



Digitalisation du diagnostic pédagogique :

de l'évolution à la révolution

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

Introduction

Les considérations autour de l'expression « salle de classe numérique » s'orientent traditionnellement en premier lieu vers des scénarios d'application pédagogiques et didactiques. Or, apprentissage et enseignement vont toujours de pair avec évaluation. Quiconque souhaite gérer avec succès les processus d'enseignement et d'apprentissage, et parvenir à une adéquation optimale entre les cours dispensés et la population scolaire se doit d'évaluer régulièrement l'état des compétences de cette population scolaire. Il va également sans dire que la certification présuppose une évaluation. Ce diagnostic pédagogique en particulier – qu'il soit formatif, sommatif ou certificatif – a énormément à gagner d'une utilisation ciblée des nouveaux médias et des nouvelles technologies, tant en termes d'efficience (par exemple, évaluation automatique, logistique simplifiée, plus d'informations collectées) que d'efficacité (par exemple, psychométrie améliorée et représentation de compétences avancées, sélection adaptative des tâches, *feedback* en temps réel, interrelation entre apprentissage et diagnostic) et d'équité (par exemple, standardisation accrue de la situation de test et de l'évaluation, interfaces utilisateurs inclusives, évaluations sans recours à la langue et multilingues).

Opportunités et défis

Vers la fin des années 1980, l'*Educational Testing Service* (ETS, Princeton), le leader mondial de « l'industrie du test », a publié une étude (Bunderson et al., 1988)

dans laquelle les auteurs examinaient les nouvelles possibilités de diagnostic pédagogique assisté par ordinateur et exposaient leurs prévisions concernant de futures évolutions. L'éventail des scénarios d'application esquissés allait du scénario minimal d'une simple substitution – soit un simple transfert pratiquement « un à un » d'évaluations papier vers un support numérique – au scénario maximal d'une véritable transformation – soit l'interrelation parfaite entre apprentissage et diagnostic sous la forme de systèmes intégrés de test et d'apprentissage « intelligents », en passant par une transposition assortie d'améliorations numériques.

Depuis la publication du rapport de l'ETS, plus de trois décennies se sont écoulées. Mis à part le fait que les choses ont évolué beaucoup plus lentement qu'anticipé à l'époque, et que des innovations technologiques révolutionnaires n'ont pas été adéquatement prises en compte pour ces projections, comme la mise en réseau mondiale et l'ubiquité d'ordinateurs mobiles (tablettes, smartphones, montres connectées), une conclusion semble s'imposer : les principaux postulats et hypothèses de l'époque, et surtout les défis qui y sont associés, sont plus actuels que jamais (voir p. ex. Quellmalz & Pellegrino, 2009 ; voir également Greiff et al, 2014 ; Redecker & Johannessen, 2013). L'étude PISA-2015 (SCRIPT & LUCET, 2016), entièrement numérisée à l'échelle mondiale et parfois controversée précisément pour cette raison (voir p. ex. Robitzsch et al., 2017 ; voir également Fischbach et al., 2016), a définitivement marqué



l'introduction de ladite transposition dans les pratiques courantes, du moins en ce qui concerne l'évaluation des systèmes. Avec l'intégration de tâches ou d'exercices interactifs pour la résolution de problèmes complexes et collaboratifs (Fiore et al., 2018 ; voir également le chapitre sur KolPS dans ce volume), la transposition va manifestement au-delà d'une simple substitution en intégrant de nouveaux construits qui nécessitent une mesure axée sur la technologie. Malgré tout, on est encore loin de l'idée d'une véritable transformation. En outre, de nombreux systèmes disponibles aujourd'hui, même en mode de substitution simple, soit le transfert direct d'une tâche vers, par exemple, l'ordinateur, présentent encore des faiblesses dans des domaines critiques tels que la sécurité, la protection des données, la facilité d'utilisation et l'expérience utilisateur (*user experience*, UX). Cette dernière est particulièrement pertinente dans la mesure où elle constitue la clé pour obtenir des données valides et équitables (mot-clé : *digital divide*, ou fracture numérique) en rendant l'utilisation des nouvelles technologies aussi intuitives, pour tous les apprenants, que les scénarios « *low-tech* » (p. ex. sur papier) (voir p. ex. Rohles et al., 2019 ; Weinerth et al., 2014). Par conséquent, nous assistons actuellement à une évolution plutôt qu'à une révolution, surtout en matière d'assurance qualité systémique, et le diagnostic pédagogique numérique n'a jusqu'à présent pu faire son entrée en salle de classe que de façon sporadique.

Des systèmes intégrés de test et d'apprentissage « intelligents »

La vision des années 1980 concernant des systèmes intégrés de test et d'apprentissage « intelligents » (Bunderson et al., 1988) ne s'est à ce jour pas encore matérialisée (voir p. ex. Baker, 2019) ; cependant, nous nous sommes quelque peu rapprochés au cours des trente dernières années. Le défi majeur est la prise en compte des traces comportementales (*log*), ou « données de processus ». Lorsqu'ils utilisent des systèmes de test numériques, les utilisateurs ne fournissent pas seulement leurs réponses comme ils le faisaient auparavant sur papier, mais produisent également des « traces » comportementales qui décrivent le chemin parcouru pour arriver au résultat final ; ces traces peuvent contenir

des indications sur ce qui est bien assimilé, mais aussi sur les échecs d'apprentissage. Dans quel laps de temps l'utilisateur a-t-il répondu ? Combien de fois la réponse a-t-elle été modifiée ? Dans quel ordre les questions ont-elles été traitées ? Dans quelle mesure et comment l'utilisateur a-t-il interagi avec l'environnement numérique et l'a-t-il exploré lors de tâches interactives ? À quelle fréquence le texte de lecture a-t-il été consulté ? Comment l'utilisateur a-t-il navigué dans les hypertextes ? Quels passages ont été affichés pendant une période particulièrement longue ou courte, et quels outils d'aide ont été utilisés ? Ces informations, et bien d'autres encore, sont contenues dans les fichiers « log ». La mise en relation de ces données avec les contenus des tâches et les propriétés inhérentes aux tâches (comme la complexité, le degré de difficulté, les exigences linguistiques) permet de reconstruire les processus de solution et de réflexion, qui sont parfois plus instructifs que le résultat du test à proprement parler – du moins en théorie.

Les systèmes intégrés de test et d'apprentissage permettent d'évaluer les données de journal en temps réel et de proposer ensuite des tâches dédiées via des systèmes de recommandation (*recommender systems*), qui correspondent exactement au niveau des connaissances de l'apprenant au moment donné. Il sera ainsi amené à faire les efforts nécessaires pour remédier à d'éventuels déficits diagnostiqués et continuer à se développer de manière générale. Le terme d'évaluation formative (par distinction avec l'évaluation sommative), ou d'*assessment for learning* (par distinction avec l'*assessment of learning*) est fréquemment utilisé pour désigner cette approche. Compte tenu de la masse de données et de la complexité des données de processus, la boîte à outils psychométriques « classique » atteint ses limites, raison pour laquelle la branche n'a d'autre choix que de se diriger progressivement vers le *machine learning* et l'intelligence artificielle (voir également Defossez et al. sur la ludification (gamification) dans ce volume). Les systèmes intégrés de test et d'apprentissage « intelligents », qui font disparaître la frontière entre les tâches d'apprentissage et de diagnostic, fonctionnent par définition dans la « zone proximale de développement » (ZPD, *zone of proximal development* ; Vygotsky, 1978) postulé il y a cent ans déjà par Vygotsky ; aujourd'hui, le terme utilisé est « apprentissage personnalisé » (*personalized learning*). L'apprentis-



sage personnalisé basé sur des données et algorithmes représente l'avenir dans la mesure où il constitue très probablement la seule réponse durable à l'hétérogénéité croissante des groupes d'apprenants – et aux défis qui en découlent (voir p. ex. Fischbach et al., 2016) – ainsi qu'à un monde en ébullition, placé sous le signe de la remise à niveau (*upskilling*) et du recyclage (*reskilling*) massifs et continus (voir p. ex. TIR Consulting Group LLC, 2016). La révolution de l'enseignement décrite, basée sur les données, qui s'annonce imminente est non seulement en passe de dissoudre les frontières entre l'enseignement, l'apprentissage et le diagnostic, mais également de redéfinir les frontières entre les sciences éducatives, les sciences (neuro)cognitives et l'informatique.

Et qu'en est-il des enseignants dans ce scénario ? Dans un monde qui évolue de plus en plus rapidement, les systèmes éducatifs ont plus que jamais besoin d'enseignants parfaitement qualifiés, hautement professionnels et (auto)critiques, dont la motivation et la déontologie n'ont d'égal que leurs compétences professionnelles. Si nous voulons que les enseignants répondent à l'image du praticien réflexif (*reflective practitioner*) de Schön (1983), il faut également créer les espaces de liberté nécessaires à cet effet. Les enseignants ne sont nullement obsolètes dans l'univers de l'apprentissage personnalisé, mais leur rôle lui aussi est amené en partie à une redéfinition. La digitalisation progressive du diagnostic pédagogique constitue non seulement une opportunité de faire progresser la « salle de classe numérique », mais aussi l'occasion idéale de repenser fondamentalement le diagnostic pédagogique, et de professionnaliser le système éducatif et ses acteurs à cet égard, car dans la pratique scolaire actuelle, l'évaluation et le diagnostic sont malheureusement – et à tort – encore trop souvent assimilés au stylo rouge. Le diagnostic n'est pas une finalité en soi. Un bon diagnostic n'est ni destructeur ni répressif, mais constructif, et l'évaluation ne constitue pas la fin, mais plutôt le début d'un nouvel apprentissage. Dans le scénario de systèmes intégrés de test et d'apprentissage « intelligents » tel que décrit, les enseignants sont avant tout des experts en enseignement et en apprentissage, des médiateurs – ou « facilitateurs » – d'opportunités d'apprentissage, qui contrôlent les processus d'apprentissage en toute conscience et de façon ciblée. La lecture et l'interprétation de données (traitées), qui sont plus

que jamais disponibles avec la digitalisation progressive du diagnostic scolaire, deviennent des compétences professionnelles incontournables (voir également Baumert et Kunter, 2006). Pour les exercices tout comme pour le diagnostic ou encore la différenciation interne, les enseignants bénéficient d'un soutien actif et se voient délestés d'une partie de la charge de travail par les systèmes intégrés de test et d'apprentissage décrits, de sorte que ces tâches professionnelles – qui continuent à constituer une grande partie de l'enseignement – peuvent être accomplies non seulement de manière plus efficiente, mais aussi plus efficace, offrant ainsi des espaces de liberté bien nécessaires. Ceux-ci peuvent ensuite être mis à profit, par exemple, pour mener une réflexion et faire évoluer sa propre pratique – la profession d'enseignant n'étant certainement pas épargnée par la nécessité d'une mise à niveau (*upskilling*) et d'un recyclage (*reskilling*) réguliers – ou tout simplement « seulement » pour planifier et mettre en œuvre un enseignement de haute qualité.


Un travail précurseur luxembourgeois

Peu de gens savent que ces 15 dernières années, le Luxembourg a joué un rôle précurseur et novateur en matière d'évaluation à l'échelle internationale en ce qui concerne le diagnostic pédagogique numérique. De vastes études comparatives internationales, telles que les enquêtes PISA et PIAAC, sont menées sur des systèmes de test luxembourgeois, et des chercheurs basés au Luxembourg participent à, voire dirigent, des groupes d'experts et des consortiums internationaux connexes. Depuis que l'Université du Luxembourg existe, le Fonds National de la Recherche (FNR) a facilité la mise en œuvre de dizaines de projets précisément dans ce domaine. On notera notamment l'octroi de deux prestigieuses bourses ATTRACT dans les domaines de l'évaluation des compétences du 21^e siècle, ainsi que de l'apprentissage numérique.

Le Luxembourg n'exporte pas seulement son expertise en matière d'évaluation numérique, mais la met bien sûr également à profit sur son propre territoire : depuis plus de dix ans, le dispositif de monitoring national connu sous le nom d'Épreuves standardisées (ÉpStan ; Martin et al., 2015 ; voir également epstan.lu) existe de manière




Fig. 1 : Infographie OASYS



OASYS
for schools

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



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

Designed **together with 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



généralisée sous forme numérisée dans les classes d'enseignement secondaire, les évaluations de cours à l'Université sont entièrement numérisées et mobiles, et les tests d'accès à la nationalité luxembourgeoise ont également été transposés sur tablettes il y a peu. Les trois exemples mentionnés – qu'il ne faut en aucun cas considérer comme une énumération exhaustive – ont tous en commun de s'appuyer sur le système d'évaluation en ligne spécialement conçu à cet effet à l'Université du Luxembourg (OASYS ; voir Fig. 1). OASYS couvre déjà tous les scénarios de substitution courants, et se dirige progressivement au-delà de la simple transposition (par exemple en intégrant de nouvelles tâches interactives de « concept mapping » numérique ; voir le chapitre sur le « concept mapping » dans ce volume). OASYS n'est pas (encore) « intelligent », mais l'analyse des données comportementales a déjà lieu. Le système combine l'expertise en matière d'évaluation du LUCET (Luxembourg Centre for Educational Testing) avec celle du groupe de recherche sur l'interaction homme-machine, et place l'utilisateur, tant en « front end » qu'en « back end », au centre du développement basé sur la recherche. C'est ainsi qu'il garantit une expérience utilisateur optimale ainsi que la meilleure qualité de données, constituant la base des processus d'apprentissage fructueux. La sécurité constitue également une priorité absolue. OASYS s'attaque donc résolument à ces lacunes critiques de qualité, souvent tenues secrètes dans la branche, de systèmes de la même génération.

Perspectives

La conception explicitement centrée sur l'utilisateur d'OASYS est l'une des principales raisons pour lesquelles le SCRIPT, et l'Université du Luxembourg ont uni leurs efforts en 2018 pour développer OASYS sous le label « OASYS4schools » (oasys4schools.lu), en collaboration étroite et continue avec les futurs utilisateurs, pour en faire la plateforme standard en matière d'évaluation numérique, d'examen électronique, de travail à domicile électronique, etc. du paysage éducatif national. Le premier grand pas vers la digitalisation du diagnostic pédagogique dans les salles de classe luxembourgeoises a ainsi été franchi. L'évolution a commencé – que suivra la révolution.

Références

- 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. Dans SCRIPT & LUCET, PISA 2015. *Nationaler Bericht Luxemburg* (p. 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. (Eds.) (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 Schulleistungsstudien: Eine Skalierung der deutschen PISA-Daten. *Diagnostica*, 63(2), 148–165.
- Rohles, B., Koenig, V., Fischbach, A. & Amadiou, 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-CG_Luxembourg-Final-Report_Long-Version.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.