Tallinn University

School of Digital Technologies
Digital Learning Games

Embedded Assessment in Learning Games
Addressing the Gap Between Theory and Practice

Master Thesis

Author: Manisha Khulbe
Supervisor: Mikhail Fiadotau

Author: ..... „ ..... „ 2018

Supervisor: ..... „ ..... „ 2018

Head of Institute: ..... „ ..... „ 2018

Tallinn 2018

Author’s Declaration

I hereby declare that, apart from work whose authors are clearly acknowledged, this document is the result of my own and original work.
This thesis has not and is not being submitted for any other comparable academic award.

The thesis has been supervised by Mikhail Fiadotau.

.....
(Manisha Khulbe)
03-05-2018

Table of Contents

Introduction

Mr.Strongman's 8th grade class are playing a video game to learn about Newton's laws of motion. They manipulate the forces acting upon game objects, in order to guide them to required goals. Mr.Strongman can follow their progress on a dashboard that displays useful statistics about each student. He can see which problems his students took the longest to solve, and which students employed a trial-and-error strategy to get through the game. The dashboard even groups students based on the misconceptions they seem to harbour about the topic. Later, he is eloquent in his praise of the dashboard to the researchers who have developed the game: Mr.Strongman and his class are part of an experiment to test the physics game and the assessment system that is at the heart of it. Over the weekend, Mr.Strongman searches for similar games to use with his class, but does not find anything approaching the assessment capabilities that he experienced during the study.

Numerous educational researchers see assessment in game based learning, and stealth assessment, in particular, as the future of digital learning games. Much has been written about the many ways in which the incorporation of internal assessment can guide along proper lines the process of learning game development and testing, provide much-needed evidence of actual learning happening during a game, make formative assessment available easily for the benefit of teachers and students, and make learning games more welcome in schools. A number of embedded assessment frameworks have been developed and tested over the last decade.

Most embedded assessment frameworks, however, have not been used either extensively or outside of experimental setups. It seems that apart from an enthusiastic but small subgroup, neither educational researchers nor game designers have yet taken to thinking of assessment as an essential part of learning games.

It may be argued, therefore, that a gap exists between what some researchers claim could prove to be an important, path-breaking ingredient of learning games, and what most other researchers and game designers think of when they set out to develop such a game. The purpose of the current study is the formulation of recommendations to bridge this gap. For this, the prerequisite will be to present a clear picture of existing embedded assessment frameworks, and the ways in which

they can advance and improve digital learning games. Once this context is provided, the next step will be to develop an understanding of how professionals working in the field of game-based learning perceive embedded assessment in learning games. Finally, relying on what is learnt, recommendations will be formulated as to the steps that can be taken to encourage the use of embedded assessment frameworks in the design and development of learning games.

The following questions guide the researcher during this study:

1. What are the embedded assessment frameworks for learning games described in recent research written in English?

2. How can the use of these assessment frameworks make learning games better and more relevant to schools?

3. What are the perceptions and practices of educational researchers and game designers when it comes to incorporating assessment in educational games?

4. How can adoption of embedded assessment frameworks be promoted?

To answer the first two questions, the author will carry out a systematic review of relevant, English language literature published in recent years. The answer to the third question is hoped to be arrived at through a survey directed at researchers and designers active in the field of game-based learning. Then, combining all that the author learns about embedded assessment frameworks and the perceptions of the game-based learning community regarding it, it is hoped that a satisfactory answer to the fourth question will be arrived at.

1 Theoretical overview

This chapter defines terms used throughout the study, and also provides information that this study is logically built upon.

1.1 Defining a game

According to Salen and Zimmerman (2003), a game is "a system in which players engage in an artificial conflict, defined by rules, that results in a quantifiable outcome". This statement succinctly expresses what is essential in many other viable definitions of the term.

This study focuses specifically on digital learning games meant for K–12 use. That is, games whose flow and rules are controlled by a computational device, and which are intended to be used for learning content or skills relevant to a school setting, and in such a setting.

1.2 Games for learning

The enthusiasm for game-based learning that we are currently witnessing is not without precedent. The history of games being used for learning is a longer one than most expect. This section traces the origin and evolution of learning games, and also provides a summary of the findings of the most recent reviews of research about the effectiveness of game-based learning.

1.2.1 A brief historical overview

Computer games and simulations have been in use for learning since the 1950s, when management games were very new and quite popular. Disenchantment with the medium was soon to follow, though, as problems associated with it came to light, and games came to occupy a minor, lacklustre place in educational institutions (Cullingford, Mawdesley, & Davies, 1979).

Ito (2009) describes how video games again captured the imagination of educators in the late 1970s. The game Space Invaders was released in 1978, and proved to be hugely popular almost overnight. Inspired by this success, the gaming industry recovered from the crash it had experienced in 1977. The appeal exercised on young

people by video games was very great, and it was only natural that educators paid attention and were desirous of harnessing the motivational and engaging aspects of video games for learning purposes. In the USA, some educators set out to find the right balance between play and learning, and created software, fun and informative at the same time, for the use of children (Ito, 2009).

In the 1980s and 1990s, more and more people came to own personal computers. Significant improvements were made in the computing speeds of these devices, and at about the same time, CD-ROMs became ubiquitous. Klopfer, Osterweil, & Salen (2009) provide an interesting account of how these changes contributed towards the establishment of the edutainment industry, which focused on combining pedagogy with gaming to make learning attractive for children. “Educational CD-ROMs” were eagerly purchased by parents anxious to give their children a head start in school. Unfortunately, the industry expanded too quickly for its own good. Edutainment titles were released hastily and in abundance, and while a few games, such as *Reader Rabbit, SimCity, and Odell Lake*, did come to enjoy a large audience, most eventually found their way to bargain bins of retails stores.

In the mid to late 1990s, the CD-ROM market shrank, with internet access becoming more widespread. Publishers who were already creating software, toys and other products for children rushed to carve out a place for themselves on the web, and offered free online content to attract consumers. Competition was intense, and products were developed and launched in a hurry, little thought being devoted to the educational aspect of edutainment. To garner attention, publishers took to licensing at high cost characters already popular among children, thus taking more funds away from already suffering R&D departments. Diminishing quality and market pressures caused the appeal of edutainment to fade in the early years of the 21st century.

In 2002, the game *America’s Army* was released. The game is a first-person shooter that allows players to virtually explore the army of the USA in order to decide whether a career as a soldier would be suitable for them. This game, developed not for entertainment but for army recruitment, inspired the founding of the Serious Games Initiative by David Rejeski and Ben Sawyer, and the serious games movement was born. The movement has since continued to grow in popularity, and new avenues for serious games have opened up, such as games for social change,

health, and of course, learning (Susi, Johannesson, & Backlund, 2007).

1.2.2 Regarding the effectiveness of game-based learning

Since the early 2000s, scholars have been very vocal about the transformative effect digital games can have on learning in school. Some of this is mere speculation not backed by persuasive evidence, and has been criticised by a number of authors (Selwyn, 2016; Westera, 2015) Other claims, however, rest on the more solid foundation of empirical research.

Clark, Tanner-Smith, & Killingsworth (2014) reviewed research on the effectiveness of games for learning, selecting papers in peer-reviewed journals published between 2000 and 2012, where participants were students in K-16 grades. From an initial selection of 3141 papers, 70 were chosen, which met adequate research method requirements. These studies showed that students who were part of game conditions performed better than those in non-game instructional conditions. It was also seen that games specifically designed with features to enhance learning performed better than standard versions of these games.

A more recent review of research into learning games identified 7117 papers published between 2009 and 2014 that purported to show that games have significant value for learning. Of these, merely 143 papers passed the quality checks imposed for the review, namely, the usage of valid and reliable research methods and measures. The positive outcomes of playing the games described in the papers were diverse, including knowledge acquisition, perceptual and cognitive benefits, affective and behaviour change, and improved social skills. (Boyle et al., 2016)

While these reviews show that proponents of game-based learning have, in recent years, put together sufficient proof about the pedagogical value of learning games to warrant interest, and further development and research in the field, other authors point out that a large portion of related research suffers from the lack of an overarching methodology that is known to, and followed by all researchers (All, Castellar, & Van Looy, 2016).

1.3 A practical categorisation of learning games

The kinds of games academics have in mind when they extol the virtues of games for

learning are not necessarily the ones that are most frequently utilised in schools. It would be useful, going forward, to have a clear picture of the sorts of games typically used in classrooms.

1.3.1 Educational games

These are games that have been developed specifically for educational purposes. One way to group such games is along a spectrum, spanning from short-form games on the one end, to long-form games on the other. As the names suggest, the difference between the two lies in the amount of time it takes to play through the game. As we shall see later, the duration of gameplay has much influence upon teachers' decisions to use the game.

Short-form games

These are very often quizzes (no different from those administered in more traditional, pen-and-paper ways) overlaid with colourful graphics and paired with some game-like elements such as stars, badges, scoreboards, etc. The game mechanics are thus quite unconnected to what has to be learnt, and are in place mainly to motivate and engage users in gameplay.

An example of such a game is *MathBlaster*, first launched in 1983, and now considered a classic. In the game, after correctly answering a simple arithmetic equation, players are rewarded by being allowed to shoot at objects.

These games are generally a far cry from the games that scholars have shown interest in as tools for learning, and research-based claims regarding their pedagogical efficacy are scant (Boyle et al., 2016). However, they can fulfil certain educational needs, by providing focused, repetitive practice (a majority of these games fall into the category of what are referred to as drill-and-practice games) and opportunity for memorisation of facts in a very specific context. In this, it can be argued that these games are influenced by behaviourism, and seek to reinforce learning through repetition and external rewards.

Long-form games

These games usually involve game mechanics thoughtfully designed to be suited and intrinsic to what is being learnt. Often, these are role-playing games which are

exploratory and open-ended in nature.

An example of long-form educational games is *Mission US: Cheyenne Odyssey*. Players step into the shoes of Northern Cheyenne boy in the late 1800s, and see from his point of view the changes wrought by newly-arrived white settlers.

It is long-form games, generally, that researchers have in mind when they make claims about the pedagogical effectiveness of learning games (Richards, Stebbins, & Moellering, 2013). It has also been shown that players are more immersed in these games where play and learning are expertly combined, than in games where game mechanics are divorced from game content (Habgood & Ainsworth, 2011).

1.3.2 Commercial off-the-shelf games

Creative and enterprising school teachers have, for many years, made commercial titles such as *Civilization*, *Minecraft*, *Roller Coaster Tycoon*, *SimCity*, *World of Warcraft*, etc., a part of their classrooms. While entertainment is the main business of these games, they can, and have been shown to, contribute towards the enhancement of academically relevant knowledge and skills.

For example, Short (2012) describes how *Minecraft*, and its educational modification, *MinecraftEDU* can be used to effectively teach scientific concepts. With its focus on realistic biomes, the game is especially suited to the study of ecology. *MineChem* is another modification with focus on chemistry and some basic physics. Improvements in higher order skills (creation and evaluation), and collaboration have also been documented as a result of playing the game (Callaghan, 2016). *SimCity* has successfully been used to encourage awareness of local government and civic participation (Tanes & Cemalcilar, 2010). Numerous similar studies exploring the pedagogical uses and benefits of different commercial games have been published.

1.3.3 Discussion

From reading this section, it would be reasonable to expect that teachers who use games in their classrooms employ them in a variety of ways, picking game titles and formats that are most suited to the task at hand. However, as we shall see next, a disconcerting picture emerges on putting together information from surveys

involving the actual use of digital games in the classroom.

1.4 Teachers' experiences with learning games

When researchers talk about cases in which games were proven to be pedagogically effective, they mostly talk about studies that have taken place under their supervision. In other words, most of the empirical research that backs the merit of games for learning is based on experimental setups. Naturalistic studies about actual game use in the "wild" are, so far, pitifully scarce, and what they have to show is not particularly encouraging for the proponents of game-based learning.

1.4.1 How are games being used in the classroom?

Combining the results of a survey of 1704 pre-service and in-service teachers, followed by interviews with 42 pre-service teachers and 110 in-service teachers, Ruggiero (2013) discovered that an overwhelming majority of participants were of the view that games should not be the main activity around which a lesson should be planned. Instead, they were in favour of using games to reward students for work done in class, and for motivating special needs students.

A survey of 694 K-8 teachers in the USA found that 74% of participants used digital games in their classrooms. This number seems promising, but the details that follow shed a different light upon the matter. It appears that while teachers do report occasionally using games for delivering mandated content, and even for assessment purposes, more than half of game-using teachers use games in their class mainly to motivate or reward lower-performing students, and are of the opinion that the most valuable quality of game-based learning is its potential to motivate low performing and special needs students (Takeuchi & Vaala, 2014).

Another survey involving 111 science teachers in the USA mirrored these findings. While 70% of the participants had previously used games with their students, they mostly used drill-and-practice games for motivational purposes, and were not well-informed regarding immersive games created specifically to improve higher order skills in the STEM (Science, Technology, Engineering, and Maths) fields (An, Haynes, D'Alba, & Chumney, 2016).

1.4.2 Discussion

It is true that games make excellent motivation tools for students, and their use in such a manner is commendable. However, the problem with this approach lies in the fact that this is a rather trivial use of games, and falls far short of what they can be used to achieve, and have achieved when tested in experimental setups. The following section is an attempt to explain why many teachers are still hesitant to use educational games in class, and why most game-using teachers, too, generally employ games in an extremely circumscribed manner.

1.5 Obstacles to the adoption of games in the classroom

A number of roadblocks stand in the way of learning game integration in schools. Some of these, such as the lack of computers and scarcity of financial resources can properly be addressed by individual schools, or relevant wings of the government. Other difficulties, even though challenging, can be eased by action on part of educational researchers and learning game developers, as will be shown later in this chapter. They are:

1.5.1 Scepticism from teachers

The value of educational games has long been a subject of speculation. Many teachers hold on to the belief that games and learning cannot mix, and video games certainly do not belong in class, where they distract students from their studies (An et al., 2016).

In a small but painstaking study, in-service history teachers played a number of short- and long-form learning games for a minimum of 20 hours over 2 months, and reflected upon their experience. All participants displayed scepticism about content learning through video games, though they recognised the games as useful motivational tools (Gaudelli & Taylor, 2011).

On analysing responses from 505 teachers in the Netherlands, Bourgonjon et al. (2013), concluded that participants were doubtful that commercial games could add any value to their classroom practices. Another prominent feature of the study was the very low rate of adoption of games by the participants.

1.5.2 Lack of games aligned to the curriculum

The demands on teachers' time are many, and require that teaching must be confined to the prescribed national and state standards. Time constraints make it difficult for teachers to plan class activities with games that do not directly contribute to the learning of what is mandatory for the school year.

Squire (2004), in his study of the use of Civilization III in classrooms noted that the game content and school curriculum requirements were at odds, and resulted in the acquisition of mere superficial information.

Takeuchi & Vaala (2014) quizzed teachers on what prevents them from employing games in their classrooms. One in every five teachers who already used games said that it was difficult to find games suited to the curriculum of the school. Among non-users, this number was higher.

The findings of An et al. (2016) affirm that teachers do not think it wise to invest time in games unless they have a definite connection to the curriculum.

1.5.3 Uncertainty regarding the role of games in the classroom

So far, neither teacher education, nor development courses for teachers include much information about learning games. This leads to teachers being unaware of the ways in which games can become part of their lesson plans, as has been recorded in a number of studies. Further, time limitations discourage teachers from experimenting with precious class hours to test whether, and how, a game can be beneficial for learning.

Barbour & Evans (2009) queried pre-service teachers regarding their perceptions of commercial video games for learning in school. On playing some games and reflecting upon the experience, most participants agreed that games could be a useful addition to their classes in the future, but also expressed worries about the inability to see how they could combine games with other class activities.

In a survey of 488 teachers across the USA, half of whom taught grades 6 to 8, it was found that a third were hesitant to use educational games because they were not sure of ways to integrate games into instruction (Fishman, Riconscente, Snider, Tsai, & Plass, 2014).

Takeuchi & Vaala, (2014) also determined that a failure to see where games fit into the classroom keeps a significant number of teachers from using the medium.

1.5.4 Discussion

Whether games find a place in classroom instruction is ultimately the decision of teachers. Therefore, teacher concerns must necessarily be addressed if learning games are to be made a part of school. Training and support from school and state authorities are obviously a necessity, but educational researchers and game designers can also make a positive difference.

A number of educational researchers are of the opinion that the merging of assessment and educational games is the way forward for game-based learning (Bellotti, Kapralos, Lee, Moreno-Ger, & Berta, 2013; Chin, Gamson, & Dukes, 2009; Shute, Ventura, Bauer, Zapata-Rivera, 2009). They claim that this approach can help improve learning games, and break down many of the barriers that keep educational games out of school. The sections that follow examine assessment and psychometric models, and then describe game-based assessment, thus setting the stage for discussion of what is central to this study: embedded assessment frameworks.

1.6 Assessment

Before discussing assessment with reference to learning games, it would be useful to describe assessment in general, along with its essential elements and features.

1.6.1 Defining Assessment

In the context of education, assessment is an essential part of learning, and comprises all activities undertaken to collect information that can provide insight into progress made towards learning objectives.

The functions of assessment are diverse. It can be used to guide instruction and meet educational goals. It also serves the social purpose of indicating merit and achievement. While both functions are important, the social, judgemental function has come to be much-maligned, and numerous authors have freely expressed their negative opinion of it and the harm it might cause to learners (Taras, 2005).

1.6.2 Summative and formative assessment

Summative assessment is what we normally think of when we hear the word test. Summative assessment is carried out to discover where a student is placed with reference to a set of objective standards. A few examples of summative assessment are end-of-term exams, tests held at completion of a study unit, and all standardised tests.

The term “formative” was first used in the context of assessment by Michael Scriven in 1967. Assessment is formative when the data it provides is used to tailor subsequent instruction and learning in a manner that serves to close the gap between a student’s current learning state and desired learning state. Examples of formative assessment are diagnostic tests performed to gauge previous knowledge before a new topic is taught, exit/admit tickets, four corners, etc.

In their thorough review of research relating to formative assessment, Black & Wiliam (1998) concluded that the addition of formative assessment activities to classroom practice leads to significant learning gains. It was seen that low achieving students and those with learning abilities stand to gain the most from such interventions.

1.6.3 Feedback

An important aspect of formative assessment is the utilisation of feedback received from the test process to support student learning. Formative assessment should ideally be a daily, routine activity in classrooms, and as several authors have been at pains to point out, students should be recognised as the teacher’s partners in the process (Stiggins, 2008). They should be provided with timely, frequent, descriptive feedback about the quality of their work, along with suggestions for how to improve the same (McTighe & O’Connor, 2005). According to Hattie & Timperley (2007), the most useful feedback is that which answers the questions, “Where am I going? How am I going? and Where to next?” Such feedback promotes students’ reflection about and ownership of their own learning, which, in turn, has been observed as being one of the most substantial factors contributing towards student success (Cannata, Haynes, & Smith, 2013)

While feedback can also be generated from the results of summative assessment, learners cannot use it for constructive purposes. This is because such feedback is infrequent, and more often than not, comes too late.

1.6.4 Assessment reliability and validity

Meyer (2010) defines test score reliability as "*the degree of test score consistency over many replications of a test or performance task"* (p. 4). That is, the same assessment taken at a different time, or a similar exam with only minor variations should yield the same results, if it is sufficiently reliable.

An assessment provides a sample of learner performance, and from this sample, we seek to draw inferences about the general ability of the learner, or learner competence. The validity of an assessment is a measure of how accurate an insight into learner competence is provided by the learner performance. In other words, does the assessment really measure what it claims to measure? (Gipps, 1994)

1.7 Psychometric models

Assessment is serious business in a number of contexts. For many kinds of assessment, the highest standards are required in matters of reliability, validity, fairness of questions, etc. Psychometry, the science of testing, measurement, assessment, and other related activities, is therefore an important field of study, and scientists devoted to improvements in testing have developed models to guide the making of high quality tests. Some of these are discussed here, as they can and have been used to add assessment to learning games, as we shall see in the next chapter.

1.7.1 Item Response Theory

To overcome the limitations of classical measurement models, psychometricians adopted and further developed the Item Response Theory (IRT) in the 1970s and 1980s. IRT is a collection of models which guide the creation and appraisal of tests for educational and psychological purposes (DeMars, 2010).

What distinguishes IRT models from classical models is the focus on individual test items rather than the test as a whole. Numerous IRT models are currently in use, but all of them are based upon the essentials described below:

An examinee's performance on a test particular test item depends upon relevant traits, or abilities of the examinee. This relationship between examinee performance and the traits that contribute towards that performance can be represented with a function that is called Item Characteristic Curve (ICC). The ICC lays down that the probability of an examinee answering an item correctly is directly proportional to the level of the ability being measured by the item.

IRT models also take into account one or more characteristics of test items in order to more fully explain examinee performance. The most commonly used item parameters and corresponding IRT models are as follows:

One-Parameter Logistic model

Also known as the Rasch model after its creator, this model assumes that the item difficulty, or chance of correct response on a test varies, and this variety should be taken into account when examinee performance is being evaluated. In the IRT, an item's difficulty is the ability at which it is expected that about half the examinees will answer the item correctly.

Two-Parameter Logistic model

This model is sometimes called the Birnbaum model. Along with the parameter of item difficulty, this model uses a second parameter of item discrimination. This parameter refers to how well an item identifies separately examinees of differing ability levels.

Three-Parameter Logistic model

Neither of the two models mentioned above take into account the fact that examinees with low abilities may make a guess when faced with test items, and there is a chance, especially with multiple-choice questions, that they will guess correctly. The Three-Parameter Logistic model does take the probability of random guessing into account.

1.7.2 Bayesian network models

Bayesian network models are graphical models which are used to represent probabilistic (but not necessarily causal) dependencies among events. A Bayesian network model is made up of nodes, each of which represents a random process,

which can take at least two values, and to which probabilities may be assigned. Nodes that are directly dependent on each other are linked with arrows. The arrow direction indicates which of the events depends upon the other.

For example, the event “smoke detector works” affects the probability of the event “fire is discovered”, which, in turn, is what the event “fire department arrives in time” depends on. This can be represented as follows with the help of a Bayesian network model, which also includes additional relevant nodes.

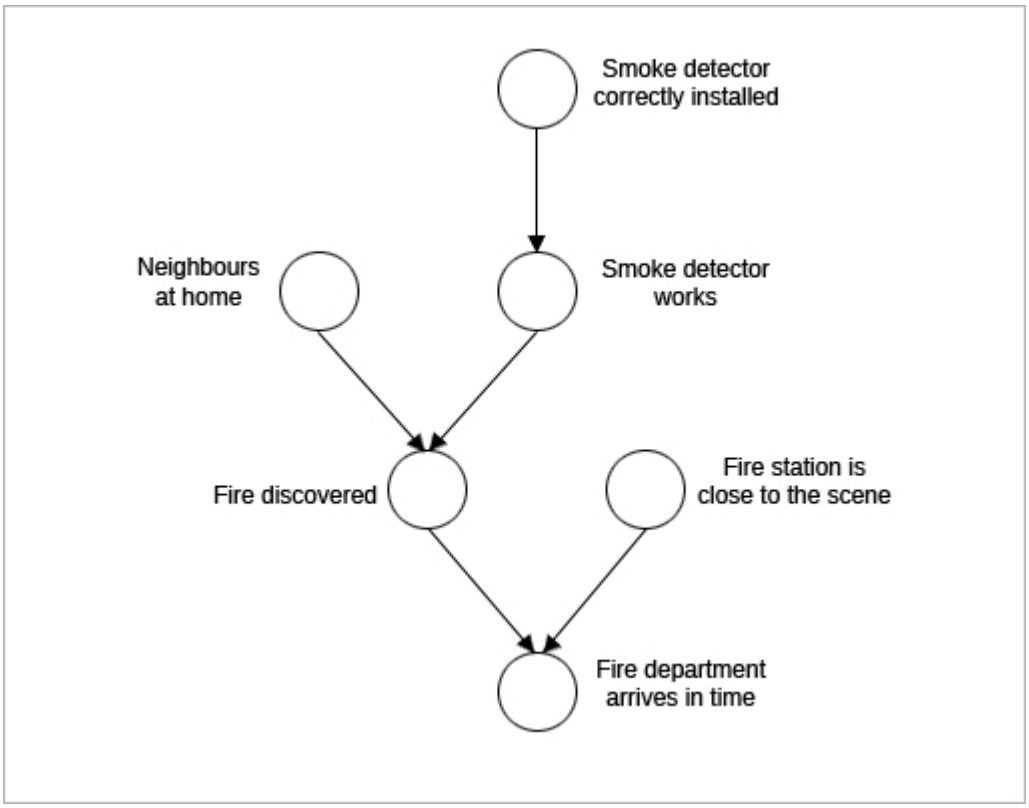

Figure 1. A sample Bayesian network model

When the probability distribution of nodes in Bayesian networks are updated, they pass along this information to their child nodes (predictive propagation), and also to their parent nodes (retrospective propagation), updating these nodes according to Bayes’ theorem. Thus, Bayesian networks can be used to predict or explain events. Turning to the Bayesian network shown above, if the probability of the smoke detector working is high, we can predict that the probability of the fire department arriving at the scene in time is also high. Starting from the opposite node, if the fire department arrives in time, the probability is high that the smoke detector did work.

Similar Bayesian networks can be created for student assessment purposes, with parent, “knowledge and skills” nodes being updated depending on data provided by child, “performance on task” nodes.

1.8 Game-based assessment

Game-based assessment comprises a number of practices that seek, in rather distinct ways, to use games to draw valid inferences about student competencies. The most widely used varieties of game-based assessment are described in this section, followed by an overview of research regarding teachers’ perceptions of such assessment.

1.8.1 External assessment

In the classroom, to gauge whether students have learnt what they should have from playing an educational game, tests, either paper- or computer-based, may be administered after gameplay. Sometimes, games come bundled with sample tests and testing ideas for teachers to use with their class. In most cases, though, teachers create their own assessments to test game-based learning. It is worth noting, though, that a majority of teachers do not assess game-based learning at all (Takeuchi & Vaala, 2014).

For research setups and game testing, participating students often have to complete a pre-test in addition to identical tests after using the game. These tests help measure whether, and to what extent, the educational game affects learning. Such tests are created by educational researchers or game developers, sometimes with the help of teachers or subject matter experts.

1.8.2 Embedded assessment and stealth assessment

Assessment that happens during gameplay is termed embedded assessment. It is also referred to as internal assessment.

Embedded assessment can be thought of as that part of an educational game which is designed to elicit pedagogically relevant information from players, to analyse this data according to pre-determined rules, and to provide the results of such analysis to stakeholders in useful, accessible formats.

A further subset of embedded assessment is stealth assessment. In stealth assessment, in-game test instruments are carefully chosen and/or designed to complement and blend in with the rest of the game. Such assessment seamlessly and invisibly melds with the gameplay, and seems a natural part of it (Shute, 2011).

The difference between stealth assessment and embedded assessment that is not stealth assessment is understood better with examples:

Mars Generation One: Argubot Academy, developed by GlassLab in collaboration with ETS, is an RPG-adventure that aims to teach students argumentation skills, and also assesses the same. Players take on the role of citizens of a city on Mars, and have to make reasoned decisions about the governance of the city. Players choose sides in a debate, and then explore the game world to collect evidence that supports their claims. For example, players may converse with NPCs, who may mention some facts pertinent to the argument. Player assistants, called argubots, are furnished with this information, and a battle of wits ensues. The player with the most sound argument wins. During the course of gameplay, player actions, such as conversations with NPCs, selection of evidence, evaluation of an opponent's claims, etc., are logged by the game. Some other relevant information is also captured, like the time spent on a task. This data is used to create a picture of player progress towards learning goals. The game has a dashboard for teachers, which displays statistics about individual students. Thus, without interfering with the player's experience of the game, stealth assessment is carried out.

An example of embedded assessment, on the other hand, is the very popular *Kahoot*. During a round of *Kahoot*, it is obvious to all participants that assessment is in progress, even though it has been gamified and made appealing with humour. The player data logged is very simple, consisting of answers, and the time taken to decide upon an answer. Teachers can later review class and individual performance reports, and tailor instruction to students' needs.

1.8.3 Making embedded assessment possible

The increased interest of educational researchers in embedded assessment in learning games is not a co-incidence. Instead, it is a result of the rise of data-driven analytics in numerous fields, including education. Data from and about learners is more

abundant and easily accessible now than ever before, with the rise of e-learning in traditional classrooms, and also on the internet. Learning games, of course, involve such data in abundance. The desire to utilise educational data in meaningful ways resulted in the emergence of the fields of Educational Data Mining (EDM) and Learning Analytics (LA).

According to Romero and Ventura (2010), the main objective of EDM is to further educational research by exploring data collected from educational settings. EDM, which combines practices from fields such as statistics, data mining, and machine learning, is used to automatically process this data, scan it for meaningful patterns, and convert it into valuable information that can be utilised by educational researchers to develop an improved understanding of learners and their interactions with educational settings (Kickmeier-Rust et al., 2014). LA, too, seeks to use data about learners and their context to guide instruction and design better educational experiences. The fields of psychology and sociology, in addition to statistics, computer science and data mining, inform practices in the LA community (Calvet Liñan & Juan Pérez, 2015).

Many methods are used by both EDM and LA communities. Prominent among these are clustering, discovery with models such as IRT, Bayesian networks, and Markov models, and distillation of data for human judgement (Baker & Yacef, 2009).

Despite their common goals and methods, EDM and LA do differ in some ways. Siemens and Baker (2012) draw distinctions between the two fields: EDM is more focused on automated discoveries than LA, in which human judgement is central. EDM is often employed to automatically adapt learning systems to learners, while LA makes relevant information available to stakeholders, aiding them in making good decisions.

1.8.4 Game creation as assessment

An increasing number of educational researchers have been writing in favour of expanding game-based learning to encompass not only learning by playing games, but also learning by making games. In this approach, inspired by constructionist theories of education, students take on the role of game designers, consolidating what is learnt, and acquiring new knowledge by creating tangible objects. This is

helpful for learning coding skills, and also lends itself well to content learning, and improvement of metacognition (Kafai & Burke, 2015).

Most commonly, after learning a new concept in class, students are provided with guidelines about creating engaging learning games, and briefed on the learning goals of the game they must create. The process of designing the game usually takes place over a number of sessions. Finally, teachers examine the creations of students to gauge their understanding of the game topic (Weitze, 2016).

1.8.5 Teachers' perceptions of game-based assessment

There is a dearth of research regarding teachers' opinions of and experiences with game-based assessment. In one of the few studies focusing on the subject, Fishman, Riconscente, Snider, Tsai, & Plass (2014) studied the use of learning game features for formative assessment in schools in the USA. 30 middle school (grades 5 to 8) teachers agreed to use one out of an offered selection of eleven learning games available on BrainPOP's Game Up portal. Following half a day of training about game use in the classroom, the teachers' experiences of use or non-use of the games for assessment purposes were recorded, and followed up with interviews.

All the games used were short-form games, focused on very specific topics, and designed for use for a single class period, or over multiple lessons. The features identified by participating teachers as capable of supporting formative assessment ranged from the very simple, such as end-of-game scores, to more complex dashboards displaying student activity and performance statistics. A significant number of teachers were concerned that game elements that could be used to conduct formative assessment were not linked to learning goals, and did not provide a true picture of student mastery of content or skills. Dashboards, while their merits could be appreciated by all participants, proved to be difficult to set up in some cases. It was concluded, thus, that teachers are interested in using games for formative assessment purposes. Further, it was observed that this interest can be fostered if games are designed with well-aligned game and learning elements, and technological improvements make these games more user-friendly and reliable.

1.8.6 Discussion

All of the forms of game-based assessment discussed in this section can be found to be useful in the classroom, depending on the context of the lesson. This study, however, will now confine itself to matters related to embedded and stealth assessment. In the following chapter, the most useful design frameworks for the creation of educational games with embedded assessment will be introduced, described and analysed. The reasons justifying such focus on embedded and stealth assessment will be spelled out in the third chapter.

2 Embedded assessment frameworks for learning games

2.1 Method

A review is the collection and critical analysis of literature about a specific subject. A review is systematic when it is conducted to answer a predetermined research question, and when criteria for work to be eligible for inclusion are drawn up in advance and strictly adhered to. It is also fundamental to the approach to document the process of selection in detail (Kitchenham, 2004).

2.1.1 Research questions

In this part of the study, the researcher aims to answer the following questions:

1. What are the embedded assessment frameworks for learning games described in recent research written in English?
2. How can the use of these assessment frameworks make learning games better and more relevant to schools? (addressed in the next chapter)

2.1.2 Search terms

Keywords used to search for papers were “assessment”, and either of the phrases “game based learning”, or “educational game”, or “learning game”, occurring in the title and/or the abstract of the work.

2.1.3 Inclusion criteria

The review took into consideration research describing embedded assessment frameworks that have been employed in designing learning games for K-12 use. A further criterion was that the resulting game should have been tested with a suitable audience, and results of the test described in sufficient detail. Only publications in English, and appearing between 2008 and 2017 were included.

2.1.4 Sources

The databases utilised for the review were Association for Computing Machinery (ACM), Science Direct, and SpringerLink. Following this search, snowballing and reverse snowballing (using Google Scholar) were employed, to discover other relevant work.

2.1.5 Selection process

To determine which papers found in databases met the inclusion criteria, abstracts of the papers were read, and other section headings briefly scanned. This screening yielded 7 papers relevant to this study. Table 1 shows details of the database search.

Table 1. Details of paper selection from databases

Database	Number of papers containing search terms	Number of papers meeting inclusion criteria
ACM	65	3
Science Direct	23	3
SpringerLink	115	1

A perusal of the reference lists of these papers resulted in the discovery of a further two papers that met the inclusion criteria. This brought the total number of papers to 9.

2.2 Results

From the papers selected for review, five embedded assessment frameworks were identified. These are described in some detail below, and examples of actual implementation are also provided.

2.2.1 Evidence Centred Design

Evidence-Centred Design (ECD) is currently the most well known and widely used design framework for the creation of learning games with embedded or stealth assessment. Robert Mislevy, Russell Almond, and Janice Lukas first developed ECD at the Educational Testing Service as a general framework to guide the creation of assessments administered in diverse formats, ranging from traditional paper-based tests, to embedded assessment in educational games.

2.2.1.1 ECD description

The process of creating an assessment with the use of ECD is long-drawn. So, only the parts of the process which are most pertinent to assessment creation for learning games are described here in some detail.

A. Domain analysis and domain modelling

Mislevy & Riconscente (2005) explain that ECD begins with domain analysis, which is a thorough review of the subject that the assessment is going to address. Important knowledge and skills in the domain of interest are studied in detail by content and instruction experts. Content standards (such as prescribed national or state curricula) and textbooks are good sources of information for domain analysis.

The next step is domain modelling, during which the information collected for domain analysis is arranged in the form of arguments that link examinee responses to the inferences that can be drawn from them about examinee abilities. Descriptions of valued abilities are created, and tasks are formulated which can extract from examinees responses that act as evidence of the aforementioned abilities.

B. Conceptual Assessment Framework

The Conceptual Assessment Framework (CAF) can be described as the "blueprint" for an assessment. Through its three component models, it defines and aligns the game content, game tasks, and rules for evaluating player performance, as shown in Figure 2.

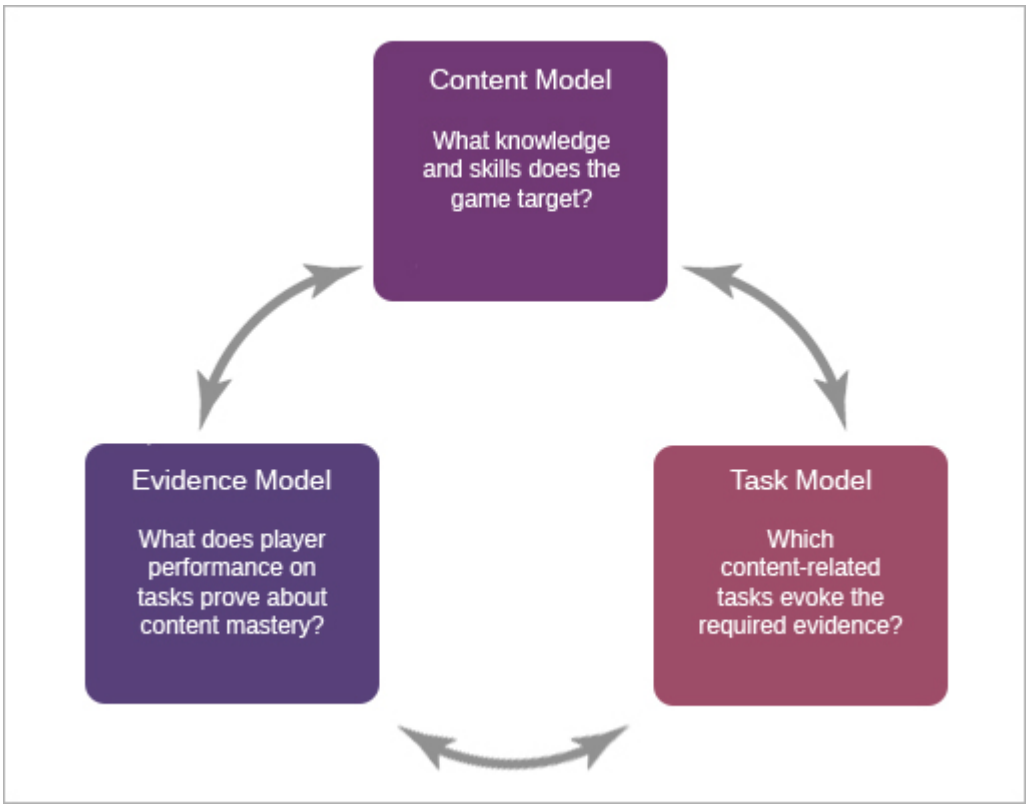

Figure 2. Creating an assessment blueprint using ECD

The Content Model - "*What are we measuring?*"
Also sometimes referred to as the competency model, or student model, this part of the ECD framework defines variables related to the knowledge, skills and attributes being measured. In other words, it contains a detailed description of the learning goals of the game, and what different levels of learner proficiency look like.

The Evidence model - "How do we measure it?", "What do different behaviours say about student competence?"
These models link the content and task model. They contain information about how a student's response for each task should be analysed, what kind of feedback should be provided to the player, and how the content model should be updated based on the analysis of all such responses put together.

The Task model - "Where do we measure it?"
The task model comprises instructions about designing different kinds of tasks that can evoke the evidence that is needed by the evidence model. In the case of an educational game, this is a description of all of the categories of challenges that the player will face.

C. Assessment implementation
In the implementation stage, the assessment and the learning game are actually created. For assessment purposes, different psychometric models can be utilised. The simplest of these models can be based on the Classical Test Theory (CTT), but educational researchers and game designers more frequently make use of latent variable models such as IRT and Bayesian networks (DiCerbo et al., 2015).

2.2.1.2 ECD usage
Results of studies selected for the review which used games built around ECD can be summarised as follows:

Kiili, Devlin, Perttula, Tuomi, & Lindstedt (2015) described their evolving work with Semideus, a learning game focusing on conceptual understanding of rational numbers. The game was tested with 66 sixth grade students, some of them in the USA, and others in Finland. Student performance in the game was compared with performance on a post-test, and it was found that the game provided as good an

estimation of students' understanding of rational numbers as the test. In a more recent study by the same researchers, the assessment system correctly identified areas of difficulty, and some misconceptions, though visualisation and reporting of these was not yet implemented (Ninaus, Kiili, McMullen, & Moeller, 2017).

Newton's Playground, a physics game for secondary school students was tested with 167 eighth- and ninth-grade students. In-game gold and silver trophies received as a result of gameplay were hypothesised by the researchers to be correlated to scores on a post-test if the game-based assessment was valid. This correlation was observed (Shute, Ventura, & Kim, 2013).

2.2.2 CyGaMEs and timed reports

Reese et al. (2012) define Cyberlearning through Game-based, Metaphor Enhanced Learning Objects (CyGaMEs) as an approach to facilitate intuitive learning by using educational games to provide learners with experiences that serve as prior knowledge. To create a CyGaMEs game, developers must painstakingly and closely map a targeted domain to create the game world. The game world borrows its structure from the targeted domain, and gameplay closely imitates the procedural activity of the targeted system. Thus, familiarity with the game prepares players to learn about the target domain, providing them with an accurate mental model of it. This mental model can later be evoked by teachers to aid learner understanding when the target domain is encountered.

To test the game, and to provide feedback to learners and teachers, the CyGaMEs team has created an assessment tool, timed report.

2.2.2.1 Timed report description

Timed reports, as the name suggests, compare, at fixed intervals of time, the current game state with the one before it, and record a score of +1, −1, or 0, with reference to player progress away or towards the game goal. Successive timed report scores are accumulated and plotted on a graph to visualise player progress through the game, which is a good measure of the evolution of the player's mental model of targeted content. The process of specifying and using timed reports can be explained as follows:

A. Choosing a time interval for the reports

Available system resources and the time taken to complete one action in the game are taken into account to choose the optimal time interval for the Timed Report.

B. Testing the timed report

To validate the Timed Report as an assessment tool, the cumulative timed report graph of a knowledgeable player is examined. The graph should be a line with a positive slope, approximating 1, depicting continuous progress towards the game goal. If this is not the case, a problem with the specification of the timed report, or misalignment of the game world and targeted domain are indicated.

2.2.2.2 Timed report usage

The game *Selene* teaches basic planetary science concepts to middle and high school students, and also assesses their understanding of game content. The game has been tested with 221 learners, and timed reports were found to be an adequate tool for tracking player understanding of targeted content. The data from the reports also formed the basis of scaffolding in the form of immediate, descriptive feedback for students. Assessment results in the form of timed report graphs of each student were also available for teachers' viewing (Reese et al., 2012).

2.2.3 Assessment for implicit knowledge

When learners have tacit (unarticulated) prior knowledge related to a concept, this knowledge can be evoked to promote meaningful learning of the concept (Merrill, 2002). Learners sometimes have tacit understanding of principles of science and the laws of nature, but often, they are lacking in it. In such a scenario, educational video games can be used to provide experiences that build and consolidate tacit knowledge. Later, using examples from the game, a teacher can bridge tacit knowledge and explicit understanding of the subject matter.

Educational researchers involved with creating games that provide such prior learning experiences have sought to measure the implicit learning taking place during gameplay. Rowe et al. (2017), creators of two physics learning games, explain that in games like theirs, where all player behaviours cannot possibly be anticipated because of the open-ended nature of gameplay, the ECD framework is of limited

usefulness as it relies on *a priori* selection of a learning trajectory. Instead, the researchers drew upon ECD, but used EDM to infer players' implicit science knowledge.

2.2.3.1 Assessment for implicit knowledge description

Rowe et al. have provided details of the process they use to identify implicit knowledge and track its evolution during gameplay. A summary of this process is provided here.

A. Coding of player behaviours

First, players' moves in the game are categorised as being either consistent or inconsistent with an understanding of the concept targeted by the game. This can be achieved by human coding of gameplay video and accompanying audio of players expressing their thoughts about in-game actions (Asbell-Clarke, Rowe, & Sylvan, 2013). When faced with a very large number of step-by-step actions from all players, a data-driven method for identifying and classifying such moves can be used, in which human judgement is omitted at first, and EDM is leveraged instead to reduce the many moves to fewer clusters of similar game states. These clusters can then conveniently be reviewed by researchers and sorted as correct answers states, or different kinds of error states. Following this coding, all actions of all students are automatically marked as being either correct answers, or a variety of errors.

B. Automating the coding of player behaviours

In cases where players' actions are coded manually, the second step is to train automated detector models to mimic human judgement and automatically categorise a given gameplay sequence as showing implicit understanding of the targeted concept, or lack of such implicit understanding.

C. Comparison with results of traditional assessment

The validity of the detector model is checked by comparing whether students exhibiting certain strategic moves in the game really do score higher on a post-test. In case the second method was chosen, the validity of the assessment is confirmed by comparing a students' coded moves and test score. The number of correct moves should be directly proportional to the assessment score, and the number of errors should negatively affect the score, if the game-based assessment is valid.

2.2.3.2 Assessment for implicit knowledge usage

Impulse, a game about Newton's first two laws of motion was tested with 299 high school students. The detector model created for assessment of learners was able to predict with sufficient accuracy students' scores on a post-test, given their game moves (Rowe et al., 2017).

Quantum Spectre, a puzzle game about the laws of reflection, was tested with 319 high school students. It was found that the embedded assessment tools yielded results mirroring those achieved by students on a post-test (Rowe et al., 2017).

2.2.4 Student PRoficiency INferrer from Game data

Some authors have sought to avoid the time-consuming and expensive process of domain-knowledge engineering that is a part of ECD. They opt instead for data-drive methods to recognise differences in learners' in-game behaviour in order to predict post-test scores, which can be linked to proficiency in game content. Student PRoficiency INferrer from Game data (SPRING) is one such approach.

2.2.4.1 SPRING description

Falakmasir, González-Brenes, Gordon and DiCerbo (2016) explain how SPRING is set up to assess student performance in a learning game. Given student game logs and results in traditional assessments, SPRING can identify differences in gameplay patterns of high- and low-performing students. Once the models used for the assessment have been fine-tuned, student scores can be predicted based on log data alone, doing away with the need to conduct traditional assessments. A very simplified description of this process is provided below:

A. Discretization

Player log data is received in the slot and filler format, and converted into discrete observations which are suited for sequence modelling, which is the next stage in the SPRING process. This discretization is achieved with the use of an unsupervised clustering algorithm, which classifies the data into a small number of clusters.

Similarly, student are divided into two clusters of high and low performers, based on median scores in a paper-based assessment. It is hypothesised that the two groups play differently.

B. Sequence modelling

A Hidden Markov Model (HMM) is then trained for each group of students, using the sequence of student actions obtained as a result of discretization. These models, using the data they have been trained with, can infer the likelihood of the occurrence of any given arbitrary sequence of actions for that group of students.

C. Feature Vector Modelling

At this stage, for each game level, the likelihood of occurrence of a student's observed sequence of events is calculated for both HMMs. Likelihood data from this step is used in a regression model that ultimately predicts the post-test scores of each student. If these are in agreement with the real scores, the game-based assessment method is valid.

2.2.4.2 SPRING usage

SPRING was evaluated using the game *Alice in AreaLand*, which aims to teach geometric measurement to sixth grade students. The study involved 129 participants, who played 11 levels of the game. Game log data and post-test scores from 80% of the students were used to train the HMMs, while those from the remaining 20% were used for test and verification. It was found that scores predicted using SPRING were positively correlated with the real assessment scores.

2.2.5 Game Traces

A player's interaction with any video game may be visualised as a series of events, with all events connected to each other in simple or complex ways. A number of authors have implemented embedded assessment in games used for learning by identifying game events which provide insight into player knowledge or skills, logging these events, and putting very simple rules in place for the evaluation of players based on event parameters.

2.2.5.1 Game Traces description

The use of the game traces methods to add assessment features to an existing learning game is a very simple and intuitive process. It can be summarised as follows:

A. Identifying universal traces

While events differ from game to game, some "universal traces" are basic and common enough to be present in practically all games and game genres. Serrano-Laguna, Torrente, Moreno-Ger, & Fernández-Manjón (2014) have identified several such events, and they may be described as below:

Game traces

Game start, game end, and game quit are simple events, but do yield useful information. For example, if a significant number of players quit the game at the same point in gameplay, a problem with the challenge difficulty may be indicated, and further testing on the part of game designers is warranted.

Phase change traces

Most games, educational or otherwise, are broken into smaller parts, such as levels, quests, sub-quests, etc. Each of these sections may be connected with others, or be self-contained. The start and end of a phase are useful traces, especially when they carry additional information such as timestamps. To illustrate, if a student restarts a phase a number of times in a short time period, it is possible that the content is difficult for them, and personal guidance from the teacher is required.

Meaningful variable traces

A number of variables are generated by the game in order to keep track of player progress. Some examples are score, lives or health bar, items in the inventory, etc. These can reveal information about a player's understanding of a game. For example, collecting all items available may be an indicator of a player's grasp of game content.

Input traces

These consist of player inputs, such as clicks and keystrokes. They can be used to create a heatmap of player interaction with game screens, yielding information valuable for game designers to improve their game.

B. Creating game-specific traces

While the universal traces described above can provide a useful overview of player progress through the game, Serrano-Laguna et al., (2014) have found that more valuable information can be obtained by building game- and context-specific traces.

For example, a rule could be created to interpret different values of a variable as different levels of student ability. Say, in a certain scenario, a score of above 80 shows mastery of a concept, while a score between 50 to 79 shows intermediate grasp of content.

C. Utilising traces

Identifying universal and game-specific traces is generally a straightforward task. Collecting, analysing and visualising this data may prove to be a more complicated process, greatly dependent upon whether a game allows the collection of trace data, and to what extent. Fortunately, some game engines do facilitate such data collection, and similar functionality can be added to open source game engines (Serrano-Laguna et al., 2014).

2.2.5.2 Game traces usage

Assessment based on game traces has been implemented using the game *Lost in Space <XML>*, which is used for learning XML. In the study, 37 students played the game, and their teachers followed their progress with the help of a visualisation tool that reported real-time learner statistics to teachers, who were then able to assist learners with problems Serrano-Laguna et al., (2014).

2.3 Discussion

The findings of the review indicate that a number of embedded assessment frameworks for learning games have been created in recent years, and demonstrated as being conducive to the creation of valid assessments. These frameworks can cater to general assessment needs, and also more specific ones, such as science assessment, or assessment of implicit knowledge.

The ECD and CyGaMEs frameworks provide the best support for game creation from scratch. They guide game designers to adopt only such practices as are most conducive to the creation of high-quality, effective educational games.

SPRING can be set up to be used with any existent game to automatically assess player performance on game concepts. Its implementation, however, is a rather complex process. Methods for assessment of implicit knowledge are similarly complicated to carry out, but make up for the limitations of ECD in certain contexts.

Both these frameworks can be used to add assessment to already existing games, but it is imperative that the games be chosen with extreme care. Only games with rich learning content and gameplay aligned with content and learning goals should be chosen.

Finally, game traces is a method that stands out because of its simplicity. Using educational game authoring tools created especially for teachers (e-Adventure is a prominent example), game traces can be implemented by teachers themselves, to test exactly what they require.

Perhaps it would not be out of place here to metion that in selecting papers for this review, the author came across a few innovative, promising embedded assessment methods that could not be included here as they had never been tested with a target audience. These methods included the use of Petri nets to record expert traces for comparison with learner traces, employment of HMMs to analyse students' inputs during game play, etc. Even greater in number were papers describing ECD-based assessments created for existing learning games. In these papers, the process of assessment creation, and the rationale behind design choices were elaborated upon, but it was not indicated that any actual testing had taken place. It is certainly to be hoped that in the future more games with assessment make their way to testing and launch.

3 Uses of embedded assessment

Drawing upon papers selected for review, the author has put together a list of the ways in which adding assessment to games has been shown to improve their value for teachers, students, researchers, and game designers.

Formative assessment tools for teachers

Nearly all the papeprs reviewed mentioned how assessment data from games can be utilised by teachers for formative assessment purposes. For example, prediction of test scores using SPRING allows automatic grouping of students, so that the needs of each group can be addressed separately (Falakmasir et al., 2016). Student misconceptions can also be identified, which gives teachers a very good idea of the lines along which subsequent instruction should proceed (Kiili et al., 2015).

In the first chapter, it was shown that teachers are often uncertain of how to incorporate games into their lessons plans, and this lack of awareness of the roles games can play in class keeps teachers from giving game-based learning a try. Learning games that include assessment remove such uncertainty, as it is obvious that they can play an important role in any classroom by serving as convenient formative assessment tools.

Personalised learning and feedback for students

Continuous assessment of student learning during gameplay means that instant feedback can be provided to learners, as was the case in the games *Newton's Playground* and *Selene* (Reese et al., 2012; Shute et al., 2013). If this feedback is well-formulated, it can aid learning greatly.

More comprehensive, end-of-level assessment reports can also be generated for students. Students can use this information about their in-game performance to become more aware of their own progress, and to recognise their weaknesses and strengths (Reese et al., 2012).

When player abilities are known, games can be designed to adapt to the learner, and always present content tailored to the understanding level of the individual (Rowe et al., 2017). For example, the game *Radix Endeavor* prompts players to undertake

additional quests to strengthen skills that they are assessed as not having gained complete mastery of (Clarke-Midura, Rosenheck, & Groff, 2015). This keeps learners in their Zone of Proximal Development. Learners can then move through the game at a pace that is right for them, approaching new topics only after mastering old ones.

Alignment of learning and gameplay

Some creators of embedded assessment frameworks write how being guided by the framework ensures alignment of learning goals, game content, and gameplay. When using frameworks like ECD and CyGaMEs, game designers must first clearly describe what the learning goals of the game are, or, what the game must test, and then plan the rest of the game in alignment with the assessment requirements. As one researcher puts it, using these frameworks "challenges designers to continuously question whether their game teaches what they think it does" (Clarke-Midura et al., 2015). Thus, alignment helps create more useful, pedagogically effective learning games.

Evidence of game-based learning

Teacher scepticism, as we have seen, is one factor keeping games out of school. It is also known that teachers prefer to use games that have been well-tested and proven to be effective in experimental setups. Properly planned, well-designed games that incorporate assessment tools provide strong proof of learning that has happened during gameplay, which should be helpful in reducing doubts about the pedagogical value of educational games.

Evaluation of learning games

Student data from games can prove to be very valuable for educational researchers and game designers. For example, clickstreams can provide an overview of problem areas for players, which can then be checked and improved as needed (Serrano-Laguna et al., 2014). Data from experts playing the game can be analysed to ensure that the game functions like it should (Reese et al., 2012).

4 Professionals' perceptions of game-based assessment

The two preceding chapters would seem to indicate that the field of game-based learning has a lot to gain by more widespread inclusion of embedded assessment in educational games. It is baffling that, in such a scenario, more practitioners of game-based learning are not creating games with assessment capabilities. The author conducted a survey to discover professionals' perceptions of game-based assessment in an attempt to find out why games with embedded assessment are not created more frequently.

4.1 Method

Opinions of individuals active in the field of game-based learning regarding the usefulness of game-based assessment and possible reasons for the low rates of creation of games with assessment capabilities were explored with the help of a questionnaire that included both closed- and open-ended questions. Open-ended questions were included to ensure that participants could express themselves freely, shedding light on aspects of the matter that had not occurred to the author.

4.1.1 Participants

Email invitations to take part in the survey were sent to 270 professionals in the field of game-based learning. 59 invitations were sent to individual email addresses, while the rest were dispatched to subscribers of a mailing list. Snowball sampling was also carried out, with survey respondents being requested to suggest more participants for the survey. Finally, 25 responses were received.

Of the respondents, 16 were between 25 to 39 years of age, and the rest were between 40 to 60 years of age. There were no location-based restrictions on the sample.

Eight participants stated they were involved with game based learning in the capacity of educational researcher cum game designer cum teacher. Seven participants described their role as that of an educational researcher, five as educational researcher plus game designer, four as educational researcher and teacher, and one as game designer. They had between 3 to 26 years of experience working with games

for learning.

4.1.2 Instruments

Data was collected using an online survey created with Google Forms (see Appendix). The survey consisted of two demographic questions, which were followed by six questions exploring participants' professional experience and involvement in game-based learning.

Participants who were familiar with the trend of game-based assessment were then directed to a set of questions designed to gauge participants' views and opinions about the topic. This section began with 5 Likert-scale items, all of which were related to participants' perceptions of the usefulness of game-based assessment. Participants could express on a 5-point scale the extent of their agreement or disagreement with the statements provided to them. The next question explored participants' opinions about why there are only a few games on the market with assessment capabilities. This was a check all that apply type of question, with the option of suggesting additional reasons if the participant desired.

The Likert-scale items were formulated by referring to literature on game-based assessment and identifying what two or more sources described as potential benefits of games with assessment capabilities (Clarke-Midura et al., 2015; Reese et al., 2012; Serrano-Laguna et al., 2014; Shute et al., 2013). The options for the check all that apply item reflected factors that are mentioned in literature as possible obstacles to the more frequent incorporation of assessment in educational games (Chin et al., 2009;). Participants were also given the option to provide more reasons as they thought fit, as the study aimed to explore participant perceptions about games with assessment capabilities.

From among participants aware of game-based assessment, those who had personal experience working on such a project were directed to two more questions. These were both open-ended questions, framed with the view of obtaining information about participants' work in game-based assessment, such as the kind of games they created, and their experiences during the project.

Participants who stated that they did not know about game-based assessment were

presented with a brief introduction to the topic. Two closed yes-no questions followed, aimed at gauging whether the participant found the idea of game-based assessment interesting and useful based on the information they had just read.

4.1.3 Data analysis

To analyse Likert-scale data, the median and Inter-Quartile Range (IQR) were calculated. The median is a measure of central-tendency, and indicates roughly what the "typical" response is. The IQR is a measure of dispersion, and shows whether responses are clustered together, or spread out over the whole possible range.

Other data was qualitative in nature, and was coded and organised into categories for better interpretation and presentation.

4.2 Results

Of the 25 survey participants, all but 3 knew about the use of embedded assessment in learning games. Among the three participants who were not familiar with game-based assessment at the time of taking the survey, only two expressed an interest in learning more about the topic.

The results presented below consist of the opinions of the 22 participants familiar with assessment in the field of game-based learning.

4.2.2 Perceptions about the usefulness of assessment in learning games

Participants familiar with the idea of game-based assessment were asked to express agreement or disagreement regarding the usefulness of games with assessment capabilities in five separate areas.

Most respondents indicated agreement with the idea that "game-based formative assessment can help teachers to easily tailor instruction to learners' needs" (Mdn = 4, IQR = 1). Seven participants were undecided, and none disagreed with the statement.

Similar responses were received regarding the ability of game-based formative assessment to encourage students to take ownership of their learning (Mdn = 4, IQR = 2). One participant disagreed with the statement, while four participants were unsure of how to answer.

Given the statement "focus on assessment helps create better educational games, as an effort has to be made to properly align learning goals, game content, and gameplay", again, a majority of respondents expressed agreement (Mdn = 4, IQR = 2). Three participants disagreed, in varying degrees, with the statement, and five were undecided.

Participants next expressed their views about the statement "educational games with incorporated assessment provide proof of learning, which can help address scepticism about the effectiveness of game-based learning". Once again, most participants agreed with this statement (Mdn = 4, IQR = 1). Two participants did not concur, and two others could not make up their minds either way.

That "game-based assessment of learners is better for gauging the effectiveness of an educational game than external, paper-based pre- and post-tests" was the last statement in this section of the survey. Most participants agreed with the statement (Mdn = 4, IQR = 2), but three disagreed in varying degrees, and seven neither agreed nor disagreed.

Participant responses are represented graphically in Figure 3.

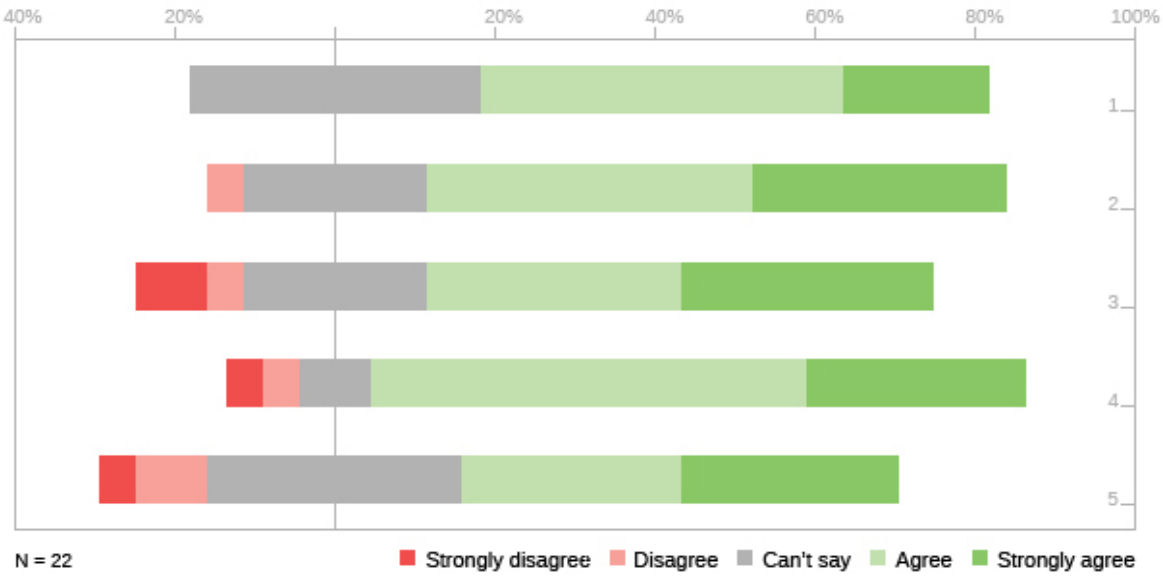

Figure 3. Perceptions of the usefulness of game-based assessment

A significant finding was the differences between the opinions of those who had

first-hand experience of creating learning games with assessment capabilities (users) and those who did not have such personal experience in the matter (non-users). All responses in the negative, except for one, came from the latter group, as depicted in Figures 4 and 5.

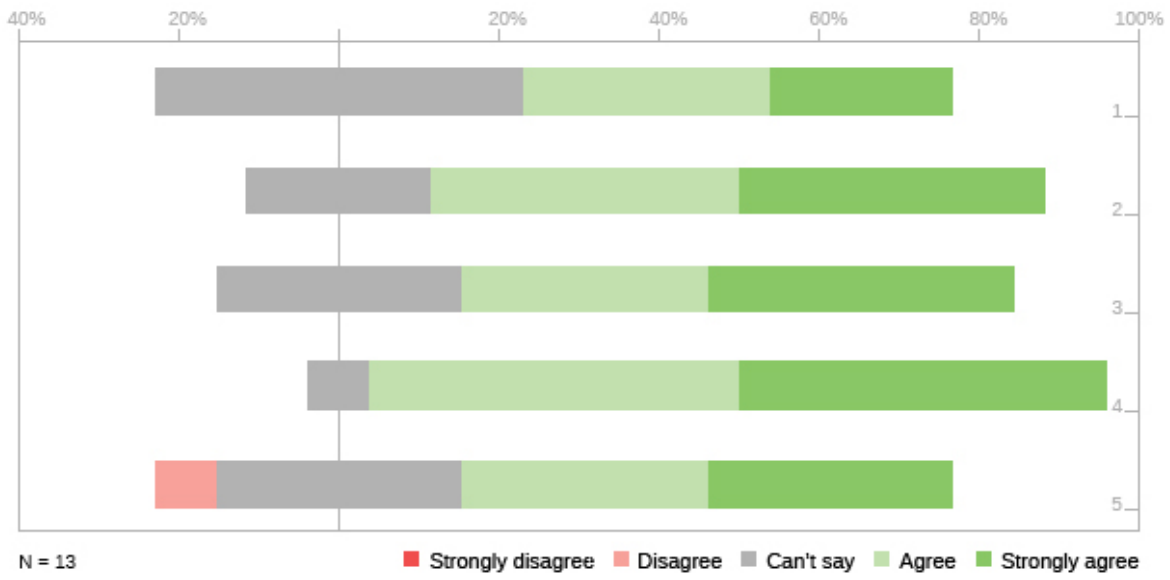

Figure 4. Users' perceptions of the usefulness of game-based assessment

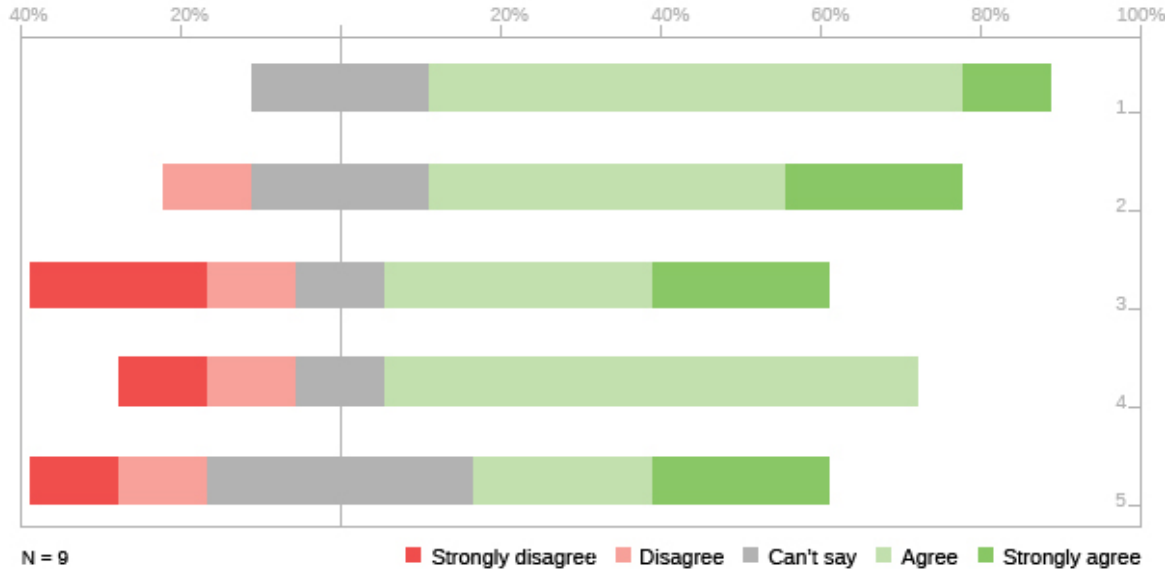

Figure 5. Non-users' perceptions of the usefulness of game-based assessment

4.2.3 Obstacles to the adoption of embedded assessment frameworks

The author had proposed, relying on available literature, that learning games with

embedded assessment are scarce, at least partly, because educational researchers and game designers are not aware of the advances recently made in the field of game-based assessment. Eight individuals agreed with this view.

Fifteen participants expressed support for the next available option, which stated that incorporating assessment into a game is a time-consuming process, and this discourages game designers from attempting the same.

A number of participants offered additional reasons to explain the paucity of learning games that include assessment tools:

a) Four participants suggested that professionals involved in learning game creation do not focus on assessment because it is very difficult to maintain the balance between entertainment, learning and assessment.

b) Two participants thought that learning games with assessment are rare because there is currently "no market" for them. It was pointed out by another participant that compared to the extra effort that designing assessment for a game requires, the increase in market value of the game is insignificant.

c) One participant also advanced the idea that assessment in learning games would make them regionally specific, narrowing down their target markets.

d) It was conjectured by another respondent that designing valid assessments is difficult, and not a skill designers/developers have.

4.2.4 Experiences of creating games with assessment capabilities

Thirteen individuals stated that they had first-hand experience in designing a learning game that could assess players. Some of them provided details of their projects, and described how they handled the process of creating a game around assessment.

Most participants stated that their project took about two to three years to complete.

The setbacks described were many:

a) It is difficult to identify in sufficient detail the knowledge and skills to be tested in the game.

b) Inferring results from player behaviours is not easy.

c) Ensuring validity of assessment, and the need for an equivalent parallel test for this purpose, are time-consuming, complex tasks.

d) Correct diagnosis of problems such as guessing and gaming the system can be a complicated process.

e) There may be difficulty in maintaining learner engagement.

f) Reluctance of teachers to use games in the classroom can be difficult to overcome.

4.3 Discussion

On analysing survey responses, it would appear that most participants view the trend of assessment using educational games in a positive light and believe it can add value to learning games.

Having the experience of actually implementing assessment in a game appeared to positively affect participants' perceptions of the usefulness of game-based assessment. It may be conjectured that a closer acquaintance with games involving embedded assessment reduces doubts about their usefulness, and, in any case, does not cause any disillusionment in those already convinced of their merits.

There are numerous perceived and real barriers to the more frequent and widespread adoption of the practice of creating learning games with assessment capabilities. Participants who were creators of such games elaborated upon these obstacles, detailing the impediments they faced while creating games with assessment capabilities.

A surprisingly high number of participants knew about game-based assessment, contrary to the author's expectations. This could indicate that there is already sufficient community awareness about the subject. On the other hand, more than a third of the participants expressed the opinion that many in the field are simply not

familiar with the idea of assessment being designed as a part of learning games. Lack of awareness, then, should be considered as a factor negatively affecting the popularity of game-based assessment. (It is possible that snowball sampling resulted in the sample being composed of participants conversant with the topic, but participants recruited in this way formed a very small portion of the sample.)

Thus, it may be concluded that while there seems to be little scepticism about the merits of game-based assessment, there is clearly a need of addressing obstacles to the creation of games with embedded assessment. This is the topic of the next chapter.

It is to be noted, though, that one impediment to the creation of games with assessment cannot be dealt with here: it seemed to be a somewhat common complaint among survey participants that maintaining the playfulness of a learning game is not easy when assessment is to be part of it, too. This, perhaps, is a problem with learning games as a whole, whether with or without assessment, and is beyond the scope of this study.

5 Recommendations

In this chapter, the author uses what has been learnt through the systematic review to address perceived obstacles to the creation of games with assessment, as voiced by survey participants. To overcome these obstacles, instructional game developers must:

Collaborate

Creating an assessment is not easy for most game designers, as it is unlikely that they have received training in such a skill. Not only does assessment creation require familiarity with psychometrics, it demands thorough knowledge of the domain being assessed. It is not surprising, therefore, that a significant number of survey participants found the making of an educational game with embedded assessment laborious, as considerable time and effort must be spent on domain research and creation of quality assessments, especially by someone lacking experience in these fields.

Collaboration with domain experts (for example, teachers who are expected to use the learning game) could prove to be a solution to this problem. Not only would a team with such expertise be able to achieve its goal more quickly, the resultant work is also likely to be more satisfactory and suited to the actual needs of teachers.

Choose the right framework

The assessment frameworks described in this study can meet many, diverse needs of educational researchers and game designers embarking upon the creation of a game with assessment capabilities. Before settling down to work with one assessment design model, instructional game developers are encouraged to make a study of all available frameworks, and weigh the pros and cons of each with reference to their upcoming project.

Figure 6 represents a simple guide to choosing embedded assessment frameworks, taking into account time limitations and lack of technical know-how. Of course, users will have to consider other factors, too, such as the aim of their learning game project.

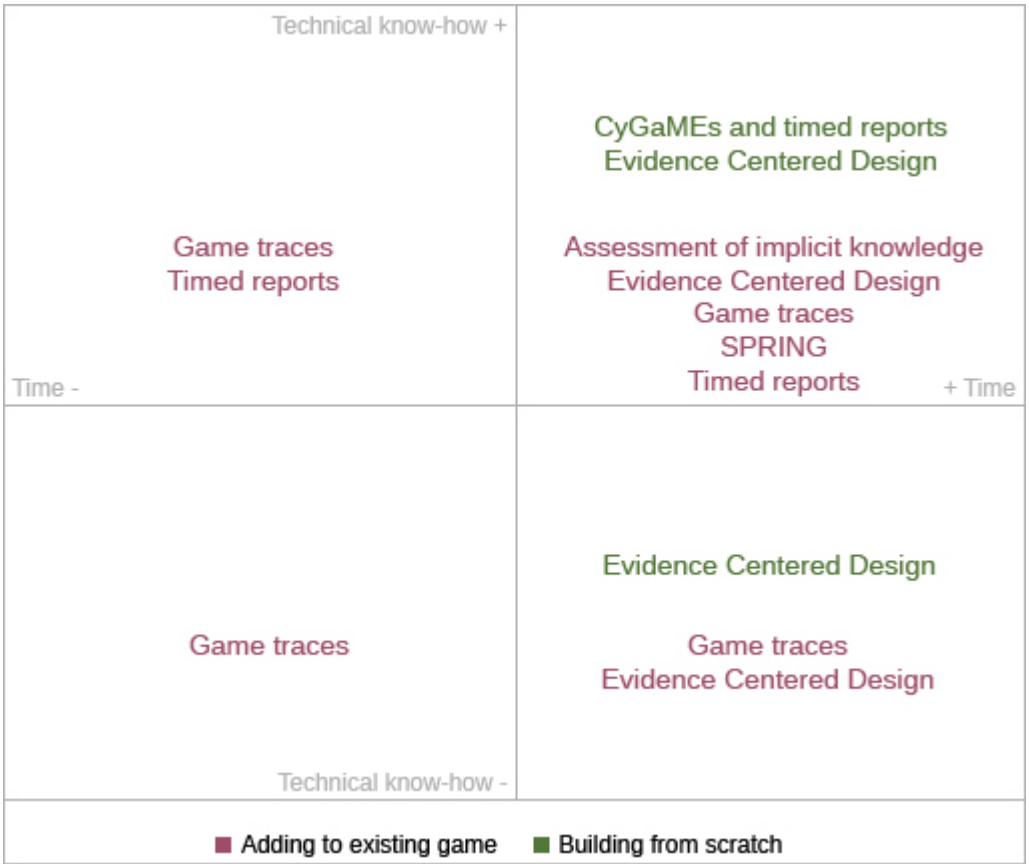

Figure 6. Choosing an embedded assessment framework

Choose the right kind of assessment

All assessment is not equal, and that is a good thing as far as game-based assessment goes. For use as summative assessment, such as in a high-stakes situation, test instrument validity and testing are essential. However, all assessment does not have to match these high standards. For example, teachers carry out formative assessment routinely, using methods such as “show of hands”, “corners” (to achieve grouping of students), “bump in the road” (students are asked to write down what they found confusing about the topic being taught), etc. These simple methods provide teachers with exactly what they need to plan further instruction.

Some survey participants described the difficulty they faced in creating valid assessments for their games, along with parallel traditional assessments for testing. This kind of rigour is certainly to be appreciated, and more of such carefully designed learning games are needed. However, it is not realistic for all educational game developers to follow the same path. Much simpler assessment can also be useful in game-based learning.

Developers of instructional games should look at assessment from various perspectives, ask target audience (teachers) for their opinions, and select or create the kind of assessment most suited to their project, taking into account the outcomes sought, team compostion, funds, time limits, etc..

Conclusion

Some teachers are sceptical about the use of games in the classroom. Many want to use them, but are unsure of how to do so. Many bemoan not being able to find games suited to their curriculum requirements. Most of the remaining teachers who do use learning games with their students see them mostly as motivational tools and not instruments to aid and enhance learning.

On the other hand, there is a small community of educational researchers and instructional game developers convincingly (at first sight) demonstrating and enthusiastically writing about how games with embedded assessment have the potential to fix all the problems listed above, and to take digital learning games to the next level.

Where, then, are all the games with assessment capabilities? An internet search throws up only a handful of results for fully functional, playable games that can boast of comprising effective assessment tools.

This puzzling situation gave rise to the current study.

The author first undertook to learn more about embedded assessment frameworks. On conducting a review of recent literature, five viable frameworks were discovered which had been used for the creation of embedded assessment for educational games. Most had been employed to create only a couple of games at most, but the results were certainly promising.

Then, the merits of embedded assessment for learning games were explored, and it was found that the uses of assessment in games are diverse and innovative. One should be wary of cure-alls, but what educational researchers have achieved with game-based assessment is certainly exciting. It seemed clear by now that serious attempts at the creation of more games with assessment capabilities are warranted.

An investigation into why such attempts are scarce followed. A survey with a small pool of participants provided useful information about professionals' perceptions of assessment incorporation in learning games. While most participants were of the opinion that game-based assessment has much to offer to improve game-based

learning, the list of obstacles to the creation of games with assessment was long.

Finally, the author created guidelines that may ease some of the difficulties survey participants mentioned as accompanying the process of creation of game-based assessment. Though simple, it is hoped that these guidelines can encourage game-based learning practitioners who are deliberating whether they should create a game with embedded assessment to commence with their efforts.

Kokkuvõte

Osa õpetajaid on õppetundides mängude kasutamise suhtes skeptilised. Paljud soovivad neid kasutada, kuid ei tea kuidas seda teha. Õpetajad avaldavad nördimust, et nad ei suuda leida õppekavaga sobituvaid mänge. Enamik järelejäänud õpetajatest, kes kasutavad hariduslikke mänge õpilaste õpetamisel, näevad neis mängudes pigem motiveerimise vahendit ja mitte õpetamise efektiivsemaks tegemise vahendit.

Samas on olemas väike rühm haridusuurijaid ja hariduslike mängude arendajaid, kes demonstreerivad (esmapilgul) veenvalt ja kirjutavad innustatult sellest, kuidas mängu sisene hindamine omab potentsiaali parandada eelpool nimetatud probleemid ning viib digitaalsed hariduslikud mängud uuele tasemele.

Kuhu jäävad sellisel juhul hindamise võimekusega mängud? Internetiotsing annab vaid väikese koguse tulemusi, mis on täielikult toimivad mängud efektiivsete hindamistööriistadega.

Eelpool kirjeldatud segadust tekitav olukord andis tõuke käesolevaks uurimuseks.

Esiteks võttis autor eesmärgiks uurida, millised on mängude sisese hindamise metoodid. Viies läbi teaduskirjanduse analüüsi, avastati viis efektiivset mudelit, mida on kasutatud hindamise integreerimisel hariduslikesse mängudesse. Enamuse puhul kasutati neid vaid mõnes mängus, kuid tulemused on paljulubavad.

Seejärel uuriti integreeritud hindamise eeliseid ning leiti, et hindamise kasutamine mängudes on mitmekülgne ja innovatiivne. See, mida haridusuurijad on mängupõhises hindamises saavutanud, tundus ühelt poole liiga hea, et olla tõsi, kuid teiselt poolt tekitas huvi. Nüüd oli autor veendunud, et tõsised ettevõtmised hindamisega mängude loomisel on kindlad tulema.

Järgnes uuring tuvastamaks, miks selliste ettevõtmiste arv on väike. Uuringuga, mis viidi läbi kitsa sihtrühmaga, koguti andmeid asjatundjate hinnangust hindamistööriistade lisamise kohta hariduslikudesse mängudesse. Samal ajal kui enamik osalejaid oli arvamusel, et mängupõhine hindamine saab pakkuda palju mängupõhise õppimise arendamisel, oli vastustes toodud hindamisega mängude loomise takistuste nimekiri pikk.

Lõpuks koostas autor uuringu põhjal juhised, mis võivad kergendada osade mängupõhise hindamise loomise takistuste, mida uuringus osalejad tõid välja , ületamist. Kuigi need on lihtsad, loodab autor, et need juhised julgustavad hariduslike mängude arendajaid, kes on kahtleval seisukohal, looma mänge integreeritud hindamisega.

References

All, A., Castellar, E. P. N., & Van Looy, J. (2016). Assessing the effectiveness of digital game-based learning: Best practices. *Computers & Education*, *92–93*, 90–103.

An, Y.-J., Haynes, L., D'Alba, A., & Chumney, F. (2016). Using Educational Computer Games in the Classroom: Science Teachers' Experiences, Attitudes, Perceptions, Concerns, and Support Needs. *Contemporary Issues in Technology & Teacher Education*, *16*(4), 415–433.

Asbell-Clarke, J., Rowe, E., & Sylvan, E. (2013). Assessment design for emergent game-based learning. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (pp. 679–684). Paris, France: New York, NY: ACM.

Baker, R., & Yacef, K. (2009). The State of Educational Data Mining in 2009: A review and future visions. *Journal of Educational Data Mining*, *1*, 3–17.

Barbour, M., & Evans, M. (2009). Making sense of video games: Pre-service teachers struggle with this new medium. Presented at the 20th international conference of the society of informational technology and teacher education, Chesapeake, VA: Chesapeake, VA: AACE.

Bellotti, F., Kapralos, B., Lee, K., Moreno-Ger, P., & Berta, R. (2013). Assessment in and of serious games: an overview. *Advances in Human-Computer Interaction*, 1–11.

Black, P., & Wiliam, D. (1998). Inside the black box: raising standards through classroom assessment. *Phi Delta Kappan*, *80*(2), 139–148.

Bourgonjon, J., De Grove, F., De Smet, C., Van Looy, J., Soetaert, R., & Valcke, M. (2013). Acceptance of game-based learning by secondary school teachers. *Computers & Education*, *67*, 21–35.

Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., … Pereira, J. (2016). An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Computers*

& Education, *94*, 178–192.

Callaghan, N. (2016). Investigating the role of Minecraft in educational learning environments. *Educational Media International*, *53*(4), 244–260.

Calvet Liñán, L., & Juan Pérez, Á. A. (2015). Educational Data Mining and Learning Analytics: Differences, similarities, and time evolution. *RUSC.Universities and Knowledge Society Journal*, *12*(3), 98–112.

Cannata, M., Haynes, K. T., & Smith, T. M. (2013). *Reaching for rigor: Identifying practices of effective high schools*. National Center on Scaling Up Effective Schools.

Chin, J., Gamson, W., & Dukes, R. (2009). Assessment in simulation and gaming: A review of the last 40 years. *Simulation & Gaming*, *40*(4), 553–568.

Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. (2014). *Digital games, design and learning: A systematic review and meta-analysis*. Menlo Park, CA: SRI International.

Clarke-Midura, J., Rosenheck, L., & Groff, J. (2015). Designing games for learning and assessment: The Radix endeavour (pp. 343–346). Presented at the 14th International Conference on Interaction Design and Children, Boston, MA: New York, NY: ACM.

Cullingford, G., Mawdesley, M. J., & Davies, P. (1979). Some experiences with computer based games in civil engineering teaching. *Computers & Education*, *3*, 159–164.

DeMars, C. (2010). *Item Response Theory*. New York, NY: Oxford University Press.

DiCerbo, K. E., Bertling, M., Stephenson, S., Jia, Y., Mislevy, R. J., Bauer, M., & Jackson, G. T. (2015). An application of exploratory data analysis in the development of game-based assessments. In C. Loh, Y. Sheng, & D. Ifenthaler (Eds.), *Serious Games Analytics* (pp. 319–342). Cham: Springer.

Falakmasir, M. H., González-Brenes, J. P., Gordon, G. J., & DiCerbo, K. E. (2016). A data-driven approach for inferring student proficiency from game activity logs. In *Proceedings of the Third ACM Conference on Learning @ Scale* (pp.

341–349). Edinburgh, Scotland: New York, NY: ACM.

Fishman, B., Riconscente, M., Snider, R., Tsai, T., & Plass, J. (2014). *Empowering educators: Supporting student progress in the classroom with digital games (Part 2)*. Ann Arbor: University of Michigan.

Gaudelli, W., & Taylor, A. (2011). Modding the Global Classroom? Serious Video Games and Teacher Reflection. *Contemporary Issues in Technology and Teacher Education, 11*(1), 70–91.

Gipps, C. (1994). Developments in educational assessment: What makes a good test? *Assessment in Education: Principles, Policy & Practice*, *1*(3), 283–292.

Habgood, M. P. J., & Ainsworth, S. E. (2011). Motivating children to learn effectively: exploring the value of intrinsic integration in educational games. *Journal of the Learning Sciences*, *20*(2), 169–206.

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, *77*(1), 81–112.

Ito, M. (2009). *Engineering play - A cultural history of children's software*. The MIT Press.

Kafai, Y., & Burke, Q. (2015). Constructionist gaming: Understanding the benefits of making games for learning. *Educational Psychologist*, *50*(4), 313–334.

Kiili, K. J. M., Devlin, K., Perttula, A., Tuomi, P., & Lindstedt, A. (2015). Using video games to combine learning and assessment in mathematics education. *International Journal of Serious Games*, *2*(4), 37–55.

Kitchenham, B. (2004). *Procedures for performing systematic reviews*. Keele University.

Klopfer, E., Osterweil, S., & Salen, K. (2009). *Moving learning games forward: obstacles, opportunities, and openness*. Cambridge, MA: Massachusetts Institute of Technology.

McTighe, J., & O'Connor, K. (2005). Seven practices for effective teaching. *Educational Leadership*, *63*(3), 10–17.

Merrill, M. D. (2002). First Principles of Instruction. *Educational Technology*

Research and Development, *50*(3), 43–59.

Meyer, P. (2010). *Reliability*. New York, NY: Oxford University Press.

Mislevy, R. J., & Riconscente, M. M. (2005). *Evidence-centered assessment design: Layers, structures, and terminology* (PADI Technical Report). Menlo Park, CA: SRI International - Center for Technology in Learning.

Ninaus, M., Kiili, K. J. M., McMullen, J., & Moeller, K. (2017). Assessing fraction knowledge by a digital game. *Computers in Human Behavior*, *70*, 197–206.

Reese, D. D., Seward, R. J., Tabachnick, B. G., Hitt, B. A., Harrison, A., & McFarland, L. (2012). Timed report measures learning: Game-based embedded assessment. In D. Ifenthaler, D. Eseryel, & X. Ge (Eds.), *Assessment in Game-Based Learning* (pp. 145–172). New York, NY: Springer.

Richards, J., Stebbins, L., & Moellering, K. (2013). *Games for a digital age: K-12 market map and investment analysis*. New York: The Joan Ganz Cooney Center at Sesame Workshop.

Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, *40*(6), 601–618.

Rowe, E., Asbell-Clarke, J., Baker, R. S., Eagle, M., Hicks, A. G., Barnes, T. M., … Edwards, T. (2017). Assessing implicit science learning in digital games. *Computers in Human Behavior*, *76*, 617–630.

Ruggiero, D. (2013). Video games in the classroom: The teacher point of view. Presented at the 8th International Conference on the Foundations of Digital Games, Chania, Greece. Retrieved from http://www.fdg2013.org/program/workshops/papers/G4L2013/g4l2013_02.pdf

Salen, K., & Zimmerman, E. (2003). *Rules of play: Game design fundamentals*. MIT Press.

Selwyn, N. (2016). Minding our language: why education and technology is full of

bullshit … and what might be done about it. *Learning, Media and Technology*, *41*(3), 437–443.

Serrano-Laguna, Á., Torrente, J., Moreno-Ger, P., & Fernández-Manjón, B. (2014). Application of learning analytics in educational videogames. *Entertainment Computing*, *5*(4), 313–322.

Short, D. B. (2012). Teaching scientific concepts using a virtual world - Minecraft. *Teaching Science*, *58*(3), 55–58.

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). Charlotte, NC: Information Age.

Shute, V. J., Ventura, M., Bauer, M., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. In U. Ritterfield, M. J. Cody, & P. Vorderer (Eds.), *Serious Games: Mechanisms and Effects* (pp. 295–321). New York: Routledge.

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.

Siemens, G., & Baker, R. (2012). Learning analytics and educational data mining: Towards communication and collaboration. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 252–254).

Squire, K. D. (2004). *Replaying history: Learning world history through playing "Civilization III"* (Doctoral dissertation). Indiana University. Retrieved from ProQuest Dissertations and Theses. (3152836)

Stiggins, R. (2008). *Assessment FOR learning, the achievement gap, and truly effective schools*. Presented at the Educational Testing Service and College Board conference, Washington, DC. Retrieved from https://www.ets.org/Media/Conferences_and_Events/pdf/stiggins.pdf

Susi, T., Johannesson, M., & Backlund, P. (2007). *Serious games - An overview*.

University of Skövde, Sweden.

Takeuchi, L., M., & Vaala, S. (2014). *Level up learning: A national survey on teaching with digital games*. New York: The Joan Ganz Cooney Center at Sesame Workshop.

Tanes, Z., & Cemalcilar, Z. (2010). Learning from SimCity: An empirical study of Turkish adolescents. *Journal of Adolescence*, *33*, 731–739.

Taras, M. (2005). Assessment - Summative and formative - Some theoretical reflections. *British Journal of Educational Studies*, *53*(4), 466–478.

Weitze, C. L. (2016). Student learning-game designs: Emerging learning trajectories (pp. 756–764). Presented at the The 10th European Conference on Games Based Learning, Paisley, Scotland: Academic Conferences and Publishing International.

Westera, W. (2015). Games are motivating, aren´t they? Disputing the arguments for digital game-based learning. *International Journal of Serious Games*, *2*(2), 3–17.

Appendix

Your Background

What is your age?

◯ 18 to 24

◯ 25 to 39

◯ 40 to 60

◯ 60 plus

Which country are you currently based in?

_____

Your Experience with Game-Based Learning

In what capacity are you associated with game-based learning? Check all that apply.

☐ Educational researcher

☐ Game designer

☐ Teacher

☐ Other _____

How long have you been working with educational games?

_____

How many educational games have you designed?

_____

Which environment have you designed educational games for?

☐ School

☐ Home

☐ Other _____

Which school section/corresponding age group have your games been targeted at? Check all that apply.

☐ Pre-school

☐ Grades 1 and 2

☐ Grades 3 to 5

☐ Grades 6 to 8

☐ Grades 9 to 12

☐ Other _____

Which subjects have your games dealt with? Check all that apply.

☐ Art

☐ Biology

☐ Chemistry

☐ Computer science

☐ History

☐ Languages

☐ Maths

☐ Physics

☐ Other _____

Are you familiar with the trend of game-based assessment?

◯ Yes (Particpiant directed to “Your Views on Game-based Assessment”)

◯ No (Participant directed to “About Game-based Assessment”)

Your Views on Game-based Assessment

Game-based formative assessment can help teachers to easily tailor instruction to learners' needs.

Strongly disagree ◯ ◯ ◯ ◯ ◯ Strongly agree

Game-based formative assessment encourages students to take ownership of their learning.

Strongly disagree ○ ○ ○ ○ ○ Strongly agree

Focus on assessment helps create better educational games, as an effort has to be made to properly align learning goals, game content, and gameplay.

Strongly disagree ○ ○ ○ ○ ○ Strongly agree

Educational games with incorporated assessment provide proof of learning, which can help address scepticism about the effectiveness of game-based learning.

Strongly disagree ○ ○ ○ ○ ○ Strongly agree

Game-based assessment of learners is better for gauging the effectiveness of an educational game than external, paper-based pre- and post-tests.

Strongly disagree ○ ○ ○ ○ ○ Strongly agree

Why, in your opinion, have only a few educational games with a focus on assessment been launched so far?

☐ I do not think this is the case.

☐ Game designers are unaware of advances in game-based assessment.

☐ Designing games with assessment is a time-consuming process.

☐ Other _____

If you can recall the titles of any educational games that can be used for assessment, please list them here.

_____

Have you had personal experience with creating an educational game with assessment capabilities?

○ Yes (Particpiant taken to “Your Experience with Game-based Assessment”)

○ No (Participants taken to “Suggestions and Comments”)

About Game-based Assessment

In recent years, game-based assessment has been the focus of a number of researchers in the field of game-based learning. This involves educational games designed with assessment in mind from the start. First, learning goals are decided upon. Next, game content and gameplay are designed to elicit from players actions that provide evidence of progress towards the learning goals. Assessment tools analyse large volumes of player data to measure student performance, and to draw inferences about students' learning states. This data is then presented in an accessible format to teachers, students, and other stakeholders.

Teachers and students can then use this data to improve instruction and learning. Game developers can use it to test their creation. Proponents of game-based learning can provide strong proof in favour of the effectiveness of educational games.

Based on the information provided above, do you think the inclusion of assessment can help create better and more useful educational games?

◯ Yes

◯ No

◯ Not sure

Would you be interested in learning more about game-based assessment?

◯ Yes

◯ No

Your Experience with Game-based Assessment

Please provide some information (game title, game website, publications about the game, etc.) about your work in game-based assessment.

_____

Please describe your experience (the length of the project, difficulties faced, lessons learnt, etc.) of working with game-based assessment.

_____

Suggestions and Comments

It has not been easy finding participants for this survey! If you know someone who is also involved in game-based learning, please provide below their name and affiliations, or their contacts details.

_____

If you have any comments, this is the place for them.

_____