FEATURES

CHALLENGES AND DEVELOPMENTS IN TECHNOLOGY-BASED ASSESSMENT: POSSIBILITIES IN SCIENCE EDUCATION

Gyöngyvér Molnár – DOI: https://doi.org/10.1051/epn/2021202
University of Szeged, Institute of Education, Research Group on the Development of Competencies

Motto: 'If you cannot measure it, you cannot improve it.' – *Lord Kelvin*

Nowadays, your smartphone offers workout programs for you if you decide to lose weight after you enter a few important parameters (*e.g.* height, weight, age and gender) and your goal. As you do sport, your smartwatch continuously monitors your time, calories burnt, pulse rate, heart rate *etc.* More accurate, personalized workout programs are provided for you based on these data. In parallel, technology can contribute to personalizing education by even predicting what types of tasks and activities would be most beneficial for different students.

From paper-based to technology-based assessment

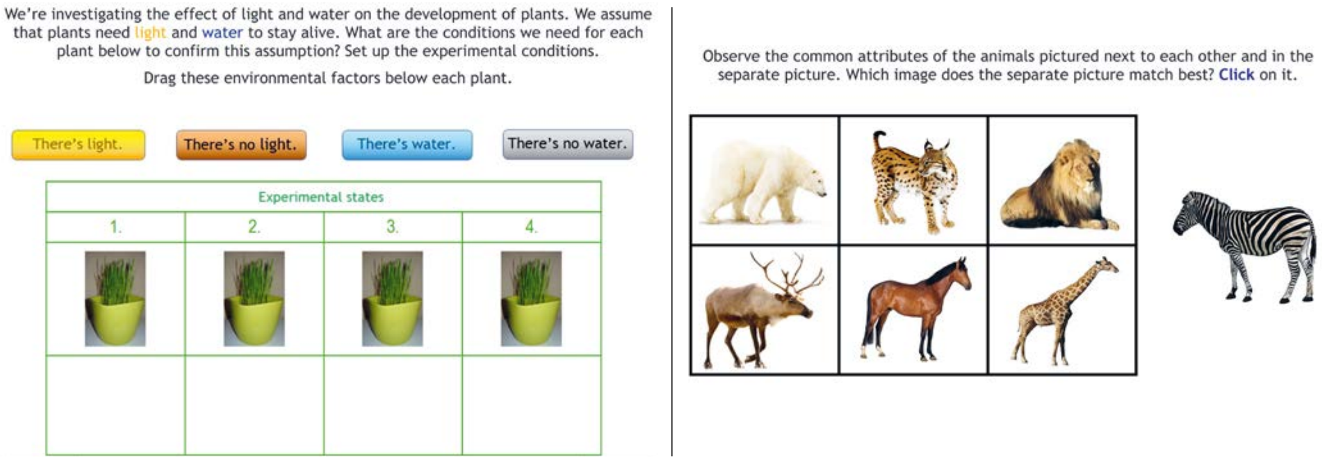
Educational assessment has been among the most dynamically developing areas in education since the turn of the millennium. In this period of time, large-scale international assessments (*e.g.* OECD PISA[1] and IEA TIMSS[2]) have become regularly administered by the world's leading test centres, resulting in a huge improvement of data transfer technology and data analysis methods. Two decades ago, paper-based assessments were the most widespread and accepted means of assessment, but due to rapid development, the tools for paper-based assessments represented serious constraints on further improvements. A qualitative change and a new kind of assessment had to be made to meet the learning and assessment needs of the 21st century.

The direction of these developments has been determined by technology, especially by computers, thus offering extraordinary opportunities. We can administer tasks in a more realistic, application-oriented, engaging and authentic context with computer-based assessment; we can use innovative item development opportunities, producing dynamic, interactive multimedia items. For example, in science education, students can engage with simulations and apply their knowledge of scientific

▲ © iStockPhoto

[1] See https://www.oecd.org/pisa/

[2] See https://www.iea.nl/studies/iea/timss



▲ **FIG. 1:** Two examples from the reasoning dimension of science (both are Grade 6-level tasks)

concepts to solve problems. We can design more valid assessments. For example, using an audio version of science tasks and instructions, we can exclude the level of reading skills from the final achievement. Technology-based assessment makes it possible to provide instant, objective, standardized feedback, thus replacing previous long feedback times, and to use adaptive test algorithms to fit the difficulty level of the tasks to the knowledge and skill level of the students (see Csapó *et al.*, 2012).

Adaptive testing makes assessment results more exact and makes assessment fully personalized. In traditional testing, each person receives the same tasks in the same order. In contrast, in adaptive testing, each person completes different tasks, with the most diagnostic power from an item bank. The results can be compared because the items are scaled and defined on a common difficulty and ability scale, even though the students took different tests. The difficulty level of tasks thus administered is tied to the ability level of the students, offering them an optimal challenge. Testing therefore does not become boring or cause anxiety. This can have a positive effect on students' interest and test-taking motivation, which is crucial for the frequent use of tests. This type of testing was even implemented in the PISA 2018 main survey for the domain of reading.

These issues – instant feedback and adaptivity – are also among the secrets that make video games so popular. We do not need to wait weeks or longer to receive feedback on our performance, and we do not need to solve problems which are too difficult for us. If we have to wait days, weeks or months for feedback or start playing the game at a level that is too difficult for us and have no success or if it is too easy and poses no challenge for us, we will never, ever play that game. These are the areas where technology-based assessment and game-based learning are converging, with numerous innovations possible in both.

Today, computer-based assessment offers more effective assessments (*e.g.* they are cheaper, the data flow is faster and safer, indicators of test goodness are higher, student motivation is higher, and feedback is quicker) than traditional paper-based or face-to-face testing. Using at least some of the advantages of computer-based assessment, international summative tests have already been transitioned from paper-and-pencil to digitally-based assessments, and all important assessments will probably follow suit within a reasonable time.

From summative to personalized diagnostic assessment

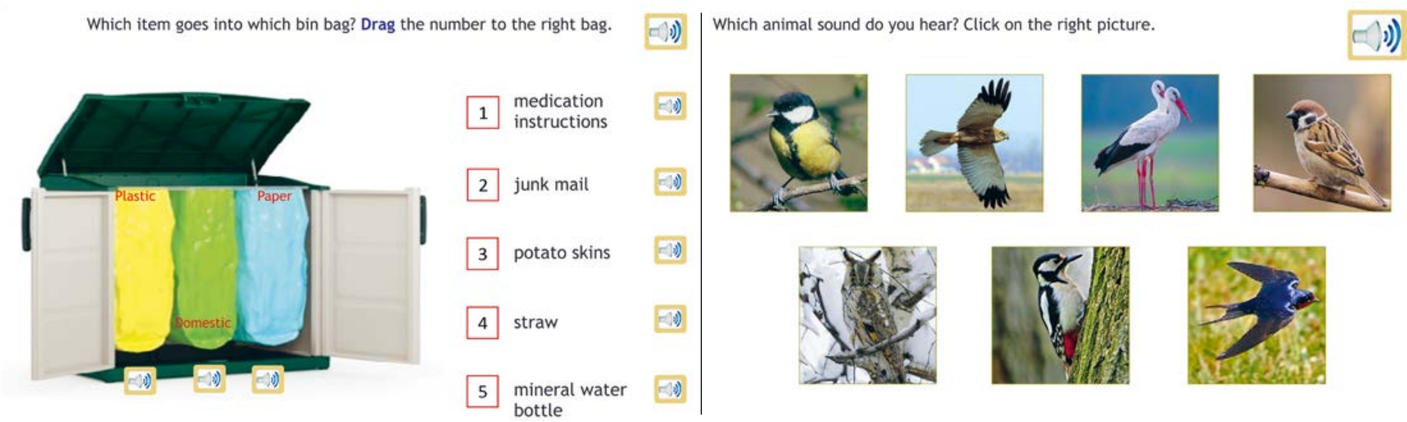
In 2021 there is no longer any question whether we can develop complex, real-world, authentic, high-quality tests. COVID-19-related school closures and digital teaching have reinforced the idea that the 'one-size-fits-all' approach is not effective, either in general or in educational assessment in particular. The almost exclusively used summative test[3] results have limited usefulness with regard to learning and teaching processes to personalize intervention and student-level feedback in general (Csapó & Molnár, 2019). They are good for accountability purposes (see Koretz, 2018) in 'normal teaching times', but they do not meet the individual needs of students. They do not provide actionable feedback for learners to aid in improving their learning process. The COVID-19-related interruptions or modifications in high-stakes national assessment provide an opportune moment to re-think the essence of assessment (Cairns, 2020) and the elimination of summative tests[4].

This crisis is a good reminder that beyond summative, high-stakes testing a more learning-centred, low-stakes approach[5], using the power of prompt, proper ● ● ●

[3] Summative assessments are used to evaluate student academic achievement at the end of a defined instructional period (e.g. a unit, course, semester or school year). Examples of summative assessments include end-of-unit tests, end-of-term tests and standardized tests for the purposes of school accountability, e.g. the SAT and Matura.

[4] Potential changes to if not elimination of the SAT: https://www.insidehighered.com/admissions/article/2021/01/25/changes-sat-prompt-discussion-future-college-board

[5] Low-stakes assessment (as opposed to high-stakes assessment) is any form of evaluation that does not heavily impact students' educational outcomes or final grades at school. The purpose of low-stakes assessment is to provide students with information on their actual performance and to provide an opportunity to improve their achievement prior to final grading.



▲FIG. 2: Two examples from the application dimension of science (both are Grade 5-level tasks)

● ● ● – that is, efficient – feedback is relevant and appropriate. This approach was not possible in the days of paper-based testing, but is now fully realisable with technology-based assessment. Like the world of video games, tasks and tests which match students' ability level (adaptive or tailored tests) and assessment with regular and objective feedback are also motivating in education, can even result in flow, and enable teachers to tailor instruction and support students' development more effectively.

Technology-based diagnostic assessments for identifying early learning difficulties in science

In the history of science education, three main goals have remained clear from the very beginning of schooling and public education up to the present day: (1) to cultivate students' minds, general cognitive abilities and thinking skills; (2) to develop usable, transferable, applicable knowledge, that is, knowledge which is applicable inside and outside of school context; and (3) to transmit content knowledge accumulated and organised according to the values of the sciences (Csapó & Szabó, 2012). In modern societies these three dimensions of learning should be present simultaneously, reinforce and interact with each other, and not compete for teaching time. However, most digital games in the sciences develop scientific knowledge/concepts, and less than one-third facilitate students' thinking skills and scientific reasoning skills[7].

The development of the eDia system by the Research Group on Learning and Instruction at the University of Szeged addresses this issue (Csapó & Molnár, 2019). The aim of the technology-based assessment system is to identify students' developmental deficiencies in the domains of reading, mathematics and science and to provide diagnostic feedback, based on an empirically validated three-dimensional model of knowledge (Molnár & Csapó, 2019, 2020) with objective reference points for teachers on their students' development.

Scientific reasoning tasks, that is, tasks from the reasoning dimension of scientific knowledge (see Figure 1), are the most universal across cultures and school systems.

Tasks in the application dimension of knowledge are embedded in relevant situations, and real life-like context illustrated by pictures, videos or simulations, which can be manipulated. In the first tasks presented in Figure 2 students can interact with an authentic problem environment using online technology. The second task presented in Figure 2 uses an item format, which it is impossible to realise with traditional techniques.

The disciplinary dimension of knowledge is measured in the eDia diagnostic system via tasks assessing the acquisition of concepts and procedures which are part of the curriculum. Figure 3 presents an example, in which pupils' disciplinary scientific knowledge is assessed in the area of Earth science.

At present, the eDia system contains more than 25,000 innovative (multimedia-supported), empirically scaled tasks – one-third of these tasks are science tasks – and provides students and teachers regular feedback from the beginning of schooling to the end of the six years of primary education. Beyond student-level achievement and national standards, teachers receive feedback on their class- and school-level results in comparison to regional and state-level achievement. At present, the system is used during the whole school year in more than 1000 elementary schools (approx. one-third of the primary schools in Hungary). Our future plan is to administer the task on an adaptive level and make better use of the impact of visualization in feedback, that is, offer teachers an interactive feedback module with highly integrated visualization techniques.

At the very beginning of the school closures caused by COVID-19, we launched a new module of the eDia system, the eDia kindergarten test module, and made the eDia teacher test module available to everyone. Both are free and available to all. The kindergarten module

[6] Flow is a mental state of being fully immersed and focused, the optimal experience, based on the classic work of Mihály Csíkszentmihályi.

[7] Scientific reasoning encompasses reasoning skills involved in generating, testing and revising hypotheses and theories.

(see ovi.edia.hu with tasks in Hungarian) contains more than 2000 innovative reading, counting and science tasks developed for kindergarten-aged children, while the teacher test module contains all the 20,000 tasks used in the diagnostic assessments (test.edia.hu). The tasks can be filtered by domain, topic, sub-topic, grade and difficulty level. Then online tests or mini-developmental training can be generated from the selected tasks with prompt feedback. All the students need to do the test is a technological device (*e.g.* a personal computer, tablet or smartphone) with an Internet connection.

Recent research findings offer a promising basis for further integration of game-based approaches to science education to enhance students' factual knowledge, its applicability and students' reasoning skills through active participation and interaction, thus finding the balance between digital learning and working with appropriate challenges. ■

Acknowledgements

This study has been conducted with support from the National Research, Development and Innovation Fund of Hungary, financed under the OTKA K135727 funding scheme.

About the Author

Gyöngyvér Molnár is a full professor and the head of Institue of Education and the deputy head of the Doctoral School of Education at the University of Szeged, Hungary. She earned her PhD in 2004. Her main areas of interest include: technology-based assessment, improving cognitive skills, and the potential for using ICT in education.

References

[1] R. Cairns, Exams tested by Covid-19: An opportunity to rethink standardized senior secondary examinations (2020). Prospects, https://doi.org/10.1007/s11125-020-09515-9

[2] B. Csapó and G. Molnár, Online diagnostic assessment in support of personalized teaching and learning: The eDia system. Frontiers in Psychology (2019).

[3] B. Csapó and G. Szabó, *Framework for diagnostic assessment of science. Budapest: Nemzeti Tankönyvkiadó* (2012, Eds.), http://pedagogus.edia.hu/sites/default/files/science_framework_english.pdf

[4] B. Csapó, J. Ainley, R. E. Bennett, T. Latour, and N. Law, . Technological issues for computer-based assessment. In P. Griffin, B. McGaw, & E. Care (Eds.), Assessment and teaching of 21st century skills, 143 (2012). New York: Springer. https://doi.org/10.1007/978-94-007-2324-5_4

[5] D. Koretz, *American Educator* **41(4)**, 22 (2018).

[6] G. Molnár and B. Csapó, *Frontiers in Psychology* **10**, 1368 (2019).

[7] G. Molnár and B. Csapó, Separating the disciplinary, application and reasoning dimensions of learning: the power of technology-based sssessment, 174. In H. Lane, S. Zvacek, & J. Uhomoibhi (Eds.), Computer Supported Education. Springer (2020).