

Goldhammer, Frank; Kröhne, Ulf; Keßel, Yvonne; Senkbeil, Martin; Ihme, Jan Marten
Diagnostik von ICT-Literacy. Multiple-Choice- vs. simulationsbasierte Aufgaben

formal und inhaltlich überarbeitete Version der Originalveröffentlichung in:

formally and content revised edition of the original source in:

Diagnostica 60 (2014) 1, S. 10-21



Bitte verwenden Sie beim Zitieren folgende URN /
Please use the following URN for citation:
urn:nbn:de:0111-pedocs-146050

Nutzungsbedingungen

Dieses Dokument steht unter folgender Creative Commons-Lizenz:
<http://creativecommons.org/licenses/by-nc/4.0/deed.de> - Sie dürfen das Werk bzw. den Inhalt vervielfältigen, verbreiten und öffentlich zugänglich machen sowie Abwandlungen und Bearbeitungen des Werkes bzw. Inhaltes anfertigen, solange Sie den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen und das Werk bzw. den Inhalt nicht für kommerzielle Zwecke verwenden.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

This document is published under following Creative Commons-License:
<http://creativecommons.org/licenses/by-nc/4.0/deed.en> - You may copy, distribute and render this document accessible, make adaptations of this work or its contents accessible to the public as long as you attribute the work in the manner specified by the author or licensor. You are not allowed to make commercial use of the work, provided that the work or its contents are not used for commercial purposes.

By using this particular document, you accept the above-stated conditions of use.



Kontakt / Contact:

peDOCS
Deutsches Institut für Internationale Pädagogische Forschung (DIPF)
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft

Akzeptierte Manuskriptfassung (nach peer review) des folgenden Artikels:

[Goldhammer, F., Kröhne, U., Keßel, Y., Senkbeil, M. & Ihme, J. M.: Diagnostik von ICT-Literacy: Multiple-Choice- vs. simulationsbasierte Aufgaben. Diagnostica, 60 \(1\). doi: 10.1026/0012-1924/a000113](#)

© Hogrefe Verlag, Göttingen 2014

Diese Artikelfassung entspricht nicht vollständig dem in der Zeitschrift veröffentlichten Artikel. Dies ist nicht die Originalversion des Artikels und kann daher nicht zur Zitierung herangezogen werden.

Die akzeptierte Manuskriptfassung unterliegt der Creative Commons License CC-BY-NC.

Zusammenfassung. ICT-Literacy legt eine performanzbasierte Erfassung nahe, also mithilfe von Testaufgaben, die interaktive (simulierte) Computerumgebungen präsentieren und eine Reaktion mittels Maus und/oder Tastatur erfordern. Dennoch kommen häufig Verfahren wie Selbstbeurteilungen oder papierbasierte Leistungstests zum Einsatz. Ziel der vorliegenden Studie war es, die psychometrischen Eigenschaften simulationsbasierter (SIM) Aufgaben mit den Eigenschaften inhaltlich paralleler Multiple-Choice (MC)-Aufgaben zu vergleichen, bei denen Screenshots als Stimulus verwendet werden. Die MC-Aufgaben, die im Rahmen der National Educational Panel Study (NEPS) entwickelt wurden, erfassen die Fähigkeit, digitale Informationen auszuwählen und abzurufen sowie grundlegende Operationen durchzuführen (Access). In einem Zufallsgruppendesign bearbeiteten 405 Jugendliche der Klassenstufe 9 die computerbasierten Access-Testitems entweder als MC-Aufgabe oder als SIM-Aufgabe sowie den simulationsbasierten Basic Computer Skills (BCS)-Test. Die Ergebnisse zeigen, dass sich die meisten MC-Aufgaben und SIM-Aufgaben hinsichtlich Schwierigkeit und Ladung unterscheiden. Übereinstimmende konvergente Validität wird durch vergleichbar hohe Korrelationen der beiden Testformen mit BCS angezeigt.

Schlüsselwörter: ICT-Literacy, Multiple-Choice-Aufgaben, simulationsbasierte Aufgaben, Messmodellinvarianz, konvergente Validität

Assessment of ICT literacy: Multiple-choice vs. simulation-based tasks

Abstract. ICT literacy suggests a performance-based assessment by means of tasks presenting interactive (simulated) computer environments and requiring responses by means of mouse and/or keyboard. However, commonly assessment procedures like self-ratings or paper-based performance measures are used. The goal of the present study was compare the psychometric properties of simulation-based (SIM) tasks with parallel multiple choice (MC) tasks presenting screenshots of software applications. The MC tasks developed for the National Educational Panel Study (NEPS) reflect the skill to select and retrieve digital information and to perform basic operations (access). In a random groups design, 405 grade 9 students completed the computer-based access items as MC task or as SIM tasks as well as the simulation-based Basic Computer Skills (BCS) test. Results show that the majority of MC tasks and SIM tasks differ in difficulty and loading. Consistent convergent validity is indicated by comparably high correlations of the two test forms with BCS.

Key words: ICT literacy, multiple-choice tasks, simulation-based tasks, measurement invariance, convergent validity

Diagnostik von ICT-Literacy: Multiple-Choice- vs. simulationsbasierte Aufgaben

Informations- und Kommunikationstechnologie (Information and Communication Technology, ICT) ist zu einem integralen Bestandteil unseres heutigen Alltags geworden. Computerbezogene Fertigkeiten stellen eine wichtige Kompetenz dar, um alltägliche informationsbezogene Anforderungen und Problemstellungen erfolgreich zu bewältigen. Sowohl im Bildungs- als auch im beruflichen Kontext ist der kompetente Umgang mit Computer und Internet (ICT-Literacy) nahezu unverzichtbar geworden (vgl. Lennon, Kirsch, Von Davier, Wagner & Yamamoto, 2003). Dementsprechend sind in manchen Ländern Unterrichtseinheiten zu ICT-Literacy zu einem Bestandteil des Lehrplans an Schulen geworden, die systematisch geplant wie auch empirisch evaluiert werden (z. B. Kim, Jung & Lee, 2008; s. auch Poynton, 2005) oder aber der Umgang mit dem Computer als Informationslieferant und Kommunikationsmedium wird als transversale Fertigkeit in traditionelle Unterrichtsfächer integriert (vgl. KMK, 2012). Die reliable und valide Diagnose von interindividuellen Unterschieden in ICT-Literacy ist somit für verschiedene angewandte pädagogische Forschungsfelder eine wichtige Anforderung.

Diagnostik von ICT-Literacy

Es existieren verschiedene Ansätze zur Messung von ICT-Literacy. Häufig werden Selbstbeurteilungen dieser Fertigkeit verwendet (z. B. in der ICT Self-efficacy Scale, Markauskaite, 2007, oder im Rahmen von PISA 2009 im ICT Familiarity Questionnaire, OECD, 2011). Obwohl Selbstbeurteilungen einen gewissen Einblick in tatsächliche Fertigkeiten ermöglichen (z. B. Ballantine, Larres & Oyerle, 2007), unterliegen sie jedoch systematischen Verzerrungen und sollten deshalb als Messungen von Selbstwirksamkeit betrachtet werden (vgl. Kuhlemeier & Hemker, 2007). Ein anderer Ansatz in der Messung von ICT-Literacy besteht in der Verwendung von Paper-Pencil-Tests mit Multiple-Choice (MC)-Aufgaben. Kuhlemeier und Hemker (2007) beispielsweise verwenden in ihrem Test „Internet skills for school“ Screenshots als Stimulusmaterial. Anhand dieser werden Aufgabenstellungen präsentiert, und die Testperson

hat anzugeben, was sie in dieser Situation am besten tun sollte. Ein ähnlicher Ansatz wird von Potosky (2007) zur Messung von Internetwissen (Test iKnow) verfolgt. Im INCOBI-R (Inventar zur Computerbildung – Revision) setzen Richter, Naumann und Horz (2010) kurze Beschreibungen von alltäglichen Problemsituationen ein. Aus vier Möglichkeiten ist jeweils die zielführende Handlungsalternative auszuwählen.

Obwohl Paper-Pencil-Tests von Computerwissen objektiver sind und im Vergleich zu Selbstbeurteilungen das tatsächliche Leistungsverhalten besser vorhersagen (Goldhammer, Naumann & Keßel, 2013), unterliegen sie dennoch gewissen Einschränkungen. Anhand von Paper-Pencil-Tests können nur deklarative und prozedurale Wissensbestände erfasst werden, die ein erfolgreiches Handeln mit Computer und Internet wahrscheinlich machen, nicht aber das computerbezogene Handeln selbst (z. B. Formatieren und Abspeichern eines Textes). Beispielsweise bleiben technische Basisfertigkeiten (z. B. Manipulieren von Menüs mittels Tastatur und Maus; vgl. Goldhammer et al., 2013) unberücksichtigt. Diese werden bei Verwendung von Paper-Pencil-Tests zwar vorausgesetzt, sind aber nicht Bestandteil der Aufgabenbearbeitung. Obwohl vor diesem Hintergrund die Entwicklung verhaltensbasierter Leistungsmaße zur Erfassung von ICT-Literacy naheliegend ist, sind derzeit nur wenige performanzbasierte Messungen verfügbar, wie z. B. der „Information and Communication Technology Literacy Test“ (iSkills) des ETS (2008; siehe auch Katz & Macklin, 2007), oder der „Basic Computer Skills (BCS) Test“ von Goldhammer et al. (2013).

Das zentrale Ziel dieses Beitrags besteht darin, unterschiedliche Operationalisierungsansätze zur Erfassung von ICT-Literacy hinsichtlich ihrer psychometrischen Eigenschaften miteinander zu vergleichen: Zum einen werden traditionelle MC-Aufgaben mit Screenshots von Softwareumgebungen als Stimulusmaterial herangezogen. Zum anderen werden inhaltsgleiche Aufgaben mit simulierten (SIM) Softwareumgebungen verwendet. SIM-Aufgaben weisen eine deutlich höhere Augenscheinvalidität als MC-Aufgaben auf. In der vorliegenden Studie sollen zudem vermutete Unterschiede in Item- und

Skalencharakteristiken sowie in der konvergenten Validität der MC- und SIM-Skala empirisch überprüft werden.

ICT-Literacy-Komponente „Access“

Für den psychometrischen Vergleich von MC- und SIM-Aufgaben wurde als grundlegende ICT-Literacy-Komponente „Access“ ausgewählt, die als Voraussetzung für weitergehende Informationsverarbeitung im ICT-Kontext verstanden werden kann. Unser Vorgehen lässt jedoch analog auch auf andere ICT-Literacy-Komponenten übertragen. Ferrari (2012) ordnet „Access“ in ihrer Übersichtsarbeit zu Frameworks Digitaler Kompetenzen dem Kompetenzbereich Informationsmanagement zu, der die Fertigkeiten umfasst, Informationen zu identifizieren, zu lokalisieren, auf diese zuzugreifen, sie abzurufen, zu speichern und zu organisieren. Die vorliegende Studie übernimmt die Access-Definition sowie daraus abgeleitete Testitems, die für das National Educational Panel Study (NEPS) formuliert wurde. Im NEPS stellt „Access“ eine von mehreren Prozesskomponenten dar und wird definiert als „knowledge of basic operations used to retrieve information (e. g., entering a search term in an internet browser, opening and saving a document)“ (Senkbeil, Ihme & Wittwer, in press). Die Annahme mehrerer Komponenten von ICT-Literacy basiert auf dem theoretischen Rahmenmodell, das vom International ICT-Literacy Panel (2007) entwickelt wurde. Dieses fokussiert kognitive und technologische Aspekte der ICT-Literacy und beschreibt verschiedene Prozesskomponenten, die gemeinsam ICT-Literacy konstituieren: "ICT literacy is using digital technology, communications tools, and/or networks to access, manage, integrate, evaluate and create information in order to function in a knowledge society" (S. 2). Das Konzept des Informationszugriffs („Access“) lässt sich theoretisch in der Leseforschung verankern (z. B. Rouet, 2006; vgl. auch Leseaspekte in Large-Scale Assessments, z. B. OECD, 2009a). Wenn z. B. eine Frage mithilfe eines Textdokumentes beantwortet werden soll, muss dafür auf Informationen unter Anwendung von Suchstrategien effektiv zugegriffen werden. Im ICT-Kontext besteht die Informationsquelle jedoch nicht mehr nur aus einem oder mehreren

Textdokumenten, sondern aus vielen über Links miteinander verbundenen Web-Seiten, Online-Foren, E-Mails, Datenbanken, etc. Entsprechend erfordert der Informationszugriff ICT-spezifische Fertigkeiten, wie etwa das Formulieren von Abfragen in einer Suchmaschine passend zum Informationsbedarf oder das zielorientierte Navigieren in Hypertexten.

„Access“ überlappt konzeptuell insofern mit basalen ICT-Fertigkeiten, als „Access“ auch einfache Operationen umfasst, die den Informationszugriff ermöglichen, wie z. B. in einem Web-Browser die Schaltflächen für Navigation zu benutzen. Basic Computer Skills (BCS) dienen in dieser Studie als konvergentes Validitätskriterium für „Access“ und werden definiert als „...the fundamental ability and speed of performing basic actions in graphical user interfaces of computers to access, collect, and provide information“ (Goldhammer, et al., 2013, S. 2). Das Konstrukt BCS ist breiter als „Access“ definiert, da es neben basalen Fertigkeiten zum Zugriff und Abruf von Information auch solche zur organisierten Sammlung und Bereitstellung von Information beinhaltet. Solche Fertigkeiten können als generisch verstanden werden, da sie über unterschiedliche Betriebssysteme und Softwareanwendungen hinweg von Relevanz sind (Goldhammer, et al., 2013; Markauskaite, 2007). Sie umfassen nur wenige Teilschritte, z. B. einen Text auswählen, kopieren und einfügen, die in unterschiedlichen Anwendungskontexten aufgeführt werden und eine Voraussetzung für fortgeschrittenere ICT-Nutzung darstellen (vgl. De Wit, Heerwegh & Verhoeven, 2012).

Multiple-Choice-Aufgaben vs. simulationsbasierte Aufgaben

MC-Aufgaben werden aufgrund ihrer Effizienz in der Erfassung von kognitiven Wissens- oder Fertigkeitsdomänen (vgl. Downing, 2006) in vielen groß angelegten Leistungsstudien (Large-Scale Assessments) verwendet. MC-Aufgaben können automatisch bewertet werden und weisen dementsprechend eine hohe Auswertungsobjektivität auf. Diese Vorteile sind bei Verwendung offener Antwortformate (constructed response, z. B. Freitext) hingegen häufig schwieriger zu realisieren (Downing, 2006; Haladyna & Downing, 2004). MC-Aufgaben können effizient kognitive Wissensdomänen abdecken und liefern reliable Messungen. Darüber

hinaus gibt es Studien, die auf Konstruktäquivalenz zwischen MC- und Constructed-response-Aufgaben unter bestimmten Bedingungen hinweisen (vgl. die Übersichtsarbeit von Rodriguez, 2003).

Aufgaben im MC-Formate haben jedoch den Nachteil, dass sie für die Erfassung komplexer Produktionsaufgaben (z. B. Schreiben) nicht geeignet sind. Zudem unterliegen sie der Gefahr, dass die korrekte Antwort durch logisches Ausschließen von Distraktoren ermittelt oder zufällig geraten werden kann, was die Konstruktvalidität mindert (Downing, 2006). Zentral für die vorliegende Studie ist, dass der MC-Ansatz keine innovativen Itemformate zulässt, um Wissen, Fertigkeiten und Fähigkeiten in realistischer und authentischer Weise zu messen (Sireci & Zenisky, 2006). Dies kann jedoch ausschlaggebend sein, um die angemessene Repräsentation des Konstruktes sicherzustellen oder zu verbessern, beispielsweise wenn das Konstrukt eine Interaktion mit dem Stimulus und ggf. dessen Manipulation erforderlich macht, wie z. B. im Falle von ICT-Literacy (vgl. auch Problemlösen unter Verwendung von ICT, OECD, 2009b).

Simulationen können als spezielle Form innovativer Itemformate verstanden werden, die erst durch computerbasiertes Testen ermöglicht werden. Sie schöpfen die Möglichkeiten einer elektronischen Aufgabendarbietung in vollem Maße aus, indem sie multimediale Inhalte, d. h. Text, Grafiken, Video und Audio, in das Stimulusmaterial einbinden und ein hohes Maß an Interaktivität zwischen Antwortverhalten und Stimulusdarbietung ermöglichen. Ein typisches Beispiel ist die Problemlösungsvignette, bei der die Testperson in einem simulierten Problemraum und in mehreren Schritten ein bestimmtes Problem lösen muss (Sireci & Zenisky, 2006; vgl. auch komplexes Problemlösen, Greiff, Wüstenberg, Molnar, Fischer, Funke & Csapo, 2013). Potenzielle Nachteile solcher Simulationen beziehen sich auf konstruktirrelevante Varianz durch das Präsentationsmedium (vgl. Effekte des Administrationsmodus, Kröhne & Martens, 2011), eingeschränkte Auswertungsobjektivität bei offenen Antwortformaten sowie längere Testzeit infolge der verstärkten Interaktionsmöglichkeiten mit dem Stimulus bzw. bei gleicher Testzeit geringere Reliabilität (s. z. B. Greiff et al. 2013; bzgl. der

Implementierungsherausforderungen s. Sireci & Zenisky, 2006; Greiff, Wüstenberg, Holt, Goldhammer & Funke, 2013).

Erfassung der ICT-Literacy-Komponente „Access“

Die Frage des passenden Antwortmodus zur Erfassung der ICT-Literacy-Komponente „Access“ lässt sich anhand von Formen der Wissensrepräsentation behandeln. Neben deklarativem Wissen, das die Beantwortung von Fragen erlaubt, und prozeduralem Wissen, das erfolgreiches Handeln ermöglicht, lässt sich jeweils nach Sachwissen, also „was ist der Fall?“, und Handlungswissen, also „was ist zu tun?“, differenzieren (Süß, 1996; s. auch Anderson, 1996). Während MC-Aufgaben geeignet sind, deklaratives Sach- und Handlungswissen sowie prozedurales Sachwissen zu erfassen, sind zur Repräsentation von prozeduralem Handlungswissen MC-Aufgaben offenbar weniger gut geeignet als simulationsbasierte Aufgaben, die tatsächliches Handeln beinhalten.

Aus theoretischer Sicht erfordert die Erfassung der ICT-Literacy-Komponente „Access“ die Abbildung individueller Unterschiede in prozeduralem ICT-Handlungswissen, z. B. indem die Testperson aktiv innerhalb einer (simulierten) ICT-Umgebung agieren muss. Zwar lässt sich in MC-Aufgaben auf der Stimulusseite die Authentizität vorgegebener Problemszenarien durch Screenshots erhöhen, die Aufgabenbearbeitung selbst ist aber durch die fehlende Interaktivität statischer Screenshots und die Vorgabe von MC-Antwortalternativen sehr beschränkt. Hinsichtlich des Antwortmodus bleibt zur Erfassung von prozeduralem (Sach-)Wissen nur die beschränkte Möglichkeit, einschrittiges Interaktionsverhalten als Antwortalternativen zu verwenden, z. B. Labels, die auf Interaktionsflächen im Screenshot verweisen oder direkt Bilder dieser Interaktionsflächen. Komplexeres, mehrschrittiges Interaktionsverhalten ließe sich lediglich in Form verbaler Beschreibungen von zielorientiertem Verhalten in den Antwortalternativen unterbringen. Während hierbei der Antwortraum bei MC-Aufgaben auf vier bis sechs Möglichkeiten beschränkt ist, ist der einer simulationsbasierten Aufgabe ungleich

größer. Ein Beispiel für eine mehrschrittig zu bewältigende Aufgabe wäre etwa, aus einer Menüstruktur das richtige Menü auszuwählen und darin ein Menü-Item anzuklicken.

Fragestellung und Hypothesen

In der vorliegenden Studie wurde die ICT-Literacy-Komponente „Access“ in Form von MC-Aufgaben und in Form von inhaltsgleichen SIM-Aufgaben erhoben. Durch ein experimentelles Zwischensubjekt-Design können Unterschiede in den Antworten der beiden zufallsäquivalenten Gruppen auf Unterschiede in der jeweils eingesetzten Testform (MC-Skala vs. SIM-Skala) zurückgeführt werden. Um darüber hinaus Unterschiede in der konvergenten Validität zu untersuchen, bearbeiteten beide Gruppen zusätzlich die „Basic Computer Skills“ (BCS)-Skala, die ein verwandtes Konstrukt erfasst.

Fragestellung 1: Klassische Itemeigenschaften und Reliabilität. Sind die Access-MC-Skala und die Access-SIM-Skala im Hinblick auf klassische Item- und Skalencharakteristiken vergleichbar? Um diese Frage zu beantworten, werden die beiden Skalen auf Itemebene hinsichtlich des Anteils korrekter Antworten und der punktbiserialen Korrelation mit dem Gesamttestwert sowie auf Skalenebene hinsichtlich der Reliabilität (bestimmt als Cronbachs α) miteinander verglichen.

Wir nehmen an, dass Access-SIM-Items schwieriger sind als Access-MC-Items, da simulierte Umgebungen mehr Antwortalternativen aufweisen bzw. in MC-Aufgaben durch die geringe Anzahl von Distraktoren das Erraten der korrekten Lösung eher möglich ist (insbesondere, falls Distraktoren ungünstig gewählt wurden, vgl. Haladyna, 2004). Zudem beinhalten SIM-Items im Vergleich zu MC-Items mehr Anforderungen. In der Regel sind mehrere Schritte durchzuführen (z. B. Menü und Menü-Item auswählen, Dialogoptionen bearbeiten und bestätigen), die zudem die Handhabung von Maus und/oder Tastatur erfordern.

Fragestellung 2: Messmodellinvarianz. In welchem Maße weisen die Access-MC-Skala und die Access-SIM-Skala dieselben Messmodelleigenschaften auf? Zur Überprüfung der Messmodellinvarianz beider Skalen wird ein liberales Zwei-Gruppen-Messmodell getestet, das

keine Zwischen-Gruppen-Restriktionen für Faktorladungen und Schwellen enthält (konfigurale Invarianz). Auf diese Weise können die Messmodellparameter der MC-Skala mit denen der SIM-Skala verglichen werden. Zusätzlich erfolgt ein Vergleich dieses liberalen Modells mit restriktiveren Modellen, in denen sowohl für Faktorladungen und Schwellen keine Zwischen-Gruppen-Variation zugelassen wird (starke Invarianz). Unter der Annahme, dass Access-SIM-Items schwieriger sind als MC-Items, sollte das Modell mit starker Messmodellinvarianz abgelehnt werden. Zur Vergleichbarkeit der Faktorladungen bestehen keine Vorannahmen.

Fragestellung 3: Konvergente Validität. Zeigen die Access-MC-Skala und die Access-SIM-Skala dieselben Zusammenhänge mit der BCS-Skala, welche das verwandten Konstrukt „Basic Computer Skills“ erfasst? Hierfür wurde die Gleichheit der Korrelationen von Access-MC bzw. Access-SIM mit BCS getestet.

Aus theoretischer Sicht nehmen wir an, dass „Access“ konvergente Validität mit BCS aufweist bzw. zwischen beiden Variablen ein starker positiver Zusammenhang besteht. Beide Konstrukte zeigen insoweit konzeptuelle Gemeinsamkeiten, als sie beide basale Operationen des Informationszugriffes im ICT-Kontext beinhalten. Infolge ihrer Operationalisierungen gehen wir aber davon aus, dass Access-MC und Access-SIM nicht dasselbe Konstrukt erfassen. SIM-Aufgaben enthalten die konstruktrelevante Anforderung, prozedurales Handlungswissen anzuwenden, was eine bessere Konstruktrepräsentation und somit validere Messung von „Access“ bedeutet. Dies sollte sich auch in einer höheren Korrelation zwischen Access-SIM und BCS als zwischen Access-MC und BCS bemerkbar machen. BCS wurde in vorliegender Studie wie Access-SIM ebenfalls anhand computerbasierter Simulationsaufgaben gemessen.

Material und Methode

Stichprobe

Insgesamt bearbeiteten 405 Schülerinnen und Schüler der Klassenstufe 9 aus unterschiedlichen Schulformen die computerbasierten Items. Die Stichprobe setzte sich aus 46 % weiblichen, 46 % männlichen und 8 % Jugendlichen unbekanntes Geschlechts zusammen.

Die Testpersonen waren zwischen 14 und 20 Jahre alt ($M = 15.8$, $SD = .82$). Die Daten wurden im Rahmen einer Pilotierungsstudie des NEPS erhoben, das durch das Bundesministerium für Bildung und Forschung (BMBF) finanziert wird. Zusätzlich zu den aufgeführten Tests beinhaltete die Testbatterie der Pilotierungsstudie weitere Instrumente, welche für die vorliegende Studie nicht von Relevanz waren. Die Erhebungen wurden klassenbasiert in den jeweiligen Schulen unter Aufsicht von einem oder zwei Testleiterinnen bzw. Testleitern durchgeführt. Die gesamte Bearbeitungszeit inkl. Pausen lag bei ca. 120 Minuten.

Aufgabenmaterial

Um eine Konfundierung zwischen der Art der Itemdarbietung (Papier vs. Computer) und dem Messansatz (MC vs. SIM) zu vermeiden, wurden nicht nur die SIM-Aufgaben, sondern auch die MC-Aufgaben computerbasiert vorgegeben. Alle Aufgaben wurden mit dem graphischen Autorenwerkzeug CBA ItemBuilder (Rölke, 2012) implementiert.

Die Testpersonen konnten zwischen den verschiedenen Aufgaben navigieren, indem sie am linken Monitorrand einen „Weiter“-Button anklickten. Dieser Vorgang wurde in einem Übungsdurchlauf zu Beginn des Tests eingeübt. Die Testpersonen wurden instruiert, die Access- und BCS-Aufgaben so schnell und so genau wie möglich zu bearbeiten.

Access-MC-Aufgaben. Die elf MC-Aufgaben wurden im Rahmen der NEPS-Studie entwickelt, um die ICT-Literacy-Prozesskomponente „Access“ zu erfassen (für eine ausführliche Darstellung siehe Senkbeil et al., in press).

In den Access-MC-Aufgaben werden Screenshots aus realen Softwareanwendungen (Textverarbeitung, Tabellenkalkulation, Bildschirmpräsentationssoftware, Web Browser, Email-Programm) präsentiert. Dabei werden in manchen Fällen mehrere Screenshots bzw. Ausschnitte aus Screenshots gleichzeitig dargeboten, um mehrere Antwortalternativen darzustellen (z. B. mehrere Aufklappmenüs). Sechs Items sind einfache Mehrfachwahlaufgaben mit vier Antwortalternativen; zwei Aufgaben haben fünf Antwortoptionen. Drei weitere Aufgaben sind „komplexe“ Mehrfachwahlaufgaben. Diese enthalten mehrere Mehrfachwahlaufgaben mit

jeweils zwei Antwortalternativen und werden zu einem Itemscore zusammengefasst (vgl. Pohl & Carstensen, in press). Eine dieser komplexen Mehrfachwahlaufgaben hat vier Antwortalternativen, zwei Aufgaben haben fünf. Die Antwortalternativen werden unterhalb des Screenshots präsentiert. Sie beschreiben entweder Handlungsalternativen, d. h. das, was man tun könnte, um die Aufgabe zu lösen, oder sie beziehen sich direkt auf Elemente des Screenshots (z. B. Labels, die angeben, welcher Button gedrückt werden soll; siehe auch Abbildung 1a).

Access-SIM-Aufgaben. Die elf Access-MC-Aufgaben wurden in inhaltsgleiche Access-SIM-Aufgaben konvertiert, in denen die Testperson tatsächlich mit einer simulierten Computerumgebung interagieren musste. Beispielsweise muss die Testperson, statt eine Antwortalternative auszuwählen, die sich auf ein bestimmtes Menü-Item im Screenshot bezieht, aktiv das richtige Menü auswählen und dann auf das korrekte Menü-Item klicken. Die simulierten Umgebungen wurden dabei von real existierender Software und Betriebssystemen abstrahiert. Sie wurden jeweils so gestaltet, dass sie bezüglich allgemeiner Merkmale interaktiver Funktionen mit realen Computerumgebungen übereinstimmen (z. B. das Klicken auf ein Menü mit der linken Maustaste öffnet die Liste der Menü-Items).

Abbildung 1 zeigt als Beispielaufgabe die computerisierte Fassung der MC-Aufgabe „Live Ticker“ (Item 9) sowie die konvertierte Aufgabe in der Access-SIM-Skala. Hier ist bei der Beantwortung anzuzeigen, wie die im Live-Ticker präsentierte Information aktualisiert wird.

BCS-Aufgaben. Basale Computerfertigkeiten wurden mit einer revidierten Fassung der Basic Computer Skills (BCS)-Skala von Goldhammer et al. (2013) erfasst. Diese umfasst 20 interaktive Items, die die Fertigkeit messen, basale Operationen in graphischen Benutzeroberflächen vorzunehmen, um auf Informationen zuzugreifen, diese zu sammeln und verfügbar zu machen. Beispiele hierfür wären etwa die Bedienung von Maus oder Tastatur, um Textdokumente zu öffnen, zu speichern, zu löschen oder zu ändern, sowie um Menüleisten, Hyperlinks, Bildlaufleisten und Hotkeys zu bedienen.

Design und Untersuchungsablauf

Es wurde ein experimentelles Zwischensubjekt-Design verwendet. Die Zuweisung der Testpersonen zu den Bedingungen Access-MC ($n = 211$) mit MC-Items und Access-SIM ($n = 194$) mit inhaltsgleichen simulationsbasierten Items erfolgte randomisiert. Eine Überprüfung der auf Schülerebene durchgeführten Randomisierung ergab, dass zwischen der MC- und der SIM-Gruppe keine bedeutsamen Unterschiede in der Verteilung der Variablen Alter, Geschlecht und Schulform bestehen.

Zwischen den Items konnte nur vorwärts navigiert werden, d. h. ein Item konnte nicht erneut besucht werden. Bei der Bearbeitung der Access-MC-Items konnte eine abgegebene Antwort durch die Wahl einer alternativen Antwortmöglichkeit (radio button) oder Deselektion einer Antwortmöglichkeit (check box) korrigiert werden, bei den Access-SIM-Items hingegen führten falsche Teilschritte unmittelbar zu einer falschen Antwort, d. h. es konnte nicht frei exploriert und ggf. im Verlauf ein fehlerhafter Bearbeitungsschritt korrigiert werden.

Die Reihenfolge der Access-Items war zwischen den Bedingungen SIM und MC und somit auch zwischen den Personen identisch. Vor der Bearbeitung der Access-Items bearbeiteten alle Testpersonen den Basic Computer Skills (BCS)-Test. Die Bearbeitungszeit der Access- und der BCS-Items betrug insgesamt etwa 20 Minuten.

Datenanalyse

Auf Basis der klassischen Testtheorie (KTT) wurden als Itemkennwerte der Anteil korrekter Antworten (p) und die punktbiseriale Korrelation (r_{it}) zwischen Itemantwort und Testwert berechnet sowie auf Skalenebene die Reliabilität in Form von Cronbachs α . Alle Kennwerte wurden mit Hilfe des Paketes ltm (Rizopoulos, 2006) im Rahmen der Statistikumgebung R (R Development Core Team, 2011) ermittelt.

Um das Ausmaß der Messmodellinvarianz für die Access-MC- und die Access-SIM-Skala zu überprüfen, wurden zunächst in einem vorbereitenden Schritt konfirmatorische Faktorenanalysen (CFA) separat für die MC- und die SIM-Gruppe durchgeführt (baseline

models, vgl. Byrne & Stewart, 2006). Mit diesem Schritt sollte sichergestellt werden, dass eventuelle Anpassungsprobleme in den Modellen zur Prüfung der Messmodellinvarianz nicht durch Fehlspezifikationen innerhalb der beiden Gruppen bedingt sind. Ergänzend zum klassischen Reliabilitätskoeffizienten α wurde die Reliabilität basierend auf den Modellparametern der (nichtlinearen) CFA-Modelle bestimmt, welche von weniger strengen Annahmen ausgeht. Das Vorgehen folgte dem Ansatz zur Bestimmung der Reliabilität in Strukturgleichungsmodellen (SEM) von Green und Yang (2009) unter Verwendung des von den Autoren veröffentlichten SAS-Codes, den wir für R angepasst haben.

Im Folgenden wurden Mehrgruppen-CFA-Modelle (s. z. B. Bollen, 1989) mit unterschiedlich starken Zwischen-Gruppen-Restriktionen bezüglich der Eigenschaften des Messmodells getestet (vgl. Muthén & Asparouhov, 2002, Muthén & Muthén, 1998–2012). Dieser Ansatz wurde im Anschluss zu einem Strukturgleichungsmodell für mehrere Gruppen ausgeweitet, um die Vergleichbarkeit konvergenter Validität zu untersuchen, d. h. die Frage der Gleichheit der Korrelation mit dem Validierungskriterium BCS. Um die Modelle mit kategorialen Indikatoren zu testen, wurde der WLSMV-Schätzer verwendet (s. z. B. Forero & Maydeu-Olivares, 2009). Die Analysen wurden mithilfe der Software Mplus (Version 7, Muthén & Muthén, 1998–2012) und der Standard (Delta)-Parametrisierung für Mehrgruppenanalysen mit kategorialen Indikatoren (vgl. Muthén & Asparouhov, 2002) durchgeführt.

Der Modellfit für ein Modell wurde dann als gut angesehen, wenn folgende Kriterien erfüllt waren (Werte in Klammern entsprechen einem noch akzeptablen Fit): Ein nichtsignifikanter χ^2 -Wert, Verhältnis aus χ^2 -Wert und Freiheitsgraden unter 2 (3), Root Mean Square Error of Approximation (RMSEA) unter .05 (.08), Comparative Fit Index (CFI) und Nonnormed Fit Index (NNFI) über .97 (.95) (vgl. Schermelleh-Engel, Moosbrugger & Müller, 2003). Für Modellvergleiche wurde die Option DIFFTEST von Mplus verwendet, da beim WLSMV-Schätzer die Differenz zwischen χ^2 -Werten nicht χ^2 -verteilt und dementsprechend eine

Korrektur erforderlich ist. Zusätzlich zu diesem statistischen Kriterium wurde als praktisches Kriterium ΔCFI herangezogen. Für die Annahme der Invarianz zwischen zwei Modellen sollte dieser Wert nicht größer als .01 sein (vgl. Byrne & Stewart, 2006).

Ergebnisse

Klassische Itemkennwerte und Reliabilität

In Tabelle 1 sind die klassischen Itemkennwerte für beide Skalen (Access-MC und Access-SIM) dargestellt, d. h. der Anteil richtiger Antworten (p) als ein Maß für die Itemschwierigkeit und die punktbiseriale Korrelation (r_{it}) zwischen Itemantwort und Testwert als Trennschärfeindex. Es zeigt sich, dass die MC-Items relativ leicht zu lösen waren: Die meisten Items wiesen Lösungshäufigkeiten von weit über .50 auf, wohingegen die SIM-Items deutlich geringe Lösungshäufigkeiten besaßen. Die mittlere Lösungshäufigkeit betrug $M(p) = .54$ für Access-MC und $M(p) = .36$ für Access-SIM. Sämtliche p -Werte waren mit einer Ausnahme (Item 7 „Search Images 2“) kleiner als die der korrespondierenden MC-Items und nur drei Werte lagen über .50. Die Trennschärfen der MC-Items waren mit Werten von durchgehend über .20, bis auf eine Ausnahme (Item 4 „Slide design“), ausreichend bis gut; bei den SIM-Items zeigten zwei Items (Item 3 „Find email“, Item 5 „Blind cc“) unzureichende Trennschärfen. Anders als bei den Lösungshäufigkeiten waren bei den Trennschärfen keine systematischen Unterschiede zwischen MC- und SIM-Items festzustellen.

Als Reliabilitätsmaß wurde Cronbachs α ermittelt. Für die Access-MC-Skala ergab sich ein Wert von $\alpha = .62$, für die Access-SIM-Skala war der Wert mit $\alpha = .49$ etwas geringer.

Messmodellinvarianz

Zur Klärung der Frage, ob die psychometrischen Eigenschaften der MC-Skala und der SIM-Skala vergleichbar sind, testeten wir Mehrgruppen-CFA-Modelle, die unterschiedlich restriktiv waren. Die beiden (zufallsäquivalenten) Gruppen unterschieden sich hinsichtlich des Operationalisierungsansatzes (MC vs. SIM).

Vorauslaufend prüften wir für jede Gruppe separat ein CFA-Modell, das jeweils eine eindimensionale Struktur annahm. Für die Access-MC-Skala wurde ein noch akzeptabler Modellfit erzielt, $\chi^2(44, N = 206) = 55.08, p = .11, RMSEA = .037, CFI = .90, NNFI = .89$, für die Access-SIM-Skala fiel der Modellfit sehr gut aus, $\chi^2(44, N = 190) = 45.70, p = .40, RMSEA = .014, CFI = .98, NNFI = .97$. Item 5 („Blind cc“) wies in beiden Gruppen keine signifikante Ladung auf, weshalb es in den folgenden Analysen ausgeschlossen wurde. Die revidierten Messmodelle zeigten einen akzeptablen bis sehr guten Modellfit, d. h. für die Access-MC-Skala, $\chi^2(35, N = 205) = 31.03, p = .66, RMSEA < .001, CFI = 1.00, NNFI = 1.04$, und für die Access-SIM-Skala, $\chi^2(35, N = 190) = 39.55, p = .27, RMSEA = .026, CFI = .94, NNFI = .92$. Für die revidierten Modelle ergaben sich vergleichbare hohe SEM-Reliabilitäten von .57 für die Access-MC-Skala und von .54 für die Access-SIM-Skala (die entsprechenden Werte für Cronbachs α lagen bei .61 und .50).

Im folgenden Schritt wurde zunächst das liberalste Zweigruppen-Modell getestet, das konfigurale Invarianz annimmt. Hierbei wurde nur die Faktor- und Ladungsstruktur zwischen den Gruppen gleichgesetzt, die Höhe der Ladungen und der Schwellen hingegen durften variieren. Aufgrund der Annahme zufallsäquivalenter Gruppen waren die Strukturparameter gleichgesetzt. Dazu wurde der Mittelwert der latenten Faktorvariablen in beiden Gruppen auf 0 fixiert, die Metrik der latenten Faktorvariablen wurde in beiden Gruppen durch Fixierung der Varianz der Faktorvariablen auf 1 festgelegt. Zur Identifikation des Modells wurden die Skalierungsfaktoren auf 1 fixiert (vgl. Muthén & Asparouhov, 2002).

Das liberale Modell konfiguraler Invarianz schnitt erwartungsgemäß zufriedenstellend ab, $\chi^2(70, N = 396) = 85.16, p = .11, RMSEA = .033, CFI = .92, NNFI = .89$. Abbildung 2 präsentiert die geschätzten Parameter der Messmodelle jeweils für Gruppe MC und SIM (Item 5, „Blind cc“ war ausgeschlossen). Teil (a) zeigt die Ladungen der Items auf dem latenten Faktor. Wie schon beim klassischen Kennwert für die Trennschärfe sind die Ladungen von

Item 3 der SIM-Skala und Item 4 der MC-Skala unzureichend; Unterschiede zwischen MC- und SIM-Items sind unsystematisch. Teil (b) präsentiert die Schwellen der Items, welche als Schwierigkeiten der Items zu interpretieren sind. Demnach sind die Items der SIM-Skala fast durchweg schwieriger. Dabei ist zu beachten, dass die Schwelle eines Items numerisch nicht identisch mit dem klassischen Schwierigkeitsparameter p ist und auch nicht dem IRT-Schwierigkeitsparameter, wobei letzterer und die Schwelle ineinander umgerechnet werden können (s. Muthén & Muthén, 2010).

Als nächster Schritt folgte die Prüfung strenger Invarianz, welche gegenüber dem Modell konfiguraler Invarianz zwischen den Gruppen auch gleiche Werte in den Ladungen und Schwellen annimmt. Erwartungsgemäß war dieses Modell als zu restriktiv zurückzuweisen, $\chi^2(80, N = 396) = 185.06, p < .01, RMSEA = .081, CFI = .42, NNFI = .34$. Ein Vergleich mit dem Modell konfiguraler Invarianz bestätigte den signifikant schlechteren Modellfit, $\chi^2(10, N = 396) = 81.48, p < .01$. Zudem wurde das praktische Kriterium für Messmodellinvarianz, $\Delta CFI \leq .01$, klar nicht erfüllt.

Wie von Muthén und Asparouhov (2002) vorgeschlagenen, haben wir auf Grundlage der Modifikationsindizes die Zwischen-Gruppen-Restriktion für Schwelle und Ladung sukzessive für einzelne Items freigesetzt. Dem von Muthén und Muthén (1998–2012) für Mplus beschriebenen Vorgehen folgend, wurde die Liberalisierung für Schwelle und Ladung simultan vorgenommen. Um auf diese Weise einen akzeptablen Modellfit zu erreichen, $\chi^2(73, N = 396) = 88.02, p = .11, RMSEA = .032, CFI = .92, NNFI = .90$, musste für alle Items mit Ausnahme der Items 2, 3 und 4 die paarweise Freisetzung erfolgen. Das im jeweiligen Schritt freigesetzte Item wies die höchsten Modifikationsindizes für Schwelle als auch Ladung auf. Gegenüber dem Modell konfiguraler Invarianz passte das so gebildete Modell nicht schlechter, $\chi^2(3, N = 396) = 3.93, p = .27$.

Zusammengenommen stützen die Ergebnisse die Annahme konfiguraler Invarianz zwischen den Messmodellen für Access-MC und Access-SIM. Erwartungsgemäß zeigte sich, dass für die meisten Items keine Gleichheit von Ladungen bzw. Schwellen (starke Invarianz) angenommen werden kann.

Konvergente Validität

Zur Frage der Vergleichbarkeit konvergenter Validität der Access-MC- und Access-SIM-Skala prüften wir, ob die Korrelationen zwischen BCS und Access-MC einerseits sowie BCS und Access-SIM andererseits empirisch unterschiedlich ausfallen. Der Vergleich von Faktorkorrelationen bzw. die Testung ihrer Invarianz setzt in Mehrgruppenanalysen üblicherweise voraus, dass die Faktorvarianzen und (weitestgehend) die Ladungen zwischen den Gruppen invariant sind (vgl. Vandenberg & Lance, 2000). Diese Voraussetzung ist nur teilweise erfüllt, wie die Analysen im vorangegangenen Abschnitt zeigten. Da sich in unserer Studie die (zufallsäquivalenten) Gruppen nur in den vorgelegten Messungen, d. h. Access-MC- vs. Access-SIM-Skala, unterscheiden und die Messungen selbst Gegenstand des Vergleichs sind, ist die Invarianz von Ladungen jedoch keine notwendige Bedingung. Potenzielle Unterschiede in den Korrelationen können eindeutig auf die unterschiedlichen Operationalisierungsansätze zurückgeführt werden.

Für die BCS-Skala ergaben sich zufriedenstellende psychometrische Eigenschaften. Die Reliabilität (berechnet über Cronbachs α) betrug für die Gesamtstichprobe .73, die Itemschwierigkeiten p variierten zwischen .23 bis .95 und die Trennschärfen r_{it} lagen zwischen .04 bis .51.

Mittels CFA wurde für die Gesamtgruppe ein BCS-Messmodell getestet. Der erzielte Modellfit fiel gemischt aus, $\chi^2(170, N = 404) = 217.25, p < .01$, RMSEA = .026, CFI = .85, NNFI = .83. Wie von Kenny (2012) diskutiert, können inkrementelle Fitindizes wie CFI und NNFI unter Umständen zur Beurteilung des Modellfits nicht angemessen sein, wenn die Korrelationen zwischen (latenten) Indikatorvariablen niedrig sind. In diesem Fall kann das Null-

(oder Unabhängigkeits-) Modell eine relativ gute Passung besitzen. Als Daumenregel untersuchten wir deshalb den RMSEA des Nullmodells, welcher .063 betrug und somit deutlich unter dem von Kenny (2012) vorgeschlagenem Schwellwert von .158 lag. Das legt nahe, dass die inkrementelle Fitindizes in diesem Fall nicht sehr informativ sind. Basierend auf der χ^2 -Statistik und dem RMSEA kann somit der Fit des Modells als akzeptabel angesehen werden. Die SEM-Reliabilität der BCS-Skala betrug .61.

Um Fragestellung 3 zu beantworten, wurden die Items der BCS-Skala und die Items der Access-Skala zusammen in einer Mehrgruppen-CFA getestet. Für das Messmodell der BCS-Skala wurde zwischen den Gruppen starke Invarianz angenommen, da die Gruppen zufallsäquivalent waren und das identische Instrument administriert wurde. Für das Messmodell der Access-Skala wurde das im vorangegangenen Abschnitt beschriebene finale Modell, das für drei Items jeweils gleiche Schwellen und Ladungen annimmt, eingesetzt.

Zunächst wurde ein Modell geschätzt, das zwischen den Gruppen MC und SIM Unterschiede in der Korrelation von „Access“ mit BCS zuließ. Wie zuvor fiel der Modellfit bis auf die inkrementellen Fitindizes gut aus, $\chi^2(831, N = 404) = 904.05, p = .04, RMSEA = .021, CFI = .87, NNFI = .87$. Für Access-MC betrug die Korrelation mit BCS $r = .84 (z = 12.26, p < .01)$, und für Access-SIM fiel sie mit $r = .80 (z = 9.76, p < .01)$ vergleichbar hoch aus. Der χ^2 -Differenztest zeigte an, dass die Gleichsetzung der beiden Korrelationen den Modellfit nicht signifikant verschlechtert, $\chi^2(1, N = 404) = 1.10, p = .29$. Daraus folgt, dass sich die Skalen Access-MC und Access-SIM bzgl. der Korrelation mit BCS nicht unterscheiden und somit von vergleichbarer konvergenter Validität zu BCS auszugehen ist.

Diskussion

Die Frage, ob sich MC-Aufgaben und inhaltsgleiche simulationsbasierte Aufgaben zur Erfassung der ICT-Literacy-Komponente „Access“ aus psychometrischer Sicht unterscheiden,

kann bejaht werden, wenngleich weniger eindeutig als erwartet. Zwar fanden sich die erwarteten Schwierigkeitsunterschiede, allerdings kein Unterschied in der konvergenten Validität mit BCS.

Unsere Ergebnisse zeigen erwartungskonform, dass die SIM-Aufgaben schwieriger sind als die inhaltsgleichen MC-Aufgaben. Die Unterschiede in den Lösungshäufigkeiten und Schwellen der Messmodelle fielen nicht gleichförmig aus, sondern variierten über die Items. Beispielsweise waren für Item 8 („Tiff“), welches die Speicherung einer Datei in einem bestimmten Format erforderte, die Itemschwierigkeiten und auch die Schwellen für beide Testformen vergleichbar. Die MC-Aufgabe beinhaltete vier Antwortalternativen (drei Buttons und ein Aufklappmenü), welche gemeinsam im Aufgabenstimulus in Form eines Screenshots vorgegeben waren. Die SIM-Aufgabe unterschied sich in der Anforderung nur darin, dass das Aufklappmenü angeklickt werden musste, um den richtigen Dateityp auszuwählen. Im Gegensatz dazu fiel der Unterschied in der Lösungshäufigkeit und den Schwellen bei Item 10 („Animals“) zwischen beiden Testformen erheblich größer aus: In dieser Aufgabe war eine Tabelle mit verschiedenen Angaben (Haustiere mit Angabe des Besitzers, der Tierart, des Geschlechts, der Fellfarbe, des letzten Besuchs) nach zwei Kriterien (Tierart und Fellfarbe) zu ordnen. Während in der MC-Aufgabe verschiedene Antwortalternativen präsentiert wurden, die jeweils zwei Sortierkriterien in unterschiedlichen Reihenfolgen darboten, musste die Testperson in der SIM-Aufgabe selbstständig das richtige Menü mit den Sortieroptionen finden und öffnen sowie die beiden korrekten Sortierungskriterien generieren.

Im ersten Beispiel (Item 8, „Tiff“) war die Information, die für die Lösung der Aufgabe benötigt wurde, im Stimulus zu erkennen und wurde in der MC- und der SIM-Aufgabe identisch dargeboten. Im zweiten Beispiel (Item 10, „Animals“) hingegen musste die Testperson in der SIM-Aufgabe die Menüauswahl selbstständig treffen und den aufgerufenen Sortierdialog richtig anwenden, während in der MC-Aufgabe die Kombinationen der Suchkriterien direkt miteinander verglichen werden konnten. Im Allgemeinen waren die Antwortmöglichkeiten in den MC-Aufgaben sehr stark beschränkt. In den SIM-Aufgaben hingegen waren die

Anforderungen insofern höher, als ICT-Basisfertigkeiten einzusetzen waren, in der Regel mehrere Bearbeitungsschritte für eine korrekte Antwort verlangt wurden und ein einfacher Ausschluss von Antwortoptionen durch logisches Schlussfolgern nicht möglich war. Die teilweise deutlichen Unterschiede in der Itemschwierigkeit sind auch darauf zurückzuführen, dass die SIM-Aufgaben nicht explorierend gelöst werden konnten bzw. für eine korrekte Lösung kein falscher Teilschritt erfolgen durfte.

Die Analyse der Messmodellinvarianz ergab, dass konfigurale Invarianz angenommen werden kann, nicht jedoch starke Invarianz, die die Gleichheit sowohl von Faktorladungen als auch von Schwellen beinhaltet. Vergleicht man die frei geschätzten Messmodellparameter (s. Abb. 2) zeigen sich vor allem bei den Schwellen deutliche und systematische Unterschiede. Insgesamt konnte nur für drei Items die Gleichheit von Schwellen und Ladungen angenommen werden. Die Unterschiede in den Ladungen bedeuten, dass die latente Variable „Access“ die beobachteten Variablen in den beiden Skalen bis auf die drei Ausnahmen in unterschiedlicher Weise erklärt. Das sagt noch nichts darüber aus, ob die beiden latenten Variablen identisch sind und jeweils dasselbe „Access“-Konstrukt widerspiegeln oder nicht. Um diese Frage näher zu beleuchten und um jeweils die Stärke konvergenter Validität zu ermitteln, untersuchten wir deshalb den Zusammenhang zwischen Access-MC und Access-SIM einerseits und der BCS-Variablen andererseits und verglichen die Korrelationen.

Erwartungskonträr zeigen die Ergebnisse, dass Access-MC genauso stark mit BCS korreliert ist wie Access-SIM. Da die Annahme gleicher Zusammenhänge also nicht falsifiziert werden konnte, gibt es keinen direkten Beleg für unsere Hypothese, dass Access-MC und Access-SIM unterscheidbare Konstrukte repräsentieren. Diese Hypothese ging von den verschiedenen Ansätzen zur Erfassung von Access aus, welche nahelegten, dass die beiden Skalen unterschiedliche Formen der Wissensrepräsentationen abdecken. Ein Grund für die vergleichbar hohe konvergente Validität von Access-MC und Access-SIM mit BCS könnte sein,

dass die Mehrheit der MC-Aufgaben Elemente der graphischen Benutzeroberfläche als Antwortoptionen darbietet. Wenn inhaltsgleiche SIM-Aufgaben infolgedessen nur relativ wenige computerbezogene Handlungen erfordern (z. B. ein oder zwei Maus-Klicks wie beispielsweise Item 8 „Tiff“) ist der Unterschied im Lösungsprozess zwischen MC- und SIM-Items vergleichsweise gering.

Die Limitationen der vorliegenden Studie implizieren weiterführende Forschungsfragestellungen und Studien. Eine wichtige Erweiterung des Untersuchungsdesigns würde darin bestehen, innerhalb der Person den Operationalisierungsansatz zu variieren, also einen Aufgabenteil als MC-Items und einen anderen als SIM-Aufgaben zu administrieren (mit Reihenfolgebalancierung für Aufgabenteil und Operationalisierungsansatz). Das würde die Möglichkeit bieten, direkt die Frage der Konstruktäquivalenz der mittels MC- und SIM-Skala erfassten Merkmale zu untersuchen. Zur Verbesserung der bisher niedrigen Reliabilität sollte die Anzahl der Aufgaben erhöht und insbesondere die neu entwickelten SIM-Varianten hinsichtlich ihrer psychometrischen Qualität weiter erprobt und verbessert werden. Aus inhaltlicher Sicht sind auch solchen MC-Items zu einem Vergleich heranzuziehen, die komplexere Handlungsanforderungen im Umgang mit dem Computer ansprechen (z. B. indem in den Antwortalternativen das Erstellen einer informativen Grafik auf der Grundlage einer vorgegebenen Tabelle beschrieben wird) und nicht einfach auszuwählende Elemente des graphischen Interfaces präsentieren. Für solche mehrschrittig zu lösenden Aufgaben wäre stärker zu erwarten, dass sich die Vorgabe von MC- oder SIM-Items nicht nur bzgl. der Schwierigkeit auswirkt, sondern auch hinsichtlich der erforderlichen Wissensbestände.

Allerdings würde man dabei auch schnell an die Grenzen des MC-Itemformats geraten, d. h. dass MC-Antwortalternativen im Gegensatz zu SIM-Aufgaben nicht beliebig komplex werden können. Das führt zurück zu der grundsätzlichen Frage, welche Operationalisierungsansätze für die Diagnose von ICT-Literacy geeignet sind. Sollen ICT-

Fertigkeiten gemessen werden, die auf prozeduralem Handlungswissen beruhen, bedarf es hinreichend komplexer Aufgaben mit (simulierten) Computerumgebungen, die von der Testperson bedient werden können.

Literatur

- Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist*, 51, 355–365.
- Ballantine, J. A., Larres, P. M. & Oyelere, P. (2007). Computer usage and the validity of self-assessed computer competence among first-year business students. *Computers & Education*, 49, 976–990.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.
- Byrne, B. M. & Stewart, S. M. (2006). The MACS approach to testing for multigroup invariance of a second-order structure: A walk through the process. *Structural Equation Modeling*, 13, 287–321.
- De Wit, K., Heerwegh, D. & Verhoeven, J. C. (2012). Changes in the basic ICT skills of freshmen between 2005 and 2009: who's catching up and who's still behind? *Education and Information Technologies*. 17, 205–231.
- Downing, S. M. (2006). Twelve Steps for Effective Test Development. In Downing, S. M. & Haladyna, T. M. (Eds.). *Handbook of Test Development* (pp. 3–25). Mahwah, NJ: Lawrence Erlbaum Associates.
- Educational Testing Service (2008). *iSkills – Information and Communication Technology Literacy Test*. Zugriff am 15. Juni 2009 unter <http://www.ets.org>
- Ferrari, A. (2012). *Digital Competence in practice: An analysis of frameworks*. Seville: JRC-IPTS. Zugriff am 9. September 2013 unter <http://ftp.jrc.es/EURdoc/JRC68116.pdf>
- Forero, C.G. & Maydeu-Olivares, A. (2009). Estimation of IRT graded models for rating data: Limited vs. full information methods. *Psychological Methods*, 14, 275–299.
- Goldhammer, F., Naumann, J. & Pfaff, Y. (in press). Assessing Individual differences in Basic Computer Skills: Psychometric characteristics of an interactive performance measure. *European Journal of Psychological Assessment*.

- Green, S. B. & Yang, Y. (2009). Reliability of summed item scores using structural equation modeling: An alternative to coefficient alpha. *Psychometrika*, *74*, 155–167.
- Greiff, S., Wüstenberg, S., Holt, D. V., Goldhammer, F. & Funke, J. (2013). Computer-based assessment of complex problem solving: concept, implementation, and application. *Educational Technology Research and Development*, *61*, 407–421. doi: 10.1007/s11423-013-9301-x
- Greiff, S., Wüstenberg, S., Molnar, G., Fischer, A., Funke, J. & Csapo, B. (2013). Complex Problem Solving in educational settings. Something beyond g: Concept, assessment, measurement invariance, and construct validity. *Journal of Educational Psychology*, *105*, 364–379.
- Haladyna, T. M. (2004). *Developing and validating multiple-choice test items*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Haladyna, T. M. & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice*, *23*, 17–27.
- International ICT Literacy Panel (2007). *Digital transformation: A framework for ICT literacy (A report of the International ICT Literacy Panel)*. Princeton, NJ: Educational Testing Service. Zugriff am 21. Januar 2008 unter http://www.ets.org/Media/Tests/Information_and_Communication_Technology_Literacy/ictreport.pdf
- Katz, I. R. & Macklin, A. S. (2007). Information and communication technology (ICT) literacy: Integration and assessment in higher education. *Journal of Systemics, Cybernetics and Informatics*, *5*, 50–55.
- Kenny, D. (2012). *Measuring Model Fit*. Zugriff am 18. Oktober 2012 unter <http://davidakenny.net/cm/fit.htm>
- Kim, J. H. & Jung, S. Y. & Lee, W. G. (2008). Design of contents for ICT literacy in-service training of teachers in Korea. *Computers & Education*, *51*, 1683–1706.

- KMK (Kultusministerkonferenz) (2012). *Medienbildung in der Schule. Beschluss der Kultusministerkonferenz vom 8. März 2012*. Bonn: KMK.
- Kröhne, U. & Martens, T. (2011). Computer-based competence tests in the National Educational Panel Study: The challenge of mode effects. *Zeitschrift für Erziehungswissenschaft, Special Issue 14*, 169–186.
- Kuhlemeier, H. & Hemker, B. (2007). The impact of computer use at home on students' Internet skills. *Computers & Education*, 49, 460–480.
- Lennon, M., Kirsch, I., Von Davier, M., Wagner, M. & Yamamoto, K. (2003). *Feasibility Study for the ICT Literacy Assessment – Report to Network A*. Zugriff am 15. Juni 2009 unter <http://www.oecd.org/dataoecd/35/13/33699866.pdf>
- Markauskaite, L. (2007). Exploring the structure of trainee teachers' ICT literacy: The main components of, and relationships between, general cognitive and technical capabilities. *Educational Technology Research and Development*, 55, 547–572.
- Muthén, B. O. & Asparouhov, T. (2002). Latent variable analysis with categorical outcomes: Multiple-Group and growth modeling In Mplus. Mplus Web Notes: No. 4. Zugriff am 18. Oktober 2012 unter <http://www.statmodel.com/download/webnotes/CatMGLong.pdf>
- Muthén, L. K. & Muthén, B. O. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Muthén, L. K. & Muthén, B. O. (2010). *IRT in Mplus*. Zugriff am 24. April 2013 unter <http://www.statmodel.com/download/MplusIRT2.pdf>
- OECD (2009a). *PIAAC Literacy: A Conceptual Framework*. OECD Education Working Paper No. 34. Paris: OECD.
- OECD (2009b). *PIAAC Problem Solving in Technology-Rich Environments: A Conceptual Framework*. OECD Education Working Paper No. 36. Paris: OECD.
- OECD. (2011). *PISA 2009 results Vol. VI. Students on line: Reading and using digital information*. Paris: OECD.

- Pohl, S. & Carstensen, C. H. (in press). Scaling of competence tests in the National Educational Panel Study – Many questions, some answers, and further challenges. *Journal for Educational Research Online*.
- Potosky, D. (2007). The Internet knowledge (iKnow) measure. *Computers in Human Behavior*, 23, 2670–2777.
- Poynton, T. A. (2005). Computer literacy across the lifespan: A review with implications for educators. *Computers in Human Behavior*, 21, 861–872.
- R Development Core Team (2011). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Zugriff am 24. April 2013 unter <http://www.R-project.org/>
- Richter, T., Naumann, J. & Horz, H. (2010). Eine revidierte Fassung des Inventars zur Computerbildung (INCOBI-R) [A revised version of the Computer Literacy Inventory]. *Zeitschrift für Pädagogische Psychologie/German Journal of Educational Psychology*, 24, 23–37.
- Rizopoulos, D. (2006). ltm: An R package for Latent Variable Modeling and Item Response Theory Analyses, *Journal of Statistical Software*, 17, 1–25. Zugriff am 24. April 2013 unter <http://www.jstatsoft.org/v17/i05/>
- Rodriguez, M. C. (2003) Construct equivalence of multiple-choice and constructed-response items: *A random effects synthesis of correlations*. *Journal of Educational Measurement*, 40, 163–184.
- Rölke, H. (2012). The item builder: A graphical authoring system for complex item development. In *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education* (pp. 344–353). Chesapeake, VA: AACE.
- Schermelleh-Engel, K., Moosbrugger, H. & Müller, H. (2003). Evaluating the fit of structural equation models: Test of significance and descriptive goodness-of-fit measures. *Methods*

of Psychological Research – Online, 8, 23–74. Zugriff am 24. April 2013. Verfügbar unter <http://www.dgps.de/fachgruppen/methoden/mpr-online/>

Senkbeil, M., Ihme, J. M. & Wittwer, J. (in press). The test of Technological and Information Literacy (TILT) in the National Educational Panel Study: Development, empirical testing, and evidence for validity. *Journal for Educational Research Online*.

Sireci, S. G. & Zenisky, A. L. (2006). Innovative item formats in computer-based testing: In pursuit of improved construct representation. In S. M. Downing & T. S. Haladyna (Eds.), *Handbook of Test Development* (pp. 329–348). Mahwah, NJ: Lawrence Erlbaum associates.

Süß, H.-M. (1996). *Intelligenz, Wissen und Problemlösen. Kognitive Voraussetzungen für erfolgreiches Handeln bei computersimulierten Problemen. Lehr- und Forschungstexte Psychologie*. Göttingen: Hogrefe.

Vandenberg, R. J. & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3, 4–69.

Tabelle 1. Charakteristiken der Access-MC-Items und Access-SIM-Items gemäß klassischer Testtheorie: Itemschwierigkeit als Anteil korrekter Antworten, p , punktbiseriale Korrelation des Itemergebnisses mit dem Testwert, r_{it} .

Item ID	Aufgabe	Access-MC		Access-SIM	
		P	r_{it}	p	r_{it}
1	Move sentence	.69	.46	.42	.39
2	Thesaurus	.36	.37	.03	.23
3	Find email	.80	.27	.51	.17
4	Slide design	.34	.19	.21	.28
5	Blind cc	.27	.40	.07	.15
6	Search images 1	.48	.62	.29	.41
7	Search images 2	.27	.47	.55	.58
8	Tiff	.63	.49	.56	.58
9	Live ticker	.72	.40	.40	.50
10	Animals	.67	.35	.04	.24
11	Sort slides	.71	.38	.32	.43

Abbildungsverzeichnisse

Abbildung 1. Ein Beispiel der (a) Access-MC-Skala und die entsprechende Aufgabe in der (b) Access-SIM-Skala (Item 9: „Live Ticker“). Die Aufgabe verlangt, die im Live-Ticker präsentierte Information zu aktualisieren.


Abbildung 2. Modellbasierte Itemcharakteristiken der Access-MC-Skala und entsprechend der Access-SIM-Skala aus der Zweigruppen-CFA mit Annahme konfiguraler Invarianz: (a) Schwellen und (b) standardisierte Ladungen. Die eingezeichneten Abweichungen entsprechen ± 1 Standardfehler.

(a)

Instruktion

Du siehst hier eine Internetseite des DFB-Pokals. Die Informationen, die du hier siehst, sind aber veraltet. Wo klickst du hin, um aktuelle Informationen zu erhalten?

Arbeite möglichst **SCHNELL** und **GENAU**.



ER62


-
-
-
-

(b)

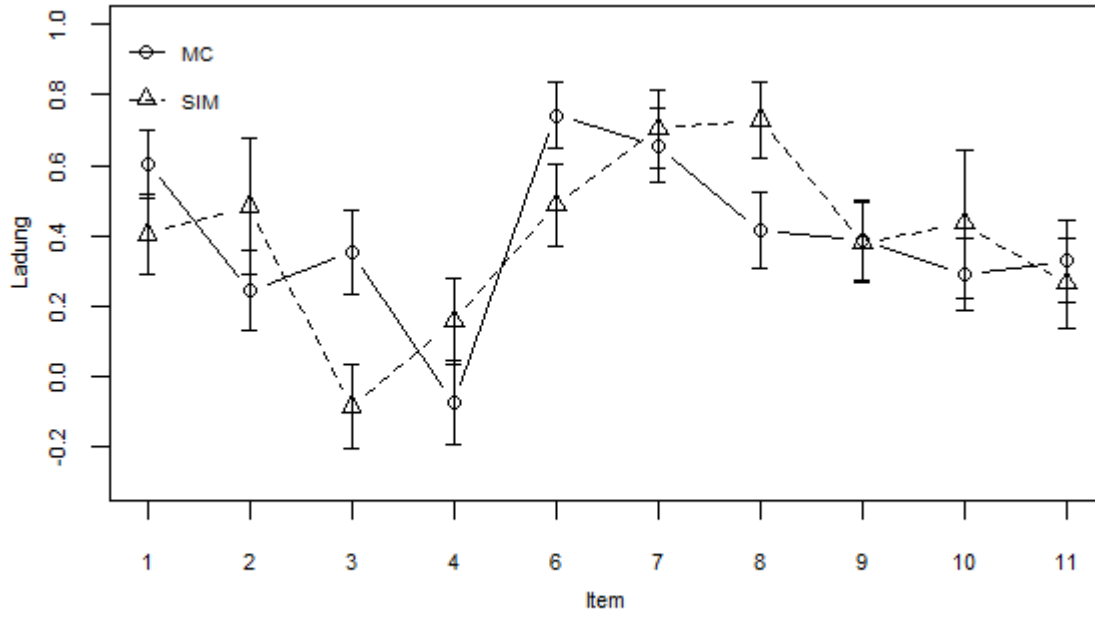
Instruktion

Du siehst hier eine Internetseite des DFB-Pokals. Die Informationen, die du hier siehst, sind aber veraltet. Wo klickst du hin, um aktuelle Informationen zu erhalten?

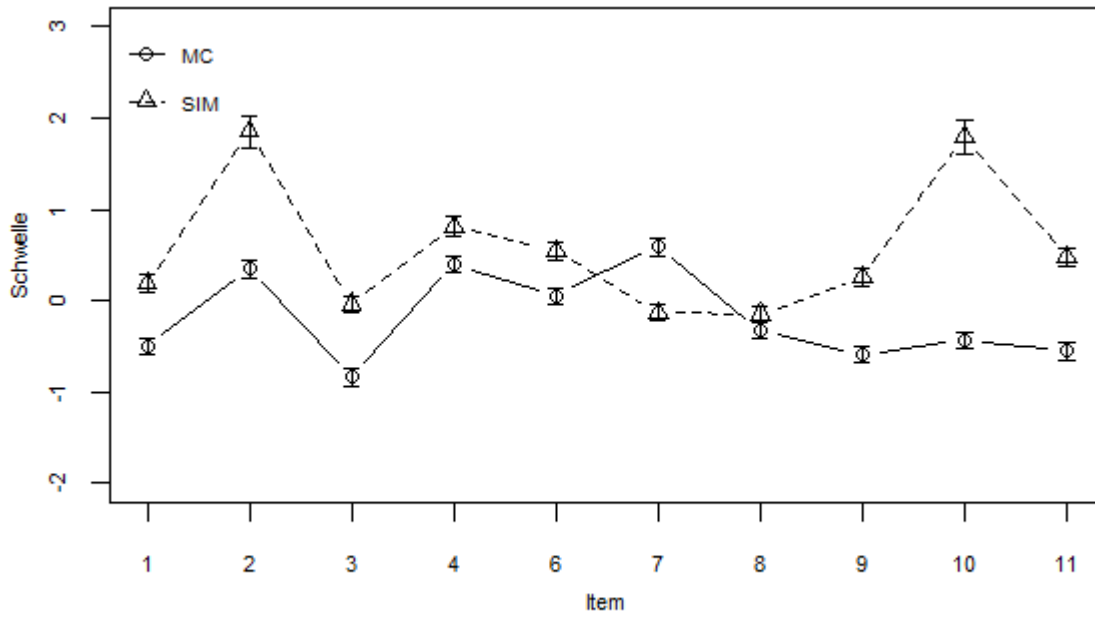
Arbeite möglichst **SCHNELL** und **GENAU**.



(a)



(b)



Die Autoren danken zwei anonymen Gutachtern dieser Zeitschrift für ihre sehr hilfreichen Kommentare zu einer früheren Fassung dieses Papiers.