

Modellierung von Kompetenzen in der beruflichen Bildung

Entwicklung und Validierung eines Kompetenzniveaumodells für das externe Rechnungswesen

KURZFASSUNG: Die Messung von Kompetenzen ist eine der wesentlichen Strömungen in der gegenwärtigen bildungspolitischen Diskussion. Zur empirischen Bestimmung von Kompetenzen sind theoretisch fundierte Kompetenzmodelle nötig, die domänenspezifisch zu formulieren sind. Die Domäne Rechnungswesen stellt ein Kernelement der Betriebswirtschaftslehre und der kaufmännischen Bildung dar. Ziel dieses Beitrags ist es, ein valides und reliables Kompetenzniveaumodell zu entwickeln, mit dem die Schwierigkeit von Anforderungssituationen im Bereich des externen Rechnungswesens eingeschätzt werden kann. Die theoretische Basis des Modells bilden die berufs- und wirtschaftspädagogische Forschung im Bereich der Kompetenzmessung sowie die Didaktik des Rechnungswesens. Demnach werden die Kompetenzdimensionen verstehens- und handlungsbasiert unterschieden. Das Kompetenzniveaumodell setzt sich aus den drei jeweils vierstufigen Schwierigkeitsdimensionen Modellierung, Informationsgewinnung und Informationsverarbeitung zusammen. Zur empirischen Prüfung des Modells dienen fünf IHK-Abschlussprüfungen von Industriekaufleuten mit insgesamt $n = 2.195$ Teilnehmern. Für die handlungsbasierte Kompetenzdimension konnten diese drei Schwierigkeitsdimensionen empirisch bestätigt werden.

Abstract: Measuring of competencies is a prominent topic in educational research. For this purpose, theoretically founded and empirically validated competence models are required. Financial accounting is a core element of business education. However, an elaborated competence model does not exist for this domain. Therefore, this paper aims in developing a competence level model for financial accounting. Our model is based on the didactics of accounting and the research in business education. Due to the latter, we distinguish between an action and a comprehension based competence dimension. The final competence level model comprises three four-level dimensions: Modelling, Gathering of information, and Processing of information. To empirically test our model, we used a sample of 2,195 industrial clerks and successfully validated our model for action based tasks.

1 Einführung

Mit dem Wandel zu einer Wissens- und Informationsgesellschaft geht die Notwendigkeit des lebenslangen Lernens einher (STAUDINGER/PATZWALD 2009, S. 13). Viele der erforderlichen Lernprozesse vollziehen sich nonformal und informell im beruflichen oder privaten Kontext (STRAKA 2000, S. 27 ff.; WEBER 2013, S. 1). Vor diesem Hintergrund scheint es geboten, auf die Ergebnisse von Lernprozessen zu fokussieren, unabhängig davon, wie und auf welche Weise sie erreicht wurden (BOHLINGER 2010, S. 37 f.). Diese Outcome Orientierung soll auf europäischer Ebene durch einen Europäischen Qualifikationsrahmen für lebenslanges Lernen (EQR) Berücksichtigung finden (FIETZ 2008, S. 3; MÜNK/SCHELTEN/SCHMID 2010, S. 8). Voraussetzung hierfür ist die valide und reliable Messung der erworbenen Kompetenzen (PREISS 2005a, S. 56; BAETHGE u. a. 2006).

Im allgemein bildenden Bereich existiert eine Vielzahl an Beiträgen, die zum Ziel haben, Kompetenzen zu messen (HARTIG/JUDE 2007, S. 17; KLIEME/LEUTNER 2006,

S. 877). Dabei zeigen sich allerdings zentrale Unterschiede zur Messung von Kompetenzen im beruflichen Kontext (HAASLER/RAUNER 2009, S. 465 f.): Im allgemein bildenden Bereich sind übergeordnete Fachideen, beispielsweise Mathematik, Gegenstand der Kompetenzmessung. Im beruflichen Kontext ist hingegen auf Handlungen, beispielsweise das Durchführen einer Lieferantenauswahl, abzu zielen (ACHTENHAGEN/WINTHER 2008, S. 121; WINTHER/ACHTENHAGEN 2008, S. 513). Das damit verbundene Konzept der beruflichen Handlungskompetenz ist als Leitziel der beruflichen Bildung anerkannt (BADER 2004, S. 20; STRAKA/MACKE 2003, S. 43). Für eine reliable Messung ist das Konstrukt der Handlungskompetenz zu spezifizieren (HARTIG 2008, S. 21 ff.; SEEBER/NICKOLAUS 2010, S. 10 f.). Vor dem Hintergrund der Outcome Orientierung wird auch im kaufmännischen Bereich auf die Definition von Weinert (2002, S. 27 f.) zurückgegriffen (SEEBER/NICKOLAUS 2010, S. 10; WINTHER 2010, S. 21). Demnach sind Kompetenzen „die bei Individuen verfügbaren oder durch sie erlernbaren 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“. Diese Konzeption von Kompetenz bildet die Grundlage für diesen Beitrag und ist Ausdruck einer kognitiven Kompetenzauffassung (HARTIG/KLIEME 2006, S. 129). Demnach sind nicht nur die Handlungsergebnisse von Interesse, sondern auch die zugrunde liegenden kognitiven Prozesse. Folglich müssen die Problemlöseprozesse bei der Kompetenzmodellierung berücksichtigt werden. Das macht die Bedeutung von Lerntheorien und der Fachdidaktik im Rahmen der Kompetenzmodellierung evident (PELLEGRINO/CHUDOWSKY/GLASER 2001, S. 44 ff.; WINTHER 2011, S. 37).

Vorliegende Studien zur Kompetenzmessung im beruflichen Bereich wie die „Untersuchung der Leistungen, Motivation und Einstellungen der Schülerinnen und Schüler der Berufsschulen und der Berufsfachschulen“ (ULME) legen die Ausdifferenzierung einer eigenständigen Rechnungswesen-Komponente nahe (SEEBER 2013, S. 156 f.). So konnte in der ULME III Studie, in der allgemeine und berufliche Kompetenzen von Bürokaufleuten am Ende der Berufsausbildung gemessen wurden, lediglich 13 % der Varianz im Sub-Fachleistungstest Rechnungswesen durch allgemeine kognitive Grundfähigkeiten erklärt werden (SEEBER 2013, S. 175 f.). Hinzu kommt die hohe Bedeutung des betrieblichen Rechnungswesens im Rahmen der kaufmännischen Bildung (PREISS 2005b, S. 61; TRAMM 2005, S. 102). Es ist als Kerndisziplin kaufmännischer Curricula (TÜRLING u. a. 2011, S. 392) in nahezu allen kaufmännischen Lehrplänen verankert.

Die Ausarbeitung theoretisch fundierter und empirisch geprüfter Kompetenzmodelle ist derzeit eine gravierende Forschungslücke im Bereich der Berufs- und Wirtschaftspädagogik (SEEBER u. a. 2010, S. 5). Gerade diese sind aber für eine valide und reliable Kompetenzmessung nötig (KLIEME u. a. 2009, S. 22–23, 81 ff.). Modelle zum Zwecke der Kompetenzmessung setzen sich dabei aus einem Struktur- und einem Niveaumodell zusammen (KLIEME/LEUTNER 2006, S. 883; NICKOLAUS 2011, S. 332 ff.). Kompetenzniveaumodelle geben Auskunft über das Schwierigkeitsniveau einer Anforderungssituation (SCHUMANN/EBERLE 2011, S. 77 f.). Das ist insofern von großer Bedeutung, weil eine Voraussetzung für die reliable Messung von Kompetenzen ein für die Testteilnehmer angemessener Schwierigkeitsgrad der Aufgaben ist (BÜHNER 2011, S. 506 f.).

Ziel dieses Beitrags ist es, auf Grundlage der Fachdidaktik des Rechnungswesens, herauszuarbeiten, welche Aufgabenmerkmale die Schwierigkeit von Anforder-

derungssituationen im externen Rechnungswesen bestimmen. Anschließend wird das entwickelte Kompetenzniveaumodell einer empirischen Prüfung unterzogen und damit validiert.

2 Stand der Forschung

Zur Didaktik des (externen) Rechnungswesens liegen zahlreiche Beiträge vor (BREUER 2011, S. 368f.; MINNAMEIER/LINK 2010, S. 114ff.; SEMBILL/SEIFRIED 2005; SEIFRIED 2002; PREISS 1999; PREISS/TRAMM 1996). Dagegen ist der Forschungsstand im Bereich der Kompetenzmessung im Rechnungswesen derzeit unbefriedigend. Im Wesentlichen sind es zwei Beiträge, die sich direkt mit der Modellierung bzw. Messung von Rechnungswesenkompetenz befassen. PREISS (2005a; 2005b, S. 69ff.) unterbreitete einen Vorschlag für ein Kompetenzmodell im Rechnungswesen, angelehnt an die PISA-Studie Mathematik. Dabei handelt es sich nicht um ein ausgearbeitetes Kompetenzmodell (ERNST 2012, S. 97). PREISS versucht mit seinem Modell die Gesamtheit des Rechnungswesens zu erfassen, beispielsweise auch Finanz- und Liquiditätsmanagement. Aufgrund dieser Breite ist es für den spezifischen Teilbereich des externen Rechnungswesens zu wenig ausdifferenziert. Zudem steht eine empirische Überprüfung des Modells von PREISS aus.

HELM und WIMMER (2012) entwickelten ein Instrument zur „Wissensüberprüfung von Basissenntnissen der Buchhaltung“. Der Arbeit liegt kein ausdifferenziertes Kompetenzmodell zugrunde. Die Erkenntnisse können deshalb nicht direkt für die eigenen Zwecke verwendet werden.

Die zahlreichen Beiträge aus dem internationalen Kontext zur Leistungsmessung im Bereich des Rechnungswesens (DAIGLE/HAYES/HUGHES 2007; STOUT u. a. 2005; STIVERS/CAMPBELL/HERMANSON 2000; DEMONG/LINDGREN/PERRY 1994) können nur bedingt nutzbar gemacht werden. Die wichtigsten Gründe sind: Die genannten Studien sind im universitären Kontext angesiedelt. Hierdurch ergibt sich eine gänzlich andere Ausrichtung, weil das Ziel gerade nicht in der Messung von beruflicher Handlungskompetenz liegt. Außerdem haben diese Studien einen sehr breiten Fokus, um so zu einer Gesamteinschätzung des Lernerfolgs in Rechnungswesenstudiengängen zu gelangen (STIVERS/CAMPBELL/HERMANSON 2000, S. 558). Dementsprechend werden auch die allgemeinen intellektuellen Fähigkeiten der Studierenden getestet. Erkennbar geht dadurch der direkte fachliche Bezug zum externen Rechnungswesen verloren.

WINTHER und ACHTENHAGEN (2008, S. 531f.) entwickelten auf entwicklungspsychologischer Basis (GELMAN/GREENO 1989) ein Kompetenzstrukturmodell für die kaufmännische Bildung. Dabei werden die beiden Dimensionen verstehensbasierte und handlungsbasierte Kompetenz unterschieden (WINTHER 2010, S. 196ff.). Verstehensbasierte Kompetenz ist die „Lösung eines definierten beruflichen Problems durch aktivierte (deklarative) Wissensbestände“, d. h. das Erbringen des Nachweises des Verständnisses zentraler domänenspezifischer Ideen. Handlungsbasierte Kompetenz ist „die prozessuale Kombination aus ‚procedural competence‘ und ‚interpretative competence‘“, d. h. die zweckadäquate Ausführung domänenspezifischer Aufgaben (WINTHER 2010, S. 200f.). Diese theoretisch begründete Aufteilung in die beiden Dimensionen ließ sich empirisch bestätigen (WINTHER/KLOTZ 2014, S. 21f.).

3 Entwicklung des Kompetenzniveaumodells

Zur Entwicklung von Kompetenzniveaumodellen existieren verschiedene Verfahren (HARSCH/HARTIG 2011, S. 10f.). Insbesondere bei unzureichendem Forschungsstand zum zu messenden Konstrukt haben heuristische Verfahren ihre Berechtigung. Im Fall des Rechnungswesens liegt aber mit dem wirtschaftsinstrumentellen Ansatz (PREISS 1999) ein fachdidaktisches Konzept vor, das es ermöglicht, Lern- und Arbeitsprozesse im Rechnungswesen zu beschreiben. Das gilt es nutzbar zu machen (ACHTENHAGEN 2012, S. 10; WEBER/ACHTENHAGEN 2014, S. 33). Der Gleichlauf von Lern- und Prüfungsprozessen ist für ein valides Assessment unabdingbar (PELLERINO 2010, S. 5).

Mit dem Beitrag von WINTHER und ACHTENHAGEN (2009, S. 543ff.) existiert für die kaufmännische Bildung bereits ein Kompetenzniveaumodell. Dieses beschreibt den Anspruch von beruflichen Anforderungssituationen über drei jeweils vierstufige Schwierigkeitsdimensionen (vgl. WINTHER 2010, S. 100ff.). Die drei Dimensionen *Funktionale Modellierung*, *Inhaltliche Komplexität* und *Art der kognitiven Beanspruchung* erwiesen sich auch in zahlreichen anderen Beiträgen als schwierigkeitsbestimmend (NICKOLAUS 2011, S. 337). Das Modell von WINTHER und ACHTENHAGEN (2009) soll mit dem wirtschaftsinstrumentellen Ansatz als fachdidaktische Grundlage verknüpft werden. Dazu wird mithilfe des wirtschaftsinstrumentellen Ansatzes der Problemlöseprozess im Rechnungswesen beschrieben, s. Abb. 1. Hieraus leiten wir dann drei jeweils vierstufige Schwierigkeitsdimensionen ab, die im Folgenden dargestellt werden.

Modellierung

Die erste Schwierigkeitsdimension ist die *Modellierung*. Diese stellt die Transformation einer Anforderungssituation in ein bearbeitungsfähiges Modell dar (AMELUNG/FUNKE 2013, S. 34; WINTHER 2010, S. 101). Dazu ist im ersten Schritt das zu lösende Problem zu identifizieren und einzugrenzen (BARIL u. a. 1998, S. 393f.; NORMAND/DUDLEY 2004, S. 64f.). Die weiteren Modellierungsschritte lassen sich mit dem wirtschaftsinstrumentellen Ansatz des Rechnungswesens beschreiben (TRAMM 2003, S. 7). Grundlage ist dabei ein allgemeines Unternehmensmodell. Die Interaktionen mit den Stakeholdern werden über Pfeile als Input- und Outputbeziehungen im Modell visualisiert. Der Wertschöpfungsprozess wird durch den Leistungsprozess, den Finanzierungsbereich und das Anlage- und Vorratsvermögen beschrieben. Hierbei führen Zugänge und Abgänge im Leistungssystem zu betragsgleichen Abgängen bzw. Zugängen im Finanzierungsbereich. Das Anlage- und Vorratsvermögen wiederum stellt ein Potential dar, das langfristig für den Leistungsprozess genutzt werden kann. Unter Rückgriff auf dieses allgemeine Unternehmensmodell ist es nun grundsätzlich möglich, sämtliche betrieblichen Prozesse abzubilden, d. h. zu modellieren. Die beschriebene Erfassung und Enkodierung ökonomischer Realität ist der erste Schritt in der Bearbeitung von Anforderungssituationen im Rechnungswesen (TÜRLING u. a. 2011, S. 398).

Die Schwierigkeit der Modellierung steigt mit zunehmender Komplexität des zugrunde liegenden Prozesses (MARZANO/KENDALL 2007, S. 11). Auf der anderen Seite kann die Schwierigkeit einer Anforderungssituation reduziert werden, indem Modellierungshilfen gegeben werden (PREISS 1999, S. 64; WINTHER/ACHTENHAGEN

2009, S. 543). Ebenso sinkt die Schwierigkeit, wenn die Aufgabe bereits häufig bearbeitet wurde (FUNKE 2003, S. 117, 178 ff.). Dieser inverse Zusammenhang zwischen dem Grad der Vertrautheit und der Schwierigkeit einer Anforderungssituation gilt für alle Schwierigkeitsdimensionen. Zusammenfassend lassen sich die nachfolgend beschriebenen vier Stufen formulieren.

Auf der niedrigsten Stufe ist keinerlei Modellierungsleistung zu erbringen. Das zu lösende Problem ist dabei bereits in seiner Input-Output-Beziehung (im allgemeinen Unternehmensmodell) dargestellt. Die zweite Schwierigkeitsstufe ist durch die Notwendigkeit der Identifikation des zugrunde liegenden Problems charakterisiert, wobei aber durch die Eingrenzung des Inhaltsbereichs eine Modellierungshilfe gegeben wird. Beispielsweise kann ein Beleg durch die Kennzeichnung als Eingangsrechnung unmittelbar dem Kreditorenbereich zugeordnet werden. Das zu lösende Problem ist damit eingegrenzt. Die dritte Stufe der Modellierung ist erreicht, wenn die zu bewältigende Anforderungssituation mehrere Inhaltsbereiche umfasst (PREISS 2005a, S. 73) oder mehr als ein Problem enthält. Die Modellierung ist ebenfalls der dritten Stufe zuzuordnen, wenn typische Schülerfehler (MINDNICH/BERGER/FRITSCH 2013, S. 68 f.; TÜRLING u. a. 2011), Lernschwierigkeiten (SCHILLER 2004; TRAMM/HINRICHS/LANGENHEIM 1996) oder Misconceptions (HAMMER 1996; SMITH/DI SESSA/ROSCHELLE 1994) Gegenstand der Anforderungssituation sind. Ein Beispiel hierfür wäre die Aufgabe, aus einer Gehaltsliste die Kosten für die Mitarbeiter zu ermitteln. Typischerweise wird hier der Überweisungsbetrag genannt. Die höchste und damit vierte Schwierigkeitsstufe ist erreicht, wenn mindestens zwei Probleme zu identifizieren sind und dabei mehrere Inhaltsbereiche betroffen sind. Ein Beispiel hierfür ist die Bearbeitung einer aktiven Rechnungsabgrenzung in einem (Modell-) Unternehmen, das diese sofort bei Rechnungseingang vornimmt. Wird dem Probanden nun eine Rechnung ohne weitere Angaben vorgelegt, muss er die Art des Belegs identifizieren (1. Problem) und diesem zusätzlich entnehmen, dass sich die Zahlung über zwei Geschäftsjahre erstreckt (2. Problem). Der Geschäftsfall betrifft dabei nicht nur die Kreditorenbuchhaltung (1. Inhaltsbereich), sondern auch den Jahresabschluss (2. Inhaltsbereich).

Informationsgewinnung

Die zweite Dimension des Niveaumodells ist die *Informationsgewinnung* (BARIL u. a. 1998, S. 394 f.; CRAWFORD/HELLIAR/MONK 2011, S. 119; SUNDEM/WILLIAMS 1992, S. 57). Sie beschreibt den Grad der Schwierigkeit, aus vorliegenden Belegen, durch Berechnungen oder heranzuziehende Quellen die zur Problemlösung notwendigen Informationen zu gewinnen. Die besondere Bedeutung dieser Dimension für das Rechnungswesen ergibt sich aus der Belegorientierung (PREISS 2005a, S. 74; WÜSTEMANN 2013, S. 21), die auch curricular gefordert ist (z. B. KMK 2002, S. 11). Belege, beispielsweise Eingangsrechnungen, dokumentieren betriebliche Prozesse (PREISS 1999, S. 155). Die hierin enthaltenen Informationen sind wiederum die Grundlage zur Erfüllung betrieblicher Aufgaben. In Abgrenzung zu einer allgemeinen literacy und numeracy bezieht sich die Informationsgewinnung auf konkret im Rechnungswesen verwendete Materialien und durchzuführende Berechnungen. Sie ist damit domänenspezifisch (COONEY 2005, S. 6).

Auf der niedrigsten Schwierigkeitsstufe der Informationsgewinnung sind sämtliche Informationen in gut strukturierter Form, beispielsweise in Tabellenform, vorgege-

ben. Trotz der geforderten Belegorientierung kann diese Stufe sinnvoll sein, um zu untersuchen, ob ein Problem auch dann nicht gelöst werden kann, wenn alle Informationen gut strukturiert vorliegen (DANIEL/EMBRETSON 2010, S. 350). Die zweite Stufe ist durch das Vorliegen eines Beleges, aus dem die Informationen gewonnen werden müssen, gekennzeichnet. Das schließt auch das Vornehmen einfacher Berechnungen (TÜRLING u. a. 2011, S. 399) zur Informationsgewinnung mit ein. Ein Beispiel wäre hier die Ermittlung des Skontobetrags aus einer Eingangsrechnung. Die Verknüpfung von Elementen bei der Informationsverarbeitung erhöht die Schwierigkeit (SWELLER 2006, S. 23). Die dritte Stufe ist demnach durch einen Rückgriff auf Informationen aus verschiedenen Quellen, beispielsweise zwei Belegen, charakterisiert. Diese sind zu kombinieren, um die zur Problemlösung nötigen Informationen zu gewinnen. Das Durchführen von komplexen Berechnungen, beispielsweise die Ermittlung von Verzugszinsen, wird ebenso der dritten Stufe zugeordnet. Die vierte Stufe der Informationsgewinnung begründet sich in der eigenständigen Auswahl und Verwendung von Quellen, die nicht der Aufgabenstellung entnommen werden können (FINLEY/WAYMIRE 2012, S. 36; JACKSON/DURKEE 2008, S. 85). Hierunter fällt insbesondere auch die eigenständige Nutzung von Gesetzestexten zur Problemlösung (CUNNINGHAM/ANDERSON 2005, S. 4).

Die Dimension der Informationsgewinnung ist dabei nur für die handlungsbaasierte Kompetenzdimension einschlägig. Bei Aufgaben der verstehensbasierten Kompetenzdimension müssen sämtliche Informationen aus dem Langzeitgedächtnis abgerufen werden (MARZANO/KENDALL 2007, S. 37). Eine Informationsgewinnung aus Belegen und Quellen ist nicht vorgesehen, weil nicht das Ausführen von Tätigkeiten, sondern der Nachweis des Verständnisses zentraler Konzepte des Rechnungswesens Gegenstand dieser Dimension ist.

Informationsverarbeitung

Zur dritten Dimension, der *Informationsverarbeitung*, liegen in Form von kognitiven Taxonomien umfangreiche Forschungsarbeiten vor (ANDERSON/KRATHWOHL 2001; BLOOM 1974; MARZANO/KENDALL 2007). Im Folgenden wird dabei systematisch zwischen verstehens- und handlungsbasierter Kompetenz unterschieden, weil sich diese hinsichtlich der Art der Informationsverarbeitung unterscheiden. Zur Beschreibung des Schwierigkeitsniveaus der verstehensbasierten Dimension verwendeten wir die vier Stufen der *New Taxonomy* von MARZANO und KENDALL (2007, S. 37 ff.). Die an Komplexität zunehmenden Stufen sind dabei: *Wiedergabe*, *Verstehen*, *Analyse* und *Anwendung* (MARZANO/KENDALL 2007, S. 35). Bei konstantem Grad der Vertrautheit entspricht das einem steigenden Schwierigkeitsgrad (MARZANO/KENDALL 2007, S. 11). Die Charakterisierung der Schwierigkeit der Informationsverarbeitung ist in der handlungsbasierten Dimension im Vergleich zur verstehensbasierten problematischer. Handlungen im Rechnungswesen, beispielsweise das Buchen, lassen sich nicht direkt den Definitionen der Taxonomiestufen zuordnen. In der handlungsbasierten Dimension wird daher zur Bildung von Schwierigkeitsstufen auf die der *New Taxonomy* zugrunde liegende Annahme Bezug genommen. Demnach steigt die Komplexität eines Prozesses mit der Anzahl der auszuführenden Schritte (MARZANO/KENDALL 2007, S. 10). Im Rechnungswesen kann hiermit ein höherer Schwierigkeitsgrad aufgrund einer steigenden Anzahl an involvierten Konten begründet werden. Diesen Überlegungen folgend, ergeben sich nachfolgend beschriebene vier Stufen.

Aufgaben, die lediglich Wissen abfragen, sind der untersten (verstehensbasierten) Stufe zugeordnet. Ein Beispiel wäre die Frage nach der Aufbewahrungsfrist von Rechnungen. Bei der handlungsbasierten Dimension fallen in die erste Kategorie einfache Kontierungen, bei denen jeweils ein Konto im Soll und Haben anzusprechen ist, z. B. bei Begleichung einer Rechnung. Die zweite Schwierigkeitsstufe bilden in der verstehensbasierten Dimension Aufgaben, bei denen ein Verständnis zur Lösung notwendig ist. Ein Beispiel ist die Frage, warum Abschreibungen zu bilden sind. In der handlungsbasierten Dimension sind hier Buchungen mit drei anzusprechenden Konten einschlägig, z. B. die Kontierung einer Kreditorenrechnung. Auf der dritten Stufe, der Analyse, ist in der verstehensbasierten Dimension ein durchzuführender Vergleich von linearer und degressiver Abschreibung einzuordnen. Kontierungen mit vier zu berücksichtigenden Konten sind in der handlungsbasierten Dimension der dritten Stufe zuzuordnen. Das ist beispielsweise bei Skontobuchungen der Fall. Für die vierte Stufe in der verstehensbasierten Dimension, der Anwendung, ist die Diskussion des Ansatzes von Drohverlustrückstellungen ein Beispiel. Der vierten Stufe der handlungsbasierten Dimension sind komplexe Buchungsstrukturen mit mehr als vier Konten, die auch mehrere Perioden betreffen können, zuzurechnen. Ein Beispiel ist die Kontierung einer Gehaltsliste, wobei vermögenswirksame Leistungen zu berücksichtigen sind. Die Entscheidung für eine Bewertungsmethode unter Beachtung der betrieblichen Ziele fällt ebenso in die vierte Stufe.

Zusammenfassend ist der dargestellte Problemlöseprozess in Abb. 1 dargestellt. Die Begriffe deklaratives und prozedurales Wissen (mental procedures) sowie Information werden dabei im Sinne von MARZANO UND KENDALL (2007, S. 23–30, 36 ff.) verwendet. Deutlich wird, dass die Informationen im Rahmen des Informationsgewinnungsprozesses in einem substitutiven Verhältnis zum deklarativen Wissen stehen können. Beispielsweise ist die Verjährungsfrist für ausstehende Rechnungen entweder bekannt oder sie muss alternativ aus einer zu identifizierenden Quelle gewonnen werden.

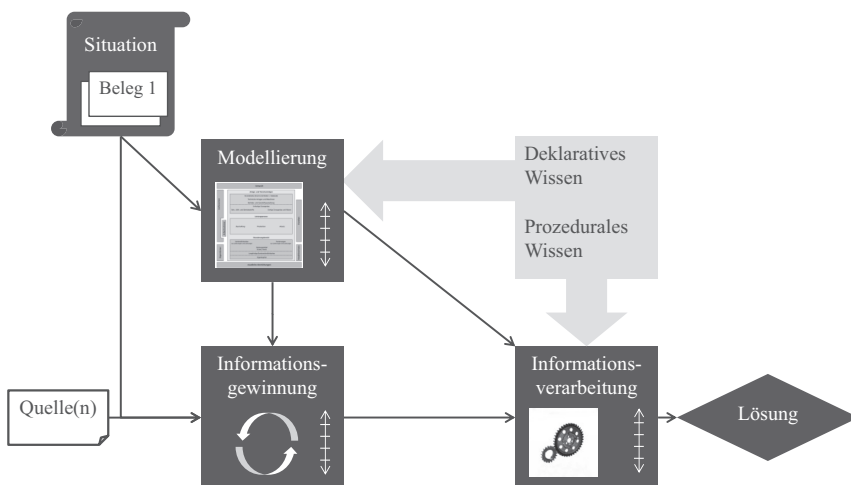


Abb. 1: Problemlöseprozess bei handlungsbasierten Anforderungssituationen

Auf Grundlage des beschriebenen Problemlöseprozesses lassen sich die in Abb. 2 festgehaltenen Schwierigkeitsstufen formulieren. Die Ausprägungen Wiedergabe, Verstehen, Analyse und Anwendung der *Informationsverarbeitung* beziehen sich dabei auf die verstehensbasierte Kompetenzdimension.

Zunahme ↑ Abnahmeschwierigkeit einer Anforderungssituation ↓	1. Modellierung	2. Informationsgewinnung	3. Informationsverarbeitung
	Mehr als ein Problem über mehrere Bereiche hinweg	Eigenständige Informationssuche und Auswertung	Anwendung/ Buchungssatz mit mehr als vier Konten
	Ein Problem über mehrere Bereiche hinweg; alternativ: mehr als ein Problem	Mehrere Quellen; alternativ: eine Quelle mit komplexen Berechnungen	Analyse/ Buchungssatz mit vier Konten
	Ein Problem im abgegrenzten Bereich	Eine Quelle inkl. einfacher Berechnungen	Verstehen/ Buchungssatz mit drei Konten
	Problem vorgegeben	Keine Informationsgewinnung nötig	Wiedergabe/ Einfacher Buchungssatz

Abb. 2: Schwierigkeitsstufen des externen Rechnungswesens

4 Validierung des Kompetenzniveaumodells

4.1 Hypothesen

Zur Validierung des Kompetenzniveaumodells werden im Folgenden aus den Eigenschaften des entwickelten Kompetenzniveaumodells Hypothesen abgeleitet, die anschließend getestet werden.

Wie dargestellt, unterscheiden sich Aufgaben der handlungsbasierten und verstehensbasierten Kompetenz grundlegend im Problemlöseprozess. Folglich sollte ein zweidimensionales Kompetenzmodell mit den Dimensionen verstehensbasiert und handlungsbasiert zu einer besseren Anpassung an die Daten als ein eindimensionales Modell führen:

- H1 Im externen Rechnungswesen führt ein zweidimensionales Kompetenzmodell mit einer Aufspaltung in verstehens- und handlungsbasierte Kompetenz zu einer höheren Anpassungsgüte als eine eindimensionale Lösung.

Die Vertrautheit der Probanden mit der Anforderungssituation ist die grundlegende Determinante der Aufgabenschwierigkeit. Sie ist für jeden Schritt im Problemlöseprozess relevant. Der Problemlöseprozess weist dabei eine hierarchische Struktur auf. Der erste Schritt ist die Modellierung. Auf dieser Grundlage sind die notwendigen Informationen zu gewinnen und abschließend zu verarbeiten. Hieraus ergeben sich folgende Hypothesen für das externe Rechnungswesen:

- H2a Ein steigender Grad an Vertrautheit mit einer Anforderungssituation beeinflusst deren Schwierigkeit negativ.

- H2b Ein zunehmender Anspruch der Modellierung erhöht die Schwierigkeit einer Anforderungssituation.
- H2c Über die Modellierung hinaus erhöht ein zunehmender Anspruch der Informationsgewinnung die Schwierigkeit einer Anforderungssituation.
- H2d Über die Modellierung und die Informationsgewinnung hinaus erhöht ein zunehmender Anspruch der Informationsverarbeitung die Schwierigkeit einer Anforderungssituation.

4.2 Stichprobe

Zur Überprüfung der Hypothesen verwendeten wir die Antworten von Industriekaufleuten in der Berufsabschlussprüfung „Kaufmännische Steuerung und Kontrolle“ (KSK). Diese stammen von der IHK München und Oberbayern.¹ Der vorliegende Datensatz umfasst fünf Prüfungen (Winter 2011/12 bis Winter 2013/14), s. Tab. 1. Die geschlossenen Aufgaben sind in Form von Single-Choice-Aufgaben gestellt, bei denen aus fünf Aussagen die zutreffende zu wählen ist. Bei den halboffenen Aufgaben ist entweder ein Rechenergebnis anzugeben oder eine Kontierung anhand vorgegebener Konten vorzunehmen. Die Antworten können allesamt maschinell ausgewertet werden, was ein Höchstmaß an Objektivität zur Folge hat. Die Antworten wurden durchgehend dichotom (richtig/falsch) kodiert. Eine polytome Kodierung wäre nur bei sehr wenigen Aufgaben überhaupt möglich. Anschließend entfernten wir domänenfremde Aufgaben, weil das entwickelte Modell und die hieraus abgeleiteten Hypothesen nur für das externe Rechnungswesen Gültigkeit beanspruchen. Im Wesentlichen handelte es sich dabei um Aufgaben des internen Rechnungswesens, des Datenschutzes und der Datensicherheit.

Tab. 1: Beschreibung der Prüfungsteilnehmer und Aufgaben des Datensatzes

	Prüfungsteilnehmer				Dimension der Aufgaben	
	Anzahl	davon weiblich		Durchschnitts- alter am Prüfungstag	verstehens- basiert	handlungs- basiert
Winter 2011/12	464	311	(67,0%)	21,0	4	12
Sommer 2012	403	262	(65,0%)	22,6	4	13
Winter 2012/13	442	297	(67,2%)	21,0	3	12
Sommer 2013	424	259	(61,1%)	22,0	4	15
Winter 2013/14	462	299	(64,7%)	21,0	3	12

Anmerkung: Jede Prüfung im Fach KSK umfasst 32 Aufgaben. Dabei sind im Durchschnitt 15,4 dem externen Rechnungswesen zuzuordnen.

1 Wir bedanken uns bei Herrn Dr. JOSEF AMANN von der IHK München und Oberbayern für die Zurverfügungstellung der Daten.

Der vorliegende Datensatz ist aus mehreren Gründen besonders dazu geeignet, das Kompetenzmodell empirisch zu prüfen: Wegen der hohen Bedeutung der Berufsabschlussprüfung für die Teilnehmer kann von einer bestmöglichen Beantwortung der Fragen ausgegangen werden. Eine Verzerrung des Antwortverhaltens aufgrund von geringer Motivation der Probanden ist damit zu vernachlässigen (WISE/DEMARS 2005, S. 8f.). Die lokale stochastische Unabhängigkeit der Antworten kann auf Personenebene als gewahrt gelten (STROBL 2012, S. 19), weil ein Abschreiben der Kandidaten voneinander in den Prüfungen konsequent unterbunden wird. Die Testteilnehmer stehen am Ende der Ausbildung, deshalb sollten alle Probanden die in Lehrplan und Ausbildungsordnung fixierten Kompetenzen erworben haben.

Darüber hinaus kann der Datensatz als repräsentativ gelten. Es sind Ausbildungsbetriebe aus insgesamt 183 Branchen vertreten. Die Unternehmensgröße, gemessen an der Zahl der Beschäftigten, streut breit, s. Tab. 2. Ferner handelt es sich beim Ausbildungsberuf des Industriekaufmanns/-frau um einen kaufmännischen Kern- bzw. Leitberuf (ACHTENHAGEN 2001, S. 11; BRÖTZ/EHRKE 2006, S. 6; WINTHER 2010, S. 197). Dadurch lassen sich die Ergebnisse potentiell auch auf andere kaufmännische Ausbildungsberufe übertragen.

Tab. 2: Unternehmensgröße der Ausbildungsbetriebe der Prüfungsteilnehmer im Datensatz

Prüfung KSK	Zahl der Beschäftigten						keine Angabe
	1 bis 9	10 bis 49	50 bis 199	200 bis 999	1.000 bis 10.000	mehr als 10.000	
Winter 2011/12	49	36	62	166	75	65	11
Sommer 2012	75	32	80	119	30	25	42
Winter 2012/13	58	30	64	147	71	69	3
Sommer 2013	67	32	87	115	26	31	66
Winter 2013/14	50	36	84	164	57	64	7
Summe	299	166	377	711	259	254	129

Mit durchschnittlich 432 Teilnehmern je Prüfung ist die Stichprobe zudem für den Einsatz von IRT-Modellen geeignet (BAETHGE u. a. 2006, S. 66).

4.3 Methode

Überprüfung der Hypothesen

Hypothese H1 kann mithilfe eines χ^2 -Tests überprüft werden. Dazu wird der -2 Log Likelihood-Wert (-2LL-Wert) eines zweidimensionalen Modells mit jenem eines eindimensionalen Modells verglichen (WU/ADAMS/WILSON 2007, S. 106).

Die Hypothesen H2a bis H2d können mithilfe einer linearen Regression überprüft werden. Aufgrund der Struktur des Problemlöseprozesses ist ein hierarchisches Regressionsmodell angemessen (EMBRETSON/DANIEL 2008, S. 337f.; EMBRETSON/GORIN 2001, S. 356f.). Dabei bilden die Aufgabenschwierigkeit die abhängige und die schwierigkeitsbestimmenden Merkmale die unabhängige Variable. Die Aufgabenschwierigkeit wird mit Methoden der IRT geschätzt. Zur Operationalisierung

der unabhängigen Variablen dienen die in Abb. 2 festgehaltenen schwierigkeitsbestimmenden Merkmale als deduktives Kategoriensystem. Damit werden die IHK-Aufgaben kodiert (EMBRETSON/GORIN 2001, S. 351). Jede Aufgabe wird dabei in den Dimensionen *Modellierung*, *Informationsgewinnung* und *Informationsverarbeitung* jeweils einer der vier möglichen Stufen zugeordnet. Hieraus werden dann für jede Dimension drei Dummy-Variablen gebildet, weil es sich bei der Einstufung um ein ordinales Maß handelt. Die Dimensionen werden schrittweise, dem Problemlöseprozess folgend, in das Regressionsmodell aufgenommen. Abschließend kann die Modellgüte mithilfe des Determinationskoeffizienten R^2 eingeschätzt werden.

Das Vorgehen, zuerst die Schwierigkeit der Aufgaben mit Methoden der IRT zu schätzen und dann diese über die schwierigkeitsbestimmenden Merkmale zu erklären, ist ein konservatives Vorgehen. Statistisch signifikante Ergebnisse zu erhalten ist dabei im Vergleich zur Verwendung eines linear logistischen Testmodells (LLTM) schwieriger (EMBRETSON/DANIEL 2008, S. 342). Im LLTM gilt folgender Zusammenhang zwischen der Wahrscheinlichkeit einer richtigen Lösung ($Y = 1$) einer Aufgabe i durch Person v und der Personenfähigkeit θ_v sowie den Eigenschaften j einer Aufgabe X_{ij} (ROST 2004, S. 254):

$$p(Y_{vi} = 1) = \frac{\exp(\theta_v + \sum_{j=1}^n \alpha + \beta_j X_{ij})}{1 + \exp(\theta_v + \sum_{j=1}^n \alpha + \beta_j X_{ij})}$$

Auf diese Weise kann der Einfluss der Aufgabeneigenschaften β_j auf die Lösungswahrscheinlichkeit einer Aufgabe geschätzt werden. Bei α handelt es sich um den Achsenabschnitt. Dummy-Variablen für die fünf Prüfungen sind nicht nötig, weil Niveauunterschiede zwischen den Prüfungen über die Personenfähigkeit θ_v erfasst werden.

Operationalisierung der unabhängigen Variablen

Im ersten Schritt kodierten wir unabhängig voneinander den Schwierigkeitsgrad der *Modellierung*, *Informationsgewinnung* und *Informationsverarbeitung* jeder IHK-Aufgabe. Die vierte Schwierigkeitsstufe wurde dabei in keiner der drei Dimensionen erreicht, s. Tab. 3. Aufgrund der Ordinalskalierung der Schwierigkeitsstufen bewerteten wir die Reliabilität unserer Kodierung mithilfe von Spearman's Rho (WIRTZ/CASPAR 2002, S. 133 ff.). Mit $Rho = 0,797$ kann diese als gegeben erachtet werden (FRÜH 2007, S. 193). Hieran anschließend diskutierten wir die ex ante abweichenden Einschätzungen, um zu einem Konsens zu gelangen (BYRON/LAURENCE 2015, S. 304).

Als Proxy für den Grad der Vertrautheit diene die Häufigkeit des Vorkommens eines Aufgabentyps in vorhergehenden Prüfungen. Dazu werteten wir für jede Prüfung die Aufgaben der vorhergehenden zehn Prüfungen inhaltlich aus. Insgesamt identifizierten wir in den 15 analysierten Prüfungen 105 Aufgabentypen. Im Mittel wurde jede Aufgabe in 2,29 der vorhergehenden zehn Prüfungen bereits in dieser Form gestellt ($SD = 2,04$). Die auf diese Weise ermittelte Häufigkeit logarithmierten wir, weil uns ein degressiver Zusammenhang von Häufigkeit und Grad der Vertrautheit plausibel erscheint.

Tab. 3: Ausprägung der schwierigkeitsbestimmenden Merkmale in den Aufgaben

Schwierigkeit	Häufigkeit verstehensbasiert			Häufigkeit handlungsbasiert		
	Stufe 1	Stufe 2	Stufe 3	Stufe 1	Stufe 2	Stufe 3
Modellierung	n/a	n/a	n/a	17	15	6
Informationsgewinnung	n/a	n/a	n/a	23	12	3
Informationsverarbeitung	1	11	4	15	17	6

Anmerkungen: Dargestellt werden nur die tatsächlich zur Auswertung herangezogenen 16 bzw. 38 Aufgaben, s. Tab. 4 und 5. Die vierte Stufe wurde in keiner Aufgabe erreicht. Verstehensbasierte Aufgaben weisen keine Modellierung und Informationsgewinnung auf.

Messung der abhängigen Variable

Zur Schätzung der Schwierigkeit der Aufgaben als abhängige Variable diente das Rasch-Modell. Der entscheidende Grund hierfür ist dessen konstante Itemtrennschärfe und die damit verbundene spezifische Objektivität (BÜHNER 2011, S. 505, 513–514 ff.; ROST 2004, S. 40, 122 ff.). D. h. bei Gültigkeit des Rasch-Modells ist die Differenz der Schwierigkeit zweier Items unabhängig von der Stichprobe, mit der diese Parameter geschätzt wurden. Diese Eigenschaft ist im vorliegenden Kontext besonders bedeutsam. Die Schwierigkeit der Items wird je Prüfung geschätzt. Ankeritems existieren nicht. Auf einer gemeinsamen Skala können die Items nur unter der Annahme der spezifischen Objektivität verortet werden. Die Niveauunterschiede, die sich zwischen Prüfungen aus dem Fehlen eines natürlichen Nullpunkts der Logit-Skala ergeben, können dabei im vorliegenden Untersuchungsdesign mit Dummy-Variablen für die fünf Prüfungen erfasst werden.

Voraussetzung für die Anwendung des Rasch-Modells ist die Einhaltung der Modellprämissen. Die Gültigkeit des Rasch-Modells wiederum ist die notwendige Voraussetzung für die Verwendung des LLTM (POINTSTINGL 2009, S. 125). Der Prozess zur Sicherstellung der Rasch-Konformität der Aufgaben wird im Folgenden dargestellt.

Um die lokale stochastische Unabhängigkeit der Aufgaben zu gewährleisten, prüften wir qualitativ für jede Aufgabe, ob diese unabhängig von anderen Aufgaben der Prüfung lösbar ist. Das war bei allen in der Stichprobe berücksichtigten Aufgaben der Fall. Zur Wahrung der Unidimensionalität der Aufgaben teilten wir, in Übereinstimmung mit H1, die Aufgaben in verstehens- und handlungsbasiert auf. Aufgaben mit einem Missfit identifizierten wir mit dem gewichteten Mean-Square-Index (MNSQ) (WILSON 2005, S. 127 ff.). Ausgeschlossen wurden dann alle Aufgaben, deren gewichteter MNSQ entweder außerhalb des Intervalls $[0,75; 1,25]$ liegt oder die eine statistisch signifikante Abweichung von der Trennschärfe des Rasch-Modells aufweisen. Die Abweichung galt dabei als signifikant, wenn der t-Wert betragsmäßig größer als 1,96 war (WINTHER 2010, S. 154). Anschließend führten wir mehrere globale Modellgeltungstests durch. Das Vorgehen orientiert sich an BÜHNER (2011, S. 546 f.). Zur Umsetzung kam die Software R (Package „eRm“ (MAIR/HATZINGER 2007) und „mixRasch“ (DALLAS/WILLSE 2014)) zum Einsatz. Für die Aufgaben der handlungsbasierten Dimension wurde zunächst ein Andersen-Likelihood-Ratio-Test (LRT) mit Mittelwert als Teilungskriterium durchgeführt (KOLLER/ALEXANDROWICZ/HATZINGER 2012, S. 181 f.). Der LRT wurde in den Prüfungen 2011/12 und 2012/13 signifikant ($\alpha =$

5 %). Mithilfe eines grafischen Verfahrens (STROBL 2012, S. 40f.) identifizierten wir jene Aufgaben mit statistisch signifikant abweichendem Schwierigkeitsparameter zwischen den beiden Gruppen und schlossen diese aus. Folglich wird der LRT nun in keiner Prüfung mehr statistisch signifikant. Mit einem Martin-Löf-Test kontrollierten wir die Itemhomogenität der Aufgaben. Teilungskriterium war dabei ebenfalls der Mittelwert. Nach Ausschluss einer Aufgabe aus Prüfung 2012 können nun alle Aufgaben als homogen gelten. Ein im Vergleich zum LRT strengerer Test zur Überprüfung von Personenhomogenität ist das Mixed-Rasch-Modell (BÜHNER 2011, S. 546). Eine Einklassenlösung weist dabei mit Ausnahme von Prüfung 2013/14 stets das geringste BIC und CAIC auf. Auf Grundlage des AIC sind stets Mehrklassenlösungen vorteilhaft. Den strengen Test auf Personenhomogenität bestehen die in der Stichprobe enthaltenen Aufgaben damit nicht. Tab. 4 fasst die Ergebnisse der Modellgeltungstests für die handlungsbasierte Dimension zusammen und zeigt die in die Stichprobe aufgenommenen Aufgaben. Insgesamt wurden 26 der in den fünf Prüfungen enthaltenen 64 handlungsbasierten Aufgaben ausgeschlossen.

Im Bereich der verstehensbasierten Dimension war aufgrund der geringen Anzahl an Aufgaben weder ein LRT noch ein Martin-Löf-Test möglich. Aus diesem Grund setzten wir trotz Bedenken (HEENE u. a. 2011, S. 290f.) ein Bootstrapping-Verfahren als globalen Modellgeltungstest ein (DAVIER 1997; ROST 2004, S. 336 ff.). Dabei nutzten wir die Software MATLAB in Verbindung mit der IRTm-Toolbox von BRAEKEN und TUERLINCKX (2009). Unter Verwendung der in den Prüfungen geschätzten Personen- und Schwierigkeitsparameter wurden $n = 1.000$ Rasch-konforme Datensätze erzeugt. Für jeden Datensatz wurde der χ^2 -Wert berechnet. Eine statistisch signifikante Abweichung vom Rasch-Modell liegt dann vor, wenn der tatsächliche χ^2 -Wert außerhalb des 95%-Konfidenzintervalls der 1.000 auf Grundlage der Simulation berechneten χ^2 -Werte liegt. Im Fall einer statistisch signifikanten Abweichung vom Rasch-Modell identifizierten wir anhand des gewichteten MNSQ weitere Aufgaben mit vergleichsweise schlechter Passung und schlossen diese aus. Das war lediglich bei einer Aufgabe in der Prüfung Winter 2012/13 nötig. Im Ergebnis tritt nun in keiner der Prüfungen eine statistisch signifikante Abweichung vom Rasch-Modell auf, s. Tab. 5. Für die handlungsbasierte Dimension s. Tab. 4.

Tab. 4: Modellgeltungstests des Rasch-Modells und in die Stichprobe aufgenommene Aufgaben: handlungsbasiert

Prüfung	Globale Modellgeltungstests			Aufgabe					
	LRT	Martin-Löf-Test	Bootstrapping (n = 1.000)	Nr.	Schwierigkeit	Standardfehler	gewichteter MNSQ	t-Wert	Trennschärfe (KTT)
Winter 2011/12	$\Delta-2LL = 13,09$ df = 7 Sig. = 0,07	$\Delta-2LL = 11,83$ df = 15 Sig. = 0,69	χ^2 -Wert = 239,54 95% CI = 334,13 Sig. = 0,27	1.2	-0,537	0,080	1,03	0,7	,58
				1.4	-3,408	0,132	0,99	-0,0	,42
				1.8	-1,784	0,088	1,02	-0,2	,57
				2.1	-2,069	0,092	1,04	0,5	,50
				2.2	-2,914	0,114	0,92	-0,7	,55
				2.3	-1,904	0,090	1,06	0,8	,51
				2.4	-1,529	0,085	0,96	-0,6	,64
				2.5	-3,326	0,129	0,96	-0,2	,46

Prüfung	Globale Modellgeltungstests			Aufgabe					
	LRT	Martin-Löf-Test	Bootstrapping (n = 1.000)	Nr.	Schwierigkeit	Standardfehler	gewichteter MNSQ	t-Wert	Trennschärfe (KTT)
Sommer 2012	Δ -2LL = 9,44 df = 6 Sig. = 0,15	Δ -2LL = 15,75 df = 11 Sig. = 0,15	χ^2 -Wert = 118,82 95% CI = 186,98 Sig. = 0,25	1.1	-4,509	0,285	1,05	0,3	,26
				1.3	-3,842	0,226	0,93	-0,4	,47
				1.4	-0,877	0,140	1,02	0,3	,68
				2.1	-3,146	0,185	0,92	-0,6	,55
				2.7	-2,601	0,163	0,89	-1,0	,65
				2.8	-2,221	0,152	1,10	1,1	,56
				4.2	-1,606	0,142	1,06	0,8	,62
Winter 2012/13	Δ -2LL = 6,47 df = 7 Sig. = 0,49	Δ -2LL = 13,32 df = 19 Sig. = 0,82	χ^2 -Wert = 426,29 95% CI = 661,47 Sig. = 0,55	1.2	-0,940	0,113	1,04	0,7	,57
				1.3	-2,302	0,142	1,04	0,4	,48
				1.6	-2,827	0,164	0,93	-0,6	,51
				2.5	-3,253	0,188	0,92	-0,5	,49
				2.6	-1,556	0,121	1,01	0,1	,57
				2.7	-1,524	0,121	1,04	0,5	,53
				2.8	-2,990	0,172	0,98	-0,1	,44
Sommer 2013	Δ -2LL = 9,27 df = 8 Sig. = 0,32	Δ -2LL = 11,05 df = 19 Sig. = 0,92	χ^2 -Wert = 604,20 95% CI = 674,32 Sig. = 0,09	4.1	-1,772	0,126	1,06	0,8	,49
				4.2	-2,151	0,136	1,07	0,8	,44
				1.2	-0,059	0,107	1,00	-0,0	,54
				1.3	-1,345	0,117	1,07	1,1	,44
				1.5	-1,300	0,116	1,00	0,0	,54
				1.7	0,937	0,117	0,96	-0,8	,55
				2.1	-1,631	0,123	1,07	1,0	,42
Winter 2013/14	Δ -2LL = 4,58 df = 3 Sig. = 0,21	Δ -2LL = 7,38 df = 5 Sig. = 0,19	χ^2 -Wert = 28,88 95% CI = 37,71 Sig. = 0,17	2.2	-2,383	0,148	1,00	0,0	,43
				2.3	-2,921	0,176	0,96	-0,2	,40
				2.5	-0,365	0,107	1,01	0,1	,55
				4.4	-0,924	0,111	1,00	0,0	,56
				1.3	-2,882	0,158	0,97	-0,2	,45
				2.1	-2,698	0,150	0,98	-0,2	,44
				2.3	-1,430	0,112	1,01	0,1	,56
4.1	-0,212	0,105	1,03	0,7	,62				
4.8	0,535	0,114	0,98	-0,4	,63				
Anzahl:				38					

Anmerkungen: CI = Konfidenzintervall. Trennschärfe KTT = Punktbiseriale Korrelation. In Prüfung 2012/13 und 2013/14 musste die Aufgabe 2.5 bzw. 2.3 aufgrund eines ungeeigneten Antwortformats vom LRT ausgeschlossen werden (BÜHNER 2011, S. 532). Die Schwierigkeitsparameter wurden mit der marginalen Maximum-Likelihood Methode geschätzt. Eine bedingte Maximum-Likelihood Schätzung führt zu identischen Ergebnissen. Aufgrund des fehlenden natürlichen Nullpunkts der Logit-Skala wurde zur Normierung die Summe der Fähigkeitsparameter je Prüfung auf Null gesetzt. Damit ist unmittelbar ersichtlich, ob eine Aufgabe für den durchschnittlichen Probanden unter psychometrischen Gesichtspunkten zu leicht ist. Der Logit-Wert der Schwierigkeit ist in diesem Fall negativ.

Tab. 5: Modellgeltungstests des Rasch-Modells und in die Stichprobe aufgenommene Aufgaben: verstehensbasiert

Prüfung	Modellfit (Bootstrapping, n = 1000)			Aufgabe					
	χ^2 -Wert	95 % CI	Sig.	Nr.	Schwierigkeit	Standardfehler	gewichteter MNSQ	t-Wert	Trennschärfe (KTT)
Winter 2011/12	6,43	16,84	0,71	1.5	-0,454	0,086	1,00	0,0	,52
				1.9	-0,934	0,074	1,00	0,0	,68
				4.4	-1,687	0,074	0,98	-0,3	,68
				4.5	-1,723	0,089	1,04	0,4	,46
Sommer 2012	3,40	17,61	0,96	1.8	-1,869	0,110	0,99	-0,2	,66
				1.9	-0,805	0,148	1,01	0,1	,42
				4.5	-2,032	0,110	1,00	0,0	,63
				4.6	-2,245	0,118	1,01	0,2	,54
Winter 2012/13	< 0,00	2,91	0,99	1.7	-1,098	0,103	0,98	-0,2	,76
				4.7	-2,617	0,103	0,98	-0,2	,72
Sommer 2013	1,74	7,78	0,65	1.9	-1,126	0,107	1,01	0,2	,60
				4.2	-1,660	0,110	0,98	-0,2	,63
				4.9	-2,204	0,102	1,00	0,0	,72
Winter 2013/14	6,59	8,55	0,12	4.2	-1,046	0,089	1,02	0,5	,65
				4.3	-1,237	0,089	1,00	0,0	,65
				4.6	-2,334	0,100	0,99	-0,1	,59
Anzahl:				16					

Anmerkungen: CI = Konfidenzintervall. Trennschärfe KTT = Punktbiseriale Korrelation. Analog zur handlungsbasierten Dimension wurden die Schwierigkeitsparameter mit der marginalen Maximum-Likelihood Methode geschätzt. Eine bedingte Maximum-Likelihood Schätzung führt zu identischen Ergebnissen. Die Personenfähigkeit je Prüfung ist auf Null normiert.

4.4 Ergebnisse

Dimensionalität

Um H1 zu prüfen, schätzten wir für alle fünf Prüfungen sowohl ein eindimensionales als auch ein zweidimensionales Rasch-Modell. Das eindimensionale Rasch-Modell unterscheidet nicht zwischen verstehens- und handlungsbasierter Dimension. Die zweidimensionale Lösung führte stets zur besseren Anpassung. Bei zwei der fünf Prüfungen ist der Unterschied der 2LL-Werte auf dem 1%-Niveau und bei zwei weiteren auf dem 5%-Niveau signifikant. Bei Prüfung 2012/13 ist mit $p = 13\%$ zumindest eine leichte Tendenz erkennbar, s. Tab. 6. Ein Grund für die fehlende statistische Signifikanz könnte bei dieser Prüfung die geringe Anzahl an verstehensbasierten Aufgaben sein; Es sind nur zwei enthalten. Zusammenfassend werten wir die Ergebnisse als Evidenz für die Notwendigkeit der Trennung der beiden Kompetenzdimensionen. H1 wird daher aufrechterhalten.

Tab. 6: Test der Dimensionalität von Rechnungswesenkompetenz

Prüfung	Eindimensionales Modell		Zweidimensionales Modell		handlungs-basiert	verstehens-basiert	Unterschied der beiden Dimensionen	
	Deviance (-2LL)	df	Deviance (-2LL)	df	EAP/PV Reliabilität	EAP/PV Reliabilität	Korrelation	Sig. ^a
Winter 2011/12	4.684,60	13	4.673,37	15	,621	,561	,870	< ,01
Sommer 2012	3.540,14	12	3.513,27	14	,591	,506	,850	< ,01
Winter 2012/13	3.872,08	12	3.868,04	14	,587	,478	,865	,133
Sommer 2013	5.226,46	13	5.220,04	15	,639	,514	,792	,040
Winter 2013/14	3.826,17	9	3.819,56	11	,468	,384	,760	,037

a. Wahrscheinlichkeit, dass das eindimensionale Modell eine genauso gute Anpassung an die Daten aufweist als das zweidimensionale Modell (p-Wert).

Anmerkung: Mit Ausnahme der p-Werte wurden alle Parameter mit der Software ACER ConQuest ermittelt.

Handlungsbasierte Dimension

Um H2 für die handlungsbasierte Dimension zu prüfen, berücksichtigten wir im ersten Schritt mögliche systematische Fähigkeits- und Schwierigkeitsunterschiede der Teilnehmer bzw. Aufgaben in den fünf Prüfungen über Dummy-Variablen. Diese haben insgesamt keinen statistisch signifikanten Einfluss auf die Schwierigkeit der Aufgaben, s. Tab. 7: Niveauunterschiede der Prüfungen. Als nächstes nahmen wir den Grad der Vertrautheit in die Regression auf, um H2a zu überprüfen. Der Grad der Vertrautheit hat dabei keinen statistisch signifikanten Einfluss auf die Aufgabenschwierigkeit, s. Tab. 7: Vertrautheit. Hypothese H2a ist damit zu verwerfen. Die Hypothesen H2b, H2c und H2d prüften wir durch die sukzessive Aufnahme der drei Schwierigkeitsdimensionen in die Regression. Jede Dimension liefert dabei einen statistisch signifikanten Beitrag zur Erklärung der Schwierigkeit, der über die zuvor berücksichtigten Faktoren hinausgeht, s. Tab. 7. Alle schwierigkeitsbestimmenden Merkmale, jeweils zwei Dummy-Variablen für die Stufen 2 und 3, haben einen statistisch signifikant positiven Einfluss auf die Schwierigkeit einer Aufgabe, s. Tab. 8. Dabei führt in allen drei Dimensionen eine höhere Stufe auch zu einer höheren prognostizierten Schwierigkeit. Die Hypothesen H2b, H2c und H2d können somit als bestätigt gelten.

Insgesamt ist die Prognosegüte des Modells mit einem adjustierten R^2 von 0,746 als gut einzuschätzen (HAIR u. a. 2014, S. 174 ff.).

Tab. 7: Lineare hierarchische Regression der Aufgabenschwierigkeit auf die schwierigkeitsbestimmenden Merkmale: handlungsbasiert

Modell	R	R ²	Adj. R ²	Änderung in R ²	Änderung in F	df1	df2	Sig. Änderung in F
1. Niveauunterschiede der Prüfungen	,486	,236	,143	,236	2,546	4	33	,058
2. Vertrautheit	,486	,236	,117	,001	0,025	1	32	,876
3. Modellierung	,658	,433	,301	,197	5,208	2	30	,011
4. Informationsgewinnung	,796	,633	,515	,200	7,618	2	28	,002
5. Informationsverarbeitung	,906	,821	,746	,188	13,687	2	26	,000

Anmerkung: Das Modell wird, dem Problemlöseprozess folgend, sukzessive erweitert. D. h. das Modell *Informationsverarbeitung* inkludiert alle anderen Modelle.

Tab. 8: Ergebnis der linearen hierarchischen Regression der Aufgabenschwierigkeit auf die schwierigkeitsbestimmenden Merkmale: handlungsbasiert

Variable	B	Standardfehler	Beta	Sig. (2-seitig)
Konstante	-4,182	,777		,000
Dummy_Prüfung_2013	0,575	,562	,205	,318
Dummy_Prüfung_2012/13	-0,096	,491	-,034	,832
Dummy_Prüfung_2012	-0,950	,567	-,309	,103
Dummy_Prüfung_2011/12	-0,301	,505	-,103	,533
Vertrautheit	0,341	,219	,178	,133
Modellierung_2	1,186	,344	,487	,007
Modellierung_3	2,453	,524	,751	,001
Informationsgewinnung_2	1,299	,453	,507	,015
Informationsgewinnung_3	2,241	,484	,507	,001
Informationsverarbeitung_2	0,731	,319	,305	,033
Informationsverarbeitung_3	1,821	,529	,558	,004

Anmerkungen: Zur Berechnung robuster Standardfehler ohne Verteilungsannahme wurde das Bootstrapping Verfahren (n = 5.000, BCa-Methode) verwendet (Breusch-Pagan-Test, p = ,153). Kollinearität stellt kein Problem dar. Alle VIF liegen deutlich unter dem kritischen Wert von 5 (HAIR u. a. 2014, S. 124 ff.).

Verstehensbasierte Dimension

Bei der verstehensbasierten Kompetenzdimension konnten wir weder einen statistisch signifikanten Einfluss des Grades der Vertrautheit noch der Informationsverarbeitung, operationalisiert über die Taxonomie von MARZANO und KENDALL (2007, S. 10 f.), feststellen, s. Tab. 9 und 10. Die Hypothesen H2a und H2d sind damit für die verstehensbasierte Dimension zu verwerfen.

Tab. 9: Lineare hierarchische Regression der Aufgabenschwierigkeit auf die schwierigkeitsbestimmenden Merkmale: verstehensbasiert

Modell	R	R ²	Adj. R ²	Änderung in R ²	Änderung in F	df1	df2	Sig. Änderung in F
1. Niveauunterschiede der Prüfungen	,529	,279	,017	,279	1,066	4	11	,419
2. Vertrautheit	,600	,360	,040	,081	1,258	1	10	,288
3. Informationsverarbeitung	,745	,554	,164	,195	1,746	2	8	,235

Tab. 10: Ergebnis der linearen hierarchischen Regression der Aufgabenschwierigkeit auf die schwierigkeitsbestimmenden Merkmale: verstehensbasiert

Variable	B	Standardfehler	Beta	Sig. (2-seitig)
Konstante	-0,934	,575		,143
Dummy_Prüfung_2013	-0,806	,577	-,573	,262
Dummy_Prüfung_2012/13	-0,936	,816	-,666	,490
Dummy_Prüfung_2012	-1,281	,665	-,696	,264
Dummy_Prüfung_2011/12	-0,809	,697	-,519	,421
Vertrautheit	0,387	,764	,267	,639
Informationsverarbeitung_2 (Verstehen)	-0,270	,533	-,206	,773
Informationsverarbeitung_3 (Analyse)	0,414	,826	,294	,701

Anmerkungen: Zur Berechnung robuster Standardfehler ohne Verteilungsannahme wurde das Bootstrapping Verfahren (n = 5.000, BCa-Methode) verwendet (Breusch-Pagan-Test, p = ,198). Kollinearität stellt ein Problem dar (VIF Informationsverarbeitung_2 = 5,393; VIF Informationsverarbeitung_3 = 6,205).

Zusammenfassend ist die Gültigkeit des Kompetenzniveaumodells für die verstehensbasierte Dimension damit abzulehnen.

4.5 Robustheit der Ergebnisse

Anpassungsgüte des LLTM

Die Robustheit der Ergebnisse kontrollierten wir mit dem LLTM. Um den Fit des LLTM beurteilen zu können, fungierte das Rasch-Modell als Referenzmodell. Zunächst betrachteten wir dabei die handlungsbasierte Kompetenzdimension. Im Vergleich zum Rasch-Modell ist die Anpassung des LLTM an die Daten dabei statistisch signifikant schlechter (2LL = 10.792 vs. 10.203; $\chi^2 = 589$, df = 31, p < ,01). Das war zu erwarten, weil sich anhand der Eigenschaften einer Aufgabe deren Schwierigkeit nicht perfekt prognostizieren lässt. Mithilfe des $\Delta^{1/2}$ -Index beurteilten wir folglich die relative Anpassungsgüte des LLTM im Vergleich zum Rasch-Modell (DANIEL/EMBRETSON 2010, S. 356; EMBRETSON/DANIEL 2008, S. 339). Um den $\Delta^{1/2}$ -Index ermitteln zu können, wurde zusätzlich ein Nullmodell geschätzt. Dort werden nur die Personenfähigkeiten, keine Schwierigkeitsparameter, in das Modell aufgenommen.

Die Anpassung an die Daten ist dementsprechend mit einem -2LL-Wert von 12.676 schlechter. Sowohl das Rasch-Modell als auch das LLTM weisen im Vergleich zum Nullmodell eine statistisch signifikant bessere Anpassung auf ($\chi^2 = 2.473$, $df = 38$, $p < ,01$) bzw. ($\chi^2 = 1.884$, $df = 7$, $p < ,01$). Hieraus leitet sich ein $\Delta^{1/2}$ -Index von ,87 ab, der für eine gute relative Anpassung des LLTM an die Daten in der handlungsbasierten Dimension spricht (ebd., S. 339). In der verstehensbasierten Dimension ist der relative Fit des LLTM mit einem $\Delta^{1/2}$ -Index von 0,39 unbefriedigend.

Handlungsbasierte Dimension

Die Ergebnisse der Schätzungen des LLTM für die handlungsbasierte Dimension finden sich in Tab. 11. Sie können als weitere Evidenz für die Hypothesen H2b, H2c und H2d gewertet werden. H2a ist, wie im Fall der linearen Regression, zu verwerfen. Bei simultaner Berücksichtigung aller schwierigkeitsbestimmenden Merkmale weist der Grad der Vertrautheit einen statistisch signifikant negativen Einfluss auf die Wahrscheinlichkeit einer korrekten Lösung auf. Das ist gleichbedeutend mit einer zunehmenden Schwierigkeit der Aufgabe bei steigender Vertrautheit.

Tab. 11: Ergebnisse der LLTM Schätzung: handlungsbasierte Dimension

Variable	B	Standard- fehler	Wald	Sig. ^a
Konstante	2,235	1,030	4,711	,030
Vertrautheit	-0,245	0,051	22,730	< ,01
Modellierung_2	-1,175	0,084	194,325	< ,01
Modellierung_3	-2,497	0,103	592,139	< ,01
Informationsgewinnung_2	-1,262	0,093	185,117	< ,01
Informationsgewinnung_3	-2,571	0,109	556,968	< ,01
Informationsverarbeitung_2	-0,766	0,071	115,754	< ,01
Informationsverarbeitung_3	-1,895	0,086	488,254	< ,01

a. p-Werte wurden mit einem Wald-Test (2-seitig) ermittelt.

Verstehensbasierte Dimension

Im Ergebnis konnte mit dem LLTM für die verstehensbasierte Dimension ein signifikant negativer Einfluss des Grads der Vertrautheit auf die Lösungswahrscheinlichkeit einer Aufgabe festgestellt werden, was zur Ablehnung von H2a führt, s. Tab. 12.

Tab. 12: Ergebnisse der LLTM Schätzung: verstehensbasierte Dimension

Variable	B	Standard- fehler	Wald	Sig. ^a
Konstante	-0,769	1,235	0,387	,534
Vertrautheit	-0,670	0,113	35,329	< ,01
Informationsverarbeitung_2 (Verstehen)	0,336	0,164	4,206	,040
Informationsverarbeitung_3 (Analyse)	-0,531	0,195	6,941	< ,01

a. p-Werte wurden mit einem Wald-Test (2-seitig) ermittelt.

Bezogen auf die *Informationsverarbeitung* ergibt sich folgendes Bild: Im Vergleich zur Referenzkategorie *Wissen* weist die Kategorie *Verstehen* eine signifikant höhere Lösungswahrscheinlichkeit auf. *Analyse*-Aufgaben haben eine statistisch signifikant geringere Lösungswahrscheinlichkeit als Wissensaufgaben. D. h. verstehensbasierte Aufgaben, zu deren Lösung Wissen nötig ist, sind schwieriger als jene, die zur erfolgreichen Bewältigung ein Verständnis voraussetzen. Insgesamt ist damit auch auf Grundlage des LLTM die Hypothese H2d und damit die Gültigkeit des Niveaumodells für die verstehensbasierte Dimension abzulehnen.

4.6 Diskussion der Ergebnisse

Die Notwendigkeit der Aufspaltung von Rechnungswesenkompetenz in eine verstehens- und eine handlungsbasierte Dimension konnte bestätigt werden. Zudem spricht gegen deren Zusammenlegung eine Korrelation der beiden Dimensionen von stets kleiner als ,90 (ABELE/GSCHWENDTNER 2010, S. 16). Zu erwähnen ist in diesem Zusammenhang die relativ geringe Reliabilität der beiden Dimensionen in den vorliegenden Prüfungen. Diese beträgt im Mittel ,58 für die handlungsbasierte Dimension und ,49 für die verstehensbasierte Dimension, s. Tab. 6. Der Grund hierfür liegt in der teilweise sehr geringen Anzahl an Aufgaben je Dimension und Prüfung. Die Trennschärfen im Sinne der klassischen Testtheorie sind für sämtliche Items zufriedenstellend (HALADYNA/RODRIGUEZ 2013, S. 344), s. Tab. 4 und 5.

Eine weitere Ursache für Mehrdimensionalität könnte in der variierenden Schwierigkeit der Aufgaben liegen (SEEBER 2008, S. 83). Das ist für die vorliegenden Daten allerdings nicht der Fall. Für jede der fünf Prüfungen führten wir in der handlungsbasierten Dimension einen Martin-Löf-Test durch. Dazu bildeten wir in den drei Dimensionen *Modellierung*, *Informationsgewinnung* und *Informationsverarbeitung* jeweils zwei Gruppen: Stufe 1 vs. Stufe 2 und 3. Unter Berücksichtigung der Alphafehler-Kumulierung (prüfungsweise) konnte zum Niveau $\alpha = 5\%$ die Nullhypothese, wonach die Subskalen dasselbe messen, in keinem Fall verworfen werden.

Die Gültigkeit des Niveaumodells wurde sowohl mit einer linearen Regression als auch mit einem LLTM geprüft. Für die drei Schwierigkeitsdimensionen sind die Ergebnisse von LLTM und linearer Regression konsistent. In der handlungsbasierten Dimension führten beide Verfahren zu strukturell identischen Ergebnissen. Die Korrelation (Pearson) der Schätzparameter von *Modellierung*, *Informationsgewinnung* und *Informationsverarbeitung* beträgt zwischen den beiden Modellen -,99. Das deckt sich mit den Ergebnissen von HARTIG (2007, S. 97). Der Vorteil des

LLTM ist, dass damit deutlich eher statistisch signifikante Einflüsse nachzuweisen sind. Die lineare Regression wiederum zeichnet sich durch den deutlich geringeren Aufwand der Datenaufbereitung und Parameterschätzung aus. Für die Festlegung von Kompetenzstufen scheint es aus diesem Grund – wie von HARTIG (2007) vorgeschlagen – angemessen, auf das konservativere Verfahren der linearen Regression zurückzugreifen.

Kontraintuitiv ist der im Fall der linearen Regression fehlende statistisch signifikante Einfluss des Grades der Vertrautheit auf die Schwierigkeit der Aufgabe. Im LLTM ist der Zusammenhang zwischen der Lösungswahrscheinlichkeit einer Aufgabe und dem Grad der Vertrautheit sogar statistisch signifikant negativ. Das widerspricht der theoretisch gut begründeten und in anderen Domänen bestätigten (NICKOLAUS/GSCHWENDTNER/ABELE 2009, S. 10) Hypothese, wonach ein höherer Grad der Vertrautheit zu einer geringeren Aufgabenschwierigkeit führt. Der wahrscheinliche Grund hierfür ist somit eine ungeeignete Operationalisierung der Vertrautheit über die logarithmierte Häufigkeit des Vorkommens einer Aufgabe in den vorhergehenden zehn Prüfungen. Weitere Analysen, bei denen der Zeitraum von zehn auf fünf und anschließend zwei Prüfungen reduziert wurde, führten zu vergleichbaren (negativen) Ergebnissen. Auch eine dichotome Kodierung, die nur unterscheidet, ob eine Aufgabe in den vergangenen zehn bzw. fünf und zwei Prüfungen bereits in dieser Form gestellt wurde oder nicht, verlief erfolglos. Oftmals waren Aufgaben zwar nicht in derselben, aber in artverwandter Form in vorhergehenden Prüfungen enthalten. Ein Beispiel für eine Artverwandtheit ist die Kontierung einer Frachtkostenrechnung und die Kontierung der Rechnung für die Instandhaltung einer Maschine. In beiden Fällen handelt es sich um Fremdleistungen, was zu identischen Kontierungen führt. Der Versuch, eine breitere Fassung von typgleichen Aufgaben zu verwenden, scheiterte an der notwendigen Objektivität und Reliabilität der Abgrenzung. Diese konnte bei einer Erweiterung auf Artverwandtheit nicht gewährleistet werden. Im Ergebnis führte keine der durchgeführten Reanalysen zu mit H2a konformen Ergebnissen. Ein inhaltlicher Grund hierfür könnte sein, dass das Bearbeiten von alten Prüfungsaufgaben im Rahmen der Prüfungsvorbereitung keine dominierende Rolle einnimmt. Mit Schulbüchern steht ein alternatives Medium zur Verfügung, das intensiv zur Prüfungsvorbereitung genutzt wird (ERNST 2012, S. 1, 9, 114, 189). Darüber hinaus existieren Bücher, die explizit auf die Prüfungsvorbereitung abzielen. Ferner bietet die IHK selbst Kurse zur Prüfungsvorbereitung an. Weil es sich beim Grad der Vertrautheit lediglich um eine Kontrollvariable handelt, wurde auf weiterführende Analysen, die über die vorliegenden Daten hinausgehen, verzichtet. Denkbar wäre in diesem Kontext, den Grad der Vertrautheit von Experten einschätzen zu lassen.

In der verstehensbasierten Kompetenzdimension konnte die Hypothese eines steigenden Anspruchs der Stufen *Wissen*, *Verstehen* und *Analyse* nicht bestätigt werden. Ein wahrscheinlicher Grund ist, dass die Stufen, anders als bei der New Taxonomy, Aufgaben über Inhaltsbereiche hinweg vergleichen. Die Frage „Was sind Abschreibungen?“ (Wissen) ist als leichter einzuschätzen als „Warum sind Abschreibungen zu bilden?“ (Verstehen). Hier setzt, konform mit der New Taxonomy, das Verstehen ein entsprechendes Wissen voraus. Demgegenüber weist die Frage nach der Höhe des maximalen degressiven Abschreibungssatzes im Jahr 2009 (Wissen) vermutlich eine geringere Lösungswahrscheinlichkeit auf als die Frage, warum Abschreibungen zu bilden sind (Verstehen). Das könnte die statistisch signifikant

höhere Schwierigkeit von Wissensaufgaben im Vergleich zu Verstehensaufgaben bei Verwendung des LLTM erklären.

Das Kompetenzniveaumodell enthält in jeder Dimension vier Stufen. Die vierte Stufe wurde von keiner einzigen IHK-Aufgabe in wenigstens einer der drei Dimensionen erreicht. Das ist insofern von Bedeutung, weil die IHK-Aufgaben insgesamt zu leicht sind. 95,1 % der handlungsbasierten Aufgaben und 100 % der verstehensbasierten Aufgaben weisen einen negativen Logit-Wert auf. Bei der gewählten Logit-Skalierung entspricht das einer Lösungswahrscheinlichkeit des durchschnittlichen Testteilnehmers von größer als 50 %. Unter Gesichtspunkten der Reliabilität ist aber eine Lösungswahrscheinlichkeit von 50 % optimal (BÜHNER 2011, S. 506 f.). Aufgaben zum Zwecke der Kompetenzmessung im externen Rechnungswesen sollten demnach einen höheren Schwierigkeitsgrad als jene der IHK-Abschlussprüfung aufweisen. Würden bereits die IHK-Aufgaben die vierte Stufe in den drei Schwierigkeitsdimensionen systematisch erreichen, würde das die Nützlichkeit des Modells für die Konstruktion von angemessen schweren Aufgaben stark einschränken.

5 Zusammenfassung, Implikationen und Ausblick

Zusammenfassung

Ziel dieses Beitrags war es, ein Kompetenzniveaumodell für das externe Rechnungswesen zu entwickeln und zu validieren. Dazu wurde der wirtschaftsinstrumentelle Ansatz als fachdidaktische Grundlage mit Forschungsergebnissen zum Problemlöseprozess im Rechnungswesen aus dem internationalen Kontext verknüpft. Bei der Kompetenzstruktur wurde, dem Stand der wirtschaftspädagogischen Forschung folgend, eine Aufspaltung in eine verstehens- und eine handlungsbasierte Kompetenzdimension vorgenommen. Die Notwendigkeit dieser Trennung konnte empirisch bestätigt werden. Das Kompetenzniveaumodell setzt sich aus drei Schwierigkeitsdimensionen zusammen. Dabei handelt es sich um die *Modellierung*, *Informationsgewinnung* und *Informationsverarbeitung*. Zur empirischen Prüfung des Niveaumodells wurden eine lineare Regression und ein LLTM eingesetzt. Die Evidenz ist dabei zweigeteilt. Im Bereich der handlungsbasierten Kompetenz beeinflussen alle drei Schwierigkeitsdimensionen die Aufgabenschwierigkeit positiv. Dort kann das entwickelte Modell als bestätigt gelten. In der verstehensbasierten Kompetenzdimension konnten die vermuteten Zusammenhänge zwischen den (vermeintlich) schwierigkeitsbestimmenden Merkmalen und der Schwierigkeit einer Anforderungssituation nicht bestätigt werden. Ein höherer Grad der Vertrautheit mit der Aufgabe führte in keiner der beiden Kompetenzdimensionen zu einer Reduktion der Schwierigkeit.

Implikationen

Der vorliegende Beitrag hat Implikationen für Curriculum, Instruktion und Assessment.

Gegenwärtige Curricula sind nach dem Lernfeldkonzept aufgebaut. Die zu erreichenden Kompetenzen bleiben dabei regelmäßig im Unklaren (MINNAMEIER 2004, S. 241 ff.). Das Kompetenzniveaumodell kann helfen, diese zu präzisieren, indem Bezug auf die drei Schwierigkeitsdimensionen genommen wird.

Im Bereich der Instruktion kann das Kompetenzniveaumodell dazu dienen, den

Unterricht zu artikulieren. Folgende Strukturierung des Unterrichts lässt sich aus dem Kompetenzniveaumodell ableiten: Ausgehend von einer komplexen Situation identifizieren die Schüler die enthaltenen Probleme und modellieren die damit verbundenen Wertströme. Im nächsten Schritt leiten die Schüler hieraus die zur Lösung notwendigen Informationen ab, die dann aus authentischen Belegen gewonnen werden. Abschließend kombinieren die Schüler die modellierten Wertströme mit den gewonnenen Informationen, um zur Lösung des Problems zu gelangen.

Im Rahmen des Assessments können die Aufgabenschwierigkeiten unter Rückgriff auf das Kompetenzniveaumodell theoriegeleitet eingeschätzt und variiert werden. Gerade die IHK könnte diese Erkenntnisse für die im Rahmen der Erstellung der Abschlussprüfung vorgenommene Bestimmung des Schwierigkeitsgrads (VOGEL 2014, S. 18) nutzen. Dadurch kann die Messgenauigkeit optimiert werden, indem für die Testteilnehmer angemessen schwierige Aufgaben ausgewählt bzw. konstruiert werden. Die Häufigkeit, mit der eine Aufgabe in dieser Form bereits in vorhergehenden Prüfungen gestellt wurde, hatte keinen negativen Einfluss auf deren Schwierigkeit. Vor diesem Hintergrund ist die praktizierte Wiederverwendung von Prüfungsaufgaben (ebd., S. 18) nicht zu beanstanden.

Ausblick

Im vorliegenden Beitrag wurde das Niveaumodell mit quantitativer Methode validiert. Mit qualitativer Methode wäre eine Validierung über Think Aloud Studien möglich (ERICSSON/SIMON 1993; WILSON 2005, S. 55 ff.). Auf diese Weise könnte die Robustheit der Ergebnisse in der handlungsbasierten Dimension geprüft werden. Ein qualitatives Design würde sich auch anbieten, um Erkenntnisse zu den Gründen für die mangelnde Gültigkeit des Niveaumodells in der verstehensbasierten Dimension zu gewinnen (COLLINS/ONWUEGBUZIE/SUTTON 2006, S. 79). Die sich hieraus ergebenden Hypothesen zu den schwierigkeitsbestimmenden Merkmalen können anschließend wiederum mit quantitativer Methode getestet werden.

6 Literatur

- ABELE, S./GSCHWENDTNER, T. (2010): Die computerbasierte Erfassung beruflicher Handlungskompetenz. Konzepte, Möglichkeiten, Perspektiven am Beispiel der KfZ-Mechatronik. In: *Berufsbildung in Wissenschaft und Praxis*, H. 1, S. 14–17.
- ACHTENHAGEN, F. (2001): Zum Konzept des zielerreichenden Lernens in der Industriekaufleuteausbildung. *Modernisierung der Berufsbildung in Europa Neue Befunde wirtschafts- und berufspädagogischer Forschung*. Opladen: Leske und Budrich.
- ACHTENHAGEN, F. (2012): The curriculum-instruction-assessment triad. In: *Empirical Research in Vocational Education & Training* 4, H. 1, S. 5–25.
- ACHTENHAGEN, F./WINTHER, E. (2008): Wirtschaftspädagogische Forschung zur beruflichen Kompetenzentwicklung. In: JUDE, N./HARTIG, J./KLIEME, E. (Hrsg.): *Kompetenzerfassung in pädagogischen Handlungsfeldern. Theorien, Konzepte und Methoden*. Bonn, Berlin: Bundesministerium für Bildung und Forschung (BMBF), S. 117–140.
- AMELUNG, D./FUNKE, J. (2013): Dealing with the uncertainties of climate engineering: Warnings from a psychological complex problem solving perspective. In: *Technology in Society* 35, H. 1, S. 32–40.
- ANDERSON, L. W./KRATHWOHL, D. R. (2001): *A Taxonomy for Learning, Teaching and Assessing*. New York: Longman.

- BADER, R. (Hrsg.) (2004): Handlungsfelder-Lernfelder-Lernsituationen. Unterrichtsgestaltung nach dem Lernfeldkonzept. Dokumentation zum BLK-Modellversuchsverbund SELUBA. Bielefeld: Bertelsmann.
- BAETHGE, M./ACHTENHAGEN, F./AREND, L./BABIC, E./BAETHGE-KINSKY, V./WEBER, S. (2006): Berufsbildungs-PISA. Machbarkeitsstudie. Stuttgart: Steiner.
- BARIL, C. P./CUNNINGHAM, B. M./FORDHAM, D. R./GARDNER, R. L./WOLCOTT, S. K. (1998): Critical thinking in the public accounting profession: aptitudes and attitudes. In: *Journal of Accounting Education* 16, 3–4, S. 381–406.
- BLOOM, B. S. (1974): *Taxonomie von Lernzielen im kognitiven Bereich*. Weinheim: Beltz.
- BOHLINGER, S. (2010): Validierung von Kompetenzen und Anerkennung von Lernergebnissen – Nationale Ansätze zur Umsetzung europapolitischer Ziele. In: MÜNK, D./SCHELLEN, A. (Hrsg.): *Kompetenzermittlung für die Berufsbildung. Verfahren, Probleme und Perspektiven im nationalen, europäischen und internationalen Raum*. Bielefeld: Bertelsmann, S. 37–50.
- BRAEKEN, J./TUERLINCKX, F. (2009): Investigating latent constructs with item response models: a MATLAB IRTm toolbox. In: *Behavior Research Methods* 41, H. 4, S. 1127–1137.
- BREUER, K. (2011): Zur Abbildung von Prozesskompetenz in quasiexperimentellen Simulationen. In: ZLATKIN-TROITSCHANSKAIA, O. (Hrsg.): *Stationen empirischer Bildungsforschung. Traditionen und Perspektiven*. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 368–382.
- BRÖTZ, R./EHRKE, M. (2006): *Thesenpapier Leitbild kaufmännisch-betriebswirtschaftlicher Berufsbildung*. http://www.bibb.de/dokumente/pdf/Thesenpapier_Leitbild.pdf. 05.01.2015.
- BÜHNER, M. (2011): *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson Studium.
- BYRON, K./LAURENCE, G. A. (2015): Diplomas, Photos, and Tchotchkes as Symbolic Self-Representations: Understanding Employees' Individual Use of Symbols. In: *Academy of Management Journal* 58, H. 1, S. 298–323.
- COLLINS, K. M./ONWUEGBUZIE, A. J./SUTTON, I. L. (2006): A model incorporating the rationale and purpose for conducting mixed methods research in special education and beyond. In: *Learning Disabilities: A Contemporary Journal* 4, H. 1, S. 67–100.
- COONEY, M. (2005): Business Information Literacy Instruction. In: *Journal of Business & Finance Librarianship* 11, H. 1, S. 3–25.
- CRAWFORD, L./HELLIAR, C./MONK, E. A. (2011): Generic Skills in Audit Education. In: *Accounting Education* 20, H. 2, S. 115–131.
- CUNNINGHAM, N. A./ANDERSON, S. C. (2005): A Bridge to FARS and Information Literacy for Accounting Undergraduates. In: *Journal of Business & Finance Librarianship* 10, H. 3, S. 3–16.
- DAIGLE, R. J./HAYES, D. C./HUGHES, K. E. (2007): Assessing Student Learning Outcomes in the Introductory Accounting Information Systems Course Using the AICPA's Core Competency Framework. In: *Journal of Information Systems* 21, H. 1, S. 149–169.
- DALLAS, A. D./WILLSE, J. T. (2014): Survey analysis with mixture Rasch models. In: *Journal of applied measurement* 15, H. 4, S. 394–404.
- DANIEL, R. C./EMBRETSON, S. (2010): Designing Cognitive Complexity in Mathematical Problem-Solving Items. In: *Applied Psychological Measurement* 34, H. 5, S. 348–364.
- DAVIER, M. VON (1997): Bootstrapping Goodness-of-Fit Statistics for Sparse Categorical Data. – Results of a Monte Carlo Study –. In: *Methods of Psychological Research Online* 2, H. 2, S. 29–48.
- DEMONG, R. F./LINDGREN, J. H./PERRY, S. E. (1994): Designing an Assessment Program for Accounting. In: *Issues in Accounting Education* 9, H. 1, S. 11–27.
- EMBRETSON, S./DANIEL, R. C. (2008): Understanding and quantifying cognitive complexity level in mathematical problem solving items. In: *Psychology Science Quarterly* 50, H. 3, S. 328–344.
- EMBRETSON, S./GORIN, J. (2001): Improving Construct Validity with Cognitive Psychology Principles. In: *Journal of Educational Measurement* 38, H. 4, S. 343–368.

- ERICSSON, K. A./SIMON, H. A. (1993): Protocol analysis. Verbal reports as data. Cambridge, Mass.: MIT Press.
- ERNST, F. (2012): Fachdidaktische Analyse von Lehrbüchern für den Rechnungswesenunterricht in Deutschland und den USA. Jena: Friedrich-Schiller-Universität.
- FIETZ, G. (2008): Flexibilisierung von Lernprozessen und „Outcome-Orientierung“ – Aspekte eines Modellprojekts im europäischen Kontext. In: *bwp@ – Berufs- und Wirtschaftspädagogik online*, Spezial 4, S. 1–9.
- FINLEY, W./WAYMIRE, T. (2012): Information Literacy in the Accounting Classroom: A Collaborative Effort. In: *Journal of Business & Finance Librarianship* 17, H. 1, S. 34–50.
- FRÜH, W. (2007): Inhaltsanalyse. Theorie und Praxis. Konstanz: UVK.
- FUNKE, J. (2003): Problemlösendes Denken. Stuttgart: Kohlhammer.
- GELMAN, R./GREENO, J. G. (1989): On the Nature of Competence: Principles for Understanding. Knowing, learning, and instruction Essays in honor of Robert Glaser. Hillsdale: Lawrence Erlbaum.
- HAASLER, B./RAUNER, F. (2009): Kompetenzmessung in der beruflichen Bildung. In: *Zeitschrift für Berufs- und Wirtschaftspädagogik* 105, H. 3, S. 465–468.
- HAIR, J. F./HULT, G. T./RINGLE, C. M./SARSTEDT, M. (2014): A primer on partial least squares structural equations modeling (PLS-SEM). Los Angeles: SAGE.
- HALADYNA, T. M./RODRIGUEZ, M. C. (2013): Developing and validating test items. New York: Routledge.
- HAMMER, D. (1996): Misconceptions or P-Prims: How May Alternative Perspectives of Cognitive Structure Influence Instructional Perceptions and Intentions? In: *The Journal of the Learning Sciences* 5, H. 2, S. 97–127.
- HARSCH, C./HARTIG, J. (2011): Modellbasierte Definition von fremdsprachlichen Kompetenzniveaus am Beispiel der Bildungsstandards Englisch. In: *Zeitschrift für Interkulturellen Fremdsprachenunterricht: Didaktik und Methodik im Bereich Deutsch als Fremdsprache* 16, H. 2, S. 6–17.
- HARTIG, J. (2007): Skalierung und Definition von Kompetenzniveaus. In: KLIEME, E./BECK, B. (Hrsg.): Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch Englisch Schülerleistungen International). Weinheim: Beltz, S. 83–99.
- HARTIG, J. (2008): Kompetenzen als Ergebnisse von Bildungsprozessen. In: JUDE, N./HARTIG, J./KLIEME, E. (Hrsg.): Kompetenzerfassung in pädagogischen Handlungsfeldern. Theorien, Konzepte und Methoden. Bonn, Berlin: Bundesministerium für Bildung und Forschung (BMBF), S. 15–25.
- HARTIG, J./JUDE, N. (2007): Empirische Erfassung von Kompetenzen und psychometrische Kompetenzmodelle. In: HARTIG, J./KLIEME, E. (Hrsg.): Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik. Eine Expertise im Auftrag des Bundesministeriums für Bildung und Forschung. Bonn, Berlin: Bundesministerium für Bildung und Forschung (BMBF), S. 17–36.
- HARTIG, J./KLIEME, E. (2006): Kompetenz und Kompetenzdiagnostik. In: SCHWEIZER, K. (Hrsg.): Leistung und Leistungsdiagnostik. Berlin: Springer, S. 127–143.
- HEENE, M./DRAXLER, C./ZIEGLER, M./BÜHNER, M. (2011): Performance of the bootstrap Rasch model test under violations of non-intersecting item response functions. In: *Psychological Test and Assessment Modeling* 53, H. 3, S. 283–294.
- HELM, C./WIMMER, B. (2012): Wie lässt sich der Lernerfolg von Schülerinnen/Schülern im Fach Rechnungswesen messen? In: *wissenplus* 5, H. 11, S. 24–29.
- JACKSON, S./DURKEE, D. (2008): Incorporating Information Literacy into the Accounting Curriculum. In: *Accounting Education* 17, H. 1, S. 83–97.
- KLIEME, E./AVENARIUS, H./BLUM, W./DÖBRICH, P./GRUBER, H./PRENZEL, M./REISS, K./RIQUARTS, K./ROST, J./TENORTH, H.-E./VOLLMER, H. J. (2009): Zur Entwicklung nationaler Bildungsstandards. Expertise. Bonn, Berlin: Bundesministerium für Bildung und Forschung (BMBF).

- KLIEME, E./LEUTNER, D. (2006): Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. In: *Zeitschrift für Pädagogik* 52, H. 6, S. 876–903.
- KMK (2002): Rahmenlehrplan für den Ausbildungsberuf Industriekaufmann/Industriekauf-frau. <http://www.kmk.org/fileadmin/pdf/Bildung/BeruflicheBildung/rlp/industriekfm.pdf>. 05.01.2015.
- KOLLER, I./ALEXANDROWICZ, R./HATZINGER, R. (2012): *Das Rasch Modell in der Praxis*. Stuttgart: UTB.
- MAIR, P./HATZINGER, R. (2007): Extended Rasch Modeling: The eRm Package for the Appli-cation of IRT Models in R. In: *Journal of Statistical Software* 20, H. 9, S. 1–20.
- MARZANO, R. J./KENDALL, J. S. (2007): *The New Taxonomy of Educational Objectives*. Thousand Oaks, CA: Corwin Press.
- MINDNICH, A./BERGER, S./FRITSCH, S. (2013): Modellierung des fachlichen und fachdidaktischen Wissens von Lehrkräften im Rechnungswesen – Überlegungen zur Konstruktion eines Testinstruments. In: FASSHAUER, U./FÜRSTENAU, B./WUTTKE, E. (Hrsg.): *Jahrbuch der berufs- und wirtschaftspädagogischen Forschung 2013*. Opladen: Barbara Budrich, S. 61–72.
- MINNAMEIER, G. (2004): Wie fruchtbar sind Lernfelder? In: *Wirtschaft und Erziehung* 6, S. 239–246.
- MINNAMEIER, G./LINK, M. (2010): Jenseits des wirtschaftsinstrumentellen Rechnungswesens – ein kognitiv struktureller und inferentieller Ansatz. In: SEIFRIED, J. (Hrsg.): *Lehr-Lern-Forschung in der kaufmännischen Berufsbildung – Ergebnisse und Gestaltungsaufgaben*. Stuttgart: Steiner, S. 107–121.
- MÜNK, D./SCHELLEN, A./SCHMID, A. (2010): Perspektiven der Kompetenzmessung: Forschungs-stand und Desiderate aus internationaler und europäischer sowie aus der Binnenpers-pektive. In: MÜNK, D./SCHELLEN, A. (Hrsg.): *Kompetenzermittlung für die Berufsbildung. Verfahren, Probleme und Perspektiven im nationalen, europäischen und internationalen Raum*. Bielefeld: Bertelsmann, S. 7–16.
- NICKOLAUS, R. (2011): Die Erfassung fachlicher Kompetenzen und ihrer Entwicklungen in der beruflichen Bildung – Forschungsstand und Perspektiven. In: ZLATKIN-TROITSCHANSKAIA, O. (Hrsg.): *Stationen empirischer Bildungsforschung. Traditionslinien und Perspektiven*. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 331–351.
- NICKOLAUS, R./GSCHWENDTNER, T./ABELE, S. (2009): Die Validität von Simulationsaufgaben am Beispiel der Diagnosekompetenz von Kfz- Mechatronikern. Vorstudie zur Validität von Simulationsaufgaben im Rahmen eines VET-LSA. Stuttgart: Universität Stuttgart.
- NICKOLAUS, R./PETSCH, C./NORWIG, K. (2013): Berufsfachliche Kompetenzen am Ende der Grundbildung in bautechnischen Berufen. Modellierung und erzielte Leistungen in Abhängigkeit der angestrebten beruflichen Profile. In: *Zeitschrift für Berufs- und Wirt-schaftspädagogik* 109, H. 4, S. 538–553.
- NORMAND, C. J./DUDLEY, L. W. (2004): Teaching Problem Solving in Accounting Courses Using a Generic Problem-Solving Model. In: *Journal of Accounting and Finance Research* 12, H. 2, S. 62–77.
- PELLEGRINO, J. W. (2010): *The Design of an Assessment System for the Race to the Top: A Learning Sciences Perspective on Issues of Growth and Measurement*. Princeton: Educational Testing Service.
- PELLEGRINO, J. W./CHUDOWSKY, N./GLASER, R. (2001): *Knowing what students know. The sci-ence and design of educational assessment*. Washington, DC: National Academy Press.
- POINTSTINGL, H. (2009): The Linear Logistic Test Model (LLTM) as the methodological founda-tion of item generating rules for a new verbal reasoning test. In: *Psychology Science Quarterly* 51, H. 2, S. 123–134.
- PREISS, P. (1999): *Wirtschaftsinstrumentelles Rechnungswesen als Kernelement kaufmän-nischer Curricula*. München: Oldenbourg.
- PREISS, P. (2005a): Entwurf eines Kompetenzkonzepts für den Inhaltsbereich Rechnungs-wesen/Controlling. In: GONON, P./KLAUSER, F./NICKOLAUS, R./HUISINGA, R. (Hrsg.): *Kompe-*

- tenz, Kognition und neue Konzepte der beruflichen Bildung. Wiesbaden: VS Verl. für Sozialwissenschaften, S. 67–85.
- PREISS, P. (2005b): Förderung kaufmännischer Kompetenzen mit Hilfe des wirtschaftsinstrumentellen Rechnungswesens – aus fachlich-curricularer Perspektive. In: SEMBILL, D./SEIFRIED, J. (Hrsg.): Rechnungswesenunterricht am Scheideweg. Lehren, lernen und prüfen. Wiesbaden: Deutscher Universitätsverlag, S. 53–97.
- PREISS, P./TRAMM, T. (Hrsg.) (1996): Rechnungswesenunterricht und ökonomisches Denken. Didaktische Innovationen für die kaufmännische Ausbildung. Wiesbaden: Gabler.
- ROST, J. (2004): Lehrbuch Testtheorie – Testkonstruktion. Bern: Huber.
- SCHILLER, G. (2004): Lernschwierigkeiten im Fach Rechnungswesen. In: Erziehungswissenschaft und Beruf, H. 3, S. 335–363.
- SCHUMANN, S./EBERLE, F. (2011): Bedeutung und Verwendung schwierigkeitsbestimmender Aufgabenmerkmale für die Erfassung ökonomischer und beruflicher Kompetenzen. In: FASSHAUER, U./FÜRSTENAU, B./WUTTKE, E. (Hrsg.): Grundlagenforschung zum Dualen System und Kompetenzentwicklung in der Lehrerbildung. Leverkusen: Budrich, S. 77–89.
- SEEBER, S. (Hrsg.) (2013): ULME III Untersuchung von Leistungen, Motivation und Einstellungen der Schülerinnen und Schüler in den Abschlussklassen der Berufsschulen. Münster: Waxmann.
- SEEBER, S./NICKOLAUS, R. (2010): Kompetenzmessung in der beruflichen Bildung. In: Berufsbildung in Wissenschaft und Praxis, H. 1, S. 10–13.
- SEEBER, S./NICKOLAUS, R./WINTHER, E./ACHTENHAGEN, F./BREUER, K./FRANK, I./LEHMANN, R./SPÖTL, G./STRAKA, G. A./WALDEN, G./WEISS, R./ZÖLLER, A. (2010): Kompetenzdiagnostik in der Berufsbildung. Begründung und Ausgestaltung eines Forschungsprogramms. In: Berufsbildung in Wissenschaft und Praxis (Beilage zu 1), S. 1–15.
- SEIFRIED, J. (2002): Selbstorganisiertes Lernen im Rechnungswesen. In: Zeitschrift für Berufs- und Wirtschaftspädagogik 98, H. 1, S. 104–121.
- SEMBILL, D./SEIFRIED, J. (Hrsg.) (2005): Rechnungswesenunterricht am Scheideweg. Lehren, lernen und prüfen. Wiesbaden: Deutscher Universitätsverlag.
- SMITH, J. P./DI SESSA, A. A./ROSCHELLE, J. (1994): Misconceptions Reconceived: A Constructivist Analysis of Knowledge in Transition. In: The Journal of the Learning Sciences 3, H. 2, S. 115–163.
- STAUDINGER, U. M./PATZWALD, K. (2009): Lebenslanges Lernen und strategisches Personalmanagement. In: Wirtschaft und Berufserziehung, H. 5, S. 13–18.
- STIVERS, B. P./CAMPBELL, J. E./HERMANSON, H. M. (2000): An Assessment Program for Accounting: Design, Implementation, and Reflection. In: Issues in Accounting Education 15, H. 4, S. 553–581.
- STOUT, D. E./BORDEN, J. P./GERMAN, M./MONAHAN, T. F. (2005): Designing and implementing a comprehensive assessment plan for a graduate accounting programme. In: Accounting Education 14, H. 4, S. 395–410.
- STRAKA, G. A. (2000): Lernen unter informellen Bedingungen (informelles Lernen) Begriffsbestimmung, Diskussion in Deutschland, Evaluation und Desiderate. In: Arbeitsgemeinschaft Qualifikations-Entwicklungs-Management (Hrsg.): Kompetenzentwicklung 2000. Lernen im Wandel-Wandel durch Lernen. Münster: Waxmann, S. 15–70.
- STRAKA, G. A./MACKE, G. (2003): Handlungskompetenz und Handlungsorientierung als Bildungsauftrag der Berufsschule – Ziel und Weg des Lernens in der Berufsschule? In: Berufsbildung in Wissenschaft und Praxis, H. 4, S. 43–47.
- STROBL, C. (2012): Das Rasch-Modell. München: Rainer Hampp Verlag.
- SUNDEM, G. L./WILLIAMS, D. Z. (1992): Changes in accounting education: preparing for the twenty-first century. In: Accounting Education 1, H. 1, S. 55–61.
- SWELLER, J. (2006): How the Human Cognitive System Deals with Complexity. In: ELEN, J./CLARK, R. E. (Hrsg.): Handling complexity in learning environments. Theory and research. Oxford: Elsevier, S. 13–25.

- TRAMM, P. T./HINRICHS, K./LANGENHEIM, H. (1996): Lernschwierigkeiten im Buchführungsunterricht. In: PREISS, P./TRAMM, T. (Hrsg.): Rechnungswesenunterricht und ökonomisches Denken. Didaktische Innovationen für die kaufmännische Ausbildung. Wiesbaden: Gabler, S. 158–221.
- TRAMM, T. (2003): Wirtschaftsinstrumentelles Rechnungswesen und die Modellierungsmethode – eine fachdidaktische Einführung. In: JOOST, D./KRIPPKE, G./TRAMM, T. (Hrsg.): Wirtschaftsinstrumentelles Rechnungswesen – Materialienband. Troisdorf: Bildungsverlag EINS, S. 4–10.
- TRAMM, T. (2005): Wirtschaftsinstrumentelles Rechnungswesen konkret. In: SEMBILL, D./SEIFRIED, J. (Hrsg.): Rechnungswesenunterricht am Scheideweg. Lehren, lernen und prüfen. Wiesbaden: Deutscher Universitätsverlag.
- TÜRLING, J. M./SEIFRIED, J./WUTTKE, E./GEWIESE, A./KÄSTNER, R. (2011): Typische Schülerfehler im Rechnungswesenunterricht - Empirische Befunde einer Interviewstudie. In: Zeitschrift für Berufs- und Wirtschaftspädagogik 107, H. 3, S. 390–407.
- VOGEL, W. (2014): Bundeseinheitliche IHK-Prüfungen am Beispiel der zentralen Aufgabenerstellung der Aka. In: Berufsbildung in Wissenschaft und Praxis, H. 3, S. 16–19.
- WEBER, S. (2013): Sense of Workplace Learning. In: Vocations and Learning, H. 6, S. 1–9.
- WEBER, S./ACHTENHAGEN, F. (2014): Fachdidaktisch gesteuerte Modellierung und Messung von Kompetenzen im Bereich der beruflichen Bildung. In: Zeitschrift für Erziehungswissenschaft 17, H. 1, S. 33–58.
- WEINERT, F. E. (Hrsg.) (2002): Leistungsmessungen in Schulen eine umstrittene Selbstverständlichkeit. Weinheim: Beltz.
- WILSON, M. (2005): Constructing Measures: An Item Response Modelling Approach. Mahwah: Lawrence Erlbaum.
- WINTHER, E. (2010): Kompetenzmessung in der beruflichen Bildung. Bielefeld: Bertelsmann.
- WINTHER, E. (2011): Kompetenzorientierte Assessments in der beruflichen Bildung – Am Beispiel der Ausbildung von Industriekaufleuten. In: Zeitschrift für Berufs- und Wirtschaftspädagogik 107, H. 1, S. 33–54.
- WINTHER, E./ACHTENHAGEN, F. (2008): Kompetenzstrukturmodell für die kaufmännische Bildung. In: Zeitschrift für Berufs- und Wirtschaftspädagogik 104, H. 4, S. 512–537.
- WINTHER, E./ACHTENHAGEN, F. (2009): Skalen und Stufen kaufmännischer Kompetenz. In: Zeitschrift für Berufs- und Wirtschaftspädagogik 105, H. 4, S. 521–556.
- WINTHER, E./KLOTZ, V. K. (2014): Spezifika der beruflichen Kompetenzdiagnostik – Inhalte und Methodologie. In: Zeitschrift für Erziehungswissenschaft 17, S1, S. 9–32.
- WIRTZ, M./CASPAR, F. (2002): Beurteilerübereinstimmung und Beurteilerreliabilität. Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen. Göttingen: Hogrefe.
- WISE, S. L./DEMARS, C. E. (2005): Low Examinee Effort in Low-Stakes Assessment: Problems and Potential Solutions. In: Educational Assessment 10, H. 1, S. 1–17.
- WU, M. L./ADAMS, R. J./WILSON, M. R. (2007): ACER ConQuest. Camberwell, Victoria: ACER Press.
- WÜSTEMANN, J. (2013): Buchführung case by case. Frankfurt am Main: Recht und Wirtschaft.

Anmerkung

Die Verfasser danken den anonymen Gutachterinnen/Gutachtern für ihre wertvollen Hinweise, die wir gerne aufgegriffen haben.

Anschrift der Autoren: STR Dipl.-Hdl. Josef Guggemos, MBR, Doktorand der Ludwig-Maximilians-Universität München, Lehrstuhl für Wirtschaftspädagogik, Ludwigstr. 28/RG, 80539 München, Mail: guggemos@bwl.lmu.de. OStR Dipl.-Hdl., Dipl.-Finanzw. Michael Schönlein, Doktorand der Ludwig-Maximilians-Universität München, Lehrstuhl für Wirtschaftspädagogik, Ludwigstr. 28/RG, 80539 München, Mail: schoenlein@bwl.lmu.de.