

# Game-Based Assessment of School Readiness Domains of 3-8-year-old-children: A Scoping Review

Stephen Amukune<sup>1,2</sup> , Karen Caplovitz Barrett<sup>3</sup>  and Krisztián Józsa<sup>1,4</sup> 

<sup>1</sup>Institute of Education, University of Szeged, Hungary

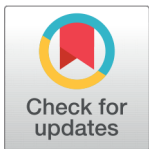
<sup>2</sup>School of Education, Pwani University, Kenya

<sup>3</sup>Department of Human Development & Family Studies, Colorado State University, USA

<sup>4</sup>Institute of Education, Hungarian University of Agriculture and Life Sciences, Hungary

## ABSTRACT

Precise assessment of school readiness is critical because it has practical and theoretical implications for children's school and life success. However, school readiness assessment mainly relies on teacher reports and a few direct evaluations requiring a trained examiner. Studies indicate that 80% of games and apps target preschool children and education, suggesting that apps are familiar and fun for this age group. Previous reviews have focused on these apps' training capability but not on their assessment of school readiness. This Scoping review examines 31 studies published from 2011-2019. The Evidence Centred Design (ECD) framework was used to evaluate game-based assessment (GBA) suitability to assess school readiness domains. Results show that it is possible to assess school readiness using GBA. Most studies assessed cognitive domains in school settings and adopted an external assessment of the tasks. However, most studies only evaluated one competency, and few intervention strategies targeted the enhancement of school readiness. Besides, few studies followed the ECD framework strictly. Implications include expanding the assessment to other school readiness domains with a real-time inbuilt assessment that conforms to the ECD framework. GBA provides a new approach to assess school readiness outside or inside the school settings in this online era.



**Received** 2021-03-10

**Revised** 2021-03-11

**Accepted** 2021-06-03

**Published** 2022-01-15

**Corresponding Author**

Stephen Amukune,  
[s.amukune@pu.ac.ke](mailto:s.amukune@pu.ac.ke)

Petőfi Road, No. 30-34, 6722,  
Szeged, Hungary

**DOI** <https://doi.org/10.7821/naer.2022.1.741>

**Pages:** 146-167

**Funding:** National Research, Development and Innovation Office, Hungary (Award:NKFI K124839)

Distributed under  
CC BY-NC 4.0

**Copyright:** © The Author(s)

**Keywords** EVIDENCE CENTRED DESIGN, GAME-BASED ASSESSMENT, PRESCHOOL EDUCATION, SCHOOL READINESS, SCOPING REVIEW

## 1 INTRODUCTION AND BACKGROUND

Children who have the necessary foundational skills and attributes for school readiness are better positioned to succeed in school and beyond (Russo, Williford, Markowitz, Vitiello, & Bassok, 2019). On the other hand, low school readiness is linked to later unemployment, criminality, and academic failure (Burchinal, Magnuson, Powell, & Hong, 2015). Therefore, precise assessment of school readiness is critical (Barrett, Józsa, & Morgan, 2017) because of the following reasons. First, at early ages, assessment of child competencies is usually

## OPEN ACCESS

formative. If correctly assessed, it can provide an opportunity for effective individualised intervention. Second, an accurate evaluation will provide parents and teachers with information to decide whether or not the child should delay school entry. Third, a reliable and meaningful assessment of school readiness helps understand a particular programme or curriculum (Suleiman, Arslan, Alhajj, & Ridley, 2016).

School readiness, to a great extent, depends on the method of assessment used. Most of these tests and reports are pencil and paper-based. These reports' value depends on the quality of information teachers, examiners, and parents can and do provide (Li, Fan, & Jin, 2019). To alleviate this challenge, a form of direct assessment that could be administered without intensive training of examiners is a computerised self-administered assessment. There has been much effort to develop technology-based assessment (Csapó, Molnár, & Nagy, 2014; Neumann & Neumann, 2019). This effort has historically been complex due to young children's low computer skills and developmental level, raising validity issues (Csapó et al., 2014; Suleiman et al., 2016). This challenge has significantly been reduced with the introduction of tablets with touchscreen technology that are highly portable, with digital measurement abilities and engaging to children (Semmelmann et al., 2016). However, children have been the most targeted group for digital games on computer tablets (Chaudron et al., 2015), suggesting that game-based assessments on computer tablets might be an effective means of directly assessing school readiness skills in young children.

It is estimated that there are more than 1000 computer-assisted interventions for children (Axelsson, Andersson, & Gulz, 2016). Moreover, about 80% of the Apple Store's best-selling apps are for pre-schoolers or education (Papadakis & Kalogiannakis, 2017). Given this heavy consumption of video games and apps, parents and teachers have consistently inquired about their effects on young children (Behnamnia, Kamsin, Ismail, & Hayati, 2018). Studies indicate that playing games is positively related to the development of cognitive skills, motivational and academic performance (e.g. Chan et al., 2017), and attention (e.g. Godwin et al., 2015). However, the application of tablets and these apps to assessing learning in children is less known (Carson, 2017), despite the potential advantages of this approach (Neumann & Neumann, 2019). Previous reviews have focused on these apps' training capability but not on assessing school readiness. The paucity of information about the efficacy of game-based tablet assessments makes it unclear whether such assessments should be recommended to support teachers and parents making critical decisions about children's education.

This scoping review aimed to establish whether the computer games, apps or game-like features used as assessment tools (from here on referred to as Game-Based Assessment (GBA)) are employed in the assessment of the following child school readiness domains: (1) cognition and general knowledge (2) approaches to learning; (3) physical well-being and motor development; (4) social and emotional development; and (5) language development (Kagan, Moore, & Bredekamp, 1995; Sabol & Pianta, 2017). The review is organised as follows. Section 2.0 presents the theoretical framework underpinning Game-Based Assessments and the rationale for doing a scoping review of Game-Based Assessment. Section 3.0 presents the review methodology, followed by the scoping review results in Section 4.0.

Finally, discussion and conclusions will be presented in Section 5.0.

## 1.1 Framework and Rationale

### 1.1.1 Theoretical Framework

The Evidence-Centered Design (ECD) (Mislevy, Steinberg, Almond, & Lukas, 2006) is instrumental in guiding the design of GBA. ECD belongs to a category of assessment frameworks referred to as principled assessment designs. These frameworks require evidence throughout the design, development and implementation, and their validity evidence is more robust than conventional assessments. Other similar frameworks include Cognitive design systems, Assessment engineering, Berkeley Evaluation and Assessment Research (BEAR) Center assessment system, and Principled design for efficacy. Among all these frameworks, the most widely recommended, implemented and researched is the ECD. Additionally, the other frameworks are often used in large item banks, secondary and undergraduate education (Ferrara, Lai, Reilly, & Nichols, 2017). In fact, the latest framework, the Principled Design for Efficacy, is an adaptation of ECD that is primarily used in the assessment of summative end of year exams (Nichols, Kobrin, Lai, & Koepfler, 2016). ECD has also been successfully implemented to manage game-based and simulation assessments challenges (Kim, Almond, & Shute, 2016).

The ECD framework asks fundamental questions common in any assessment: “what, where, and how are we measuring, and how much do we need to measure” (Kim et al., 2016, p. 3). ECD answers these questions in four models. First is the student or competency or proficiency model in the latest publications (e.g. Almond et al., 2015) that stipulates the competencies and other student attributes that we want to measure, in this case, school readiness domains. Second is the task model that indicates the set of activities that the learner will undertake to demonstrate those domains. The task model answers the question where (during what activities) do we measure the competencies? The third is the evidence model that connects the student’s activities to the competence that we wanted to know about the learner. This model provides specific metrics to answer the question: How do we measure the domains based on the task completion of activities representing the construct under investigation? The evidence model is composed of two components, the scoring model and the measurement model. The connection between work products from learner activities and evidence from students’ performance makes the assessment valid (DiCerbo, 2017). The competence, task and evidence model is also referred to as the Conceptual Assessment Framework (CAF) (Mislevy et al., 2006). The fourth component is the Assembly model. This model stipulates how the CAF models will work together to generate enough evidence to measure the construct under investigation (Almond, Mislevy, Steinberg, Yan, & Williamson, 2015). There are proposals to expand ECD by incorporating learning into the four models to form an expanded ECD or e-ECD, although they have not been actualised (Arieli-Attali, Ward, Thomas, Deonovic, & Davier, 2019).

### 1.1.2 Why Game-Based Assessment?

Children are naturally playful, and therefore games are crucial in their development (Bento & Dias, 2017). With the advent of GBA, teachers and researchers can assess knowledge and various skills and abilities that are difficult to determine using traditional assessment methods by integrating them into those games. Players also experience motivation, behaviour change and deep engagement in these games, providing more reasons for this medium's success (Chan et al., 2017). GBA focuses on collecting, analysing, and extracting information from data obtained while playing serious games. This concept is borrowed from Educational Data Mining (EDM), also known as Learning Analytics (Alonso-Fernández, Calvo-Morata, Freire, Martínez-Ortiz, & Fernández-Manjón, 2019).

The use of games in assessment has many advantages. First, they can adopt a real-life scenario that the learner can relate to, thus increasing the learner's motivation, assessment accuracy and reducing dropout rates and test anxiety (Barab, Gresalfi, & Ingram-Goble, 2010). Secondly, touch screen technology emulates children's constructivist mode of learning (Orfanakis & Papadakis, 2014). A study across Britain, Australia, New Zealand and the US reported that 2-5-year-olds could operate apps better than biking or shoe lacing (Grose, 2013). Besides, many computer games share some common characteristics with academic assessments: evidence identification as proof of knowledge and its accumulation; presentation and finalising of activities to accomplish some goal, and presentation of another, usually more challenging activity once one completes the previous activity (Mislevy, Behrens, Dicerbo, & Levy, 2012). Usually, to play a game, a player must apply various competencies or other attributes (e.g. creativity, problem-solving, persistence, and collaboration), so success in playing it could provide a measure of those domains and other learning outcomes (Caballero-Hernández, Palomo-Duarte, & Doderó, 2017). On the other hand, Klerk, Veldkamp, and Eggen (2015) reported two shortcomings of GBA. Firstly, the interaction of sound, contrasting colours and graphics can affect a child's concentration, especially in a high stakes assessment. Secondly, the amount of process data generated during a game is enormous, making it challenging to identify the elements under investigation.

There are three types of GBA. First is scoring game-related success measures, such as obstacles overcome, targets achieved, or the time taken to complete a task (Chaudy, Connolly, & Hainey, 2013). The second is an external assessment that uses tools such as pre-post questionnaires, debriefing interviews, essay and knowledge maps, and test scores from multiple-choice questions (Caballero-Hernández et al., 2017). The third is an embedded assessment based on player response data, such as the use of click streaming or log file analysis and information trails (Ifenthaler, Eseryel, & Ge, 2012). These assessment types are also integrated into GBA in six main approaches: adopting assessment models, monitoring states, quests, non-invasive assessment, quizzes, and peer assessment (Chaudy et al., 2013).

## 1.2 Related Work

To better establish the need for this scoping review, we first set out to determine whether there are other similar reviews in the same area. To identify any relevant, suitable previous studies in the same field, we searched in the English language, "Game-based assessment"

OR “Game Learning Analytics” AND “systematic mapping” OR “systematic literature” OR “scoping review” AND “children” OR “childhood” OR “school readiness domains” OR “skills” OR “knowledge” OR “abilities” in the following databases: Web of Science, ScienceDirect, SpringerLink, Scopus, ERIC, PsycINFO, and IEEE Xplore. The search resulted in ten different systematic reviews that at first appeared close to the current review. Only two systematic reviews had very close objectives to the present scoping review. One review by Caballero-Hernández et al. (2017) focused on skill assessment in learning experiences based on serious games. However, this review was different from the current one because it did not address the evaluation of school readiness domains but concentrated on skills assessed by serious games alone for formative or summative assessments in secondary schools. The second review by de Klerk et al. (2015) evaluated the psychometrics of Simulation-Based Assessment in higher education competencies, a different type of assessment and a very different age group compared to the current study. To be informed of subsequent similar reviews, we registered alerts on similar topics with google alerts, ScienceDirect and Web of Science.

## 2 METHODS

A scoping review is a type of research synthesis that “aims to map the literature on a particular topic or research area and provide an opportunity to identify key concepts; gaps in the research; and types and sources of evidence to inform practice, policymaking, and research” (Daudt et al., 2013, p. 8). We adopted the following steps in conducting the scoping review: (1) identifying the research questions; (2) identifying relevant studies; (3) selecting studies; (4) charting the data; and (5) collating, summarising, and reporting the results (Arksey and O’Malley, 2005, p. 22).

### 2.1 Identifying the Research Questions

This scoping review aimed to identify how GBA has been implemented in the assessment of school readiness domains. The following research questions guided this scoping review:

- RQ1: What are the main characteristics of studies in GBA of school readiness domains? Which countries are involved? Which knowledge, skills or abilities related to school readiness are assessed? Are these assessments done in schools or outside schools?
- RQ2: What measurement type and instruments does each assessment adopt? What are the psychometric properties of these tools?
- RQ3: What type of performance data analyses are employed by these studies? Are these analyses on the process or product data?
- RQ4: How is the outcome of the GBA used to enhance the development of school readiness domains?

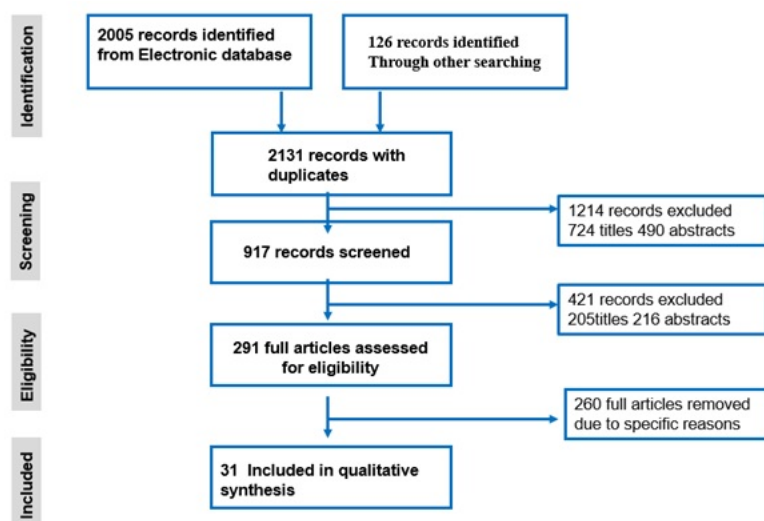
The answer to these research questions will help advise teachers, parents and game developers on the level at which GBA can be implemented to assess school readiness, the gaps

available and how to seal these gaps.

## 2.2 Identifying Relevant Studies

For inclusion and exclusion, we adopted a similar procedure by Caballero-Hernández et al. (2017, p. 46). We adopted the following exclusion criteria : (i) Out of scope: articles earlier than 2010 were excluded since serious games began effectively after 2010 (Ifenthaler et al., 2012); (ii) Unsupported language: languages other than English; (iii) Off matter: textbooks on general assessment and test theories but not a game-based assessment; (iv) Duplicated: article already included from another database; (v) Off topic: assessments of other subjects or ages other than children. We followed a systematic search in the following databases, PsycINFO, ERIC, SCOPUS, ACM Digital library, Science Direct, Web of Science, IEEE Xplore, and SpringerLink. We designed a Boolean search as follows; “Game-based assessment” OR “game learning analytics” OR “children apps” AND “validation” OR “evaluation” AND “school readiness domains: “cognitive” OR “approaches to learning” OR “language” OR physical OR “socio-emotional” OR “numeracy” OR “science” OR “knowledge” OR “skills” OR “abilities” OR “education” AND “children” OR “childhood”. Figure 1 shows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram (Moher, Liberati, Tetzlaff, Altman, & Group, 2009) that we followed during this review.

Springer Link had the highest number of studies and conference proceedings in GBA, followed by Web of Science, as shown in Table 3. Most of the studies identified focused more on middle school and above.



**Figure 1** Flowdiagram of the study selection process

## 2.3 Selection of the Studies

All the selected databases were collected together in Zotero Electronic Referencing Management software. Within this database, we selectively implemented the inclusion and exclusion criteria. We coded all the articles into the school readiness domains (Kagan et al., 1995; Sabol & Pianta, 2017). This coding gave a broad range of articles to search and opportunities for further investigation. Table 1 shows the results of the studies that met the inclusion and exclusion criteria.

**Table 1** Selected empirical studies and journal avenues

Item no.	Study	Country study conducted	Journal source
1	Hsu, Tsai, and Liang (2011)	Taiwan	Journal of Science Education and Technology
2	Bottino, Ricerche, Ott, and Tavella (2014)	Italy	International Journal of Game-Based learning
3	Craig, Derosier, and Watanabe (2015)	USA and Japan	Games for Health Journal
4	Godwin et al. (2015)	USA	International Journal of Gaming and Computer-mediated Simulations
5	Geurts et al. (2015)	Belgium	Advances in Game-based Learning
6	Agudo, Rico, and Sanchez (2016)	Spain	Journal of Universal Computer Science
7	Moyer-Packenham et al. (2016)	USA	Mathematics Research Journal
8	Ninaus, Kiili, McMullen, and Moeller (2016)	Finland	Games and Learning Alliance
9	Gresalfi, Rittle-Johnson, Loehr, Nichols, and Mccracken (2016)	Sweden	AERA Online
10	Axelsson et al. (2016)	USA	Journal of Educational Psychology
11	Aragon-Mendizabal, Aguilar-Villagran, Navarro-Guzman, and Howell (2017)	Mexico	Anales de Psicología
12	Jozsa, Barrett, and Morgan (2017)	Hungary	Electronic Journal of Research in Educational Psychology
13	Ow and Tan (2017)	Malaysia	IEEE Conference on e-Learning, e-Management and e-Services (IC3e)
14	Chiu and Hsieh (2017)	Taiwan	EURASIA Journal of Mathematics, Science & Technology Education
15	Puolakanaho and Latvala (2017)	Finland	Human Technology
16	Loachamín-Valencia, Juan, Méndez-López, and Pérez-Hernández (2018)	Spain	Artificial Intelligence and Bioinformatics
17	Behnamnia et al. (2018)	Malaysia	Social-Informatics and Telecommunications Engineering
18	DeRosier and Thomas (2018)	USA	Journal of Applied Developmental Psychology
19	Lee et al. (2018)	USA	Frontiers in Paediatrics
20	Mironcika et al. (2018)	Netherlands	Proceeding of the Twelfth International Conference on Tangible, Embedded and Embodied Interaction
21	Obradovic, Sulik, Finch, and Tirado-Strayer (2018)	USA	Journal of Applied Developmental Psychology
22	Miller (2018)	Canada	International Journal of STEM Education
23	Mainieri, Azevedo, Braga, and Omar (2018)	Brazil	Latin American Conference on Learning Technologies (LACLO)
24	Cohrsen and Niklas (2019)	Australia	Assessment for Effective Intervention
25	Ford, Kim, Brown, Aber, and Sheridan (2019)	Lebanon-Niger	Research in Comparative and International Education
26	Neumann and Neumann (2019)	Australia	International Journal of Research & Method in Education

*Continued on next page*

Table 1 continued

Item no.	Study	Country study conducted	Journal source
27	Bhavnani et al. (2019)	India	Global Health Action
28	Shih, Kuo, and Lee (2019)	Taiwan	Educational Psychology
29	Rauschenberger, Lins, Rousselle, Hein, and Fudickar (2019)	Germany	ACM International Conference proceeding
30	Willoughby, Piper, Oyanga, and King (2019)	Kenya	Developmental Science
31	Xin, Jian, and Zhiyu (2018)	China	Acta Psychologica Sinica

## 2.4 Data Analysis

The authors coded all the articles according to the school readiness domains. After coding, we pooled studies that assessed similar school readiness domains together. Where there was a lack of consensus among the authors, we sought a third opinion. Then we checked the article's suitability for assessing the domain claimed guided by the Evidence Centred Design (ECD: Mislevy et al., 2006) framework. The authors discussed their suitability based on the evidence provided for studies that did not categorically state the models as stipulated by the ECD framework.

## 3 RESULTS

### 3.1 General Characteristics of the Empirical Studies on GBA

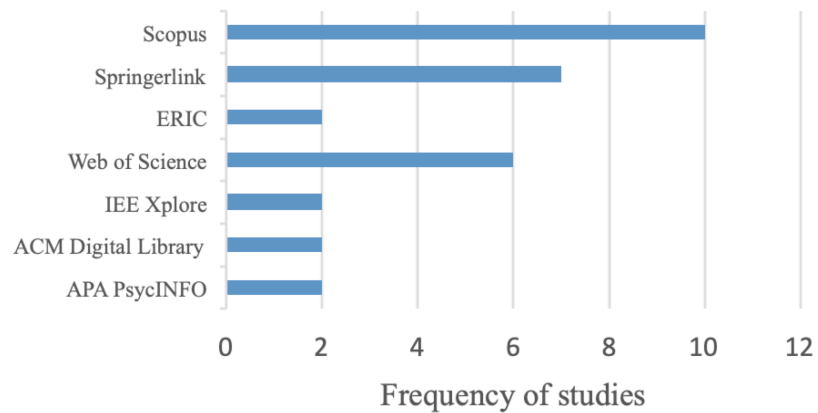
Following the inclusion and exclusion criteria, we included 31 articles for the review. Scopus had the highest number of articles ( $n = 11$ ). PsycINFO, ACM, IEEE Xplore and ERIC had the least –two studies each, as shown in Figure 2. However, some databases shared some articles and to avoid duplication; the articles were registered once. The studies came from different parts of the world. Europe had the highest representation with  $n=10$  studies, followed by North America  $n = 8$ , Asia  $n = 8$ , Australia and South America  $n = 2$  and Africa  $n = 1$  as shown in Table 1. The most represented country was the USA ( $n = 7$ ), followed by Taiwan ( $n = 3$ ), then Australia, Finland, Malaysia and Spain had  $n = 2$  studies each.

#### 3.1.1 Distribution of the Empirical Studies by Year

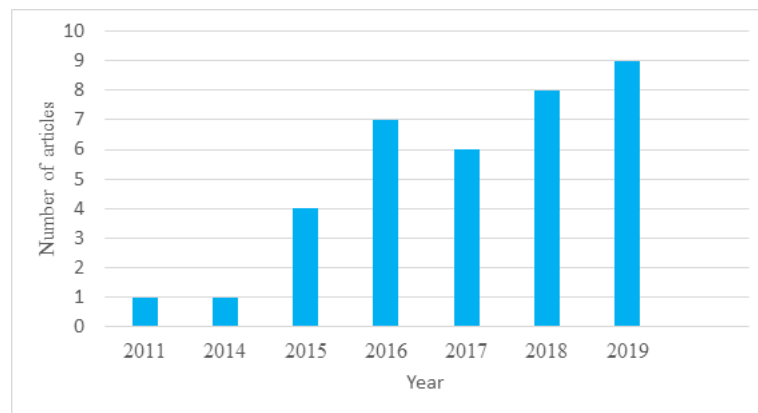
During 2011-2019, the number of empirical studies on GBA of school readiness increased ninefold, as shown in Figure 3. In 2012 and 2013, no studies were found that directly targeted school readiness domains for children of 3-8 year old. Nevertheless, there is an upward trend of studies in addressing school readiness assessment.

We also investigated the settings where these studies were conducted. Only one study by Bhavnani et al. (2019) was done outside school settings in the children's homes. All the other studies were conducted in school settings but offline and were not formalised in the school curriculum as preferred assessment methods. GBA has an opportunity to provide a different form of assessment, even in homes and online, especially when schools are not accessible in the event of a pandemic.





**Figure 2** Distribution of studies in the selected databases



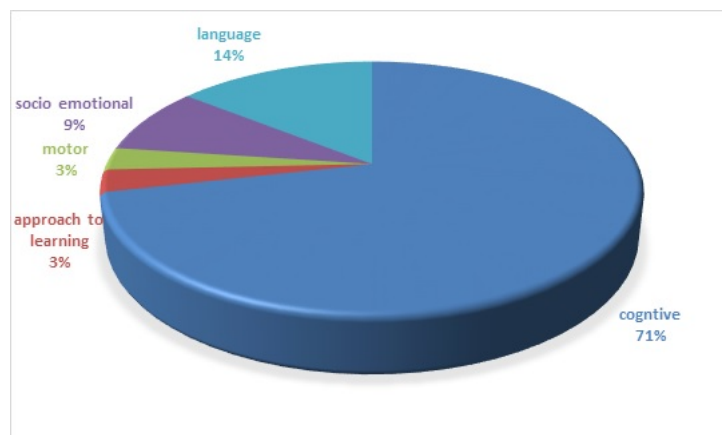
**Figure 3** The trend of studies in GBA of school readiness domains from 2011-2019.

### 3.2 School Readiness Domains Assessed using GBA

We identified each school readiness domain considered in each study. Most empirical studies assessed cognitive domains ( $n = 25$ ). Cognitive competency refers to the ability of the child to process information. It is divided into two: subject-specific and general cognitive skills. The subject-specific cognitive domain is mostly supported by teaching and learning through a particular curriculum. The subject-specific cognitive skills featured in the empirical studies were arithmetic ( $n = 8$ ), reading or letter recognition ( $n = 5$ ), English ( $n = 1$ ) and science ( $n = 1$ ). On the other hand, general cognitive skills are not necessarily taught in a classroom situation, but they are essential in problem-solving (Suleiman et al., 2016). The studies in this review assessed, memory ( $n=1$ ), critical thinking ( $n=1$ ), attention ( $n=1$ ) and reasoning ( $n=2$ ). Five studies assessed executive functions, which involve working memory and cognitive switching, and could be construed as general cognitive skills, but they also involve inhibitory control and are sometimes viewed as approaches to learning.

Socio-emotional development was only measured in  $n = 3$  studies, with two studies evaluating the social aspect and one on emotions, as shown in Table 3 and Figure 4. Socio-emotional development involves children's understanding and regulation of emotions and behaviour and skills for interacting with others at school, all of which are important for participating effectively in classroom activities. Physical well-being and motor development featured in only one study that addressed fine motor skills. This domain is significant in early childhood since it facilitates learners' writing and manipulative play (e.g., constructing puzzles) during teaching and learning.

As mentioned earlier, executive functions can be viewed either as general cognitive skills or approaches to learning. If viewed as approaches to learning, there were six studies in total on this school readiness domain. However, only one GBA study was conceptualised to study approaches to learning Jozsa et al. (2017). Approaches to learning is an umbrella term for traits that help children learn. They include focus, enthusiasm, flexibility, persistence, and motivation. Recent studies have demonstrated the significance of approaches to learning on academic performance, perseverance when faced with challenging tasks, problem-solving creatively, and children's socio-emotional development (Hunter, Bierman, & Hall, 2018).



**Figure 4** School readiness domains assessed by GBA

### 3.3 Task Performance in GBA Studies

Based on the ECD framework (Mislevy et al., 2006), for each study, we coded whether it had a specific task for the school readiness domains and the nature of the evidence that was provided for that readiness domain. A task is defined as “a unit of activity that the student attempted, which produces a work product” (Kim et al., 2016, p. 4). In GBA, it can be a multiple-choice question, game level or, in some instances, a very complex interaction. How we measure that task constitutes the evidence model (Klerk et al., 2015). Most GBA offered three types of assessments of these tasks: external, embedded and game scoring during the game activity. The work products sometimes were captured as observed data, either process or a product of the task performance. The majority of the external assessments only

assessed the final product, but the embedded type produced both process and product data (see Table 3). All the studies  $N = 31$  had explicit tasks provided, although the number of tasks varied depending on the competency's nature. For external assessment, the students were provided with MCQ or interviews after the GBA. Three studies compared the effectiveness of GBA and traditional tests, Hsu et al. (2011), Loachamín-Valencia et al. (2018) and Neumann and Neumann (2019). These studies concluded that GBA is better than pen and paper formats. Other studies with intervention strategies had a similar conclusion, as shown in Table 3.

### 3.4 Analysis of Performance Data Based on the Tasks

Tasks performed by learners can either be analysed as a process or a final product. These different types of measures require different methods of analysis. Two branches of research have emerged in educational assessment on this front. The first is educational data mining techniques that explore the relationship between competency and performance-based tasks (Rupp, Gushta, Mislevy, & Shaffer, 2010). The second emerging branch is computational psychometrics that employs complex models such as Bayesian Networks that use performance data to make probabilistic decisions about learners' skills (e.g. Almond et al., 2015; Arieli-Attali et al., 2019). Some studies in the current review did not explicitly declare the method of performance data analysis they adopted. We investigated  $n = 12$  studies that purely adopted the embedded type of assessment. In this type of assessment, the log file registers the process and the various performances' product data. Of these 12 studies, only four studies (Craig et al. (2015), Lee et al. (2018), Puolakanaho and Latvala (2017) and Willoughby et al. (2019)) indicated the psychometric or statistical model they employed to score or process the performance data, as shown in Table 3. The other studies reported how the output was integrated into their data analysis strategies instead (e.g. Ford et al., 2019 indicated they adopted SAS).

Nevertheless,  $n = 8$  studies provided how the scoring was done, and  $n=10$  provided a model of the relationship between the competency to be assessed and the observed work product, as shown in Table 2. Ten studies adopted external assessment;  $n=8$  had external and embedded assessment types, while  $n=14$  studies assumed only embedded type of assessment. The most common external assessment tools were the multiple-choice questions and the interview. Most questionnaires targeted teachers and parents, while the one for children only had simple images to indicate whether they enjoyed the game or not. For the embedded type, the log file analysis was the most preferred method of assessment. Unfortunately, only three studies indicated the program they will deploy to analyse the log files: Microsoft excel, high-order DINA model for the problem-solving and scoring algorithm.

**Table 2** GBA Types and Psychometric Properties of the Studies

Item no.	GBA type	Tasks	Observed variable	Analyses	Psychometric evidence
1	external	MCQ test and interview on shadow	process	N	N

*Continued on next page*

*Table 2 continued*

Item no.	GBA type	Tasks	Observed variable	Analyses	Psychometric evidence
2	external and embedded	MCQ test on transversal skills	process	N	N
3	embedded	embedded test of six types of social skills	process/product	scoring algorithm	validated
4	in game and external	smile questionnaire	process	N	concurrent validity
5	external	phonological test	process	N	N
6	external	questionnaire and classroom observation	process	N	N
7	external	interview test(pre-post), quantitized videos	process	N	N
8	in-game and external	embedded	process and product	N	N
9	embedded	questionnaire	product	N	N
10	external	embedded test	product and process	scoring formula	N
11	embedded	early numeracy test	product	N	N
12	embedded	embedded measure of mastery motivation and EF	process	N	reliability and validity
13	embedded	embedded test on six critical thinking skills	process	N	N
14	external	questionnaire and embedded test	product	N	N
15	external and ingame	pencil and paper tasks and in-game scoring computer-based tasks	process and product	statistical model	correlation with other tasks
16	external and embedded	embedded and compared with traditional	process	r-software	concurrent validity
17	external and embedded	pen and paper; in-game scoring	product and process	N	validity
18	embedded	embedded algorithm scores student menu selections and in-game social behaviours	process	scoring algorithm	validity
19	embedded	Weschler tests and embedded cognitive TAG tests	process	bayesian nets	validity
20	embedded	embedded fine motor tasks	process	N	N
21	embedded	embedded EF tasks	process	response time analysis	reliability and validity
22	external	score test on numerical cast on a screen	product	N	N
23	external	questionnaire	process	N	validity
24	external and ingame	pre-post numerical tests	product	N	N
25	embedded	embedded EF tasks analysis	process	SAS	N
26	in-game scoring and external	completed receptive tests delivered via a tablet and pencil and paper	process	inbuilt scoring	reliability and validity
27	external	questionnaire and interviews	product	N	reliability and validity
28	embedded	problem-solving tasks measured by HO-DINA model	process and product	HO-DINA model	N
29	embedded	embedded telemetry/dyslexia	process	N	N

*Continued on next page*

*Table 2 continued*

Item no.	GBA type	Tasks	Observed variable	Analyses	Psychometric evidence
30	embedded	embedded EF tasks analysis	process	confirmatory factor analysis	reliability and validity
31	embedded	log file analysis and pre-post test	process	machine learning algorithm	N

Note: N - Not provided

### 3.5 Psychometric Evidence Provided by the Empirical Studies

We also investigated the Psychometric properties described by the studies of GBA. Of the articles selected, n =16 studies did not indicate any psychometric evidence in their studies. Studies that addressed general cognitive skills, mainly executive functions, had an elaborate test for psychometric properties (Ford et al., 2019; Obradovic et al., 2018; Willoughby et al., 2019). These studies all determined correlations with the existing tools for assessing executive functions to establish concurrent validity. Some studies (e.g. Obradovic et al., 2018) further reported the ecological and predictive validity of the game-based assessment tool.

### 3.6 School Readiness Intervention Strategies Adopted by GBA Studies

Studies with an intervention component adopted experimental or quasi-experimental designs to test the intervention's efficacy in school settings. Studies that adopted survey methods had a larger sample size compared to the ones that adopted experimental techniques. The smallest sample size was 5 participants, while the largest, which assessed executive functions as one of the general cognitive skills, had 1480 participants. However, of all interventions targeting 3-8-year-old children, only n = 8 studies aimed their intervention towards school readiness (Table 3). The n = 10 studies that addressed subject-specific cognitive skills such as science, arithmetic, and English aimed to solve the challenge of poor performance and attitude towards those subjects. However, support of individual learning of subject areas contributes positively to school readiness. Two studies (Rauschenberger et al. (2019) and Geurts et al. (2015)) aimed to identify the risk of developing dyslexia in children as they prepare to join the school and possible early intervention strategy to adopt.

**Table 3** Study designs and Intervention Strategies of the Studies

Item no.	Intervention	Domain assessed	Study design	Sample Size
1	Acquisition of scientific concepts regarding light and shadow	science	experimental	50
2	Triggering and sustaining students' reasoning and logical abilities	reasoning	quasi-experimental	60
3	N	social skills	correlational	497
4	N	sustained attention	correlational	27
5	detect a high risk for developing dyslexia	phonological	correlational	25

*Continued on next page*

Table 3 continued

Item no.	Intervention	Domain assessed	Study design	Sample Size
6	English acquisition	English-EFL	descriptive	35
7	N	arithmetic	mixed methods	100
8	Validate conceptual knowledge of fractions	arithmetic	quasi(pre-post)	54
9	Early learning of equivalence and ordering infractions	arithmetic	correlational	65
10	N	executive functions	correlational	161
11	Reduce the risk of developing problems learning mathematics	arithmetic	quasi-experimental	156
12	N	mastery motivation &EF	correlational	247
13	N	six critical thinking skills	correlational	20
14	Improve maths achievement and attitudes	fractions	quasi(pre-post)	100
15	Predict early reading skills-slow and normal	phonological	quasi(pre-post)	57
16	Auditory stimuli for assessing short-term spatial memory	spatial reasoning	correlational	35
17	Collaboration through social apps	social skills	quasi(pre-post)	32
18	N	socio-emotional skills	correlational	270
19	N	cognitive	correlational	40
20	Games can be introduced in smart toy to improve fine motor skills to improve handwriting	fine motor skills	correlational	30
21	N	executive functions	correlational	269
22	Increase engagement of learners in the teaching of maths	arithmetic	mixed methods	13
23	Practising the arithmetic operations of addition, subtraction and multiplication	arithmetic	mixed methods	20
24	Supporting children's engagement with mathematical ideas and numeracy development.	arithmetic	quasi(pre-post)	79
25	N	inhibitory control and working memory	correlational	850
26	Measure early literacy skills	the letter, word, and numeral skills	quasi(pre-post)	99
27	N	six cognitive skills	correlational	100
28	Performance of computational estimation in addition and subtraction	computational skills	correlational	729
29	Early detection of dyslexia	reading and writing	correlational	5
30	N	executive functions	correlational	1480
31	N	arithmetic	quasi(pre-post)	360

Note: N - Not provided; EFL - English as a foreign language; EF-Executive Functions

## 4 DISCUSSION

Despite the importance of school readiness, there has been a heavy debate regarding what it means to be ready for school (UNICEF, 2012). This theoretical understanding has implications for how school readiness will be assessed. Some studies recommended assessing readiness of school, community, family, and the child, while others focused on the child's readiness to learn or perform at the school level (Stein, Veisson, Oun, & Tammemäe, 2019). For this scoping review, we focused on child-centred school readiness and examined the domains stipulated by the National Educational Goals Panel (Kagan et al., 1995; Sabol & Pianta, 2017).

Many instruments have been developed to assess school readiness, both paper and pencil, and technology-based (Csapó et al., 2014). However, in this online era, the use of GBA

to assess children domains has received surprisingly little attention and yet over 80% of children's apps target pre-schoolers. Since children love games and most apps and video games are produced for children (Behnamnia et al., 2018), we sought to determine if studies on GBA have considered assessing these school readiness domains. To achieve this, we based our review on evidence centred design (Mislevy et al., 2006).

Over 70% of the studies we reviewed assessed cognition and general knowledge. This is due to overemphasis on intellectual factors that affect academic performance rather than non-intellectual factors such as approaches to learning (Li et al., 2019). The other competency frequently assessed but not widely featured in GBA is the socio-emotional domain with only two studies. Although socioemotional skills are malleable, their assessment has usually been done using the teacher's behavioural ratings, requiring training to interpret and suffering from psychometric challenges (DeRosier & Thomas, 2018). A performance-based assessment like a GBA could identify the socio-emotional needs of the children directly. Another cognitive competence that did not receive the attention it deserves was the language domain, with only 5 (14%) of the studies measuring this competency. There was only one study on fine motor development to prepare children to write in readiness for school. However, games that influence physical activity, popularly known as active video games or "exergames", have impacted intellectual skills, executive functions, and health outcomes (Merino-Campos & Castillo-Fernández, 2016).

To assess the school readiness domains, each study provided a series of tasks for the children. Ten studies (33%) provided tasks in a pre-post experimental design. This approach does not provide enough evidence to analyse what happens in the game itself. This design has been criticised for not analysing the complex process/performance data that can be used to inform 21st-century educational skills (Suleiman et al., 2016). To solve this "black box" issue, we need to measure GBA in real-time with automated scoring. One of the frequently employed analyses in such situations is Bayesian Networks. Bayesian networks allow one to make a probabilistic statement of latent variables under investigation based on the observed variables (Almond et al., 2015). This further supports learning by providing information about the processes underlying performance on the assessment, enabling other types of learner attributes to be measured and supported (Klerk et al., 2015; Suleiman et al., 2016). This type of assessment is interwoven into the fabric of the gameplay, such that children imagine they are enjoying playing a game, but in the background, complex skills and domains are assessed, which is referred to as stealth assessment (Kim et al., 2016). The videogame industry seems to have taken advantage of learning analytics or educational data mining by providing non-disruptive tracking methods that visualise the process of playing the game (Carvalho et al., 2015). If GBA were to adopt similar procedures, teachers could be more comfortable identifying children's challenges in processing information, problem-solving approach, creativity, critical thinking, and other process variables rather than scores alone (Serrano-Laguna, Manero, Freire, Fernández-Manjón, & Serrano-Laguna, 2017). Although 14 (42%) studies had embedded tasks into the game, only 3 (10%) indicated their stealth assessment's nature.

The studies also presented different intervention strategies to help children on their journey to school. Two studies offered unique solutions to detect and remedy dyslexia—a learning disorder prevalent among 5-15% of children, affecting reading and writing (American Psychological Association, 2013). These studies employed the Human Centred Design approach to develop systems and experiments.

Only four subjects were featured in the subject-specific cognitive domain: science, arithmetic, reading/pre-reading, and English, and almost all GBAs just assessed one of these subjects. However, other subjects are more commonly offered for children in preschool, such as music and creative arts; moreover, accurate arithmetic and English grammar are rarely included in preschool curricula. Neumann and Neumann's (2019) study was unique since it offered numeracy and literacy skills assessment. The other study that assessed more than one school readiness attribute was Jozsa et al. (2017). This study assessed two approaches to learning (executive functions and mastery motivation) and pre-academic skills, specifically letter and number recognition. All the other studies assessed only one competency or skills related to one competency alone. Despite all studies adopting most of the ECD framework components, only one study by DeRosier and Thomas (2018) recognised this framework's utility and actively implemented it.

## 5 CONCLUSIONS

The number of studies focussing on GBA of school readiness domains is increasing, although there is too much focus on cognitive domains at the expense of non-cognitive domains that are very useful in developing 21<sup>st</sup> Century skills. The ECD framework can guide game designers to improve these assessments' psychometric properties during their development. Most GBA that adopted embedded assessments did not indicate how the performance data would be analysed for both process and product data. However, those that employed external assessment in a pre-post design had implemented positive intervention strategies to improve school readiness. Additionally, the majority of these studies were carried out in school settings. GBA has a unique opportunity to be applied both in formal classes and at homes. In situations where children cannot attend schools due to pandemics, school enhancement programmes can continue at home. We strongly recommend using GBA in combination with other established instruments to give the teacher and parents a broader spectrum to make correct decisions concerning the child.

### 5.1 Limitations

There are over 150 definitions of school readiness (UNICEF, 2012). Therefore, we chose the most widely used and cited by the National Education Goals Panel in the US. There are debates about what it means to be ready for school. Secondly, we included studies that adopted game-like features in the assessment of the domains. Some studies did not strictly meet a serious game's qualities, although it adopted some game-like features to evaluate the school readiness domains.



## 5.2 Implications for Research, Policy and Practice

Although the wide use of serious games by children of 3-8 years old, very few studies have focused on more than three school readiness domains. There is a need to address other domains and give them more prominence, comparable to the cognitive domain. Despite the evidence centred design being available since 2003, only one study recognised its utility. This provides more evidence for the reliability and validity of such assessments. Most of these studies targeted school settings; however, schools are not always accessible, especially during pandemics such as COVID-19. In this online era, GBA provides an opportunity to diversify assessment outside the school and enhance school readiness.

## ACKNOWLEDGEMENTS

Funded by: National Research, Development and Innovation Office, Hungary

Funder Identifier: <http://dx.doi.org/10.13039/501100018818>

Award: NKFI K124839

## REFERENCES

- Agudo, J. E., Rico, M., & Sanchez, H. (2016). Design and Assessment of Adaptive Hypermedia Games for English Acquisition in Preschool. *Journal of Universal Computer Science*, 22(2), 161–179.
- Almond, R. G., Mislevy, R. J., Steinberg, L. S., Yan, D., & Williamson, D. M. (2015). *Bayesian networks in educational assessment*. Springer. <https://doi.org/10.1007/978-1-4939-2125-6>
- Alonso-Fernández, C., Calvo-Morata, A., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B. (2019). Applications of data science to game learning analytics data: A systematic literature review. *Computers & Education*, 141, 103612–103612. <https://doi.org/10.1016/j.compedu.2019.103612>
- American Psychological Association. (2013). Cautionary Statement for Forensic Use of DSM-5. *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). American Psychiatric Publishing. <https://doi.org/10.1176/appi.books.9780890425596.744053>
- Aragon-Mendizabal, E., Aguilar-Villagran, M., Navarro-Guzman, J. I., & Howell, R. (2017). Improving number sense in kindergarten children with low achievement in mathematics. *Anales De Psicología*, 33(2), 311–318. <https://doi.org/10.6018/analesps.33.2.239391>
- Arieli-Attali, M., Ward, S., Thomas, J., Deonovic, B., & Davier, A. A. (2019). The Expanded Evidence-Centered Design (e-ECD) for Learning and Assessment Systems: A Framework for Incorporating Learning Goals and Processes Within Assessment Design. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.00853>
- Arksey, H., & Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19–32. <https://doi.org/10.1080/1364557032000119616>
- Axelsson, A., Andersson, R., & Gulz, A. (2016). Scaffolding Executive Function Capabilities via Play-&Learn Software for Preschoolers. *Journal of Educational Psychology*, 108(7), 969–981. <https://doi.org/10.1037/edu0000099>
- Barab, S. A., Gresalfi, M., & Ingram-Goble, A. (2010). Transformational play: Using games to position person, content, and context. *Educational Researcher*, 39(7), 525–536. <https://doi.org/10.1080/00131392.2010.500000>

- [.org/10.3102/0013189X10386593](https://doi.org/10.3102/0013189X10386593)
- Barrett, K. C., Józsa, K., & Morgan, G. A. (2017). New computer-based mastery motivation and executive function tasks for school readiness and school success in 3 to 8-year-old children. *Hungarian Educational Research Journal*, 7(2), 86–105. <https://doi.org/10.14413/HERJ/7/2/6>
- Behnamnia, N., Kamsin, A., Ismail, M. A. B., & Hayati, A. (2018). Game-based social skills apps enhance collaboration among young children: A case study. *International Conference for Emerging Technologies in Computing iCETiC 2018: Emerging Technologies in Computing* (pp. 285–291). [https://doi.org/10.1007/978-3-319-95450-9\\_24](https://doi.org/10.1007/978-3-319-95450-9_24)
- Bento, G., & Dias, G. (2017). The importance of outdoor play for young children's healthy development. *Porto Biomedical Journal*, 2(5), 157–160. <https://doi.org/10.1016/j.pbj.2017.03.003>
- Bhavnani, S., Mukherjee, D., Dasgupta, J., Verma, D., Parameshwaran, D., Divan, G., ... Patel, V. (2019). Development, feasibility and acceptability of a gamified cognitive Developmental assessment on an E-Platform (DEEP) in rural Indian pre-schoolers-A pilot study. *Global Health Action*, 12(1). <https://doi.org/10.1080/16549716.2018.1548005>
- Bottino, R. M., Ricerche, C. N. D., Ott, M., & Tavella, M. (2014). Serious Gaming at School: Reflections on Students' Performance, Engagement and Motivation. *International Journal of Game-Based Learning*, 4(1), 21–36.
- Burchinal, M., Magnuson, K., Powell, D., & Hong, S. S. (2015). Early Childcare and Education. In R. M. Lerner (Ed.), *Handbook of Child Psychology and Developmental Science* (pp. 1–45). John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118963418.childpsy406>
- Caballero-Hernández, J. A., Palomo-Duarte, M., & Doderó, J. M. (2017). Skill assessment in learning experiences based on serious games: A Systematic Mapping Study. *Computers & Education*, 113, 42–60. <https://doi.org/10.1016/j.compedu.2017.05.008>
- Carson, K. L. (2017). Reliability and Predictive Validity of Preschool Web-Based Phonological Awareness Assessment for Identifying School-Aged Reading Difficulty. *Communication Disorders Quarterly*, 39(1), 259–269. <https://doi.org/10.1177/1525740116686166>
- Carvalho, M. B., Bellotti, F., Berta, R., Gloria, A. D., Sedano, C. I., Hauge, J. B., ... Rauterberg, M. (2015). An activity theory-based model for serious games analysis and conceptual design. *Computers & Education*, 87, 166–181. <https://doi.org/10.1016/j.compedu.2015.03.023>
- Chan, K. Y. G., Tan, S. L., Hew, K. F. T., Koh, B. G., Lim, L. S., & Yong, J. C. (2017). Knowledge for games, games for knowledge: Designing a digital roll-and-move board game for a law of torts class. *Research and Practice in Technology Enhanced Learning*, 12(1), 1–20. <https://doi.org/10.1186/s41039-016-0045-1>
- Chaudron, S., Beutel, M., Navarrete, V. D., Dreier, M., Fletcher-Watson, B., Heikkilä, A., ... Marsh, J. (2015). *Young Children (0-8) and digital technology: A qualitative exploratory study across seven countries*. JCR/ISPRA.
- Chaudy, Y., Connolly, T., & Hainey, T. (2013). Specification and design of a generalised assessment engine for GBL applications. *7th European Conference on Games-Based Learning, Portugal*.
- Chiu, F. Y., & Hsieh, M. L. (2017). Role-Playing Game-Based Assessment to Fractional Concept in Second Grade Mathematics. *EURASIA Journal of Mathematics*, 13(4), 1075–1083. <https://doi.org/10.12973/eurasia.2017.00659a>
- Cohrssen, C., & Niklas, F. (2019). Using mathematics games in preschool settings to support the development of children's numeracy skills. *International Journal of Early Years Education*, 27(3). <https://doi.org/10.1080/09669760.2019.1629882>
- Craig, A. B., Derosier, M. E., & Watanabe, Y. (2015). Differences Between Japanese and US Children's Performance on "Zoo U": A Game-Based Social Skills Assessment. *Games for Health Journal*, 4(4), 285–294. <https://doi.org/10.1089/g4h.2014.0075>
- Csapó, B., Molnár, G., & Nagy, J. (2014). Computer-based assessment of school readiness and

- early reasoning. *Journal of Educational Psychology*, 106(3), 639–650. <https://doi.org/10.1037/a0035756>
- Daudt, H. M., Van Mossel, C., & Scott, S. J. (2013). Enhancing the scoping study methodology: A large, inter-professional team's experience with Arksey and O'Malley's framework. *BMC Medical Research Methodology*, 13(1), 48. <https://doi.org/10.1186/1471-2288-13-48>
- DeRosier, M. E., & Thomas, J. M. (2018). Establishing the criterion validity of Zoo U's game-based social emotional skills assessment for school-based outcomes. *Journal of Applied Developmental Psychology*, 55, 52–61. <https://doi.org/10.1016/j.appdev.2017.03.001>
- DiCerbo, K. E. (2017). Building the Evidentiary Argument in Game-Based Assessment. *Journal of Applied Testing Technology*, 18(S1), 7–18.
- Ferrara, S., Lai, E., Reilly, A., & Nichols, P. (2017). Principled Approaches to Assessment Design, Development, and Implementation. In A. A. Rupp & J. Leighton (Eds.), *The Handbook of Cognition and Assessment*. Wiley. <https://doi.org/10.1002/9781118956588.ch3>
- Ford, C. B., Kim, H. Y., Brown, L., Aber, J. L., & Sheridan, M. A. (2019). A cognitive assessment tool designed for data collection in the field in low- and middle-income countries. *Research in Comparative and International Education*, 14(1), 141–157. <https://doi.org/10.1177/1745499919829217>
- Geurts, L., Abeele, V. V., Celis, V., Husson, J., Audenaeren, L. V., Den, ... Ghesquière, P. (2015). DIESEL-X: A Game-Based Tool for Early Risk Detection of Dyslexia in Preschoolers. In J. Torbeyns, E. Lehtinen, & J. Elen (Eds.), *Describing and Studying Domain-Specific Serious Games*. Springer. [https://doi.org/10.1007/978-3-319-20276-1\\_7](https://doi.org/10.1007/978-3-319-20276-1_7)
- Godwin, K. E., Lomas, D., Koedinger, K. R., & Fisher, A. V. (2015). Monster Mischief: Designing a video game to assess selective sustained attention. *International Journal of Gaming and Computer-Mediated Simulations*, 7(4), 18–39. <https://doi.org/10.4018/IJGCMS.2015100102>
- Gresalfi, M. S., Rittle-Johnson, B., Loehr, A. M., Nichols, I. T., & Mccracken, C. (2016). *Slicing and Bouncing: Can Implicit Digital Games Support Transfer to Traditional Assessments as Well as Explicit Digital Games*. AERA Online Paper Repository.
- Grose, M. (2013). *The Good and the Bad of Digital Technology for Kids*. Retrieved from [http://www.lawley.wa.edu.au/upload/pages/parenting-resources-insight/insights\\_learning\\_technology.pdf?1445403034](http://www.lawley.wa.edu.au/upload/pages/parenting-resources-insight/insights_learning_technology.pdf?1445403034)
- Hsu, C. Y., Tsai, C. C., & Liang, J. C. (2011). Facilitating Preschoolers' Scientific Knowledge Construction via Computer Games Regarding Light and Shadow: The Effect of the Prediction-Observation-Explanation (POE) Strategy. *Journal of Science Education and Technology*, 20(5), 482–493. <https://doi.org/10.1007/s10956-011-9298-z>
- Hunter, L. J., Bierman, K. L., & Hall, C. M. (2018). Assessing Non-cognitive Aspects of School Readiness: The Predictive Validity of Brief Teacher Rating Scales of Social-Emotional Competence and Approaches to Learning. *Early Education and Development*, 29(8), 1081–1094.
- Ifenthaler, D., Eseryel, D., & Ge, X. (2012). Assessment for game-based learning. *Assessment in game-based learning* (pp. 1–8). Springer. [https://doi.org/10.1007/978-1-4614-3546-4\\_1](https://doi.org/10.1007/978-1-4614-3546-4_1)
- Jozsa, K., Barrett, K. C., & Morgan, G. A. (2017). Game-like Tablet Assessment of Approaches to Learning: Assessing Mastery Motivation and Executive Functions. *Electronic Journal of Research In Educational Psychology*, 15(3), 665–695. <https://doi.org/10.14204/ejrep.43.17026>
- Kagan, S. L., Moore, E., & Bredekamp, S. (1995). *Reconsidering children's early development and learning: Toward common views and vocabulary*. DIANE Publishing.
- Kim, Y. J., Almond, R. G., & Shute, V. J. (2016). Applying Evidence-Centered Design for the Development of Game-Based Assessments in Physics Playground. *International Journal of Testing*, 16(2), 142–163. <https://doi.org/10.1080/15305058.2015.1108322>
- Klerk, S. D., Veldkamp, B. P., & Eggen, T. J. (2015). Psychometric analysis of the performance

- data of simulation-based assessment: A systematic review and a Bayesian network example. *Computers & Education*, 85, 23–34. <https://doi.org/10.1016/j.compedu.2014.12.020>
- Lee, K., Jeong, D., Schindler, R. C., Hlavaty, L. E., Gross, S. I., & Short, E. J. (2018). Interactive Block Games for Assessing Children's Cognitive Skills: Design and Preliminary Evaluation. *Frontiers in Pediatrics*, 6. <https://doi.org/10.3389/fped.2018.00111>
- Li, L., Fan, J., & Jin, Z. (2019). Comparing multimethod assessment of approaches to learning among preschool children: Direct measure, teacher report, and parent report. *Psychology in the Schools*, 56(8), 1271–1286. <https://doi.org/10.1002/pits.22274>
- Loachamín-Valencia, M., Juan, M. C., Méndez-López, M., & Pérez-Hernández, E. (2018). Using a serious game to assess spatial memory in children and adults. *LNCS*, 10714, 829–829. [https://doi.org/10.1007/978-3-319-76270-8\\_55](https://doi.org/10.1007/978-3-319-76270-8_55)
- Mainieri, B. O., Azevedo, V. L., Braga, P. H. C., & Omar, N. (2018). Development and assessment of an adaptive difficulty arithmetic game-based learning object. *Proceedings - 13th Latin American Conference on Learning Technologies, LACLO 2018* (pp. 232–239). <https://doi.org/10.1109/LACLO.2018.00051>
- Merino-Campos, C., & Castillo-Fernández, H. (2016). The Benefits of Active Video Games for Educational and Physical Activity Approaches: A Systematic Review. *Journal of New Approaches in Educational Research*, 5(2), 115–122. <https://doi.org/10.7821/naer.2016.7.164>
- Miller, T. (2018). Developing numeracy skills using interactive technology in a play-based learning environment. *International Journal of Stem Education*, 5. <https://doi.org/10.1186/s40594-018-0135-2>
- Mironcika, S., Toussaint, H., Schipper, A. D., Kröse, B., Brons, A., & Schouten, B. (2018). Smart toys design opportunities for measuring children's fine motor skills development. 2018-January. *The Twelfth International Conference* (pp. 349–356). Retrieved from <https://doi.org/10.1145/3173225.3173256>
- Mislevy, R. J., Behrens, J. T., Dicerbo, K. E., & Levy, R. (2012). Design and Discovery in Educational Assessment: Evidence-Centered Design, Psychometrics, and Educational Data Mining. *Journal of Educational Data Mining*, 4(1), 11–48. <https://doi.org/10.5281/zenodo.3554641>
- Mislevy, R. J., Steinberg, L. S., Almond, R. G., & Lukas, J. F. (2006). Concepts, terminology, and basic models of evidence-centred design. In D. M. Williamson, R. J. Mislevy, & I. I. Bejar (Eds.), *Automated scoring of complex tasks in computer-based testing* (pp. 15–47). Lawrence Erlbaum Associates.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med*, 6(7).
- Moyer-Packenham, P. S., Bullock, E. K., Shumway, J. F., Tucker, S. I., Watts, C. M., Westenskow, A., ... Jordan, K. (2016). The role of affordances in children's learning performance and efficiency when using virtual manipulative mathematics touch-screen apps. *Mathematics Education Research Journal*, 28(1), 79–105. <https://doi.org/10.1007/s13394-015-0161-z>
- Neumann, M. M., & Neumann, D. L. (2019). Validation of a touch screen tablet assessment of early literacy skills and a comparison with a traditional paper-based assessment. *International Journal of Research & Method in Education*, 42(4), 385–398. <https://doi.org/10.1080/1743727X.2018.1498078>
- Nichols, P., Kobrin, J. L., Lai, E., & Koepfler, J. (2016). The role of theories of learning and cognition in assessment design and development. In A. A. Rupp & J. P. Leighton (Eds.), *The hand-book of cognition and assessment: Frameworks, methodologies, and applications* (pp. 15–40). John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118956588.ch2>
- Ninaus, M., Kiili, K., McMullen, J., & Moeller, K. (2016). A Game-Based Approach to Examining Students' Conceptual Knowledge of Fractions. In R. Bottino, J. Jeuring, & R. Veltkamp (Eds.),

- Games and Learning Alliance. GALA 2016* (Vol. 10056). Springer. [https://doi.org/10.1007/978-3-319-50182-6\\_4](https://doi.org/10.1007/978-3-319-50182-6_4)
- Obradovic, J., Sulik, M. J., Finch, J. E., & Tirado-Strayer, N. (2018). Assessing students' executive functions in the classroom: Validating a scalable group-based procedure. *Journal of Applied Developmental Psychology*, 55(SI), 4–13. <https://doi.org/10.1016/j.appdev.2017.03.003>
- Orfanakis, V., & Papadakis, S. (2014). A new programming environment for teaching programming. A first acquaintance with Enchanting. *The 2nd international virtual Scientific Conference - Scieconf 2014, EDIS* (pp. 268–273). University of Zilina.
- Ow, S. H., & Tan, C. M. (2017). Using a computer game to assess the critical thinking skills of preschoolers: A pilot study. *IEEE Conference on E-Learning, e-Management and e-Services (IC3e)* (pp. 151–156). <https://doi.org/10.1109/IC3e.2017.8409254>
- Papadakis, S., & Kalogiannakis, M. (2017). Mobile educational applications for children: What educators and parents need to know. *International Journal of Mobile Learning and Organisation*, 11(3), 256–277. <https://doi.org/10.1504/IJMLLO.2017.085338>
- Puolakanaho, A., & Latvala, J. M. (2017). Embedding preschool assessment methods into digital learning games to predict early reading skills. *Human Technology*, 13(2), 216–236. <https://doi.org/10.17011/ht/urn.201711104212>
- Rauschenberger, M., Lins, C., Rousselle, N., Hein, A., & Fudickar, S. (2019). Designing a new puzzle app to target dyslexia screening in pre-readers. *ACM International Conference Proceeding Series* (pp. 155–159). <https://doi.org/10.1145/3342428.3342679>
- Rupp, A. A., Gushta, M., Mislevy, R. J., & Shaffer, D. W. (2010). Evidence-centered Design of Epistemic Games: Measurement Principles for Complex Learning Environments. *The Journal of Technology, Learning and Assessment*, 8(4).
- Russo, J. M., Williford, A. P., Markowitz, A. J., Vitiello, V. E., & Bassok, D. (2019). Examining the validity of a widely-used school readiness assessment: Implications for teachers and early childhood programs. *Early Childhood Research Quarterly*, 48, 14–25. <https://doi.org/10.1016/j.ecresq.2019.02.003>
- Sabol, T. J., & Pianta, R. C. (2017). The state of young children in the United States: School readiness. In E. Votruba-Drzal & E. Dearing (Eds.), *The Wiley Handbook of Early Childhood Development Programs, Practices, and Policies* (pp. 3–17). Wiley. <https://doi.org/10.1002/9781118937334.fmatter>.
- Semmelmann, K., Nordt, M., Sommer, K., Röhnke, R., Mount, L., Prüfer, H., ... Weigelt, S. (2016). U Can Touch This: How Tablets Can Be Used to Study Cognitive Development. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.01021>
- Serrano-Laguna, A., Manero, B., Freire, M., Fernández-Manjón, B., & Serrano-Laguna, A. (2017). A methodology to assess the effectiveness of serious games and infer player learning outcomes. *Multimedia Tools and Applications*, 77(2), 2849–2871. <https://doi.org/10.1007/s11042-017-4467-6>
- Shih, S. C., Kuo, B. C., & Lee, S. J. (2019). An Online Game-Based Computational Estimation Assessment Combining Cognitive Diagnostic Model and Strategy Analysis. *Educational Psychology*, 39(10), 1255–1277. <https://doi.org/10.1080/01443410.2018.1501468>
- Stein, K., Veisson, M., Oun, T. ., & Tammemäe, T. (2019). Estonian preschool teachers' views on supporting children's school readiness. *Education 3-13*, 47(8), 920–932. <https://doi.org/10.1080/03004279.2018.1539113>
- Suleiman, I., Arslan, M., Alhadj, R., & Ridley, M. (2016). The power of online genetic algorithm in stealth assessment for school readiness. *Journal of Computers In Education*, 3(2), 209–246. <https://doi.org/10.1007/s40692-016-0054-5>
- UNICEF. (2012). School Readiness: A conceptual framework. United Nations Children's Fund

Education Section, Programme Division.

- Willoughby, M. T., Piper, B., Oyanga, A., & King, K. (2019). Measuring executive function skills in young children in Kenya: Associations with school readiness. *Developmental Science*, e12818. <https://doi.org/10.1111/desc.12818>
- Xin, S., Jian, L., & Zhiyu, F. (2018). Using game log-file to predict students' reasoning ability and mathematical achievement: An application of machine learning. *Acta Psychologica Sinica*, 50(7), 761–770. <https://doi.org/10.3724/SP.J.1041.2018.00761>