

Digitalisierung

Bibliografie:

Stéphanie Berger und Urs Moser:
Adaptives Lernen und Testen.

journal für lehrerInnenbildung, 20 (1), 42-52.

https://doi.org/10.35468/jlb-01-2020_03

Gesamtausgabe online unter:

<http://www.jlb-journallehrerinnenbildung.net>

<https://doi.org/10.35468/jlb-01-2020>

ISSN 2629-4982

journal für lehrerInnenbildung
j l b
no. 1
2020

03

*Stéphanie Berger
und Urs Moser*

Digitalisierung im Unterricht

Zieht man für die Beurteilung der Wirksamkeit von Schule und Unterricht die Metaanalyse von John Hattie bei, dann deutet wenig auf die Notwendigkeit einer Digitalisierung im Unterricht hin. Zu den wirkungsvollsten Einflussgrößen auf den schulischen Lernerfolg gehören unter anderem die Selbsteinschätzung des eigenen Leistungsniveaus oder die formative Evaluation des Unterrichts. Faktoren, welche auf die seit geraumer Zeit angekündigte digitale Bildungsrevolution hinweisen könnten, wie etwa der Einsatz des Computers im Unterricht oder webbasiertes Lernen, folgen im hinteren Teil der Rangliste der Einflussgrößen (Hattie, 2009). Die in der Metaanalyse zitierten Studien sind allerdings etwas in die Jahre gekommen und werden dem aktuellen Digitalisierungstrend kaum gerecht. Auch wenn die digitale Revolution die Schule in ihrer Ausrichtung nicht grundsätzlich verändern wird – das professionelle Handeln von Lehrerinnen und Lehrern und die Interaktion mit den Schülerinnen und Schülern lässt sich durch Technik nicht ersetzen (Lipowsky, 2006) –, sind digitale Technologien zur Unterstützung des Lehr-Lern-Prozesses in Zukunft kaum mehr wegzudenken.

Dass die Nutzung digitaler Technologien für wirksamen Unterricht durchaus von Bedeutung ist, zeigt Hattie mit seiner aus den Erkenntnissen der Metaanalyse abgeleiteten Unterrichtstheorie, die er als *Visible Learning* bezeichnet (Hattie, 2014). Unterricht ist dann erfolgreich, wenn Lehrpersonen das Lernen mit den Augen der Schülerinnen und Schüler sehen und diese den Lernprozess aus der Perspektive der Lehrperson betrachten. Lehrpersonen nutzen empirische Belege, um ein fachbezogenes Feedback zum Lernprozess zu geben, sodass Schülerinnen und Schüler entsprechend dem kumulativen Aufbau von Wissen und Können Lernfortschritte erzielen und diese auch selbst feststellen können. Qualifiziertes Feedback ist eine Voraussetzung, damit Lernen sichtbar gemacht werden kann. Dieser Prozess kann durch digitale Technologie unterstützt werden, weil sich das Feedback erstens durch die Anwendung von adaptiven Algorithmen personalisieren lässt, zweitens auf anstehende Lernschritte hinweist und drittens sich mit Bezug zum kumulativen Aufbau von Wissen und Können empirisch verorten lässt.

John Hattie hat deshalb bereits vor längerer Zeit das *Electronic Assessment Tool for Teaching and Learning* (e-asTTle) zur Beurteilung

von fachlichen Leistungen und Lernfortschritten entwickelt, das von Lehrpersonen in Neuseeland eingesetzt wird (Brown, 2013). Digitale Feedbacksysteme führen aber nicht nur zu personalisiertem und kompetenzorientiertem Feedback, sondern auch zu einer großen Menge an ökologisch validen Daten, die für die Forschung im Bereich der Pädagogischen Psychologie von größtem Interesse sind (Tomasik & Moser, 2019). Die Digitalisierung eröffnet für die Unterrichtspraxis wie für die Wissenschaft neue Möglichkeiten, die das Unterrichten erleichtern und die Modellierung des kumulativen Aufbaus des Lernens in einer noch nie dagewesenen Feinheit erlauben.

Personalisierung des Lernens

Das große Potenzial der Digitalisierung liegt darin, dass Lernergebnisse, Lernfortschritte und der kumulative Aufbau von Wissen und Können aufgrund von mathematischen Modellen empirisch gesichert und sichtbar gemacht werden können. Während früher für die unabhängige Erfassung des Lernstands Papiertests eingesetzt wurden, können heutzutage Aufgaben in einer digitalen Itembank abgelegt werden, die so umfangreich ist, dass sich das Wissen und Können in unterschiedlichen Domänen und auf verschiedenen Klassenstufen in Form von Aufgaben über eine sehr große Zeitspanne, beispielsweise über die gesamte obligatorische Schulzeit hinweg, abbilden lassen. Die Aufgaben werden mit didaktisch relevanten Metadaten zum Inhalt, zu den kognitiven Anforderungen und zur Schwierigkeit verknüpft. Diese Metadaten dienen dazu, den Schülerinnen und Schülern Aufgaben zuzuweisen, die ihren Fähigkeiten entsprechen, und Rückmeldungen zum Lernstand und zum Lernfortschritt zu generieren, die sich mit Bezug zum kumulativen Aufbau von Wissen und Können interpretieren lassen.

Die Lehrperson wird mit der Nutzung einer digitalen Itembank darin unterstützt, ein personalisiertes und qualifiziertes Feedback anhand von objektiven Informationen über den Lernstand und den Lernfortschritt zu geben. Objektive Informationen bzw. Daten über den Lernstand dienen als Grundlage für die Definition von Lernzielen und für die Festlegung von Unterrichtsmaßnahmen (*Data-Based Decision Making*; Schildkamp, Lai & Earl, 2013). Je genauer der Lernstand erfasst werden kann, desto besser können auch die Lernziele und der Unter-

richt personalisiert werden. Weitere Daten zu späteren Messzeitpunkten erlauben es, den Lernfortschritt zu überprüfen und somit im Sinne von Hattie das Lernen sichtbar zu machen. Das Feedback liefert jedem Schüler/jeder Schülerin wertvolle Informationen über die Effektivität seines/ihres Lernens und führt gleichzeitig die nächsten inhaltlichen Schritte vor Augen.

Die Auswahl passender Aufgaben kann in einer Itembank von Lehrpersonen, von den Schülerinnen und Schülern oder von einem Algorithmus gesteuert werden (z. B. Tomasik, Berger & Moser, 2018). Auf der Basis der Metadaten können jene Aufgaben identifiziert werden, die bezüglich Inhalt und Schwierigkeit zuverlässige Antworten auf Fragen zum Lehr-Lern-Prozess liefern. Adaptive Algorithmen haben zusätzlich den Vorteil, dass sie aufgrund des Lösungsverhaltens die Fähigkeiten der Schülerinnen und Schüler fortwährend berechnen und jeweils eine nächste Aufgabe zur Bearbeitung auswählen können, die optimal auf die aktuellen Fähigkeiten abgestimmt ist (van der Linden & Glas, 2010; Wainer, 2000).

Werden anstelle von einzelnen Aufgaben gleich ganze Aufgabengruppen (Segmente), die einen homogenen Schwierigkeitsgrad aufweisen, adaptiv ausgewählt, so spricht man von *Multistage Testing* (Yan, Von Davier & Lewis, 2016). Multistage Testing ist in der Adaptivität zwar etwas weniger flexibel und daher auch etwas weniger effizient als die adaptive Auswahl jeder Aufgabe. Seine Vorteile sind allerdings, dass Schülerinnen und Schüler Aufgaben innerhalb einer Aufgabengruppe überarbeiten und dass inhaltlich zusammengehörende Aufgaben (z. B. mehrere Aufgaben zu einem Lesetext) gemeinsam präsentiert werden können. Sofern die Itembank eine genügend große Anzahl an Aufgaben enthält, hat jede Schülerin und jeder Schüler die Möglichkeit, individuell Aufgaben zu lösen, die ihrem oder seinem Lernstand entsprechen. Zudem können die Ergebnisse mit Bezug zum kumulativen Aufbau des Lehr-Lern-Prozesses auf einer einheitlichen, vertikalen (d. h. klassenstufenübergreifenden) Skala zurückgemeldet werden (Young, 2006). Eine optimale Passung zwischen Aufgabenschwierigkeit und Fähigkeit wirkt sich zudem positiv auf die Motivation der Schülerinnen und Schüler aus, da sie durch die ausgewählten Aufgaben gefordert, aber nicht überfordert werden (Assenburg & Frey, 2013).

Die digitale Itembank hat somit ein großes Potenzial für die diagnostische Erfassung von individuellen Stärken und Schwächen (*Diagnostic Testing*; van der Kleij, Vermeulen, Schildkamp & Eggen, 2015).

Denn nicht nur die Informationen zum „richtigen“ oder „falschen“ Lösen einer Aufgabe liefern Hinweise zu den Fähigkeiten. Vielmehr können falsche Antworten beziehungsweise die Wahl falscher Antwortoptionen bei geschlossenen Frageformaten weiter analysiert werden, um detailliertere Informationen über die Fähigkeiten, die kognitiven Prozesse und die Problemlösestrategien zu erhalten (de la Torre, 2009). Dies kann beispielhaft an der Mathematikaufgabe

$$2\frac{4}{7} - \frac{7}{12}$$

illustriert werden. Auf der Grundlage von kognitiven Theorien und typischen Schüler*innenantworten können für diese Aufgabe fünf Attribute (Skills) definiert werden (Mislevy, 1995):

1. Ein Ganzes ausleihen
2. Einfache Subtraktion von Brüchen
3. Reduzieren/Vereinfachen
4. Ganze Zahlen separieren
5. Ganze Zahlen in Brüche umwandeln

Um die Aufgabe richtig zu lösen, müssen alle fünf Attribute erfolgreich umgesetzt werden. Falsche Antworten können Hinweise darauf geben, welche dieser Attribute bereits beherrscht werden und welche nicht. Je mehr Aufgaben mit diesen Attributen einer Schülerin/einem Schüler vorgelegt werden, desto genauere Fähigkeitsprofile können ausgewiesen werden.

Psychometrische Grundlagen

Die Metadaten von Aufgaben in einer digitalen Itembank erlauben es, komplexe psychometrische Modelle einzusetzen, die wiederum eine Voraussetzung dafür sind, dass Ergebnisse über verschiedene Schülerinnen und Schüler und Zeitpunkte hinweg verglichen werden können, auch wenn die Schülerinnen und Schüler unterschiedliche Aufgaben aus der Itembank lösen. Die Grundlage für die intelligente Nutzung einer Itembank und den Einsatz von adaptiven Algorithmen bildet meist die *Item-Response-Theory* (IRT), die im europäischen Sprachraum auch als *probabilistische Testtheorie* bezeichnet wird (Rost, 2004). Probabilistische Modelle nehmen eine stochastische Beziehung zwischen dem *Antwortverhalten* einer Person, der *Fähigkeit* der Person und der

Schwierigkeit der Aufgabe an. Diese Beziehung wird durch eine mathematische, nicht-lineare logistische Funktion dargestellt, wobei die Fähigkeit der Personen und die Schwierigkeit der Aufgaben auf der gleichen Skala abgebildet werden und die Erfolgswahrscheinlichkeit beim Lösen einer Aufgabe im einfachsten Modell – dem *Rasch-Modell* – vom Personenparameter, der die Fähigkeit der Person repräsentiert, und vom Schwierigkeitsparameter, der die Schwierigkeit der Aufgabe abbildet, abhängt (ebd.).

Entsprechend den Parametern im Rasch-Modell erfüllen IRT-Modelle zwei Funktionen. Zum einen können die Aufgaben anhand ihrer Schwierigkeiten auf einer einheitlichen Skala empirisch verortet – d. h. kalibriert – werden. Je mehr Schülerinnen und Schüler eine Aufgabe bearbeiten, desto genauer kann ihre Schwierigkeit bestimmt werden. Die Anzahl der Aufgaben, die eine Skala und damit eine bestimmte Fähigkeit repräsentieren, ist theoretisch unbegrenzt. Somit können jederzeit neue Aufgaben zu einer bestehenden Itembank und dem zugrundeliegenden IRT-Modell hinzugefügt werden. Die Schwierigkeit der neuen Aufgaben wird aufgrund des Erfolgs beim Lösen von neuen und bereits kalibrierten Aufgaben in Abhängigkeit der Fähigkeit der Schülerinnen und Schüler berechnet. Die Berechnung der Schwierigkeit erfolgt weitgehend online während die Schülerinnen und Schüler die Aufgaben lösen (*Online-Kalibrierung*, z. B. Verschoor, Berger, Moser & Kleintjes, 2019).

Zum andern dienen IRT-Modelle und entsprechend kalibrierte Aufgaben in der Itembank als Basis für die Schätzung der Fähigkeiten der Schülerinnen und Schüler. Je näher die Aufgabenschwierigkeit bei der tatsächlichen Fähigkeit liegt und je mehr Aufgaben bearbeitet werden, desto genauer kann die Fähigkeit bestimmt werden (Rost, 2004). Durch Algorithmen für adaptives Testen kann die Aufgabenauswahl und somit die Messgenauigkeit optimiert werden (van der Linden & Glas, 2010). Der Algorithmus berechnet fortlaufend die Fähigkeit der Schülerinnen und Schüler und wählt auf dieser Basis Aufgaben mit einer Schwierigkeit aus, die der aktuellen Fähigkeit entspricht. Dieser approximative Prozess dauert so lange, bis es bei der Berechnung der Fähigkeit zu nahezu keinen Schwankungen mehr kommt und diese zuverlässig ausgewiesen werden kann (Kubinger, 2003).

Komplexere IRT-Modelle erlauben den Einbezug weiterer Parameter (z. B. Trennschärfe, Ratewahrscheinlichkeit oder Antwortzeit) oder die Abbildung multidimensionaler Konstrukte (van der Linden,

2016). Für detailliertes diagnostisches Feedback werden sogenannte *Cognitive Diagnosis Models* (de la Torre, 2009) eingesetzt, die einen Bezug zwischen verschiedenen kognitiven Fähigkeiten und den Attributen einer Aufgabe herstellen (Minchen & de la Torre, 2018). Die Schülerinnen und Schüler erhalten folglich nicht eine Angabe zum Lernstand in einem Fähigkeitsbereich, sondern ein diagnostisches Profil kognitiver Stärken und Schwächen. Die Entwicklung der zugrundeliegenden Aufgaben und deren Metadaten ist allerdings relativ aufwändig, da alle möglichen Antwortoptionen jeder Aufgabe hinsichtlich der kognitiven Fähigkeiten kategorisiert werden müssen (de la Torre, 2009).

Lernfortschritte sichtbar machen

Eine Itembank mit kalibrierten Aufgaben, die auf der Grundlage der Item-Response-Theorie (insbesondere auf dem einfachen Rasch-Modell) basiert, verfügt über mehrere Eigenschaften, die für die Feedback-Unterstützung des Lehr-Lern-Prozesses besonders wertvoll sind. Eine dieser Eigenschaften betrifft die Unabhängigkeit der Berechnung der Fähigkeit einer Person von der Aufgabenauswahl aus einem hypothetischen Aufgaben-Universum, was als spezifische Objektivität bezeichnet wird. Dahinter steht die Überlegung, dass jeder Test nur eine sehr begrenzte Anzahl an Aufgaben umfassen kann, das Ergebnis aber nicht nur etwas über die Fähigkeit zum Lösen dieser Aufgaben aussagen soll, sondern generell über die Fähigkeit, die zum Lösen von Aufgaben dieser Art vorausgesetzt wird (Rost, 2004). Die spezifische Objektivität ist eine Voraussetzung für das adaptive Testen, denn die Auswahl der Aufgaben berücksichtigt den Leistungsstand einer Person und unterscheidet sich deshalb von Person zu Person. Sämtliche Aufgaben der Itembank, die dem Rasch-Modell entsprechen, sind jedoch Indikatoren der zu messenden Fähigkeit im Sinne einer latenten Variable. Daher ist die gemessene Fähigkeit nicht von der Aufgabenauswahl abhängig. Der Person wird demnach unabhängig von den gelösten Aufgaben ein Feedback zur Fähigkeit gegeben. Zugleich heißt dies, dass ein Feedback an mehrere Personen nicht davon abhängt, welche Aufgaben sie im Einzelnen gelöst haben. Ein direkter Vergleich mehrerer Feedbacks wird trotz Bearbeitung unterschiedlicher Aufgaben möglich.

Aufgrund der spezifischen Objektivität ist es zudem möglich, Lernfortschritte entsprechend dem kumulativen Aufbau von Wissen und Können sichtbar zu machen. Weil die Aufgaben Indikatoren für eine latente Variable sind und aufgrund ihrer Schwierigkeit auf einem Kontinuum lokalisiert sind, kann die Itembank in regelmäßigen Abständen formativ genutzt werden. Die Fähigkeit von Schülerinnen und Schülern wird mit der Zeit steigen, der adaptive Algorithmus wird entsprechend schwierigere Aufgaben auswählen, das Feedback erfolgt jedoch zur gleichen Fähigkeit (latente Variable). Das bedeutet, dass die Metrik und in diesem Sinne der Beurteilungsmaßstab konstant bleiben, was eine Voraussetzung für den Nachweis von Lernfortschritten ist. Mehrere Feedbacks lassen sich somit direkt miteinander vergleichen. Ein weiterer Vorteil dieses Modells ergibt sich daraus, dass sowohl das Ergebnis des Schülers bzw. der Schülerin als auch die Schwierigkeit der Aufgaben auf der gleichen Skala abgebildet werden. Die Skala ist so konstruiert, dass bei einer Entsprechung von Schwierigkeit und Fähigkeit die Erfolgswahrscheinlichkeit 50 Prozent ($p = 0,5$) beträgt (Rost, 2004). Die Beziehung zwischen Lösungswahrscheinlichkeit, Fähigkeit des Schülers/der Schülerin und Schwierigkeit der Aufgabe ist für das Feedback von entscheidender Bedeutung. Dadurch lässt sich das Feedback mit Bezug zu einem Kompetenzmodell interpretieren, das den Aufbau von Wissen und Können zusammenfasst und illustriert (Fleischer, Koeppen, Kenk, Klieme & Leutner, 2013). Die einzelnen Aufgaben stellen das Bindeglied zwischen der Fähigkeit des Schülers/der Schülerin und dem kumulativen Aufbau des Wissens und Könnens im Kompetenzmodell dar, was in der zugrundeliegenden Theorie nicht zuletzt auch mit dem Namen *Item-Response* angedeutet wird.

In diesem Kontext wird Feedback konzipiert als eine Rückmeldung zum Stand des Lernens, welche den Schülerinnen und Schülern zeigt, was sie bereits beherrschen und was als Nächstes kommt. Feedback hat ohne Bezug zu einem Kompetenzmodell keine Wirkung, sondern muss als Teil des Lehr-Lern-Prozesses verstanden werden. Es orientiert sich am Inhalt, mit dem sich Schülerinnen und Schüler beschäftigen, und liefert eine Grundlage für das Planen des Lernens. Zugleich stärkt ein solches Feedback die Selbsteinschätzung des eigenen Leistungsstandes, die sich als besonders wirksamer Einflussfaktor für den Lernerfolg erweist (Hattie & Timperley, 2007). Digitale Technologien sind dann wirksam, wenn sie die Lehrperson in ihrer aktiven Rolle unterstützen und effektives Unterrichtshandeln optimieren.

Ausblick

Lernen lässt sich aufgrund von intelligenten Algorithmen adaptiv und kompetenzorientiert organisieren. Das Potenzial von digitalen Technologien erscheint verführerisch. Ohne grundlegende methodische Kenntnisse, wie die in Form von Zahlen oder Deskriptoren sichtbaren Lernfortschritte adäquat interpretiert werden (*Assessment Literacy*), und ohne die Diskussion über mögliche unerwünschte Auswirkungen auf die Lehr-Lern-Kultur (*Teaching/Learning to the test*) droht allerdings die Gefahr, dass der Einsatz digitaler Technologien ein wirkungsloses Versprechen bleibt (McElvany & Rjosk, 2013). Digitale Technologien können zwar einen substanziellen Beitrag für erfolgreichen Unterricht leisten, jedoch nur dann, wenn die Lehrpersonen entsprechend ausgebildet werden.

Vielversprechend ist das Potenzial von digitalen Technologien auch für die Gewinnung neuer Erkenntnisse zum Lehr-Lern-Prozess. Die Nutzung einer Itembank zur Personalisierung des Lernens und zum Sichtbarmachen von Lernfortschritten geht zwangsläufig mit der Generierung einer großen Anzahl ökologisch valider Daten einher und eröffnet der Forschung in der Pädagogischen Psychologie somit ganz neue Möglichkeiten. Auf Grundlage dieser *Big Data* können Lernverläufe über die gesamte Schulzeit hinweg und aufgrund einer domänenübergreifenden Perspektive modelliert werden (Sin & Muthu, 2015). Diese Modelle bilden nicht nur eine Basis für die Messung und die Vorhersage der Fähigkeit von Schülerinnen und Schülern, sondern auch für das fachdidaktische Vorgehen von Lehrpersonen im Unterricht.

Literatur

- Asseburg, R. & Frey, A. (2013). Too hard, too easy, or just right? The relationship between effort or boredom and ability-difficulty fit. *Psychological Test and Assessment Modeling*, 55 (1), 92-104.
- Brown, G. T. L. (2013). AsTTle – A national testing system for formative assessment: How the national testing policy ended up helping schools and teachers. In S. Kushner, M. Lei & M. Lai (Eds.), *Advances in Program Evaluation: Volume 14. A national developmental and negotiated approach to school self-evaluation* (pp. 39-56). Bradford: Emerald Group Publishing Limited. doi:10.1108/S1474-7863(2013)0000014003

- de la Torre, J. (2009). A cognitive diagnosis model for cognitively based multiple-choice options. *Applied Psychological Measurement*, 33 (3), 163-183. doi:10.1177/0146621608320523
- Fleischer, J., Koeppen, K., Kenk, M., Klieme, E. & Leutner, D. (2013). Kompetenzmodellierung: Struktur, Konzepte und Forschungszugänge des DFG-Schwerpunktprogramms. *Zeitschrift für Erziehungswissenschaft*, 16 (1), 5-22. doi:10.1007/s11618-013-0379-z
- Hattie, J. (2009). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. London/New York: Routledge.
- Hattie, J. (2014). *Lernen sichtbar machen für Lehrpersonen*. Überarbeitete deutschsprachige Ausgabe von „Visible Learning for Teachers“ besorgt von Wolfgang Beywl und Klaus Zierer. Hohengehren: Schneider Verlag.
- Hattie, J. & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77 (1), 81-112.
- Kubinger, K. D. (2003). Adaptives Testen. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 1-9). Weinheim & Basel: Beltz.
- Lipowsky, F. (2006). Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler. *Zeitschrift für Pädagogik*, 51, 47-70.
- McElvany, N. & Rjosk, C. (2013). Wann kann Kompetenzdiagnostik negative Auswirkungen haben? *Zeitschrift für Erziehungswissenschaft*, 16 (1), 65-70.
- München, N. & de la Torre, J. (2018). A general cognitive diagnosis model for continuous-response data. *Measurement: Interdisciplinary Research and Perspectives*, 16 (1), 30-44. doi:10.1080/15366367.2018.1436817
- Mislevy, R. (1995). Probability-based inference in cognitive diagnosis. In P. D. Nichols, S. F. Chipman & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 43-71). Hillsdale, NJ: Erlbaum.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion* (2. Aufl.). Bern: Huber.
- Schildkamp, K., Lai, M. K. & Earl, L. (2013). *Data-based decision making in education: Challenges and opportunities*. Dordrecht: Springer.
- Sin, K. & Muthu, L. (2015). Application of big data in education data mining and learning analytics: A literature review. *ICTACT Journal on Soft Computing*, 5 (4), 1035-1049.
- Tomasik, M. J., Berger, S. & Moser, U. (2018). On the development of a computer-based tool for formative student assessment: Epistemological, methodological, and practical issues. *Frontiers in Psychology*, 9 (2245), 1-17.
- Tomasik, M. J. & Moser, U. (2019). Den digitalen Wandel in der Pädagogischen Psychologie gestalten. *Psychologische Rundschau*, 70 (2), 133-134.
- van der Kleij, F. M., Vermeulen, J. A., Schildkamp, K. & Eggen, T. J. H. M. (2015). Integrating data-based decision making, Assessment for Learning and diagnostic testing in formative assessment. *Assessment in Education: Principles, Policy & Practice*, 22 (3), 324-343. doi:10.1080/0969594X.2014.999024
- van der Linden, W. J. (Ed.). (2016). *Handbook of item response theory. Volume one: Models*. Boca Raton: CRC Press.
- van der Linden, W. J. & Glas, C. A. W. (Eds.). (2010). *Elements of adaptive testing*. New York, NY: Springer.
- Verschoor, A. V., Berger, S., Moser, U. & Kleintjes, F. (2019). On-the-fly calibration in computerized adaptive testing. In B. Veldkamp & C. Sluijter (Eds.), *Theoretical and practical advances in computer-based educational measurement* (pp. 307-323). Cham: Springer Open.

- Wainer, H. (Eds.). (2000). *Computerized adaptive testing: A primer* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Yan, D., Von Davier, A. A. & Lewis, C. (Eds.). (2016). *Computerized multistage testing: Theory and applications*. Boca Raton: CRC Press.
- Young, M. J. (2006). Vertical scales. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 469-485). Mahwah, NJ: Lawrence Erlbaum Associates.



Stéphanie Berger, Dr. phil.,
Institut für Bildungsevaluation,
assoz. Institut der Universität Zürich.
Arbeitsschwerpunkte:
Messung schulspezifischer Kompetenzen,
Computerized Adaptive Testing und Multistage Testing,
Vertikale Skalierung auf der Basis
der Item Response Theory

Stephanie.Berger@ibe.uzh.ch



Urs Moser, Prof. Dr., Titularprof. für das Gebiet Pädagogik
mit Schwerpunkt empirische Bildungsforschung,
Institut für Bildungsevaluation,
assoz. Institut der Universität Zürich.
Arbeitsschwerpunkte:
Evaluationsforschung im Bildungswesen,
Methoden der Leistungsmessung,
Schnittstellen des Bildungssystems

Urs.Moser@ibe.uzh.ch