



Digitalisierung der pädagogischen Diagnostik:

von Evolution zu Revolution

Antoine Fischbach, Samuel Greiff, Pedro Cardoso-Leite & Vincent Koenig

Einleitung

Bei dem Stichwort „digitales Klassenzimmer“ gehen die Überlegungen traditionell erst einmal in Richtung pädagogisch-didaktische Anwendungsszenarien. Wo gelernt und gelehrt wird, muss jedoch auch evaluiert werden. Wer Lehr-Lern-Prozesse erfolgreich steuern will und eine optimale Passung zwischen Unterricht und Schülerschaft erreichen möchte, muss notgedrungen regelmäßig feststellen, wie es um die Kompetenzen dieser Schülerschaft steht. Dass Zertifikation Evaluation voraussetzt, versteht sich ebenfalls von selbst. Insbesondere diese pädagogische Diagnostik – egal ob formativ, summativ oder zertifikativ – kann enorm von einem zielgerichteten Einsatz neuer Medien und Technologien profitieren, und zwar in puncto Effizienz (z. B. automatische Auswertung, vereinfachte Logistik, höherer Informationsgehalt), Effektivität (z. B. verbesserte Psychometrie und Abbildung von höheren Kompetenzen, adaptive Aufgabenauswahl, Echtzeitfeedback, Verquickung von Lernen und Diagnostik) und Fairness (z. B. erhöhte Standardisierung der Testsituation und Auswertung, inklusive Benutzeroberflächen, sprachfreie und multilinguale Assessments).

Opportunitäten und Herausforderungen

Bereits Ende der 80er Jahre publizierte *Educational Testing Service* (ETS), Princeton, weltweiter Marktführer im Bereich der sogenannten „Testindustrie“, ein Forschungspapier (Bunderson et al., 1988), in dem die

Autoren die neuen Möglichkeiten computergestützter pädagogischer Diagnostik diskutierten und zukünftige Entwicklungen prophezeiten. Das Spektrum der umrissten Anwendungsszenarien reichte vom Minimalszenario der simplen Substitution – also einer einfachen, quasi Eins-zu-eins-Übertragung von papierbasierten Assessments auf ein digitales Medium – über Transposition mit digitalen Erweiterungen und Verbesserungen zum Maximalszenario einer regelrechten Transformation (der perfekten Verquickung von Lernen und Diagnostik in Form von „intelligenten“ integrierten Test-und-Lern-Systemen).

Seit der Publikation des ETS-Berichts sind nunmehr über drei Jahrzehnte vergangen. Sieht man einmal davon ab, dass sich ohnehin alles wesentlich langsamer entwickelt hat als damals prognostiziert und bahnbrechende technologische Innovationen wie die globale Vernetzung und die universelle Verfügbarkeit mobiler Rechner (Tablets, Smartphones, Smartwatches) nicht adäquat einkalkuliert wurden, so kann man eigentlich nur schlussfolgern, dass die damaligen Kernaussagen und -annahmen, vor allem aber auch die damit assoziierten Herausforderungen heute aktueller sind denn je (s. z. B. Quellmalz & Pellegrino, 2009; s. a. Greiff et al., 2014; Redecker & Johannessen, 2013). Spätestens mit der weltweit komplett digitalisierten – und mitunter genau deswegen kontrovers diskutierten – PISA-2015 Studie (s. z. B. Robitzsch et al., 2017; s. a. Fischbach et al., 2016; SCRIPT & LUCET, 2016) ist die besagte



Transposition, zumindest was Systemevaluationen angeht, im Mainstream angekommen. Mit der Integration von interaktiven Aufgaben zum komplexen und kollaborativen Problemlösen (Fiore et al., 2018; s. a. Krieger et al. in diesem Band) geht die Transposition auch eindeutig über eine einfache Substitution im Sinne einer Hinzunahme von neuen Konstrukten, die auf eine technologiebasierte Erfassung angewiesen sind, hinaus, von einer Transformation kann aber noch keine Rede sein. Darüber hinaus schwächen viele der heute verfügbaren Systeme – sogar im simplen Substitutionsmodus, also der direkten Übertragung einer Aufgabe auf z. B. den Computer – noch in kritischen Bereichen, wie Sicherheit, Datenschutz, Benutzerfreundlichkeit und Benutzererfahrung (*user experience; UX*). Letzteres ist insofern höchst relevant, als dass es den Schlüssel zu validen und fairen Daten – Stichwort *digital divide* – darstellt, indem es die Nutzung der neuen Technologien für alle Lernenden genauso intuitiv gestaltet wie wir das von *low-tech*-Szenarien (z. B. papierbasierten) kennen (s. z. B. Rohles et al., 2019; Weinerth et al., 2014). Wir beobachten derzeit demnach vielmehr eine Evolution als eine Revolution, und auch dies vor allem in der systemischen Qualitätssicherung, denn die digitale pädagogische Diagnostik konnte das Klassenzimmer selbst bis dato nur sporadisch erreichen.

„Intelligente“ integrierte Test-und-Lern-Systeme

Die 80er-Jahre-Vision von „intelligenten“ integrierten Test-und-Lern-Systemen (Bunderson et al., 1988) bleibt auch heute noch unerreicht (s. z. B. Baker, 2019). Wir sind dem Ziel in den letzten dreißig Jahren allerdings etwas näher gekommen. Das vermeintliche Zauberbrot heißt „Log-“ oder „Prozessdaten“. Bei der Benutzung von digitalen Testsystemen hinterlassen die Nutzerinnen und Nutzer nicht nur, wie auch vorher, auf Papier ihre Antworten, sondern darüber hinaus auch Spuren, die den Weg dorthin beschreiben und Auskunft darüber geben könnten, wo gelernt oder eben auch konkret gescheitert wird. Wie schnell wurde geantwortet? Wie oft wurde die Antwort angepasst? In welcher Reihenfolge wurden die Fragen abgearbeitet? Wie viel und wie wurde bei interaktiven Aufgaben mit der Aufgabenumgebung interagiert und experimen-

tiert? Wie häufig wurde der Lesetext konsultiert? Wie wurde durch Hypertexte navigiert? Welche Passagen wurden besonders lange (oder kurz) angezeigt, welche Hilfestellungen wurden genutzt? Diese Informationen, und noch viel mehr, stecken in den besagten Log-Daten. Setzt man diese Daten dann in Bezug zu den Aufgabeninhalten und den inhärenten Aufgabeneigenschaften (z. B. Komplexität, Schwierigkeitsgrad, Sprachanforderungen), so lassen sich Lösungs- und Denkprozesse rekonstruieren, welche mitunter informativer sind als das eigentliche Testergebnis – so zumindest die Theorie.

In integrierten Test-und-Lern-Systemen werden Log-Daten in Echtzeit ausgewertet und über sogenannte *recommender*-Systeme dann jeweils dedizierte Aufgaben administriert, welche den Lerner genau dort abholen, wo dieser gerade steht. Sie verlangen ihm genau das ab, was er benötigt, um eventuell diagnostizierte Defizite aufzuarbeiten und sich generell weiterzuentwickeln. Dies wird häufig mit dem Begriff *formatives Assessment* (in Abgrenzung zum *summative Assessment*) oder *assessment for learning* (in Abgrenzung zu *assessment of learning*) umschrieben. Bei der Datenmasse und Komplexität von Prozessdaten stößt der „klassische“ psychometrische Werkzeugkasten an seine Grenzen, weshalb sich das Feld notgedrungener Weise progressiv in Richtung *machine learning* und *artifizielle Intelligenz* orientiert (s. a. Defossez et al. zu *Gamification* in diesem Band). „Intelligente“ integrierte Test-und-Lern-Systeme, in denen die Grenze zwischen Lern- und Diagnostikaufgaben schwindet, operieren per Definition in der bereits vor hundert Jahren von Vygotsky postulierten „Zone der nächsten Entwicklung“ (ZPD; *zone of proximal development*; Vygotsky, 1978); heutzutage heißt das dann *personalized learning*. Personalisiertes, daten- und algorithmengestütztes Lernen ist insofern die Zukunft, als dass es mit hoher Wahrscheinlichkeit die einzige nachhaltige Antwort auf zunehmend heterogenere Gruppen von Lernern – und die damit einhergehenden Herausforderungen (s. z. B. Fischbach et al., 2016) – sowie auf eine schnellebige Welt des massiven und kontinuierlichen *up- und reskilling* (s. z. B. TIR Consulting Group LLC, 2016) liefert. Die umrissene datengestützte Bildungsrevolution, an deren Schwelle wir derzeit stehen, wird nicht nur die Grenzen zwischen Unterrichten, Lernen und Diagnostik verwischen, sondern auch



die Grenzen zwischen Erziehungs-, Neuro-, Kognitions- und Computerwissenschaften neu definieren.

Und die Lehrkräfte? In einer sich immer schneller drehenden Welt brauchen Bildungssysteme mehr denn je hervorragend ausgebildete, hochgradig professionelle, (selbst)kritische Lehrkräfte, deren Motivation und Berufsethik dem Professionswissen in nichts nachsteht. Wollen wir den von Schön (1983) postulierten *reflective practitioner*, so müssen wir auch die dafür notwendigen Freiräume schaffen. Lehrkräfte sind keinesfalls obsolet in einer Welt des *personalized learning*, aber auch ihre Rolle wird partiell neu definiert. Die progressive Digitalisierung der pädagogischen Diagnostik ist nicht nur eine Chance, um das „digitale Klassenzimmer“ voranzutreiben, sondern auch die perfekte Gelegenheit, um pädagogische Diagnostik von Grund auf neu zu denken und das Bildungssystem sowie dessen Akteure diesbezüglich zu professionalisieren. Denn in der aktuellen schulischen Praxis werden leider – und auch fälschlicherweise – noch allzu oft Evaluation und Diagnostik mit dem Rotstift gleichgesetzt. Diagnostik ist kein Selbstzweck. Gute Diagnostik ist nicht destruktiv und repressiv, sondern konstruktiv, und Evaluation ist nicht das Ende, sondern vielmehr der Anfang neuen Lernens. In dem von uns skizzierten Szenario „intelligenter“ integrierter Test-und-Lern-Systeme sind Lehrkräfte in erster Linie Lehr-Lern-Experten, Vermittler (sogenannte *facilitators*) von Lernopportunitäten, welche Lernprozesse ganz bewusst und gezielt steuern. Lesen und Interpretieren von (aufbereiteten) Daten, welche mit der progressiven Digitalisierung der schulischen Diagnostik mehr denn je zur Verfügung stehen, werden zu unabdinglichen professionellen Kompetenzen (s. a. Baumert & Kunter, 2006). Beim Üben, in der Diagnostik und in der Binnendifferenzierung werden die Lehrkräfte von den umrissenen integrierten Test- und-Lern-Systemen aktiv unterstützt und entlastet, so dass diese professionellen Aufgaben, welche nach wie vor einen Großteil von Schule ausmachen, nicht nur effizienter, sondern auch effektiver durchgeführt werden können und dadurch dringend benötigte Freiräume entstehen. Letztere können dann beispielsweise zur Reflexion und Weiterentwicklung der eigenen Praxis – die Notwendigkeit des regelmäßigen *up- und reskilling* bleibt besonders dem Lehrberuf nicht erspart – oder ein-

fach „nur“ zur Planung und Durchführung von qualitativ hochwertig(er)em Unterricht genutzt werden.

Pionierarbeit in und aus Luxemburg

Zurück ins Hier und Jetzt. Die wenigsten wissen, dass der Standort Luxemburg in der internationalen *Assessment*-Welt seit nunmehr anderthalb Jahrzehnten Synonym für Pionierarbeit und Innovation im Bereich der digitalen pädagogischen Diagnostik ist. Große internationale Vergleichsstudien, wie PISA und PIAAC, laufen auf luxemburgischen Testsystmen, und in Luxemburg basierte Forscher partizipieren in damit verbundenen internationalen Expertengruppen und Konsortien bzw. leiten diese. Der nationale Forschungsbund FNR hat seit Gründung der Universität Luxemburg dutzende Projekte in genau diesem Bereich ermöglicht, darunter unter anderem zwei prestigeträchtige ATTRACT-Stipendien in den Bereichen „Testung von 21st Century Skills“ und „digitales Lernen“.

Luxemburg exportiert nicht nur seine digitale *Assessment*-Expertise, sondern setzt sie selbstverständlich auch im eigenen Land ein: Das nationale Bildungsmonitoring *Épreuves Standardisées* (ÉpStan; Martin et al., 2015; s. a. epstan.lu) ist seit nunmehr über einem Jahrzehnt in den Sekundarschulklassen flächendeckend digitalisiert, Kursevaluationen an der Universität sind komplett digital und mobil, und die Zugangstests zur luxemburgischen Nationalität haben ebenfalls rezent eine Transposition auf Tablets vollzogen. Die drei genannten Beispiele – die keinesfalls als eine erschöpfende Auflistung zu verstehen sind – haben allesamt gemein, dass sie auf dem an der Universität Luxemburg eigens hierfür entwickelten *Online Assessment System* (OASYS; s. Abb. 1) laufen. OASYS deckt bereits heute alle gängigen Substitutionsszenarien ab und geht progressiv über einfache Transposition hinaus (z. B. durch die Integration von neuartigen interaktiven Aufgaben zu digitalem Konzept-Mapping; s. Rohles & Backes in diesem Band). „Intelligent“ ist OASYS (noch) nicht, aber Log-Datenanalyse findet bereits statt. Das System kombiniert die *Assessment*-Expertise des LUCET (*Luxembourg Centre for Educational Testing*) mit der Forschungsgruppe zur Mensch-Computer-Interaktion;



Abb. 1: OASYS Infografik



OASYS
for schools

A powerful **webtool** to build **digital student evaluations**

Designed **together with teachers**

Discover OASYS4schools, an Online-Assessment System designed with and for teachers



Benefit from **traditional question formats** and discover **innovative formats** such as concept maps and versatile drag and drop

Available for **all schools** in Luxembourg

Developed by **SCRIPT & University of Luxembourg**

OASYS: A complete system to support your work



- Multi-platform
- State-of-the-art web technologies
- Support of traditional & innovative item formats
- Multi-device
- State-of-the-art user-centred-design
- Multilingual

What is OASYS?

OASYS is an **Online-Assessment System**, a testing and exam platform. It is a computer- and web-based **e-testing** solution that allows for easy building, delivering and sharing your tests and to answer them just as simply.

Originally developed for the **national school monitoring programme Épreuves Standardisées (ÉpStan)**. OASYS presents successful applications in **large-scale, experimental and survey data collections** since 2010.



It was developed and designed by LUCET and COSA/HCI, an **interdisciplinary** research-based team with over a decade of experience in computer-based assessment, from the University of Luxembourg.

It was designed following a **user-centric process**, with and for users, taking into account their needs and expectations.

UNIVERSITÉ DU LUXEMBOURG  **SCRIPT** Service de Coordination de la Recherche et de l'Innovation pédagogiques et technologiques





es stellt den Benutzer und die Benutzerin – *Front-* wie *Back-End* – ins Zentrum der forschungsbasierten Entwicklung, um so optimale UX und dadurch bestmögliche Datenqualität und Lernprozesse zu gewährleisten. Sicherheit wird ebenfalls großgeschrieben. OASYS nimmt demnach ganz bewusst kritische Qualitätslücken, die branchenüblich oft verschwiegen werden, von Systemen der gleichen Generation in Angriff.

Ausblick

Das explizit benutzerzentrierte Design von OASYS ist einer der Hauptgründe, warum das SCRIPT und die Universität Luxemburg sich 2018 zusammengetan haben, um OASYS unter dem Label „OASYS4schools“ (oasys4schools.lu) im engen und kontinuierlichen Austausch mit den zukünftigen Benutzern und Benutzerinnen u. a. zur Standard-e-Assessment-, Standard-e-Exams- und Standard-e-Hausaufgaben-Plattform der nationalen Bildungslandschaft auszubauen. Der erste große Schritt in Richtung einer Digitalisierung der pädagogischen Diagnostik in Luxemburgs Klassenzimmern ist demnach gemacht. Die Evolution hat begonnen, und die Revolution kann kommen.

Literatur

- Baker, R. S. (2019). Challenges for the future of educational data mining: The Baker learning analytics prizes. *Journal of Educational Data Mining*, 11(1), 1–17.
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520.
- Bunderson, C. V., Inouye, D. K. & Olsen, J. B. (1988). *The four generations of computerized educational measurement* [ETS Research Report]. ETS.
- Fiore, S. M., Graesser, A. & Greiff, S. (2018). Collaborative problem-solving education for the twenty-first-century workforce. *Nature Human Behaviour*, 2(6), 367–369.
- Fischbach, A., Ugen, S. & Martin, R. (2016). Bilanz nach zwei vollen Erhebungszyklen. In *SCRIPT & LUCET, PISA 2015. Nationaler Bericht Luxemburg* (S. 13–19). Luxembourg: MENJE.
- Greiff, S., Wüstenberg, S., Csapó, B., Demetriou, A., Hautamäki, J., Graesser, A. C. & Martin, R. (2014). Domain-general problem solving skills and education in the 21st century. *Educational Research Review*, 13, 74–83.
- Martin, R., Ugen, S. & Fischbach, A. (Hrsg.) (2015). *Épreuves Standardisées: Bildungsmonitoring für Luxemburg. Nationaler Bericht 2011 bis 2013*. Luxembourg: University of Luxembourg, LUCET.
- Quellmalz, E. S. & Pellegrino, J. W. (2009). Technology and testing. *Science*, 323(5910), 75–79.
- Redecker, C. & Johannessen, Ø. (2013). Changing assessment – Towards a new assessment paradigm using ICT. *European Journal of Education*, 48(1), 79–96.
- Robitzsch, A., Lüdtke, O., Köller, O., Kröhne, U., Goldhammer, F. & Heine, J.-H. (2017). Herausforderungen bei der Schätzung von Trends in Schuleistungsstudien: Eine Skalierung der deutschen PISA-Daten. *Diagnostica*, 63(2), 148–165.
- Rohles, B., Koenig, V., Fischbach, A. & Amadieu, F. (2019). Experience matters: Bridging the gap between experience- and functionality-driven design in technology-enhanced learning. *Interaction Design and Architecture(s) Journal*, 42, 11–28.
- Schön, D. (1983). *The Reflective Practitioner: How professionals think in action*. Temple Smith.
- SCRIPT & LUCET. (2016). *PISA 2015. Nationaler Bericht Luxemburg*. Luxembourg: Imprimerie Centrale.
- TIR Consulting Group LLC. (2016). *The 3rd industrial revolution strategy study for the Grand Duchy of Luxembourg*. https://www.troisiemerevolutionindustrielle.lu/wp-content/uploads/2016/11/TIR-Strategy-Study_Short.pdf.
- Vygotsky, L. S. (1978). *Mind in society. The development of higher psychological processes*. Cambridge: Harvard University Press.
- Weinerth, K., Koenig, V., Brunner, M. & Martin, R. (2014). Concept maps: A useful and usable tool for computer-based knowledge assessment? A literature review with a focus on usability. *Computers & Education*, 78, 201–209.