 11). In comparison to other quality criteria (objectivity and reliability) the examination of validity is more complex because standardized, routine procedures for supporting validity do not exist.

Classically, four main types of validity are distinguished: content or internal validity, criterion or external validity, statistical validity, and construct validity (Bortz and Döring 2006). Internationally, the concept of validity has changed over the years as the field of educational and psychological measurement has developed, and so far, there is no general agreement with regard to what exactly validity is and how it should be measured (e.g., Kane 2001; Lissitz and Samuelsen 2007; for a recent overview see; Newton and Shaw 2016). One common view within contemporary discussions of validity is that in addition to the traditional sources of validity, such as evidence based on test content and expert opinions (traditionally called *content validity*) and evidence based on relations to other variables (an extension of the traditional *criterion validation*), response processes and consequences are also to be accepted as important sources of validity evidence that should be included in the validation process. Support for this view appears in the most recent edition of the *Standards for Educational and Psychological Testing* jointly published by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education (2014). Overall, these standards emphasize that "these sources of evidence may illuminate different aspects of validity, but they do not represent distinct types of validity. Validity is a unitary concept. It is the degree to which all the accumulated evidence supports the intended interpretation of test scores for the proposed use" (pp. 13–14). Furthermore, the published standards emphasize the need to integrate various sources of validity evidence: "A sound validity argument integrates various strands of evidence into a coherent account of the degree to which existing evidence and theory support the intended interpretation of test scores for specific uses" (AERA, APA and NCME 2014, p. 21).

Following these standards, we examine evidence for test content as the most important source of validity and evidence for the relations to external variables. Content validity can be obtained, according to the standards,

> ... from an analysis of the relationship between the content of a test and the construct it is intended to measure.... The content specification carefully describes the content in detail, often with a classification of areas of content and types of items. Evidence based on test content can include logical or empirical analyses of the adequacy with which the test content represents the content domain and of the relevance of the content domain to the proposed interpretation of test scores. Evidence based on content can also come from expert judgements of the relationship between parts of the test and the construct. (AERA, APA and NCME 2014, p. 14).

Furthermore, the standards mention that evidence about content "can be used, in part, to address questions about differences in the meaning or the interpretation of test scores across relevant sub-groups of test takers" (AERA, APA and NCME 2014, p. 15). Analyses are therefore needed that explore the extent to which *construct underrepresentation*—i.e., the "degree to which a test fails to capture important aspects of the construct" (p. 12)—or *construct-irrelevance*—i.e., "the degree to which test scores are affected by processes that are extraneous to the test's intended purpose" (AERA, APA and NCME 2014, p. 12)—may give an advantage to subgroups of test takers. This new kind of evidence for validity covers the traditionally distinguished *content validity* and *construct validity*.

A particular variation for providing evidence for construct validity is the *technique of related groups* (Bortz and Döring 2006, p. 201). The selection characteristic used in this technique is the affiliation to composed groups, upon which the same test is carried out in order to measure different specifications of the construct. In other words, by deliberately constructing underrepresentation, it would become possible to study whether specific groups of test takers have an advantage and perform better or have a disadvantage and therefore perform more poorly. In this study the related groups are selected by their empirically assumed expertise so that it will be relatively easy to immediately identify errors in mathematics. The test results should differ between participants who are able to perceive student errors under time limitation and those who are not (yet) able to do so. The group with a high qualification in MCK and MPCK should perform better in the test than those without any expertise in the area of MCK and MPCK. This particular technique has already been used in various studies in mathematics education, which are described briefly as follows.

The COACTIV study differentiated empirically between content knowledge (CK) and pedagogical content knowledge (PCK) of mathematics teachers in Germany. To test the validity of the CK and PCK measurements in that study, researchers used contrast groups to compare the results of COACTIV subjects with those of university students majoring in mathematics, biology and chemistry teachers, senior high school students, and mathematics teacher candidates at the end of their university education (Krauss et al. 2008). The mathematical content knowledge of all groups measured was at the levels hypothesized with mathematics majors outperforming all other groups and high school students as well as biology and chemistry teachers performing the worst. In contrast, the data did not provide sufficient evidence for PCK as a distinct knowledge category of (future) mathematics teachers because the mathematics majors did well also on this test component (Krauss et al. 2008). In the empirical approach developed by Ball et al. (2008), contrast groups were used to provide evidence for the construct validity of a test of the professional knowledge of primary teachers, non-teachers of similar professions (e.g., nurses), and professional mathematicians (Hill et al. 2007). Participants were studied in cognitively oriented interviews to capture subject matter knowledge (SCK) and knowledge of content and students (KCS). In the interviews, 40% of the participants across all groups solved the items referring to KCS by using mathematical argumentation, which revealed that the student-focused material in those items was dominated by the underlying mathematical content; as a result, the items could not be considered as good measures of KCS (Ball et al. 2008).

Charalambous (2016) evaluated the knowledge of preservice elementary teachers, inservice elementary teachers and university students with strong mathematical background. The three groups all performed well on the pure mathematics content knowledge test while the results about mathematics pedagogical content knowledge were much in line with those of Krauss et al. (2008). Contrary to expectation, no significant differences could be identified between the two teacher groups and the university students in mathematically intensive departments. According to these results, MKT is measured too strongly in a holistic way and does not adequately reflect group differences. Charalambous (2016) concluded that more studies are needed to determine how teachers apply their specific professional knowledge and which role teachers' experience plays in developing teacher competence.

In addition, relations to other variables external to the test are examined. "External variables may include measures of some criteria that the test is expected to predict, as well as relationships to other tests hypothesized to measure the same constructs, and tests measuring related or different constructs" (AERA, APA and NCME 2014, p. 16). Relations between test scores that aim to evaluate the same or similar constructs provide convergent evidence, whereas relations between test scores intending to measure different constructs deliberately offer discriminant evidence. A special case of relations are *test-criterion relationships*, which measure "how accurately the test scores predict criterion performance" (AERA, APA and NCME 2014, p. 17) and are called in classical terminology *criterion validity*.

In this study, evidence for criterion validity is provided using the external criterion of the GPA of the university entrance qualification (so-called Abitur). This criterion has already been used in various studies in mathematics education, also in our own research in the TEDS-FU context. Blömeke et al. (2014) provided evidence that the university entrance qualification predicted fast perception of student errors mediated by the lower-secondary teachers' knowledge of mathematics content and mathematics pedagogical content.

3 Research aim and hypotheses

Using techniques based on the approaches described above, this study aims to provide evidence for the construct validity of the interpretations of the instrument used to measure fast perception of student errors. Specifically, the validity assessment will be based on differences in the performance of contrast groups of test takers.

To achieve this goal, five contrast groups with varying expertise in mathematical knowledge, mathematics pedagogical content knowledge, teaching experience, and reaction time (based on age) were selected: future mathematics teachers in their master's study (Sample 1: future teachers), teachers with four years of experience (Sample 2: early career teachers), teachers with more than four years of teaching experience (Sample 3: experienced teachers), university students studying mathematics as their degree program (Sample 4: university mathematics majors), and high school students (Sample 5: high school students).

This study investigated the following hypotheses:

Hypothesis 1 Among all five contrast groups, the group of early career teachers will perform best in the test of fast perception of student errors. This hypothesis is based on three assumptions: early career teachers' experience in the classroom will enable them to perceive student errors more quickly than groups with less (future teachers) or without any teaching experience (mathematics majors and high school students), while the mathematical content knowledge members of this group gained during their teacher training is more recent—and therefore easier to recall—than that of experienced teachers, and their younger age allows them to react more quickly than this group.

Hypothesis 2 Among all five contrast groups, the group of high school students will perform worst in the test of fast perception of student errors due to lack of any teaching experience with these, and they will rank lowest in terms of both mathematical content knowledge and mathematics pedagogical content knowledge, as their training in both facets is the least developed of all participating groups.

Hypothesis 3 The three groups of future teachers, experienced teachers and mathematics majors will perform between the groups of early career teachers and high school students regarding fast perception of student errors.

Hypothesis 4 The results of the test of fast perception of student error can be predicted using students' GPA results from their university entrance qualification, which is known to be, in general, a strong indicator for academic success. This hypothesis is based on the results of previous studies, such as TEDS-M and TEDS-FU, which have shown that the grade of the Abitur turned out to be a valid predictor of academic success and fast perception of student error mediated by MCK and MPCK.

4 Design of the study and methodology

This section describes the instrument used to measure the fast perception of student errors, which was implemented as part of the national TEDS-FU study, as well as the participants of the five contrast samples.

4.1 Description of the test instrument

The fast perception of student errors is part of the range of facets that constitute teacher competence. The test instrument used to measure the fast perception of student errors consists of 16 items, all of which are structured in the same way. This uniform presentation helps ensure that the instrument takes a uniform approach to the assessment of the different types of student errors and, thus, avoids methods bias.

The instrument was implemented online, and the tests took place at a self-chosen location with an internet connection (mainly at home). Since the participants were spread out across Germany, the online tool proved to be the ideal instrument for the test, since it allowed a widely disbursed sample of participants to complete the test in their own familiar environments, thus minimizing any distractions.

All student errors were derived from mathematical topics that constitute main themes of lower-secondary-level mathematics curriculum in Germany. Before the study began, participants were asked to think about their experience dealing with typical student errors, then an introductory item was shown to allow the participants to become acquainted with the test format. The test contained as specific feature an *anticipating phase*, i.e., before each student error was shown, the topic area from which the student error would come was shown to the participants (e.g., 'addition of fractions' in the case of Fig. 3, Item 1). The participants were asked to think about possible student errors coming from this mathematical topic. This phase was not strongly time-restricted—the participants could think as long as they considered appropriate, up to 5 min. This test design was developed to come closer to the reality of a classroom, where teachers know the topics they are teaching and have probably thought about common student errors or misconceptions in advance.

Once the participants decided to proceed with each item, they were shown three different student answers, of which one presented a typical student error that needed to be identified, while the other two answers were correct. The test was implemented in such a way that the participants were required to answer within four seconds. The maximum length of four seconds per item was determined by a group of 15 experts consisting of researchers from mathematics education and experienced teachers. The purpose of the time restriction was to prevent the participants from calculating the items in the given time, as this was not the purpose of the assessment; rather, the test was intended to measure the time it took to perceive the errors. If the test taker exceeded the time given, the response was considered incorrect. Practically, the test taker had to place one of three fingers of the left hand on each of the keys "s", "d," and "f" on a computer keyboard; these letters corresponded with the three given students answers, and the key indicating the erroneous student response had to be entered within the given time. Guess probability was 33% with each item.

For illustration purposes, two items of the test are displayed below (see Fig. 3). The first item addresses the mathematical extension of addition from integers to rational numbers. According to Padberg (2009), this is an area of difficulty for many students that will continue to reappear throughout the course of students' school careers. Wartha (2007) points out that nearly half of the errors made by students attempting to add two fractions are based on the approach of adding numerators and denominators separately. Additionally, students often develop their own set of rules when faced with the problem of adding two fractions (Padberg 2009).

The second item is less researched empirically, but well-known to practitioners. In German mathematics teaching, quadratic equations are usually solved with a specific formula after having standardized the quadratic equation to its

Fig. 3 Example items of "addition of two fractions" and "application of the solution formula for quadratic equations"

Nr.	Announcement	s	d	f
1	Fractions: Addition of two fractions	$\frac{1}{2}+\frac{1}{4}=\frac{3}{4}$	$\frac{2}{7}+\frac{3}{7}=\frac{5}{14}$	$\frac{1}{3}+\frac{2}{3}=1$
4	Application of the p-q formula for quadratic equation $x^2 + 8x - 4 = 0$	$x^2 + 8x - 4 = 0$		
		$x_{1,2} = -4 \pm \sqrt{16+4}$	$x_{1,2} = -4 \pm \sqrt{16-(-4)}$	$x_{1,2} = -4 \pm \sqrt{16-4}$

normal form. The typical student error results while applying this formula $x_{1,2} = -\frac{p}{2} \pm \sqrt{\left(\frac{p}{2}\right)^2 - q}$. In addition to the probably missing standardization (Ritter and Voß 2015), the plus/minus sign ahead of the "q" is of importance—a double-sign issues are likely to be encountered there, as shown in the example. Another difficulty is the p-q formula itself; it is, like other formulae and mathematical strategies, often applied without referring to its meaning (Allmendinger et al. 2013).

In sum, a variety of potential errors could be anticipated by the participant. These errors can be traced back to either the usage of the wrong concept by the student when adding fractions, or to the wrong application of the solution formula of quadratic equations, especially concerning the usage of the algebraic sign.

4.2 Samples

As already mentioned within the discussion of this study's research goals, five different groups of participants were included in the validation study. Three groups were composed of teachers who were in different stages of their careers (see Table 1, Samples 1–3) and therefore had different levels of teaching experience. Two more groups were made up of students: one of university mathematics majors, the other, high school students. The two student groups were chosen to serve as contrast groups to the group of early career teachers included in the original TEDS-FU study. The university mathematics majors (Sample 4) are assumed to represent the highest accumulated level of MCK, whereas the senior high school students (Sample 5) are assumed to represent the lowest accumulated level of MCK. All three groups of teachers were assumed to fall somewhere between these two contrast sample groups.

This sample design was based on research in teacher expertise (Li and Kaiser 2011), and groups were chosen to reflect the development of teachers' expertise from the start of their initial training to later stages in their career. Furthermore, the increase in reaction time may reflect the increasing age of participants.

In the following we describe the various samples (Table 2).

Sample 1 The participating future teachers were in the first phase of their teaching career at the University of Hamburg. They had already completed their bachelor's degree for teaching at the primary or secondary level. All participants (n = 42) were in an early stage of completing their master's degree, and the average age among them was 24 years (min.: 21; max.: 36; SD: 0.4). Males made up 11 percent of participants in this sample; 89 percent were female. They had completed their university entrance qualification (Abitur) with an average of 2.1 (SD = 0.4), which is above the average of the North German Federal States, from where most students are originating, which is currently 2.5; the highest mark is 1.0.

Sample 2 The participants of TEDS-FU had originally participated in TEDS-M at the end of their German teacher education studies and had agreed to take part in a follow-up study 3.5 years later. They were spread all over Germany. Of the 171 data sets provided, 137 were valid for this test component. The participants' teaching experience ranged from three to four years, excluding their teaching traineeship.

Table 1 Samples and their sizes

	Mathematician			Non- Mathematician
Teacher	Sample 1: Future teachers (n = 42)	Sample 2: Early career teachers from TEDS-FU (n = 137)	Sample 3: Experienced teachers from TEDS-Instruct (n = 113)	–
Non-teacher	Sample 4: University mathematics majors (n = 151)			Sample 5: Senior high school students (n = 62)

Table 2 Demographic details of the five sample groups

Sample	N	Age (years)	Gender	University entrance qualification
Sample 1: Future teachers	42	24 (SD = 0.4)	89% female	2.1 (SD = 0.4)
Sample 2: Early career teachers (data from TEDS-FU)	137	32 (SD = 5.9)	63% female	2.1 (SD = 0.6)
Sample 3: Experienced teachers (data from TEDS-Instruct)	113	39 (SD = 10.4)	50% female	1.71 (SD = 0.5)
Sample 4: University mathematics majors	151	21 (SD = 4.06)	26% female	1.87 (SD = 0.4)
Sample 5: Senior high school students	62	17–21 (estimated)	37% female	–

Their mean age was 32 years (SD = 5.9). About two-thirds of the sample were female (63%); all participants had completed their university entrance qualification with an average of 2.1 (SD = 0.6).

Sample 3 These participants, experienced teachers from the German Federal State of Hamburg, took part in TEDS-Instruct, another follow-up study of TEDS-M. 113 participant data sets were deemed valid for use in the analysis. The participants indicated a broad range of teaching experience. Just over half of the participants in the sample were female (50.4%); the mean age was 39 years (SD = 10.4). Participants in this sample had completed their university entrance qualification with an average of 1.71 (SD = 0.5).

Sample 4 These participants were studying either at the University Hamburg or the University of Bremen. They were at the end of their bachelor study and aiming for a degree in mathematics. Their participation was voluntary and by the invitation of the university lecturer, who was asked to help recruit participants for the study. This group had an average age of 21 years (SD = 4.06); 26 percent of its members were female, and they had completed its university entrance qualification with an average of 1.87 (SD = 0.4). During their bachelor program, the students had participated in 4.7 (SD = 3.1) mathematics courses.

Sample 5 The high school students were tested in two mathematics courses at a comprehensive school in Hamburg. The participants ranged in age from 17 to 21 years. This sample was 37 percent female, and all students in the class participated in the survey.

Table 3 Sample and their correct answers

Sample no.	Sample	N	Median (min.; max.)
4	University mathematics majors	151	9 (0;16)
1	Future mathematics teachers	42	8 (4;13)
2	Early career teachers (TEDS-FU)	137	7 (0;14)
3	Experienced teachers (TEDS-Instruct)	113	6 (0;13)
5	Senior high school students	62	3 (0;9)

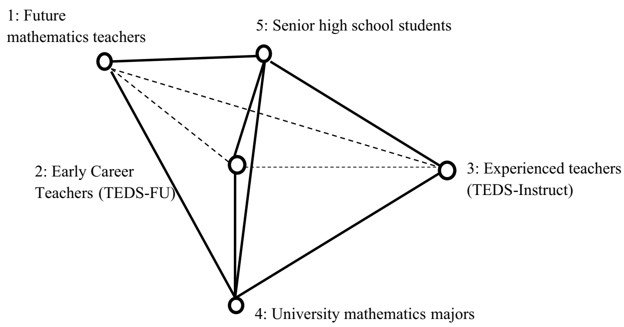


Fig. 4 Pairwise comparison of the different samples [dashed lines mean no significant difference; continuous lines indicate significant difference]

4.3 Methodology

In order to compare the test results across groups, a certain level of *measurement invariance* is a necessary requirement. Testing for the equivalency of the measurement ensures that the same construct is measured in every sub-sample. Measurement invariance is often tested by constraining parameters in a confirmatory factor analysis, evaluating the model fit, and comparing it with the fit of a less restrictive model. Followed the approach taken by Krauss et al. (2008), subscales were constructed that represent typical student errors in handling fractions, functions, or algorithmic or arithmetic problems (four items each). In all subscales, a participant was able to achieve a score between 0 and 4. These subscales were then used as indicators through the process of item-parceling (see Bandalos and Finney 2009). The indicators were used to conduct a series of confirmatory factor analyses with lavaan 0.5 series (Rosseel 2017) under R version 3.4.3. First, the same measurement model was applied to all groups—i.e., the construct was operationalized by the same items for every group. In this first step, the parameters (such as factor loadings and intercepts) were estimated freely. Second, the factor loadings were constrained, which results in so-called metric invariance. Metric invariance is required if one is to conclude that the same construct is being examined in all subpopulations (see Schwab and Helm 2015 for more information). In a third step, further parameters such as intercepts and residual variances may be constrained (scalar invariance). Scalar invariance ensures that all the items were measured on the same level. Hence, absolute comparisons on the latent variable between groups become possible.

5 Results

A multi-group confirmatory factor analysis with subscales as indicators showed an almost perfect fit for both configural ($\chi^2 = 8.64$, $df = 10$, $p = .56$, CFI = 1.00, RMSEA = 0.00) and metric invariance ($\Delta\chi^2 = 16.95$, $\Delta df = 12$, $p = .77$). However, restricting the model further to scalar invariance resulted in a significantly worse model fit ($\Delta\chi^2 = 81.42$, $\Delta df = 12$, $p < .01$). Comparisons of mean group results have therefore to be used with care whereas the prerequisites for a quasi-experimental design with non-equivalent control groups as applied in this study are satisfied.

The following explores the question of the degree to which the evidence for construct validity of the fast perception of student errors can be supported based on the performance analysis of different sample groups (see Table 3; Fig. 4).

In contrast to the expectations in first hypothesis, the early career teachers did not perform best on the test of fast perception of student errors; their results came in between the university mathematics majors and high school students. Hypothesis 2 could be confirmed though, because the group of high school students received the lowest results, with a median of 3 correct answers—lower than the probability of guessing (5 correct answers). Also Hypothesis 3 was supported by the data with respect to future and experienced mathematics teachers who came in between mathematics majors and high school students.

A Kruskal–Wallis H test showed that there was a statistically significant difference in the average scores received on the timed test between the groups, $\chi^2(4) = 101.374$, $p = .000$, with a rank score of 313.62 for university mathematics majors, 260.98 for future mathematics teachers, 255.44 for early career teachers (TEDS-FU), 217.80 for experienced teachers (TEDS-Instruct) and 105.15 for senior high school students.

A pairwise comparison, shown in Fig. 4, analyzes the average ranking of the groups and examines statistical significance. The displayed knots represent the groups, bold connections describe highly significant differences between the groups, and the dashed lines show that the connected groups do not differ significantly. These results show that the groups of teachers do not significantly differ from each other, whereas the two contrast groups—the senior high school students and the university mathematics majors—differ significantly from all other groups.

All participants in the original study and in the control groups were asked to disclose their age. Although 258 participants answered this question, none of the high school students did, so they were excluded from this analysis and its result. The results point to a negative correlation between success on the test and age, $r_s = -0.227$ ($p = .000$), meaning that as the age of the test takers increases, test scores tend to decrease.

Concerning the fourth hypothesis—the ability to predict the speed of student-error perception based on university entrance qualification scores—the following results were achieved. Data on 359 participants show a small but highly significant relation between university entrance examination results and achievement in the test on the fast recognition of student errors, $r_s = 0.14$ ($p < .001$). These results are in line with those of previous studies such as TEDS-M (Blömeke et al. 2010) or TEDS-FU (Blömeke et al. 2014).

5.1 Summary and discussion of the main results

On average, the university mathematics majors received the highest scores in the test of fast perception of student errors, which is similar to, though not exactly the same as, the results in the COACTIV study. This result indicates the importance of sound mathematics content knowledge by mathematics teachers for perceiving student errors.

The results found among the various groups of teachers were significantly better than the results of the high school students, which supports the high relevance of mathematical content knowledge and of mathematics pedagogical knowledge for the fast perception of student errors; the high school students were expected to possess low knowledge in both domains.

Results among the various groups of practicing and future teachers were not very distinct from one another, which implies that teaching experience in our case does not play a strong role in determining the speed of student-error perception. One reason for this may be that the student errors in the test instrument were not embedded in the type of pedagogically rich situation that occurs in the classroom; one in which the teachers would have anticipated possible student errors already in their lesson preparation and where the knowledge of strengths and weaknesses of the individual student often plays a significant role. Outside this pedagogical environment, mathematical content knowledge seems to play the most relevant role.

University entrance qualification results could be used as an external variable to predict the test results of participants. Except for the high school students, who had not yet passed this examination at the test time, evidence for the validity of the instrument could be provided by significant correlations between academic success at the end of school and the results in the test of fast perception of student errors, for every sample in the study. The results for fast perception of student errors could confirm the assumed ranking of the groups according to their expertise and therefore confirm the evidence for validity based on relations to an external variable, namely university entrance qualification exam results, as extension of the criterion validity of the fast perception of student errors.

Concerning limitations of the study, it can be seen that only parts of the error perception cycle of Heinrichs and Kaiser (2018) were tested. Neither reasons for the perceived student error nor possible pedagogical measures were included in the test, although they belong to a comprehensive understanding of diagnostic competence as a sub-facet of mathematics pedagogical competence. To include these more situated aspects of error perception, more contextualized items would be necessary—such as those used by, for example, Heinrichs and Kaiser (2018). However, doing so could affect the clarity of the results seen here, as various other knowledge facets and indeterminable influences would come into play. Finally, it has to be considered that "it is commonly observed that the validation process never ends, as there is always additional information that can be gathered to more fully understand a test and the inference that can be drawn from it" (AERA, APA, and NCME 2014, p. 21–22).

As another limitation we have to point out that scalar measurement invariance did not exist, which limits the possibility of carrying out group mean comparisons. However, a contrast group approach can be regarded as a type of quasi-experimental design with non-equivalent groups, where metric invariance is sufficient.

Overall, the fast perception of student errors in pedagogical situations creates important opportunities to assess the level of students' knowledge and to help them depart from this level by focusing, if necessary, on students' misconceptions or other learning difficulties. Therefore, being able to perceive a student's error quickly is a key competence for quality-oriented instruction and needs to be stressed in teacher education, by fostering not only mathematics pedagogical content knowledge but also mathematical content knowledge as bases for this perception.

References

Allmendinger, H., Lengnink, K., Vohns, A., & Wickel, G. (2013). *Mathematik verständlich unterrichten: Perspektiven für Unterricht und Lehrerbildung*. Wiesbaden: Springer.

Altmann, A. F., & Nückles, M. (2017). Empirische Studien zu Qualitätsindikatoren für den diagnostischen Prozess. In A. Südkamp & A.-K. Praetorius (Eds.), *Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen* (pp. 142–149). Münster: Waxmann.

Ball, D., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59*(5), 389–407.

Bandalos, D. L., & Finney, S. J. (2009). Item parceling issues in structural equation modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 269–296). Mahwah: Erlbaum.

Binder, K., Krauss, S., Hilbert, S., Brunner, M., Anders, Y., & Kunter, M. (2018). Diagnostic Skills of mathematics teachers in the COACTIV study. In T. Leuders & K. Philipp K., & J. Leuders (Eds.), *Diagnostic competence of mathematics teachers* (pp. 33–53). Mathematics Teacher Education. Cham: Springer.

Blömeke, S., Busse, A., Kaiser, G., König, J., & Suhl, U. (2016). The relation between content-specific and general teacher knowledge and skills. *Teaching and Teacher Education, 56*(May), 35–46.

Blömeke, S., Gustafsson, J.-E., & Shavelson, R. J. (2015). Beyond dichotomies: Competence viewed as a Continuum. *Zeitschrift für Psychologie, 223*(1), 3–13.

Blömeke, S., Kaiser, G., & Lehmann, R. (2010). *(Eds.). TEDS-M 2008—Professionelle Kompetenz und Lerngelegenheiten angehender Mathematik-Lehrkräfte für die Sekundarstufe I im internationalen Vergleich*. Münster: Waxmann.

Blömeke, S., König, J., Busse, A., Suhl, U., Benthien, J., Döhrmann, M., & Kaiser, G. (2014). Von der Lehrerausbildung in den Beruf—Fachbezogenes Wissen als Voraussetzung für Wahrnehmung, Interpretation und Handeln im Unterricht. *Zeitschrift für Erziehungswissenschaft, 17*(3), 509–542.

Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation: Für Human- und Sozialwissenschaftler* (4. edn.). Heidelberg: Springer.

Brühwiler, C. (2017). Diagnostische und didaktische Kompetenz als Kern adaptiver Lehrkompetenz. In: A. Südkamp, & A.-K. Praetorius (Eds.), Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen (pp. 123–134). Münster: Waxmann.

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

Charalambous, C. Y. (2016). Investigating the knowledge needed for teaching mathematics. *Journal of Teacher Education, 67*(3), 220–237.

Heinrichs, H., & Kaiser, G. (2018). Diagnostic competence for dealing with students' errors: Fostering diagnostic competence in error situations. In T. Leuders, K. Philipp & J. Leuders (Eds.), *Diagnostic competence of mathematics teachers* (pp. 79–94). Cham: Springer.

Heinze, A. (2004). Zum Umgang mit Fehlern im Unterrichtsgespräch der Sekundarstufe I. *Journal für Mathematik-Didaktik, 25*(3–4), 221–244.

Hill, H. C., Ball, D., & Schilling, S. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education, 39*(4), 372–400.

Hill, H. C., Dean, C., & Goffney, I. M. (2007). Assessing elemental and structural validity: Data from teachers, non-teachers, and mathematicians. *Measurement: Interdisciplinary Research & Perspective*, 5(2–3), 81–92.

Hoth, J., Döhrmann, M., Kaiser, G., Busse, A., König, J., & Blömeke, S. (2016). Diagnostic competence of primary school mathematics teachers during classroom situations. *ZDM, 48*(1), 41–53.

Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement, 38*(4), 319–342.

Krauss, S., Baumert, J., & Blum, W. (2008). Secondary mathematics teachers' pedagogical content knowledge and content knowledge: Validation of the COACTIV constructs. *ZDM—The International Journal on Mathematics Education, 40*(5), 873–892.

Krauss, S., & Brunner, M. (2011). Schnelles Beurteilen von Schülerantworten: Ein Reaktionszeittest für Mathematiklehrer/innen. *Journal für Mathematik-Didaktik, 32*(2), 233–251.

Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M., & Jordan, A. (2008). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology, 100*(3), 716–725.

Leuders, T. (2001). *Qualität im Mathematikunterricht in der Sekundarstufe I und II*. Berlin: Cornelsen Scriptor.

Leuders, T., Philipp, K. & Leuders, J. (2018). Diagnostic competence of mathematics teachers. Cham: Springer.

Lindmeier, A. M., Heinze, A., & Reiss, K. (2013). Eine Machbarkeitsstudie zur Operationalisierung aktionsbezogener Kompetenz von Mathematiklehrkräften mit videobasierten Maßen. *Journal für Mathematik-Didaktik, 34*(1), 99–119.

Lissitz, R. W., & Samuelsen, K. (2007). A suggested change in terminology and emphasis regarding validity and education. *Educational Researcher, 36*(8), 437–448.

Madden, D. J. (2007). Ageing and visual attention. *Current Directions in Psychological Science, 16*, 2, 70–74.

Newton, P., & Shaw, S. (2016). Disagreement over the best way to use the word 'validity' and options for reaching consensus. *Assessment in Education: Principles, Policy & Practice, 23*(2), 178–197.

Oser, F., Hascher, T., & Spychiger, M. (1999). Lernen aus Fehlern: Zur Psychologie des "negativen" Wissens. In W. Althof (Ed.), *Fehlerwelten. Vom Fehlermachen und Lernen aus Fehlern* (pp. 11–41). Wiesbaden: VS Verlag für Sozialwissenschaften.

Padberg, F. (2009). *Didaktik der Bruchrechnung*. Heidelberg: Springer.

Pankow, L., Kaiser, G., Busse, A., König, J., Blömeke, S., Hoth, J., & Döhrmann, M. (2016). Early career teachers' ability to focus on typical students errors in relation to the complexity of a mathematical topic. *ZDM, 48*(1–2), 55–67.

Radatz, H. (1980a). *Fehleranalysen im Mathematikunterricht*. Wiesbaden: Vieweg + Teubner Verlag.

Radatz, H. (1980b). Students' errors in the mathematical learning process: A survey. *For the Learning of Mathematics, 1*, 16–20.

Reisman, F. K. (1976). *A guide to the diagnostic teaching of arithmetic*. Columbus: Merrill.

Ritter, S., & Voß, U. (2015). *Erfolgreich Starten ins Ingenieurstudium: Grundlagen der Mathematik anwendungsorientiert erklärt*. Berlin: Springer Vieweg.

Rosseel, Y. (2017). *Package* "*lavaan*": *Latent variable analysis. Comprehensive R Archive Network*. Retrieved from http://lavaan.org. Accessed 23 Dec 2017.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103*(3), 403–428.

Schlag, B. (1993). Elderly drivers in Germany—Fitness and driving behavior. *Accident Analysis and Prevention, 25*(1), 47–55.

Schoy-Lutz, M. (2005). *Fehlerkultur im Mathematikunterricht: Theoretische Grundlegung und evaluierte unterrichtspraktische Erprobung anhand der Unterrichtseinheit "Einführung in die Satzgruppe des Pythagoras"*. Hildesheim: Franzbecker.

Schwab, S., & Helm, C. (2015). Überprüfung von Messinvarianz mittels CFA und DIF-Analysen. *Empirische Sonderpädagogik, 7*(3), 175–193.

Shulman, L. S. (1986). Those who understand: knowledge growth in teaching. *Educational Researcher, 15*(2), 4–14.

Südkamp, A., & Praetorius, A.-K. (Eds.). (2017). *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen*. Münster: Waxmann.

Swan, M. (2004). Making sense of mathematics. In I. Thompson (Ed.), *Enhancing primary mathematics teaching and learning* (pp. 111–124). Maidenhead: Open University Press.

Türling, J. M. (2014). *Die professionelle Fehlerkompetenz von (angehenden) Lehrkräften: Eine empirische Untersuchung im Rechnungswesenunterricht*. Wiesbaden: Springer.

Wahl, D., Weinert, F. E., & Huber, G. L. (1984). *Psychologie für die Schulpraxis: Ein handlungsorientiertes Lehrbuch für Lehrer*. München: Kösel.

Wartha, S. (2007). Längsschnittliche Untersuchungen zur Entwicklung des Bruchzahlbegriffs. Hildesheim: Franzbecker.

7. Zentrale Ergebnisse

In diesem Abschnitt werden die zentralen Ergebnisse der Dissertation dargestellt, indem die formulierten Fragestellungen aus dem zweiten Abschnitt kurz dargestellt und anschließend beantwortet werden. Weitere Ergebnisse finden sich in den Beiträgen.

7.1. inhaltliche Komplexität und Antizipationszeit der Items

Jedes Item des zeitbegrenzten Tests zur schnellen Fehlererkennung enthält einen ersten Teil, in dem der Schülerfehler inhaltlich angekündigt wird und zu dem Themengebiet typische Schülerfehler antizipiert werden sollen. Der erste Beitrag mit dem Titel „Early Career Teachers' ability to focus on typical students' errors in relation to the complexity of a mathematical topic", erschienen in der Zeitschrift *ZDM Mathematics Education*, fokussiert diese Phase der Beantwortung der Items und stellt zunächst die Frage, ob die inhaltliche Komplexität erkannt werden kann. Weiterhin wird gefragt, ob ein Zusammenhang zwischen der inhaltlichen Komplexität und der Antizipationszeit seitens der Teilnehmer(innen) gibt.

Zur inhaltlichen Komplexität wurde zunächst ein Expertenrating durchgeführt, um zu zeigen, dass es einen inhaltlichen Unterschied bezüglich der 16 Itemankündigungen gibt. Dies konnte gezeigt werden. Die Expertiseforschung zeigt, dass Expertinnen und Experten die Fähigkeit haben, wahrgenommene Informationen nach relevant und irrelevant (vgl. Livingston und Borko 1989) zu filtern und bei Bedarf entsprechend zu fokussieren. Weiterhin wird das Konstrukt der schnellen Fehlererkennung entwickelt und die Zusammenhänge mit dem mathematischen und dem mathematikdidaktischen Wissen verdeutlicht, welches Lehrkräfte besitzen müssen, um Schülerfehler zu identifizieren. Der Artikel untersucht die Forschungsfrage nach dem Zusammenhang zwischen der gemessenen Antizipationszeit und der Kompetenz, Schülerfehler zu identifizieren. Dabei wird die Hypothese aufgestellt, dass die richtige Antworthäufigkeit und der Komplexität der Itemankündigung zusammenhängen.
Es konnte ein Unterschied in der Antworthäufigkeit bzgl. der Komplexität der Items festgestellt werden. Die Lehrkräfte, die richtig geantwortet haben, waren in der Lage, die Komplexität zu erkennen und haben entsprechend längere Zeit antizipiert. Bei den falsch Antwortenden waren die Antizipationszeiten bei komplexen und nicht komplexen Items gleich lang, sodass keine Fokussierung festgestellt werden konnte. Es konnte demnach auch bei den Junglehrkräften ein Unterschied bzgl. der Fähigkeit zum Fokussieren festgestellt werden.

In den Ergebnissen wird eine Nähe zu der Wissensfacette des mathematischen Wissens deutlich, die in einem weiteren Beitrag zu untersuchen ist. Ebenso wird die erste Phase der Itembeantwortung, die Antizipation, untersucht. Die zweite Phase ebenso wie die Untersuchung der Nähe zu MCK wird im zweiten bzw. im dritten Beitrag analysiert.

7.2. Merkmale zur Schwierigkeitsbestimmung der Items

Der zweite Beitrag mit dem Titel „Ein zeitbeschränkter Test zur schnellen Erkennung von Schülerfehlern durch Junglehrkräfte – Qualitative Merkmale zur Schwierigkeitsbestimmung von Items“, erschienen in der *mathematica didactica*, fokussiert die zweite Phase der Itembeantwortung: die Identifizierung der falschen Schülerlösung. Im Rahmen des Beitrags wird die Frage nach schwierigkeitsgenerierenden Merkmalen der Items beantwortet. Der zweite Schritt widmet sich der Frage nach der Stabilität der Merkmale.

Die Ergebnisse zeigen, dass die Items in schwer und leicht zu lösende Items zu unterscheiden sind. Aus der Literatur können Merkmale abgeleitet werden, mit denen die Itemschwierigkeit inhaltlich erklären werden kann.

So kann anhand des Merkmals der „curricularen Relevanz“ die curriculare Einführung des Themas und die Häufigkeit des curricularen Auftretens des Themas im Mathematikunterricht der folgenden Jahrgangsstufen untersucht werden. Grundsätzlich werden die Inhalte der Schulmathematik im deutschen Mathematikunterricht im Sinne des Spiralprinzips zu verschiedenen Zeitpunkten und in verschiedenen Kontexten wieder aufgegriffen (Brinkmann 2002). Fehler zu Themenfeldern, die in den unteren Klassenstufen unterrichtet werden, können folglich häufiger von Lehrkräften wahrgenommen werden als jene, die erst in höheren Schulstufen eingeführt werden. Testitems, die typische Fehler zu Themenfeldern repräsentieren, die in unteren Klassenstufen unterrichtet werden, sollten somit zu leichteren Items gehören im Vergleich zu jenen, die in höheren Klassenstufen zum ersten Mal unterrichtet werden. Im Sinne des Spiralprinzips sollten ebenfalls solche Items leichter lösbar sein, die nach ihrer Einführung häufig unterrichtet werden, da sie in diversen Themenfeldern involviert sind.

Das Merkmal der „strukturellen Komplexität“ spiegelt den strukturellen Aufbau der Items wider. Anhand des genannten Merkmals werden die Zeichen als Objekte – ihre Definition und ihre Grammatik (Rechenregeln/Gesetze) – näher betrachtet. Unter Bezug auf die in den KMK-Bildungsstandards formulierte allgemeine Kompetenz, „mit den symbolischen, formalen und technischen Elementen der Mathematik umzugehen“, um Aufgaben bearbeiten zu können (vgl. Leiß & Blum 2007, S. 47), kann davon ausgegangen werden, dass die Zahlen und Symbole, aus denen die Items bestehen, seitens der Teilnehmenden zunächst wahrgenommen, dann interpretiert und schließlich im Gesamtzusammenhang beantwortet werden müssen. Dabei ist anzumerken, dass mathematische Ausdrücke häufig in Gruppen bzw. Cluster zusammengefügt werden, die einer anderen Logik folgen als etwa die der geschriebenen Sprache (vgl. Dyrvold 2016, S. 11).

Das Merkmal der curricularen Relevanz ist – wie oben ausgeführt – am robustesten: Items aus Themengebieten, die in einer früheren Klassenstufe bereits unterrichtet wurden, sind leicht(er) zu lösen. Das Merkmal der strukturellen Komplexität bedeutet, dass ein Item, das viele Zeichen beinhaltet und zu dessen Lösung verschiedene Regeln angewandt werden müssen, als schwer gelten muss.

Die dritte Frage, die in diesem Beitrag beantwortet wird, gilt der Stabilität der Merkmale. Diese Frage wurde mit der Anwendung der Merkmale auf den zeitbegrenzten Test zur schnellen Fehlererkennung für die Primarstufe beantwortet. Beide Merkmale konnten, wie im

Beitrag gezeigt, auf die 15 Items des Primarstufen-Tests angewandt werden. Die Ergebnisse der Teilnehmer(innen) stimmen mit den vorhergesagten Ergebnissen aus der Analyse der Merkmale überein, sodass hier von einem ersten Schritt der Validierung ausgegangen werden kann.
Nach wie vor unbeantwortet ist die Frage des ersten Beitrages nach der inhaltlichen Nähe zu der Wissensfacette MCK, ebenso konnte die Validität des Tests noch nicht endgültig gezeigt werden.

7.3. Validierung des Instruments

Die vierte und letzte Frage dieser Dissertation ist die Frage nach der Validität des Tests zur schnellen Fehlererkennung, die mittels Kontrastgruppenvergleich beantwortet werden soll.

Der Beitrag zu dieser Fragestellung mit dem Titel „Perception of student errors under time limitation: are teachers faster than mathematicians or students?" ist 2018 in der Zeitschrift *ZDM Mathematics Education* erschienen.

Im Rahmen der Validierung werden Kontrastgruppen vorgestellt, die unterschiedliche Wissensausprägungen bezüglich MCK und MPCK inkorporiert haben. Als Kontrastgruppen werden grundständige Mathematikstudierende, zukünftige Mathematiklehrkräfte, Teilnehmer(innen) von TEDS-FU, Teilnehmer(innen) von TEDS-Unterricht (vgl. Schlesinger & Jentsch 2016) und Oberstufenschüler(innen) mit dem Test zur schnellen Fehlererkennung getestet. Weiterhin werden externe Variablen als Indikatoren zur Validierung herangezogen.

Nach der Feststellung, dass ein Gruppenvergleich zulässig ist, da dasselbe Konstrukt gemessen wird, kann eine Reihenfolge bezogen auf das Abschneiden im Test bestimmt werden. Aus dieser Reihenfolge geht hervor, dass die Mathematikstudierenden am besten abschneiden, ihnen folgen die Lehrkräfte, während die Oberstufenschüler(innen) die Gruppe bilden, die am schlechtesten abgeschnitten hat. Ergänzend ist festzustellen, dass die Gruppen der Lehrkräfte je nach Lehrerfahrungen sich in ihrem Abschneiden nicht signifikant voneinander unterscheiden.

Die Ergebnisse weisen darauf hin, dass die Validität gezeigt werden konnte, indem die kontrastierenden Gruppen überhaupt miteinander verglichen werden durften. Weiterhin konnte der erwartete Zusammenhang mit der Wissensfacette MCK nachgewiesen werden.

8. Diskussion

Im Rahmen der Diskussion sollen die Grenzen des Instruments aufgezeigt werden und es wird ein Ausblick auf weitere forschungsbezogene Fortführungen gegeben, um den noch offenen Fragen nachzugehen.

8.1. Grenzen der Studie

Die Studie TEDS-FU sowie ihre Ergänzungsstudien haben Grenzen, die im Folgenden kritisch zu reflektieren sind. Die Studie ist eine Feldstudie, die internetbasiert in ganz Deutschland durchgeführt wurde. Es gab somit keine Testleitung, die für Rückfragen zur Verfügung gestanden hätte. Auch hätten in einer klassischen Laborstudie störende Umwelteinflüsse von den Teilnehmer(inne)n ferngehalten werden können. Des Weiteren war die Teilnahme an der Studie freiwillig, indem die Teilnehmer(innen) von TEDS-M gebeten wurden, sich erneut zur Verfügung zu stellen. Unter diesen Voraussetzungen ist von einer positiven Verzerrung der Ergebnisse auszugehen, insbesondere, da sich hauptsächlich für die Grundschulstudie jene Teilnehmer(innen) meldeten, die Mathematik als Fach studiert haben. Des Weiteren kamen in dem zeitbeschränkten Test 16 Items aus der Sekundarstufe zum Einsatz, die aus unterschiedlichen mathematischen Disziplinen stammen. Diese Auswahl war allerdings nicht gleichverteilt, so gab es beispielsweise nur ein einziges Item aus der Stochastik, im Unterschied zu jeweils mehreren Items aus anderen mathematischen Themengebieten. Als weiterer Kritikpunkt ist anzumerken, dass nicht erhoben wurde, in welchem Bundesland bzw. an welcher Universität oder Pädagogischen Hochschule die Ausbildung zur Lehrkraft erfolgt war und folglich unklar blieb, welche Lerngelegenheiten universitärer Art für den Umgang mit Schülerfehlern in Anspruch genommen worden waren.
Im Rahmen der Validierungsstichproben konnte aus ökonomischen Gründen und um höhere Teilnehmer(innen)zahlen zu erreichen, kein Test zum mathematischen Wissen (MCK) und mathematikdidaktischen Wissen (MPCK) durchgeführt werden. Die Teilnehmer(innen) wurden – wie oben dargestellt – aufgrund äußerer Kriterien für die Stichprobe rekrutiert, weshalb davon auszugehen ist, dass das angenommene Wissen überdurchschnittlich hoch ausgeprägt war.
Ein weiterer Aspekt liegt in der Betrachtung der Schülerlösungen bzw. in den Einzelheiten der Schülerlösung, die betrachtet werden, um den Fehler zu finden und das Item zu lösen. Auf Basis der Daten kann keine Analyse durchgeführt werden, welche Einzelheit(en) die Teilnehmer(innen) fokussiert haben.
Weiterhin ist kritisch darauf hinzuweisen, dass die Testkonstruktion nur auf die reine Fehlerwahrnehmung ausgerichtet ist; die anderen Elemente des Diagnosekreislaufs (vgl. Heinrichs 2015) wie beispielsweise der genaue Umgang mit dem Fehler oder dessen mögliche Erklärungen bleiben unberücksichtigt, die eigentlich der Diagnose zugehörig sind.

8.2. Ausblick

Aus den Analysen wird deutlich, dass die Antizipationsphase bei der schnellen Fehlererkennung weiterer Forschung bedarf, da deren Einfluss auf die Fehlererkennung aktuell noch nicht endgültig geklärt zu sein scheint.
Kleinere Ergänzungsstudien weisen darauf hin, dass sich die adäquate, umfassende Antizipation des Fehlers positiv auf die schnelle Fehlererkennung auswirkt. Es sollte qualitativ genauer untersucht werden, welche Fehler Lehrkräfte erwarten, um das Testinstrument auf Basis diesbezüglicher Erkenntnisse zu validieren und weiterzuentwickeln.

Daneben könnte in einem weiteren Follow-up von TEDS-FU eine Längsschnittstudie implementiert werden. In der Studie TEDS-Unterricht wurden erfahrene Lehrkräfte getestet, die in dem zeitbeschränkten Test keine guten Leistungen erbracht haben, obwohl sie als Expert(inn)engruppe gelten konnten. Die Teilnehmer(innen) von TEDS-FU könnten, ungeachtet der oben beschriebenen Grenzen, erneut kontaktiert werden und die Tests zum mathematischen und mathematikdidaktischen Wissen ebenso durchlaufen wie den veränderten Test zur schnellen Fehlererkennung, um ihre Leistungsentwicklung festzustellen und sie mit den erfahreneren Lehrkräften aus TEDS-Unterricht zu vergleichen.

Angesichts der hier präsentierten aktuellen Studienergebnisse ist ebenfalls darüber nachzudenken, inwieweit Lehrveranstaltungen zur Wahrnehmung von Schülerfehlern und zum konstruktiven Umgang mit diesen Fehlern in das Curriculum der Lehrerausbildung einschließlich der Lehrerfortbildung integriert werden sollten. Erste positive Ergebnisse für eine solche Intervention berichtet beispielsweise Heinrichs (2015) in ihrer Dissertation. Larrain und Kaiser (2019) stellen darüber hinaus das Konzept einer Diagnose-Schulung für angehende Lehrkräfte vor, bei der sich diese schon während ihrer berufspraktischen Ausbildung zu typischen Schülerfehlern austauschen und voneinander lernen. Gerade in dieser Nähe zur Unterrichtspraxis kann auf die Notwendigkeit des positiven, konstruktiven Umgangs mit Fehlern im Unterricht eingegangen werden, die nicht aus der Perspektive der Defizitorientierung wahrgenommen werden sollten, sondern als Lerngelegenheiten, die für die einzelnen Schüler(innen), die eine fehlerhafte Lösung präsentieren, und für die gesamte Lerngruppe als Lernchance genutzt werden sollten.

9. Literaturverzeichnis

AERA, APA &NCME = American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.

Altmann, A. F., & Nückles, M. (2017). Empirische Studien zu Qualitätsindikatioren für den diagnosischen Prozess. In A. Südkamp & A.-K. Praetorius (Eds.), *Pädagogische Psychologie und Entwicklungspsychologie: Vol. 94. Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen* (1st ed., S. 142–149). Münster: Waxmann.

Ball, D. & Bass, H. (2002). Toward a Practice-Based Theory of Mathematical Knowledge for Teaching. In E. Simmt & B. Davis (Hrsg.), *Proceedings of the 2002 Annual Meeting of the Canadian Mathematics Education Study Group*. (S. 3–14).

Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, *9* (4), 469–520.

Beutelspacher, A. (2008). Über die Unmöglichkeit und die Notwendigkeit von Fehlern in der Mathematik. In R. Caspary (Ed.), *Herder Spektrum: Vol. 5892. Nur wer Fehler macht, kommt weiter. Wege zu einer neuen Lernkultur* (1st ed., S. 86–96). Freiburg im Breisgau: Herder.

Binder, K., Krauss, S., Hilbert, S., Brunner, M., Anders, Y., & Kunter, M. (2018). Diagnostic Skills of mathematics teachers in the COACTIV study. In T. Leuders & K. Philipp K., & J. Leuders (Eds.), *Diagnostic competence of mathematics teachers* (pp. 33–53). Mathematics Teacher Education. Cham: Springer.

Blanck, B. (2006). Entwicklung einer Fehleraufsuchdidaktik und Erwägungsorientierung - unter Berücksichtigung von Beispielen aus dem Grundschulunterricht. *Schweizerische Zeitschrift für Bildungswissenschaften*, *28* (1), 63–83.

Blömeke, S., Gustafsson, J.-E., & Shavelson, R. J. (2015). Beyond Dichotomies. *Zeitschrift für Psychologie*, *223*(1), 3–13. https://doi.org/10.1027/2151-2604/a000194

Blömeke, S., König, J., Busse, A., Suhl, U., Benthien, J., Döhrmann, M., & Kaiser, G. (2014). Von der Lehrerausbildung in den Beruf – Fachbezogenes Wissen als Voraussetzung für Wahrnehmung, Interpretation und Handeln im Unterricht. *Zeitschrift für Erziehungswissenschaft*, *17*(3), 509–542. https://doi.org/10.1007/s11618-014-0564-8

Blömeke, S., Kaiser, G. & Lehmann, R. (Hrsg.) (2010). *TEDS-M 2008 - Professionelle Kompetenz und Lerngelegenheiten angehender Primarstufenlehrkräfte im internationalen Vergleich.* Münster: Waxmann.

Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation: Für Human- und Sozialwissenschaftler* (4., überarb. Aufl., [Nachdr.]). *Springer-Lehrbuch Bachelor, Master*. Heidelberg: Springer-Medizin-Verl.

Bortz, J., & Döring, N. (2016). *Forschungsmethoden und Evaluation: Für Human- und Sozialwissenschaftler* (5., überarb. Aufl.). *Springer-Lehrbuch*. Heidelberg: Springer.

Brinkmann, A. (2002). *Über Vernetzungen im Mathematikunterricht - eine Untersuchung zu linearen Gleichungssystemen in der Sekundarstufe I.* Duisburger Elektronische Texte.

Bromme, R. (1997). Kompetenzen, Funktionen und unterrichtliches Handeln des Lehrers. In F. E. Weinert, N. Birbaumer & C. F. Graumann (Hrsg.), *Psychologie des Unterrichts und der Schule* (Bd. 3, S. 177–212). Göttingen: Hogrefe Verl. für Psychologie.

Brühwiler, C. (2017). Diagnostische und didaktische Kompetenz als Kern adaptiver Lehrkompetenz. In A. Südkamp & A.-K. Praetorius (Eds.), *Pädagogische Psychologie und*

Entwicklungspsychologie: Vol. 94. Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen (1st ed., S. 123–134). Münster: Waxmann.

Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3., aktualisierte und erweiterte Auflage). *Always learning*. München: Pearson Deutschland.

Buschang, R. E., Chung, G. K. W. K., Delacruz, G. C., & Baker, E. L. (2012). Validating Measures of Algebra Teacher Subject Matter Knowledge and Pedagogical Content Knowledge. *Educational Assessment*, *17*(1), 1–21. https://doi.org/10.1080/10627197.2012.697847

Charalambous, C. Y. (2016). Investigating the Knowledge Needed for Teaching Mathematics. *Journal of Teacher Education*, *67*(3), 220–237. https://doi.org/10.1177/0022487116634168

Dyrvold, A. (2016). *Difficult to read or difficult to solve ? The role of natural language and other semiotic resources in mathematics tasks.* Umeå: Print & Media.

Gruber, H., Mack, W., & Ziegler, A. (1999). Wissen und Denken: Eine problematische Beziehung. In H. Gruber, W. Mack, & A. Ziegler (Eds.). *Wissen und Denken. Beiträge aus Problemlösepsychologie und Wissenspsychologie* (S. 7–16). Wiesbaden: Deutscher Universitäts-Verlag.

Hartig, J., Frey, A., & Jude, N. (2012). Validität. In H. Moosbrugger & A. Kelava (Eds.), *Testtheorie und Fragebogenkonstruktion* (S. 143–171). Berlin Heidelberg: Springer-Verlag. https://doi.org/10.1007/978-3-642-20072-4_7

Heinrichs, H. (2015). *Diagnostische Kompetenz von Mathematik-Lehramtsstudierenden: Messung und Förderung*. Wiesbaden: Springer VS.

Heinze, A. (2004). Zum Umgang mit Fehlern im Unterrichtsgespräch der Sekundarstufe I. *Journal für Mathematik-Didaktik*, *25*(3-4), 221–244. https://doi.org/10.1007/BF03339324

Herppich, S., Altmann, A. F., Wittwer, J., & Nückles, M. (2017). Förderung von Instuktionsstrategien zum verbesserten Diagnostizieren im Eins-zu-Eins-Tutoring. In A. Südkamp & A.-K. Praetorius (Eds.), *Pädagogische Psychologie und Entwicklungspsychologie: Vol. 94. Diagnostische Kompetenz von Lehrkräften. Theoretische und methodische Weiterentwicklungen* (1st ed., S. 203–208). Münster: Waxmann.

Hill, H. C., Ball, D., & Schilling, S. (2008). Unpacking Pedagogical Content Knowledge: Conceptualizing and Measuring Teachers' Topic-Specific Knowledge of Students. *Journal for Research in Mathematics Education*, *39*(4), 372–400.

Hill, H. C., Dean, C., & Goffney, I. M. (2007). Assessing Elemental and Structural Validity: Data from Teachers, Non-teachers, and Mathematicians. *Measurement: Interdisciplinary Research & Perspective*, *5*(2-3), 81–92. https://doi.org/10.1080/15366360701486999

Hoth, J. (2016). *Situationsbezogene Diagnosekompetenz von Mathematiklehrkräften: eine Vertiefungsstudie zur TEDS-Follow-Up-Studie.* Springer Spektrum, Wiesbaden.

Jenßen, L., Dunekacke, S., & Blömeke, S. (2015). Qualitätssicherung in der Kompetenzforschung. Empfehlungen für den Nachweis von Validität in Testentwicklung und Veröffentlichungspraxis. In S. Blömeke & O. Zlatkin-Troitschanskaia (Eds.), *Zeitschrift für Pädagogik. Beiheft: Vol. 61. Kompetenzen von Studierenden* (S. 11–31). Weinheim: Beltz Juventa.

Kaiser, G., Busse, A., Hoth, J., König, J., & Blömeke, S. (2015). About the Complexities of Video-Based Assessments. Theoretical and Methodological Approaches to Overcoming Shortcomings of Research on Teachers’ Competence. *International Journal of Science and Mathematics Education, 13* (2), 369–387.

KMK Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Hrsg.). (2015). Lehrerbildung für eine Schule der Vielfalt. Gemeinsame Empfehlung von Hochschuldirektorenkonferenz und Kultusministerkonferenz vom 12.03.2015). http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2015/2015_03_12-Schule-der-Vielfalt.pdf (letzter Zugriff am 9.12.2019)

Krauss, S., Baumert, J., & Blum, W. (2008). Secondary mathematics teachers' pedagogical content knowledge and content knowledge: Validation of the COACTIV constructs. *ZDM*, *40*(5), 873–892. https://doi.org/10.1007/s11858-008-0141-9

Krauss, S., Blum, W., Brunner, M., Neubrand, M., Baumert, J., Kunter, M.,. . . Elsner, J. (2011). Konzeptualisierung und Testkonstruktion zum fachbezogenen Professionswissen von Mathematiklehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 135–162). Berlin: Waxmann: Waxmann Verlag.

Krauss, S., & Brunner, M. (2011). Schnelles Beurteilen von Schülerantworten: Ein Reaktionszeittest für Mathematiklehrer/innen. *Journal für Mathematik-Didaktik*, *32*(2), 233–251. https://doi.org/10.1007/s13138-011-0029-z

Lange, K., Kleickmann, T., & Möller, K. (2012). Die Bedeutung des fachdidaktischen Wissens von Lehrkräften für Lernfortschritte von Schüler(innen) und Schülern im Sachunterricht der Grundschule. In H. Bayrhuber, U. Harms, B. Muszynski, B. Ralle, M. Rothgangel, L.H. Schön,. . . H.G. Weigang (Eds.), *Fachdidaktische Forschungen: Vol. 2. Formate fachdidaktischer Forschung. Empirische Projekte - historische Analysen - theoretische Grundlegungen* (S. 315–334). Münster: Waxmann.

Larrain, M., Kaiser, G. (2019). *Analysis of Students' Mathematical Errors as a Means to Promote Future Primary School Teachers' Diagnostic Competence.* Uni-pluriversidad.

Leiß, D., & Blum, W. (2007). Beschreibung zentraler mathematischer Kompetenzen. In W. Blum, C. Drüke-Noe, R. Hartung, & O. Köller. *Bildungsstandards Mathematik: konkret. Sekundarstufe I: Aufgabenbeispiele, Unterrichtsanregungen, Fortbildungsideen* (3. Aufl.) (S. 33-50). Berlin: Cornelsen Scriptor.

Leuders, T. (2001). *Qualität im Mathematikunterricht in der Sekundarstufe I und II* (1. Aufl.). Berlin: Cornelsen Scriptor.

Leuders, T., Philipp, K. & Leuders, J. (2018). *Diagnostic competence of mathematics teachers.* Cham: Springer.

Lindmeier, A. M., Heinze, A., & Reiss, K. (2013). Eine Machbarkeitsstudie zur Operationalisierung aktionsbezogener Kompetenz von Mathematiklehrkräften mit videobasierten Maßen. *Journal für Mathematik-Didaktik*, *34*(1), 99–119. https://doi.org/10.1007/s13138-012-0046-6

Loewenberg Ball, D., Thames, M. H., & Phelps, G. (2008). Content Knowledge for Teaching: What Makes It Special? *Journal of Teacher Education*, *59*(5), 389–407. https://doi.org/10.1177/0022487108324554

McElvany, N., Schroeder, S., Hachfeld, A., Baumert, J., Richter, T., Schnotz, W., . . .Ullrich, M. (2009). Diagnostische Fähigkeiten von Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, *23* (3), 223–235.

Moosbrugger, H., & Kelava, A. (Eds.). (2012). *Testtheorie und Fragebogenkonstruktion*. Berlin Heidelberg: Springer-Verlag.

Newton, P., & Shaw, S. (2016). Disagreement over the best way to use the word 'validity' and options for reaching consensus. *Assessment in Education: Principles, Policy & Practice*, *23*(2), 178–197.

Oser, F., Hascher, T., & Spychiger, M. (1999). Lernen aus Fehlern: Zur Psychologie des "negativen" Wissens. In W. Althof (Ed.), *Fehlerwelten. Vom Fehlermachen und Lernen aus Fehlern* (S. 11–41). Wiesbaden: VS Verlag für Sozialwissenschaften.

Oser, F., & Spychiger, M. (2005). *Lernen ist schmerzhaft: Zur Theorie des negativen Wissens und zur Praxis der Fehlerkultur*. Weinheim, Basel: Beltz Verlag.

Pankow, L., Kaiser, G., Busse, A., König, J., Blömeke, S., Hoth, J., & Döhrmann, M. (2016). Early Career Teachers' ability to focus on typical students errors in relation to the complexity of a mathematical topic. *ZDM*, *48*(1-2), 55–67. https://doi.org/10.1007/s11858-016-0763-2

Prediger, S., Wessel, L., Tschierschky, K., Seipp, B., & Özdil, E. (2013). Diagnose und Förderung: am Beispiel Mathematiklernen bei Deutsch als Zweitsprache. In C. Selter & S. Hußmann (Eds.), *Diagnose und individuelle Förderung in der MINT-Lehrerbildung. Das Projekt dortMINT* (S. 171–192). Münster [u.a.]: Waxmann.

Reisman, F. K. (1976). *A guide to the diagnostic teaching of arithmetic* (6. [printing]). Columbus, Ohio: Merrill.

Reiss, K., & Hammer, C. (2013). *Grundlagen der Mathematikdidaktik*. Basel: Springer Basel.

Rölke, H. (2012). The Item Builder: A graphical authoring system for complex item development. In T. Bastiaens & G. Marks (Eds.). *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education* (pp. 344–353). Chesapeake: AACE.

Schlesinger, L., & Jentsch, A. (2016). Theoretical and methodological challenges in measuring instructional quality in mathematics education using classroom observations, *ZDM*, *48* (1), 29-40.

Schnell, R., Hill, P. B., & Esser, E. (2011). *Methoden der empirischen Sozialforschung* (9., aktualisierte Aufl.). München: Oldenbourg.

Schoy-Lutz, M. (2005). *Fehlerkultur im Mathematikunterricht: Theoretische Grundlegung und evaluierte unterrichtspraktische Erprobung anhand der Unterrichtseinheit "Einführung in die Satzgruppe des Pythagoras". Texte zur mathematischen Forschung und Lehre: Vol. 39*. Hildesheim, Berlin: Franzbecker.

Schumacher, R. (2008). Der produktive Umgang mit Fehlern: Fehler als Lerngelegenheit und Orientierungshilfe. In R. Caspary (Ed.), *Herder Spektrum: Vol. 5892. Nur wer Fehler macht, kommt weiter. Wege zu einer neuen Lernkultur* (1st ed., S. 49–72). Freiburg im Breisgau: Herder.

Seidel, T., & Prenzel, M. (2003). Mit Fehlern umgehen - Zum Lernen motivieren. *Praxis der Naturwissenschaften - Physik*, *51*(1), 30–34.

Sherin, M. & van Es, E. (2002). Using Video to Support Teachers' Ability to Interpret Classroom Interactions. In D. Willis, J. Price & N. Davis (Eds.), *Proceedings of SITE 2002--Society for Information Technology & Teacher Education International Conference* (pp. 2532-2536). Nashville, Tennessee, USA: Association for the Advancement of Computing in Education (AACE). Retrieved December 18,

Shulman, L. S. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher*, *15*(2), 4–14. https://doi.org/10.3102/0013189X015002004

Steuer, G. (2014). *Fehlerklima in der Klasse. Zum Umgang mit Fehlern im Mathematikunterricht*. Wiesbaden: Springer VS.

Südkamp, A., & Praetorius, A.-K. (Eds.). (2017). *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen* (1. Auflage). *Pädagogische Psychologie und Entwicklungspsychologie: Vol. 94*. Münster: Waxmann.

Swan, M. (2004). Making sense of mathematics. In I. Thompson (Ed.), *Enhancing primary mathematics teaching and learning* (S. 111–124). Maidenhead: Open University Press.

Terhart, E. (2011). Lehrerberuf und Professionalität. Gewandeltes Begriffsverständnis - neue Herausforderungen. In: Helsper, W. (Hrsg.). Pädagogische Professionalität. *Zeitschrift für Pädagogik* - 57. Beiheft. Weinheim, 202-224.

Türling, J. M. (2014). *Die professionelle Fehlerkompetenz von (angehenden) Lehrkräften: Eine empirische Untersuchung im Rechungswesenunterricht*. Wiesbaden: Springer.

Wahl, D., Weinert, F. E., & Huber, G. L. (1984). *Psychologie für die Schulpraxis: Ein handlungsorientiertes Lehrbuch für Lehrer*. München: Kösel.

Weinert, F. E. (2001). Qualifikation und Unterricht zwischen gesellschaftlichen Notwendigkeiten, pädagogischen Visionen und psychologischen Möglichkeiten. In W. Melzer & U. Sandfuchs (Hrsg.), *Was Schule leistet. Funktionen und Aufgaben von Schule.* (S. 65–86). Weinheim: Juventa.

Wuttke, E., & Seifried, J. (2017). Competence, Teacher Competence and Professional Error Competence: An Intruduction. In E. Wuttke & J. Seifried (Eds.), *SpringerBriefs in Education. Professional Error Competence of Preservice Teachers. Evaluation and Support* (S. 1–13). Cham: Springer International Publishing; Imprint: Springer.

Zusammenfassung

Die kumulative, publikationsbasierte Dissertation befasst sich mit der Analyse und Validierung des Tests zur schnellen Fehlererkennung, der im Rahmen von TEDS-FU (Teacher Education and Development Study – Follow Up) zum Einsatz gekommen ist. Hierfür werden zunächst das der Arbeit zugrundeliegenden Begriffsverständnis der Professionellen Kompetenz von Lehrkräften insbesondere bezogen auf Schülerfehler und deren Wahrnehmung erläutert. Weiterhin wird die Studie TEDS-FU als Nachfolger der internationalen Vergleichsstudie TEDS-M vorgestellt und gegenüber anderen relevanten Studien abgegrenzt. Abschließend wird die Güte des Testinstruments mittels Kontrastgruppenvalidierung ermittelt.

Zwei Aspekte des TEDS-FU-Instruments werden genauer analysiert: einerseits die Antizipationszeit der Probandinnen und Probanden und der Zusammenhang der Antizipationszeit mit der Komplexität der Ankündigung des folgenden Items im Test sowie andererseits die inhaltliche Erklärung der Itemschwierigkeit anhand der Evaluation der Schülerlösung. Die Validierung erfolgt mittels kontrastierender Gruppen bezüglich des mathematischen Wissens (Mathematical Content Knowledge – MCK) und des mathematikdidaktischen Wissens (MPCK – Mathematical Pedagogical Content Knowledge).

Im Rahmen von drei Zeitschriftenbeiträgen werden die genannten Inhalte untersucht, wobei die Studie TEDS-FU den Ausgangspunkt der Analysen bildet. Die im Rahmen der Dissertation eingereichten Forschungen wurden im Rahmen von weiteren Studien wie TEDS-Unterricht vertieft. Die Ergebnisse zeigen, dass Berufsanfängerinnen bzw. -anfänger, obwohl sie über nur vier Jahre Berufserfahrung verfügten, ebenso wie langjährige Expertinnen und Experten dazu in der Lage waren, einzelne Informationen in komplexen Situationen zu fokussieren. Damit konnten zwei stabile Merkmale identifiziert werden, mit denen die Itemschwierigkeit inhaltlich erklärt werden kann: curriculare Relevanz einerseits und strukturelle Komplexität andererseits.

Nach Aussagen zur Testgüte werden die Schwächen des Instruments identifiziert, verbunden mit Lösungsansätzen als Ausgangspunkt weiterer Forschungen.

English Summary

The publication-based thesis deals with the analysis and validation of a test for rapid student error recognition that was used in the context of TEDS-FU (Teacher Education and Development Study - Follow Up). For this purpose, the underlying conceptual frameworks of professional teacher competence especially concerning student errors and their perception are presented first. Subsequently, the TEDS-FU study is presented as the successor study of the international comparative TEDS-M and in distinction to other studies. Finally, the quality of the test instrument is determined by means of contrast group validation.

Two aspects of the instrument TEDS-FU are analyzed in more detail: on the one hand, the subjects' anticipation time and the relation between the anticipation time and the complexity of the announcement of the following test item; on the other hand, the contextual explanation of the item difficulty based on the evaluation of the student responses. Validation is carried out by contrast groups regarding mathematical content knowledge (MCK) and mathematical pedagogical content knowledge (MPCK).

The described content is examined within the framework of three journal articles and the study TEDS-FU serves as the starting point for these analyses. Additional research was conducted in the context of the TEDS studies and by means of further studies. The results show that early career teachers with only four years of professional teaching experience were as capable as experienced teachers concerning the focusing in complex situations. Thereby two stable characteristics could be identified to explain item difficulty in terms of content: curricular relevance on the one hand and structural complexity on the other.

Concluding, the statements on test quality are followed by a discussion of the remaining open questions, furthermore the weaknesses of the instrument are identified, presented by potential solutions as initial point for further research.

Liste der Publikationen von Lena Pankow

Pankow, L., Kaiser, G., Busse, A., Hoth, J., Döhrmann, M., König, J. & Blömeke, S.(2015). Wahrnehmung von Schülerfehlern unter Zeitdruck als Aspekt von professioneller Kompetenz berufstätiger Mathematiklehrkräfte. *Beiträge zum Mathematikunterricht 2015*, 688-691.

Pankow, L. & Benecke, K. (2016). Wahrnehmung von Schülerfehlern unter Zeitdruck - Ergebnisse aus TEDS-FU. *Beiträge zum Mathematikunterricht 2016*, 727-730.

Pankow, L., Kaiser, G., Busse, A., König, J., Hoth, J., Döhrmann, M., & Blömeke, S. (2016). Early career teachers' ability to focus on typical students errors in relation to the complexity of a mathematical topic. *ZDM Mathematics Education, 48*(1), 55-67.

Pankow, L. (2017). Merkmale zum Vergleich zeitbegrenzter Tests zum schnellen Erkennen von Schülerfehlern. *Beiträge zum Mathematikunterricht 2017*, 745-748.

Pankow, L., Kaiser, G., König, J. & Blömeke, S. (2018). Perception of student errors under time limitation: are teachers faster than mathematicians or students? *ZDM Mathematics Education, 50*(4), 631–642.

Pankow, L. & Kaiser, G. (2018). Ein zeitbeschränkter Test zur schnellen Erkennung von Schülerfehlern durch Junglehrkräfte – Qualitative Merkmale zur Schwierigkeitsbestimmung von Items. *mathematica didactica, 41*(2), 147-162.

Erklärung über die Eigenständigkeit der Dissertation

Eidesstattliche Versicherung

Hiermit versichere ich, dass die Dissertation von mir selbst verfasst wurde, und dass keine anderen als die angegebenen Hilfsmittel von mir genutzt wurden.

5. Januar 2020	(Signatur entfällt aus datenschutzrechtlichen Gründen)
Datum	Unterschrift

Erklärung

Ich erkläre hiermit, dass ich keine kommerzielle Promotionsberatung in Anspruch genommen habe, und dass ich mich bisher keiner weiteren Doktorprüfung unterzogen habe. Insbesondere habe ich die Dissertation in der gegenwärtigen oder einer anderen Fassung an keiner anderen Fakultät eingereicht.

5. Januar 2020	(Signatur entfällt aus datenschutzrechtlichen Gründen)
Datum	Unterschrift