ALN ASSESSMENT LEARNING NETWORK

LEARNING POINT

How can stealth assessment in games measure and support learning?

In general, stealth assessment is an approach that embeds ongoing formative assessment opportunities deeply into a game (or other engaging digital learning environment), blurring the distinction between learning and assessment (see Figure 1). While playing the game, students continually produce rich sequences of actions as data points that are captured in log files.

Stealth assessments can be a viable alternative to traditional summative assessment (assessment of learning) methods that have remained the same across decades. It can also be an engaging form of formative assessment (assessment for learning) in which data is collected throughout a course of learning with the goal to provide targeted support to the learner along the way.

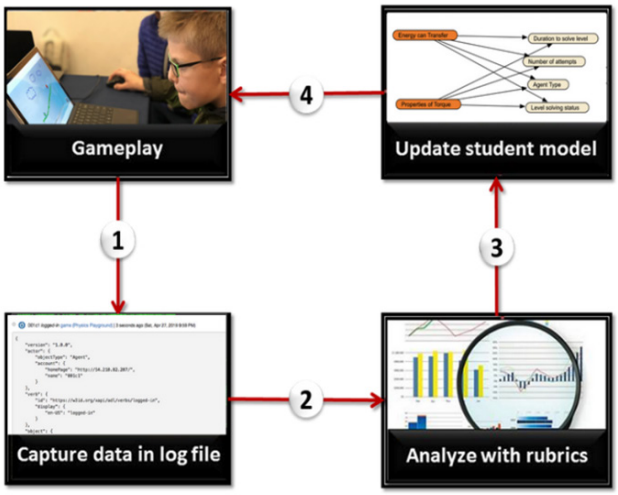

Figure 1. Stealth assessment process

Why use stealth assessment?

The traditional selected-response tests given on and scored by computer find favor because they are easy to deliver and score, but they have some limitations. For instance, they have a limited number of items so can't fully assess what has been taught or learned in class. They also just measure learning at a single point in time summarized by an overall "score" with rarely any constructive feedback. And as described in Shute et al. (2016), these tests:

- predict badly,
- don't measure all relevant determinants of important criteria related to learning,
- are subject to coaching,
- can be biased against members of racial and ethnic minority groups, and
- are subject to motivational differences among students.

Consequently, such assessments can leave lower-performing students even further demotivated and further behind. Stealth assessment offers a potential solution to these problems. Data logs captured through stealth assessment are automatically scored by in-game rubrics, then aggregated in real-time by Bayesian networks or other statistical models, which show evolving estimates of mastery levels on targeted competencies. Shute, Lu, and Rahimi (2021) provide more information on the particular steps needed to develop and validate a stealth assessment.

Evidence-centered design

Moving from log file data to valid inferences about competency states is an important yet quite difficult task. Stealth assessment addresses this issue by employing the evidence-centered design (ECD) framework (Mislevy, Steinberg, & Almond 2003). ECD provides a way to reason about assessment design and student performance.

There are four core ECD models:

(1) **Competency Model**—operationalizing the construct we want to assess (e.g., conceptual physics understanding, creativity, persistence) and defining the claims to be made about students' competencies;

(2) **Evidence Model**—automatically scoring and accumulating valid evidence (i.e., observables) of a claim about students' competencies (i.e., unobservables);

(3) **Task Model**—detailing the nature and form of tasks (e.g., game levels) that will elicit the evidence needed for the evidence model; and

(4) **Assembly Model**—specifying the number, types, and sequencing of tasks.

> **Stealth assessment represents an evidence-based approach that discreetly assesses students' learning progression while they are engaged with a highly interactive and immersive environment, then provides learning supports, as needed.**

In contrast with traditional summative assessment, digital game-based assessment methods like stealth assessment have the following merits:

- They are fun and engaging for most people and can reduce test anxiety;
- They allow for recording students' interactions in detail (i.e., via the accumulation of log data generated by keystrokes, mouse clicks, choice patterns, and so on), which can be used to analyze students' learning progress; and
- They can be designed to provide ongoing feedback and other real-time learning supports, which traditional assessment does not.

Moreover, the psychometric qualities of stealth assessment (e.g., validity, reliability, fairness) are at least on par with typical summative assessments. So, when using games with stealth assessment to measure and support learning, students are likely to be engaged and learn. For example, in one study (Shute et al., 2020) researchers designed and validated a stealth assessment of physics understanding in the game Physics Playground. Not only did the students learn physics as a function of gameplay, but they also enjoyed the experience, rating it on average a 4 on a 1–5 scale (1 = strongly dislike to 5 = strongly like). Moreover, there were no differences in terms of either learning gains or enjoyment by gender or ethnicity.

Features of well-designed stealth assessment

Stealth assessment represents an evidence-based approach that discreetly assesses students' learning progression while they are engaged with a highly interactive and immersive environment, then provides learning supports, as needed. It thus aims to blur the boundaries between gameplay, learning, and assessment using unobtrusive methods (e.g., log files and eye tracking) to continually collect student data and examine their progression of cognitive and non-cognitive variables throughout the game. Assessment is part of gameplay.

The main aspects of a well-designed stealth assessment include:

(1) the use or creation of a technology-rich environment (e.g., a digital game);
(2) the application of ECD to design, develop, and implement the core models described above,
(3) the embedding of the stealth assessment into the code of a technology-rich environment; and
(4) the creation of capabilities for the system to provide formative assessment/feedback, and/or adaptivity —of game levels, supports, and so on.

Possible hurdles

Despite the various features of using stealth assessment in games to support learning in school and at home, there are some hurdles to surmount before it can become mainstream. Assessment design frameworks, like ECD, represent a design methodology but not a panacea, so more research is needed to figure out how to create common measurements from diverse environments. That is, it's important to figure out how to interpret evidence where the activities may be the same but the contexts in which students are working are different (e.g., working alone vs. working with another student).

Another hurdle involves figuring out a way to resolve privacy, security, and ownership issues regarding student information. The privacy/security issue relates to the accumulation of student data from disparate sources. The main issue boils down to this: information about individual students may be at risk of being shared far more broadly than is justifiable. And being aware of the often high-stakes consequences associated with tests, many parents, and other stakeholders may fear that the data collected could later be used against the students.

In any case, constructing the envisioned ubiquitous and unobtrusive stealth assessments across multiple learner dimensions, with data accessible by diverse stakeholders, could yield educational benefits.

First, the time spent administering tests, handling make-up exams, and going over test responses is not very conducive to learning. Given the importance of time on task as a predictor of learning, reallocating those test-preparation activities into ones that are more educationally productive would provide potentially larger benefits to almost all students. Second, by having assessments that are continuous and ubiquitous, students are no longer able to "cram" for an exam. Although cramming can provide good short-term recall, it is a poor route to long-term retention and transfer of learning.

Conclusion

Traditional assessment practices in school can lead to assessing students in a manner that may conflict with their long-term success. With a continuous assessment model in place, the best way for students to do well is to do well every day. This shift in assessment mirrors the national shift toward evaluating students based on acquired competencies.

References

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). Focus article: On the structure of educational assessments. Measurement: Interdisciplinary Research & Perspective, 1(1), 3–62. https://doi.org/10.1207/S15366359MEA0101_02

Shute, V. J., Leighton, J. P., Jang, E. E., & Chu, M-W. (2016). Advances in the science of assessment. Educational Assessment, 21(1), 34-59.

Shute, V. J., Lu, X., & Rahimi, S. (2021). Stealth assessment. In J. M. Spector (Ed.), The Routledge Encyclopedia of Education (pp. 1-9). London, UK: Taylor & Francis group.

Shute, V. J., Rahimi, S., Smith, G., Ke, F., Almond, R., Dai, C-P, Kamikabeya, R., Liu, Z., Yang, X., & Sun, C. (2020). Maximizing learning without sacrificing the fun: Stealth assessment, adaptivity, and learning supports in educational games. Journal of Computer-Assisted Learning, 127-141. doi: 10.1111/jcal.12473

Resources for more learning

Maximizing learning without sacrificing the fun: Stealth assessment, adaptivity, and learning supports in educational games
Valerie Shute, et al. journal of Computer-Assisted Learning (2020)
myweb.fsu.edu/vshute/pdf/JCAL2020.pdf