International Society of the Learning Sciences

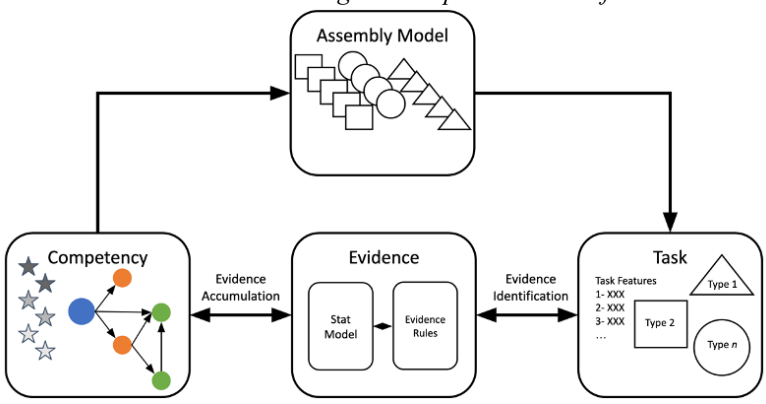
Stealth Assessment: A Systematic Review of the Literature

Abstract: Stealth assessment is embedded within technology-rich environments, like games, to measure and support learning. In this paper, following our review of a range of stealth-assessment studies, we discuss how stealth assessment is being used and where it may be headed. We identified 93 relevant studies consisting of 41 journal articles, 27 conference papers, 14 book chapters, 10 dissertations, and 1 book. These studies included participants ranging from third grade students to adults. We briefly discuss our findings in this paper.

Introduction

Stealth assessment (Shute, 2011) uses methods and technologies to collect and analyze learners' interaction data and make real-time inferences of learning based on the data. Digital learning environments employing stealth assessment can help researchers to accurately assess learners' competencies, and adapt the learning environment to fit learners' needs (Shute & Rahimi, 2017). Such adaptivity is closely linked to learning, engagement, and motivation theories (e.g., Csikszentmihalyi, 1990; Deci & Ryan, 2012; Vygotsky, 1978). The backbone of stealth assessment is evidence-centered design (ECD; Mislevy et al., 2003) which provides the requisite models for the system. There are four core models (see Figure 1). First, the Competency Model (CM) defines the set of knowledge and skills of interest, along with their sub-facets (unobservables) and their relationships to each other. When defining the competency model, researchers respond to the question of *what* to assess. Second, the Evidence Model (EM) identifies appropriate indicators (observables) in the game that provide evidence for the CM variables via statistical linkages (i.e., the statistical model). When defining the EM, researchers answer the question of *how* to assess. Third, the Task Model (TM) involves the creation of various task types that can elicit the evidence needed for the evidence model. One can think of a TM as a template from which a game developer can instantiate as many instances as needed. Finally, the Assembly Model (AM) allows researchers to arrange various tasks together with various difficulty levels, sufficient per competency, to be delivered to the learners. Moreover, the AM includes rules for adaptivity, personalization of the learning supports and other features of the game environment. Researchers respond to the question of *how much to assess* by defining the AM.

Figure 1
The four core ECD models used in the design and implementation of stealth assessments.



After more than a decade of research using this method to assess and support various competencies, across different learners and in various settings, in this in-progress study, reviewed how stealth assessment is being used and where it may be headed. We address the following research questions:

(1) What are the publication trends and contexts of research that use stealth assessment?
(2) What is the purpose of the stealth assessment research?
(3) What types of validity are more common in stealth assessment studies?

Method

To address the research questions, we conducted a systematic review using the updated PRISMA (Page et al., 2021) method. To provide comprehensive coverage of the literature, two search phases were conducted: database and reference search. In the database search, the following search string was created: "stealth assessment" OR "game-based assessment" OR "embedded assessment" OR "evidence-based assessment" OR "computer-based

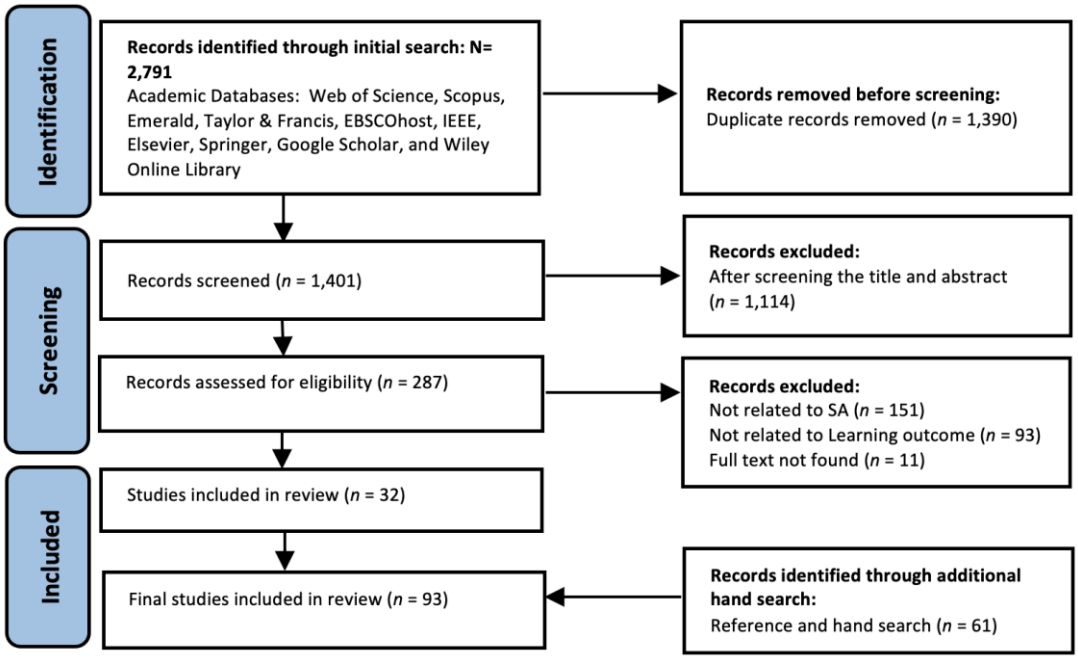
assessment for learning" OR "evidence-centered design" AND ("game" OR "online-learning" OR virtual reality") OR "stealth assessment" AND ("game" OR "online-learning" OR virtual reality"). We searched through the following databases: Web of Science, Scopus, Emerald, Google Scholar, Taylor & Francis, EBSCOhost, IEEE, Elsevier, Springer, and Wiley Online Library.

The inclusion criteria were as follows: 1) the studies should be published between 2004-2022; 2) the language of the publication should be in English; 3) stealth assessment should be the focus of the study; 4) the studies should target learning outcomes. Following the PRISMA guidelines, we included 32 studies for review. Finally, through hand searching within various papers' references and conducting additional targeted search (e.g., identifying studies from scholars who we knew were employing stealth assessment methods), we included an additional 61 studies for review, reaching our total number of studies to 93 (see Figure 2).

To extract the data, we created a spreadsheet to collect and summarize focal features per study. At this stage, we entered the information of each included paper in separate columns (e.g.,, title, manuscript type, field of study, name of the game, statistical modeling approach, target audience, sample, etc.). In this study, both qualitative and quantitative analyses were used. First, the basic features of our dataset provided a descriptive analysis of the papers. Second, a content analysis method was conducted to obtain an in-depth understanding of each included study to examine our research questions. Two researchers independently extracted the data from each study and met to resolve their disagreements in an iterative process.

Figure 2

PRISMA flow diagram detailing the steps in the identification and screening of sources.



Results

Addressing RQ1 (i.e., the publication trends of stealth assessment), we identified 93 studies (41 journal articles; 27 conference papers; 14 book chapters; 10 dissertations; 1 book). These studies included participants ranging from third grade students to adults. The competencies that were assessed in those studies (see a sample of studies in Table 1) included: (a) hard-to-measure competencies such as creativity, persistence, problem solving, computational thinking, risk taking, safety and emergency readiness; and (b) content knowledge and skill acquisition, such as mathematics, physics, genetics, geometry, reading and writing, and ratio and proportional reasoning. The studies included in this review came from various fields of study including computer science, educational technology, learning sciences, bioengineering, and applied mathematics. Moreover, 81 studies used or designed a game, while 11 studies used a simulation or an immersive learning environment to assess their competency of interest. Regarding RQ2 (i.e., the purpose of stealth assessments in each study), so far, we were able to categorize 60 studies into three categories: (a) validational studies (n = 42); (b) studies that used the stealth assessment estimates for purposes of providing adaptivity or feedback to students (n = 4); and studies

that only discussed the design phase of a stealth assessment (n = 14). This result suggests that most of the current stealth assessment studies are validational. Finally, regarding RQ3 (i.e., the type of validity approaches in stealth assessment studies), from the 60 studies that we coded, most of the studies (n = 24) used a *convergent* validity approach (i.e., correlational analysis between the stealth assessment estimates and the external measures); 18 studies used other types of validation methods (i.e., predicting the posttest and classifying accuracy). The remaining studies did not specify any validational measures.

Table 1

Selected sample of the 93 studies that used stealth assessment.

First Author	Year	Field	Game/LE	Competency	EL	N	Use
DeKosier, M. E.	2012	Health	Zoo U	Social Skills	3‴ & 4‴ g	187	V
Shute, V. J.	2013	ET / LS	PP	Physics	8‴ & 9‴ g	154	V
DiCerbo, K. E.	2014	ET / LS	Poptropica	Persistence	6 to 14 y	892	V
Snow, E., L.	2015	ET / LS	iSTART-2	Students' Agency	College	70	V
Kim, K.	2016	ET / LS	Semideus	Fraction	6‴ g	51	V
Shute, V. J.	2016	ET / LS	Use Your Brainz	Problem solving	7‴ g	47	V
Snow, E., L.	2016	ET / LS	iSTART-2	Self-explanation	H	40	V
Antoniou, P. E.	2017	Health	Serious talk	Tech acceptance	Adults	21	V
Akram, B.	2018	CS/AI	ENGAGE	Problem solving	M	244	V
Georgiadis, K.	2019	ET / LS	abcdeSIM	Medical Caring	Adults	267	V
de-Juan-Ripoll, C.	2020	BE	Spheres & Shield	Risk Taking	Adults	58	V
Shute, V. J.	2020	ET / LS	PP	Physics	9‴ to 11‴ g	263	V
Chen, F.	2020	ET / LS	Raging Skie	Weather Phenomena	5‴ g	460	V
Henderson, N.	2020	CS/AI	Geniventure	Genetics	M & H	462	V
Shute, V. J.	2021	ET / LS	PP	Creativity	9‴ & 8‴ g	167	V
Gupta, A.	2021	CS / AI	Crystal Island	Microbiology	M	119	V

Note. EL = Educational Level; ET = Educational Technology, LS = Learning Sciences, CS = Computer Sciences, AI = Artificial Intelligence, AM = Applied Mathematics, AS = Applied Sciences, BE = Bioengineering, PP = Physics Playground (Shute et al, 2019), PS = Problem Solving, CA = Cyberbullying Awareness, E = Elementary School, M = Middle School, H = High School, V = Validation, and D = Design.

Discussion & Conclusion

In this systematic review, we included 93 studies that have used stealth assessment to assess a competency within a digital environment, typically a game. Our findings indicate that studies that have used a stealth assessment methodology focused on a diverse set of competencies, included a range of target audiences, and came from multiple fields of studies. This finding shows that this methodology has been adopted and adapted by researchers in multiple contexts to assess and, in some cases, support learning. Although some studies included uses of the stealth assessment estimates in real time (e.g., for adaptation or providing personalized feedback to the learners), most of the studies are still at the validation stage where they describe the design of a stealth assessment for a particular competency using a validational approach (e.g., convergent validity) to indicate the accuracy of their assessment. After more than a decade of work in this area, it seems that stealth assessment can move to a new phase which is using the stealth assessment estimates to enhance learning (e.g., by adaptivity or personal feedback and support). Finally, our results indicate that most of the studies used a game as the vehicle in which to embed their stealth assessment. However, eleven studies used other learning environments. Despite the common understanding about stealth assessment, it is not only bound to games. We hope to see more studies that use this methodology in advanced, technology-rich learning environments in the future. As indicated earlier, this is an in-

progress project, and we will have a more complete dataset by the time of the conference with the aim to report findings on additional research questions (e.g., related to misconceptions about stealth assessment, innovations to optimize the methodology), and we will provide a meta-analysis of the convergent validity correlations reported across studies.

References

Akram, B., Min, W., Wiebe, E., Mott, B., Boyer, K. E., & Lester, J. (2018, July). Improving stealth assessment in game-based learning with LSTM-based analytics. *International Conference on Educational Data Mining*.

Antoniou, P. E., Siountas, A., Zilidou, V. I., & Bamidis, P. D. (2017). Virtual scenarios for stealth assessment of the elderly: Perceptions and acceptance of technology-based health and wellness interventions. *2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)*, 696–701.

Chen, F., Cui, Y., & Chu, M.-W. (2020). Utilizing game analytics to inform and validate digital game-based assessment with evidence-centered game design: A case study. *International Journal of Artificial Intelligence in Education*, *30*(3), 481–503. https://doi.org/10.1007/s40593-020-00202-6

Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. Harper and Row.

Deci, E. L., & Ryan, R. M. (2012). Self-determination theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology, Vol. 1* (pp. 416–436). Sage Publications Ltd. https://doi.org/10.4135/9781446249215.n21

DeRosier, M. E., CraigAshley, B., & Sanchez, R. P. (2012). Zoo U: a stealth approach to social skills assessment in schools. *Advances in Human-Computer Interaction*, 1–7. https://doi.org/10.1155/2012/654791

DiCerbo, K. E. (2014). Game-Based assessment of persistence. *Journal of Educational Technology & Society*, *17*(1), 17–28. JSTOR.

de-Juan-Ripoll, C., Soler-Domínguez, J. L., Chicchi Giglioli, I. A., Contero, M., & Alcañiz, M. (2020). The spheres & shield maze task: A virtual reality serious game for the assessment of risk taking in decision making. *Cyberpsychology, Behavior, and Social Networking*, *23*(11), 773–781. https://doi.org/10.1089/cyber.2019.0761

Georgiadis, K., Faber, T., & Westera, W. (2019). Bolstering Stealth Assessment in Serious Games. In A. Liapis, G. N. Yannakakis, M. Gentile, & M. Ninaus (Eds.), *Games and Learning Alliance* (pp. 211–220). Springer International Publishing. https://doi.org/10.1007/978-3-030-34350-7_21

Gupta, A., Carpenter, D., Min, W., Rowe, J., Azevedo, R., & Lester, J. (2021). Multimodal multi-task stealth assessment for reflection-enriched game-based learning. *MAIED@ AIED*, 93–102.

Henderson, N., Kumaran, V., Min, W., Mott, B., Wu, Z., Boulden, D., Lord, T., Reichsman, F., Dorsey, C., Wiebe, E., & Lester, J. (2020, July). Enhancing student competency models for game-based learning with a hybrid stealth assessment framework. *International Educational Data Mining Society*. https://eric.ed.gov/?id=ED607823

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.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Systematic reviews*, *10*(1), 1-11.

Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. In S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 503–524). Information Age Publishers.

Shute, V., Almond, R., & Rahimi, S. (2019). *Physics Playground* (1.3) [Computer software]. https://pluto.coe.fsu.edu/ppteam/pp-links/

Shute, V. J., & Rahimi, S. (2017). Review of computer-based assessment for learning in elementary and secondary education: Computer-based assessment for learning. *Journal of Computer Assisted Learning*, *33*(1), 1–19. https://doi.org/10.1111/jcal.12172

Shute, V. J., & Rahimi, S. (2020). Stealth assessment of creativity in a physics video game. *Computers in Human Behavior*, *116*, 106647. https://doi.org/10.1016/j.chb.2020.106647

Shute, V. J., Rahimi, S., Smith, G., Ke, F., Almond, R., Dai, C., Kuba, 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*, *37*(1). https://doi.org/10.1111/jcal.12473

Shute, V. J., Ventura, M., & Kim, Y. J. (2013). Assessment and Learning of Qualitative Physics in Newton's Playground. *The Journal of Educational Research*, *106*(6), 423–430. https://doi.org/10.1080/00220671.2013.832970

Shute, V. J., Wang, L., Greiff, S., Zhao, W., & Moore, G. (2016). Measuring problem solving skills via stealth assessment in an engaging video game. *Computers in Human Behavior*, *63*, 106–117.

Snow, E. L., Likens, A. D., Allen, L. K., & McNamara, D. S. (2016). Taking control: Stealth assessment of deterministic behaviors within a game-based system. *International Journal of Artificial Intelligence in Education*, *26*(4), 1011–1032. https://doi.org/10.1007/s40593-015-0085-5

Vygotsky, L. S. (1978). *Mind in society: The development of higher mental processes*. Harvard University Press.