Validation of game scenarios for the assessment of professional competence: Development of a serious game for system managers in training

Validation of Game Scenarios for the Assessment of Professional Competence: Development of a Serious Game for System Managers in Training

Hans G.K. Hummel[a], Rob J. Nadolski[a], Desirée Joosten- ten Brinke[a], Liesbeth K.J. Baartman[b]

[a]Welten Institute
Research Centre for Learning, Teaching and Technology
Faculty of Psychology and Educational Sciences
Open University of the Netherlands
[hans.hummel, rob.nadolski, desiree.joosten-tenBrinke]@ou.nl

[b]University of Applied Sciences of Utrecht
liesbeth.baartman@hu.nl

ABSTRACT

Serious games hold potential for fostering the acquisition of more complex problem solving skills in professional practice. However, until now the empirical evidence on these workplace learning effects of serious games has remained rather scarce. Therefore such games have hardly been adopted for assessment purposes. This article argues why a validation method is needed that points out and controls what and where learners are learning from games. The core of the method entails mapping the learning activities on the performance indicators and outputs, as derived from the formal attainment levels in vocational education. In this study we have elaborated and applied a validation method for the development of a scenario-based assessment game for system managers in (secondary vocational) education. The method provides a general procedure, practical guidelines, and assessment forms, that can be used beyond this educational context and domain by those interested in more dynamic and motivating ways to assess the acquisition of complex skills in workplace learning.

Keywords

Serious games, seamless assessment, validation method, professional competence, game scenarios

1. INTRODUCTION

Several authors have argued the strength of games as assessment engines [1, 2]. Gee and Schaffer argue that games are good learning engines because they are first good assessment engines [1]. Interest in and use of serious games for learning has grown over the last decade, but until now the empirical evidence on the professional learning effects of serious games has remained rather scarce. As a consequence such games have hardly been adopted for assessment purposes yet.

For the true adoption of serious games for both learning and assessment, we first of all need to employ some type of validation method that makes us understand better what a learner is learning from playing the game, to what degree, and in which contexts while at the same time no sacrifices are made to reliability, and validity of assessment and to the core essence of the highly dynamic interactive nature of games.

The study presented in this article will describe a method to validate game scenarios for the assessment of professional competence, and describe the application of this method on an assessment game that was developed for system managers within secondary vocational education. Core to the approach and developed game is that all performance indicators (as were derived from the formal attainment level) have been clearly mapped on the learning activities and outputs (within the game scenario). We will describe this validation method and argue why our approach can be useful beyond this educational context and domain for those interested in more dynamic and motivating ways to formatively assess professional competence in action.

The remainder of this introduction will now further explain the need for seamless assessment using scenario-based gaming (section 1.1), explain the need for more transparency using a validation method (section 1.2), and introduce the educational context and assessment game (and its scenario) we have used for this study (section 1.3). The validation method itself (section 2) and the game obtained by applying the method (section 3) will then be elaborated and presented in subsequent sections. We will conclude (section 4) with an evaluation of this validation method and suggest future research.

1.1 Seamless assessment in games

The main challenge involved with creating games that assess competencies key to workplace learning is to consider their highly dynamic interactive nature, being unobtrusive to the player, while not sacrificing reliability and validity in the assessment process. The integration of formative assessment within game play should be 'seamless'. Gee and Shaffer expect games to reform current educational assessment (mainly facts and knowledge), and lead to radical transformation towards learning for 21st century skills [1]. As they and other educationists state it: "Assessment is the tail that wags the dog of learning". Assessment of learning is the process of using data to demonstrate that stated learning objectives are actually being met by a learner [3, 4]. Creating scenarios with learning activities closely aligned with the learning objectives is key in ensuring learning goals will be met. In other words, assessments need to be aligned with learning objectives and with the learning activities (i.e. constructive alignment [5, 6]). As a consequence, the domain of assessment is in transition from a perspective with an emphasis on summative assessment to a more balanced assessment program in which summative assessment is balanced with formative assessments. Redecker et al. describe the stepwise development from 1st generation in the

1990s (automated administration and scoring) and 2nd generation in the 2000s (more adaptive) to 3rd generation from 2010 (continuous, unobtrusive, more formative assessment), which is supposed to further include behavioral tracking in immersive and game-based environments [7]. For several formative assessment methods, like giving feedback, feed up and feed forward, working with rubrics or self and peer assessment, evidence is available that formative assessment is effective for learning [8]. However, as stated before, a major impediment for exploiting games for the formative assessment of more complex skills purposes is the current lack of proof on the efficacy and impact of serious games on learner achievement [9, 10]. This type of learning and assessment requires more complex, seamless but also transparent validation methods and assessment procedures, which we will present in this article. According to Corti: "Serious games will only grow as an industry if the learning experience is definable, quantifiable and measurable" [11].

1.2 Validation methods and assessment

Validation is the process of building arguments to support the claims and decisions that are made from assessment scores [12]. Validation methods evaluate whether assessment achieves its purposes, i.e. the fitness for purpose[13]. Fitness for purpose encompasses the way results of an assessment are interpreted and used by the educators and students. A validation model provides information whether the assessment is in line with the learning objectives and the learning scenario. This implicates that assessments are representative for and balanced over the learning objectives. Validation has to be argument-based using two kinds of arguments [12]. Interpretive arguments specify the proposed interpretations and uses of scores and are used as a starting point for validation. This includes the analyses of performance indicators and the learning activities. Validity arguments then evaluate the plausibility of these interpretations and uses by evaluating to which extend performance indicators are covered by learning activities and the availability of assessment procedures, instructions and forms. A validation method to assess complex skills therefore has to involve different kinds of evidence, like the implementation of assessment procedures, the translation of the learning objectives into the learning scenario, the expert judgments and the documentation.

With the implementation of competence-based education comes the need for other, more dynamic forms of assessment. More classical forms of testing and assessment have gradually been replaced by so called competence assessment programs (CAP), where the mere application of classical criteria for reliability and validity no longer suffices. Such programs and the examination projects within also need to comply to the new demands of competence-based assessment, like acceptability, authenticity, meaningfulness, cognitive complexity, fairness, fitness for purpose, reproducibility of decision, educational consequences, self-regulated learning, transparency, comparability and costs and efficiency [14]. For this educational context, the general quality of education and assessment is considered to be problematic by the various stakeholders involved [15]. Evidently, serious games offer great potential for CAP as they provide highly engaging and dynamic environments with authentic tasks at the core for the development of professional competence.

The qualitative problem with assessment is largely caused by the lack of clear design criteria and standards for examination which make that the various examination projects differ largely and are hard to compare. Another important aspect that has hampered the uptake of more dynamic forms of education and assessment (like serious games) is the lack of sufficient evidence-based research into these innovations, even though research did reveal that the way assessment is conducted is a major determinant of reaching graduation.

1.3 Example game: Events Agency Galema

Secondary vocational education is (in the Netherlands) largely offered by so called Regional Education Centres, large training institutes that on the average serve about 30,000 students each. The attainment levels for each profession and educational level (of which there are four) are documented in so called Qualification Dossiers which have been accredited on a national level. The Stichting Praktijk Leren (SPL) is the Dutch Foundation on work-based learning that operates closely together with branch organizations for various professions, and has the responsibility to stimulate, coordinate and coach the development of more innovative ways of professional training and assessment in secondary vocational education. Recently SPL decided to aim for an integral, transparent and proven system of examination projects that covers all Core Tasks within the Qualification Dossiers. To validate such assessments currently two instruments are available and used for their design: quality criteria for CAP [14], and frameworks of the educational inspection [15]. The development of the learning and assessment games is done by applying the EMERGO game platform [16]. Eventually SPL strives to have each core task assessed by a game. The curriculum for training System Managers on attainment level 4 has been taken as first pilot, one learning game and one assessment game have been developed so far. This study deals with the developed assessment game which is called 'Events Agency Galema' (name of the case and virtual contractor**).**

The examination project 'Events Agency Galema' is based on a practical case that has to be done within a virtual company 'ITadvice4U'. This means that students are largely assessed while carrying out tasks on their computer. The game is based on a scenario with consecutive learning activities that have to be carried out within the virtual company by guidance of a virtual coach, and partly by having face-to-face talks with the teacher in real life. The main task that is given to the student: develop a new system for project management for a agency that organizes events. For this, the student performs a needs-analysis, distills a functional and technical design of the new system, draws up a plan for developing the new system, tests a first version, and writes a test report. This all yields a total study load of about two days to pass the assessment game.

2. METHOD

This section will briefly introduce the validation method we used and its four steps (section 2.1), then explain the first two steps (Performance Indicators and Game Scenario) in section 2.2, and on the last two steps (Mapping and Assessment Procedures) in section 2.3. The next section will present the results of applying this validation method on the Galema game.

2.1 Validation method

The validation method essentially is comprised of executing following four steps procedure: (1) Analyze the Qualification Dossier, with having Performance Indicators as its outcome; (2) Develop learning activities, with having a detailed Game Scenario

as its outcome; (3) Evaluate to which extend performance indicators are covered by learning activities, with having a Mapping of intended performance on activity; and (4) Distill Assessment procedures, instructions and forms. The method is not merely consecutive, but iterative as well. For instance, evaluation takes place in various rounds, leaving opportunity to adjust the game scenario. The core of the method can be depicted as in Figure 1.

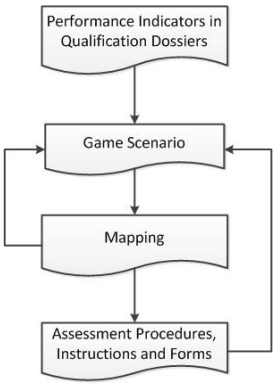

Figure 1. Stepwise validation method

2.2 Performance indicators and game scenario

For Step 1 we analyze the Qualification Dossier. As stated before, the attainment levels and performance indicators of vocational education for various professionals and levels are nationally documented and accredited in so called Qualification Dossiers. The structure of each Qualification Dossier is comprised of Core Tasks, that each contain Work Processes. Each Work Process is described with Performance Indicators and Wanted outcomes. The assessment game under study aims at the core task 1 'Develop (parts of) information -or media systems' which is comprised of five work processes. For brevity reasons, we only look at the first work process ('Analyze the needs of the contractor'). This process has two outcomes (i.e., a full and correct overview of (O1): the information needs of the contractor organization; and (O2): the conditions and possibilities within the organization) and six performance indicators P1 up till P6 (see Table 1). Step 1 ends by filling a validation table with four columns: performance indicator; place of occurrence within the scenario, information the game - if applicable - contains for the assessment, and information the document output or face-to-face talk - if applicable - contain for the assessment (see Table 1 which is already filled for the game example that is further described in Section 3). The third and fourth column of this table will reveal if and which performance indicators have to be assessed beyond the digital part of the game (i.e., computer program) and how. The second and third column will reveal which activities of the scenario will be used for assessment purposes. The third column describes the information the computer program contains for assessment purposes, like logging data on progress, sent mails and document outcomes.

For Step 2 we need to have a fully elaborated and adjusted game scenario. At this point it is good to further define scenario-based serious games as simulated task environments, which have been modeled after real-life situations that often include a sequence of learning activities that involve complex decision making, problem solving strategies, intelligent reasoning and other complex cognitive skills. Such games are often based on professional or academic role adoption and modeled after expert behavior. Students are left in charge to deal with complex problems according to professional or scientific standards. Real-life situations display ambiguity and conflicting information and offer a large degree of freedom. The EMERGO approach and toolkit is dedicated towards such scenario-based games, and has been used for the development of the scenario and game under study [16]. Before game development actually starts, for each activity is identified how students are expected and allowed to perform: what does the student do, with whom, with what tools and resources, and with which support (teacher, fellow student, or embedded in the game)? Does task performance result in a product, and if so, how will this be evaluated? Is a sufficient result needed before students can carry on? Which interactions with other participants and the digital part of the game are foreseen during and after carrying out activities? All (possible) interactions for each activity are exhaustively described, also in terms of required tools and resources.

2.3 Mapping and assessment procedures

For Step 3 a number of iterative evaluation rounds to establish the content validity are carried out in which the performance indicators will be mapped on the game scenario. The performance indicators for core task 1 (Develop (parts of) information- or media systems) were used as they could be derived and formulated by SPL based on the Qualification Dossier. Two assessment experts mapped indicators on activities and outputs as contained in the game scenario, using Table 1 independent from each other. In case not all indicators could be mapped, this was reported back to the project team which then decided either to incorporate the assessment of more indicators in the scenario or leave them out.

For Step 4, clear instructions are needed for the teachers / assessors that will be using the assessment game. In this case some performance indicators are left out of the digital part of the game and will be assessed during face-to-face talks. As results of Step 4, Assessment forms are developed for each core task (and the individual scoring on performance indicators for each work processes), as well as for the overall assessment that refers to a weighted sum of the performance scores on all five work processes and constitutes the final output of the validation method.

3. RESULTS

This section provides the results of applying the validation method on the Galema game. Again, we first describe the first two steps in section 3.1, and then the last two steps in section 3.2.

3.1 Game activities for assessment

Two assessment experts found that most performance indicators could be mapped on activities in the (adjusted version) of the game scenario. Some Work Processes could only be partly mapped on the scenario. And for some Performance Indicators it was decided they could better be assessed completely beyond the computer program (but still as integral part of the game scenario) by means of a face-to-face talk with the teacher (i.e., the game role that is indicated with the label 'Mr. Jonkman'). The Validation table for work process 1.1 is provided in Table 1.

Eventually, a detailed scenario of about 50 pages containing 55 learning activities could be agreed upon (Step 2), which could be used for the evaluation / mapping rounds in Step 3.

3.2 Assessment procedure and instructions

During Step 3, for each work process, a scoring model could be derived after it was decided what performance indicators were assessed (where and how). Such scoring models also clarify to what extend the assessor can use information obtained from outcomes (like written needs analysis) or contained in the computer program (like reports sent or logging of actions).

Attainment of each performance indicator is assessed by either I (insufficient), S (Sufficient) or G (Good). It was further decided and documented (in the assessment manual) that several criteria should be considered by the teacher when assessing work processes (for example: task is clearly described; the current way of working in projects is clearly described; problems of the current system are clearly mentioned; demands on the new system are clearly mentioned; wishes (may haves) and requirements (must haves) are clearly distinguished). Furthermore, the assessment manual contains example questions for the face-to-face talks and provides information for the game-role the teacher has to fulfill.

Table 1. Validation for work process 1.1 (Analyze the needs of the contractor)

Performance indicators	Content validation (place in scenario / activity student)	Assessment Information (system)	Assessment Information (in documents or by Jonkman)
(P1) Collect sufficient information by both interviewing and document analysis.	Virtual talks with employees Galema; F2F talk with Mr. Jonkman: Must prepare questions		F2F talk with Mr. Jonkman: Does student pose relevant and sufficient question?
(P2) Ask for the ideas and needs of employees to get a good overview of the information need within the organization	Virtual talks with employees Galema; F2F talk with Mr. Jonkman: Must prepare questions.		F2F talk with Mr. Jonkman: Does student pose questions about opinions, ideas and needs?
(P3) Consider the wishes of the client in relation with the possibilities when determining the information needs	Make a needs-analysis		Needs-analysis: Does student weigh the wishes and possibilities?
(P4) Show plan to relevant others and adjust them when appropriate	Send report talk with Mr. Boekhorst to him ; Send reports of all talks to coach; F2F-talk with Mr. Jonkman: discuss ideas and adjust analysis; Send needs-analysis to Jonkman, coach and Galema	Report talk with Boekhorst been send to him?; All reports sent to coach?; Has needs-analysis been send to Jonkman, coach and Galema?	F2F-talk with Mr. Jonkman: Does student respond adequately to comments?
(P5) Acquire a full and correct overview of business processes and information streams	Make needs-analysis		Needs-analysis: Does it show practice correctly and completely?
(P6) Verify correctness of acquired information, structure information, and consider conclusions by using available facts and weighing pros and cons.	Make needs-analysis; Report talk with Boekhorst: Verify with him if it is a correct reflection of actual practice	Report sent to Boekhorst requesting him to check for correctness?	Needs-analysis: is document correct and complete with clear structure?

4. CONCLUSION AND DISCUSSION

This study shows it is indeed possible to develop and apply a validation method to validate game scenarios for assessment purposes. Preliminary experiences reveal that an assessment game that results from this validation is indeed more transparent, better documented, and can be more effectively compared and organized. Both students and teachers find this more dynamic way of assessment more motivating and effective. Two teachers that used this assessment game over the last months (with 20 students) report that both the preparation and execution of the examination project is now less labor-intensive.

However, some of the performance indicators were not suitable for e-assessment (i.e., the digital part of the game). Therefore, the face-to-face component is still required. A blended approach (both virtual and face-to-face) with students and teacher "stepping in and out" of the digital part of the game did not appear to be problematic for students and teachers. Current gaming platforms do not yet cater for valid and reliable in-game assessment of all types of activities. For example, the assessment of the more 'soft' communication competence is beyond scope although there are some promising developments with respect to speech recognition and emotion recognition that alleviates the work of the teacher and can prevent students from struggling too long on ineffective learning paths [17].Validating the content of game scenarios seems to be an important line of future research, and can ensure that serious games are better warranted against the current criticism of not being transparent enough for assessment purposes. The assessment in this case study seems to result in comparable and more efficient assessments. Such advances in adaptive serious games with "embedded assessment" make better visible how learners develop skills and monitor their success, and thus provide teachers with new insights that help them improve their teaching and tutoring. It has remained beyond the scope of this study (which is mainly descriptive) to investigate the impact of different design mechanism upon students' and teachers' opinions with respect to assessment and students' skill development and success. We are currently preparing a study with a larger group of participants in which we will examine the impact of different game guidance

mechanics towards students' success. Another limitation of this study is that we do not have enough proof that such assessment games are sufficiently warranted towards fraud on the long run, when larger numbers of students study the same cases. Although we cannot fully exclude such risks, it needs more attention in the design and exploitation. Furthermore, the positive effects of studying just one assessment game (of two days) will be snowed under when the remainder of the curriculum is still classically tested. For this reason SPL is now developing assessment games for all core tasks within the piloted curriculum. Finally, we also have to see if results found within the domain of system management are generalizable towards other domains.

5. ACKNOWLEDGEMENTS

Development of this game was funded by SPL, and we thank Ton Remeeus (director) and Martin van Kollenberg (ICT coordinator) for their support. The project was carried out under responsibility of the Learning Media program lead by Wim Westera. Game scenario, game and assessment procedure were developed with and between Marc Hector (content), Henk van den Brink (educational design), Hub Kurvers (programming), and the last two authors (validation). Finally we thank the first two teachers and small group of students at ROC Zadkine that used the game.

6. REFERENCES

[1] Gee, J.P., & Shaffer, D.W. (2010). Looking Where the Light is Bad: Video Games and the Future of Assessment. *Phi Delta Kappa International EDge*, 6(1).

[2] Ifenthaler, D., Eseryel, D., & Ge, X. (2012). *Assessment in Game-Based Learning: Foundations, innovations and perspectives*. New York: Springer.

[3] Black, P., & Wiliam, D. (1998). Assessment and Classroom learning. *Assessment in Education, 5*, 7-74.

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

[5] Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32*, 347-364.

[6] Biggs, J. (2003). *Teaching for quality learning at university*. Maidenhead: SRHE.

[7] Redecker, C., Punie, Y. & Ferrari, A. (2012). *eAssessment for 21st century learning and skills*. In: Ravenscroft, A., Lindsteadt, S. Kloos, C.D. & Hernandez-Leo, D. (Eds.) 21st Century Learning for 21st Century Skills. Proceedings EC-TEL, Saarbrücken, 2012, 292-305. Heidelberg: Springer.

[8] Sluijsmans, D. M. A., Joosten-ten Brinke, D., & Van der Vleuten, C. P. M. (2013). *Toetsen met leerwaarde. Een reviewstudie naar effectieve kenmerken van formatief toetsen*. [Formative assessment. A review study on effectiveness of formative assessment]. Den Haag: NWO.

[9] Young, M.F., Slota, S., Cutter, A.B., Jalette, G., Mullin, G., Lai, B., Simeoni, Z., Tran, M., & Yukhymenko, M. (2012). Our Princess Is in Another Castle: A Review of Trends in Serious Gaming. *Review of Educational Research, 82*(1), 61-69.

[10] 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. Ritterfeld, M. Cody, & M. Vorderer (Eds.), *Serious Games: Mechanisms and Effects* (pp. 295-321). New York: Routledge.

[11] Corti (2011). Proof of Learning: Assessment in Serious Games. Last accessed: 4 May 2013, http://www.gamasutra.com/view/feature/2433/proof_of_learning_assessment_in_.php

[12] Kane, M. (2006). Validation. In R. Brennan (Ed.), *Educational measurement*, 4th ed. (pp. 17–64), Westport, CT: American Council on Education and Praeger.

[13] Van der Vleuten, C. P. M., Schuwirth, L. W. T., Driessen, E. W., Dijkstra, J., Tigelaar, D., Baartmand, L. K. J., & Van Tartwijk, J. (2012). A model for programmatic assessment fit for purpose. *Medical Teacher, 34*, 205-214.

[14] Baartman, L. K. J. (2008). *Assessing the assessment: Development and use of quality criteria for Competence Assessment programs*. Thesis. University Utrecht: Utrecht, The Netherlands.

[15] Inspectie van het Onderwijs (2009). Boekhouder of wakend oog. Verslag van een onderzoek bij examencommissies in het hoger onderwijs over de garantie van het niveau. Inspectierapport 2009-16 (april). Available at: www.onderwijsinspectie.nl/actueel/publicaties/Boekhouder+of+wakend+oog.html

[16] Nadolski, R. J., Hummel, H. G. K., Van den Brink, H. J., Hoefakker, R., Slootmaker, A., Kurvers, H., and Storm, J. 2008. EMERGO: Methodology and Toolkit for Efficient Development of Serious Games in Higher Education. *Simulations & Gaming*. 39, 3 (Sep. 2008), 338-352. DOI = http://sag.sagepub.com/content/39/3/338.full.pdf+html

[17] Bahreini, K., Nadolski, R., and Westera, W. 2014. Towards Multimodal Emotion Recognition in E-learning Environments. *Interactive Learning Environments*. May 2014, 1-16. DOI= http://dx.doi.org/10.1080/10494820.2014.908927.