

FLORIAN BERDING / ANKE BECKMANN / VICTORIA KÜR TEN

Modellieren mit dem Rechnungswesen

Entwicklung eines rasch-konformen Messverfahrens für erfolgswirksame Vorgänge und didaktische Implikationen

Modelling within accounting

Development and validation of a measurement instrument and didactical implications

KURZFASSUNG: Das Rechnungswesen stellt ein abstraktes System aus Regeln und Objekten dar. Es soll Kaufleute bei der Gestaltung wirtschaftlicher Prozesse unterstützen. Damit Kaufleute die Potentiale dieses „Werkzeuges“ nutzen können, müssen sie in der Lage sein, reale ökonomische Phänomene mit dem Rechnungswesen zu modellieren. Ziel des vorliegenden Beitrages ist es auf der Grundlage einer Stichprobe von 443 Lernenden dualer und vollzeitschulischer Bildungsgänge einen Test zur Erfassung der Modellierungsfähigkeit erfolgswirksamer Vorgänge zu entwickeln.

Der entwickelte Test besteht aus 10 Aufgaben und erweist sich als raschkonform mit einer Reliabilität von .657. Dabei sind die Eigenschaften des Tests vom Bildungsgang (dual vs. vollzeitschulisch) unabhängig. Darüber hinaus lassen sich erwartungskonforme Zusammenhänge zu Motivation, Grundvorstellungen und Noten nachweisen, welche die Konstruktvalidität belegen. Es werden Richtlinien zur Konstruktion von Modellierungstests abgeleitet.

Die Ergebnisse zeigen, dass Aufgaben mit Belegen deutlich schwieriger sind als Aufgaben mit photorealistischen Abbildungen. Der Beitrag schließt daraus, dass die Rechnungswesendidaktik nicht bei dem Einbezug von Belegen stehen bleiben sollte und gibt praktische Hinweise für eine stärkere Anbindung an reale ökonomische Phänomene im Unterricht.

Schlagworte: Modellierungsfähigkeit, Grundvorstellungen, Motivation, Noten, Rechnungswesen, Belege

ABSTRACT: Accounting is an abstract system of symbols and rules. It is an instrument for commercial professionals in order to organize business processes. Using this tool requires professionals to develop modelling competencies. Against this backdrop the current paper aims to develop an instrument for measuring the modeling competence of learners.

Based on a sample of 443 apprentices and learners within vocational education and training the results indicate 10 tasks to form a rasch-conform measurement instrument. The reliability of the test is .657. The properties of the test items do not differentiate between learners of the dual system and full-time vocational schools. Relationships with motivation, basic ideas, and marks of learners are proved.

The findings indicate that the task with vouchers are more difficult than task with iconic representations. Thus, the paper argues that vouchers have to be introduced to accounting classes carefully and should be complemented by realistic representations of business processes.

Keywords: modeling competence, basic ideas, motivation, marks, accounting, vouchers

1 Einleitung

Der Rechnungswesenunterricht gilt traditionell als ein Kernelement der kaufmännischen beruflichen Bildung (vgl. PREISS/TRAMM 1996a, S. V). Konkret zeigt sich dies an dem hohen quantitativen Anteil in den Ausbildungsordnungen der kaufmännischen Ausbildungsberufe von durchschnittlich 38 % (vgl. BRÖTZ et al. 2015, S. 93), wobei hier die vollzeitschulischen Bildungsgänge wie z. B. das Berufliche Gymnasium Wirtschaft, die Berufsfachschule Wirtschaft oder die Fachoberschulen Wirtschaft noch nicht eingerechnet sind.

Den konkreten Rechnungswesenunterricht erleben Lernende in der Regel als Herausforderung. So stellt beispielsweise die Studie von SEEBER (2007, S. 117 ff.) bei angehenden Bürokaufleuten fest, dass ihnen Rechnungswesen im Vergleich zur allgemeinen Wirtschaftslehre (bestehend aus Betriebswirtschaftslehre, Volkswirtschaftslehre und Recht) deutlich schwerer fällt. Korrespondierend dazu berichten Studien der vergangenen Jahre immer wieder von hohen Lernschwierigkeiten und ermitteln als Ursachen u. a. den hohen Abstraktionsgrad der Inhalte, die geringe Anschaulichkeit und fehlende Bezüge zur Praxis bzw. Lebenswirklichkeit der Lernenden (vgl. z. B. PAWLIK 1980; TRAMM/HINRICHS/LANGENHEIM 1996, S. 188, S. 192 f.; TÜRLING 2014, S. 170 ff.).

Die besondere Relevanz dieser drei Ursachen wird zusätzlich dadurch untermauert, dass die Studie von FINDEISEN (2017, S. 238, S. 275 f.) zeigt, dass es Masterstudierenden der Wirtschaftspädagogik nur in geringem Maße gelingt, in ihre Erklärungen angemessene Repräsentationen für die Symbole und Zusammenhänge des Rechnungswesens einzubringen. Auch die Studie von TÜRLING (2014, S. 184 f.) zum Umgang mit Fehlern macht deutlich, dass Wirtschaftspädagog(inn)en unabhängig von ihrem Ausbildungsstatus bevorzugt eine Einschränkung der Lösungsmöglichkeiten, das Einbringen von Leitfragen und die Evaluation der Lösungen vornehmen, während Visualisierungen in einem mittleren Maße und ein Wechsel der Symbol- bzw. Objektebene, die Einbindung von Vorwissen und das Herstellen von Lebensweltbezügen nur selten zum Einsatz kommen. ERNST (2012, S. 73) stellt in einer Schulbuchanalyse schließlich fest, dass Illustrationen vorwiegend fachspezifische Themen fokussieren und die Lebenswelt der Lernenden sowie den betriebs- und volkswirtschaftlichen Kontext unberücksichtigt lassen.

Vor diesem Hintergrund ist es nicht verwunderlich, wenn in der nahen Vergangenheit kontrovers über die Stärken und Schwächen verschiedener didaktischer Ansätze diskutiert wurde, wie vor allem die Auseinandersetzung mit der Bilanzmethode und dem wirtschaftsinstrumentellen Rechnungswesenunterricht in der Zeitschrift „Wirtschaft und Erziehung“ belegt (vgl. z. B. BURKHARDT et al. 2014; PLINKE 2014). ERNST (2014, S. 17) fordert in dieser Diskussion empirische Studien zur Wirkung verschiede-

This material is under copyright. Any use outside of the narrow boundaries of copyright law is illegal and may be prosecuted.

This applies in particular to copies, translations, microfilming as well as storage and processing in electronic systems.

© Franz Steiner Verlag, Stuttgart 2020

ner didaktischer Konzepte, um Lehrkräften die begründete Auswahl eines Ansatzes zu erleichtern bzw. zu ermöglichen.

Bislang liegen jedoch nur wenige solcher Wirkungsstudien vor, die sich auf verschiedene Erfolgsindikatoren stützen (vgl. z. B. BOULEY 2017; SEIFRIED 2004; SEIFRIED/SEMBILL 2004; SEIFRIED et al. 2005). So verwendet beispielsweise die Studie von SEIFRIED (2004, S. 116, S. 131 f.) zur Ermittlung des Faktenwissens und der Buchungskompetenz einen Test nach dem Vorbild von Kammerprüfungen, während die Problemlösekompetenz mit einem komplexen Verfahren ermittelt wird. Die Studie von BOULEY (2017, S. 202 ff.) erfasst hingegen das tägliche Rechnungswesenwissen, das buchungsspezifische Wissen sowie das beleggestützte Wissen. Diese Studien geben damit wichtige Einblicke in die unterschiedlichen Facetten der beruflichen Handlungskompetenz von Lernenden im Bereich Buchführung und ihrer Förderung durch bestimmte didaktische Konzepte.

Die Kompetenzfacette, reale ökonomische Phänomene mit dem Rechnungswesen modellieren zu können, wurde bislang jedoch selten berücksichtigt. Dabei stellt sie eine wichtige Facette dar, wie HELM (2017) herausarbeitet. In dem entwickelten kognitiven Denkmodell steht die Tätigkeit des „Modellierens“ am Anfang sämtlicher Bearbeitungsschritte zur Bewältigung von Buchungsaufgaben und umfasst das sinnerfassende Lesen, die Einnahme verschiedener Perspektiven und das Erkennen zugrunde liegender Prozesse und Tätigkeiten (vgl. HELM 2017, S. 41). Einen anderen Ansatz wählen GUGGEMOS und SCHÖNLEIN (2015) in einer Verbindung des Kompetenzmodells von WINTHER und ACHTENHAGEN (2008) und dem wirtschaftsinstrumentellen Ansatz (PREISS/TRAMM 1990). In ihrem Kompetenzniveaumodell für das externe Rechnungswesen stellt die zu erbringende Modellierungsleistung neben der Informationsgewinnung und -verarbeitung eine eigenständige schwierigkeits- bzw. anforderungsbestimmende Dimension dar. Im Gegensatz zu HELM (2017) wird die Modellierungsleistung hier weniger über das Erkennen zugrunde liegender Handlungen operationalisiert, sondern über die Anzahl und Komplexität zu lösender Probleme, mit der geringsten Stufe, d. h. der Vorgabe eines Problems, bis zur höchsten Stufe, d. h. der Identifikation mehrerer Probleme, die über verschiedene Bereiche verteilt sind (vgl. GUGGEMOS/SCHÖNLEIN 2015, S. 527 f., S. 531).

Die Fähigkeit zur Modellierung von realen Phänomenen mit Begriffen und Verfahren des Rechnungswesens nimmt direkten Bezug zu den Zielen, die mit dem Unterricht verfolgt werden. So ist die Aufgabe des betrieblichen Rechnungswesens „die Erfassung, Speicherung und Verarbeitung von betriebswirtschaftlich relevanten quantitativen Informationen über angefallene oder geplante Geschäftsvorgänge und -ergebnisse [fett im Original]“ (SCHIERENBECK 2003, S. 505), um durch Dokumentation, Planung und Kontrolle wirtschaftliche Entscheidungen zu unterstützen und abzusichern (vgl. SCHIERENBECK 2003, S. 505; WÖHE/DÖRING 2005, S. 809). Demnach soll der Unterricht dazu beitragen, dass „die Lernenden die Funktion dieses Lernbereichs im Hinblick auf die Führung und Steuerung sowie auf das Controlling in der Unternehmung verstehen und die einzelnen Prozesse aus dieser ganzheitlichen Sichtweise anwenden können.“ (DUBS 2008, S. 26 f.; vgl. auch PREISS/TRAMM 1996b, S. 226; SEIFRIED 2002, S. 107).

This material is under copyright. Any use outside of the narrow boundaries of copyright law is illegal and may be prosecuted.

This applies in particular to copies, translations, microfilming as well as storage and processing in electronic systems.

© Franz Steiner Verlag, Stuttgart 2020

Die konkrete Bedeutung dieses Anspruchs für die Modellierungsfähigkeit lässt sich durch eine Analogie zur Mathematikdidaktik verdeutlichen. Ähnlich wie das Rechnungswesen stellt die Mathematik ein System aus abstrakten Regeln, Verfahren und Objekten dar. Gleichzeitig besteht in der Mathematik ebenfalls wie im Rechnungswesen der Anspruch, dass dieses abstrakte System ein wirkungsvolles Werkzeug bzw. Instrument zum Verstehen und Gestalten konkreter Phänomene in der Wirklichkeit bereitstellt (vgl. z. B. VOM HOFE 2003, S. 4). Damit wird es aus mathematik- und rechnungswesendidaktischer Sicht notwendig, reale Phänomene in abstrakte Modelle zu übersetzen, in diesen Modellen Lösungen zu erarbeiten und die Lösungen wieder mit realen Bedeutungen zu verknüpfen (vgl. z. B. BERDING 2019, S. 90 ff.; VOM HOFE 2003, S. 5; SCHUPP 1988).

Das Ziel des vorliegenden Beitrages ist es vor diesem Hintergrund ein standardisiertes Testverfahren zu entwickeln, welches die Modellierungsfähigkeit mit Hilfe des Rechnungswesens erfasst, um diese wichtige Kompetenzfacette in zukünftigen Wirkungsstudien berücksichtigen und den Blick auf die Wirkung didaktischer Konzepte über bereits bestehende Testverfahren hinaus erweitern zu können. Dabei wird die Entwicklung eines rasch-konformen Tests angestrebt, da diese Tests für didaktische Wirkungsstudien eine Reihe von wichtigen Vorteilen aufweisen, wie z. B. die Subgruppeninvarianz. Dieses Merkmal bedeutet, dass sich die Eigenschaften der Testaufgaben nicht systematisch zwischen verschiedenen Personengruppen unterscheiden (vgl. KOLLER/ALEXANDROWICZ/HATZINGER 2012, S. 62; STROBL 2015, S. 42). Praktisch bedeutet dies beispielsweise, dass der Test für Lernende in einem dualen oder vollzeitschulischen Bildungsgang gleichermaßen funktioniert und nicht eine Gruppe benachteiligt, d. h. faire und objektive Lernerfolgsmessungen erlaubt.

Bevor die eigentliche Testentwicklung und Ergebnisse dargestellt werden, beginnt der Beitrag mit einer Vertiefung des theoretischen Hintergrunds. Er endet mit einer Diskussion der Ergebnisse und stellt auch didaktische Konsequenzen dar.

2 Theoretischer Hintergrund

2.1 Modellieren im Rechnungswesen

Für ROSKI (1986, S. 86 f.) ist ein Modell „eine für einen bestimmten Zweck gebildete, vereinfachende Abbildung eines als System aufgefaßten Realitätsausschnitts“. Ähnlich definieren auch HAUSSER und LUCHKO (2011, S. 3) ein Modell als „ein Konzept zur Darstellung eines komplexen realen Systems oder Prozesses. Es beschreibt die in einem bestimmten Kontext wichtigen Eigenschaften oder Verhaltensmuster des entsprechenden Modellierungsobjektes.“

Diese Definitionen verweisen bereits auf die drei zentralen Eigenschaften und Funktionen von Modellen, wie sie auch bei STACHOWIAK (1973, S. 131 ff.) zu finden sind (vgl. auch HOLZMÜLLER/BANDOW 2010, S. VII; TÖLLNER et al. 2010, S. 8 f.). (1) Die wichtigste Aufgabe von Modellen ist die Abbildung von realen Phänomenen (Abbil-

dungsmerkmal). Diese Abbildung erfolgt (2) in der Regel durch eine Vereinfachung bzw. Komplexitätsreduzierung der realen Gegebenheiten. Dabei werden nur die Eigenschaften im Modell berücksichtigt, die für den jeweiligen Zweck relevant sind (Verkürzungsmerkmal). Modelle sind demnach (3) stets an einen Zweck gebunden (pragmatisches Merkmal). Aufgrund dieser Eigenschaften dienen Modelle in der Praxis z. B. dem Planen, Gestalten und Verbessern von Systemen und Prozessen (vgl. HOLZMÜLLER/BANDOW 2010, S. VII f.). Ziel ist es u. a. mit Modellen Erkenntnisse zu gewinnen, die für Entscheidungen von Relevanz sind (vgl. TÖLLNER et al. 2010, S. 17).

Die Fähigkeit zur Modellentwicklung lässt sich in einer ersten Annäherung mit Hilfe der Mathematikdidaktik beispielsweise durch MAASS (2006, S. 117) wie folgt definieren: „Modelling competencies include skills and abilities to perform modelling processes appropriately and goal-oriented as well as the willingness to put these into action.“ GREEFRATH et al. (2013, S. 18) verwenden hingegen ein breiteres Verständnis und betrachten die Modellierungskompetenz als „die Fähigkeit, die jeweils nötigen Prozessschritte beim Hin- und Herwechseln zwischen Realität und Mathematik problemadäquat auszuführen sowie gegebene Modelle zu analysieren oder vergleichend zu beurteilen.“

An dieses Verständnis von Modellierungsfähigkeit anknüpfend lässt sich für den Rechnungswesenunterricht der Prozess des Modellierens mit Hilfe des Modells der Grundvorstellungen darstellen (BERDING 2019), das auf den Arbeiten von VOM HOFE (2003) sowie SCHUPP (1988, S. 11) aus der Mathematikdidaktik basiert. Das Modell ist in Abbildung 1 dargestellt.

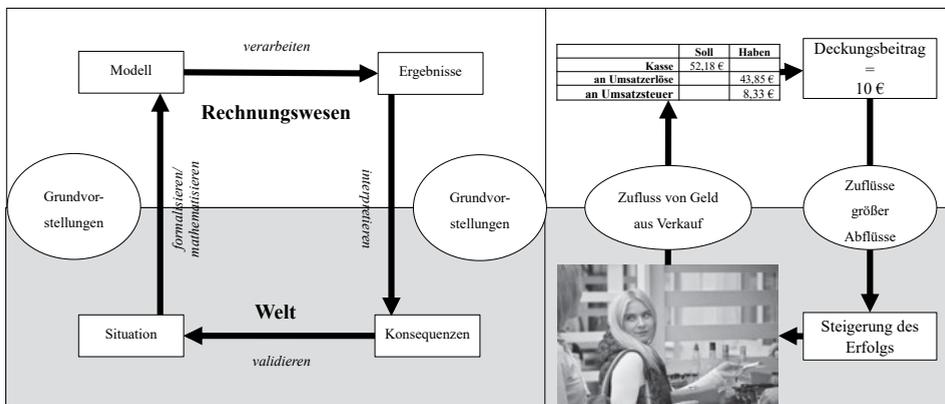


Abb. 1: Modell der Grundvorstellung und des Modellierens im Rechnungswesen in Anlehnung an BERDING 2019, S. 148; VOM HOFE 2003, S. 5; SCHUPP 1988, S. 11)

So sind Kaufleute bei der Verwendung des Rechnungswesens gefordert, ein reales ökonomisches Phänomen in das Rechnungswesen zu übersetzen. Beispielsweise muss der Verkauf von Waren gegen Bargeld, wie er in der Abbildung 1 auf der rechten Seite dargestellt ist, zunächst in das Rechnungswesen übertragen werden. Das Ergebnis einer solchen Übertragung kann z. B. ein entsprechender Buchungssatz sein. Hier besteht die

besondere Anforderung, dass das reale Phänomen von sich aus keinen expliziten Hinweis gibt, wie es zu übertragen ist. Vielmehr sind Kaufleute gefordert, eine Situation zu deuten und zu interpretieren und auf der Grundlage dieser Deutung passende Objekte, Begriffe und Verfahren aus dem Rechnungswesen auszuwählen, um das reale Phänomen entsprechend abzubilden. Im Beispiel aus Abbildung 1 könnte ein Buchungssatz eine solche stark vereinfachte „Übersetzung“ bzw. ein abstraktes Modell des Verkaufsvorganges darstellen (siehe das Verkürzungsmerkmal von Modellen).

Die so vollzogene Formalisierung und Übersetzung des Verkaufsvorganges ermöglicht es nun, innerhalb des Rechnungswesens weitere Operationen vorzunehmen, die Kaufleuten tiefgehende Erkenntnisse für die Gestaltung von realen Leistungsprozessen bieten (siehe die pragmatische Funktion von Modellen). So lässt sich mit Hilfe der Informationen aus dem Buchungssatz unter Hinzunahme weiterer Daten beispielsweise der Deckungsbeitrag des Verkaufsvorganges berechnen. Damit Kaufleute diese im Rechnungswesen erstellte Information für reale Zwecke nutzen können, müssen sie dem konkreten Deckungsbeitrag jedoch wieder eine inhaltliche Bedeutung zuweisen, d. h. eine Übersetzung aus dem Rechnungswesen in die Realität vornehmen. Dieser Rückbezug könnte beispielsweise darin bestehen, dass mit dem Verkaufsvorgang insgesamt mehr Geldzuflüsse als Geldabflüsse in das Unternehmen verbunden sind, der Vorgang also zum Erfolg des Unternehmens beigetragen hat.

Die Analyse der Modellierungskompetenz ist sowohl als Zielgröße des Unterrichts als auch zur Ermittlung der Ausgangsbedingungen für den Rechnungswesenunterricht aus fachdidaktischer Perspektive relevant. So wird als Zielgröße für den Rechnungswesenunterricht seit längerem nicht mehr ein wirtschaftspropädeutischer oder situationsbezogener-funktionaler Ansatz, sondern ein wirtschaftsinstrumenteller Ansatz befürwortet, der darauf abzielt, den Lernenden mit dem Rechnungswesen ein Werkzeug zur Gestaltung einer Unternehmung zur Verfügung zu stellen (vgl. PREISS/TRAMM 1996b, S. 225 f.). Genau dies ist die Aufgabe von Modellen in der Praxis, indem sie z. B. die Planung, Gestaltung und Verbesserung von Systemen und Prozessen unterstützen (vgl. HOLZMÜLLER/BANDOW 2010, S. VII f.). Ziel ist es u. a. mit Modellen Erkenntnisse zu gewinnen, die für Entscheidungen von Relevanz sind (vgl. TÖLLNER et al. 2010, S. 17). Dies zeigt sich auch konkret in der Diskussion zum Rechnungswesenunterricht, wenn verschiedene Perspektiven auf ein Unternehmen und die Abbildung dieser Perspektiven im Rechnungswesen diskutiert werden. So wird in der Bilanzmethode die Sichtweise von Eigenkapitalgeber(inne)n und im wirtschaftsinstrumentellen Rechnungswesenunterricht die Perspektive eines Unternehmens eingenommen, was zu verschiedenen Modellen von Unternehmensprozessen führt (vgl. ERNST 2014; STOMMEL 2014). Gerade weil Modelle stets an einen Zweck gebunden sind und sich die Modellierungskompetenz darin äußert, verschiedene Modelle zu vergleichen und zu beurteilen (vgl. dazu GREEFRATH et al. 2013, S. 18), bietet das Konstrukt der Modellierungskompetenz die Chance zu einer Weiterentwicklung der Didaktik des Rechnungswesens, indem die Lernenden dazu befähigt werden, dass Modell zu wählen, das ihnen zur Lösung eines konkreten Problems am hilfreichsten ist. Neben der Zielperspektive ist die Modellierungsfähigkeit der Lernenden aber auch eine wichtige Grundlage für die Unterrichtsplanung

und Erklärung von Lernprozessen. So zeigt beispielsweise die Studie von TÜRLING et al. (2011, S. 402), dass Lernenden oft bereits zu Beginn eines Modellierungsprozesses, also beim ökonomischen Verständnis und der Auswahl der Fachtermini, Fehler unterlaufen. Diese Problematik lässt sich mit Hilfe der Modellierungsfähigkeit erklären und gezielt im Unterricht durch Betrachtung der Modellierungsschritte aufarbeiten.

Im Vergleich zur Mathematikdidaktik gibt es in der Rechnungswesendidaktik, speziell im Bereich der Buchführung, jedoch eine Besonderheit. So sind in der Buchführung diese Übersetzungen mit Hilfe von Belegen vorzunehmen. Entsprechend gehen SEIFRIED, TÜRLING und WUTTKE (2010, S. 145 f., S. 147) davon aus, dass zunächst der reale Vorgang mit Hilfe passender Belege zu *repräsentieren* ist. Die Darstellung eines Vorganges mit Hilfe eines Beleges stellt den ersten Abstraktionsschritt eines realen Prozesses dar, bevor auf dieser Grundlage weitere Formalisierungen vorgenommen werden können. Abstrahieren bedeutet hier, bestimmte individuelle Eigenschaften eines konkreten Phänomens wegzulassen, sodass die verbleibenden Eigenschaften auf mehrere Phänomene zutreffen (vgl. TÖLLNER et al. 2010, S. 18).

Aufgrund dieses Zwischenschrittcharakters haben Belege in neueren Arbeiten zur Rechnungswesendidaktik eine besondere Relevanz. So kommt ihnen beispielsweise im wirtschaftsinstrumentellen Ansatz eine Mittlerfunktion zwischen realen Prozessen einerseits und dem Informationssystem andererseits zu (vgl. PREISS/TRAMM 1996b, S. 252 ff.). Insgesamt übernehmen Belege vier Aufgaben (vgl. PREISS/TRAMM 1996b, S. 253):

- 1) Belege abstrahieren einen realen Vorgang und nehmen bereits eine erste Reduzierung auf die relevanten Aspekte des Vorganges vor.
- 2) Belege verdichten einen Vorgang durch Verwendung der Fachsprache.
- 3) Belege dokumentieren einen Vorgang.
- 4) Belege stellen den Ausgangspunkt für weitere Abstraktions- und Verdichtungsvorgänge dar.

Übereinstimmend mit dieser hohen Bedeutung leiten SEIFRIED und SEMBILL (2004, S. 378) die konsequente Verwendung von Belegen als didaktisches Prinzip aus ihrem Projekt zum selbstorganisiertem Lernen im Rechnungswesen ab. Die Verwendung von Belegen ist jedoch noch aus einem weiteren Grund für den geplanten Modellierungstest relevant. So werden in der Buchhaltung reale Vorgänge vor allem auf der Grundlage von Belegen erfasst. Die Verwendung von Belegen im Modellierungstest sorgt somit für eine Anschlussfähigkeit an die buchhalterische Praxis.

2.2 Konstruktion des Erhebungsinstruments

Um den Modellierungsprozess aus Abbildung 1 vollständig abzudecken und zudem die Relevanz von Belegen adäquat zu berücksichtigen, wurden insgesamt vier verschiedene Typen an Testaufgaben für den Bereich erfolgswirksamer Vorgänge konzipiert.

Aufgaben der Kategorie 1 (siehe dazu Abbildung 2) präsentieren den Proband(inn)en photorealistische Abbildungen von möglichen Prozessen und Situationen in Zusam-

Aufgabe In welcher der folgenden Situationen entstehen Aufwendungen für ein Unternehmen? Kreuzen Sie diese an.

		
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Abb. 2: Beispieltetaufgabe der Kategorie 1

menhang mit erfolgswirksamen Vorgängen. Im Sinn des Modells der Grundvorstellungen geben diese Aufgaben eine „reale“ Situation vor, die in das Rechnungswesen zu übersetzen ist, und decken den grau hinterlegten Bereich in Abbildung 1 ab.

Aufgabe Welche der folgenden Buchungssätze beschreiben einen Aufwand für ein Unternehmen? Kreuzen Sie diese an.

		Soll	Haben
<input type="checkbox"/>	Abschreibungen an technische Anlagen und Maschinen	12.500,00 €	12.500,00 €
<input type="checkbox"/>	Rohstoffe Umsatzsteuer an Bank	36.000,00 € 6.840,00 €	42.840,00 €

Abb. 3: Beispieltetaufgabe der Kategorie 2

Bei den Aufgaben der Kategorie 2 ist die Situation hingegen genau umgekehrt (siehe dazu Abbildung 3). Diese Aufgaben decken den weiß hinterlegten Bereich in Abbildung 1 ab und geben ein abstraktes Objekt aus dem Rechnungswesen vor, welches mit inhaltlicher Bedeutung zu versehen ist.

Eine direkte Verbindung wird hingegen mit den Aufgaben der Kategorien 3 und 4 vorgenommen. Beispiele für diese Aufgaben zeigen die Abbildung 4 und 5.

So wird mit den Aufgaben der Kategorie 3 ein realer Vorgang durch ein Foto vorgegeben und eine Auswahl dazu (nicht-)passender Belege dargestellt. Damit deckt dieses Format sowohl den weiß als auch grau hinterlegten Bereich des Modells der Grundvorstellungen ab (siehe dazu Abbildung 1). Lernende sind hier gefordert, die vorgegebene

Aufgabe Im Folgenden sehen Sie eine Reihe von Fotos, die Aktivitäten in einem Unternehmen zeigen. Bitte wählen Sie zu jedem Foto die Belege aus, die zu der Situation passen. Es können mehrere Belege zu einem Foto passen.



Quittung

Währung: Betrag in Zahlen
EUR: 43,85

Netto: 8,33
Steuer: 35,52
Gesamt: 43,85

zweihundfünfzig Euro und achtzehn Cent

Gewürzwunder e. K.

Esig und Öl Geschenkkorb

Oldenburg, Datum: 16.09.2018

Gewürzwunder e. K.

Steuerkonto

Pos.	Menge	Anlassbeschreibung	Einheitspreis (€)	Gesamt (€)
1		Esig und Öl Geschenkkorb	43,85	43,85
		Kartenzahlung		43,85
		20% USt		8,33
		Buchungsbetrag		52,18

Datum	Umsatz	Betrag
02.10.2018	Bank Z	3.540,69 +
01.09.	Strom September	34,00 -
01.09.	Wafel Hofwerkstraße 4	600,00 -
14.09.	Leuchtröhre, Kapazitoristik GmbH, AH 4965	123,99 -
13.09.	Rechnung 18/209 Gewürzwunder e. K.	52,18 -
30.09.	Gehalt Monat September	2564,79 +
	Zwischensumme	5.246,59 +

Abb. 4: Beispieltestaufgabe der Kategorie 3

Aufgabe Im Folgenden sehen Sie einen Buchungssatz und mehrere Bilder. Bitte kreuzen Sie jeweils die Bilder an, die zu dem Buchungssatz passen. Es können mehrere Bilder passen. Es kann aber auch vorkommen, dass kein Bild passend ist.

Buchungssatz		Soll	Haben
Gehälter und Löhne an Bank		6.300,00 €	6.300,00 €







Abb. 5: Beispieltestaufgabe der Kategorie 4

Situation und die Belege zu interpretieren sowie beide Bereiche korrekt aufeinander zu beziehen.

Auf dem gleichen Prinzip basieren die Aufgaben der Kategorie 4. Der Unterschied liegt hier darin, dass nicht Belege, sondern Buchungssätze im Fokus der Betrachtung stehen, der Abstraktionsgrad also höher angesetzt ist. Zudem ist der Ausgangspunkt nicht ein „realer Vorgang“, sondern ein abstraktes Objekt. Die Lernenden sind hier gefordert, abstraktes Objekt und dargestellte reale Vorgänge korrekt aufeinander zu beziehen.

In der Gesamtheit decken die vier Aufgabenkategorien vollständig den Modellierungsprozess ab, wie er in Abbildung 1 dargestellt ist. Sie fokussieren damit vor allem auf den Wechsel zwischen Rechnungswesen und realen Phänomenen, wie er sinngemäß bei der mathematischen Modellierungskompetenz von besonderer Bedeutung ist (vgl. GREEFRATH et al. 2013, S. 18). Da das Modell der Grundvorstellungen (BERDING 2019) bzw. der Prozess des Modellierens nach SCHUPP (1988) Ausgangspunkt der Testentwicklung ist, sind die einzelnen Testitems nach dem sog. rationalen Konstruktionsansatz konzipiert (vgl. dazu z. B. BÜHNER 2011, S. 93; KAPLAN/SACCUZZO 2013, S. 344). Insgesamt liegen 41 Testitems für die Validierung vor. Diese verteilen sich auf sieben Aufgaben der Kategorie 1, zwölf der Kategorie 2, sieben der Kategorie 3 und 15 der Kategorie 4.

2.3 Hypothesen für die Testvalidierung

Um den entwickelten Test einer strengen Evaluation zu unterziehen, kommt neben den klassischen Kriterien zur Reliabilität vor allem die Prüfung der Annahmen eines dichotomen Raschmodells zum Einsatz (vgl. dazu z. B. KOLLER/ALEXANDROWICZ/HATZINGER 2012; STROBL 2015). Zudem wird ein besonderer Fokus auf die Prüfung der Validität gelegt, die sich in die Inhaltsvalidität, Kriteriumsvalidität und Konstruktvalidität unterteilen lässt (vgl. DÖRING/BORTZ 2016, S. 470 f.). Während die Inhaltsvalidität nicht statistisch geprüft werden kann, sondern sich danach bemisst, in welchem Maß inhaltliche und theoretische Überlegungen leitend bei der Testkonstruktion waren, prüft die Kriteriumsvalidität Zusammenhänge zwischen den Testwerten und *direkt* beobachtbaren Merkmalen ab (vgl. DÖRING/BORTZ 2016, S. 470). Da direkt beobachtbare Merkmale zur Modellierungsfähigkeit nicht vorliegen, konzentriert sich die Prüfung in dieser Studie auf die Konstruktvalidität. „Ein Test ist konstruktvalid, wenn aus dem zu messenden Zielkonstrukt theoretisch und/oder empirisch gut fundierte Hypothesen ableitbar sind, die anhand der Testergebnisse bestätigt werden können.“ (DÖRING/BORTZ 2016, S. 471). Folglich sind für die Evaluation Prüfungshypothesen zu formulieren.

Zusammenhänge zu Grundvorstellungen: Das Modell der Grundvorstellungen postuliert, dass Grundvorstellungen ein Scharnier zwischen dem abstrakten System Mathematik bzw. Rechnungswesen einerseits und realen Phänomenen andererseits bilden (vgl. z. B. BERDING 2019, S. 90 ff.; VOM HOFE 2003). Sie beschreiben inhaltliche Interpretationen von abstrakten Objekten, Begriffen und Verfahren (vgl. PREDIGER 2009, S. 170) und besitzen fünf zentrale Eigenschaften (vgl. im Folgenden BERDING 2018, S. 7; VOM HOFE 1992, S. 347; 1995, S. 97 f.; 2003, S. 6):

- *Sinnkonstituierung:* Grundvorstellungen geben abstrakten Begriffen, Symbolen und Konstrukten eine Bedeutung.
- *Repräsentation:* Grundvorstellungen dienen der Entwicklung visueller, bildlicher bzw. mentaler Vorstellungen, die ein mentales Handeln mit den Begriffen, Symbolen und Konstrukten erlauben.

- *Anwendung*: Grundvorstellungen erlauben es dem Individuum, Begriffe, Symbole und Konstrukte durch das Erkennen sachlicher bzw. formeller Strukturen auf die Wirklichkeit anzuwenden.
- *Netzartig*: Verschiedene Grundvorstellungen bilden ein veränderbares, flexibles Netz an Vorstellungen.
- *Vielfältig*: Es sind stets mehrere Vorstellungen notwendig, um einen Begriff, ein Symbol oder ein Konstrukt vollständig abzudecken.

Grundvorstellungen gelten damit als eine Voraussetzung für den zielführenden Umgang mit abstrakten Begriffen (vgl. BÜCHTER/HENN 2010, S. 30). Allerdings sind Grundvorstellungen nicht mit dem Fachwissen einer Person gleichzusetzen. Vielmehr stellen sie im Sinne der Konzeptwechsel- bzw. Konzeptrekonstruktionsforschung individuelle Konzepte zu abstrakten Objekten dar (vgl. BERDING/RIEBENBAUER/SLEPCEVIC-ZACH 2019).

Studien aus der Mathematikdidaktik zeigen, dass ca. 50 % der beim Lösen von Aufgaben auftretenden Fehler auf inadäquate oder fehlerhafte Grundvorstellungen zurückzuführen sind (vgl. VOM HOFE et al. 2009, S. 140 ff.; PEKRUN et al. 2006, S. 35 f.). Studien aus dem Rechnungswesen belegen weiterhin, dass Grundvorstellungen nach Berücksichtigung von Einflussfaktoren wie Lernstrategien oder Motivation zwischen 15 % und 30 % der Varianz in den Rechnungswesennoten aufklären können (vgl. BERDING 2019, S. 225). Vor diesem Hintergrund sollte es mit Hilfe des Modellierungstests möglich sein, die folgenden Hypothesen bei einer validen Messung zu bestätigen:

- (G1) *Grundvorstellungen zum „Aufwand“ und „Ertrag“ erklären die Variation in der Modellierungsfähigkeit.*
- (G2) *Formale Buchungsstrategien stehen in keinem Zusammenhang zur Modellierungsfähigkeit.*

Zusammenhänge zu Motivation: Die Metastudie von SCHIEFELE, KRAPP und SCHREYER (1993) mit 127 Stichproben belegt einen mittelstarken Einfluss von Interesse auf die Schulleistung. Ebenso kommt die Meta-Studie von CERASOLI, NICKLIN und FORD (2014) auf der Grundlage der Daten von ca. 212.000 Teilnehmenden zu dem Schluss, dass die intrinsische Motivation einen mittleren bis starken Einfluss auf die Qualität einer Leistung und nur einen schwachen bis mittleren Einfluss auf die Quantität von Leistung nimmt. Die Ergebnisse zeigen ferner auch, dass intrinsische Motivation besser zur Vorhersage qualitativer Leistung geeignet ist als extrinsische Motivation, während extrinsische Motivation die Quantität der Leistung besser vorhersagen kann als intrinsische Motivation (vgl. CERASOLI/NICKLIN/FORD 2014, S. 994 f.). Folglich ist davon auszugehen, dass die Motivation in Zusammenhang mit den Testwerten des Modellierungstests stehen sollte.

Eine genauere Klärung des Zusammenhangs zwischen Motivation und Modellierungstest kann die Selbstbestimmungstheorie der Motivation leisten, da sie eine feine Unterscheidung der Qualität von Motivation erlaubt. Sie umfasst neben der intrinsischen Motivation verschiedene Varianten der extrinsischen Motivation, die sich nach

dem Grad der Selbstbestimmung unterscheiden (vgl. DECI/RYAN 1993; 2012; RYAN/DECI 2002). Während die externe und introjizierte Regulation als eher fremdbestimmt wahrgenommen werden, gelten die identifizierte und integrierte Regulation als selbstbestimmt. Ebenso stellt die intrinsische Motivation eine selbstbestimmte Motivationsvariante dar.

Die Selbstbestimmungstheorie der Motivation postuliert drei Bedürfnisse, die erfüllt sein müssen, damit ein Individuum eine (eher) selbstbestimmte motivationale Regulation entwickelt: (1) Das Bedürfnis nach Autonomie, d. h. selbstbestimmt über die eigenen Handlungen entscheiden zu können, (2) das Bedürfnis nach Kompetenzerleben, d. h. sich selbst als wirksam und fähig zur Beeinflussung der sozialen Umwelt wahrzunehmen, sowie (3) das Bedürfnis nach sozialer Eingebundenheit, d. h. sich mit anderen Menschen verbunden zu fühlen (vgl. DECI & RYAN 1993, S. 229; 2012, S. 87; RYAN & DECI 2002, S. 7 f.). Verschiedene Studien konnten in diesem Zusammenhang zeigen, dass die Motivation umso mehr selbstbestimmt bzw. umso weniger fremdbestimmt ist, je mehr diese drei Grundbedürfnisse erfüllt werden (vgl. z. B. BERDING 2019; KRAMER/PRENZEL/DRECHSEL 2000). In der Konsequenz ist zu erwarten, dass eine höhere Modellierungsfähigkeit es einem Individuum erlaubt, häufiger und besser Aufgaben erfolgreich zu lösen. Das Individuum kann sich somit als wirksam erleben und erfüllt sein Bedürfnis nach Kompetenzerleben. Dadurch ist eine Bedingung für die Entwicklung einer eher selbstbestimmten Motivation erfüllt, die ihrerseits bessere Leistungen bzw. Erfolge beim Erwerb der Modellierungskompetenz ermöglicht. Es entsteht ein sich verstärkender Spiralprozess. Demnach sollten sich bei valider Messung folgende Hypothesen bestätigen lassen:

- (M₁) *Je mehr eine Person Rechnungswesen fremdbestimmt lernt, desto geringer ist ihre Modellierungsfähigkeit.*
- (M₂) *Je mehr eine Person Rechnungswesen selbstbestimmt lernt, desto größer ist ihre Modellierungsfähigkeit.*

Zusammenhänge zu Noten: Die Modellierungsfähigkeit als ein Maßstab für den Lernerfolg im Rechnungswesenunterricht sollte zudem in einem Zusammenhang mit den Noten stehen, die ebenfalls als ein Lernerfolgsindikator angesehen werden können. Es ist von folgender Hypothese auszugehen:

- (L) *Je besser die Modellierungsfähigkeit einer Person ist, desto bessere Noten erzielt sie im Unterricht.*

Allerdings darf der Zusammenhang bei dieser Hypothese nicht zu stark sein, da andernfalls der zu entwickelnde Test das Gleiche messen würde wie die Noten. Noten erfassen jedoch auch andere Konstrukte, z. B. die mündliche Leistung im Unterricht oder das Fachwissen. Im Sinn der Diskriminanzvalidität, die verstanden werden kann als „the degree to which a construct is discriminable (e. g., uncorrelated) from, and non-redundant with, other constructs“ (LOUNSBURY/GIBSON/SAUDARGAS 2006, S. 139), darf folglich kein starker Zusammenhang vorliegen. Die zur Validierung genutzten Daten und Methoden werden im folgenden Abschnitt beschrieben.

3 Methodik

3.1 Stichprobe

Zur Entwicklung des Messinstruments nahmen insgesamt 443 Lernende aus 24 Klassen der beruflichen Bildung teil. Die 227 Frauen und 216 Männer waren durchschnittlich $M = 19,63$ ($SD = 2,851$) Jahre alt. 145 von ihnen besuchten einen schulischen Bildungsgang, d. h. 29 besuchten die Fachoberschule Wirtschaft, 97 das Berufliche Gymnasium Wirtschaft und 19 die Berufsfachschule Wirtschaft. Die verbleibenden 298 Lernenden durchliefen eine duale Ausbildung. Es handelte sich dabei um 45 angehende Kaufleute für Büromanagement, 86 Groß- und Außenhandelskaufleute sowie um 167 angehende Industriekaufleute. Von diesen waren zum Zeitpunkt der Erhebung 145 im ersten, 67 im zweiten und 86 im dritten Jahr ihrer Ausbildung. In der gesamten Stichprobe verfügen drei Teilnehmende über einen Hauptschulabschluss, 214 über die mittlere Reife, 80 über die Fachhochschulreife und 140 über die allgemeine Hochschulreife (6 fehlende Werte).

Um die Eigenschaften des Tests auch auf ihre Stabilität zu prüfen, wurde diese Gesamtstichprobe mittels Zufallsauswahl in zwei Stichproben aufgeteilt. Dabei wurde die Zufallsauswahl so vorgenommen, dass die beiden Typen an Bildungsgängen (dual vs. schulisch) in beiden Stichproben vertreten sind. Substichprobe 1 umfasst insgesamt 105 Frauen und 111 Männer mit einem Durchschnittsalter von $M = 19,58$ ($SD = 2,538$) Jahren. 72 gehören einem schulischen (Fachoberschule Wirtschaft, berufliches Gymnasium Wirtschaft) und 144 einem dualen Bildungsgang an (Kaufleute für Büromanagement, Groß- und Außenhandel sowie Industriekaufleute). Es handelt sich um Lernende aus insgesamt 12 Klassen.

Substichprobe 2 umfasst 122 Frauen und 105 Männer mit einem Durchschnittsalter von $M = 19,67$ ($SD = 3,126$) Jahren. 73 absolvieren einen schulischen Bildungsgang (Berufliches Gymnasium Wirtschaft, Berufsfachschule Wirtschaft) und 154 einen dualen Bildungsgang (Kaufleute für Büromanagement, Groß- und Außenhandel sowie Industriekaufleute). Es handelt sich um Lernende aus insgesamt 12 Klassen.

3.2 Erhebungsinstrumente

Neben dem zu entwickelnden Modellierungstest wurden weitere Daten zur Validierung des neuen Messverfahrens erhoben, die auf eine Prüfung der abgeleiteten Validitätshypothesen abzielen.

Motivationsfragebogen: Zur Erhebung der psychologischen Grundbedürfnisse und verschiedenen Motivationsvarianten auf der Grundlage der Selbstbestimmungstheorie der Motivation (DECI/RYAN 1993; 2012; RYAN/DECI 2002) kam der von PRENZEL et al. (1996) entwickelte Fragebogen für die kaufmännische Bildung zum Einsatz. Der Fragebogen umfasst insgesamt 37 Aussagen (z. B. „Beim Lernen von Rechnungswesen verging die Zeit wie im Flug.“). Die Teilnehmenden sollten auf einer Skala von 0 = „nie“

This material is under copyright. Any use outside of the narrow boundaries of copyright law is illegal and may be prosecuted.

This applies in particular to copies, translations, microfilming as well as storage and processing in electronic systems.

© Franz Steiner Verlag, Stuttgart 2020

bis 5 = „sehr häufig“ bewerten, wie oft die jeweilige Aussage auf den von ihnen erlebten Rechnungswesenunterricht zutrifft. Der Fragebogen erfasst die drei psychologischen Grundbedürfnisse „Soziale Eingebundenheit“, „Kompetenzerleben“ und „Autonomieerleben“ ebenso wie die eher fremdbestimmten (Amotiviert, externe Regulation, injizierte Regulation) und selbstbestimmten Motivationsvarianten (identifizierte Regulation, intrinsische Regulation, interessierte Regulation).

Lernerfolg: Als Indikator für den Lernerfolg wurden die Noten im Rechnungswesen erhoben. Zu diesem Zweck sollten die Kaufleute für Büromanagement ihre Zensuren in den Lernfeldern 06 „Wertströme erfassen und beurteilen“, 10 „Wertschöpfungsprozesse erfolgsorientiert steuern“, die Industriekaufleute in den Lernfeldern 03 „Wertströme und Werte erfassen und dokumentieren“, 04 „Wertschöpfungsprozesse analysieren und beurteilen“, 08 „Jahresabschluss analysieren und bewerten“ sowie 11 „Investitions- und Finanzierungsprozesse planen“, die angehenden Groß- und Außenhandelskaufleute in den Lernfeldern 04 „Geschäftsprozesse als Werteströme erfassen, dokumentieren und auswerten“ sowie 11 „Unternehmensergebnisse aufbereiten, bewerten und nutzen“ angeben. Lernende der Fachoberschule Wirtschaft wurden gebeten ihre Ergebnisse in dem Lerngebiet „Werte und Werteströme unter Einsatz einer integrierten ERP-Software erfassen, darstellen und auswerten“, Lernende des Beruflichen Gymnasiums im Fach BRC und im Bereich Buchführung und Lernende der Berufsfachschule Wirtschaft im Lernfeld 04 „Wertströme dokumentieren und auswerten“ zu berichten. Wurde mehr als eine Note angegeben, dient der auf ganze Zahlen gerundete Mittelwert als Indikator für den Lernerfolg.

Grundvorstellungen: Schließlich wurden die Grundvorstellungen der Lernenden mittels Diagnoseaufgaben für die Begriffe „Aufwand“ und „Ertrag“ nach den von BERDING (2019) entwickelten Prinzipien erhoben. Die Diagnoseaufgabe zum Begriff „Aufwand“ präsentiert den Lernenden mit einem Buchungssatz ein abstraktes Objekt aus dem Rechnungswesen. Die Lernenden sollen hier ausführlich eine Situation oder Tätigkeit aus einem Unternehmen schildern, die zu dem Buchungssatz passen könnte und erklären, wie sich die Tätigkeit auf den Unternehmenserfolg auswirkt. Die Aufgabe für den Begriff „Ertrag“ gibt eine Ausgangsrechnung vor, die von den Lernenden in das Rechnungswesen zu übersetzen ist. Dabei werden die Teilnehmenden aufgefordert, ihre Gedanken und Bearbeitungsschritte einer neuen Auszubildenden zu erläutern. Die inhaltsanalytische Auswertung der individuellen Lösungen erfolgte mit dem Kodierschema von BERDING (2019). Die konkrete Auswertung wurde von zwei Codierer(inne)n vorgenommen und erzielte in einem Pretest hohe Übereinstimmungswerte für die einzelnen Kategorien, wie die Tabelle 1 zeigt. So liegt der Anteil übereinstimmender Codierungen bei mindestens 77,8 %.

Die auf dieser Datengrundlage erzielten Ergebnisse werden im Folgenden dargestellt.



Tab. 1: Inter-Coder-Reliabilität für die inhaltsanalytische Auswertung der Lösungen zu den Diagnoseaufgaben ($N = 18$) als übereinstimmende Einschätzungen

Kategorien für den Begriff „Aufwand“		Kategorien für den Begriff „Ertrag“	
<i>Grundvorstellungen</i>		<i>Grundvorstellungen</i>	
Beschaffungsvorstellung	.944	Liquiditätszuflussvorstellung	.889
Inputvorstellung	.778	Liquiditätsforderungsvorstellung	.833
Verbrauchsvorstellung	1.000	Wertabflussvorstellung	.833
Zerstörungsvorstellung	1.000	Bleibevorstellung	1.000
Umwandlungsvorstellung	.944	Differenzvorstellung	.889
Liquiditätsabflussvorstellung	.944	Erfolgssteigerungsvorstellung	.944
Lagerentnahmevorstellung	.889	Eigenkapitalsteigerungsvorstellung	1.000
Eigenkapitalminderungsvorstellung	1.000	Umsatzerlösvorstellung	.778
Erfolgsminderungsvorstellung	.944		
Neutralitätsvorstellung	.778		
Materialentnahmevorstellung	1.000		
Lagerbewertungsvorstellung	1.000		
<i>Formale Strategien</i>		<i>Formale Strategien</i>	
Kontenart	1.000	Kontenart	.944
Kontenseite	.944	Kontenseite	.889
Gegenkonto	1.000	Gegenkonto	.944
Abschluss über GuV	1.000	Abschluss über GuV	.944
Gegenpol	1.000	Gegenpol	1.000

4 Ergebnisse

4.1 Prüfung der Annahmen des Raschmodells in Teilstichprobe 1 und Optimierung des Tests

Die Prüfung der Annahmen des Rasch-Modells erfolgt mit dem Paket *eRm* (extended Rasch modeling) von MAIR, HATZINGER und MAIER (2018) für das Programm R. Im Rahmen der Modellprüfung müssen mehrere Hypothesen am gleichen Datensatz geprüft werden, was eine Justierung des Alpha-Niveaus erforderlich macht (vgl. KOLLER/ALEXANDROWICZ/HATZINGER 2012, S. 162 ff.). Aus diesem Grund wird der Empfehlung von KOLLER, ALEXANDROWICZ und HATZINGER (2012, S. 162) gefolgt und das globale α -Niveau auf 10 % festgelegt. Die jeweiligen Grenzwerte des Alpha-Niveaus werden passend zu den Tests in den folgenden Tabellen offengelegt.

Während der Prüfung dieser Annahmen mussten schrittweise Aufgaben aus dem Test entfernt werden, da sie nicht mit den Eigenschaften eines Rasch-Modells vereinbar

waren. Es wurden so lange Aufgaben aus dem Test entfernt, bis eine für die Teilstichprobe 1 raschkonforme Zusammenstellung von Aufgaben vorlag. Aus Platzgründen werden nur für diese finale Lösung die statistischen Kennwerte präsentiert. Informationen zu den entfernten Aufgaben werden an späterer Stelle dieses Beitrages gegeben.

Die Darstellung beginnt mit den parametrischen Testverfahren. Bei Gültigkeit des Rasch-Modells sind die Eigenschaften der Aufgaben unabhängig von den Personen. Folglich dürfen sich die Parameter der Items nicht systematisch zwischen verschiedenen Personengruppen unterscheiden (sog. Subgruppeninvarianz) (vgl. KOLLER/ALEXANDROWICZ/HATZINGER 2012, S. 62; STROBL 2015, S. 42). Um dies zu überprüfen, kommt der Andersen-Test für drei Kriterien zur Anwendung. Es wird getestet, ob das Geschlecht, die Leistungsfähigkeit der Personen (Gruppenbildung über Median-Split) sowie die Zugehörigkeit zu einem Bildungsgang (duale vs. schulische Bildungsgänge) unterschiedliche Itemparameter bewirken (vgl. z. B. BÜHNER 2011, S. 531 f.). Da drei Kriterien geprüft werden, beträgt das zu verwendende Alpha-Niveau für die globale Prüfung rund 3,33 % (vgl. KOLLER/ALEXANDROWICZ/HATZINGER 2012, S. 168). Die Ergebnisse sind in Tabelle 2 zusammengefasst.

Tab. 2: Ergebnisse der parametrischen Tests für Teilstichprobe 1 ($n = 216$)

	Leistungsfähigkeit der Personen		Geschlecht der Personen		Bildungsgang der Personen		Aufgabenkategorie
	Globale Kennwerte, Grenzwert für $\alpha = .033$						
	$\chi^2(15) = 15.936, p = .386$		$\chi^2(15) = 20.176, p = .165$		$\chi^2(15) = 7.336, p = .948$		
	Lokale Kennwerte, Grenzwert für $\alpha = .0021$						
	z-Werte	p-Werte	z-Werte	p-Werte	z-Werte	p-Werte	
Ao1	0.862	.389	1.271	.204	-1.042	.298	2
Ao2	0.312	.755	-1.149	.250	0.653	.514	2
Ao3	-0.703	.482	0.865	.387	-1.033	.302	2
Ao4	1.986	.047	0.378	.706	0.322	.747	2
Ao5	-0.169	.866	-0.228	.820	-0.045	.964	2
Ao6	1.148	.251	-1.104	.270	1.131	.258	1
Ao7	-1.462	.144	0.876	.381	0.022	.982	1
Ao8	1.138	.255	0.446	.656	0.633	.507	1
Ao9	-0.417	.677	1.730	.084	-0.136	.892	1
A10	0.164	.870	1.456	.146	-0.134	.893	3
A11	1.335	.182	-1.862	.063	-0.404	.686	3
A12	0.998	.318	-1.838	.066	-0.308	.758	3
A13	-0.609	.542	-0.327	.744	-0.718	.473	3
A14	0.354	.723	-1.414	.157	-0.925	.355	4
A15	-0.481	.630	0.858	.391	1.341	.180	4
A16	-1.838	.066	-0.625	.532	0.078	.938	4

Martin-Löf-Test: $\chi^2(63) = 60.454, p = .568$, Grenzwert für $\alpha = .10$

Hinweise: Die Justierung des Grenzwertes für α erfolgte nach KOLLER, ALEXANDROWICZ und HATZINGER (2012, S. 168 f.)

Tabelle 2 zeigt, dass die Andersen-Tests auf globaler Ebene nicht signifikant werden. Dies bedeutet, dass sich die Aufgabeneigenschaften zwischen Leistungsstarken und Leistungsschwachen, zwischen Frauen und Männern und zwischen Mitgliedern eines dualen oder vollzeitschulischen Bildungsgangs *nicht* systematisch unterscheiden. Diesen Befund bestätigt auch der Wald-Test auf lokaler Ebene für die einzelnen Aufgaben. So ist der Test für kein Item signifikant.

Mit Hilfe des Martin-Löf-Tests werden nicht Personen in Gruppen aufgeteilt, sondern Aufgaben und getestet, ob diese zu den gleichen Personenparametern führen, sog. Itemhomogenität (vgl. BÜHNER 2011, S. 538; KOLLER/ALEXANDROWICZ/HATZINGER 2012, S. 90). Ein Vergleich leichter und schwerer Aufgaben (Median-Split) zeigt ein nicht signifikantes Ergebnis. Dies bedeutet, dass leichte und schwere Aufgaben nicht zu unterschiedlichen Personenparametern führen.

Um diese Resultate weiter abzusichern, zeigt Tabelle 3 die Ergebnisse quasi-exakter Tests sowie die Prüfungsergebnisse weiterer Eigenschaften des Raschmodells.

Tab. 3: Ergebnisse der quasi-exakten Tests für Teilstichprobe 1 ($n = 216$)

	Leistungsfähigkeit der Personen		Geschlecht der Personen		Bildungsgang der Personen		
Globale Kennwerte, Grenzwert für $\alpha = .033$, Prozedur T10							
	p = .502		p = .162		p = .963		
Lokale Kennwerte, Grenzwert für $\alpha = .0021$, Prozedur T4							
	p-Werte		p-Werte		p-Werte		
	Für Leistungs- schwächere leichter als für	Für Leistungs- starke leichter als für	Für Männer leichter als für Frauen	Für Frauen leichter als für Männer	In der dualen Ausbildung leichter als in schulischen Bildungsgängen	In schulischen Bildungsgängen leichter als in der dualen Ausbildung	Trennschärfe Grenzwert für $\alpha = .006$ Prozedur Tpbis
A01	.318	.774	.122	.922	.877	.203	.320
A02	.540	.617	.897	.179	.295	.816	.906
A03	.868	.210	.208	.871	.873	.206	.675
A04	.074	.960	.379	.723	.406	.703	.101
A05	.724	.436	.597	.551	.551	.592	.611
A06	.226	.867	.879	.189	.141	.918	.348
A07	.962	.093	.234	.877	.565	.597	.800
A08	.210	.861	.332	.750	.280	.835	.448
A09	.799	.303	.046	.976	.591	.528	.751
A10	.612	.503	.091	.951	.590	.525	.532
A11	.199	.884	.980	.045	.705	.432	.300
A12	.283	.815	.980	.036	.665	.451	.306
A13	.829	.291	.685	.438	.791	.324	.613
A14	.519	.590	.924	.121	.853	.249	.077
A15	.805	.388	.251	.876	.121	.946	.619
A16	.989	.023	.767	.339	.474	.648	.927
Itemhomogenität (Aufgabenschwierigkeit): p = .344, Grenzwert $\alpha = .10$, Prozedur Tmd							
Globale Prüfung der lokalen stochastischen Unabhängigkeit: p = .340, Grenzwert $\alpha = .10$, Prozedur T11							
Lokale Prüfung der lokalen stochastischen Unabhängigkeit:							
120 Paarvergleiche, min p > .0008, Grenzwert $\alpha = .0008$, Prozedur T1, keine signifikanten Paare							
Prüfung der Mehrdimensionalität:							
120 Paarvergleiche, min p > .0008, Grenzwert $\alpha = .0008$, Prozedur T1 m, keine signifikanten Paare							
Hinweise: Die Justierung des Grenzwertes für α erfolgte nach KOLLER, ALEXANDROWICZ und HATZINGER (2012, S. 168f.)							

Auch die quasi-exakten Tests zeigen auf lokaler und globaler Ebene keine signifikanten Resultate und bestätigen damit, dass sich die Aufgabeneigenschaften nicht zwischen Leistungsstarken und Leistungsschwachen, zwischen Frauen und Männern sowie zwischen Mitgliedern eines dualen und vollzeitschulischen Bildungsgangs unterscheiden. Die Eigenschaft der Itemhomogenität ist durch den nicht-signifikanten Test ebenfalls gegeben.

Tabelle 3 legt darüber hinaus noch weitere Informationen offen. So ist die Trennschärfe der einzelnen Aufgaben nicht signifikant, d. h. die Trennschärfe ist mit den Annahmen eines Raschmodells vereinbar. Ebenso führen die Prüfungsprozeduren T_1 und T_{11} zu nicht-signifikanten Ergebnissen und bestätigen das Vorliegen lokaler stochastischer Unabhängigkeit. Gleiches gilt für die Mehrdimensionalität mittels der Prozedur T_{1m} .

Insgesamt ist davon auszugehen, dass die 16 verbleibenden Aufgaben einen raschkonformen Test bilden. Vier der verbleibenden Aufgaben entfallen auf die Kategorie 1, fünf auf die Kategorie 2, vier auf die Kategorie 3 und drei auf die Kategorie 4. Da diese Eigenschaften speziell an der Teilstichprobe 1 entwickelt wurden, ist die Stabilität der Ergebnisse durch die gleichen Prüfschritte an der Teilstichprobe 2 zu validieren.

4.2 Prüfung des optimierten Tests von Teilstichprobe 1 an Teilstichprobe 2

Tabelle 4 stellt die Testergebnisse für die parametrischen Verfahren für Teilstichprobe 2 dar. Es wird deutlich, dass die Aufgaben auf lokaler wie globaler Ebene für die Teilkriterien „Leistungsfähigkeit“ und „Geschlecht“ nicht signifikant werden. Dies bestätigt die Ergebnisse von Stichprobe 1 und zeigt an, dass die Aufgabeneigenschaften sich nicht zwischen Leistungsstarken und Leistungsschwachen sowie zwischen Frauen und Männern unterscheiden. Allerdings wird der Andersen-Test für den Bildungsgang und der Wald-Test für die Aufgaben A_{04} , A_{05} , A_{08} und A_{11} signifikant, während A_{09} das Signifikanz-Niveau nur knapp verpasst. Dies bedeutet, dass diese Aufgaben sich in ihren Eigenschaften für Lernende dualer und vollzeitschulischer Bildungsgänge voneinander unterscheiden. Für das Kriterium „Bildungsgang“ sind die Annahmen des Rasch-Modells nicht erfüllt.

Zu den gleichen Analyseergebnissen kommen auch die quasi-exakten Testverfahren, deren Ergebnisse Tabelle 5 zusammenfasst. Dies gilt insbesondere für die Aufgaben A_{04} , A_{05} , A_{08} und A_{11} beim Wald-Test für die Bildungsgänge. Während die Itemhomogenität, die Trennschärfen der einzelnen Aufgaben und die Mehrdimensionalität erfüllt sind, zeigen die Prozeduren T_1 und T_{11} eine Verletzung der lokalen stochastischen Unabhängigkeit sowohl auf lokaler wie auch globaler Ebene an. Problematisch sind hier die Aufgabenpaare A_{02} und A_{05} sowie A_{06} und A_{09} .

Obwohl sich eine ganze Reihe der Annahmen des Raschmodells für die Teilstichprobe 2 bestätigen lassen, zeigen die Verletzungen bei der lokalen stochastischen Unabhängigkeit und der Subgruppeninvarianz bezüglich der Bildungsgänge die Notwendigkeit einer weiteren Optimierung des Tests an. Zu diesem Zweck sind die Aufgaben A_{04} ,

Tab. 4: Ergebnisse der parametrischen Tests für Teilstichprobe 2 ($n = 227$)

	Leistungsfähigkeit der Personen		Geschlecht der Personen		Bildungsgang der Personen		Aufgabenkategorie
Globale Kennwerte, Grenzwert für $\alpha = .033$							
	$\chi^2(15) = 11.720,$ $p = .700$		$\chi^2(15) = 5.590,$ $p = .986$		$\chi^2(15) = 72.605,$ $p = .000$		
Lokale Kennwerte, Grenzwert für $\alpha = .0021$							
	z-Werte	p-Werte	z-Werte	p-Werte	z-Werte	p-Werte	
Ao1	1.514	.130	0.308	.758	1.930	.054	2
Ao2	-0.912	.362	-1.091	.275	1.837	.066	2
Ao3	-0.731	.465	0.243	.808	0.194	.846	2
Ao4	1.268	.205	-0.637	.524	-4.664	.000*	2
Ao5	-1.199	.231	-0.557	.577	3.015	.003*	2
Ao6	-1.076	.282	1.136	.256	1.148	.251	1
Ao7	0.603	.547	-0.530	.596	-0.719	.472	1
Ao8	0.303	.762	0.977	.329	-3.253	.001*	1
Ao9	0.210	.834	0.138	.891	-2.963	.003	1
A10	-.170	.865	0.480	.631	-0.768	.442	3
A11	0.483	.629	-0.630	.529	3.086	.002*	3
A12	-0.921	.357	0.212	.832	-1.106	.269	3
A13	1.643	.100	0.402	.688	1.943	.052	3
A14	-0.380	.704	0.685	.493	-0.962	.336	4
A15	-0.069	.945	-0.321	.748	0.118	.906	4
A16	0.089	.929	-0.215	.830	0.721	.471	4

Martin-Löf-Test: $\chi^2(63) = 56.799, p = .696$, Grenzwert für $\alpha = .10$

Hinweise: Die Justierung des Grenzwertes für α erfolgte nach KOLLER, ALEXANDROWICZ und HATZINGER (2012, S. 168 f.)

Ao5, Ao8, Ao9, A11 sowie A12 aus dem Test auszuschließen. In der Folge beinhaltet der Test 10 korrekt zu lösende Aufgaben. Dabei entfallen zwei Aufgaben auf die Kategorie 1, drei Aufgaben auf die Kategorie 2, zwei Aufgaben auf die Kategorie 3 und drei Aufgaben auf die Kategorie 4. Die Eigenschaften dieser finalen Fassung werden im Folgenden Abschnitt tiefergehend analysiert.



Tab. 5: Ergebnisse der quasi-exakten Tests für Teilstichprobe 2 ($n = 227$)

Leistungsfähigkeit der Personen		Geschlecht der Personen		Bildungsgang der Personen			
Globale Kennwerte, Grenzwert für $\alpha = .033$, Prozedur T10							
p = .752		p = .981		p = .000			
Lokale Kennwerte, Grenzwert für $\alpha = .0021$, Prozedur T4							
p-Werte		p-Werte		p-Werte			
Für Leistungs- schwächere leichter als für Leistungs- starke	Für Leistungs- starke leichter als für Leistungs- schwächere	Für Männer leichter als für Frauen	Für Frauen leichter als für Männer	In der dualen Ausbildung leichter als in schulischen Bildungsgängen	In schulischen Bildungsgängen leichter als in der dualen Ausbildung	Trennschärfe für $\alpha = .006$ Prozedur Tpbis	
A01	.085	.945	.474	.634	.029	.988	.375
A02	.868	.245	.916	.168	.050	.975	.938
A03	.833	.266	.442	.656	.482	.636	.466
A04	.132	.919	.827	.260	1.000	.000*	.259
A05	.911	.157	.788	.328	.001*	1.000	.996
A06	.908	.153	.159	.886	.129	.920	.911
A07	.376	.790	.779	.360	.789	.343	.475
A08	.458	.650	.229	.842	.999	.002*	.166
A09	.494	.626	.531	.570	.999	.003	.538
A10	.638	.485	.397	.710	.764	.323	.390
A11	.429	.673	.791	.313	.001*	1.000	.039
A12	.884	.179	.529	.585	.870	.208	.514
A13	.083	.952	.415	.724	.029	.991	.170
A14	.697	.410	.312	.779	.848	.236	.876
A15	.615	.624	.731	.451	.536	.652	.270
A16	.523	.598	.645	.448	.280	.817	.295

Itemhomogenität (Aufgabenschwierigkeit): $p = .304$, Grenzwert $\alpha = .10$, Prozedur Tmd

Globale Prüfung der lokalen stochastischen Unabhängigkeit: $p = .000$, Grenzwert $\alpha = .10$, Prozedur T11

Lokale Prüfung der lokalen stochastischen Unabhängigkeit:
120 Paarvergleiche, $\min p > .0008$, Grenzwert $\alpha = .0008$, Prozedur T1: Zwei signifikante Paare (2,5) (6,9)

Prüfung der Mehrdimensionalität:
120 Paarvergleiche, $\min p > .0008$, Grenzwert $\alpha = .0008$, Prozedur T1 m: kein signifikantes Paar

Hinweise: Die Justierung des Grenzwertes für α erfolgte nach KOLLER, ALEXANDROWICZ und HATZINGER (2012, S. 168 f.)



4.3 Abschließende Analyse des finalen Entwurfs und Prüfung der Validität

Die Tabellen 6 und 7 berichten die Prüfschritte des Raschmodells für die finale Schlussfassung auf der Grundlage des gesamten Datensatzes bestehend aus Teilstichprobe 1 und 2.

Tab. 6: Ergebnisse der parametrischen Tests für die finale Fassung über die gesamte Stichprobe ($N = 443$)

	Leistungsfähigkeit der Personen		Geschlecht der Personen		Bildungsgang der Personen		Aufgabenkategorie
	Globale Kennwerte, Grenzwert für $\alpha = .033$						
	$\chi^2(9) = 7.707, p = .564$		$\chi^2(9) = 6.210, p = .719$		$\chi^2(9) = 9.536, p = .389$		
	Lokale Kennwerte, Grenzwert für $\alpha = .003$						
	z-Werte	p-Werte	z-Werte	p-Werte	z-Werte	p-Werte	
A01	-0.413	.679	1.107	.309	0.197	.844	2
A02	-1.578	.115	-1.755	.079	1.449	.147	2
A03	0.401	.689	0.619	.536	-0.999	.318	2
A06	0.417	.677	-0.053	.957	1.189	.235	1
A07	-0.376	.707	0.157	.875	-0.784	.453	1
A10	1.911	.056	1.175	.240	-1.068	.286	3
A13	1.117	.264	-0.099	.921	0.519	.604	3
A14	0.650	.516	-0.568	.570	-1.809	.070	4
A15	-0.207	.836	0.330	.742	0.830	.406	4
A16	-0.118	.906	-0.724	.469	0.153	.878	4

Martin-Löf-Test: $\chi^2(24) = 19.027, p = .751$, Grenzwert für $\alpha = .10$

Hinweise: Die Justierung des Grenzwertes für α erfolgte nach KOLLER, ALEXANDROWICZ und HATZINGER (2012, S. 168 f.)

Insgesamt erfüllt die finale Fassung alle Testverfahren, sodass vom Vorliegen eines raschkonformen Modellierungstests ausgegangen werden kann.

Zur weiteren Prüfung der Eigenschaften des Tests wird eine konfirmatorische Faktoranalyse über die 10 Aufgaben des finalen Entwurfs gerechnet. Da die Aufgaben nur die Ausprägungen „0“ für „nicht gelöst“ bzw. „1“ für „korrekt gelöst“ annehmen können, kommt der WLSMV-Schätzer zum Einsatz (Weighted Least Square Estimates with standard errors and Mean- and Variance adjusted chi-square statistics). Während der klassische ML-Schätzer für metrische Variablen zu verwenden ist, kommt der WLSMV-Schätzer bei kategorialen Daten wie im vorliegenden Fall zum Einsatz (vgl. FINNEY/DISTEFANO 2013, S. 474 ff.; URBAN/MAYERL 2014, S. 66). Die CFA kommt zu folgenden Kennwerten für die globale Passung des Messmodells: korrigiertes $\chi^2(35) = 38.453$, $p = .316$, RMSEA: .015, RMSEA CI90 [.000; .038], CFI: .975, WRMR: .769. Nach HU und BENTLER (1999, S. 27) sollte der RMSEA unter .060 sowie der CFI über .950 für

Tab. 7: Ergebnisse der quasi-exakten Tests für die finale Fassung auf der Grundlage der gesamten Stichprobe ($N = 443$)

Leistungsfähigkeit der Personen		Geschlecht der Personen		Bildungsgang der Personen			
Globale Kennwerte, Grenzwert für $\alpha = .033$, Prozedur T10							
p = .642		p = .793		p = .431			
Lokale Kennwerte, Grenzwert für $\alpha = .003$, Prozedur T4							
p-Werte		p-Werte		p-Werte			
Für Leistungs- schwächere leichter als für Leistungs- starke	Für Leistungs- starke leichter als für Leistungs- schwächere	Für Männer leichter als für Frauen	Für Frauen leichter als für Männer	In der dualen Ausbildung leichter als in schulischen Bildungsgängen	In schulischen Bildungsgängen leichter als in der dualen Ausbildung	Trennschärfe Grenzwert für $\alpha = .01$ Prozedur Tpbis	
A01	.831	.241	.197	.866	.432	.656	.660
A02	.976	.060	.967	.056	.105	.943	.941
A03	.513	.593	.352	.728	.855	.199	.282
A06	.498	.604	.575	.515	.100	.939	.188
A07	.777	.389	.496	.615	.818	.284	.682
A10	.064	.959	.138	.909	.847	.212	.117
A13	.255	.841	.599	.484	.309	.769	.146
A14	.409	.681	.787	.312	.973	.047	.440
A15	.712	.494	.420	.690	.250	.843	.694
A16	.719	.377	.814	.267	.443	.667	.914
Itemhomogenität (Aufgabenschwierigkeit): p = .333, Grenzwert $\alpha = .10$, Prozedur Tmd							
Globale Prüfung der lokalen stochastischen Unabhängigkeit: p = .273, Grenzwert $\alpha = .10$, Prozedur T11							
Lokale Prüfung der lokalen stochastischen Unabhängigkeit: 45 Paarvergleiche, min p > .002, Grenzwert $\alpha = .002$, Prozedur T1: keine signifikanten Paare							
Prüfung der Mehrdimensionalität: 45 Paarvergleiche, min p > .002, Grenzwert $\alpha = .002$, Prozedur T1m: keine signifikanten Paare							
Hinweise: Die Justierung des Grenzwertes für α erfolgte nach KOLLER, ALEXANDROWICZ und HATZINGER (2012, S. 168 f.)							

eine Modellpassung liegen. Der WRMR sollte hingegen unter .90 sein (vgl. URBAN/MAYERL 2014, S. 90, S. 99). Im vorliegenden Fall sind diese Grenzwerte stets erfüllt, sodass eine globale Passung des Messmodells gegeben ist.

Zur Prüfung der lokalen Modellpassung berichtet Tabelle 8 die standardisierten Faktorladungen. Diese besitzen stets ein positives Vorzeichen und sind auf dem 1 %-Niveau signifikant, sodass eine lokale Passung des Messmodells vorliegt. Insgesamt bestätigt die CFA die Faktorstruktur des Tests.

Auf der Grundlage der CFA ist es möglich, Reliabilitätswerte zu berechnen, die um spezielle Reliabilitätswerte für Raschmodelle zu ergänzen sind. Dabei erzielt der Modellierungstest (1) eine Anova-Reliabilität von .609, (2) eine Andrichs-Reliabilität von .358, (3) ein Cronbachs α von .483, (4) eine Split-Half-Reliabilität r_{tt} von .429 sowie eine

Tab. 8: Standardisierte Faktorladungen, Lösungsquoten und Schwierigkeitsparameter der Aufgaben

Aufgabe	Standardisierte Faktorladung	Lösungsquote	Schwierigkeitsparameter
A01	.412	.483	0.939
A02	.527	.797	-0.646
A03	.367	.625	0.295
A06	.340	.616	0.337
A07	.418	.858	-1.110
A10	.268	.393	1.346
A13	.308	.770	-0.472
A14	.399	.580	0.504
A15	.433	.887	-1.392
A16	.524	.646	0.197

Hinweise: Je größer der Wert beim Schwierigkeitsparameter, desto schwerer ist die Aufgabe. Alle Faktorladungen sind auf dem 1 %-Niveau signifikant.

Faktorreliabilität von .657. Während die Faktorreliabilität den geforderten Grenzwert von mindestens .600 erfüllt (vgl. BAGOZZI/YI 1988, S. 80), liegt der Wert von Cronbachs α unterhalb des üblichen Grenzwertes in der Psychologie von .700 (vgl. dazu ADAMS/LAWRENCE 2015, S. 90).

Die mit Hilfe des Raschmodells berechneten Eigenschaften der Aufgaben sowie die damit geschätzten Fähigkeiten der teilnehmenden Lernenden fasst Abbildung 6 zusammen.

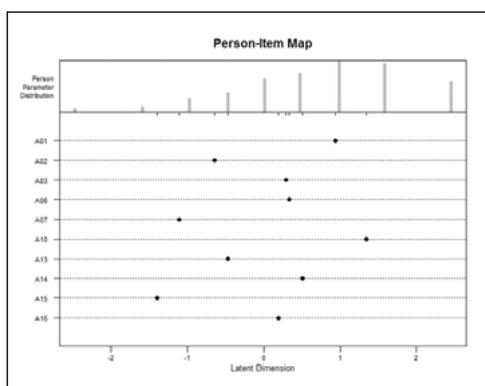


Abb. 6: Personen- und Aufgabenwerte auf der Grundlage der gesamten Stichprobe

Die durchschnittliche Fähigkeitsausprägung sowie der durchschnittliche Schwierigkeitsgrad der Aufgaben werden in Abbildung 6 als Wert „0“ dargestellt. Je größer der Wert ist, desto fähiger ist eine Person bzw. desto schwerer ist die jeweilige Aufgabe. Die

This material is under copyright. Any use outside of the narrow boundaries of copyright law is illegal and may be prosecuted.

This applies in particular to copies, translations, microfilming as well as storage and processing in electronic systems.

© Franz Steiner Verlag, Stuttgart 2020

grauen Balken repräsentieren die Fähigkeiten der Lernenden und die schwarzen Punkte die Schwierigkeit der Aufgaben. Die Abbildung verdeutlicht, dass die meisten Lernenden eine eher hohe Fähigkeitsausprägung besitzen (grauer Balken bei „1“). Zudem wird deutlich, dass die Aufgabe A15 am leichtesten und die Aufgabe A10 am schwersten ist. Insgesamt decken die Aufgaben ein breites Schwierigkeitsspektrum ab.

Die mit Hilfe des Modellierungstests gemessenen Fähigkeiten berichtet Tabelle 9 differenziert nach Bildungsgang und Geschlecht. Dabei können maximal 10 Punkte erreicht werden.

Tab. 9: Mittelwert und Standardabweichung auf dem Modellierungstest für verschiedene Untersuchungsgruppen

	Männlich	Weiblich	Gesamt
Kaufleute für Büromanagement	M = 6.86 SD = 2.507	M = 5.97 SD = 1.923	M = 6.24 SD = 2.134
Industriekaufleute	M = 7.49 SD = 1.466	M = 7.11 SD = 1.549	M = 7.29 SD = 1.518
Kaufleute für Groß- und Außenhandel	M = 6.56 SD = 1.974	M = 5.88 SD = 1.771	M = 6.29 SD = 1.915
Fachoberschule Wirtschaft	M = 6.67 SD = 1.875	M = 4.88 SD = 1.996	M = 5.62 SD = 2.111
Berufliches Gymnasium	M = 6.94 SD = 1.883	M = 6.38 SD = 1.669	M = 6.68 SD = 1.800
Berufsfachschule Wirtschaft	M = 5.43 SD = 1.272	M = 4.92 SD = 2.275	M = 5.11 SD = 1.941
Gesamt	M = 6.98 SD = 1.838	M = 6.34 SD = 1.871	M = 6.65 SD = 1.880

Es wird deutlich, dass die Lernenden der gesamten Stichprobe im Durchschnitt 6,65 der möglichen 10 Punkte erzielen und mit einer Standardabweichung von ca. 1,9 eine große Variation bei den Lernenden vorliegt.

Auffallend ist zudem die mittlere Aufgabenschwierigkeit der verschiedenen Kategorien. So erzielen Aufgaben der Kategorie 1 $M = -.387$ ($SD = 1.023$), Kategorie 2 $M = .196$ ($SD = 0.797$), der Kategorie 3 $M = .437$ ($SD = 1.286$) und der Kategorie 4 $M = -.230$ ($SD = 1.018$). Aufgaben, die sich vor allem auf Buchungssätze stützen (Kategorie 2) oder Bilder und Belege präsentieren (Kategorie 3) sind demnach deutlich schwieriger als Aufgaben, die nur Bilder präsentieren (Kategorie 1) oder Bilder und Buchungssätze (Kategorie 4).

Zur Einschätzung der Zusammenhänge der Ergebnisse des Modellierungstests mit anderen relevanten Konstrukten berichtet Tabelle 10 die bivariaten Korrelationen.

Die Korrelationen belegen einen gemessen an den Richtlinien von COHEN (1988, S. 79 f.) schwachen bis mittleren Zusammenhang zu den Noten. Je besser die Modellierungsfähigkeit der Lernenden, desto bessere Noten erzielen sie auch im Rechnungswesen.

Tab. 10: Korrelationen

	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) Modellierungsfähigkeit	.017	.096*	.040	-.121*	-.014	-.013	.020	.052	-.064	-.183**
(2) Soziale Eingebundenheit		.622**	.499**	-.198**	-.102*	.150**	.210**	.313**	.323**	-.190**
(3) Kompetenzerleben			.560**	-.300**	-.214**	.193**	.276**	.403**	.373**	-.214**
(4) Autonomieerleben				-.274**	-.218**	.090	.224**	.357**	.344**	-.119*
(5) Amotiviert					.562**	-.379**	-.480**	-.539**	-.422**	.419**
(6) Externale Regulation						-.348**	-.452**	-.359**	-.305**	.363**
(7) Introjizierte Regulation							.471**	.331**	.239**	-.286**
(8) Identifizierte Regulation								.483**	.478**	-.330**
(9) Intrinsische Regulation									.690**	-.331**
(10) Interessierte Regulation										-.183**
(11) Noten										1

Hinweise: * p < .05; ** p < .01

sen. Zur Analyse der Einflüsse von Lernprozessen auf die Modellierungsfähigkeit fasst Tabelle 11 die Ergebnisse einer multiplen Regressionsanalyse zusammen.

Tab. 11: Zusammenfassung der multiplen, hierarchischen Regressionsanalyse (N = 320)

Prädiktor	b	β	ΔR^2	ΔR^2_{kor}
<i>Alter</i>	.038		.016*	.013
<i>Geschlecht</i>	-.780**		.036**	.033
<i>Typ Bildungsgang</i>	-.562*		.011 ⁺	.008
<i>Motivation</i>			.037*	.019
Amotivation	-.208	-.117		
Externale Regulation	-.106	-.060		
Introjizierte Regulation	-.082	-.042		
Identifizierte Regulation	-.049	-.026		
Intrinsische Regulation	.323*	.212		
Interessierte Regulation	-.395**	-.229		
<i>Formale Buchungsstrategien</i>			.021	-.008
Kontenart – Aufwand	-.285			
Kontenseite – Aufwand	-.136			
Gegenkonto – Aufwand	1.245			
Abschluss über GuV – Aufwand	.307			
Gegenpol – Aufwand	-2.544			
Kontenart – Ertrag	.045			

Prädiktor	b	β	ΔR^2	ΔR^2_{kor}
Kontenseite – Ertrag	.253			
Gegenkonto – Ertrag	-.081			
Abschluss über GuV – Ertrag	-.053			
Gegenpol – Ertrag	.789			
<i>Grundvorstellungen</i>			.102*	.049
Beschaffungsvorstellung	.587			
Inputvorstellung	.514*			
Verbrauchsvorstellung	.302			
Zerstörungsvorstellung	-2.266*			
Umwandlungsvorstellung	.310			
Liquiditätsabflussvorstellung	-1.178**			
Lagerentnahmeverstellung	-.038			
Eigenkapitalminderungsvorstellung	-.644			
Erfolgsminderungsvorstellung	.087			
Neutralitätsvorstellung	.265			
Materialentnahmeverstellung	-.091			
Lagerbewertungsvorstellung	.266			
Liquiditätszuflussvorstellung	-.223			
Liquiditätsforderungsvorstellung	-.295			
Wertabflussvorstellung	-.191			
Bleibevorstellung	.304			
Differenzvorstellung	.067			
Erfolgssteigerungsvorstellung	.218			
Eigenkapitalsteigerungsvorstellung	.118			
Umsatzerlösvorstellung	.147			

Hinweise:

Referenzkategorie Geschlecht: 0 Männer, 1 Frauen

Referenzkategorie Bildungsgang:

0 dualer Ausbildungsgang, 1 vollzeitschulischer Bildungsgang

Die Tolerance ist stets größer .10 und der Variance Inflation Factor ist stets kleiner 10.

* $p < .05$; ** $p < .01$; † $p < .10$

Demnach erklärt das Alter ca. 1,6 %, das Geschlecht 3,6 %, die Motivation 3,7 % und die Grundvorstellungen ca. 10,2 % der Unterschiede in der Modellierungsfähigkeit auf. Anhand der Regressionsgewichte lassen sich diese Zusammenhänge weiter konkretisieren. So erzielen Männer höhere Werte als Frauen. Zudem sind die erzielten Werte auf den Modellierungstest umso größer, je mehr die Lernenden intrinsisch und desto geringer je mehr die Lernenden Rechnungswesen interessiert lernen. Lernende mit der Input-

This material is under copyright. Any use outside of the narrow boundaries of copyright law is illegal and may be prosecuted.

This applies in particular to copies, translations, microfilming as well as storage and processing in electronic systems.

© Franz Steiner Verlag, Stuttgart 2020

vorstellung erzielen schließlich höhere Werte auf dem Modellierungstest als Lernende mit der Zerstörungs- oder Liquiditätsabflussvorstellung. Der nachfolgende Abschnitt diskutiert die Ergebnisse.

Zusammenfassung und Diskussion der Resultate

Einschätzung der Testeigenschaften: Der vorliegenden Studie gelingt es, einen rasch-konformen Modellierungstest zu entwickeln. Dass Raschmodelle sehr hohe Ansprüche an ein Testverfahren stellen, zeigt sich daran, dass sich in der finalen Schlussfassung nur noch 10 der ursprünglichen 41 Testaufgaben wiederfinden. Eine besondere Stärke des entwickelten Testverfahrens ist dabei die Einhaltung der Raschannahmen für Geschlecht, Leistungsfähigkeit und Bildungsgang. Dies bedeutet, dass die Testaufgaben gleichermaßen für leistungsstarke und leistungsschwache Teilnehmer/-innen, für Frauen und Männer sowie Lernende dualer und vollzeitschulischer Bildungsgänge funktionieren. Dies ist ein wesentlicher Vorteil gegenüber Noten als Indikatoren für den Lernerfolg. Während eine „2“ in einer dualen Ausbildung etwas anderes bedeuten kann als eine „2“ in einem vollzeitschulischen Bildungsgang, sind die Testwerte des Modellierungstests davon unabhängig, d. h. die Testwerte sind durch diese Faktoren *nicht* systematisch verzerrt. In diesem Sinn ist der Test fair und erfüllt so das wichtige Nebengütekriterium der Testfairness (vgl. dazu z. B. DÖRING/BORTZ 2016, S. 449). Wie Abbildung 6 zeigt, decken die Aufgaben zudem ein breites Schwierigkeitsspektrum ab.

Eine weitere Stärke des entwickelten Tests ist es, dass sich die meisten der Testeigenschaften in der Substichprobe 2 ohne Veränderungen an den Testaufgaben replizieren lassen. Dies deutet auf die Stabilität der Testeigenschaften hin. Lediglich das Kriterium „Bildungsgang“ konnte nicht unmittelbar in der Teilstichprobe 2 repliziert werden. Eine mögliche Ursache für diesen Befund kann darin bestehen, dass in der ersten Teilstichprobe die vollzeitschulischen Bildungsgänge aus dem Beruflichen Gymnasium und der Fachoberschule bestanden, während diese Gruppe in der zweiten Teilstichprobe das Berufliche Gymnasium und die Berufsfachschule Wirtschaft umfasste. Wie die abschließenden Analysen in Tabelle 7 und 8 zeigen, konnte durch die finalen Veränderungen eine für die gesamte Stichprobe rasch-konforme Aufgabenzusammenstellung entwickelt werden. Jedoch sind diese Testeigenschaften in weiteren Studien auf ihre Stabilität zu prüfen.

Darüber hinaus zeichnet sich der Test in der konfirmatorischen Faktoranalyse durch eine hohe Modellpassung aus, wie die Erfüllung sämtlicher Kriterien der globalen und lokalen Modellpassung zeigen. Auch die Validierungshypothesen für die Konstruktvalidität konnten in der vorliegenden Studie grundsätzlich belegt werden. So geht eine gestiegene Modellierungsfähigkeit mit besseren Noten im Rechnungswesen einher. Die Korrelation ist mit einem Betrag von .183 nach COHEN (1988, S. 79 f.) nur als schwach bis mittel einzustufen. Dies ist ein Beleg für die diskriminante Validität des Tests. Darüber hinaus erklären vor allem die Grundvorstellungen der Lernenden die Modellierungsfähigkeit. So vereinen die Grundvorstellungen im Vergleich zu Geschlecht, Typ des Bil-

dungsgangs und Motivation die größte aufgeklärte Varianz auf sich. Dies gilt auch noch dann, wenn das R^2 um die hohe Anzahl der berücksichtigten Vorstellungen korrigiert wird. Die aufgeklärte korrigierte Varianz liegt immer noch bei ca. 5 %. Ebenso tragen die formalen Buchungsstrategien nicht zur Erklärung der Modellierungsfähigkeit bei. Dies entspricht den theoretischen Erwartungen aus dem Modell der Grundvorstellungen (BERDING 2019, VOM HOFE 1995; 2003), sind für Modellierungen doch inhaltliche Vorstellungen von den abstrakten Begriffen die kritischen Aspekte.

Insgesamt gelingt es dem entwickelten Test wichtige Facetten der Modellierungsfähigkeit für erfolgswirksame Vorgänge abzubilden. Dies lässt sich damit begründen, dass in dem Test sämtliche Bereiche des Modells der Grundvorstellungen durch verschiedene Aufgaben abgedeckt werden und sowohl Aufgaben zum Ertrag als auch zum Aufwand enthalten sind. So fordert der Test die Lernenden beispielsweise dazu auf, reale Situationen, die durch Fotos repräsentiert werden, mit den richtigen Belegen oder Buchungssätzen in Verbindung zu bringen und umgekehrt. Damit fokussiert der Test deutlich den Teil der Modellierungsfähigkeit, der auf den Wechsel zwischen konkreter und abstrakter Ebene abzielt, wie ihn die Definition von GREEFRATH et al. (2013, S. 18) betont. Weiterentwicklungen dieses Tests sollten Module ergänzen, die einen Vergleich und eine Bewertung von verschiedenen Modellierungen fordern sowie motivationale Aspekte erfassen (vgl. dazu et al. GREEFRATH 2013, S. 18; MAASS 2006, S. 117), um ein noch vollständigeres Bild der Modellierungskompetenz zu ermöglichen.

Uneinheitlich sind hingegen die Befunde zur Motivation. So erklären die eher fremdbestimmten Motivationsvarianten die Modellierungsfähigkeit nicht. Bei den eher selbstbestimmten Varianten ist das Ergebnis uneinheitlich. So gehen eine gestiegene intrinsische Motivation mit einer höheren Modellierungsfähigkeit und eine gestiegene interessierte Motivation mit einer reduzierten Modellierungsfähigkeit einher. Erklären lässt sich dieser Effekt möglicherweise mit methodischen Einflüssen. So vereint die intrinsische Motivation möglicherweise die vorteilhaften Aspekte der interessierten Motivation auf sich (vgl. dazu z. B. BÜHNER/ZIEGLER 2009, S. 690). Dieser Zusammenhang sollte in zukünftigen Studien mit weiteren Motivationsinstrumenten abgesichert werden.

Allerdings weist der Test auch Schwächen auf, vor allem bei dem Gütekriterium der Reliabilität. So erfüllen die Werte für die verschiedenen Reliabilitäten nicht durchgehend den in der Psychologie üblichen Grenzwert von mindestens .700 (vgl. dazu ADAMS/LAWRENCE 2015, S. 90). Lediglich die Faktorreliabilität mit einem Wert von .657 erfüllt den Grenzwert von .600 (vgl. z. B. BAGOZZI/YI 1988, S. 80). Cronbachs α gehört allerdings zu den Gütekriterien der ersten Generation. Für diesen Koeffizient gelten sehr hohe Werte nahe 1 als problematisch, da dies ein Zeichen für inhaltlich zu ähnliche, d. h. redundante Aufgaben darstellt (vgl. WEIBER/MÜHLHAUS 2014, S. 137). Zudem hängt der Wert von Cronbachs α auch von der Anzahl der berücksichtigten Items ab und nutzt deutlich zu restriktive Annahmen, wie z. B. dass alle Aufgaben die gleiche Reliabilität aufweisen (vgl. WEIBER/MÜHLHAUS 2014, S. 137, S. 142). Geringe Werte für die Reliabilität können somit anzeigen, dass die Aufgaben im vorliegenden Test sehr unterschiedliche Facetten des Konstrukts erfassen. Die explizite Schätzung von Messfehlern ist hingegen erst mit den Kriterien der zweiten Generation möglich, zu

denen die Faktorreliabilität gehört (vgl. WEIBER/MÜHLHAUS 2014, S. 130, S. 143). Methodisch ist damit die Schätzung der Reliabilität über die Faktorreliabilität aussagekräftiger als über Cronbachs α . Für die praktische Anwendung ist vor diesem Hintergrund zunächst von einem zuverlässigen Instrument auszugehen. Zukünftige Entwicklungsarbeiten sollten aber weiterhin die Reliabilitätssteigerung verfolgen.

Einordnung des Tests für didaktische Wirkungsstudien: Aufgrund der Einhaltung der Raschannahmen eignet sich der Test für Wirkungsstudien verschiedener didaktischer Konzepte über verschiedene Bildungsgänge hinweg in besonderem Maße, da die Testeigenschaften unabhängig von dualen und vollzeitschulischen Bildungsgängen sind. Dadurch wird es möglich, didaktische Konzepte in verschiedenen Bildungsgängen einzusetzen und die Wirkung der Konzepte mit einem einheitlichen Indikator für den Lernerfolg zu testen.

Vor diesem Hintergrund erweitert der Test die Prüfungsmöglichkeiten im Rahmen von Wirkungsstudien und ergänzt die Tests zum Fachwissen (vgl. BOULEY 2017; SEIFRIED 2004) und zur Problemlösekompetenz (vgl. SEIFRIED 2004) um die Modellierungsfähigkeit. Für Wirkungsstudien wird jedoch je nach Themengebiet die Entwicklung weiterer Modellierungstests notwendig sein, da der vorliegende Beitrag den Schwerpunkt auf erfolgswirksame Vorgänge im Rahmen des externen Rechnungswesens legt. Für die Konstruktion solcher Testverfahren können folgende Richtlinien zu Grunde gelegt werden.

- Der Test sollte Aufgaben aller Kategorien enthalten, um den Modellierungsprozess vollständig abzudecken. Konkret bedeutet dies, dass Aufgaben zu entwickeln sind, die (a) *nur* reale Phänomene vorgeben, (b) *nur* abstrakte Objekte aus den jeweiligen Themen darstellen und (c) die reale Phänomene *und* abstrakte Objekte miteinander in Beziehung setzen.
- Bei der Prüfung der Rasch-Annahmen ist mit einer hohen Rate von ca. 75 % der Testitems zu rechnen, die aus dem Test zu entfernen sind. Für die Planung einer Entwicklungsstudie sollten daher viermal so viele Aufgaben einbezogen werden, wie in der endgültigen Testversion gewünscht sind.

Fachdidaktische Einordnung der Ergebnisse: Die Eigenschaften der Testaufgaben geben zudem wichtige Hinweise für die weitere didaktische Forschung im Rechnungswesen. So sind die Aufgaben, welche nur photorealistische Abbildungen (Kategorie 1) enthalten leichter als Aufgaben mit photorealistischen Abbildungen und Buchungssätzen (Kategorie 4). Diese Aufgaben sind wiederum leichter als Aufgaben nur mit Buchungssätzen (Kategorie 2). Am schwersten sind Aufgaben mit photorealistischen Abbildungen und Belegen (Kategorie 3).

Die Rangfolge der Aufgabenschwierigkeiten zeigt, dass die Anbindung an reale Phänomene den Lernenden leichter fällt als die Bewältigung von Aufgaben nur mit Buchungssätzen. So nimmt der Schwierigkeitsgrad ausgehend von der Vorgabe realistischer Abbildungen (Kategorie 1) über die Kombination von Buchungssätzen mit Abbildungen (Kategorie 4) bis hin zur Bearbeitung reiner Aufgaben mit Buchungssätzen zu (Kategorie 2). Der Schwierigkeitsgrad nimmt mit der Anzahl abstrakter Elemente zu.

Unerwartet ist hingegen, dass die Aufgaben der Kategorie 3, welche Belege und realistische Abbildungen verbinden, die schwersten Aufgaben darstellen. So empfiehlt gerade die Wirkungsstudie von SEIFRIED und SEMBILL (2004, S. 378) die durchgehende Verwendung von Belegen. Auch im wirtschaftsinstrumentellen Rechnungswesenunterricht kommt Belegen als Bindeglied eine zentrale didaktische Bedeutung zu (vgl. PREISS/TRAMM 1996b, S. 252 ff.). So zeigt beispielsweise die Studie von TÜRLING et al. (2011, S. 401), dass Belegbuchungen eher nicht zu den besonders fehleranfälligen Themengebieten aus Sicht von Lehrkräften gehören. Für die befragten Fachlehrer/-innen in der Studie von TRAMM, HINRICHS und LANGENHEIM (1996, S. 179) zählen Belegbuchungen hingegen eher zu den fehleranfälligen Bereichen. Auch bei den Lernenden aus der Studie von SEIFRIED (2004, S. 49) nimmt das Buchen nach Belegen einen mittleren Schwierigkeitsgrad ein. Die subjektiven Einschätzungen der Fachlehrer/-innen und Lernenden entspricht dem empirisch tatsächlich feststellbaren Schwierigkeitswert deutlich besser als die Einschätzungen der Lehrkräfte aus der Studie von TÜRLING et al. (2011).

Die Studie von TÜRLING et al. (2011, S. 402) zeigt jedoch, dass bei der Bearbeitung von Aufgaben Belege zwar eine geringe Fehleranfälligkeit in den Einschätzungen von Lehrkräften aufweisen, allerdings nennen die Lehrkräfte das ökonomische Verständnis und das Formalisieren und Mathematisieren als besonders fehleranfällige Schritte bei der Bearbeitung. Gerade diese beiden Bereiche werden aber mit dem Modellierungstest auf der Grundlage des Modells der Grundvorstellungen adressiert und können den hohen Schwierigkeitsgrad der Aufgaben der Kategorie 3 erklären.

In der Konsequenz übernehmen Belege damit zwar einerseits eine wichtige didaktische Funktion, wie PREISS und TRAMM (1996b, S. 252 ff.) herausarbeiten. Andererseits trägt die Verwendung von Belegen nicht automatisch zu einem „einfacheren“ Buchführungsunterricht bei, denn auch Belege stellen bereits Abstraktionen von realen Prozessen dar (vgl. PREISS/TRAMM 1996b, S. 253). Für die Rechnungswesendidaktik bedeutet dies, dass Belege gleichermaßen sorgsam in den Unterricht einzuführen sind wie Buchungssätze, Bilanzen, Kalkulationsverfahren usw.

Darüber hinaus zeigt der im Vergleich geringe Schwierigkeitsgrad bei den Aufgaben mit photorealistischen Abbildungen, dass der Bezug im Rechnungswesenunterricht zu ökonomischen Phänomenen eher über photorealistische Darstellungen und Animationen erfolgen sollte. Dies untermauert auch die Meta-Studie aus der Mediendidaktik von BERNEY und BÉTRANCOURT (2016, S. 155, S. 159), die aufzeigt, dass photorealistische Animationen lernförderlicher sind als abstrakte Darstellungen wie z. B. über Symbole, Diagramme und Tabellen. Demnach sollte die Rechnungswesendidaktik nicht bei der Inklusion realitätsnaher Belege stehen bleiben, sondern neue Technologien nutzen, um reale Phänomene in den Rechnungswesenunterricht auf einem geringeren Abstraktionslevel zu integrieren. Für die Planung eines solchen Unterrichts schlägt BERDING (2019, S. 289) bereits ein Analyseschema vor. Weitere konkrete Umsetzungsmöglichkeiten könnten sein:

- *Fotoaufträge*: Lernende erstellen mit ihren Smartphones Bilder von Prozessen aus ihrem Ausbildungsbetrieb, die hinter den abstrakten Objekten (z. B. Buchungssätzen,

This material is under copyright. Any use outside of the narrow boundaries of copyright law is illegal and may be prosecuted.

This applies in particular to copies, translations, microfilming as well as storage and processing in electronic systems.

© Franz Steiner Verlag, Stuttgart 2020

Belegen, Kalkulationen) stehen könnten. In einer aufwendigeren Variante könnten auch Lern- und Erklärvideos durch die Lernenden erstellt werden. Diese lassen sich dann im Unterricht thematisieren.

- *Dynamische Arbeitsblätter*: Es werden dynamische Arbeitsblätter eingesetzt, die über die photorealistischen Abbildungen oder Videosequenzen Prozesse aus einem Unternehmen passend zu den abstrakten Objekten offenlegen und eine Interaktion mit diesen erlauben.
- *Erläuterungsaufgaben*: Lernende werden über Aufgabenstellungen dazu aufgefordert, mögliche Prozesse hinter den abstrakten Objekten zu beschreiben, zu erläutern und in ihrer Wirkung auf ein Unternehmen darzustellen.

Fachdidaktische Analyse der nicht aufgenommenen Aufgaben: Darüber hinaus können auch die *nicht* in den Test aufgenommenen Aufgaben Implikationen für die Weiterentwicklung der Rechnungswesendidaktik bieten. Bei den Aufgaben der Kategorie 1, die vor allem Bilder mit möglichen relevanten Situationen zum Aufwand und Ertrag präsentierte, mussten vor allem Bilder entfernt werden, die keine typischen Situationen aus einem Industrieunternehmen zeigten. So mussten beispielsweise Bilder entfernt werden, die Personen bei der Büroarbeit in einer eher lockeren Atmosphäre außerhalb eines Bürogebäudes präsentierten. Auch ein Bild, bei dem der Energieaufwand für die Beleuchtung eines Verkaufsräumens in einem Modeunternehmen erkannt werden sollte, musste aus dem Test entfernt werden. Darüber hinaus waren vor allem Bilderaufgaben zu eliminieren, die Ertragssituationen aus dem Alltag zeigten, beispielsweise der Verkauf eines frisch aufgebrühten Kaffees. Die Verletzung der Rasch-Annahmen durch diese Aufgaben impliziert, dass es bei den Lernenden deutliche Unterschiede in der Fähigkeit gibt, abgebildete Situationen ökonomisch zu interpretieren, was zu den Studienergebnissen von TÜRLING et al. (2011, S. 402) passt, wonach Fehler vor allem beim ökonomischen Verständnis auftreten. Didaktisch lässt dies die Implikation zu, dass der Unterricht stärker *variierende* Situationen aus der Praxis thematisieren sollte, die auch das Erkennen und *ökonomische Deuten* von realen Situationen einfordern.

Bei der Kategorie 2 mussten beispielsweise Aufgaben aus dem Test entfernt werden, bei denen in Buchungssätzen Miete als Aufwand für ein Unternehmen zu erkennen war. Zudem mussten alle Aufgaben aus dem Test entfernt werden, wo die Bezeichnung „Aufwand“ bereits in den angesprochenen Konten auftauchte. Abweichungen von den Eigenschaften eines Rasch-Modells lassen sich hier möglicherweise damit erklären, dass diese Aufgaben nicht die Modellierungskompetenz, sondern vor allem die Fähigkeit zum Erfassen von Textinformationen abbildeten. Vergleichbares gilt auch für die Aufgaben zum Ertrag.

Bei den Aufgaben der Kategorie 3, die Situationen und mögliche Belege zu den Situationen präsentierten, wichen vor allem solche Aufgaben von den Annahmen eines Raschmodells ab, die keine Standardsituationen aus einem Industrieunternehmen zeigten. So sind ein Verkaufsvorgang und ein Herstellungsprozess von Autos im Test verblieben. Situationen hingegen, in denen Material aus dem Lager mit einem Gabelstapler entnommen, ein Verkaufsraum beleuchtet oder ein Auto repariert wurde, konnten nicht im Sinne der Rasch-Annahmen von den Lernenden mit Aufwand und Ertrag in Ver-

bindung gebracht werden. Dies untermauert die Folgerungen aus der Analyse zu den Aufgaben der Kategorie 1.

Vergleichbare Strukturen lassen sich auch bei den Aufgaben der Kategorie 4 feststellen, bei denen Buchungssätze und Abbildungen realer Situationen aufeinander zu beziehen waren. So sind klassische Situationen aus dem Rechnungswesenunterricht im Test verblieben, aber Kombinationen, die z. B. eine Person bei der Büroarbeit oder dem Zusammenbau von Maschinen zeigten, konnten nicht im Sinne des Raschmodells mit einem Buchungssatz zu Löhnen und Gehältern in Verbindung gebracht werden. Vergleichbares gilt auch für die Abschreibung eines fahrenden LKWs. Auch dies lässt die didaktische Implikation zu, dass im Rechnungswesenunterricht die Schulung ökonomischer Deutungen verschiedener realer Situationen in den Fokus rücken sollte, damit die Lernenden wirtschaftliche Begriffe zur Modellierung verschiedener realer Situationen nutzen können.

Zusammenfassung: Am Ende dieser Studie liegt ein validierter Test vor, der (1) die Kompetenzfacette des Wechsels zwischen konkreter und abstrakter Ebene als Teil der Modellierungskompetenz fokussiert, (2) die Modellierungsfähigkeit grundsätzlich valide erfasst, (3) eine ausreichende Reliabilität aufweist, (4) in Zusammenhang mit den Rechnungswesennoten steht, aber (5) tatsächlich etwas anderes als Noten erfasst (diskriminante Validität). Der validierte Test (6) erhebt die Modellierungsfähigkeit *ohne* Bevor- und Benachteiligung von Lernenden verschiedener Leistungsstufen, Bildungsgänge und verschiedenen Geschlechts und ist in diesem Sinn fair. Darüber hinaus konnten (7) erwartungskonforme Zusammenhänge zu den Grundvorstellungen ermittelt werden, d. h. für das erfolgreiche Modellieren im Rechnungswesen kommt es vor allem auf inhaltliche Vorstellungen an.

Aus didaktischer Perspektive zeigt die Studie, dass (8) Belege in den Rechnungswesenunterricht ebenfalls bewusst und didaktisch aufbereitet eingeführt werden sollten. Zudem (9) zeigt die Analyse der ausgeschlossenen Aufgaben, dass eine Förderung der ökonomischen Wahrnehmung durch den gezielten und variierenden Einsatz von realitätsnahen Fotos oder Videos sinnvoll erscheint, damit Lernende abstrakte Begriffe zur Beschreibung der Realität und zur Lösung realer Probleme nutzen können.

Limitationen und Ausblick: Aufgrund des Designs der Studie ist es nicht möglich gewesen, die Testeigenschaften bezüglich des allgemeinbildenden Schulabschlusses zu prüfen. Ebenso konnten bisherige Messinstrumente für Wirkungsstudien nicht in die Studie mit einbezogen werden. Zukünftige Arbeiten sollten daher die Zusammenhänge zu anderen Messinstrumenten prüfen und vor allem auch den späteren beruflichen Erfolg der Lernenden über Längsschnittstudien erfassen, um die prognostische Validität des Modellierungstests zu evaluieren.



Literatur

- ADAMS, K. A. / LAWRENCE, E. K. (2015). *Research methods, statistics, and applications*. Thousand Oaks: Sage.
- BAGOZZI, R. P. / YI, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, 16(1), 74–94.
- BERDING, F. (2019). *Rechnungswesenunterricht. Grundvorstellungen und ihre Diagnose*. Augsburg: Hampp.
- BERDING, F. / RIEBENBAUER, E. / SLEPCEVIC-ZACH, P. (2019). Grundvorstellungen angehender Wirtschaftspädagog(inn)en im Rechnungswesen. Verknüpfungen mit Lernprozessen und Unterricht im Vergleich zwischen Deutschland und Österreich. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 115(3), 363–398.
- BERNEY, S. / BÉTRANCOURT, M. (2016). Does animation enhance learning? A meta-analysis. *Computers & Education*, 2016, vol. 101, p. 150–167.
- BOULEY, F. (2017). *Kompetenzerwerb im Rechnungswesenunterricht*. Wiesbaden: Springer.
- BRÖTZ, R. / KOCK, A. / ANNEN, S. / SCHAAL, T. (2015). Gemeinsamkeiten und Unterschiede der kaufmännischen Ausbildungsberufe. In R. BRÖTZ / F. KAISER (Hrsg.), *Kaufmännische Berufe – Charakteristik, Vielfalt und Perspektiven* (S. 91–106). Bielefeld: Bertelsmann.
- BÜCHTER, A. / HENN, H.-W. (2010). *Elementare Analysis. Von der Anschauung zur Theorie*. Heidelberg: Spektrum.
- BÜHNER, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3. Aufl.). München: Pearson.
- BÜHNER, M. / ZIEGLER, M. (2009). *Statistik für Psychologen und Sozialwissenschaftler*. München: Pearson.
- BURKHARDT, F. / HINSCH, H. / KOSTEDE, W. / WESSELOH, H. (2014). Neues Rechnungswesen: Süßes Nervengift oder ernst zu nehmender didaktischer Ansatz für den Rechnungswesenunterricht – eine Antwort auf Axel Stommel. *Wirtschaft und Erziehung*, 66(6), 223–235.
- CERASOLI, C. / NICKLIN, J. M. / FORD, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological Bulletin*, 140(4), 980–1008.
- COHEN, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Psychology Press.
- DECI, E. L. / RYAN, R. M. (1993). Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. *Zeitschrift für Pädagogik*, 39(2), 223–238.
- DECI, E. L. / RYAN, R. M. (2012). Motivation, personality, and development within embedded social contexts: An overview of Self-Determination Theory. In R. M. RYAN (Ed.), *The Oxford handbook of human motivation* (pp. 85–107). Oxford: Oxford University Press.
- DÖRING, N. / BORTZ, J. (2016). *Forschungsmethoden und Evaluation* (5. Aufl.). Heidelberg: Springer.
- DUBS, R. (2008). Der Unterricht im Fach Rechnungswesen. In C. RUDOLPHI (Hrsg.), *Max Boemle: Festschrift zum 80. Geburtstag* (S. 21–37). Zürich: SKV.
- ERNST, F. (2012). Illustrationen in Rechnungswesenbüchern. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 108(1), 63–80.
- ERNST, F. (2014). Bilanzmethode oder Wirtschaftsinstrumentelles Rechnungswesen: Welche Methode ist die beste? *Wirtschaft und Erziehung*, 66(1), 16–18.
- FINDEISEN, S. (2017). *Fachdidaktische Kompetenzen angehender Lehrpersonen*. Wiesbaden: Springer.
- FINNEY, S. J. / DISTEFANO, C. (2013). Nonnormal and categorical data in structural equation modeling. In G. R. HANCOCK / R. O. MUELLER (Eds.), *Structural equation modeling* (2nd ed.; pp. 439–492). Charlotte: Information Age Publishing.

- GREEFRATH, G. / KAISER, G. / BLUM, W. / BORROMEO FERRI, R. (2013). Mathematisches Modellieren – Eine Einführung in theoretische und didaktische Hintergründe. In R. BORROMEO FERRI / G. GREEFRATH / G. KAISER (Hrsg.), *Mathematisches Modellieren für Schule und Hochschule* (S. 11–37). Wiesbaden: Spektrum.
- GUGGEMOS, J. / SCHÖNLEIN, M. (2015). Modellierung von Kompetenzen in der beruflichen Bildung. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 111(4), 524–551.
- HAUSSER, F. / LUCHKO, Y. (2011). *Mathematische Modellierung mit MATLAB*. Spektrum: Heidelberg.
- HELM, C. (2017). Welche Denkschritte durchlaufen Schüler/innen beim Erstellen von Buchungssätzen? *Wissenplus*, 1–16/17, 38–41.
- HOLZMÜLLER, H. H. / BANDOW, G. (2010). Einleitung. In G. BANDOW / H. H. HOLZMÜLLER (Hrsg.), *„Das ist gar kein Modell!“* (S. VII–XIV). Wiesbaden: Gabler.
- HU, L. / BENTLER, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling*, 6(1), 1–55.
- KAPLAN, R. M. / SACCUZZO, D. P. (2013). *Psychological testing: Principles, applications, and issues*. 8th ed. Belmont, CA: Wadsworth.
- KOLLER, I. / ALEXANDROWICZ, R. / HATZINGER, R. (2012). *Das Rasch Modell in der Praxis*. Wien: Facultas.
- KRAMER, K. / PRENZEL, M. / DREHSEL, B. (2000). Lernmotivation in der kaufmännischen Ausbildung aus der Perspektive von Auszubildenden unterschiedlicher Berufe. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 96(2), 207–227.
- LOUNSBURY, J. W. / GIBSON, L. W. / SAUDARGAS, R. A. (2006). Scale development. In F. T. L. LEONG / J. T. AUSTIN (Eds.), *The psychology research handbook* (pp. 125–146). Thousand Oaks: Sage.
- MAASS, K. (2006). What are modelling competencies? *Zentralblatt für Didaktik der Mathematik*, 38(2), 113–142.
- MAIR, P. / HATZINGER, R. / MAIER, M. J. (2018). *eRm: Extended Rasch Modeling*. 0.16–2. URL: <http://erm.r-forge.r-project.org/> [26.04.2018].
- PAWLIK, W. (1980). Fachdidaktik des Unterrichts im Rechnungswesen. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 76(1), 33–42.
- PEKRUN, R. / VOM HOFE, R. / BLUM, W. / GÖTZ, T. / WARTHA, S. / FRENZEL, A. / JULLIEN, S. (2006). Projekt zur Analyse der Leistungsentwicklung in Mathematik (PALMA). In M. PRENZEL / L. ALLOLIO-NÄCKE (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule* (S. 21–53). Münster: Waxmann.
- PLINKE, W. (2014). Essay zur Kakophonie in der (wirtschaftsinstrumentellen) Rechnungswesendidaktik, der die langersehnte Euphorie nicht gelingen will. *Wirtschaft und Erziehung*, 66(8), 306–311.
- PREDIGER, S. (2009). Verstehen durch Vorstellen. In T. LEUDERS / L. HEFENDEHL-HEBEKER / H. G. WEIGAND (Hrsg.), *Mathe Magische Momente* (S. 166–175). Berlin: Cornelsen.
- PREISS, P. / TRAMM, T. (1990). Wirtschaftsinstrumentelle Buchführung – Grundzüge eines Konzepts der beruflichen Grundqualifikation im Umgang mit Informationen über Mengen und Werte. In F. ACHTENHAGEN (Hrsg.), *Didaktik des Rechnungswesens* (S. 13–94). Wiesbaden: Gabler.
- PREISS, P. / TRAMM, T. (1996a). Vorwort. In P. PREISS / T. TRAMM (Hrsg.), *Rechnungswesenunterricht und ökonomisches Denken* (S. V–VI). Wiesbaden: Gabler.
- PREISS, P. / TRAMM, T. (1996b). Die Göttinger Unterrichtskonzeption des wirtschaftsinstrumentellen Rechnungswesens. In P. PREISS / T. TRAMM (Hrsg.), *Rechnungswesenunterricht und ökonomisches Denken* (S. 222–323). Wiesbaden: Gabler.

- PRENZEL, M. / KRISTEN, A. / DENGLER, P. / ETTLE, R. / BEER, T. (1996). Selbstbestimmt motiviertes und interessiertes Lernen in der kaufmännischen Erstausbildung. *Zeitschrift für Berufs- und Wirtschaftspädagogik, Beiheft 13*, 108–127.
- ROSKI, R. (1986). Einsatz von Aggregaten – Modellierung und Planung. Berlin: Duncker & Humblot.
- RYAN, R. M. / DECI, E. L. (2002). Overview of Self-Determination Theory: An organismic dialectical perspective. In E. L. DECI / R. M. RYAN (Eds.), *Handbook of self-determination research* (pp. 3–33). Rochester: University of Rochester Press.
- SCHIEFELE, U. / KRAPP, A. / SCHREYER, I. (1993). Metaanalyse des Zusammenhangs von Interesse und schulischer Leistung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 25(2), 120–148.
- SCHIERENBECK, H. (2003). *Grundzüge der Betriebswirtschaftslehre* (16. Aufl.). München: Oldenbourg.
- SCHUPP, H. (1988). Anwendungsorientierter Mathematikunterricht in der Sekundarstufe I zwischen Tradition und neuen Impulsen. *Der Mathematikunterricht*, 34(6), 5–16.
- SEEBER, S. (2007). Berufsspezifische Fachleistungen in ausgewählten Berufen des Bereichs Wirtschaft und Verwaltung am Ende der Berufsausbildung. In R. LEHMANN / S. SEEBER (Hrsg.), *ULME III – Untersuchung von Leistungen, Motivation und Einstellungen der Schülerinnen und Schüler in den Abschlussklassen der Berufsschulen* (S. 107–157). Hamburg.
- SEIFRIED, J. (2002). Selbstorganisiertes Lernen im Rechnungswesen. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 98(1), 104–121.
- SEIFRIED, J. (2004). *Fachdidaktische Variationen in einer selbstorganisationsoffenen Lernumgebung. Eine empirische Untersuchung im Rechnungswesenunterricht*. Wiesbaden: Deutscher Universitätsverlag.
- SEIFRIED, J. / SEMBILL, D. (2004). Rechnungswesenunterricht neu gestalten – Empirische Befunde aus einem Kooperationsprojekt zwischen Schule und Hochschule. *Wirtschaft und Erziehung*, 54(11), 371–379.
- SEIFRIED, J. / TÜRLING, J. M. / WUTTKE, E. (2010). Professionelles Lehrerhandeln – Schülerfehler erkennen und für Lernprozesse nutzen. In J. WARWAS / D. SEMBILL (Hrsg.), *Schule zwischen Effizienzkriterien und Sinnfragen* (S. 137–156). Baltmannsweiler: Schneider Verlag Hohengehren.
- SEIFRIED, J. / WOLF, K. D. / KLÜBER, C. / SEMBILL, D. (2005). Die Kompatibilität curricularer und methodischer Modellierung als notwendige Bedingung für Unterrichtsqualität. In D. SEMBILL / J. SEIFRIED (Hrsg.), *Rechnungswesenunterricht am Scheideweg* (S. 123–142). Wiesbaden: Deutscher Universitätsverlag.
- STACHOWIAK, H. (1973). *Allgemeine Modelltheorie*. Wien: Springer.
- STOMMEL, A. (2014). Bilanzmethode oder „Neues Rechnungswesen“: Welche Methode fördert das ökonomische Denken? *Wirtschaft und Erziehung*, 66(4), 148–154.
- STROBL, C. (2015). *Das Rasch-Modell* (3. Aufl.). München: Hampp.
- TÖLLNER, A. / JUNGSMANN, T. / BÜCKER, M. / BRUTSCHEK, T. (2010). Modelle und Modellierung. In G. BANDOW / H. H. HOLZMÜLLER (Hrsg.), „Das ist gar kein Modell!“ (S. 3–21). Wiesbaden: Gabler.
- TRAMM, T. / HINRICHS, K. / LANGENHEIM, H. (1996). Lernschwierigkeiten im Buchführungsunterricht. In P. PREISS / T. TRAMM (Hrsg.), *Rechnungswesenunterricht und ökonomisches Denken* (S. 158–221). Wiesbaden: Gabler.
- TÜRLING, J. M. (2014). *Die professionelle Fehlerkompetenz von (angehenden) Lehrkräften. Eine empirische Untersuchung im Rechnungswesenunterricht*. Wiesbaden: Springer.
- TÜRLING, J. M. / SEIFRIED, J. / WUTTKE, E. / GEWIESE, A. / KÄSTNER, R. (2011). ‚Typische‘ Schülerfehler im Rechnungswesenunterricht – Empirische Befunde einer Interviewstudie. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 107(3), 390–407.

- URBAN, D. / MAYERL, J. (2014). *Strukturgleichungsmodellierung. Ein Ratgeber für die Praxis*. Wiesbaden: Springer.
- VOM HOFE, R. (1992). Grundvorstellungen mathematischer Inhalte als didaktisches Modell. *Journal für Mathematik-Didaktik*, 13(4), 345–364.
- VOM HOFE, R. (1995). *Grundvorstellungen mathematischer Inhalte*. Heidelberg: Spektrum.
- VOM HOFE, R. (2003). Grundbildung durch Grundvorstellungen. *mathematik lehren*, 118, 4–8.
- VOM HOFE, R. / HAFNER, T. / BLUM, W. / PEKRUN, R. (2009). Die Entwicklung mathematischer Kompetenzen in der Sekundarstufe – Ergebnisse der Längsschnittstudie PALMA. In A. HEINZE / M. GRÜSSING (Hrsg.), *Mathematiklernen vom Kindergarten bis zum Studium* (S. 125–146). Münster: Waxmann.
- WEIBER, R. / MÜHLHAUS, D. (2014). *Strukturgleichungsmodellierung* (2. Aufl.). Berlin: Springer.
- WINTHER, E. / ACHTENHAGEN, F. (2008). Kompetenzstrukturmodell für die kaufmännische Bildung. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 104(4), 511–538.
- WÖHE, G. / DÖRING, C. (2005). *Einführung in die Allgemeine Betriebswirtschaftslehre* (22. Aufl.). München: Vahlen.

Bildnachweise

- <https://pixabay.com/de/photos/frauen-teamarbeit-team-gesch%C3%A4ft-1209678/>
- <https://picjumbo.com/white-delivery-van-in-motion-driving-on-highway/>
- <https://www.pexels.com/de-de/foto/alkohol-bier-bierflaschen-container-533353/>
- <https://www.pexels.com/de-de/foto/burgersteig-business-draussen-einkaufen-919453/>
- <https://picjumbo.com/purchasing-a-product-with-credit-card-on-e-commerce-online-store/>
- <https://www.pexels.com/de-de/foto/business-der-umsatz-drinnen-einkaufen-1005638/>
- <https://pixabay.com/de/photos/zahlung-bargeld-transaktion-gr%C3%BCn-1438142/>
- <https://pixabay.com/de/photos/stoffrollen-fabrik-material-1767504/>
- <https://pixabay.com/de/photos/schleifer-werkzeuge-arbeitnehmer-2175150/>
- <https://pixabay.com/de/photos/industrie-plasma-fabrik-stahl-2805773/>

PD DR. FLORIAN BERDING

Carl von Ossietzky Universität Oldenburg, Ammerländer Heerstraße 114–118,
26129 Oldenburg, florian.berding@uni-oldenburg.de

ANKE BECKMANN (B. A.)

Carl von Ossietzky Universität Oldenburg

VICTORIA KÜR TEN (B. A.)

Carl von Ossietzky Universität Oldenburg

