

SANDRA BLEY / ANN CATHRICE GEORGE

## **Kognitive Diagnosemodelle zur Begleitung individualisierter Lehr- und Lernprozesse: Neue Möglichkeiten aus einem alternativen Ansatz?**

**Cognitive diagnostic models to accompany individualized teaching and learning processes: New possibilities from an alternative approach?**

**KURZFASSUNG:** Sowohl die Kategorien des guten Unterrichts als auch die Ausführungen zu den Lehrerbildungsstandards zeigen die Notwendigkeit detaillierter Informationen über Lern- und Leistungsprozesse von Lernenden auf, um diese effizient und effektiv fördern und unterstützen zu können (u. a. die „Diagnosefähigkeit“). In der Ausbildungspraxis greifen Lehrende dabei zumeist auf langjähriges Erfahrungswissen zurück. In diesem Beitrag soll ein kognitives Diagnosemodell vorgestellt und dessen Potenziale für eine evidenz-basierte individuelle, kompetenztheoretisch fundierte und valide Lern- und Leistungsförderung hervorgehoben und diskutiert werden. Dieses erfolgt am Beispiel der Vermittlung von Intrapreneurship-Kompetenz im Rahmen der Ausbildung von Industriekaufleuten. Hierzu wird zunächst das Konzept kognitiver Diagnosemodelle sowie deren zugrundeliegenden Annahmen und Umsetzung am konkreten Beispiel eines technologie-basierten Performance Assessments dargestellt. Anhand der konkreten Analyseergebnisse werden die erhaltenen Informationen fachdidaktisch interpretiert und im Hinblick auf ihre Potenziale zur Förderung individueller Lern- und Leistungsprozesse diskutiert.

**Schlagnote:** Kognitive Diagnosemodelle – Lehr-Lernprozesse – Intrapreneurship – Assessment – Kompetenz – Skills

**ABSTRACT:** The categories of good teaching and the teaching standards both indicate a need for detailed information on learning and performance processes. This detailed information should support an efficient and effective support of learners (e.g., “competence to diagnose”). In daily business teachers mostly rely on many years of experience. In this article we first give an overview of the approach of cognitive diagnostic models. Especially we highlight and discuss the model’s potentials for an evidence-based individual, theoretically grounded and valid learning and performance support. Second, we exemplify the method of cognitive diagnosis models by measuring intrapreneurship competence within the vocational education training of industrial clerks. The example includes an explanation of underlying assumptions as well as their implementation through a technology-based performance assessment. Finally, the outcomes of the cognitive diagnostic model on intrapreneurship competence are presented and interpreted in terms of their potential to promote individual learning and performance processes.

*Key words:* cognitive diagnostic models – sub-competencies – teaching and learning processes – intrapreneurship – assessment

## 1. Einführung

Der gesellschaftliche Anspruch einer zunehmenden Differenzierung im Unterricht sowie einer Individualisierung von Lernprozessen erwächst aus den vielfach belegten Erkenntnissen, dass individualisierte Unterstützung von Lern- und Entwicklungsprozessen zu größtmöglichem Erfolg führt (HATTIE/TIMPERLEY 2007; BILLET 2002; TYNJÄLÄ 2013). Dementsprechend ist die diagnostische Kompetenz des Bildungspersonals auch in nationalen sowie internationalen Standards zur Professionalisierung von Lehrenden verankert (KMK 2004; OSER 1997; INTERSTATE TEACHER ASSESSMENT AND SUPPORT CONSORTIUM 2011). Danach sollen Lehrende im Hinblick auf den Lehrerbildungsstandard bzw. „Kompetenz 8“ Leistungen von Schülerinnen und Schülern auf der Grundlage transparenter Beurteilungsmaßstäbe erfassen. In der praktischen Umsetzung sollten sie (A) Aufgabenstellungen kriteriengerecht (lernzielgerecht) konstruieren und diese adressatengerecht formulieren können (d.h. valide Instrumente zur detaillierten Erfassung domänenspezifischer Lern- und Leistungsprozesse entwickeln), (B) Leistungsüberprüfungen als konstruktive Rückmeldung über die weitere eigene Unterrichtstätigkeit nutzen (d.h. zielgerichtet didaktische Maßnahmen ergreifen) sowie (C) Leistungsstände adressatengerecht transparent aufzeigen (d.h. individuelle Kompetenzprofile identifizieren) und Perspektiven für das weitere Lernen entwickeln (d.h. individuelles informatives Feedback geben) (KMK 2004, S. 11; BAUMERT/KUNTER 2006; HELMKE/HOSENFELD/SCHRADER 2004; ARTELT/GRÄSEL 2009; SCHRADER 2009).

Um jedoch seitens des Bildungspersonals derart differenziert unterrichten bzw. individualisiert Lernprozesse begleiten zu können, sind vielfältige Informationen über den Lern- und Leistungsprozess notwendig. Während bisher Lehrkräfte diesbezüglich auf ihre jeweiligen individuellen Erfahrungen zurückgreifen, soll hier mit dem Ansatz kognitiver Diagnosemodelle ein Verfahren vorgestellt werden, dass ein evidenz-basiertes Vorgehen zur Gestaltung von u. a. individualisierten Lehr- und Lernprozessen erlaubt.

Gegenwärtig vorliegende Instrumente und Verfahren in (i) der Ausbildungspraxis wie Alltagsassessments, (ii) der pädagogisch-psychologischen Individualdiagnostik sowie (iii) der „derzeitigen“ beruflichen Kompetenzmessung mittels Item Response Theorie, mit denen aktuell versucht wird, die notwendigen Informationen zu gewinnen, weisen dabei unterschiedliche Vor- und Nachteile auf, um diesem Anspruch gerecht zu werden (vgl. auch BÖGEHOLZ/EGGERT 2013):

In (i) der schulischen und/oder betrieblichen Ausbildungspraxis angewandte Verfahren und Instrumente zur summativen wie formativen Leistungsfeststellung in Form von Tests, Hausaufgabenkontrollen, Präsentationen, Projektarbeiten, Feedback, Portfolios usw. zeichnen sich durch eine hohe Praktikabilität aus. Bei einem

hohen Grad an ökologisch-inhaltlicher Validität, vernachlässigen diese Verfahren jedoch eine Reihe weiterer Validitätskriterien (u. a. kognitive/strukturelle/prädiktive, siehe u. a. MESSICK 1989; AMERICAN EDUCATIONAL RESEARCH ASSOCIATION/AMERICAN PSYCHOLOGICAL ASSOCIATION/NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION 2014)<sup>1</sup>. Nach Wild und Krapp (2006) mangelt es diesen **Alltagsassessments** an einer Standardisierung, denn ob ein Bündel von Lernaufgaben tatsächlich die intendierten Lernprozesse auslöst und Kompetenzentwicklung ermöglicht, wird zwar angenommen aber empirisch nicht geprüft. Weiterhin unterliegen schulische Alltagsassessments den Annahmen der klassischen Testtheorie. Demnach zählt für die Bewertung einer Leistung ausschließlich die Summe der erzielten Punkte, nicht jedoch welche konkreten Aufgaben (mit welchem Schwierigkeitsgrad, zu welchem Inhaltsbereich oder welche konkreten Kompetenzfacetten betreffend) gelöst werden. Personen mit einer geringen Punktzahl müssen nach dieser Logik das vollständige Bündel von Aufgaben erneut vorgelegt bekommen, um ihre Kompetenz weiter zu entwickeln. Dies ist weder effizient noch entspricht es dem gewünschten förderlichen Individualisierungsgedanken.

**Individualdiagnostische Verfahren** der pädagogischen Psychologie (ii) sind im Gegensatz zu Alltagsassessments standardisiert und zeichnen sich durch ihren hohen Detailgrad an Informationen aus (ebd.). Damit sind pädagogische Instrumente der Individualdiagnostik in der Lage, in standardisierter Form individuelle Stärken-Schwächen-Profile von Lernenden zu erfassen. Zentraler Nachteil dieser Verfahren für eine fachdidaktische Anwendung in berufs- und wirtschaftspädagogischen Kontexten ist ihr fehlender Situationsbezug. D. h. diese Verfahren basieren nicht auf einem kompetenztheoretischen Verständnis von sich gegenseitig bedingenden Personen- und Situationsmerkmalen, sondern deren theoretische Grundlage fokussiert nur auf die Person als einzigen Einflussfaktor für eine gezeigte bzw. nicht gezeigte Leistung (NICKOLAUS/SEEBER 2013). Messtheoretisch basieren individualdiagnostische Verfahren der pädagogischen Psychologie zumeist auf klassischer Testtheorie, welche, wie oben erläutert, andere Schwierigkeiten im individualdiagnostischen Sinne mit sich bringt (WILD/KRAPP 2006).

Die **berufliche Leistungsmessung** (iii) ist kompetenztheoretisch fundiert und bedient sich – zumindest in weiten Bereichen (vgl. vor allem die ULME-Studien: SEEBER 2005; LEHMANN/SEEBER 2007 sowie die vom Bundesministerium geförderte ASCOT-Initiative BECK/LANDENBERGER/OSER 2016) – der Verfahren der probabilistischen Testtheorie (u. a. WINTHER/ACHTENHAGEN 2009; BAETHGE/SEEBER 2016; WEBER/BECK/KÖLLER 2016; WINTHER 2010) und erlaubt standardisierte Rückmeldungen. Probabilistische Verfahren beziehen im Gegensatz zur klassischen Testtheorie neben der erbrachten Gesamtpunktzahl auch das konkrete Antwortverhalten der Testteilnehmer auf jede Aufgabe in die Diagnose mit ein. Somit ist es möglich, eine gezieltere Auswahl von Aufgaben (z. B. nur schwere) zu treffen und

1 Abgesehen von Problemen der Objektivität (insbes. der Durchführungsobjektivität) und Reliabilität.

damit zielführende Lernprozesse anzustoßen. Zu den bekanntesten Verfahren der probabilistischen Testtheorie gehören das Rasch Model, das Partial Credit Model oder das 2 PL Model, welche der eindimensionalen Item-Response-Theorie (IRT) zugeordnet werden. Diese Modelle beinhalten die grundlegende Annahme, dass menschliche Kompetenzentwicklung als ein lineares Kontinuum von geringer bis hoher Fähigkeitsausprägung angesehen wird und dass sich Aufgabenschwierigkeiten und Personenfähigkeiten auf ein- und derselben Skala abbilden lassen (STROBL 2012; EMBRETSON/REISE 2000). Instrumente die mittels eindimensionaler IRT-Verfahren ausgewertet werden beinhalten allerdings häufig keine detaillierten individualisierten Kompetenzinformationen. D. h. der Fähigkeitswert eines Lerners lässt zwar Urteile darüber zu, welche Aufgaben welcher Schwierigkeitsstufe dieser mit hoher Wahrscheinlichkeit lösen kann, nicht jedoch welche individuellen Stärken und Schwächen dieser Fähigkeitswert beinhaltet. Das Ziel dieser Verfahren liegt eher auf Kompetenzbeschreibungen einer Population und Gruppenvergleichen – sowie auch im Kontext von Large-Scale-Assessments.

**Kognitive Diagnosemodelle** (cognitive diagnostic models: CDMs; DIBELLO/ROUSSOS/STOUT 2007) können die Vorteile probabilistischer Testverfahren und individualdiagnostischer Instrumente vereinen. Sie ermöglichen damit standardisierte, kompetenztheoretisch-fundierte und individualdiagnostische Aussagen, um Lernprozesse individuell begleiten zu können. Dabei wird das latente Konstrukt der übergreifenden Kompetenz über korrespondierende Kompetenzfacetten (hier nachfolgend Skills genannt) operationalisiert, die wiederum aufgabenbasierte beobachtbare Lern- bzw. Leistungsprozesse beschreiben. Das Beherrschen bzw. Nichtbeherrschen eines jeden Skills wird als Stärken-Schwächen-Profil für jeden Lerner ausgegeben (LEIGHTON/GIERL 2007b). Damit lassen sich über CDMs wahrscheinlichkeitsbasierte Informationen über die folgenden drei Aussagen generieren: (1.) den Prozentsatz der Population (z. B. aller Auszubildenden im Beruf Industriekaufmann/-frau), die einen spezifischen Skill beherrscht (**Skills**), (2.) über den Prozentsatz der Population, der ein spezifisches Stärken-Schwächen-Profil besitzt (**Skill-Klassen**) und (3.) über individuelle Stärken-Schwächen-Profile für jeden Lerner (**Kompetenzprofile**). Aus den Stärken-Schwächen-Profilen der Lernenden kann in detaillierter Form auf deren zugrundeliegende Lern- und Leistungsprozesse geschlossen werden, da davon ausgegangen wird, dass beherrschte Skills (Stärken) im Gegensatz zu nicht beherrschten Skills (Schwächen) im Lösungsprozess mit großer Wahrscheinlichkeit eingesetzt wurden (HELM u. a. 2015; GEORGE/ROBITZSCH 2015). Somit sind CDMs in der Lage, u. a. Kapazitäten Lernender zu beschreiben und kompetenztheoretische Grundlage zur Erklärung von Leistungsgraden anzubieten (mittels 1. Skills und 2. Skill-Klassen) Es besteht weiterhin die Möglichkeit passgenaue Instruktionen für Lernende mit gleichen Stärken-Schwächen-Profilen zu planen (mittels 2. Skill-Klassen und 3. aggregierten individuellen Kompetenzprofilen) und sie erlauben auf den jeweiligen Fähigkeitsstand eines Lernenden zugeschnittene Instruktionen und individualdiagnostische Empfehlungen und Selektionen für weitere Lernprozesse und Bildungswege (mittels 3. individuellen Kompetenzprofilen) (LEIGHTON/GIERL 2007b).

Im Rahmen dieses Beitrages soll anhand der kaufmännischen Kernkompetenz innovativen Handelns im Sinne einer „Intrapreneurship (IP)-Kompetenz“ ein CDM vorgestellt und die gewonnenen Ergebnisse im Hinblick auf die Potenziale für die Gestaltung von Lehr- und Lernprozessen – auch individualisierten – diskutiert werden. Grundlage der Studie sind die Entwicklungen und Datensätze des vom BMBF-finanzierten CoBALIT-Teilprojekts im Rahmen der ASCOT-Initiative (WEBER u. a. 2016a). Dieser Beitrag versteht sich daher als eine Art „Spin-off“ dieser als Large-Scale Assessment angelegten Studie. Mit der Anwendung von CDMs wählen wir einen alternativen Ansatz zu den vorliegenden IRT-basierten Ergebnissen und generieren zudem detaillierte individualdiagnostische Informationen, die inhärent Rückschlüsse über zugrundeliegende Leistungsprozesse im IP ermöglichen. (RUPP/TEMPLIN/HENSON 2010; DIBELLO/ROUSSOS/STOUT 2007; DE LA TORRE/MINCHEN 2014).

Das vorliegende Projekt fokussiert im Speziellen auf die IP Kompetenz von Auszubildenden zum Industriekaufmann/zur Industriekauffrau (WEBER u. a. 2014; BLEY/WIETHE-KÖRPRICH/WEBER 2015; WEBER u. a. 2015; WEBER u. a. 2016a; WEBER u. a. 2016b), da in diesem Ausbildungsgang ein nicht unerheblicher Teil der Jugendlichen im kaufmännischen Bereich ausgebildet wird (u. a. BUNDESINSTITUT FÜR BERUFSBILDUNG 2012) und der Ausbildungsberuf zu Industriekaufleuten als „Querschnittsberuf“ gilt (ACHTENHAGEN/WINTHER 2009).

Im Folgenden wird zunächst die theoretische Fundierung von CDMs vorgestellt, dann ein konkretes Beispiel für die Entwicklung eines CDMs im Kontext der Ausbildung einer Intrapreneurship-Kompetenz gegeben und schließlich fachdidaktische Potenziale für die Gestaltung von Lehr- und Lernprozessen auf den drei Ebenen (Population, Gruppe, Individuum) sowie im Hinblick auf drei Stakeholder der beruflichen Bildung (a) Programmorganisatoren, (b) Bildungspersonal und (c) Lernende hervorgehoben und kritisch diskutiert.

## 2. Theoretische Einbettung

### 2.1. Kognitive Diagnosemodelle (CDMs)

Ziel von CDMs ist die Extrahierung multidimensionaler individueller Stärken-Schwächen-Profile (RUPP/TEMPLIN/HENSON 2010; DIBELLO/ROUSSOS/STOUT 2007). Im vorliegenden Beitrag wird das übergreifende latente Merkmalskonstrukt als „(IP)-Kompetenz“, die multiplen Dimensionen als „(IP-)Skills“ bezeichnet. Psychometrisch betrachtet weisen CDMs Überschneidungen mit anderen psychometrischen Modellen auf: mit probabilistischen Modellen (z. B. die Messung latenter Merkmale oder die wahrscheinlichkeitsbasierte Analyse von Antwortmustern) sowie mit konfirmatorischen Faktorenanalysen (z. B. die komplexe Ladungsstruktur und die à priori zu spezifizierende Skill-Item-Zuordnungen) (KUNINA-HABENICHT/RUPP/WILHELM 2012). Diese Skill-Item-Zuordnungen (d. h. die klare Spezifizierung von einzusetzenden Skills zur Lösung einer bestimmten Aufgabe) werden in CDMs in der sogenannten Q-Ma-

trix (TATSUOKA 1983) zusammengefasst. Der Validität der à priori zu bestimmenden Q-Matrix kommt eine besondere Bedeutung zu und sollte daher möglichst elaboriert untersucht werden (BLEY in Begutachtung). Zentrale Unterschiede zwischen CDMs und den üblicherweise eingesetzten IRT-Modellen sind die Folgenden (siehe auch DE LA TORRE/MINCHEN 2014): (i) Die Dimension (Detailgrad) der theoretischen Fundierung und damit der erzielbaren Assessmentinformationen (IRT: ein bzw. wenige Dimensionen vs. CDM: multiple Kompetenzfacetten/Skills). (ii) Die Theoriebasis für fachdidaktisch relevante Informationen (IRT: Kompetenzniveaus werden **nach** der Modellschätzung abgeleitet vs. CDM: Kompetenzprofile auf Basis à priori definierter Skills und Q-Matrix, die in die Modellschätzung einfließt). Und (iii) die Ausprägung der Personenfähigkeiten (IRT: kontinuierlich vs. CDM: dichotom oder polytom). Im hier betrachteten dichotomen Fall werden daher die Skills in den Stärken-Schwächen-Profilen (KUNINA-HABENICHT/RUPP/WILHELM 2012) als Master („Skill-Beherrscher“) und Non-Master („Skill-Nicht-Beherrscher“) angegeben.

Die grundlegende Idee von CDMs ist, von den Skills, die zur Lösung eines Items benötigt werden (bekannt durch die Q-Matrix), und den empirischen Antworten der Personen, zu schließen, welche Skills die Testteilnehmer beherrschen<sup>2</sup>. Somit ermöglichen die Antwortmuster der Lernenden auf die Items und die Spezifizierung welche Skills zum Lösen der Items benötigt werden eine Klassifizierung der Lernenden als Master oder Non-Master für jeden Skill. Das nachfolgende Beispiel erläutert die CDM-Idee auf die hier fokussierte Domäne IP. Dabei sei unter Vorgriff auf die später präsentierten Validierungsbemühungen<sup>3</sup> zum diagnostischen IP Messinstrument bemerkt, dass IP Skills als nicht kompensierbar<sup>4</sup> angesehen werden: Wenn ein Auszubildender alle für ein Item notwendigen Skills beherrscht, so wird er/sie mit hoher Wahrscheinlichkeit dieses Item korrekt lösen. Wenn der Auszubildende jedoch nur einen der benötigten Skills nicht besitzt – z. B. domänenspezifische Tools anzuwenden („TOOL“) –, dann wird er/sie Items, deren Lösung den Skill „TOOL“ beinhaltet, mit hoher Wahrscheinlichkeit nicht korrekt lösen. Umgekehrt betrachtet: Wenn ein Auszubildender durchgängig an jenen Items scheitert, die die Beherrschung von „TOOL“ fordern, dann kann mit hoher Wahrscheinlichkeit diagnostiziert werden, dass dieser Auszubildende die domänenspezifischen Tools („TOOL“) nicht richtig anwenden kann.

2 An dieser Stelle unterscheiden sich die Annahmen von CDMs von jenen der derzeitigen Berufs- und Wirtschaftspädagogik, die eine domänenspezifische Aufgabe jeweils als vollständige Handlung (von der Wahrnehmung bis zur Reflektion) begreift und entsprechend alle Skills für eine Aufgabe als notwendig erachtet (OSER/SALZMANN/HEINZER 2009). Die Community ist sich allerdings auch einig, dass bei einer detaillierten Beschreibung kontextbasierte berufliche Situationen enorm variieren und das daher – obwohl alle Skills grundsätzlich notwendig sind – bestimmte Skills eine wichtigere Rolle spielen als andere. Unter der Annahme detaillierter diagnostischer Zwecke – wie sie mit dem CDM verfolgt werden – ist der Fokus auf die wichtigsten Skills essentiell. Die Identifikation der wichtigsten Skills erfolgt im Rahmen einer elaborierten Domänenanalyse sowie der Verbindung empirischer Evidenzen zur Sicherstellung der Validität (siehe Abschnitt 2.2)

3 Siehe dazu auch Abschnitt 2.2.

4 Skills sind nicht kompensierbar, wenn das Fehlen/Nicht-Beherrschen eines Skills nicht durch die Anwendung eines anderen Skills ausgeglichen werden kann (u. a. GEORGE/ROBITZSCH 2015).

CDMs bieten drei Ergebnisse (1. Skills, 2. Skill-Klassen, 3. Kompetenzprofile), die Schlussfolgerungen auf drei Ebenen erlauben: (I) Population, (II) Gruppe und (III) Individuum. Skills (1.) und Skill-Klassen (2.) werden dabei maßgeblich auf Populationssebene (I) relevant, die individuellen Kompetenzprofile (3.) sind für jedes einzelne Individuum (III) interessant und können auf aggregierter Ebene auch gruppenspezifische (II) Anwendungen ermöglichen.

- (I.) **Population:** Je nach gewählter Stichprobe und Erhebungsdesign, lässt sich für die entsprechende Population feststellen, wie viel Prozent der Lernenden über die einzelnen **Skills** (1.) bzw. über ein bestimmtes Stärken-Schwächen-Profil (= 2. **Skill-Klassen**) verfügen. Maximal gibt es  $2^K$  ( $K$  = Anzahl der Skills) verschiedene Skill-Klassen. Diese Ergebnisse ermöglichen Aussagen darüber, welche Skill-Klassen für die Domäne besonders wichtig sind. Die Bedeutung einer Skill-Klasse ergibt sich einerseits daher, dass sie überhaupt vorkommt (denn einige Skill-Klassen/Stärken-Schwächen-Profile sind ggf. nicht vertreten) und andererseits über den prozentualen Anteil dieser Skill-Klassen innerhalb der Population. Didaktisch ließen sich auf dieser Informationsbasis thematische und instruktionale Fokussierungen vornehmen und relevante, aber in der Population noch wenig ausgeprägte Skill-Klassen instruktional aufgreifen.
- (II.) **Gruppe:** Indem die individuellen **Kompetenzprofile** (3.) für eine Gruppe (z. B. Klasse) aggregiert werden, erlauben sie sowohl vergleichende Informationen mit den **Skill-Klassen** (2.) der Population aber auch adaptive lernförderliche Maßnahmen auf Instruktionsebene. So ließen sich diese Kriterien wie auch sozial bewertbaren Leistungsüberprüfungen für eine konstruktive Rückmeldung über die weitere eigene Unterrichtstätigkeit nutzen; so z. B. durch die Bildung homogener oder heterogener Lerngruppen sowie Maßnahmen der Binnendifferenzierung.
- (III.) **Individuum:** Individuelle **Kompetenzprofile** (3.) geben Auskunft darüber, welche Skills der/die Lernende beherrscht bzw. (noch) nicht beherrscht. Didaktisch bietet diese Information Grundlage für ein individuelles informatives Feedback, in dem Leistungsstände adressatengerecht transparent aufgezeigt aber auch Perspektiven für das weitere Lernen entwickelt werden.

Valide Schlussfolgerungen dieser Art sind nur unter der Annahme eines zuverlässigen zugrundeliegenden Messinstruments zulässig. Nach MESSICK (1989) kann von einem zuverlässigen Instrument nur dann gesprochen werden, wenn verschiedene Aspekte der Validität (insbesondere inhaltliche, kognitive und strukturelle) bestimmt und im Sinne empirischer Evidenzen integrativ interpretiert werden können. Die Einhaltung derartig strenger Designprinzipien für Assessments fordern auch CDM-Assessment-Experten (wie u. a. DIBELLO/ROUSSOS/STOUT 2007; RUPP/TEMPLIN/HENSON 2010; LEIGHTON/GIERL 2007a). Grundlegende Annahmen dieser Designprinzipien sind einerseits, dass ein Schließen auf latente Merkmalsausprägungen nur auf Basis beobachtbarer Evidenzen („evidence-based reasoning“: PELLEGRINO/CHUDOWSKY/GLASER 2001) erfolgen darf und dass andererseits die Kernbestandteile eines

Assessments (Kompetenzmodell, Aufgabenmodell und statistisches Messmodell) kohärent aufeinander abgestimmt sind (für eine intensive Diskussion siehe auch BLEY in Begutachtung oder GEORGE/BLEY/PELLEGRINO in Begutachtung). Empfohlene Rahmenkonzepte, um diese Kohärenz sicherzustellen, sind u. a. der „Assessment Triangle“ (C-O-I) bestehend aus Cognition (Kompetenzmodell), Observation (Aufgabenmodell) und Interpretation (statistisches Messmodell) (PELLEGRINO/CHUDOWSKY/GLASER 2001) sowie deren praktische Schritt-für-Schritt Umsetzung, wie zum Beispiel im Evidence-Centred-Design (ECD: MISLEVY/RISCONSCENTE 2006). Das ECD gliedert sich in die folgenden Schritte: 1. Spezifizierung des intendierten Assessmentziels, 2. Spezifizierung eines Modells latenter Teilkompetenzen/Skills, 3. Entwicklung und Validierung der Assessmentitems/-aufgaben sowie Spezifizierung und Validierung der Q-Matrix, 4. Spezifikation des psychometrischen Modells aus der Familie der kognitiven Diagnosemodelle, 5. Prüfung der Modellgüte und 6. Entwicklung eines Rückmeldesystems.

Die vorliegenden Befunde zur Anwendung von CDM-Prozeduren unterstreichen die Notwendigkeit dieses aufwendigen Entwicklungs- und Validierungsprozesses: Der häufig unternommene Versuch, existierende Messinstrumente, die originär für die Auswertung mittels IRT konstruiert wurden, für CDM Analysen zu verwenden (sogenannte „Retrofitting“ Studien), führte zumeist zu inakzeptablen Klassifikationsergebnissen (GIERL/CUI 2008; LEIGHTON/GIERL/HUNKA 2004). Gründe hierfür können z. B. in einer häufig mangelnden Abstimmung der Facetten des Assessment Triangles gefunden werden: Die Items (Observation: Schritt 3 im ECD) sind nicht in der Lage, die ex post spezifizierten Skills (Cognition: Schritt 2) abzubilden, sodass die Spezifikation des CDMs als psychometrisches Messmodell (Interpretation: Schritt 4) in inakzeptablen Modellgüteparametern resultiert (Schritt 5) und somit keine diagnostische Informationen für ein (individuelles) Feedback zur Förderung der intendierten Lernziele abgeleitet werden kann (Schritt 1 und Schritt 6). Andere Versuche unterschätzten die Bedeutung der in Aufgaben angewendeten domänenspezifischen Lehr- und Lernprozesse als Grundlage für die Spezifizierung der Q-Matrix (Schritt 3), die à priori definiert unmittelbar ins Messmodell einfließt (LEUDERS 2014). Aktuell gibt es nur wenige Messinstrumente, die den besonderen Herausforderungen von CDMs gerecht werden (DE LA TORRE/MINCHEN 2014). Die existierenden gut funktionierenden Instrumente beschreiben oftmals sehr enge Inhaltsbereiche (wie bspw. die Subtraktion von Brüchen: TATSUOKA 1983; LEIGHTON/GIERL/HUNKA 2004; TORRE 2008; TORRE/DOUGLAS 2004) mit gut untermauerten Theorien zum Lehren und Lernen in den Inhaltsbereichen und sehr schmal definierten Skills (z. B. „den größten gemeinsamen Nenner finden“ oder „eine ganze Zahl in einen Bruch umwandeln“). Nach Kenntnis der Autoren gibt es für komplexe berufliche Domänen (wie z. B. dem IP) bisher keine Assessments, die diagnostische Informationen zur Beschreibung von Leistungsprozessen ermöglichen.

Mit dem nachfolgenden Beispiel stellen wir das Vorgehen und die Ergebnisse der CDM-Assessmententwicklung anhand der 6 ECD-Schritte vor (für detaillierte Darstellungen zur Spezifizierung der Q-Matrix siehe BLEY, in Begutachtung; zu de-

taillierten Aspekten der CDM-basierten Berechnung der Kompetenz-Profile siehe GEORGE/ BLEY/ PELLEGRINO in Begutachtung).

## 2.2. Entwicklung eines kognitiven Diagnose-Assessments für Intrapreneurship

Auch wenn die folgende Darstellung der Designprinzipien zur Konstruktion des IP Assessments einer linearen Schritt-für Schritt-Anleitung folgt, war das Vorgehen stark iterativ (DIBELLO/ROUSSOS/STOUT 2007). Entsprechend sind in den nachfolgenden Schritten z. T. auch Vor- und Rückgriffe auf andere Schritte erforderlich.

### Schritt 1: Spezifizierung des intendierten Assessmentzieles

Ziel des Assessments soll es sein, diagnostische Informationen über Leistungsprozesse im IP von Auszubildenden zum Industriekaufmann/zur Industriekauffrau am Ende ihrer beruflichen Erstausbildung zu erhalten, die sowohl einen detaillierten Blick auf die in der Testpopulation beherrschten Kompetenzen erlauben wie auch in adaptiven und individualisierten Instruktionen Niederschlag finden sollen (u. a. NEUMANN 2013).

### Schritt 2: Beschreibung eines Modells latenter IP Skills

Für die Spezifikation latenter diagnostizierbarer Skills, muss zunächst (i) das übergreifende Konstrukt IP definiert sowie (ii) deren Legitimation in der Zielgruppe nachgewiesen werden. Folglich sind mittels einer elaborierten Domänenanalyse (iii) kritische Situationen bzw. typische Herausforderungen im IP bei Auszubildenden zu identifizieren und (iv) die dafür benötigten Skills auf einem angemessenen Detailgrad zu spezifizieren. Als „Spin-off“ Projekt zum Forschungsprojekt von Weber und Kollegen (WEBER u. a. 2014; WEBER u. a. 2016a) wird sowohl auf deren Definition als auch auf Materialien und Ergebnisse der Domänenanalyse<sup>5</sup> zurückgegriffen.

IP wird im Sinne eines innovativen Verhaltens (RUPPRECHT u. a. 2011) definiert (i) und lässt sich präzisieren über das selbstständige und eigenverantwortliche Initiieren innovativer Projektideen und die Umsetzung dieser Ideen durch kooperative Projektarbeit (WUNDERER 2003; KORUNKA u. a. 2009). (ii) Bereits seit 2002 ist die Ausbildung von IP-Kompetenzen in den Ordnungsmitteln für den Beruf Industriekaufmann/-frau verankert (WEBER u. a. 2014). Typische IP Projekte in der Ausbildungspraxis (iii) sind z. B. innovative Events zu planen und durchzuführen wie

5 Die umfangreiche Domänenanalyse, durchgeführt von Weber, Bley und Kollegen (BLEY/WIETHE-KÖRPRICH/WEBER 2015; WEBER u. a. 2014; WEBER u. a. 2015) umfasste das **Curriculum** (Ausbildungsordnung, Lehrplan), die **Instruktion** (Schulbücher: N = 6; Klassenraumbesichtigungen mit anschließenden Schülerbefragungen und Lehrerinterviews) sowie das **Assessment** (Analysen von IHK Projektberichten (N = 205) im Rahmen der mündlichen Abschlussprüfungen). Zusätzlich wurden Stellenanzeigen (N = 437) hinsichtlich der ökologischen Validität aber auch wissenschaftliche Literatur über Intra-/Entrepreneurship und innovatives Verhalten aus den Bereichen BWL, Psychologie und Pädagogik (N = 17) untersucht.

u. a. ein neues Produkt einzuführen oder eine Modenschau zu veranstalten. Repräsentative Herausforderungen innerhalb dieser Projekte sind Gelegenheiten wahrzunehmen, neue Ideen zu generieren, nach Informationen zu suchen und diese zu bewerten oder begründete Entscheidungen zu treffen. Das auf der Basis dieser umfassenden evidenz-basierten Domänenanalyse konzipierte Kompetenzmodell (iv) von Weber und Kollegen (WEBER u. a. 2015) umfasst drei Ebenen mit steigendem Detailgrad: zwei Dimensionen (Ideengenerierung und Umsetzung einer Projektidee), sechs Projektphasen (z. B. das Analysieren von Projektideen oder die Informationsrecherche) und ein detailliertes Bündel von 14 IP Skills (z. B. das Sequenzieren von Aspekten oder der Umgang mit Tools). Insbesondere die Analyse von IHK Abschlussprojekten erbrachte Evidenzen für die einzelnen Skills (z. B. die Nutzung einer Tabellenkalkulation als Evidenz für den Gebrauch von Tools oder die Auswahl eines Lieferanten als Evidenz für die Recherche und Verarbeitung von Informationen) wie sie auf dem Ausbildungslevel erbracht werden. Da die Anzahl der Skills zur Spezifizierung einer Domäne sowohl den benötigten Stichprobenumfang zur Schätzung eines CDMs wie auch die Präzision der Schätzung negativ im praktischen Sinne beeinflussen, wurden die 14 IP Skills in einem aufwendigen Validierungsprozess (BLEY in Begutachtung) auf ein Bündel von 7 IP Skills reduziert<sup>6</sup> (siehe Tabelle 1). Die Skill-Komprimierung erfolgte auf Basis verschiedener empirischer Evidenzen, die aus einer Triangulation verschiedener Quellen (9 Experten bestehend aus 5 Lehrkräften und 4 Auszubildern sowie 26 Auszubildende), Erhebungsmethoden (Expertenrating bzw. Studien Lauten Denkens) und Auswertungsmethoden (Auszählen von Ratings/Nennungen bzw. ein Vergleich von Lösern und Nicht-Lösern) generiert wurden (BLEY in Begutachtung).

Tab. 1: Beschreibung der 7 IP Skills

<b>Skills</b>	<b>Skill Beschreibung</b>
WAHR	Wahrnehmen von IP Chancen und Problemen
ANALY	Analysieren von IP Situationen
IDEE	Generieren von IP Ideen
SEQU	Sequenzieren von Aspekten
INFO+FACH	Verarbeiten von ökonomischen Fachbegriffen/Routinen und recherchieren und bewerten von Informationen
TOOL	Anwenden kaufmännisch/betriebswirtschaftlicher Tools
ENT+VERK	Treffen und Verkaufen von ökonomisch begründeten Entscheidungen

6 Wenn sich z. B. zeigte, dass einige Skills häufig zusammen in einer Aufgabe benötigt wurden (z. B. „das Verarbeiten von ökonomischen Fachbegriffen/Routinen“ (FACH) und „das Recherchieren und Bewerten von Informationen“ (INFO)), wurden diese schließlich in einem Skill (INFO+FACH) zusammengefasst (siehe Tabelle 1) (BLEY, in Begutachtung).

### Schritt 3: Entwicklung und Validierung der Assessmentitems/-aufgaben sowie Spezifizierung und Validierung der Q-Matrix

Ein valides Assessment zur Beschreibung von beruflichen Kompetenzen impliziert eine authentische Modellierung beruflicher Handlungen, in denen Auszubildende denken und handeln wie an ihren Arbeitsplätzen (DARLING-HAMMOND u. a. 2013; JANESICK 2006; PELLEGRINO/DIBELLO/GOLDMAN 2016; SHAVELSON 2012; WEBER u. a. 2014; WINTHER/ACHTENHAGEN 2009). Dementsprechend wurden im Rahmen des ASCOT-Projekts 22 IP Items<sup>7</sup> entwickelt (WEBER u. a. 2016a), die typische IP Herausforderungen, wie sie im Rahmen der Domänenanalyse identifiziert wurden, reflektieren. Außerdem wurden alle Items als Erweiterung technologiebasiert in die Unternehmens-Plattform ALUSIM GmbH (ACHTENHAGEN/WINTHER 2014) implementiert. Die Plattform ist einem real existierenden mittelständischen Unternehmen nachempfunden und enthält ERP-gestützte Systeme (ERP: = Enterprise Resource Planning) sowie domänenspezifische Endbenutzerwerkzeuge (wie z. B. Tabellenkalkulations- oder Textverarbeitungsprogramme). Videobasierte Szenarien begleiten durch die Situationen. Die Testpersonen (die Auszubildenden) übernehmen die Rolle eines Auszubildenden der ALUSIM GmbH und lösen Aufgaben/Items zweier Unternehmensprojekte, die sie per E-Mail als Arbeitsaufträge erhalten.

Abbildung 1 präsentiert das technologiebasiert eingebettete Item 3. Der Auszubildende erhält per E-Mail den Auftrag einen GANTT-Plan für das im Teammeeting diskutierte neue Projekt zu erstellen. Die notwendigen Unterlagen (das unternehmensinterne Template, das Teammeeting-Protokoll mit den definierten Arbeitspaketen sowie ein GANTT-Plan aus dem Vorprojekt) werden zur Verfügung gestellt. Für eine erfolgreiche Lösung des Items spezifizierte das fachdidaktisch erfahrene Forscherteam drei Skills als notwendig: INFO+FACH, SEQU und TOOL. Zum einen muss der Auszubildende die fachspezifische Routine des GANTT-Plans und deren Anwendung kennen sowie relevante zeitliche Informationen der Arbeitspakete sind aus dem Teammeeting-Protokoll zu extrahieren (INFO+FACH). Unter Berücksichtigung zeitlicher Ressourcen und Abhängigkeiten sind die Arbeitspakete zu sequenzieren (SEQU) und durch Nutzung der Markierfunktion in der Tabellenkalkulation in das Unternehmenstemplate einzutragen (TOOL) (ebd.). Entsprechend wurden für alle 22 Items die zur Lösung benötigten Skills spezifiziert und in einer initialen Q-Matrix zunächst aus der Perspektive der Forscher aufgestellt (BLEY in Begutachtung). Die anschließende Überprüfung sowohl der Qualität der Items als auch der initialen Q-Matrix ergab, dass alle 22 Items mindestens einen IP Skill zur Lösung benötigten. Zwei weitere Items wurden neu konstruiert, da der inhaltlich sehr bedeutsame Skill „IDEE“ in den Aufgaben bisher nicht ausreichend repräsentiert war (BLEY in Begutachtung; WEBER u. a. 2016b).

7 Die Aufgaben zur Messung von IP in der beruflichen Erstausbildung wurden für zwei unterschiedliche Zielsetzungen konstruiert (1. wie vom Projektgeber intendiert mittels IRT und 2. zur Generierung diagnostischer Aussagen mittel CDM). Obwohl beide Assessments (IRT vs. CDM) die gleichen Aufgaben enthalten, ist deren Aufgabenzuschnitt verschieden.

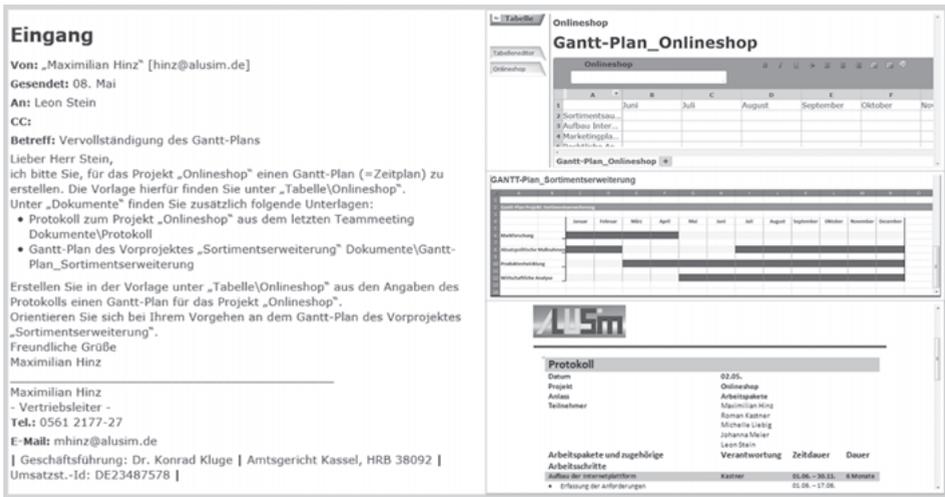


Abb. 1: Beispiel Item 3: „Erstellung eines GANTT Plans“ (Quelle: WEBER u. a. 2014)

Da es keinen direkten Zugang zu den kognitiven Prozessen der Auszubildenden beim Lösen der Aufgaben gibt, wurde eine Q-Matrix aus einer Triangulation verschiedener Quellen, Erhebungs- und Auswertungsmethoden erarbeitet. Die empirische Modellgüteprüfung (unter Einbezug verschiedener probabilistischer Modellgüteparameter; siehe auch Tabelle 2) erbrachte, dass diese Q-Matrix (Abbildung 2) unter allen diskutierten Alternativen die beste Abbildung realer Lösungsprozesse darstellt (BLEY in Begutachtung). In der ursprünglichen Entwicklung der Q-Matrix wurde sich bemüht, jeden Skill durch eine Mindestanzahl an Items abzudecken. Im Prozess der Validierung ergaben sich durch Zusammenfassungen von Skills bzw. Items diesbezüglich Anpassungen.

Item	WAHR	ANALY	IDEE	SEQU	TOOL	INFO+FACH	ENT+VERK
1	0	0	0	1	0	0	0
2	0	0	0	1	0	0	0
3	0	0	0	1	1	1	0
4	0	0	0	0	0	1	0
5	0	0	0	0	0	1	0
6	0	0	0	0	1	1	0
7	0	0	0	0	0	1	0
8	0	1	0	0	0	0	1
9	0	1	0	0	0	1	0
10	0	0	0	0	0	1	1
11	0	1	0	0	0	1	0
12	0	1	0	0	0	0	0
13	0	0	0	0	0	1	1
14	1	1	0	0	0	1	0

Item	WAHR	ANALY	IDEE	SEQU	TOOL	INFO+FACH	ENT+VERK
15	0	0	0	0	0	1	1
16	1	1	0	0	0	1	0
17	0	1	0	0	0	0	0
18	1	1	1	0	0	0	0
19	0	1	0	0	0	1	1
20	0	0	0	0	0	1	1
21	0	0	1	0	0	0	0
22	1	1	1	0	0	0	0

Anmerkungen: 1 = notwendig; 0 = nicht notwendig

Abb. 2: Q-Matrix<sup>8</sup> (BLEY in Begutachtung)

#### Schritt 4: Spezifizierung des psychometrischen Modells aus der Familie der kognitiven Diagnosemodelle

Zur Analyse des IP-Assessments wurde das Deterministic Input Noisy And Gate Model (DINA; Haertel 1989), ein sparsames, einfach zu interpretierendes und nicht-kompensatorisches Modell (alle zugewiesenen Skills müssen zur Lösung eines Items beherrscht werden) aus der Familie der CDMs, gewählt (GEORGE/BLEY/PELLEGRINO in Begutachtung).

#### Schritt 5: Prüfung der Modellgüte

Die Modelgüteprüfung wird mit ähnlichen Maßen, wie in der IRT (DE LA TORRE/MINCHEN 2014) üblich, vorgenommen. Alle Analysen wurden mit dem R (R CORE TEAM 2016) Paket CDM (GEORGE u. a. 2016) durchgeführt und evaluiert. Der absolute Modellfit eines CDM Modells kann anhand des standardized root mean square residual (SRMSR; MAYDEY-OLIVARES 2013) bewertet werden. Nach MAYDEU-OLIVARES signalisieren SRMSR-Werte kleiner als 0,05 einen guten Fit zwischen Modell und Daten.

Insgesamt weist das DINA-Modell einen guten globalen Modellfit auf (SRMSR = 0,046). Die Itemschwierigkeiten (Tabelle 2) liegen im Mittel bei 0,542 (d. h. im Mittel lösen 54,2 % der Auszubildenden die Items) und weisen eine Spanne von 0,174 bis 0,909 auf. Dies deutet auf einen hinsichtlich der Schwierigkeiten sehr ausgewogenen Test hin. Die Itemparameter Guessing- und Slipping-Parameter weisen keine Auffälligkeiten auf. Die Item-Diskriminationen zeigen mit einem Mittelwert von 0,403 eine zufriedenstellende Trennung der Items zwischen Auszubildenden, die die zugewiesenen Skills beherrschen bzw. nicht beherrschen. Interpretiert an der Faustregel, dass RMSEA Item Fit Werte kleiner als 0,05 einen guten Fit aufweisen (KUNINA-HABENICHT/RUPP/WILHELM 2009), sind die Items bis auf wenige Ausnahmen geeignet ( $M = 0,047$ ,  $Min = 0,031$ ,  $Max = 0,072$ ,  $SD = 0,010$ ). Die paarweisen Korre-

8 Im Prozess der strukturellen Validitätsprüfung wurden zwei hoch korrelierende Itempaare zusammengefasst, sodass die hier präsentierte Q-Matrix 22 Items statt 24 Items (BLEY in Begutachtung) umfasst.

Tab. 2: Deskription der Itemschwierigkeiten, der Itemparameter (Guessing, Slipping, Diskrimination), des RMSEA Itemfit und den Korrelationen zwischen den 7 Skills im DINA Modell.

	Min	Max	M	SD
Itemschwierigkeit	0,174	0,909	0,542	0,205
Guessing	0,006	0,564	0,264	0,161
Slipping	0,011	0,747	0,328	0,206
Item-Diskrimination	0,115	0,644	0,403	0,151
RMSEA Itemfit	0,031	0,072	0,047	0,010
Skill-Korrelation	0,028	0,991	0,608	0,244

Anmerkungen: Min = minimaler Wert, Max = maximaler Wert, M = Mittelwert, SD = Standardabweichung.

lationen zwischen den sieben Skills liegen minimal bei 0,028 („IDEE“ und „TOOL“) und maximal bei 0,991 („INFO“ und „SEQU“). Der mittlere Wert von 0,608 kann als Hinweis dafür gewertet werden, dass es sich bei den Skills tatsächlich um empirisch unterscheidbare Konstrukte handelt. Es wird auf eine weitere Zusammenfassung der Skills verzichtet, um den Informationsgehalt nicht weiter einzuschränken. Gemessen an der Gesamtheit der betrachteten psychometrischen Kriterien gibt es keine Bedenken an der strukturellen Validität des Instrumentes. Somit kann das Modell zur Diagnose der Fähigkeiten von Auszubildenden eingesetzt werden.

### Schritt 6: Entwicklung eines Rückmeldesystems

Dieser Schritt beinhaltet die Entwicklung eines Rückmeldesystems für das neu entwickelte Instrument. Das im Schritt 1 formulierte Assessmentziel war die Generierung diagnostischer Informationen über Leistungsprozesse im IP von Auszubildenden zum Industriekaufmann/zur Industriekauffrau. Der vorliegende Beitrag unternimmt den Versuch, die Potenziale dieser diagnostischen Informationen auf den verschiedenen Ebenen (I. Population, II. Gruppe, III: Individuum) für unterschiedliche Interessengruppen nutzbar zu machen. Entsprechend werden im Abschnitt 2.3 zunächst theoretische Überlegungen für ein Rückmeldesystem formuliert. Deren Prüfung erfolgt im Ergebnisteil.

#### 2.3. Theoretische Überlegungen zu Rückmeldungen aus Assessments

Nach (LEISS/RAKOCZY 2013) lassen sich drei Interessengruppen für Assessment-Rückmeldungen benennen: (a) Programmorganisatoren, (b) Bildungspersonal und (c) Lernende, die aufgrund ihrer unterschiedlichen Zielsetzungen unterschiedliche Informationen benötigen (KÖHLER/SCHRADER 2004):

Programorganisatoren (a) wie Schulämter, Ministerien oder Träger betrieblicher Ausbildung (BIBB/IHK) benötigen einen angemessenen Aggregationsgrad der gewonnenen Ergebnisse (Populationsebene), um die Qualität der Programme beur-

teilen und entwickeln zu können und ggf. organisatorische sowie inhaltliche (curriculare) Modifikationen zu realisieren (Populationsebene) (LEISS/RAKOCZY 2013; KOHLER/SCHRADER 2004).

Eine prominente Facette professionellen Könnens bildet die diagnostische Kompetenz von Bildungspersonal (b) (BAUMERT/KUNTER 2006). Diese Kompetenz umfasst (A) die kriterien- und adressatengerechte Formulierung kognitiv aktivierender Aufgaben auf Basis evidenzbasierter Kompetenzstrukturen (Aufgabenkonstruktion), (B) den zielgerichteten Einsatz von didaktischen Maßnahmen für differenzierte maßgeschneiderte Fördermaßnahmen auf Basis von Leistungsüberprüfungen (Differenzierung) und (C) transparente und adressatengerechte Rückmeldung individueller Leistungsstände, um Perspektiven zum Weiterlernen zu eröffnen (KMK 2004; PANT 2013). Dementsprechend sind Informationen sowohl auf einer populationsbasierten Ebene (I) zur Erstellung kognitiv aktivierender Aufgaben (A), auf einer gruppenbasierten Ebene (II) zur differenzierten Unterrichtsgestaltung (B) als auch auf einer individuellen Ebene (III) zur transparenten Rückmeldung von Leistungsständen (C) für das Bildungspersonal relevant.

Nach der Selbstbestimmungstheorie von DECI und RYAN (1993) können Rückmeldungen aus Lernergebnissen insbesondere dann förderlich sein, wenn damit ihr grundlegendes Bedürfnis nach Kompetenz unterstützt wird. D.h. den Lernenden wird aufgezeigt, welche Anforderungen sie bereits erfüllen und welche Potenziale zum Weiterlernen es gibt. Durch diese individuelle statt einer sozial vergleichenden Rückmeldung wird zugleich dem Bedürfnis einer autonomen, selbstbestimmten Lernaktivität entsprochen (LEISS/RAKOCZY 2013). Folglich benötigen Lernende (c) lernförderliche, informative und transparente Feedbacks (III. Individualebene), um ihre Distanz zum Lernziel einschätzen zu können und Potenziale zum Weiterlernen zu erkennen (HATTIE/TIMPERLEY 2007).

### 3. Forschungsfragen

Ziel dieser Studie ist es, Möglichkeiten fachdidaktischer Nutzung von CDMs für die IP Kompetenz bei Auszubildenden zum Industriekaufmann/zur Industriekauffrau herauszuarbeiten. Dafür werden im ersten Teil (FF 1) die CDM Assessmentinformationen für das vorliegende Instrument erarbeitet. Im zweiten Teil (FF 2) werden die erzielten Ergebnisse an die Erfordernisse der unterschiedlichen Interessengruppen gespiegelt und dem Nutzen derzeitiger IRT-basierter Verfahren der beruflichen Kompetenzdiagnostik gegenübergestellt.

FF 1: CDMs bieten Assessmentinformationen:

- FF 1.1 Skills: Wie viel Prozent der Auszubildenden beherrschen die einzelnen IP-Skills?
- FF 1.2 Skill-Klassen: Welche bedeutsamen Skill-Klassen lassen sich identifizieren?
- FF 1.3 Kompetenzprofile: Welche IP-Skills beherrscht ein Auszubildender?

FF 2: Nutzbarmachung unterschiedlicher Assessmentinformationen aus CDM Analysen:

- FF 2.1 (a) Welche Informationen aus CDM Analysen bieten Potentiale für Programmorganisatoren?
- FF 2.2 (b) Welche Informationen aus CDM Analysen bieten Potentiale für Bildungspersonal?
- FF 2.3 (c) Welche Informationen aus CDM Analysen bieten Potentiale für Lernende?

#### 4. Daten und methodisches Vorgehen

##### Design, Stichprobe und Durchführung

Für die Beantwortung der Forschungsfrage 1 wird auf Daten aus der ASCOT-Forschungsinitiative aus dem Verbundprojekt CoBALIT (Competencies in the Field of Business and Administration, Learning, Instruction, and Transition: (WEBER u. a. 2016a) zurückgegriffen. Die Erhebung wurde im Herbst 2014 als Eingruppen-Posttest-Design in sieben Bundesländern (Baden-Württemberg, Bayern, Hessen, Niedersachsen, Nordrhein-Westfalen, Rheinland-Pfalz, und Thüringen) durchgeführt. In die Analysen werden 919 Probanden im Alter von durchschnittlich 21 Jahren einbezogen. Von den 317 männlichen und 554 weiblichen Auszubildenden (48 ohne Angabe) besitzt ein Großteil eine Hochschulzugangsberechtigung (Abitur/Fachhochschulreife: 578, mittlere Reife: 279, Hauptschulabschluss/sonstiger Abschluss: 12, keine Angabe: 50). Die Ausbildung erfolgt in kleinen (1–50 Mitarbeiter:  $n = 105$ ), mittleren (51–500 Mitarbeiter:  $n = 499$ ) und großen Betrieben (mehr als 500 Mitarbeiter:  $n = 282$ ) Unternehmen ( $n = 33$  ohne Angabe).

Die Erhebung wurde von geschulten Testleitern durchgeführt, dauerte vier Stunden und fand während der regulären Unterrichtszeit statt. Sie umfasst neben der Bearbeitung der simulierten Arbeitsaufträge zum IP (ca. 100 min) auch Aufgaben zu kaufmännischen Geschäfts- und Steuerungsprozessen, zur wirtschaftsbürgerlichen Kompetenz, zu den Basisqualifikationen Mathematik und Lesen, sowie zum sozio-kulturellen Kontext (WINTHER u. a. 2016; WEBER u. a. 2016a).

##### Analyse

Zur Beantwortung der FF 1 wird das Deterministic Input Noisy And Gate Model (DINA; Haertel 1989) gewählt (siehe auch Abschnitt 2.2, Schritt 5). Zur Beantwortung der FF 2 wird der Nutzen der in FF 1 erzielten Ergebnisse fachdidaktisch für die einzelnen Interessengruppen analysiert und Ergebnissen derzeitiger IRT-Verfahren gegenübergestellt.

## 5. Ergebnisse

### 5.1. Diagnostische Assessment-Informationen (FF 1)

**FF 1.1 (Skills):** Für die Beantwortung der FF 1.1 wurden die Beherrschungs-Wahrscheinlichkeiten der sieben einzelnen IP-Skills berechnet (Tabelle 3). Die Auszubildenden zur Industriekauffrau/zum Industriekaufmann weisen am Ende ihrer beruflichen Erstausbildung für die sieben definierten IP Skills Beherrschungswahrscheinlichkeiten zwischen 55 % und 85 % auf. Auch wenn dieses Ergebnis aus diagnostischer Sicht zunächst unbefriedigend erscheint, ist es doch mit Blick auf den Ausbildungsfortschritt der untersuchten Gruppe als sehr positiv zu bewerten: Auszubildende am Ende ihrer beruflichen Erstausbildung sollten einen möglichst hohen Fähigkeitsstand aufweisen. Die Standardfehler der geschätzten Wahrscheinlichkeiten sind als unauffällig zu bewerten. Besonders schwer fällt den Auszubildenden der Skill „ENT-VERK“, dessen Beherrschungs-Wahrscheinlichkeit unter 60 % liegt. Ebenfalls vergleichsweise geringe Beherrschungs-Wahrscheinlichkeiten weisen die Skills „IDEE“ (75 %) und „WAHR“ (77 %) auf.

Tab. 3: Beherrschungs-Wahrscheinlichkeiten und Standardfehler der IP Skills

Skill $\alpha_k$	Beherrschungs-Wahrscheinlichkeit	Standardfehler
WAHR	0,767	0,039
ANALY	0,794	0,034
IDEE	0,750	0,029
SEQU	0,805	0,017
INFO-FACH	0,803	0,022
TOOL	0,851	0,019
ENT-VERK	0,545	0,033

Dieser Befund ist fachdidaktisch plausibel. So zeigte sich dieser Befund auch vergleichbar in einer wissenschaftlichen Re-Analyse von schriftlichen IHK-Projektberichten (WEBER u. a. 2015) – als Teilprodukt der Abschlussprüfung zur Industriekauffrau/zum Industriekaufmann. Hierbei wurden die Projektberichte hinsichtlich der Teilkompetenzen (i. S. der Skills) von IP mittels Frequenzanalyse ausgewertet. Diese ergab, dass für die Skills „WAHR“, „IDEE“ und „ENT-VERK“ in Relation zu anderen Skills eher selten Evidenzen gefunden und diese entsprechend kodiert werden konnten. Ein weiterer Grund für die im Vergleich zu den anderen Skills niedrige Beherrschungs-Wahrscheinlichkeit könnten mangelnde Lerngelegenheiten sein. An die ASCOT-Haupterhebung angeschlossen war eine Befragung der unterrichtenden Lehrer zu Lerngelegenheiten im Lernfeld 12. Hierbei wurden in Analogie zu einem typischen Projektablauf sowie in Analogie zu den Inhalten des Lernfeldes und zum definierten Kompetenzmodell sechs Aspekte hinsichtlich ihrer Intensitäten im Unterricht bewertet: „Projektziel und -beschreibung“, „Entwicklung einer Projektidee“,

„Projektplanung“, „Projektdurchführung“, „Projektrisiken und -bewertung“ und „Verkaufen von Projekten und Projektideen“. Die drei Phasen „Projektziel und -beschreibung“, „Entwicklung einer Projektidee“ und „Verkaufen von Projekten“, welche unmittelbar mit den Skills „WAHR“, „IDEE“ und „ENT-VERK“ korrespondieren, wiesen dabei kleinere Intensitäten auf als die durchschnittlich bewertete Intensität.

**FF 1.2 (Skill-Klassen):** Im Hinblick auf die Identifikation der Skill-Klassen lassen sich theoretisch aus den sieben Skills  $2^7 = 128$  Skill-Klassen bilden. Es wurden 10 Skill-Klassen mit Auftretenswahrscheinlichkeit größer als 1 % identifiziert (Tabelle 4).

Tab. 4: Skill-Klassen mit Beherrschungs-Wahrscheinlichkeiten größer 1 %

lfd. Nr.	Beherrschungs- Wahrscheinlichkeiten (Standardfehler)	Skill-Klassen						
		WAHR	ANALY	IDEE	SEQU	INFO + FACH	TOOL	ENT + VERK
1	0,485 (0,036)	1	1	1	1	1	1	1
2	0,054 (0,017)	1	1	0	1	1	1	0
3	0,042 (0,014)	1	0	1	1	1	1	0
4	0,038 (0,018)	1	1	1	0	0	0	0
5	0,033 (0,010)	1	0	0	1	1	1	0
6	0,032 (0,015)	0	0	1	1	0	1	0
7	0,028 (0,001)	0	0	0	1	1	1	0
8	0,028 (0,007)	0	0	0	0	0	0	0
9	0,027 (0,013)	0	1	1	0	1	1	0
10	0,025 (0,016)	0	1	0	1	1	1	1

Anmerkung: 0 = wird nicht beherrscht, 1 = wird beherrscht

Diese 10 Skill-Klassen umfassen 79,2 % der Population. Die mit Abstand am häufigsten vorkommende Klasse ist „111111“. Die 48,5 % Auszubildenden, die dieser Gruppe zugeordnet werden, beherrschen alle Skills. Alle nachfolgenden Skill-Klassen (ab lfd. Nr. 2) sind unvollständige Kompetenzprofile, d.h. einige Skills werden beherrscht andere nicht. 2,8 % der Auszubildenden beherrschen keinen der sieben IP-Skills (Skill-Klasse 8). Wiederum sei bemerkt, dass der hohe prozentuale Anteil eines vollständigen Kompetenzprofils („111111“) unter Beachtung der untersuchten Gruppe (Auszubildende am Ende ihrer beruflichen Erstausbildung) als positiv zu bewerten ist. Dieser Befund korrespondiert mit der Prüfungsstatistik der IHK<sup>9</sup>. Bei allen unvollständigen Kompetenzprofilen (ab lfd. Nr. 2), beherrschen z.B. Personen der zweithäufigsten Skill-Klasse (mit 5,4 % Beherrschungswahrscheinlichkeit) alle Skills außer „IDEE“ und „ENT+VERK“. Personen der Skill-Klasse 7 beherrschen genau jene Skills („SEQU“, „INFO+FACH“ und „TOOL“) die erforderlich sind, um die im

9 <http://pes.ihk.de/berufsauswahl.cfm?Auswahl=1>

Abschnitt 2.2 (Schritt 3) eingeführte Beispielaufgabe des GANTT-Plan erfolgreich lösen zu können. Aber auch Personen der Skill-Klassen 1, 2, 3, 5 und 10 sollten diese Aufgabe beherrschen.

**FF 2.3 (Kompetenzprofil):** Individuelle Kompetenzprofile auf Basis des probabilistischen DINA-Modells (Abbildung 3) umfassen für jeden der sieben IP-Skills eine individuelle Skill-Beherrschungs-Wahrscheinlichkeit (zu beachten ist, dass das Modell nicht nur die einzelnen Skills und die damit zusammenhängenden Aufgaben, sondern auch Zusammenhänge zwischen den Skills berücksichtigt).

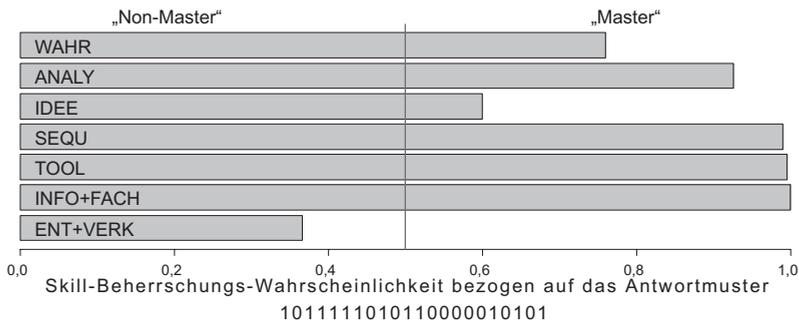


Abb. 3: individuelles Kompetenzprofil

Das in Abbildung 3 dargestellte Kompetenzprofil ergibt sich aus dem Antwortmuster „1011111010110000010101“ auf die 22 IP Items. Beherrschungs-Wahrscheinlichkeiten größer als 0,5 können nach diesem Vorschlag als beherrscht („Master“) interpretiert werden (hier: „WAHR“, „ANALY“, „IDEE“, „SEQU“, „TOOL“ und „INFO+FACH“); Beherrschungs-Wahrscheinlichkeiten kleiner als 0,5 als nicht beherrscht („Non-Master“; hier: „ENT+VERK“).

## 5.2. Potentiale CDM-basierter Assessment-Informationen (FF 2)

Die Ergebnisse aus dem CDM Assessment zur mehrdimensionalen Messung von IP-Kompetenzen in Form von Skills sollen hinsichtlich der Nutzbarkeit für verschiedene Interessengruppen analysiert werden. Dafür werden für jede Interessensgruppe (FF 2.1–2.3) die CDM-Assessmentinformationen (FF 1.1–1.3) diskutiert (vgl. auch Tabelle 5) und IRT-basierten Assessmentinformationen gegenübergestellt.

**FF 2.1 (Programmorganisatoren):** Für Entscheidungsträger auf Programmebene sind insbesondere Assessment-Informationen über die Population von Relevanz. Hier bieten sich sowohl die Ergebnisse durchschnittlicher Beherrschungs-Wahrscheinlichkeiten einzelner Skills (FF 1.1) als auch über bedeutsame Skill-Klassen (FF 1.2) an. Die IP CDM Ergebnisse liefern die detaillierte und doch populationsbasierte Information, dass Herausforderungen, die die IP-Skills „ENT+VERK“, „IDEE“ und „ANALY“ insgesamt am schwersten zu bewältigen sind (FF 1.1). Diese Informa-

Tab. 5: Potentiale CDM- basierter Assessment-Informationen

		Art der Assessmentinformation	
		<b>Skills (FF 1.1)</b> Prozentsätze durchschnittlicher Beherrschungs-Wahrscheinlichkeiten theoretisch fundierter detaillierter Skills	<b>Kompetenzprofile (FF 1.3)</b> Individuelle Stärken-Schwächen-Profile auf fundierter Basis (Skills und Q-Matrix)
		<b>Skill-Klassen (FF 1.2)</b> Prozentsätze bedeutsamer Stärken-Schwächen-Profile auf fundierter Basis (Skills und Q-Matrix)	
<b>Nutzen der Assessmentinformationen für unterschiedliche Interessengruppen</b>			
<b>Ziele der Interessengruppen</b>		<b>Gruppe</b>	<b>Individuum</b>
<b>Programm-organisatoren (FF 2.1)</b>	Bildungsmonitoring/ Curricula	die Aufsummierung der individuellen Kompetenzprofile einer Gruppe/Klasse ermöglicht Vergleiche mit anderen Gruppen/Klassen	<i>ein einzelnes Profil ist i. d. R. von untergeordnetem Interesse</i>
<b>Bildungspersonal (FF 2.2)</b>	Aufgabenkonstruktion (A) Didaktische Maßnahmen (B) Perspektiven zum Weiterlernen (C)	aggregierte individuelle Kompetenzprofile liefern gruppenspezifische Hinweise für adaptive didaktische Maßnahmen (B) (detaillierte Klassifizierung) Gruppenumfänge können mit bedeutsamen Skill-Klassen verglichen werden	individuelle Stärken (beherrschte Skills)-Schwächen (nicht beherrschte Skills) Rückmeldungen beschreiben die Distanz zum Lernziel und zeigen Möglichkeiten zum Weiterlernen auf (C)
<b>Lernende (FF 2.3)</b>	Distanz der eigenen Leistung zum Lernziel	Vergleich des individuellen Kompetenz-Profiles mit aggregierten Kompetenzprofilen der eigenen Gruppe/Klasse (Sozialnorm)	Vergleich des individuellen Kompetenz-Profiles mit dem Lernziel (Kriteriumsnorm = Liste an Skills) /im Zeitverlauf (Bezugsnorm)

tion lässt sich sowohl vergleichend (zwischen Programmen, Klassen etc.) als auch kriterienorientiert (am Lern-/Programmziel) bewerten. Im Vergleich zu kontinuierlichen Personenfähigkeiten-Verteilungen aus (eindimensionalen) IRT Analysen, die i. d. R. auf eher grobe Fähigkeitsdimensionen rekurrieren, liefern CDM-basierte Informationen stärker detaillierte (d. h. mehrdimensionale) Auskünfte diagnostizierbarer Skills einer Domäne (hier: sieben IP Skills), die kriterienorientierte Hinweise für inhaltliche Fokussierungen in Programmen und Curricula liefern. So wäre seitens der Programmorganisation z. B. anzudenken, Herausforderungen, die eine begründete Entscheidung sowie die Generierung neuer Ideen in den Mittelpunkt setzen, besonders einzufordern.

FF 1.2 liefert Informationen über bedeutsame Skill-Klassen im IP – also darüber, welche unterschiedlichen Stärken-Schwächen-Profile häufig vorkommen. Auch diese können seitens der Programmorganisation vergleichend und inhaltlich fokussierend genutzt werden. Bezogen auf die Kompetenz von Industriekaufleuten am Ende ihrer Berufsausbildung zeigt sich, dass fast die Hälfte der Auszubildenden alle Skills der IP-Kompetenz beherrscht und ein weiterer Teil von ca. 10 % fünf von sieben IP-Skills beherrscht (Skill-Klassen 2, 3 und 10). Insgesamt kristallisieren sich neun relevante Stärken-Schwäche-Profile heraus, die Anregungen zum Weiterlernen oder zur Konstruktion von Unterstützungsmaterialien anbieten. Stärken-Schwächen-bezogene Aussagen aus IRT-Analysen werden i. d. R. in Form von Kompetenzniveau-modellen abgebildet (u. a. BAETHGE/SEEGER 2016). Dabei werden Personenfähigkeiten in, von der Fähigkeit her ansteigende Leistungsniveaus gegliedert (i. d. R. drei bis fünf), die einen linearen Kompetenzzuwachs implizieren. Dies bedeutet, es wird angenommen, dass Lernende einer höheren Niveaustufe mit hoher Wahrscheinlichkeit auch Aufgaben einer niedrigeren Niveaustufe beherrschen. Diese Annahme könnte diskutiert werden (DE LA TORRE/KARELITZ 2009). Die Niveaustufen lassen sich anhand schwierigkeitsgenerierender Aufgabenmerkmale (z. B. kognitive Anforderungen oder domänenspezifische Komplexität; vgl. z. B. KLOTZ 2015) beschreiben, aus denen wiederum auf dahinterliegende Fähigkeiten geschlossen wird. Alternativ zu den basierend auf IRT-Analysen entwickelten Niveaustufen liefern CDMs damit Skill-Klassen, die keine Hierarchieannahme implizieren.

**FF 2.2 (Bildungspersonal):** Für das Bildungspersonal sind alle CDM Ergebnisse (FF 1.1 – 1.3) sinnvoll und hilfreich. Informationen auf Skill-Ebene (FF 1.1) können genutzt werden, um kognitiv aktivierende Aufgaben (A) auf Basis evidenzbasierter Befunde zu konstruieren. Der evidenzbasierte Befund der CDM-Analysen ist das Set der sieben diagnostizierbaren IP Skills, die in Form unterschiedlicher Skill-Bündel in die Konstruktion von Aufgaben einfließen (siehe Q-Matrix, Abb. 2). Aufgrund der Eindimensionalitätsannahme liefern IRT Analysen eine entsprechende Aufgliederung nicht. Wertvolle Informationen zur Aufgabenkonstruktion (A) liefern in CDMs auch die Assessment-Informationen aus den Skill-Klassen (FF 1.2). Aus der Verteilung der Skill-Klassen lässt sich ableiten, welche Aufgaben besonders benötigt werden. So zeigt sich, dass „ENT+VERK“ häufig zusammen mit „ANALY“ nicht beherrscht wird (Skill-Klasse 3). Laut Q-Matrix (Abb. 2) wären hierfür insbesondere Aufgaben wie

in Item 8 oder Item 19 formuliert, zu konstruieren. Eine vergleichbare Information aus derzeitigen IRT-Analysen könnten wiederum die Ergebnisse und Beschreibungen auf den Niveaustufen liefern. Dennoch liefern die Niveaustufen i. d. R. eine eher übergreifende Beschreibung der Kompetenz und können somit nur grobe Anregungen für die Aufgabenkonstruktion geben. Zudem sind die Niveaustufen aus IRT Modellen aufgrund der Eindimensionalität immer mit einem bestimmten Bereich von Aufgabenschwierigkeiten verbunden. Ein weiterer Unterschied zwischen den beiden Modelltypen ist, dass die Kompetenzniveaustufen in IRT-Modellen auf Basis der empirischen Parameterschätzungen identifiziert werden, während bei CDMs die Zuweisung von Skills zu Aufgaben zunächst rein theoriegeleitet vorgenommen wird und dann über die Q-Matrix in die Schätzung des empirischen Modells einfließt (DE LA TORRE/MINCHEN 2014). Zusammenfassend lässt sich festhalten, dass CDM-basierte Assessmentinformationen im Vergleich zu IRT-basierten Verfahren eine detailliertere Zielgruppenorientierung vermögen, die zudem direkt auf dem Kompetenzmodell basiert. Entsprechend sind seitens des Bildungspersonals mit CDM Informationen (Q-Matrix und Skill-Klassen) passgenauere Neukonstruktionen von Aufgaben auf Basis theoretisch fundierter Skills möglich.

Die Aggregation individueller Kompetenzprofile (gruppenbasierte Informationen) bieten darüber hinaus die Grundlage für didaktische Maßnahmen (B) und differenzierenden Unterricht (MEYER 2011a, 2011b). Naheliegender ist eine differenzierte Unterweisung in homogenen Gruppen mit explizit für ihre Schwächen erarbeiteten Aufgaben (z. B. Item 8 und 10 für die Personen mit dem Stärken-Schwächen-Profil der Skill-Klasse 3). Denkbar sind aber auch heterogene Gruppen oder die Jigsaw-Methode (ARONSON/PATNOE 2011), um kooperative Lernprozesse anzuregen. Hierfür würde sich z. B. anbieten eher kreative Lerner mit einem Stärken-Schwächen-Profil der Skill-Klasse 4 („WAHR“, „ANALY“ und „IDEE“) mit Lernenden der Skill-Klasse 10 („ANALY“, „SEQU“, „INFO+FACH“, „TOOL“ und „ENT+VERK“), also jenen, die eher in der Lage sind IP-Projekte umzusetzen, zu kombinieren. Ihre Stärken-Schwächen-Profile sind konträr ausgeprägt, sodass ein Lernen voneinander möglich wird. Auf Basis von IRT-basierten Informationen sind didaktische Maßnahmen für unterschiedliche Niveaustufen denkbar. Eine heterogene Gruppenbildung um voneinander zu lernen wäre motivational unausgeglichen, da die Hierarchieannahme impliziert, dass Lernende höherer Stufen nicht von Lernenden niedrigerer Stufen lernen können (da sie die niedrigeren Stufen ja bereits beherrschen sollten).

Aus motivationaler Perspektive sowie aus Sicht der Lerneffektivität sind Rückmeldungen in Form von individuellen Stärken-Schwächen-Profilen (FF 1.3) die Königsklasse der Lernbegleitung (C) seitens des Bildungspersonals (HATTIE/TIMPERLEY 2007). Ausgehend von dem unter FF 1.3 dargestellten beispielhaften Stärken-Schwächen-Profil kann der Lernbegleiter hier zurückmelden, dass der Lernende bereits die genau spezifizierbaren sechs der sieben IP-Skills beherrscht. Zudem kann er ihm zielgerichtet Lernaufgaben, die insbesondere den Skill „ENT-VERK“ erfordern (z. B. analog zu den Items 8, 10, 13, 15, 19 und/oder 20) anbieten, und ihm somit konkrete Anregungen zum Weiterlernen zu eröffnen. Eine personenbezogene IRT-basierte

Rückmeldung erfolgt in Form eines einzelnen Scores (Logitwert). Dieser Score kann einer Kompetenzstufe, die sich über Aufgabenmerkmale beschreiben lässt, zugeordnet werden. Auf Basis des einzelnen Scores innerhalb einer Kompetenzstufe lassen sich Stärken-Schwächen-Abwägungen vergleichend zu niedrigeren bzw. höheren Stufen anstellen. Sodass zum Weiterlernen Aufgaben der nächst höheren Kompetenzstufe angeboten werden müssten.

**FF 2.3 (Lernende):** Die wertvollste Information für einen Lerner ist, welche Teilaspekte (hier in Form von Skills) eines Lernziels er bereits erreicht hat (= Stärken) und bei welchen Facetten/Skills er vom angestrebten Lernziel abweicht (= Schwächen) (ebd.). Das positive Kompetenzerleben (DECI/RYAN 1993) über das bereits Erreichte mit Hilfe eines individuellen Stärken-Schwächen-Profiles (Kriteriumsnorm) wie es in FF 1.3 angeboten wird, ist demnach einer sozial-vergleichenden Rückmeldung, wie es personenbezogene IRT-Assessment-Informationen vermögen, deutlich vorzuziehen (Autonomie: ebd.). Der individuelle Personenscore aus IRT-Analysen wird dafür einer Kompetenzniveaustufe zugeordnet, deren aufgabenbasierte Beschreibung, vergleichend zu höheren bzw. niedrigeren Stufen erfolgen kann. Der formative Einsatz CDM-basierter Instrumente kann ergänzend die Abbildung von Lernentwicklungen (Bezugsnorm) unterstützen.

Des Weiteren kann der Lernende sein individuelles Kompetenzprofil auch sozial vergleichend nutzen (Sozialnorm), indem er dieses den durchschnittlichen Beherrschungswahrscheinlichkeiten einzelner Skills der Population (FF 1.1) oder den Prozentsätzen bedeutsamer Skill-Klassen (FF 1.2) gegenüberstellt. Aus motivationaler Sicht wird dies – insbesondere für die Lernschwächeren – weniger bedeutsam sein (HATTIE/TIMPERLEY 2007).

## 6. Zusammenfassung, Diskussion und Ausblick

Ziel des Beitrages war es, anhand eines CDM-Beispiels zur Entwicklung der Intrapreneurship-Kompetenz bei Industriekaufleuten am Ende ihrer Ausbildung Informationen zu generieren, die für die praktische Umsetzung professionellen Lehrerhandelns wie beispielsweise im Sinne des Lehrerbildungsstandards/„Kompetenz 8“ zur „individuellen Diagnostik“ notwendig sind. Dafür wurde (i) ein CDM gestütztes Assessment entwickelt, (ii) die CDM Assessment-Informationen für IP bei Auszubildenden generiert und (iii) deren Potentiale für unterschiedliche Interessengruppen analysiert.

Für die (i) Assessment-Entwicklung wurde das ECD von MISLEVY und RISCO-SCENTE (2006) als Konstruktionsrahmen verwendet: Die Darstellungen bieten einen Schritt-für-Schritt-Rahmen der es ermöglicht, ein kohärentes und damit valides Assessment für ein evidenz-basiertes Schlussfolgern unter Verknüpfung der drei Teile Cognition, Observation, Interpretation zu konstruieren (vgl. GEORGE/BLEY/PELLEGRINO in Begutachtung). Die Evidenzen für die Validität lassen sich wie folgt klassifizieren: Curricular-inhaltliche Validität bietet die elaborierte Domänenanalyse von Weber und Kollegen (WEBER u. a. 2014). Auf Basis der aufwendigen Herleitung

und Prüfung der diagnostizierbaren Skills, der korrespondierenden Aufgaben sowie der Skill-Aufgaben-Zuschreibungen (Q-Matrix) liegen Evidenzen für die kognitive Validität vor (BLEY in Begutachtung). Die strukturelle Validität ist über die Maße zur Modellgüte abgesichert (GEORGE/BLEY/PELLEGRINO in Begutachtung). Zur prädiktiven/kriterialen Validität konnten erste Plausibilitätsprüfungen (u. a. mit Ergebnissen aus IHK-Projektberichten, WEBER u. a. 2015) für die durchschnittlichen Beherrschungs-Wahrscheinlichkeiten der Skills (FF 1.1) und die Skill-Klassen (FF 1.2) angestellt werden. Über Replikationsstudien sowie längsschnittlich angelegte Interventionsstudien könnten diese Befunde weiter abgesichert werden. Wir gehen aufgrund dieser Summe an Evidenzen davon aus, dass es gelungen ist, ein Instrument zur Generierung diagnostischer Assessment-Informationen für ein komplexes Konstrukt aus der Berufs- und Wirtschaftspädagogik (IP) zu entwickeln.

Die CDM-Ergebnisse (ii) für IP bei Auszubildenden zum Industriekaufmann/ zur Industriekauffrau legen nahe, dass die sieben spezifizierten Skills gut bis sehr gut beherrscht werden (FF 1.1). Ca. 50 % der Auszubildenden verfügen am Ende ihrer Erstausbildung über ein vollständiges IP Kompetenzprofil bestehend aus den sieben Skills (FF 1.2). Weitere neun Skill-Klassen ergeben sich aufgrund ihrer Beherrschungs-Wahrscheinlichkeiten als besonders bedeutsam für die Domäne. Individuelle Kompetenzprofile (FF 1.3) bieten personenbezogene Stärken-Schwächen Informationen.

Die mittels CDMs gewonnenen diagnostischen Assessmentinformationen wurden (iii) im Hinblick auf ihr Potential für die Gestaltung individualisierter Lehr- und Lernprozesse – wie dieses in den Lehrerbildungsstandards „individuelle Diagnostik“ gefordert wird – analysiert. Dabei wurden die diagnostischen Assessmentinformationen auf der Populations-, der Gruppen- und der Individualebene im Hinblick auf ihre Nützlichkeit und ihren Unterschied im Vergleich zu derzeitigen IRT-Ergebnissen für die drei Stakeholder Programmorganisatoren (FF 2.1), Bildungspersonal (FF 2.2) und Lernende/Azubis (FF 2.3) diskutiert. So ließen sich die mittels des CDMs gewonnenen diagnostischen Assessmentinformationen nutzen, um (A) Aufgabenstellungen lernzielgerecht zu konstruieren und diese adressatengerecht formulieren zu können. Zum Beispiel sind angepasst an Personen in Skill-Klasse 2 (FF 1.2) Aufgaben zur Förderung der eher schwach beherrschten Skills „IDEE“ und „ENT+VERK“ (FF 1.1) in Analogie zu den Aufgaben aus der Q-Matrix, die explizit diese Skills erfordern, zu konstruieren. Aus Leistungsüberprüfungen (B) können Informationen identifiziert werden, die als konstruktive Rückmeldung über die weitere eigene Unterrichtstätigkeit dienen: z. B. zur Anwendung der Jigsaw-Methode für Lernende mit den konträren Stärken-Schwächen-Profilen der Skill-Klassen 4 und 10 (FF 1.2). Die Möglichkeit individueller Kompetenzprofile, lässt informative Feedbacks sowie Perspektiven zum Weiterlernen zu: z. B. von dem angestrebten Lernziel – IP – werden bereits 6 Skills beherrscht; der Skill „ENT+VERK“ muss noch weiter geübt werden (= spezifisches Kompetenzprofil für ein ausgewähltes Antwortmuster; FF1.3). Diagnostische Assessmentinformationen lassen auch Rückschlüsse auf makrodidaktische Entscheidungen

zu, wie z. B. das Schaffen von weiteren Lerngelegenheiten für weniger ausgeprägte IP-Kompetenzfacetten (insbesondere „ENT+VERK“, FF 1.1).

Als Unterschied von CDMs gegenüber derzeit üblicherweise in der beruflichen Kompetenzmessung genutzten (eindimensionalen) IRT-basierten Assessment-Informationen kristallisierten sich insbesondere drei Aspekte heraus: (a) die Mehrdimensionalität (Detailgrad der erzielbaren Informationen), (b) die theoriegeleitete (= in Kohärenz zum Kompetenzmodell Erzeugung von Stärken-Schwächen-Profilen) bzw. zur Erstellung zielgruppenadäquater neuer Aufgaben und (c) dass den Stärken-Schwächen-Profilen im Vergleich zu aus IRT Modellen abgeleiteten Niveaustufen keine Hierarchieannahme unterliegt.

Die skizzierten Ausführungen zu Potentialen von CDMs insbesondere für die Begleitung individueller Lehr-/Lernprozesse und deren Unterschied gegenüber derzeitigen IRT-Modelle beziehen sich auf zentrale Aspekte beider Modelle. Es gibt sowohl auf Seiten der IRT als auch in der Familie der CDMs spezifische Modelle (z. B. IRT: das Raschmodell; CDM das Fusion-Modell), die bestimmte Eigenschaften enthalten (z. B. IRT: spezifische Objektivität, CDMs: Kompensierbarkeit zwischen den Skills). Diese können jedoch eher als Nuancen der Modelle betrachtet werden, während die grundlegenden Merkmale hier dargestellt wurden.

Die aufgezeigten Potentiale von CDMs liefern Anregungen diagnostische Kompetenzen angehender Lehrkräfte/Ausbilder im beruflichen Bereich unter Einsatz der vorliegenden Informationen auszubilden und zu evaluieren. Für die makrodidaktische Praxis werden Anregungen für inhaltliches Fokussieren und für Weiterentwicklungen von Ordnungsmitteln eröffnet. Lehrkräfte können auf mehrdimensionale evidenzbasierte Befunde zur Förderung und Ausbildung von Intrapreneurship-Kompetenz bei kaufmännischen Auszubildenden zurückgreifen. Es werden Anregungen für die Nutzung dieser Informationen gegeben.

Zusammenfassend konnten die Möglichkeiten und Potentiale kognitiver Diagnosemodelle aufgezeigt werden, die es insbesondere mit Blick auf FF 2.2 und 2.3 in einem nächsten Schritt in der schulpraktischen Anwendung zu evaluieren gilt. Damit einher gehen Konkretisierungen der Einsatzhäufigkeit, der fachmethodischen Zusammenhänge sowie der Einbindung der beiden Lernorte (Schule und Betrieb). Da projektbezogenes innovativ unternehmerisches Denken und Handeln – wie hier als IP-Kompetenz zusammengefasst – eine zukunftsorientierte kaufmännische Kompetenz ist, wäre zudem der Einsatz des Instrumentes in weiteren Berufen außer dem Industriekaufmann sehr gut denkbar (z. B. Medienkaufmann, Informatikkaufmann, etc.), aber noch zu prüfen. Gleichwohl der Entwicklungsprozess für CDMs zwar aufwendig ist und eine starke Zusammenarbeit zwischen verschiedenen Disziplinen erfordert, scheint sich das Engagement für zentrale und innovative „Kernkompetenzen“ (wie hier IP) und große Populationen (wie hier Industriekaufleute) dennoch zu lohnen.

## Literatur

- ACHTENHAGEN, F. / WINTHER, E. (2009): Konstruktvalidität von Simulationsaufgaben: Computergestützte Messung berufsfachlicher Kompetenz – am Beispiel der Ausbildung von Industriekaufleuten. Bericht an das Bundesministerium für Bildung und Forschung (350600). [https://www.bmbf.de/pubRD/Endbericht\\_BMBF09.pdf](https://www.bmbf.de/pubRD/Endbericht_BMBF09.pdf). 02.03.2015.
- ACHTENHAGEN, F. / WINTHER, E. (2014): Workplace-based competence measurement: developing innovative assessment systems for tomorrow's VET programmes. In: *Journal of Vocational Education & Training* 66, H. 3, S. 281–295.
- AMERICAN EDUCATIONAL RESEARCH ASSOCIATION/AMERICAN PSYCHOLOGICAL ASSOCIATION/NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION (2014): Standards for educational and psychological testing. Washington, DC: American Educational Research Association.
- ARONSON, E. / PATNOE, S. (2011): Cooperation in the classroom. The jigsaw method. London: Printer & Martin.
- ARTELT, C. / GRÄSEL, C. (2009): Diagnostische Kompetenz von Lehrkräften. In: *Zeitschrift für Pädagogische Psychologie* 23, H. 4, S. 157–160.
- BAETHGE, M. / SEEBER, S. (2016): Die gemeinsame theoretische und methodische Basis der ASCOT-Projekte. In: BECK, K. / LANDENBERGER, M. / OSER, F. (Hrsg.): Technologiebasierte Kompetenzmessung in der beruflichen Bildung – Resultate aus dem Forschungsprogramm ASCOT. Bielefeld: Bertelsmann, S. 15–32.
- BAUMERT, J. / KUNTER, M. (2006): Stichwort: Professionelle Kompetenz von Lehrkräften. In: *Zeitschrift für Erziehungswissenschaft* 9, H. 4, S. 469–520.
- BECK, K. / LANDENBERGER, M. / OSER, F. (Hrsg.) (2016): Technologiebasierte Kompetenzmessung in der beruflichen Bildung – Resultate aus dem Forschungsprogramm ASCOT. Bielefeld: Bertelsmann.
- BILLET, S. (2002): Toward a workplace pedagogy: guidance, participation, and engagement. In: *Adult Education Quarterly* 53, H. 1, S. 27–43.
- BLEY, S. (in Begutachtung): Developing and validating a technology-based diagnostic assessment using the evidence-centred game design-approach. an example of intrapreneurship competence.
- BLEY, S. / WIETHE-KÖRPRICH, M. / WEBER, S. (2015): Formen kognitiver Belastung bei der Bewältigung technologiebasierter authentischer Testaufgaben. Eine Validierungsstudie zur Abbildung von beruflicher-Kompetenz. In: *Zeitschrift für Berufs- und Wirtschaftspädagogik* 111, H. 2, S. 268–294.
- BÖGEHOLZ, S. / EGGERT, S. (2013): Welche Rolle spielt Kompetenzdiagnostik im Rahmen von Lehr-Lernprozessen? In: *Zeitschrift für Erziehungswissenschaft* 16, H. 1, S. 59–64.
- BUNDESINSTITUT FÜR BERUFSBILDUNG (2012): Datenreport 2012. <https://datenreport.bibb.de/html/dr2012.html>. 02.09.2016.
- DARLING-HAMMOND, L. / HERMAN, J. / PELLEGRINO, J. W. / ABEDI, J. / ABER, J. L. / BAKER, E. / BENNET, R. / GORDON, E. / HAERTEL, E. H. / HAKUTA, K. / HO, A. / LINN, R. L. / PEARSON, P. D. / POPHAM, J. / RESNICK, L. B. / SCHOENFELD, A. H. / SHAVELSON, R. J. / SHEPARD, L. A. / SCHULMAN, L. / STEELE, C. M. (2013): Criteria for high-quality assessment. Stanford, CA: Stanford Center for Opportunity Policy in Education.
- DE LA TORRE, J. / KARELITZ, T. M. (2009): Impact of Diagnosticity on the Adequacy of Models for Cognitive Diagnosis under a Linear Attribute Structure. A Simulation Study. In: *Journal of Educational Measurement* 46, H. 4, S. 450–469.

- DE LA TORRE, J. / MINCHEN, N. (2014): Cognitively diagnostic assessments and the cognitive diagnosis model framework. In: *Psicología Educativa* 20, H. 2, S. 89–97.
- DECI, E. L. / RYAN, R. M. (1993): Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. In: *Zeitschrift für Pädagogik* 39, H. 2, S. 223–238.
- DIBELLO, L. V. / ROUSSOS, L. A. / STOUT, W. F. (2007): Review of cognitively diagnostic assessment and a summary of psychometric models. In: RAO, C. R. / SINHARAY, S. (Hrsg.): *Psychometrics*. Amsterdam, Boston: Elsevier, S. 979–1030.
- EMBRETSON, S. E. / REISE, S. (2000): *Item response theory for psychologists*. Mahwah, N.J.: Erlbaum.
- GEORGE, A. C. / BLEY, S. / PELLEGRINO, J. W. (in Begutachtung): On combining instructionally relevant results from diagnostic assessments and cognitive diagnosis models. An example of intrapreneurship.
- GEORGE, A. C. / ROBITZSCH, A. (2015): Cognitive diagnosis models in R: A didactic. In: *The Quantitative Methods for Psychology* 11, H. 3, S. 189–205.
- GEORGE, A. C. / ROBITZSCH, A. / KIEFER, T. / GROSS, J. / ÜNLÜ, A. (2016): The R Package CDM for cognitive diagnosis modeling. In: *Journal of Statistical Software* 74, H. 2, S. 1–24.
- GIERL, M. J. / CUI, Y. (2008): Defining characteristics of diagnostic classification models and the problem of retrofitting in cognitive diagnostic assessment. In: *Measurement: Interdisciplinary Research & Perspectives* 6, H. 4, S. 263–268.
- HATTIE, J. / TIMPERLEY, H. (2007): The power of feedback. *Review of Educational Research*. In: *Review of Educational Research* 77, S. 81–112.
- HELM, C. / BLEY, S. / GEORGE, A. C. / POCRNJA, M. (2015): Potentiale kognitiver Diagnosemodelle für den berufsbildenden Unterricht. In: STOCK, M. / SCHLÖGL, P. / SCHMID, K. / MOSER, D. (Hrsg.): *Kompetent – wofür? Life-Skills – Beruflichkeit – Persönlichkeitsbildung*. Innsbruck: StudienVerlag, S. 206–224.
- HELMKE, A. / HOSENFELD, I. / SCHRADER, F.-W. (2004): Vergleichsarbeiten als Instrument zur Verbesserung der Diagnosekompetenz von Lehrkräften. In: ARNOLD, R. / GRIESE, C. (Hrsg.): *Schulleitung und Schulentwicklung*. Baltmannsweiler: Schneider Verl. Hohengehren, S. 119–144.
- INTERSTATE TEACHER ASSESSMENT AND SUPPORT CONSORTIUM (2011): *Model core teaching standards. A resource for state dialogue*. Washington, DC: Council of Chief State School Officers.
- JANESICK, V. J. (2006): *Authentic assessment*. New York: Lang Primer.
- KLOTZ, V. (2015): *Diagnostik beruflicher Kompetenzentwicklung. Eine wirtschaftsdidaktische Modellierung für die kaufmännische Domäne*. Berlin, Heidelberg: Springer.
- KMK (2004): *Standards für die Lehrerbildung. Bildungswissenschaften (Beschluss der Kultusministerkonferenz vom 16.12.2004)*. [http://www.kmk.org/fileadmin/veroeffentlichungen\\_beschluesse/2004/2004\\_12\\_16-Standards-Lehrerbildung.pdf](http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf). Accessed 21 Aug 2014. 20.09.2016.
- KOHLER, B. / SCHRADER, F.-W. (2004): Ergebnisrückmeldung und Rezeption. Von der externen Evaluation zur Entwicklung von Schule und Unterricht. In: *Themenheft der Empirischen Pädagogik* 18, H. 1.
- KORUNKA, C. / FRANK, H. / LUEGER, M. / EBNER, M. L. (2009): Entwicklung und Prüfung eines Modells zur Förderung von Intrapreneurship in der dualen Berufsausbildung. In: *Zeitschrift für Personalpsychologie* 8, H. 3, S. 129–146.
- KUNINA-HABENICHT, O. / RUPP, A. A. / WILHELM, O. (2009): A practical illustration of multidimensional diagnostic skills profiling: Comparing results from confirmatory factor analysis and diagnostic classification models. In: *Studies in Educational Evaluation* 35, 2–3, S. 64–70.

- KUNINA-HABENICHT, O. / RUPP, A. A. / WILHELM, O. (2012): The impact of model misspecification on parameter estimation and item-fit assessment in log-linear diagnostic classification models. In: *Journal of Educational Measurement* 49, H. 1, S. 59–81.
- LEHMANN, R. H. / SEEBER, S. (Hrsg.) (2007): *ULME III. Untersuchung von Leistungen, Motivation und Einstellungen der Schülerinnen und Schüler in den Abschlussklassen der Berufsschulen*. Hamburg: HIBB.
- LEIGHTON, J. P. / GIERL, M. J. (Hrsg.) (2007a): *Cognitive diagnostic assessment for education. Theory and applications*. Cambridge, New York: Cambridge University Press.
- LEIGHTON, J. P. / GIERL, M. J. (2007a): Why cognitive diagnostic assessment? In: LEIGHTON, J. P. / GIERL, M. J. (Hrsg.): *Cognitive diagnostic assessment for education. Theory and applications*. Cambridge, New York: Cambridge University Press, S. 3–18.
- LEIGHTON, J. P. / GIERL, M. J. / HUNKA, S. M. (2004): The attribute hierarchy model for cognitive assessment: A variation on Tatsuoka's rule-space approach. In: *Journal of Educational Measurement*, S. 205–237.
- LEISS, D. / RAKOCZY, K. (2013): Wie können Ergebnisse der Kompetenzdiagnostik in Forschungsprojekten sinnvoll zurückgemeldet werden? In: *Zeitschrift für Erziehungswissenschaft* 16, H. 1, S. 81–88.
- LEUDERS, T. (2014): Modellierungen mathematischer Kompetenzen – Kriterien für eine Validitätsprüfung aus fachdidaktischer Sicht. In: *Journal für Mathematikdidaktik* 35, H. 1, S. 7–48.
- MAYDEY-OLIVARES, A. (2013): Goodness-of-fit assessment of item response theory models. In: *Measurement: Interdisciplinary Research and Perspectives* 11, S. 71–137.
- MESSICK, S. (1989): Validity. In: LINN, R. L. (Hrsg.): *Educational measurement*. New York: American Council on Education, S. 13–103.
- MEYER, H. (2011a): *Unterrichtsmethoden*. Berlin: Cornelsen Scriptor.
- MEYER, H. (2011b): *Was ist guter Unterricht? Mit 65 Min.-Vortragsvideo (DVD)*. Berlin: Cornelsen Scriptor.
- MISLEVY, R. J. / RISCOSENTE, M. M. (2006): Evidenced-centered assessment design: Layers, concepts and terminology. In: DOWNING, S. M. / HALADYNA, T. M. (Hrsg.): *Handbook of test development*. Mahwah, NJ: Erlbaum, S. 61–90.
- NEUMANN, K. (2013): Mit welchem Auflösungsgrad können Kompetenzen modelliert werden? In welcher Beziehung stehen Modelle zueinander, die Kompetenz in einer Domäne mit unterschiedlichem Auflösungsgrad beschreiben? In: *Zeitschrift für Erziehungswissenschaft* 16, H. 1, S. 35–39.
- NICKOLAUS, R. / SEEBER, S. (2013): Berufliche Kompetenzen. Modellierungen und diagnostische Verfahren. In: FREY, A. / LISSMANN, U. / SCHWARZ, B. (Hrsg.): *Handbuch Berufspädagogische Diagnostik*. Weinheim: Beltz, S. 166–195.
- OSER, F. K. (1997): Standards in der Lehrerbildung. Teil 1 Berufliche Kompetenzen, die hohen Qualitätsmerkmalen entsprechen. In: *Beiträge zur Lehrerbildung* 15, H. 1, S. 26–37.
- OSER, F. K. / SALZMANN, P. / HEINZER, S. (2009): Measuring the competence-quality of vocational teachers: An advocacy approach. In: *Empirical Research in Vocational Education and Training* 1, S. 65–83.
- PANT, H. A. (2013): Wer hat einen Nutzen von Kompetenzmodellen? In: *Zeitschrift für Erziehungswissenschaft* 16, H. 1, S. 71–79.
- PELLEGRINO, J. W. / CHUDOWSKY, N. / GLASER, R. (2001): *Knowing what students know: The science and design of educational assessment*. National Academy Press.
- PELLEGRINO, J. W. / DIBELLO, L. V. / GOLDMAN, S. R. (2016): A framework for conceptualizing and evaluating the validity of instructionally relevant assessments. In: *Educational Psychologist* 51, H. 1, S. 59–81.

- R CORE TEAM (2016): R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
- RUPP, A. A. / TEMPLIN, J. L. / HENSON, R. A. (2010): Diagnostic measurement. Theory, methods, and applications. New York: Guilford Press.
- RUPPRECHT, M. / MULDER, R. H. / GRUBER, H. / NEUMANN, W. (2011): Fostering innovative behaviour and dealing with diversity within consulting teams. In: IPOB (Hrsg.): The future of knowledge-intensive service work. Theory and practice of managing human and organisational resources; based upon the symposium (June 17–18, 2010 at the Carl von Ossietzky University Oldenburg, Germany). Marburg: Metropolis, S. 153–182.
- SCHRADER, F.-W. (2009): Anmerkungen zum Themenschwerpunkt Diagnostische Kompetenz von Lehrkräften. In: Zeitschrift für Pädagogische Psychologie 23, 3–4, S. 237–245.
- SEEBER, S. (2005): Input-Controlling in der beruflichen Bildung. Aspekte der Bestimmung von Lernausgangslagen. In: Unterrichtswissenschaft 33, H. 4, S. 314–333.
- SHAVELSON, R. J. (2012): Assessing business planning competence using the collegiate learning assessment. In: Empirical Research in Vocational Education and Training 4, H. 1, S. 77–90.
- STROBL, C. (2012): Das Rasch-Modell. Eine verständliche Einführung für Studium und Praxis. München: Hampp.
- TATSUOKA, K. K. (1983): Rule space: An approach for dealing with misconceptions based on item response theory. In: Journal of Educational Measurement 20, H. 4, S. 345–354.
- TYNJÄLÄ, P. (2013): Toward a 3-P Model of Workplace Learning: a Literature Review. In: Vocations and Learning 6, H. 1, S. 11–36.
- WEBER, S. / BECK, K. / KÖLLER, O. (2016): Potentiale in der beruflichen Bildung sichtbar machen. In: Unterrichtswissenschaft 44, H. 2, S. 106–113.
- WEBER, S. / DRAXLER, C. / BLEY, S. / WIETHE-KÖRPRICH, M. / WEISS, C. / GÜRER, C. (2016): Der Projektverbund CoBALIT – Intrapreneurship: Large Scale-Assessments in der kaufmännischen Berufsausbildung. In: BECK, K. / LANDENBERGER, M. / OSER, F. (Hrsg.): Technologiebasierte Kompetenzmessung in der beruflichen Bildung – Resultate aus dem Forschungsprogramm ASCOT. Bielefeld: Bertelsmann, S. 75–92.
- WEBER, S. / TROST, S. / WIETHE-KÖRPRICH, M. / WEISS, C. / ACHTENHAGEN, F. (2014): Intrapreneur: An entrepreneur within a company – an approach on modeling and measuring intrapreneurship competence. In: WEBER, S. / OSER, F. K. / ACHTENHAGEN, F. / FRETSCHENER, M. / TROST, S. (Hrsg.): Becoming an entrepreneur. Rotterdam, Boston, Taipei: Sense Publishers, S. 256–287.
- WEBER, S. / WIETHE-KÖRPRICH, M. / BLEY, S. / WEISS, C. / ACHTENHAGEN, F. (2015): Intrapreneurship-Verhalten an kaufmännischen Arbeitsplätzen. Analysen von Projektberichten. In: Empirische Pädagogik Sonderheft: Ökonomische Kompetenzen in Schule, Ausbildung und Hochschule 28, H. 1, S. 84–105.
- WEBER, S. / WIETHE-KÖRPRICH, M. / BLEY, S. / WEISS, C. / DRAXLER, C. / GÜRER, C. (2016b): Modellierung und Validierung eines Intrapreneurship-Kompetenz-Modells bei Industriekaufleuten. In: Unterrichtswissenschaft 44, H. 2, S. 149–168.
- WILD, K.-P. / KRAPP, A. (2006): Pädagogisch-psychologische Diagnostik. In: KRAPP, A. / WEIDENMANN, B. (Hrsg.): Pädagogische Psychologie. Weinheim, Basel: Beltz, S. 525–574.
- WINTHER, E. (2010): Kompetenzmessung in der beruflichen Bildung. Bielefeld: Bertelsmann.
- WINTHER, E. / ACHTENHAGEN, F. (2009): Skalen und Stufen kaufmännischer Kompetenz. In: Zeitschrift für Berufs- und Wirtschaftspädagogik 105, H. 4, S. 521–556.



- WINTHER, E. / SEEBER, S. / FESTNER, D. / SANGMEISTER, J. / LIEDKE, M. (2016): Large Scale-Assessments in der kaufmännischen Berufsbildung (CoBALIT). In: BECK, K. / LANDEBERGER, M. / OSER, F. (Hrsg.): Technologiebasierte Kompetenzmessung in der beruflichen Bildung – Resultate aus dem Forschungsprogramm ASCOT. Bielefeld: Bertelsmann, S. 55–73.
- WUNDERER, R. (2003): Führung und Zusammenarbeit. Eine unternehmerische Führungslehre. München: Luchterhand.

DR. SANDRA BLEY, MBR

Ludwig-Maximilians-Universität München, Munich School of Management | Fakultät für Betriebswirtschaft, Institut für Wirtschaftspädagogik, Ludwigstraße 28/RG, D-80539 München, Mail: [bley@bwl.lmu.de](mailto:bley@bwl.lmu.de)

DR. ANN CATHRICE GEORGE

Bundesinstitut für Bildungsforschung, Innovation & Entwicklung des österreichischen Schulwesens (BIFIE) Alpenstrasse 121, A-5020 Salzburg, Mail: [a.george@bifie.at](mailto:a.george@bifie.at)

