

## **Berufsbildungsstandards und Kompetenzmodellierung im Fach Rechnungswesen.**

Analysen zu den berufsbildenden mittleren und höheren Schulen (BMHS) in Österreich.

Mag. Dr. Helm, Christoph, Universitätsassistent, Johannes Kepler Universität Linz,  
Altenberger Straße 69, 4040 Linz, Österreich, Tel. +43 (0)732 2468 7257, Fax +43 (0)732 24  
68 7280, christoph.helm@jku.at

### **Abstract**

Mangels psychometrisch erprobter Tests stellt dieser Beitrags eine Instrument zur Erfassung der Entwicklung der Schülerkompetenzen in der Domäne Rechnungswesen (RW, 9. - 11. Schulstufe der BMHS) vor. Dabei wird auf eine skizzierte Domänenanalyse und auf Ergebnisse psychometrischer Analysen berufsbildungsstandardskonformer Aufgaben (Item Response Theory) eingegangen. Die Ergebnisse (N = 993 zu  $t_1$ ) zeigen zufriedenstellende Reliabilität und Validität der Testinstrumente. So korreliert die Leistungsentwicklung am stärksten mit den RW- und Mathematiknoten, während Leistungsindikatoren aus anderen Fächern weniger stark mit der RW-Leistung assoziiert sind. Ein empirisches Kompetenzniveaumodell wirft im Ausblick die Frage auf, ob die hohen Ansprüche der Berufsbildungsstandards realistisch sind.

Testentwicklung, Rechnungswesentest, Kompetenzmodell, Berufsbildungsstandards

### **1. Österreich – Land der Berge, Äcker, Dome und Berufsbildung**

Seit Beginn der 1990er-Jahre und insbesondere nach den hinter den Erwartungen zurückgebliebenen Ergebnissen österreichischer Schüler/-innen im Rahmen der PISA-Untersuchung zu Beginn des 21. Jahrhunderts erlebt das österreichische Schulsystem eine Phase der Veränderung der Steuerungslogik, hin zur Output-Orientierung. Davon blieb auch der in Österreich stark ausgeprägte berufsbildende Sektor – 49,6 Prozent bzw. 70,8 Prozent aller Schüler/-innen der 9. bzw. 10. Schulstufe besuchen eine berufsbildende Schule – nicht verschont. Beispielsweise werden seit 2005 Berufsbildungsstandards von der Sektion für Berufsbildung des Bundesministeriums für Bildung (aktuell: BMBF) erarbeitet ([www.bildungsstandards.berufsbildendeschulen.at](http://www.bildungsstandards.berufsbildendeschulen.at)). Diese Regelstandards sollen vor allem die Kompetenzorientierung im Unterricht unterstützen. Auch einschlägige Sammelbände, die sich dem Thema Kompetenzorientierung in der Berufsbildung widmen, sind erschienen (bspw. PAECHTER u.a. 2012; MOSER/PICHLER 2014). Allerdings gilt für das Land der Berge, anders als für Deutschland, wo sich bereits einige (ASCOT-)Forschungsprojekte mit der Messung berufsrelevanter Kompetenzen befassen (für eine Übersicht siehe SEEBER et al. 2010), dass hierzulande kaum vergleichbare Bemühungen vorliegen. Dies gilt insbesondere

für die kaufmännischen Domänen, deren Lehre auf Sekundarstufe II im Land am Strome eine lange Tradition genießt (STOCK 2014). Während für das Fach „allgemeine Betriebswirtschaftslehre“ Testkonstruktionsversuche erfolgten (SAGEDER 2003), besteht ein Forschungsdesiderat im Bereich des Faches „Rechnungswesen“ (im Folgenden RW). Um diesem beizukommen widmet sich der vorliegende Beitrag der Frage „Wie kann die Kompetenzentwicklung von Schüler/-inne/n in RW über mehrere Schulstufen der BMHS in Österreich hinweg erfasst werden?“ Zielsetzung des Beitrags ist daher die Darstellung der Entwicklung eines standardisierten Verfahrens zur curriculumnahe Erfassung der Lernergebnisse bzw. der Kompetenzen, die wiederum Basis für belastbare Aussagen zu Effekten pädagogischer Handlungsprogramme und den Prädiktoren der Kompetenzentwicklung bilden.

Der Beitrag leitet in Abschnitt 2 mit dem Bedarf an einem domänenspezifischen Messinstrument ein, stellt in Abschnitt 3 die Entwicklung des WBBs („Wissensüberprüfung von Basiskennnissen der Buchhaltung“) entlang des „Evidence-Centered Design“-Ansatz (ECD; MISLEVY/RICONSCENTE 2006) vor und zeigt in Abschnitt 4 anhand ausgewählter Rasch-Modellgeltungskontrollen sowie uni- und multidimensional IRT-Analysen einerseits die Erreichung psychometrischer Qualitätsmerkmale und andererseits die faktorielle Struktur des WBBs auf. Darauf aufbauend wird in Abschnitt 5 anhand des Vertical Scaling-Ansatzes gezeigt, wie die schulstufenspezifischen WBB-Testversionen auf eine gemeinsame latente Skale transformiert werden, sodass eine Gegenverrechnung der IRT-Scores und somit die Erfassung der Kompetenzentwicklung möglich ist. In Abschnitt 6 und 7 werden Validitätsanalysen (in Hinblick auf Inhalts-, Konstrukt- und Kriteriumsvalidität) dargestellt sowie ein empirisches Kompetenzniveaumodell (in Anlehnung an HARTIG 2007) zur qualitativen Interpretation der Schülerleistungen abgeleitet. Dieses Modell wird abschließend in Abschnitt 8 vor dem Hintergrund der Kompetenzorientierung und der österreichischen Berufsbildungsstandards reflektiert.

## **2. Gibt es geeignete Testverfahren zur Erfassung von Schülerkompetenzen im Bereich der Berufsbildung?**

In Österreich ist man derzeit bemüht, die Kompetenzorientierung auch in der Berufsbildung voranzutreiben: Zum Beispiel wurden 2014/2015 neue Lehrpläne für die Handelsakademien eingeführt, die bereits über eine relativ detaillierte Lehrzieloperationalisierung verfügen. Auch eine standardisierte Reife- und Diplomprüfung soll in 2015/16 verpflichtend eingeführt werden. Entgegen diesen Bemühungen gibt es jedoch keine Initiativen, die aufzeigen, inwiefern Lehrpersonen tatsächlich kompetenzorientiert unterrichten bzw. inwiefern die Kompetenzen der Schüler/-innen durch die Output-orientierten Maßnahmen beeinflusst sind.

Solche Nachweise sind zurzeit de facto auch nicht möglich, da es kaum psychometrisch erprobte Messinstrumente zur Erfassung der Schülerkompetenzen im berufsbildenden Bereich gibt. Dies gilt nicht nur für Österreich: WINTHER (2010, S. 5) führt allgemein an, dass die Kompetenzmodellierung und -messung im berufsbildenden Bereich ein Forschungsdesiderat darstellt. Obwohl es im deutschsprachigen wie im angloamerikanischen Raum immer wieder Bemühungen gab und gibt, kaufmännisches Wissen durch psychologisch fundierte Messinstrumente abzubilden (bspw. LEHMANN/SEEGER 2007; Winther 2010; BECK/KRUMM 1998; SAGEDER 2003; SCHUMANN et al. 2010; BOTHE/WILHELM/BECK 2005 etc.), eignen sich die existierenden Messinstrumente für die Berufsbildung in Österreich nur bedingt, da sie entweder auf das Curriculum für kaufmännisch Auszubildende in der Bundesrepublik Deutschland abgestimmt sind (LEHMANN/SEEGER 2007; WINTHER 2010) und/oder volks- und betriebswirtschaftliches (Allgemein)wissen und Denkleistungen abbilden (BECK/KRUMM 1998; SCHUMANN et al. 2010) bzw. für die Altersgruppe der Studierenden entwickelt wurden (BOTHE/WILHELM/BECK 2005; ZLATKIN-TROITSCHANSKAIA/KUHN 2010). In allen Fällen decken die Testverfahren inhaltliche Bereiche ab, die nicht explizit Teil der BMHS-Curricula sind und daher nur bedingt Aussagen über Kompetenzen dieser Schülergruppe erlauben, die auf das Curriculum und den Unterricht zurückgeführt werden können. Aus diesen Gründen erscheint es notwendig, ein Messinstrument zu entwickeln, das der österreichischen Berufsbildungspraxis und der dort praktizierten Kompetenzorientierung einerseits sowie psychometrischen Kriterien andererseits gerecht wird.

### **3. Entwicklung eines Verfahrens zur Wissensüberprüfung von *Basiskonzepten* der *Buchhaltung (WBB)***

MISLEVY und RISCOEN (2006) schlagen für die Testkonstruktion den „Evidence-Centered Design“-Ansatz (ECD) vor, in dessen Zentrum – wie der Name sagt – „Beweise“ stehen. Diese „Beweise“ sollen das Schließen von beobachtetem Verhalten (z. B. ein Antwortmuster eines/r Probanden/-in in einem Test) auf vermutete, zugrundeliegende Kompetenzen argumentativ untermauern (Assessment Argument). Im ersten Schritt des ECD-Designs, der Domänenanalyse, werden diese „Beweise“ gesammelt, um sie im zweiten Schritt, der Domänenmodellierung zum Assessment Argument zusammenzuführen. Die weiteren Schritte des ECD-Designs befassen sich mit Fragen der operativen Umsetzung des Tests.

#### *Domänenanalyse*

Für die Analyse einer Domäne sind verschiedene Dimensionen leitend. Eine zentrale Leitlinie ist nach WINTHER (2010, S. 60) der Bezug zum Kontext, weil sich Kompetenzen v. a.

in der erfolgreichen Bewältigung von Anforderungen in spezifischen Situationen (in einem Berufsfeld) zeigen. Aus diesem Grund werden im Rahmen der Domänenanalyse häufig die Handlungen und Situationen analysiert, „in denen die Lernenden/Auszubildenden ihre Kenntnisse und Fertigkeiten unter Beweis stellen“ (ebd.), was sich insbesondere bei schulischen Settings anbietet. Ähnlich wie bei der Entwicklung von Curricula ist danach zu fragen, welche Anforderungen Absolvent/-innen bewältigen können sollten, welche Kompetenzen sie dazu benötigen und wie diese Kompetenzen vermittelt werden. Dies führt zur zweiten Leitlinie. Neben dem Kontext im Sinne des Berufsfeldes ist der Kontext im Sinne der Lernumgebung der Schüler/-innen zentral. Ein Merkmal von Domänen ist, dass sich domänenspezifische Inhalte in ihre kognitiven Strukturen von jenen anderer Domänen unterscheiden und dass infolgedessen auch für ihre Vermittlung bestimmte instruktionale Vorgehensweisen adäquater sind (Fachdidaktik). Daher ist für die Testkonstruktion zu klären, wie Lernende in der Domäne Wissen erwerben, Konzepte und Methoden anwenden und Lehrende bei der Vermittlung vorgehen (a.a.O., S. 82). Praktisch gesehen kann daher die Domäne mit dem „Ziel- und Inhaltsbereich eines Unterrichtsfachs“ gleichgesetzt werden, wobei sie mehr ist als nur ein bestimmtes „Verhalten in eng begrenzten Situationen (z. B. Verbuchen von Beschaffungen)“ (a.a.O., S. 83). Für schulische Kompetenztests bedeutet dies, dass sich die Itemkonstruktion am Curriculum und an der Unterrichtspraxis orientiert (a.a.O., S. 75). Im Rahmen der Analyse der Domäne entlang dieser beiden Leitlinien können unterschiedliche Überlegungen und Fragestellungen ins Zentrum rücken (WINTHER 2010, S. 85ff.; vgl. auch MISLEVY/RISCONSCENTE, 2006):

---

„Was sind die zentralen Denkfiguren, Begriffe und Zugänge einer Domäne?“ Welche Inhalte, Konzepte, Terminologien, Werkzeuge und repräsentative Formen werden von Personen verwendet, die in dieser Domäne arbeiten?

---

*Systematik der DOPPIK, Aktiva, Passiva, Imparitätisches Realisationsprinzip, Lieferung auf Ziel, Kontenpläne, T-Konten-Darstellung, Wertstromorientierung, ...*

---

„Was sind die Lern- und Handlungsziele des Unterrichts? Über welches Wissen/welche kognitive Struktur und Arbeitstechniken müssen die Lernenden verfügen, um der Anforderungssituation gerecht zu werden?“

---

*z. B. BMUKK (2013, S. 60): „Die Schülerinnen und Schüler können Bezugs- und Versandkosten verbuchen“; Kooperatives Offenes Lernen als Handlungsziel (BMUKK 2004); Arbeitstechniken: den Kontenrahmen und Kontenplan anwenden können*

---

„Welche Modelle machen sich Lernende von der Realität? In welchem Kontext und unter welchen instruktionalen Bedingungen wird das Wissen erworben?“ Wie erfolgt der Wissensaufbau? Wird deklaratives, prozedurales, strategisches und/oder soziales Wissen angewendet? Welche kognitiven Prozesse laufen bei der Aufgabenlösung ab?

---

*Die RW-Fachdidaktik verweist auf eine Unterrichtspraxis, die auf Übung und Repetition ausgerichtet ist (GÖTZL/JAHN/HELD 2013; SEIFRIED 2009; PÄTZOLD et al. 2003; SEEGER/SQUARRA 2003; REINISCH 2005) und zudem schulbuchlastig ist (Schneider 1993; WILBERS 2014), was v. a. in Österreich zur Dominanz von Aufgabenstellungen unterer Taxonomiestufen führt (HASELGRUBER 2015).*

---

---

HELM et al. (2015) verweisen auf zehn kognitive Schritte, die Schüler/-innen beim Erfassen von Buchungssätzen durchlaufen (z. B. die Überführung der Aufgabenstellung in die „Sprache des RWs“).

---

Inwiefern sind die im Unterricht gestellten Aufgaben authentische Simulationen der Berufspraxis? Was sind typische Anforderungen, die Proband/-inn/en bewältigen müssen?

---

Obwohl neuere Auflagen von Schulbüchern versuchen die Realität stärker zu modellieren (bspw. durch das Modelleinzelunternehmen Thomas Radler), dominieren verkürzte Realitätsabbildungen in Form von Textaufgaben. Sieht man vom Übungsfirmenunterricht und dem Computer-unterstützten RW ab, so kann man argumentieren, dass Unterricht, der sich am Schulbuch orientiert, die Berufspraxis kaum authentisch simuliert.

---

Diese Fragen führen zu einem konzeptuellen Verständnis der Domäne und stellen gleichzeitig die Basis für die Definition des zu erfassenden Konstrukts dar, das sich über die zu messende Fähigkeit und den Kontext, in dem die Messung stattfindet, definieren lässt (vgl. WINTHER 2010, S. 74). Für den WBB kommt die Domänenanalyse zum Eingangsunterricht in den BMHS zur folgenden *Konstruktdefinition*, die das zentrale Lehrziel im Eingangsunterricht abbildet:

*Der WBB misst die Fähigkeit das System der Doppelten Buchhaltung (vgl. bspw. GROHMANN-STEIGER/SCHNEIDER/EBERHARTINGER 2008, S. 52ff.; vgl. BMUKK 2004, S. 40) anwenden zu können.*

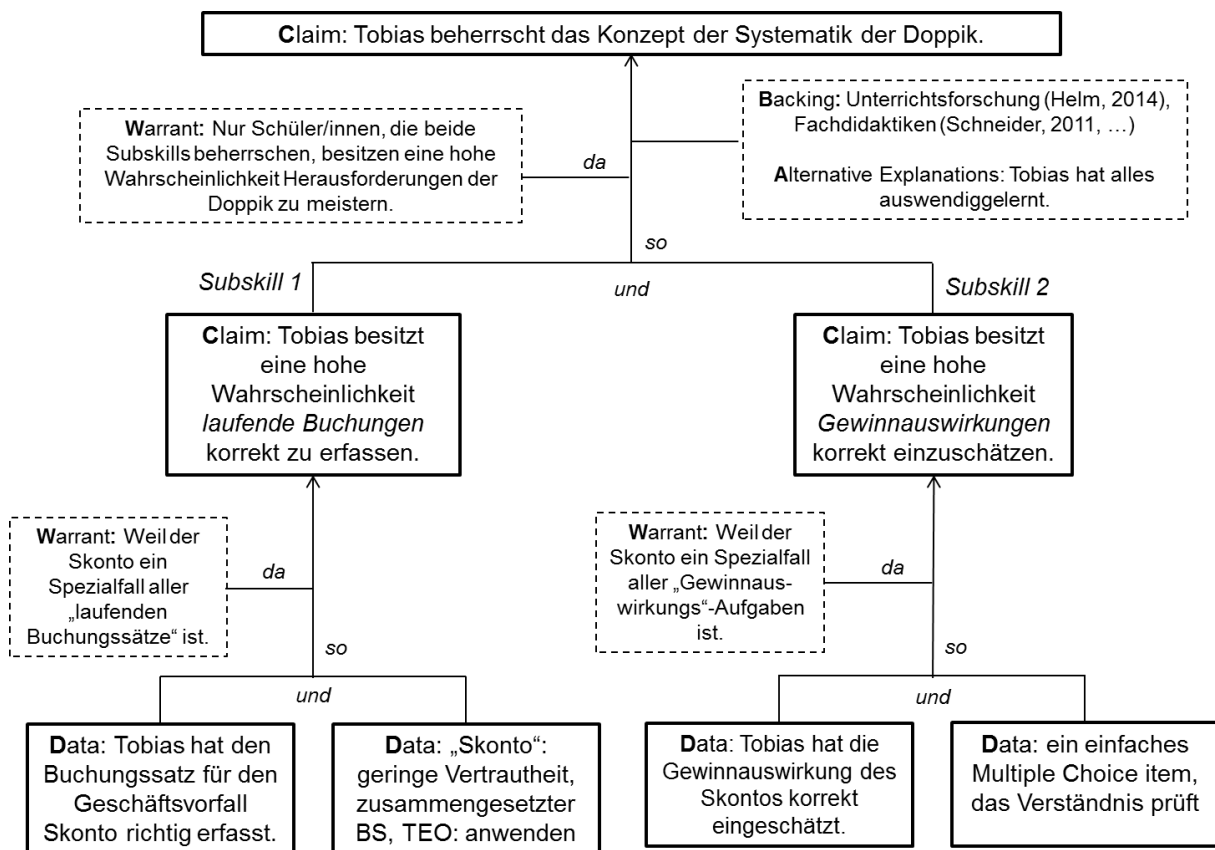
#### *Domänenmodellierung/Assessment Argument*

Im zweiten Schritt des ECD-Designs soll Überzeugungsarbeit geleistet werden. Wenn Schlüsse über Schülerfähigkeiten gezogen werden, dann nur vor dem Hintergrund gut begründeter Vorstellungen (Assessment Argument) über (1) den Zusammenhang zwischen latenter Kompetenz und Antwortverhalten und (2) zwischen Aufgabe und Antwortverhalten (MISLEVY/RICONSCENTE 2006; WINTHER 2010). Diese Assessment Arguments setzen sich aus

- *Claims* (Ziel des Assessments, d. h. Aussagen über die Schülerfähigkeiten im Bereich des RWs),
- *Data* (Antwortmuster im WBB und Informationen über Merkmale von Items) und
- *Warrants* (eine Erklärung dafür, warum das beobachtete Verhalten [Data] Evidenz für die Inferenzen über die Schülerfähigkeiten [Claims] sind) zusammen.

MISLEVY und RICONSCENTE (2006) empfehlen Strukturbäume, um das Assessment Argument verständlich zu repräsentieren:

Abbildung 1: Strukturbaum zur Visualisierung des Assessment Arguments zum WBB



Die zentrale Annahme ist daher, dass Schüler/-innen, die die sogenannte Buchungslogik einmal verinnerlicht haben, viele gängige Aufgabenstellungen des RWs lösen können. Unter der Buchungslogik werden Grundgesetze (z. B. Geldmittelverwendung und -aufbringung, Soll-Haben-Gleichheit etc.) verstanden, die bei der Verbuchung berücksichtigt werden müssen. Ob Schüler/-innen über diese Fähigkeit verfügen, zeigt sich u. a. darin Geschäftsvorfälle korrekt verbuchen und deren Auswirkungen auf den Unternehmenserfolg korrekt einzuschätzen zu können.

Wie Abbildung 1 zeigt, spielen auch die (Aufgaben erschwerenden) Merkmale eines Items eine zentrale Rolle für die Inferenzen über Schülerfähigkeiten. Daher sollten nach WINTHER (2010) a priori die konstruierten Items auf einer eindimensionalen Skala heuristisch verortet werden: „es ist einzuschätzen, ob es sich um ein leichtes Item oder um ein schweres handelt, und es ist zu beschreiben, auf welche Komponenten sich die Unterschiede des Schwierigkeitsniveaus zurückführen lassen“ (ebd.). Als Komponenten führt WINTHER (2010, S. 99ff., vgl. SEEBER et al. 2010, S. 7) die in den Aufgabenstellungen geforderte Modellierungsleistung (den Vertrautheitsgrad), die dargestellte inhaltliche Komplexität und den kognitiven Anspruch der Aufgaben (z. B. nach BLOOM) an. Im konkreten Fall der WBB-Testitems wurde diese „Komponentenanalyse“ allerdings erst ex post im Rahmen der Erstellung des empirischen Kompetenzniveauomodells (vgl. Abschnitt 7) durchgeführt. Im

Rahmen der Testentwicklung wurden die beiden in Österreich dominierenden Schulbücher (Manz und Trauner Verlag) vor dem Hintergrund des Kompetenzmodells, das den Berufsbildungsstandards „HAS Wirtschaft“ und „Entrepreneurship und Management für HAK“ zugrunde liegt, analysiert (vgl. HELM/WIMMER 2012; KIRCHHOFER 2013; HASELGRUBER 2015). Dieses Modell enthält Inhaltsdimensionen für das Fach RW (z. B. laufende Geschäftsfälle, Personalverrechnung etc.), die mit Handlungsdimensionen in Anlehnung an ANDERSON/KRATHWOHL (2001) kombiniert wurden (BMUKK 2012, 2010). In Summe führte diese Vorgehensweise zu insgesamt 127 unterschiedlichen Items.

Tabelle 1: Domänenbereiche der WBB-Tests

<i>WBB 1 (9. Schulstufe)</i>		<i>WBB 2 (10. Schulstufe)</i>		<i>WBB 3 (11. Schulstufe)</i>	
Bilanzpostenzuordnung	(8)	Ankeritems	(10)	Ankeritems	(12)
Kontenarten	(8)	Theoriefragen	(4)	laufende GF o. Beleg	(3)
Kontenseite	(8)	Kontenseite	(8)	Inventur	(1)
laufende GF o. Beleg	(13)	Um- u. Nachbuchung	(7)	Personalverrechnung	(6)
laufende GF m. Beleg	(7)	Vorratsbewertung	(4)	Kostenrechnung	(6)
Inventur	(2)	Personalverrechnung	(4)	Um- u. Nachbuchung	(6)
Privatentnahme	(2)	laufende GF o. Beleg	(8)	Gewinnauswirkung	(4)
Gewinnauswirkung	(15)				

*Anmerkung.* Aufgrund schulautonomer Lehrpläne mussten für die 10. und 11. Schulstufe jeweils drei Testversionen erstellt werden, die unterschiedlichen Schwerpunkten (Handel, Hotel- und Gastgewerbe, Belegverbuchung) gerecht wurden. Zudem zeigte sich in der Pilotstudie, dass die Kategorien Privatentnahme (2) und Kontoarten (8) aufgrund testtheoretischer Analysen, aber auch Verständnisproblemen bei den Schüler/-inne/-n nicht für die Messung der RW-Kompetenz geeignet sind. Sie wurden nicht in die hier berichtete Studie mitaufgenommen. Der Bereich Kostenrechnung wurde in der 11. Schulstufe aus curricularen Gründen mitaufgenommen, obwohl er nicht direkt mit dem interessierenden Konstrukt korrespondiert.

Schritt 3 des ECD-Designs beschäftigt sich mit operativen Überlegungen, die im Rahmen der student, evidence und task models gestellt werden. Das *student model* beschreibt die Annahme, wie die Antwortmuster zur latenten Fähigkeit statistisch in Beziehung gebracht werden (ob linear oder logistisch; ob eindimensional oder mehrdimensional). Im vorliegenden Fall wird dem WBB ein eindimensionales Konstrukt sowie eine logistische Konstrukt-Antwort-Beziehung (d. h., das 1-parametrische Rasch-Modell) zugrunde gelegt. Das *evidence model* beschreibt die Bewertungsregeln bzw. scoring rules für die Evaluation der Beobachtungen. Im WBB werden alle Testitems mit 0 (falsch) und 1 (korrekt) kodiert. Nur teilrichtige Buchungssätze oder Berechnungen werden falsch gewertet. Das *task model* beschreibt schließlich, wie die Items und ihr Antwortformat aussehen. Zu Beginn des WBBs werden die Proband/-inn/-en durch eine kurze Unternehmensbeschreibung in die Situation des/der Buchhalters/-in versetzt. Die Items des WBBs selbst werden möglichst schulbuchnahe und wie in Schularbeiten üblich präsentiert:

(1) Du kaufst für dein Unternehmen Handelswaren auf Ziel (3300) um EUR 14.500,00 + 20 Prozent Umsatzsteuer.

Buchungssatz	+, -, =

#### *Die Implementierung des Assessments und das Assessment Delivery*

Die letzten beiden Schritte des ECD-Designs beschreiben die Implementierung und Datenrückmeldung. Der WBB wurde bisher als 45-minütiger Papier-Bleistift-Test eingesetzt, wobei Taschenrechner und Kontenplan zur Bearbeitung verwendet werden durften. Gegenwärtig befindet sich auch eine computer-adaptive Testversion in Konstruktion (<http://adaptivetesting.ce.jku.at/>). Die erhobenen Daten und berechneten Schülerkompetenzen wurden auf Individual- und Klassenebene aufbereitet und auf der Onlineplattform [www.edumetrics.at](http://www.edumetrics.at) den getesteten Proband/-inn/en verschlüsselt zur Verfügung gestellt. Diesem Assessment Delivery wurde zur inhaltlichen Interpretation das in Abschnitt 7 dargestellte Kompetenzniveaumodell zugrunde gelegt.

#### **4. Ausgewählte IRT-Analysen zur Prüfung psychometrischer Qualitätsmerkmale**

*Stichprobe.* Die verschiedenen Papier-Bleistift-Versionen des WBBs wurden jeweils im Frühjahr an einer Pilotstichprobe (2012: n = 397 Schüler/-innen in 16 Klassen, 2013: n = 226 Schüler/-innen in 9 Klassen, 2014: n = 210 in 11 Klassen) erprobt, modifiziert und im Rahmen der LOTUS-Studie (HELM 2014) am Ende der 9., 10. und 11. Schulstufe an einer Panelstichprobe mit 575 Schüler/-innen ( $t_1$ ) aus 24 Klassen eingesetzt. Die ausgewählten Klassen stammen aus sieben verschiedenen Bundesländern in Österreich (ohne Kärnten und Burgenland). Dabei dominieren die weiblichen Schüler/-innen deutlich, ihr Anteil beträgt rund 75 Prozent. Dieser hohe Anteil lässt sich auf den hohen Anteil an Klassen des Schultyps „Höhere Lehranstalt für wirtschaftliche Berufe“, der fast ausschließlich von weiblichen Schüler/-inne/n besucht wird, zurückführen. Das Durchschnittsalter bei Schuleintritt betrug 14,5 Jahre (SD = rund 9 Monate).

*Statistische Analysen.* Die Schülerantworten wurden 0/1-kodiert und mittels der R-Pakete „ltm“ (RIZOPOULOS 2006) und „eRm“ (MAIR/HATZINGER/MAIER 2011) folgenden psychometrischen Analysen unterzogen: Jede der sieben Testversionen wurde globalen, inferenzstatistischen Tests (Pearson- $\chi^2$ -Test mit Bootstrap-Analyse, Andersen-Likelihood-Ratio-Test: LRT) sowie grafischen Modellgeltungskontrollen unterzogen. Im Rahmen letzterer werden die Itemschwierigkeiten von Subgruppen, die durch Trennung der Gesamtstichprobe einerseits am Mittelwert und andererseits am Schultyp entstehen, gegeneinander geplottet. Auf Itemebene wurden Infit-Statistiken berechnet und die



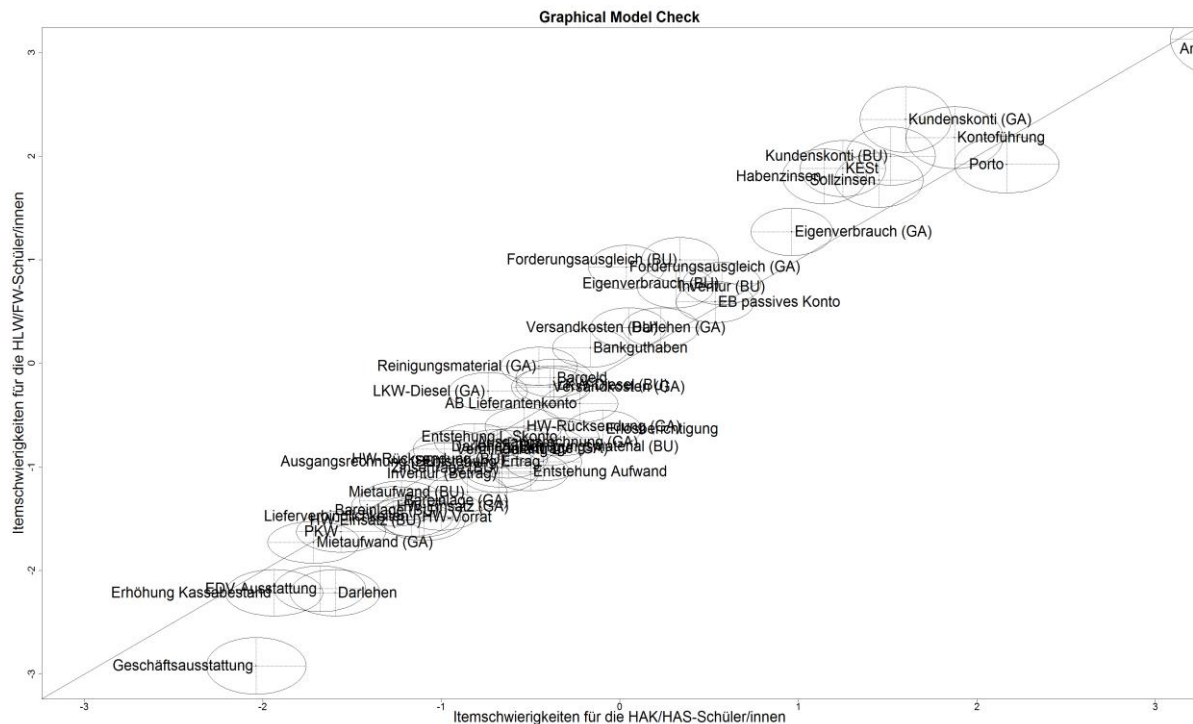
beobachtete sowie modellkonforme Itemcharakteristikkurve (ICC) grafisch inspiziert. Der Eindimensionalitätstest sowie multidimensionalen Item Response Modelle (CHALMERS 2012) wurden für die Prüfung der Faktorenstruktur herangezogen. Aus Platzgründen werden im Folgenden exemplarisch Analyseergebnisse für den WBB 1 dargestellt.

### *WBB 1 Analysen zur 9. Schulstufe*

*Globale Tests:* Eine erste Prüfung mittels LRT zeigt, dass das Rasch-Modell nicht zur Datenstruktur der 53 WBB 1-Items passt. Der LRT weist einen p-Wert unter dem Cut-off von .05 auf (LRT[52, N = 993] = 562.95,  $p < .05$ ), d. h. die Nullhypothese, dass das Rasch-Modell gilt, muss verworfen werden. Um Rasch-Modell-Konformität zu erlangen, wurden auf Itemebene ebenfalls inferenzstatistische und grafische Modellgeltungskontrollen durchgeführt, um nicht-modellkonforme Items zu identifizieren und auszuschneiden. Die Signifikanztests und visuellen Inspektionen führten zu dem Ergebnis, dass die Items der Bereiche „Bilanzerstellung“ (8) und „Kontenseite“ (8) ausgeschieden werden mussten. Es ist zu beachten, dass das hohe N (= 993) der Stichprobe schneller zu signifikanten Modellverletzungen führt. Letztlich zeigt aber auch das Ausscheiden dieser Itemkategorien keine Modellgeltung für die restlichen Daten (LRT[29, N = 993] = 77.29,  $p < .05$ ). Erst wenn die verbleibenden WBB 1-Items in die zwei theoretisch vermuteten Subskills „Geschäftsfälle verbuchen“ und „Gewinnauswirkung einschätzen“ unterteilt werden, kann jeweils für beide Itemkategorien Modellgeltung festgestellt werden (Kategorie „Geschäftsfälle verbuchen“: LRT[16, N = 993] = 22.73,  $p = .12$ , Pearson  $\chi^2$  mit Bootstrap:  $p = .05$ ; Kategorie „Gewinnauswirkung einschätzen“: LRT[11, N = 993] = 13.74,  $p = .25$ , Pearson  $\chi^2$  mit Bootstrap:  $p < .05$ ). Die Annahme, dass sich diese beiden Kategorien auf zwei unterschiedliche latente Fähigkeiten zurückführen lassen, lässt sich einerseits durch das unterschiedliche Antwortformat (offen vs. geschlossen) und andererseits durch die Tatsache begründen, dass die Gewinneinschätzungsitems nicht immer im Unterricht geübt werden, weshalb sie möglicherweise eher eine stärker verständnisorientierte Fähigkeit erfordern. In Abbildung 2 sind die geschätzten Aufgabenschwierigkeiten für die Subgruppen der Schultypen HAK/HAS und HLW/FW dargestellt. Die Kreise bzw. Ellipsen bilden die Konfidenzintervalle. Berücksichtigt man den großen Stichprobenumfang, der dazu führt, dass die Ellipsen sehr klein sind, so kann argumentiert werden, dass alle Items für beide Teilstichproben gleich schwer zu lösen sind bzw. hier keine signifikanten Unterschiede bestehen. Zudem zeigt sich, dass vor allem die Items einen Misfit aufweisen, die ohnehin den laut globalen, inferenzstatistischen Tests auszuschneidenden Itemkategorien „Bilanzzuordnung“ und „Kontoseite“ angehören. Einzig die Items zum Forderungsausgleich und zum Kundenskonti bevorteilen die HLW/FW-Schüler/-innen leicht. Die hier nicht abgebildete Modellkontrolle nach dem Splitkriterium „Mittelwert des Rohtestwertes“ verweist

nur auf das Item „Mietaufwand (GA)“, das die leistungsfähigeren Schüler/-innen systematisch bevorzugt.

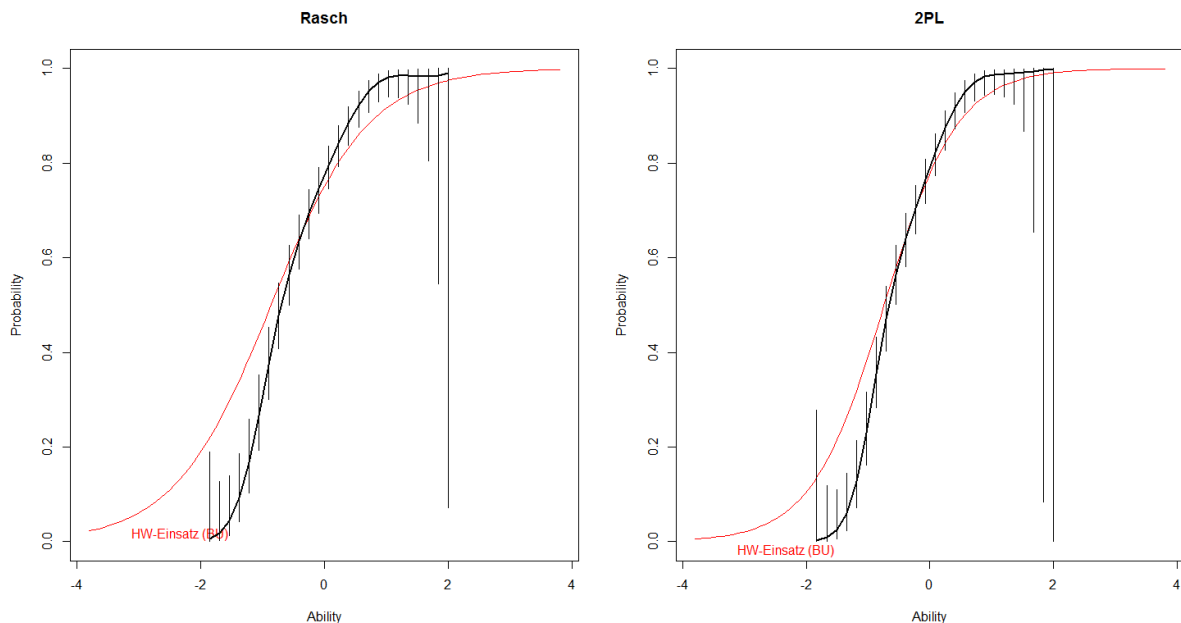
Abbildung 2: Grafische Modellgeltungskontrolle zu den WBB 1-Items (Splitkriterium: Typ)



*Rasch-Modell vs. komplexere Modell:* Ein LRT zeigt weiters, dass das 2-parametrische Modell (unterschiedliche Trennschärfen je Item werden zugelassen) besser zur Datenstruktur der WBB 1-Items passt als das Rasch-Modell (Dimension “Fall verbuchen”):  $LRT[13, N = 993] = 47.43, p < .05$ ; Dimension “GA einschätzen”:  $LRT[12, N = 993] = 64.69, p < .05$ . Dagegen fittet ein 3-parametrisches Modell (ein Rateparameter wird berücksichtigt) nicht signifikant besser (Dimension “Fall verbuchen”:  $LRT[14, N = 993] = 4.07, p = .99$ ; Dimension “GA einschätzen”:  $LRT[13, N = 993] = 14.76, p = .32$ ). D. h., die Berücksichtigung des Rateparameters ist (v. a. bei den Gewinneinschätzungsitems) nicht notwendig. Somit konkurrieren das Rasch-Modell und das 2-parametrische Modell. Mit (BÜHNER 2006, S. 349) ist darauf hinzuweisen, dass der hier verwendete LRT Nachteile hat, u. a. von der Stichprobengröße abhängig ist. Daher sollten auch informationstheoretische Maße (z. B. Bayes Information Criterion: BIC), die die Modellkomplexität mitberücksichtigen, bei der Modellauswahl herangezogen werden. Die BIC-Werte sprechen jeweils für die Beibehaltung des Rasch-Modells, da die Werte niedriger liegen als bei Annahme des 2-parametrischen Modells („Geschäftsfälle verbuchen“: 14627.63 vs. 14670.00; „Gewinnwirkung einschätzen“: 11904.60 vs. 11922.71). Für die meisten Items kann zudem grafisch gezeigt werden (siehe die Beispiele „HW-Einsatz (BU)“ in Abbildung 3), dass

sich die Passung zwischen beobachteten Daten (dunkle Linie) und zwischen den unter Modellannahmen vorhergesagten Daten (helle Linie) kaum verbessert.

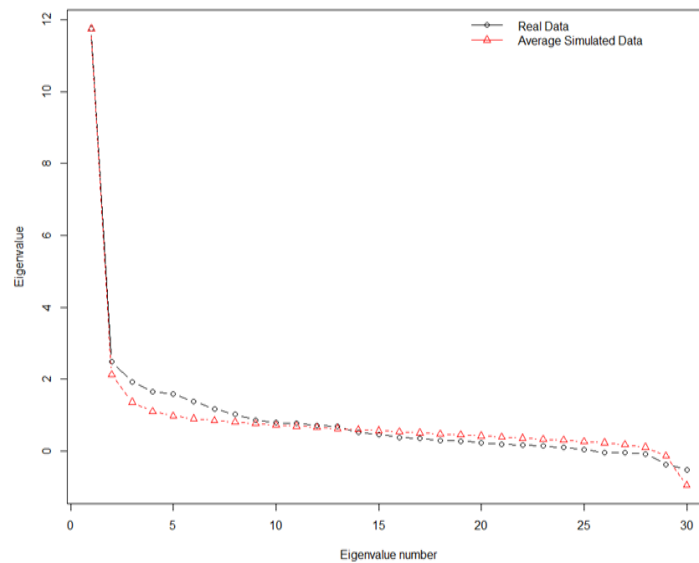
Abbildung 3: Misfit zum WBB-1-Item „HW-Einsatz (BU)“



*Auf Itemebene:* zeigt die Inspektion der Infit t-Statistik, dass nur die Aufgaben zum Kundenskonti und die Einschätzung der Gewinnauswirkung zur Anzahlung außerhalb der Grenzen liegen. Auch die Aufgaben zur Ausgangsrechnung liegen knapp außerhalb des Intervalls. Diese Items folgen nicht dem Rasch-Modell, werden aber dennoch in den Folgeanalysen aus inhaltlichen Gründen (z. B. stellt die Skontoverbuchung einen zentralen Unterrichtsinhalt der 9. Schulstufe dar) beibehalten.

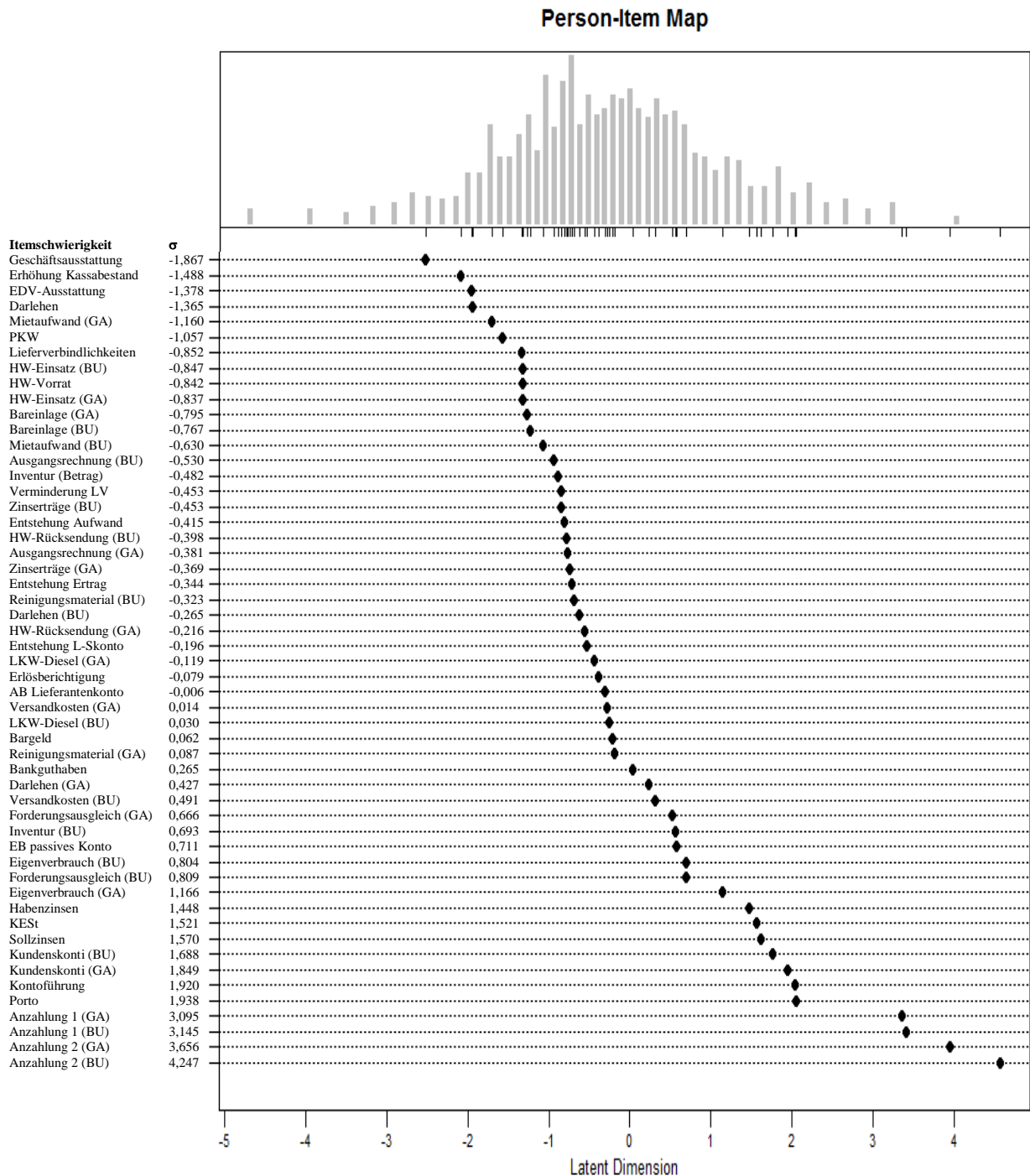
*Faktorenstruktur:* Nachdem Rasch-Modellgeltung für die WBB 1-Items festgestellt werden konnte, kann die strukturelle Annahme, die durch die Teilung der Aufgaben in zwei Bereiche vorgenommen wurde, überprüft werden. Die konfirmatorische Analyse zeigt, dass beide Faktoren eine hohe interne Konsistenz besitzen ( $\alpha$  .85 bzw. .86) und dass ein 2-Faktoren-Modell die Daten besser widerspiegelt als ein 1-Faktoren-Modell ( $LRT[1, N = 993] = 163.19$ ,  $p < .05$ , BIC: 28753.93 (1 Faktor) vs. 28597.64 (2 Faktoren)). Allerdings ist die Korrelation zwischen beiden Fähigkeiten mit  $r = .85$  derart hoch, dass von einer Trennung in zwei Kompetenzdimensionen abgesehen werden kann. Diese Entscheidung wird von der Prüfung auf Eindimensionalität gestützt, die auf einen zentralen Faktor verweist. Diesem Faktor gegenüber spielen andere Faktoren, die ebenfalls einen Eigenwert über 1 besitzen eine nachgeordnete Rolle (vgl. Abbildung 4).

Abbildung 4: Screeplot zur Prüfung auf Unidimensionalität des WBB 1



*Itemschwierigkeiten:* Abbildung 5 gibt die Aufgabenschwierigkeiten in aufsteigender Reihenfolge wider. Im oberen Abschnitt der Abbildung ist ein Histogramm dargestellt, das die Verteilung der Schülerfähigkeiten anzeigt. Es ist gut ersichtlich, dass die WBB 1-Items vor allem im Fähigkeitsbereich von -2 bis +2 messen. Für Schüler/-innen, deren Fähigkeit außerhalb dieses Intervalls liegt, besitzt der Test nur wenige Items und misst daher auch weniger verlässlich, wie auch die Testinformationsfunktion in Abbildung 6 zeigt. Wie erwartet erweisen sich die Verbuchung der Anzahl, des Kundenskonto sowie des Bankkontoauszugs zum 31.12. (Habenzinsen, KEST, Sollzinsen, Kontoführung, Porto) und des Eigenverbrauchs als besondere Herausforderung für die Schüler/-innen. Überraschend erscheint jedoch, dass auch der Ausgleich einer Forderung mittels Banküberweisung für die Schüler/-innen eine Schwierigkeit darstellt – hier wurde des Öfteren fälschlicherweise die Umsatzsteuer mitverbucht. Ein ähnliches Bild zeigt sich für die Einschätzung der Gewinnauswirkungen.

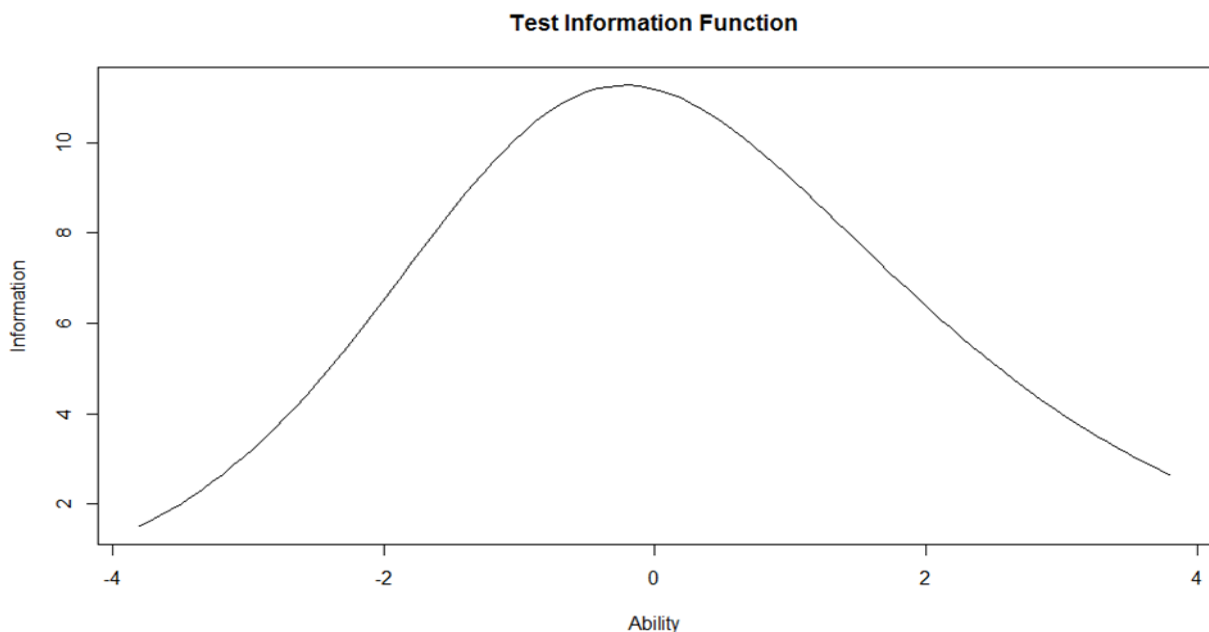
Abbildung 5: Person-Item-Map für den WBB 1



**Objektivität und Reliabilität.** Anhand der grafischen Modellgeltungskontrolle konnte gezeigt werden, dass der WBB über Objektivität im Sinne von Item- und Personenhomogenität verfügt. Die Testinformationsfunktion (TIF) zeigt die Reliabilität des WBB 1 über die gesamte Fähigkeitsskala an, d. h. wie die Messgenauigkeit des Tests in Abhängigkeit der Personenfähigkeit variiert. Es zeigt sich, wie bereits erwähnt, dass – wie bei allen Tests, die auf Durchschnittspopulationen abstellen – in den Randbereichen die Messgenauigkeit abnimmt, da dort einerseits weniger Schüler/-innen anzutreffen sind und andererseits

weniger Items für diese Fähigkeitsbereiche erstellt wurden. Aufgrund der sehr schwierigen Items zur Anzahlung ist im oberen Bereich die Messgenauigkeit des WBB 1 besser als im unteren Extrembereich, weshalb die TIF leicht asymmetrisch ist. Dass der Gipfel der TIF leicht unter dem 0-Punkt der Fähigkeitskala liegt, zeigt zudem an, dass der Test am genauesten die Fähigkeiten jener Schüler/-innen misst, die um den mittleren Fähigkeitsbereich (bzw. leicht darunter) liegen.

Abbildung 6: Testinformationsfunktion zum WBB 1



Die psychometrischen Analysen zu den WBB 2 und WBB 3 verweisen – nach Ausschluss einiger Items – ebenfalls auf Rasch-Konformität sowie Eindimensionalität der Tests (siehe HELM 2014).

## 5. Vertical Scaling

Aufgrund schulautonomer Lehrpläne mussten für die 10. und 11. Schulstufe jeweils drei verschiedene Varianten des WBBs eingesetzt werden. Diese sechs Testversionen wurden mittels Vertical Scaling (KOLEN/BRENNAN 2004) auf eine mit dem WBB 1 gemeinsame latente Skala transformiert. Dazu wurden die Itemschwierigkeiten der WBB 2- bzw. WBB 3-Tests neu kalibriert, d. h. auf einer mit den WBB 1-Items gemeinsamen Schwierigkeitsskala platziert, wobei die zehn bzw. zwölf Ankeritems, die sowohl im WBB 1 als auch in der zu transformierenden Testversion enthalten sind, als Referenzpunkt dienten. Durch Anpassung der Skalen Mittelwerte und der Standardabweichungen weisen nun alle Testformen

denselben theoretischen Nullpunkt auf und sind somit gegeneinander verrechenbar. Die Berechnungen wurden mit dem R-Paket „plink“ (WEEKS 2010) durchgeführt.

## 6. Analysen zur Validität der WBB-Testinstrumente

Die *inhaltliche Validität* des Messinstruments wurde bereits durch Orientierung am ECD-Design bzw. am Domänen- und Kompetenzmodell der Berufsbildungsstandards gewährleistet. Nicht nur die strenge curriculare Orientierung sondern auch die gemeinsame Erarbeitung der Testversionen mit erfahrenen Lehrpersonen führte zu einer hohen inhaltlichen Validität der Messinstrumente.

Die *Konvergenzvalidität* (hoher Zusammenhang zwischen zwei Messverfahren, die dasselbe Konstrukt abbilden) wird über die Korrelationen zwischen den WBB-Versionen bestimmt. Tabelle 2 zeigt, dass die Zusammenhänge mit  $r = .48$  bis  $.63$  relativ hoch sind, bedenkt man, dass sie schulstufenbedingt unterschiedliche Curricula abdecken. Gleichzeitig verweisen die Korrelationen auf eine gewisse Stabilität der Leistungsfähigkeit der Schüler/-innen über die drei Schuljahre hinweg, der messfehlerbereinigte Stabilitätskoeffizient liegt bei über  $r > .70$ . Die *Diskriminanzvalidität* (niedriger Zusammenhang zwischen zwei Messverfahren, die unterschiedliche Konstrukte erfassen) wird durch Korrelationen zwischen den WBB-Test-Werten und den Mathematiktestwerten bestimmt. Wie erwartet korrelieren die WBB-Versionen mit der Mathematikleistungsfähigkeit der Schüler/-innen weit niedriger ( $r = .30$  bis  $.38$ ).

Die *Kriteriumsvalidität* wird schließlich über die Zusammenhänge der WBB-Testversionen mit den Schulnoten der Schüler/-innen bestimmt. Die in Tabelle 3 berichteten Zusammenhänge bestätigen die Annahme, dass die Noten in den Fächern RW, Mathematik und Betriebswirtschaftslehre (als domänenverbundene Kompetenz, Winther 2010) am stärksten mit den Testwerten in den WBB-Versionen korrelieren. Die Sprachenfächer weisen dagegen – ebenfalls wie erwartet – einen niedrigeren Zusammenhang auf. In Summe kann mit diesen Analysen eine zufriedenstellende Validität der WBB-Testversionen belegt werden.

Tabelle 2: Konstruktvalidität

Korrelationen der WBB-Werte mit Außenkriterien		1	2	3	4
1	Mathematik (TIMSS-Items zu Beginn der 9. Schulstufe)	-			
2	WBB 1 (Ende 9. Schulstufe, N = 606)	.31	-		
3	WBB 2 (Ende 10. Schulstufe, N = 360)	.30	.55	-	
4	WBB 3 (Ende 11. Schulstufe, N = 286)	.38	.48	.63	-

Tabelle 3: Kriteriumsvalidität

Gemittelte Korrelationen der erhobenen Schulnoten mit den WBB-Werten			
	9. Stufe	10. Stufe	11. Stufe
	$r_{\text{Note-WBB1}}$	$r_{\text{Note-WBB2}}$	$r_{\text{Note-WBB3}}$
Mathematiknote <sup>a</sup>	-.29	-.27	-.28
Englischnote <sup>a</sup>	-.17	-.12	-.14
Deutschnote <sup>a</sup>	-.15	-.14	-.13
Rechnungswesen <sup>b</sup>	-.43	-.45	-.50
Rechnungswesen <sup>c</sup>	-.42	-.37	-.43
Deutsch <sup>b</sup>	-.25	-.34	-.34
Deutsch <sup>c</sup>	-.22	-.23	-.27
Betriebswirtschaftslehre <sup>b</sup>	n.a.	-.41	-.42
Betriebswirtschaftslehre <sup>c</sup>	n.a.	-.25	-.33
Mathematik <sup>b</sup>	n.a.	-.44	-.40
Mathematik <sup>c</sup>	n.a.	-.41	-.37

Anmerkungen. Die Korrelationen wurden je Schulklasse (N = 14) berechnet. Die hier berichteten Korrelationswerte stellen jeweils Mittelwerte dieser 14 Korrelationen dar. Auf der 9. - 11. Schulstufe beträgt das N je Klasse im Median 24 bis 26 (Min = 18-19, Max = 30-36).

<sup>a</sup>Selbstberichtete Note aus dem Abschlusszeugnis der Zubringerschule in der 8. Stufe.

<sup>b</sup>Selbstberichtete Note aus dem Jahreszeugnis in der 9., 10. bzw. 11. Schulstufe.

<sup>c</sup>Selbstberichtete Note der letzten Schularbeit in der 9., 10. bzw. 11. Schulstufe.

n.a. = Fach wird in dieser Schulstufe nicht unterrichtet

## 7. Ein Kompetenzniveaumodell für das Fach Rechnungswesen

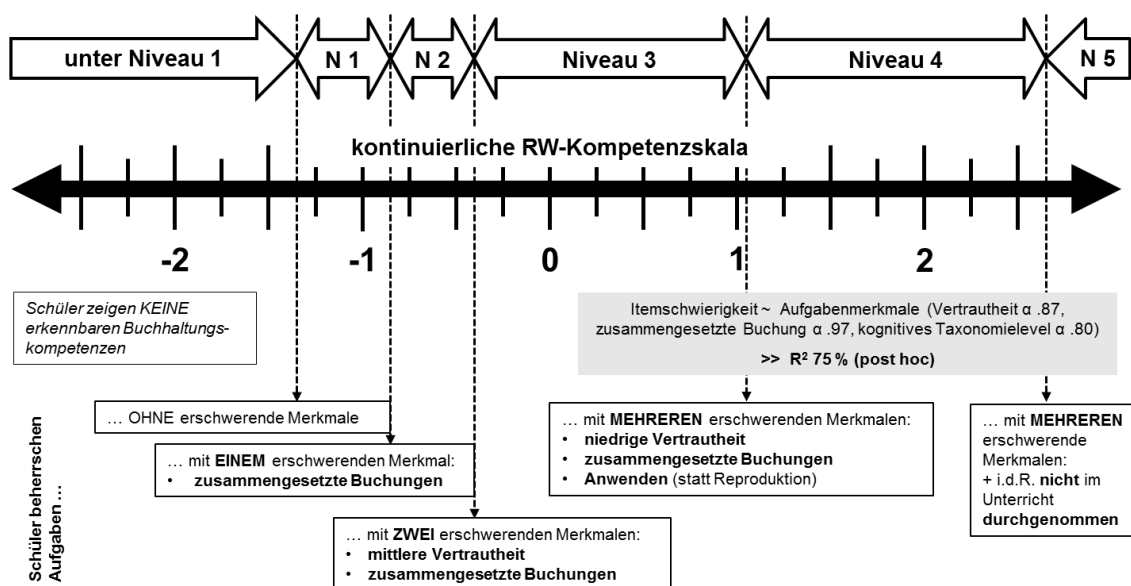
Um die Leistung der Schüler/-innen nicht nur quantitativ, sondern auch qualitativ interpretieren zu können, wurde ein empirisches Kompetenzniveaumodell erstellt. Kompetenzniveaumodelle verfolgen das Ziel, „einer kriteriumsorientierten Beschreibung der erfassten Kompetenzen“ (HARTIG 2007, S. 86) von Schüler/-inne/-n. Es interessiert, „über welche spezifischen Kompetenzen Schüler auf einem bestimmten Niveau verfügen bzw. welche fachbezogenen Leistungsanforderungen sie bewältigen können“ (ebd.). Dazu wird die Kompetenzskala anhand „schwierigkeitsbestimmender Charakteristika“ der Items in Skalenabschnitte unterteilt (a.a.O., S. 88). Konkret wurde im ersten Schritt jedes Testitem von zwei praxiserfahrenen Wirtschaftspädagoginnen in Hinblick auf drei aufgabenerschwerenden Merkmale (vgl. die Komponentenanalyse in Abschnitt 3: Vertrautheit, zusammengesetzte Buchung erforderlich, kognitives Anforderungsniveau) eingeschätzt. Die Interrater-Reliabilitäten betragen .87, .97 und .80 für die drei Merkmale. Die hohe Reliabilität des zweiten Merkmals lässt sich dadurch erklären, dass objektiv sehr einfach feststellbar ist, ob ein Buchungssatz eine zusammengesetzte Buchung erfordert oder nicht.

Nach HARTIG (2007, S. 90) werden diese Aufgabenmerkmale dummy-codiert (Merkmalsausprägung liegt vor oder liegt nicht vor) in eine multiple lineare Regression zur Vorhersage der Itemschwierigkeiten überführt. In Anlehnung an WINTHER (2010, S. 236) werden die so ermittelten Regressionsgewichte zur Berechnung der Schwellenwerte herangezogen. Am Beispiel des WBB 1 für die 9. Schulstufe (die Analysen für die 10. Schulstufe kommen zu einem sehr ähnlichen Kompetenzniveaumodell) soll dies verdeutlicht



werden (Abbildung 7): Die errechnete Regressionskonstante (-1.399 Logits) bildet die erste Schwelle. Schüler/-innen mit einem Kompetenzwert darunter verfügen über keine erkennbaren Buchhaltungskennnisse, da sie selbst Aufgaben ohne erschwerende Merkmale nicht lösen können. Auf Niveau 1, im Bereich zwischen -1.399 und -0.817 (Regressionskonstante + erstes Regressionsgewicht) liegen Schüler/-innen, die Aufgaben ohne erschwerende Merkmale lösen können. Auf Niveaustufe 2 zwischen -0.817 und -0.443 (-0.817 + zweites Regressionsgewicht) Logits befinden sich Schüler/-innen, die Aufgaben lösen können, die über das Schwierigkeitsmerkmal „zusammengesetzte Buchungen“ verfügen. Die Beschreibung der Kompetenzniveaus setzt sich entsprechend fort.

Abbildung 7: Kompetenzniveaumodell zum WBB 1



Eine relative Häufigkeitsverteilung zeigt, dass 25, 18, 8, 36, 11 bzw. 2 Prozent der mit dem WBB 1 getesteten 993 Schüler unter Niveau 1 bzw. auf Niveau 1, 2, 3, 4, und 5 liegen. In Summe wirft dies ein sehr ernüchterndes Licht auf die Leistungsfähigkeit der BMHS-Schüler/-innen am Ende der 9. Schulstufe. Allerdings darf ich dabei nicht vergessen werden, dass es sich hier um eine low-stakes Testung handelt, die keinerlei Konsequenzen für die Schüler/-innen nach sich zieht und entsprechend die Testmotivation eine Rolle spielt.

## 8. Eine Zusammenschau und kritische Reflexion

Ziel der hier dargestellten Untersuchung ist die Erstellung eines Instrumentes zur Messung der Entwicklung der Schülerkompetenzen in der Domäne RW in der 9. bis 11. Schulstufe der BMHS in Österreich. Aufbauend auf die hier skizzierte Domänenanalyse wurden verschiedene Testversionen konstruiert, die schulbuchtypische sowie curriculums- und berufsbildungsstandardkonforme Aufgaben verwenden, um Aussagen über die

Schülerkompetenzen zu treffen. Psychometrische Analyse in Abschnitt 3 zeigten, dass nach dem Ausscheiden einiger nicht modellkonformer Items zufriedenstellende Reliabilität sowohl auf Test als auch auf Itemebene erreicht werden konnte. Auch die Validitätsanalysen in Abschnitt 6 verweisen auf zufriedenstellende Befunde, so korreliert die Leistungsentwicklung – wie erwartet – am stärksten mit der Mathematiknote und den RW-Noten, während Leistungsindikatoren aus anderen Fächern weniger stark mit der RW-Leistung assoziiert sind. Im letzten Abschnitt wurde ein empirisches Kompetenzniveaumodell entwickelt, das inhaltliche Interpretationen der Schülerleistungen vor dem Hintergrund der drei Merkmale, Modellierungsanforderung (Vertrautheit), Komplexität des Buchungssatzes sowie kognitive Herausforderung der Aufgabe, erlaubt. Dieses Kompetenzniveaumodell zeigt deutlich, dass überzogene Erwartungen, die auch im Rahmen der Kompetenzorientierung entstehen können, nur schwierig empirisch haltbar sind. So können die Ergebnisse zum Kompetenzniveaumodell auf der 9. Schulstufe als Hinweise für einen stark repetitiven Charakter des RW-Unterrichts interpretiert werden, da nur wenig Schüler/innen (der vorliegenden Untersuchung) fähig sind Aufgaben zu lösen, die höhere kognitive Herausforderungen stellen. Inwiefern damit bereits dem Ruf nach Kompetenzorientierung genüge getan ist muss kritisch diskutiert werden. Gleichsam soll das hier vorgestellte Modell nur Ausgangspunkt für weitere Forschungsbemühungen darstellen. Vor allem in Hinblick auf den praktischen, fachdidaktischen Einsatz im Unterricht erscheint es notwendig, stärker diagnostisch orientierte Messinstrumente und Kompetenzmodelle zu entwickeln. Eine vielversprechende Entwicklung stellt hier der Ansatz der Cognitive Diagnosis Models dar (z. B. HELM/TROST/GEORGE/POCRNJA 2015). Dieser stärker an den Denk- und Lernschritten von Schüler/innen orientierter Ansatz stünde auch stärker im Einklang mit der aktuellen Kompetenzorientierung in der österreichischen Berufsbildung bzw. der Lehrplanreform an den BMHS, die die bisher eher grob formulierten Unterrichtsthematiken in feinkörnigere Unterrichtsziele übersetzt.

## Literatur

- ANDERSON, L. W.; KRATHWOHL, D.: A Taxonomy for Learning, Teaching, and Assessing. A Revision of Bloom's Taxonomy of Educational Objectives. Longman 2001
- BECK, K.; KRUMM, V.: Wirtschaftskundlicher Bildungs-Test (WBT). Gottingen 1998
- BMUKK: Lehrplan der Handelsakademie. Online im WWW:  
[https://www.hak.cc/files/syllabus/598\\_HAK%20LP%202004%20-%20Anlage%201.pdf](https://www.hak.cc/files/syllabus/598_HAK%20LP%202004%20-%20Anlage%201.pdf) 2004
- BMUKK: HAS Wirtschaft. Bildungsstandards in der Handelsschule. Kompetenzmodell, Deskriptoren. Kompetenzraster. Exemplarische Unterrichtsbeispiele. Online im WWW:  
[http://www.berufsbildendeschulen.at/fileadmin/content/bbs/AGBroschueren/Pilotbroschue\\_HAS\\_Wirtschaft\\_Oktober\\_20121031.pdf](http://www.berufsbildendeschulen.at/fileadmin/content/bbs/AGBroschueren/Pilotbroschue_HAS_Wirtschaft_Oktober_20121031.pdf) 2012
- BOTHE, T.; WILHELM, O.; BECK, K.: Assessment of declarative business administration knowledge: Measurement development and validation. Humboldt-Universität zu Berlin 2005.
- BÜHNER, M.: Einführung in die Test- und Fragebogenkonstruktion. München 2006.
- CHALMERS, P.: mirt: Multidimensional Item Response Theory. R package version 0.2.0. Online im WWW: <http://CRAN.R-project.org/package=mirt> 2012
- GÖTZL, M.; JAHN, R. W.; HELD, G.: Bleibt alles anders!? Sozialformen, Unterrichtsphasen und echte Lernzeit im kaufmännischen Unterricht. In: Berufs- und Wirtschaftspädagogik Online, 24 (2013), S. 1-22
- GROHMANN-STEIGER, C.; SCHNEIDER, W.; EBERHARTINGER, E.: Einführung in die Buchhaltung im Selbststudium. Wien 2008
- HARTIG, J.: Skalierung und Definition von Kompetenzniveaus. In: BECK, B. (Hrsg.): Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch-Englisch-Schülerleistungen-International). Weinheim 2007, S. 83-99
- HASELGRUBER, T.: Testkonstruktion zur Erfassung der Kompetenzen von Schüler/inne/n der 11. Schulstufe im Fach Rechnungswesen. (Unveröffentlichte Diplomarbeit) Johannes Kepler Universität. Linz 2015
- HELM, C.; BLEY, S.; GEORGE, A.-C.; POCRNJA, M.: Potentiale kognitiver Diagnosemodelle für den berufsbildenden Unterricht. In: STOCK, M., SCHLÖGL, P.; SCHMID, K.; MOSER, D. (Hrsg.): Kompetent – wofür? Life-Skills – Beruflichkeit – Persönlichkeitsbildung. Innsbruck 2015, S. 206-224
- HELM, C.; WIMMER, B.: Wie lässt sich der Lernerfolg von Schülerinnen/Schülern im Fach Rechnungswesen messen? In: wissenplus, 5-11/12 (2012), S. 24-29
- KIRCHHOFER, J.: Wie lässt sich der Lernerfolg von Schüler/inne/n im Fach Rechnungswesen messen? Entwicklung eines Messinstruments für den zweiten Jahrgang. (Unveröffentlichte Diplomarbeit) Johannes Kepler Universität. Linz 2013
- KOLEN, M. J.; BRENNAN, R. L.: Test equating, scaling, and linking. Methods and practices. Statistics for social science and public policy. New York 2004
- LEHMANN, R.; SEEGER, S.: ULME III. Untersuchung der Leistungen, Motivation und Einstellungen der Schülerinnen und Schüler in den Abschlussklassen der Berufsschulen. Berlin 2007
- MAIR, P.; HATZINGER, R.; MAIER, M.: eRm: Extended Rasch Modeling. R package version 0.14-0 2011
- MISLEVY, R. J.; RICONSCENTE, M. M.: Evidence-Centered Assessment Design. In: DOWNING, S. M.; HALADYNA, T. M. (Hrsg.): Handbook of test development. Mahwah, N.J. 2006, S. 61-90.
- MOSER, D.; PICHLER, E. (Hrsg.): Kompetenzen in der Berufspädagogik. Graz 2014
- PAECHTER, M.; STOCK, M.; SCHMÖLZER-EIBINGER, S.; SLEPCEVIC-ZACH, P.; WEIRER, W. (Hrsg.): Handbuch Kompetenzorientierter Unterricht. Weinheim 2012
- PÄTZOLD, G.; KLUSMEYER, J.; WINGELS, J.; LANG, M.: Lehr-Lern-Methoden in der beruflichen Bildung. Eine empirische Untersuchung in ausgewählten Berufsfeldern. Oldenburg 2003
- REINISCH, H.: Gibt es aus historischer Perspektive konstante Leitlinien in der Diskussion um das Rechnungswesen? In: SEMBILL, D.; SEIFRIED, J. (Hrsg.): Rechnungswesenunterricht am Scheideweg. Lehren, lernen und prüfen. Wiesbaden 2005, S. 15-31.

- RIZOPOULOS, D.: ltm: An R package for Latent Variable Modelling and Item Response Theory Analyses. In: Journal of Statistical Software, 17 (2006) 5, S. 1-25.
- SAGEDER, J.: Die Wirtschaft geht uns alle an! Unveröffentlichte Testversion. Linz 2003
- SCHNEIDER, W. (Hrsg.): Komplexe Methoden im betriebswirtschaftlichen Unterricht. Festschrift für Hans Krasensky zum 90. Geburtstag. Wien 1993
- SCHUMANN, S. et al.: Inhaltsauswahl für den Test zur Erfassung ökonomischen Wissens und Könnens im Projekt „Ökonomische Kompetenzen von Maturandinnen und Maturanden (OEKOMA)“. Universität Zürich 2010.
- SEEBER, S. et al. 2010: Kompetenzdiagnostik in der Berufsbildung. In: Berufsbildung in Wissenschaft und Praxis - BWP (2010) 1, S. 1-15
- SEEBER, S.; SQUARRA, D.: Lehren und Lernen in beruflichen Schulen. Schülerurteile zur Unterrichtsqualität. Frankfurt am Main 2003
- SEIFRIED, J.: Unterricht aus der Sicht von Handelslehrern. Frankfurt am Main 2009
- STOCK, M.: Kaiserin Maria Theresia und die Wirtschaftspädagogik. In: wissenplus, 5-13/14 (2014), S. 13-17
- WEEKS, J.: plink: An R Package for Linking Mixed-Format Tests Using IRT-Based Methods. In: Journal of Statistical Software, 35 (2010) 12, S. 1-33
- WILBERS, K.: Wirtschaftsunterricht gestalten. Eine traditionelle und handlungsorientierte Didaktik für kaufmännische Bildungsgänge. Berlin 2014
- WINTHER, E.: Kompetenzmessung in der beruflichen Bildung. Bielefeld 2010
- ZLATKIN-TROITSCHANSKAIA, O.; KUHN, C.: Messung akademisch vermittelter Fertigkeiten und Kenntnisse von Studierenden bzw. Hochschulabsolventen: Analyse zum Forschungsstand: Arbeitspapiere Wirtschaftspädagogik: Nr. 56. Mainz 2010