Content Validity of Game-based Assessment: Case study of a Serious Game for ICT managers in training

Citation for published version (APA):

Hummel, H., Joosten-ten Brinke, D., Nadolski, R., & Baartman, L. (2016). Content Validity of Game-based Assessment: Case study of a Serious Game for ICT managers in training. *Technology, Pedagogy and Education*, *26*(2), 225-240. https://doi.org/10.1080/1475939X.2016.1192060

DOI:
10.1080/1475939X.2016.1192060

Document status and date:
Published: 06/06/2016

Document Version:
Peer reviewed version

Document license:
CC BY-NC-ND

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:

https://www.ou.nl/taverne-agreement

Take down policy
If you believe that this document breaches copyright please contact us at:

pure-support@ou.nl

providing details and we will investigate your claim.

Downloaded from https://research.ou.nl/ on date: 04 Dec. 2022

Open Universiteit
www.ou.nl

Content Validity of Game-based Assessment:

Case study of a Serious Game for ICT managers in training

Hans G. K. Hummel[1], Desirée Joosten-ten Brinke[1],

Rob J. Nadolski[1] and Liesbeth K. J. Baartman[2]

[1]Open Universiteit of the Netherlands

[2]University of Applied Sciences of Utrecht, the Netherlands

Submitted: January 26, 2015

Revision: September 10, 2015

Author note

Correspondence concerning this article should be addressed to Hans G. K. Hummel, Welten Institute: research Centre for Learning, Teaching and Technology, Open Universiteit Nederland (OUNL), Valkenburgerweg 177, 6419 AT Heerlen, The Netherlands, E-mail: hans.hummel@ou.nl

Content Validity of Game-based Assessment:

Case study of a Serious Game for ICT managers in training

Submitted: January 26, 2015

Revision: September 10, 2015

Abstract

Serious games foster the acquisition of complex problem solving skills. Assessment of such skills should be in line with instruction, and within a serious game environment its content validity should equal face-to-face assessment. Research on assessment in serious gaming has remained rather scarce. This article shows how assessment can be implemented in serious gaming in a way that assures content validity. The core of our validation method entails mapping learning activities (as contained in the game scenario) on performance indicators and outputs (as derived from formal attainment levels). We present how we have elaborated and applied the method on an assessment game for ICT managers in secondary vocational education. We describe the procedure and extent to which this assessement is content valid compared to face-to-face assessment.

Keywords

serious games, seamless assessment, content validation method,

professional competence, game scenarios

1. Introduction

Playing games in education is generally known for its contribution to improving motoric skills or gaining knowledge about certain topics. Less known is that serious games may also foster the acquisition of more complex cognitive skills, like problem solving and intercultural communication (e.g., Guillén-Nieto & Aleson-Carbonell, 2012; Yang, 2012). Interest in and use of educational games has grown over the last decades. Findings of a review study on empirical evidence for the potential positive impacts of gaming revealed that playing computer games is linked to a range of perceptual, cognitive, behavioural, affective and motivational impacts and outcomes (Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012). Most frequent outcomes and impacts were knowledge acquisition, content understanding, as well as affective and motivational outcomes.

The review showed that the games are used for learning and not for assessment purposes. Based on summative assessment, important decisions about students are made in terms of fail or pass. The assessment method, whether paper-and-pencil, face-to-face or by serious gaming, should therefore not bias that decision. As a consequence, serious games have hardly been adopted for summative assessment purposes. For the true adoption of serious games for both learning and assessment, we first of all need to ensure that the quality of an assessment with serious gaming equals (or outperforms) the quality of other proven assessment methods. Assessment during game play should be seamless in order not to interrupt the game experience while therefore keeping the learner in a state of flow (Csikszentmihalyi, 1990; Shute, Ventura, Bauer & Zapata-Riviera, 2009). In other words, assessment should be (as) unobtrusive (as possible) to the player, while at the same time not sacrificing reliability and validity.

The study presented in this article will describe how summative assessment was implemented in a serious game, and will evaluate the content validity of this assessment in relation to a previously used face-to-face assessment. The remainder of this introduction will explain the role of serious games in education, explain specific criteria when using serious games for summative assessment, discuss the content validity of serious games, describe the specific educational context for this study, and present our research question. The specific game for this study (section 2.1) and the validation method (section 2.2) will be elaborated in the second section. After presenting the results on applying the validation method in section 3, final section 4 concludes with an evaluation of the method and suggestions for future research.

Serious gaming in the educational context

To adequately function in their jobs, future employees need to be able to deal with complex problems and dilemmas. Games that aim to foster the acquisition of such skills are called 'serious' to denote that they are *not just* fun to play, but *also* hold potential as cognitive tools for learning (e.g., Michael & Chen, 2006). The use of games can be a valuable way to engage students in learning, as it fits well to their daily computer use. Student engagement is strongly associated with academic achievement. Thus, the combination of school and gaming has potential to increase learning, especially for lower performing, disengaged students (Shute, Ventura, Bauer & Zapata-Rivera, 2009). In a competence based learning environment, serious gaming could be of interest in combination with a balanced set of different assessment types in an assessment program (Baartman & Gulikers, 2014). Single assessment methods are not valid for assessing the complex nature of competences, including professional knowledge, skills and attitudes and their application in complex occupational settings (Van der Vleuten *et al*., 2012).

Serious games and summative assessment

Gee and Shaffer (2010) expect games to reform current educational assessment, and lead to radical transformation towards learning for 21st century skills. With many other other educationalists, they state that "Assessment is the tail that wags the dog of learning". Assessment is the process of using data to demonstrate that stated learning objectives are actually being met by a learner (Black & Wiliam, 1998; Chin, Dukes & Gamson, 2009). For using serious games as summative assessment methods, it is important that the activities of the serious game (i.e., the scenario) align with the learning objectives that are to be assessed. Alignment, as used by Biggs (1996; 2003), refers to the need for clear and logical connections between the intended outcomes of the learning experience and the practices and processes utilized in the learning environment. The intended learning outcomes have to be described in terms of performances that represent a suitable cognitive level. Performances mentioned in the objectives are thus used to systematically align the teaching methods with the assessment. Several authors have argued for the strength of games as assessment engines (Gee & Shaffer, 2010; Ifenthaler, Eseryel, & Ge, 2012). Gee and Shaffer argue that games are good learning engines because they are first of all good assessment engines. The use of a serious game as a learning environment that fits with the learning objectives asks for an assessment in a comparable environment.

A main challenge involved with creating games in which the learning has to be assessed, is to consider the dynamic nature of the game assessment which has to be as unobtrusive as possible to the player. Such 'seamless' assessment at the same time should not sacrifice reliability and validity.

In the design of the assessment it should be clear whether the objective of the assessment is formative (i.e., to deliver information that serves to improve further learning), or summative (i.e., to deliver information that serves to make a 'high stakes' decision). In the context of formative assessment, implementations show the benefits of serious gaming. Students spend more time on tasks which may lead to higher learning outcomes, they evaluate the use of the games as stimulating and the context of the tasks often as more authentic (Connoly *et al*., 2012). However, research on the use of serious gaming for *summative* assessments is scarce.

Furthermore, the use of computers for summative assessment purposes, called e-assessment, has been mostly limited to assessment of lower order cognitive skills. Redecker, Punie and Ferrari (2012) describe the history of e-assessment. The first generation e-assessments in the 1990s were focused on automated administration and scoring. Around 2000, the second generation e-assessments introduced computer adaptive testing (e.g., Eggen, 2011). From 2010 onwards, focus shifted to continuous, unobstrusive and more formative e-assessments, which are supposed to include behavioral tracking in immersive and game-based environments (Bellotti, Kapralos, Lee, Moreno-Ger & Berta, 2013). To assess higher order cognitive skills, creating scenarios with more complex learning activities is expected to provide possibilities. However, a major impediment for exploiting games for summative assessment purposes is the current lack of proof on the content validity of such serious games (e.g., Wouters, Van Nimwegen, Van Oostendorp & Van der Spek, 2013).

Content validity of serious games

Validation methods evaluate whether an assessment achieves its purposes. They determine the so called 'fitness for purpose' (Van der Vleuten *et al*., 2012), which encompasses the way results of

an assessment are interpreted and used by educators and students. A validation method provides information on whether the (game) assessment aligns with the learning objectives and the activities of the learning scenario (within the game). This implicates that assessments need to represent the learning objectives. Higher order thinking skills are often assessed using performance assessments. These are observed achievements with a professional judgment (Stiggins, 1987). Performance assessments can be performed in real working environments (authentic assessment) or in virtual, more or less controlled but work-based, learning environments, for example simulations. It is not always evident to what extent performance demonstrated in virtual learning environments will also be showed in real working environments (Straetmans, 2014). However, game-based assessment is assumed to be more efficient than assessment in a real-life working environment because the work-based tasks can be better selected and aligned to learning objectives than in real working environments where practical constraints often hamper such alignment. Furthermore, games offer safe and controlled environments where students can have prolongued varied practice with scaffolded guidance towards goal achievement (e.g., Aldrich, 2009). To evaluate the validity of a performance assessment, different kinds of evidence have to be taken into account. It is the process of building arguments to support the claims and decisions that are made from assessment scores (Kane, 2006). It involves the translation of learning objectives into assessment tasks in the learning scenario, the assessment criteria, the instructions for students, teachers and workplace supervisors, the expert judgments, and the documentation that provides information about the content validity. When game-based assessment is part of a larger assessment procedure, a distinction can be made between the content validity of the game itself (i.e., the activities put in the scenario that represent the learning objectives) and the content validity of the entire

assessment procedure that may encompass other non-game-based activities as well (e.g., face-to-face role play).

Educational context of the study

This study on game-based assessment took place in secondary vocational education in the Netherlands. This kind of education is largely offered by so called Regional Education Centers, large training institutes that on the average serve about 30,000 students each. Learning objectives for each profession are documented in national qualification profiles, for example a qualification as ICT manager. Each qualification profile consists of a set of core tasks with work processes and performance indicators. The qualification profiles have been accredited on a national level. Students enter secondary educational vocational education at the age of 16 and the program takes four years of study. At the Regional Education Center where this study took place, yearly around 200 students enroll to follow ICT management education (courses with traditional assessment), of which 20 participated in this study and studied the course with game-based assessment.

The qualification 'ICT manager' consists of four core tasks: (1) Develop (parts of) information- or media systems; (2) Implement (parts of) information- or media systems; (3) Manage (parts of) information- or media systems; and (4) Organize an (existing) helpdesk. The assessment game under study is aimed at the first core task, which in its turn comprises of five work processes: 1.1 Set clear the information needs of the client; 1.2 Create a functional design for these systems; 1.3 Create a technical design for these systems; 1.4 Create a test plan; and 1.5 Realize a test environment. In the first column of Table 1 performance indicators for work process 1.1 are given.

The traditional assessment consists of an assignment on paper to develop the (parts of an) information system for an imaginary company, and takes place in a school setting with other students doing the same or another assessment, with assessors available for task specific questions and for judging the performance. The traditional summative assessment has been accepted and accredited by the Inspectorate of Education. During this traditional assessment, several assessors play one or more roles to assess the performance indicators, for example the role as client or the role as financial manager of the fictive company. All student performances are observed and scores on all of performance indicators are noted down on several assessment forms. This traditional assessment lasted two days, with the assessor filling in the final form and score for each student at the end of the last day. This assessment procedure was time consuming and administratively overloaded the assessors. The purpose of the serious gaming environment was to reduce the administratieve load for the assessors by letting the computer take over time consuming task. This e-assessment should be as valid and reliable as the traditional format.

This study compares the traditional summative assessment with the game-based version, based on the same learning objectives. The research question for this study is whether the content validity of the game-based assessment equals the content validity of the traditional summative assessment.

2. Method

In order to answer the research question, the serious game first had to be developed. In section 2.1 we describe the serious game under study. In section 2.2 we describe both the design of the game-based summative assessment and the method of evaluating its content validity.

2.1. The serious game under study

The game-based assessment was realized by using the EMERGO game platform (Nadolski *et al*., 2008). EMERGO entails both a methodology and generic toolkit for developing and delivering scenario-based serious games that are aimed at the acquisition of complex cognitive skills. A game scenario is the basis for the game development. The work processes with their performance indicators were 'translated' into game scenarios. These scenarios and their learning activities warrant the application of knowledge and skills in professional contexts. Core to our method for content validity is that performance indicators are mapped on the learning activities and outputs within the game scenario.

FIGURE 1 ABOUT HERE

The game scenario under study is based on the 'Events Agency Galema' case which has to be studied within a virtual company that builds IT solutions. This means that students are largely assessed while carrying out tasks on their computer. The game they play is based on a scenario with consecutive learning activities that partly have to be carried out within the virtual company, and partly by having talks with the assessor in real life. A virtual coach (see Figure 1) welcomes students as new employee for 'ITadvice4U', a fictive IT company that exists 20 years. The coach (Mr. Alaoui) is one of its employees, and (in the role of superior) introduces the company and assigns the student with the following task: develop a new system for project management for the Galema agency that organizes events. In this (also virtual) events agency the student will encounter other employees, like the director, the secretary, the project leader, a financial expert (Mr. Jonkman), and a software developer. The student first has to carry out a needs-analysis by

interviewing a number of employees of the agency, and by studying a number of documents, in order to analyze current IT problems and possible solutions. Most interviews are conducted within the game. For these interviews, students choose appropriate experts and questions from a list, which are answered in video messages. One interview is conducted face-to-face with the teacher playing Mr. Jonkman (who also has this role of financial expert within the game). This face-to-face interview involves the specific communication skills that are difficult to assess within the serious game, like reacting appropriate to strange questions. On the basis of the interview and information found, the student writes a needs-analysis as output of this activity (3 hours). In the next activities, the student uses the needs-analysis to distill a functional and technical design of the new system, which are discussed with the teacher / Mr. Jonkman in a face-to-face setting (5 hours). Finally the student draws up a plan for developing the new system for project management (2 hours), develops and tests a first version, and writes a test report about his (all students so far were male) findings (6 hours), yielding a total study load of about two days to complete the assessment game. Instead of four role-players in the face-to-face version, just one assessor/role-player was necessary in the game-based assessment.

2.2. The game design and method for evaluating content validity

The game design and method for evaluating content validity comprise the following four steps: (1) Analyze the qualification profile for ICT manager to decide how the performance indicators should be demonstrated by the students; (2) Develop the game scenario in which the assessment tasks are executed; (3) Design assessment procedures, instructions and forms; and (4) Evaluate to which extent performance indicators are covered by the assessment tasks. The method is not

merely linear 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 2.

FIGURE 2 ABOUT HERE

Step 1: Translating performance indicators into activities for assessment

For step 1 the performance indicators were analyzed. To reduce the work load of the assessor, all performance indicators were evaluated to decide whether they could be assessed within the game by the computer (using logging data), by using a delivered product, or by a person. Table 1 gives an overview of the performance indicators for work process 1.1 (e.g., P1 collect sufficient information); place of occurrence within the scenario (e.g., for P1 to have interviews with virtual employees, and a face-to-face talk with teacher); information the game contains for the assessment (e.g., for P1 there is none); and information the document output or face-to-face talk contain for the assessment (e.g., for P1 during the face-to-face talk teacher may decide if relevant questions have been raised). The third and fourth column of this table reveal which performance indicators have to be assessed beyond the computer program and how (e.g., for P1 this will be done by having a face-to-face talk with teacher). The second and third column 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 instance, work process 1.4 can use the following information as contained in the computer program: Has a report of the talk with the secretary be sent?; Has a report of all talks with the supervisor been sent?; Have comments made

by Jonkman been adapted?; and Was the needs analysis shown to Jonkman, the supervisor and director?

Step 2: Development of the game scenario

Step 2 consisted of the development of 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 modelled 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 modelled 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 (e.g., Nadolski *et al*., 2008) is dedicated for such scenario-based games, and has been used for the development of the scenario and game under study. The design phase of this approach results in a *detailed scenario* document via the intermediate *framework scenario* and *ingredients scenario*, with each step providing more detail and completeness.

A framework scenario describes the global activities students carry out during the game. This enables the game design team to identify a first series of activities without getting overwhelmed by details. This scenario is a blue print with issues to be further worked upon, resembling a construction drawing for a building. Within the ingredients scenario, for each activity we identify how students are to perform: what does the student do, with whom, with what tools and resources, and with which support (teacher, fellow student, or embedded)? All (possible) interactions for each activity are exhaustively described, but not yet in terms of

required tools and resources. In the ingredients scenario all the assessment tasks have to be clear. Finally, the detailed scenario describes each activity exhaustively in terms of required tools and resources for their actual performance. For instance, if students can interview a person, all interview questions need to be identified. At this point, we identify which materials and tools are already available or still need to be developed.

Step 3: Development of assessment procedures, instructions and forms

In this third step, clear instructions are needed for teachers, assessors and students that will be using the assessment game. As determined in step 1, some performance indicators are left out of the game and assessed during face-to-face talks. Because the teacher plays both the role of assessor as of Mr. Jonkman (the financial expert in the game environment), these situations have to be congruent. The teacher executes a number of face-to-face talks with students in this role, and he then has to behave according to this role. An assessment manual with some tips and instructions is available that, for instance, offers some basic questions / answers for Jonkman that have been worked out to describe this role. Besides, assessment forms were developed for each core task (and the individual scoring of work processes, see Table 2), as well as for the overall assessment as final output of the validation method.

Step 4: Evaluation of the mapping

For step 4 a number of iterative evaluation rounds to establish the content validity were carried out in which the performance indicators are mapped on the game scenario. Two assessment experts mapped indicators on activities and outputs as contained in the game scenario, using Table 1 independently from each other. In case not all indicators could be mapped, this was fed

back to the project team which then decided either to incorporate the assessment of more indicators in the scenario or leave them out. In this case, the first evaluation round showed that not all indicators could by assessed within the game, and it was decided that some of these performances (partially) required an assessment by a teacher during a face-to-face talk. In order to have these talks, the student has to "step out" of the scenario and afterwards "step in" again. For determining the content validity, following aspects were assessed by two assessment experts (being co-authors of this article): the extent to which performance indicators could be mapped on the game scenario; the extent to which these performances could be assessed by products delivered through game play (and by procedures beyond the game); the perceived efficiency as compared to traditional assessment; and the perceived validity as compared to traditional assessment.

To evaluate whether the users experienced the e-assessment as valid as the original assessment, eight assessors were interviewed. We invited 15 assessors (the total pool of assessors for the traditional assessment) and 8 of them volunteered to participate. They all had a long time experience (three to ten years) with the traditional assessment and were all assigned as assessors for the assessment with the serious game. Two of them did already participate in the new assessment game, six of them evaluated the new assessment based on a full game scenario and a demonstration. In the group interview, the assessors were asked: (1) whether each of the performance indicators was assessed by the assessors; (2) what the student had to deliver after each work process; and (3) which criteria were used for a pass score. The outcomes of the interviews were compared with the evaluation of the assessment in the serious game.

3. Results

This section provides the results of applying step 4 of the method presented in the previous section and the interviews with assessors, to establish whether the content validity for the game-based assessment (and the entire assessment procedure, including face-to-face interviews with the teacher) is equal to the traditional assessment. For the 'Event Agency Galema' case, the content validity of the game-based assessment has been determined in a number of evaluation rounds.

Evaluation of the mapping of performance indicators

Most performance indicators (twenty out of thirty-two) could be fully mapped on activities in the game (i.e., the content validity of the game itself). Five indicators could only be partly mapped on the scenario, like 'Uses insight during the technical design in an appropriate way' (work process 1.3, indicator 3) for which it was not clear how this could be made visible in the game. For seven performance indicators it was decided they could better be assessed 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 content validity of the entire assessment procedure). These were performance indicators like 'Asks for an opinion, ideas, and needs of others to get a complete picture of the information needs in the organization' (work process 1.1, indicator 2) and 'Explains the functional design in an understandable way which depends on the audience' (work process 1.2, indicator 5). Typical for such performance indicators is that the student is able to empathize with the audience. To assess such empathy, a human assessor appears to be necessary. Although only the first work process is elaborated and used as worked example for this article, we can mention that all other indicators that were hard or impossible to integrate in the game scenario also dealt with communication (like asking others for opinions and showing, defending

and discussing outcomes). Work process 1.5 (Realize a test environment) appeared most problematic to map. It became clear that the performance indicators for this work process required a high degree of tuning with all stakeholders, which appeared hard to include in the game scenario. In total, all indicators of the four work processes could be mapped, either on learning activities within the game (content validity of the game itself) or on face-to-face talks beyond the e-assessment (included in the entire assessment procedure). As worked example, the evaluation of the content validity for work process 1.1 is provided in Table 1.

INSERT TABLE 1 ABOUT HERE

Game-based assessment products and procedures beyond the game

Table 2 presents the scoring model (for work process 1.1) that was derived after mapping the (six) performance indicators. Table 2 also makes clear to what extent the assessor can use information obtained in outcomes from game play (like a written needs analysis) or as contained in logging data (like tracking of reports sent).

INSERT TABLE 2 ABOUT HERE

Attainment of performance indicators is scored by either U (unsufficient), S (Sufficient) or G (Good). It was further decided and documented (in the assessment manual) that following (ten) performance indicators for the outcome of work process 1.1 (information needs analysis) could be assessed by using the needs analysis report delivered by game play: (a) task is clearly

described; (b) reason for task assignment is clearly described; (c) the current way of working in projects is clearly described; (d) problems of the current system are clearly mentioned; (e) demands on the new system are clearly mentioned; (f) wishes (may haves) and requirements (must haves) are clearly distinguished; (h) a conclusion about the needs of client is drawn based on requirements and wishes; (i) in the conclusion pros and cons have been carefully balanced; and (j) the description of the needs analysis is clearly structured.

As explained before, sometimes the student has to step out of the computer program to explain or defend outcomes to Mr. Jonkman during a face-to-face talk. Because Mr. Jonkman is a financial expert who does not have much technological background, the student has to be able to clearly present his analyses or designs without using technological jargon (work process 1.2, indicator 5). During this interview the teacher will - in Jonkmans' role - assess the student's ability to raise the right questions and use clear language. Besides instructions and forms, the assessment manual also contains about twenty basic questions and answers for face-to-face talks with Mr. Jonkman / the teacher, that provides background information about this case-role, like:

Can you tell how projects currently are managed?

> Yes I can. We have been working with Excel sheets for ages. They contain the various tasks and planning. Team members use them to keep track of their projects. For the use of me and my financial colleague I have built a large Access Database, in which we neatly keep track of everything. But in order to do so, others do have to supply us with the information needed in time. And that unfortunately often is not the case.

Can you tell what you do when a new project starts?

> For us a new project starts when the quotation has been signed. This quotation contains the total amount of money the client will pay. Then the project leader will provide me with a list of estimated costs. The difference between this estimate and the amount in the bid is, roughly speaking, our profit. The project leader periodically has to supply me information about which tasks have been completed, so that I can keep track of what has been spent. I mostly receive that information by mail. However, most project leaders are not that precise and tend to forget this.

Efficiency as compared to traditional assessment

The two teachers who used this assessment game over the last year (with 20 students) report that both the preparation and execution of the assessment have become less labor-intensive. For instance, some laborious pacing mechanisms to receive output over time have been successfully automated within the scripting of the game play.

Evaluation of the validity and role by content experts

We conducted interviews with the content experts and assessors to evaluate their role in realizing content validity of the game-based assessment as compared to the traditional assessment. These assessors found the products as delivered by game play to be appropriate evidence for required students' skills on work processes. Overall this game-based assessment was considered more transparent when compared to the traditional assessment. However, the interviews showed different views on the product to be delivered in the fifth work process, which already appeared hardest to include in the game scenario. Some assessors stated that it was more important to produce a prototype instead of a test environment, and that teachers should steer the format of such a prototype. Some of them argued that the performance indicators for the technical design could not be standardized. In the traditional assessment, some assessors preferred to use a more holistic judgment based on the delivered product and only used the performance indicators in cases of doubt. Others preferred to use decision rules like "five out of seven performance indicators have to be scored sufficient to pass". Although the assessors felt competent to take the final decision, all assessors indicated that in the traditional assessment it was sometimes a difficult process to decide whether a student should pass or fail. As all the products in the

traditional assessment were assessed by just one assessor, this difference in vision on how to use the performance indicators is a real problem.

4. Conclusion and discussion

This study shows that it is indeed possible to develop game-based assessment with a content validity that at least equals the traditional assessment. The study reveals that the game under study could cover most of the performance indicators as defined for the core task. The entire assessment procedure, including face-to-face talks with the teacher, could include all performance indicators. Our evaluation with assessors and teachers shows that game-based assessment that uses the content validation method seems to be more transparent, better documented, and more efficiently organized when compared to traditional assessment. The use of game-based assessment also seems to decrease the risks of subjective and arbitrary scoring by assessors. For instance, when discussing the mapping of the performance indicators on the learning scenario, teachers had to be more concrete about their use of performance indicators. For the teachers, it became more transparent that game-based assessment uses the same performance indicators for all students. Despite the fact that teachers are trained in the use of performance indicators, the interviews showed that they used the indicators in different ways in the traditional assessment.

This case study also revealed that some of the indicators were not suitable for inclusion in e-assessment. For example, if a student has to show that he is capable in 'collecting sufficient information by an interview' (performance indicator P1), he has to do an interview. Therefore, the face-to-face component was still required. Fortunately, such a blended approach in the game scenario (virtual and face-to-face) with students and teachers "stepping in and out" the computer

program did not appear to be problematic, neither for students nor teachers. Further research is needed, however, to provide more firm conclusions on students' and teachers' experiences with e-assessments. Specific gaming platforms simply do no cater for all possible types of learning activities. In this case study for example, the gaming platform appeared less suitable for the assessment of more 'soft' communication skills.

It has remained beyond the scope of this study (which is mainly descriptive) to study *why* students and teachers like this way of assessment and *how* they exactly develop skills and monitor their success. Another limitation of this study is that we do not have enough proof that such game-based assessments are sufficiently warranted towards fraud on the long run, when larger numbers of students use the same cases. Although each student has to deliver his own assessment products, this needs more attention in the design. Finally, we have still to see if results found within the domain of ICT management are generalizable towards other domains and educational levels. After this study, this and threeother Regional Education Centers have developed and evaluated similar games and e-assessment for other core tasks and educational levels in the domain of ICT management, which can be interpreted as preliminary proof for succesful uptake and generalizability of the validation method presented here. The development of a detailed game scenario can be a cumbersome process. In this case for quite some time no sufficient content expertise could be made available during the design phase, which has seriously hampered progress. This experience again has showed the importance of multi-disciplinary collaboration (both content, didactical, assessment and technological experts working together) when developing game-based assessments using the EMERGO platform. For this case a detailed scenario of about 50 pages containing 55 learning activities could be agreed upon (Step 2 of the method), which was used for the evaluation and mapping rounds in Step 4.

Validating the content of game scenarios for assessment purposes 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 summative assessment. According to Corti (In Michael & Chen, 2011, p. 34): "Serious games will only grow as an industry if the learning experience is definable, quantifiable and measurable".

Acknowledgements

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

References

Aldrich. C. (2009). *The complete guide to serious games and simulations*. Somerset, NJ: Wiley.

Annetta, L. A., Minogue, J., Holmes, S. Y., & Cheng, M. (2009). Investigating the impact of video games on high school students' engagement and learning about genetics. *Computers & Education*, *53*, 74–85. doi:10.1016/j.compedu.2008.12.020

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. http://dx.doi.org/10.1155/2013/136864

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

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

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

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper Perennial.

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

Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education*, *59*(2), 661-686.

Eggen, T. J. H. M. (2011). Computerized classification testing with the Rasch model . *Educational Research and Evaluation*,*17*, 361-371 doi:10.1080/13803611.2011.630528

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).

Guillén-Nieto, V., & Aleson-Carbonell, M. (2012). Serious games and learning effectiveness: The case of "It's a Deal!". *Computers & Education, 58*, 435-448.

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

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

Michael, D., & Chen, S. (2006). *Serious games: games that educate, train and inform*. Thomson Course Technology, Boston, MA.

Michael, D. & Chen, S. (2011). *Proof of Learning: Assessment in Serious Games.* [Web log post] Retrieved December, 9 2014, from http://www.gamasutra.com/view/feature/2433/proof_of_learning_assessment_in_.php

Nadolski, R. J., Hummel, H. G. K., Van den Brink, H. J., Hoefakker, R., Slootmaker, A., Kurvers, H., & Storm, J. (2008). EMERGO: methodology and toolkit for efficient development of serious games in higher education. *Simulations & Gaming,39*, 338-352.

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

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. *Serious games: Mechanisms and effects*, 295-321.

Stiggins, R. J. (1987). Design and Development of Performance Assessments. *Educational Measurement: Issues and Practice*, *6*(3), 33-42. doi: 10.1111/j.1745-3992.1987.tb00507.x

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

Wouters. P., Van Nimwegen, C., Van Oostendorp, H., & Van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology, 105*, 249-265.

Yang, Y. T. C. (2012). Building virtual cities, inspiring intelligent citizens: Digital games for developing students' problem solving and learning motivation. *Computers & Education, 59*, 365-377.

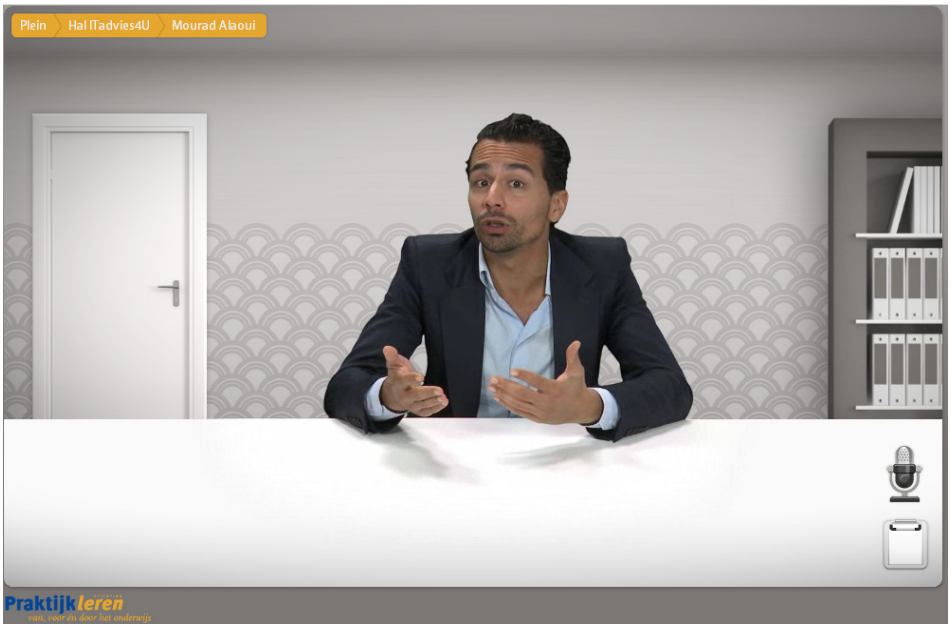

Figure 1. Screen of the game with Mr. Alaoui (supervisor) explaining the task to the student

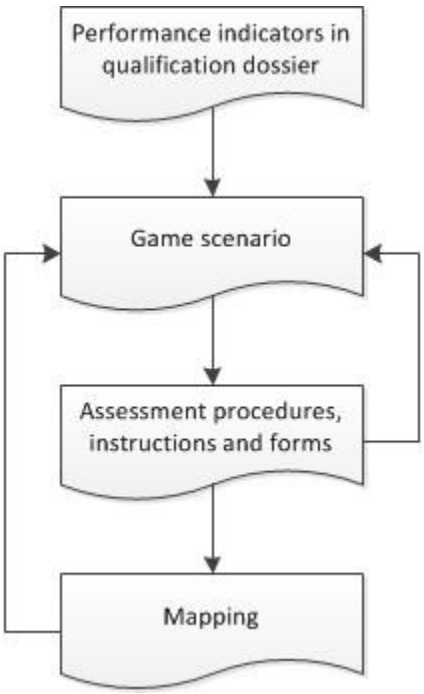

Figure 2. Stepwise design method

Table 1. *Validation Table (for work process 1.1)*

Performance indicators	**Content validation (place in scenario / activity student)**	**Assessment Information contained in system**	**Assessment Information contained in documents or by Jonkman**
(P1) Collect sufficient information by both interviewing and document analysis.	Virtual talks with employees Galema: Mrs. Galema, Mr. Boekhorst, Mrs. Vos 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 it 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?

Table 2. *Sub-scales and scoring form (for work process 1.1)*

Performance indicators	**A. Assessment by Jonkman**	**B. Assessment based on outcomes**	**C. Assessment based on logged data**	**Assessment Score**
(P1) Collect sufficient information by both interviewing and document analysis.	Has student prepared and raised relevant and sufficient questions?	-	-	U / S / G
(P2) Ask for the ideas and needs of employees to get a good overview of the information need within the organization	Has student prepared and raised questions about opinions, ideas and needs?	-	-	U / S / G
(P3) Consider the wishes of the client in relation with the possibilities when determining the information needs	-	Needs-analysis: rational weighing of possibilities?	-	U / S / G
(P4) Show plan to relevant others and adjust it when appropriate	Does student communicate needs-analysis? Respond adequately to comments Jonkman? Adjust plans when requested?		- Report Boekhorst send to him? - All reports send to coach? - Needs-analyses send to Jonkman, coach and Galema?	U / S / G
(P5) Based on the collected information acquire a full and correct overview of the business processes and information streams within the organization		Needs-analysis: complete and correct?		U / S / G
(P6) Verify the correctness of acquired information, structure this information, and consider possible conclusions rationally by using available facts and carefully weighing pros and cons.		Needs-analysis: complete and correct document with clear structure?	Report send to Boekhorst requesting to check for correctness?	U / S / G