

Deciding when to trust online test scores

Assessment and feedback

Clare Walsh, Data Science Tutor,
University of Southampton, UK

What's the idea?

Educational assessments are, at the best of times, an imprecise science. We also know that checking performance against success criteria is an important part of the learning process. Successful assessment design requires time, and skills that are given scant attention in Initial Teacher Training courses (Carter, 2015). Third party online assessment tools are a solution.

What does it mean?

Results from one approach to assessment may not predict success in another. Multiple-choice tests continue to be popular among nations that favor a statistical approach to test validity, but not all assessments are structured that way. The Welsh Skills Challenge Certificate (SCC), for example, has a curriculum driven paradigm, and a range of evidence collection methods. A test of factual recall, such as the multiple-choice questions in Kahoot tests, may only predict how students will perform in similar, multiple-choice tests.

How does it work in practice?

Machines may be coded with certain peculiarities. Free text answers may be sensitive to spelling, or upper and lower case. Natural Language Processing, a

subfield of Artificial Intelligence, is now a mature technology used to score free text in adult formal tests, such as the PTE-A university entrance language test. It is worth stress-testing scoring systems for response tolerance. Do they punish spelling, regional accents, soft voices, or even just being female?

Some assessments are criterion referenced, and student performance is referenced solely against a list of achievements. Other tests are cohort referenced, meaning that the children's ability is judged in comparison to the ability of other children taking the test at the same time (Baird et al., 2018). If you are using tests or games with online real time competition, their performance will inevitably reflect not only the ability of the child you are interested in, but also the ability of their competitors. The test or game environment may also interfere.

The difficulty of test questions within the same component of a syllabus can vary widely. The classroom practice of awarding '1' for every tick gives the impression of measuring units of cognition (Almond, Mislevy, Steinberg et al., 2015), an assumption an assessor would interrogate rigorously. A tick just shows that students answered that particular question correctly, and does not necessarily indicate that they are performing at the correct level.

Children need to be socialized into the process of completing school leaving exams (Mislevy, 2018). They may not realise that they are playing the role of test taker when they 'play' online tests or games. We observed behaviors such as allowing a friend to win, wandering off, or simply wanting to see what happens when they lose. If intended results are needed, it's important to share your expectations.

Online test providers need to invest significant resources into coding engaging environments, and curriculum coverage may be sacrificed as a result. There are, for example, a very large number of games, such as Memrise, that teach and test entry level languages. As the curriculum gets broader, coverage starts to drop off, leading to gaps in the syllabus where you have no data.

As with all assessments, there will be difficult compromises when using online 3rd party assessment tools. It is important that teachers consider the limitations of the testing tools that they are using.

WANT TO KNOW MORE?

- Almond RG, Mislevy RJ, Steinberg LS et al. (2015) Introduction. In: Bayesian Networks in Educational Assessment. New York: Springer, pp. 3–18.
- Baird J (2018) The meaning of national examinations standards. In: Baird J, Isaacs T, Opposs D et al. (eds) *Examination standards: how measures and meanings differ around the world*. London: IOE Press.
- Carter A (2015) Carter review of initial teacher training (ITT). Available at: <https://www.gov.uk/government/publications/carter-review-of-initial-teacher-training> (accessed 12 January 2021).
- Mislevy RJ (2018) *Socio-cognitive foundations of educational measurement*. Abingdon: Routledge.

CHARTERED
COLLEGE OF
TEACHING