

# Retrospective cognitive feedback for progress monitoring in serious games

**Rob J. Nadolski and Hans G. K. Hummel**

Rob J. Nadolski is assistant professor of technology enhanced learning with a focus on effective skill-based education and training using innovative educational technologies such as serious games, more specifically addressing improved learner support via technology means. Hans G. K. Hummel is associate professor with a research focus on technology enhanced learning of workplace-based professional skills such as serious games. Address for correspondence: Rob J. Nadolski, Faculty of Psychology and Educational Sciences, Welten Institute (Research Center for Learning, Teaching and Technology), Open University of the Netherlands; Valkenburgerweg 177; 6419 AT Heerlen, The Netherlands. Email: rob.nadolski@ou.nl

## Abstract

Although the importance of cognitive feedback in digital serious games (DSG) is undisputed, we are facing some major design challenges. First of all, we do not know to which extent existing research guidelines apply when we stand the risk of cognitive feedback distorting the delicate balance between learning and playing. Unobtrusive cognitive feedback has to be interspersed with gameplay. Second, many effective solutions for providing cognitive feedback we do know might simply be too costly. To face both challenges, this study offers an efficient approach for providing unobtrusive and retrospective cognitive feedback (RCF) in DSG. This approach was applied onto a game where feedback messages were triggered via simple rules about learners' questioning behavior on four dimensions. We found the experimental condition including such RCF to yield better learning outcomes while maintaining similar motivation.

## Introduction

Experts agree that feedback is crucial for fostering learning by playing digital serious games (DSG) but disagree about which feedback is most effective for learning (eg, Kapp, 2012; Koster, 2005). Conflicting research findings (both classroom learning and e-learning) are largely due to the fact that feedback factors interact with factors related to tasks, learners, feedback providers, feedback messages, learning context and way of feedback provision (Azevedo & Bernard, 1995; Bangert-Drowns, Kulik, Kulik & Morgan, 1991; Hattie & Gan, 2011; Hattie & Timberley, 2007; Kluger & DeNisi, 1996; Kulhavy & Wager, 1993; Narciss, 2013; Shute, 2008; Thurlings, Vermeulen, Bastiaens & Stijnen, 2013). Also, the function of feedback can vary, with main functions of feedback being (a) cognitive, (b) motivational and (c) meta-cognitive. Our study focuses on *retrospective cognitive feedback* (RCF) in DSG based on guidelines derived from feedback research. Such RCF presents feedback messages that enable learners to monitor their own *cognitive progress*. Such feedback is natural part of the game scenario (ie, unobtrusive) and based upon *reasoning with user-specific interaction data* collected during game-play through applying a small set of simple rules afterwards (ie, retrospective). Two issues stand out when designing such unobtrusive and RCF DSG.

First, we need to critically evaluate the extent in which more general research findings about cognitive feedback could be translated to guidelines for designing DSG more specifically. We want to

### Practitioner Notes

What is already known about this topic

- Digital serious games (DSG) can have positive impacts on learning outcomes.
- Cognitive feedback is crucial for fostering more complex skills learning.
- The use of educational technology for cognitive feedback has proven to be effective, but often labor intensive to develop and restricted to problems with clear solutions.
- Feedback might be too obtrusive for game players.

What this paper adds

- A proof-of-concept for applying effective retrospective cognitive feedback (RCF) within a skill-based DSG at minimal costs.
- Evidence-informed design guidelines for cognitive feedback in DSG.
- A practical and feasible approach for developing and exploiting RCF in more complex skill-based DSG that is not obtrusive to the game play.

Implications for practice and/or policy

- The study offers design guidelines and an approach for developing and exploiting RCF in skill-based DSG. The approach taken on RCF is effective and at the same time does not sacrifice motivation as feedback is provided as a natural part of the game scenario.
- RCF that supports learners to monitor their individual progress by applying a small set of simple rules is less complex and costly than most current approaches, which are technically grounded in costly and complicated learner progress modeling. This new approach for calibrated learning could therefore stimulate meaningful adoption of DSG in formal education, and not remain restricted to problems with clear solutions.

maintain advantages of playful learning (like active participation and motivation) over regular e-learning, and at the same time warrant an optimal balance between learning and playing (Harteveld, 2011; Schell, 2008). *Less obtrusive cognitive feedback* (provided as natural part of the gaming scenario) is required when compared to other types of e-learning. Regarding findings from feedback research, we also need to consider that cognitive feedback is provided timely (not too late nor too early) and in an adequate amount (not too much nor too less).

Second, we need to evaluate the extent to which RCF caters for complex skills acquisition through DSG. For instance, this might only be the case when error analysis is quite straightforward and when tasks have clear solutions (eg, Narciss, 2013). *Economic and feasible cognitive feedback* is required when conducting practice-based research and implementing DSG in formal curricula.

Our approach will therefore focus on RCF as type of feedback that will not interfere with the gaming experience because it can be provided as natural part of the game scenario, and that can be *efficiently* generated without investing much additional developmental effort. For the purposes of this study we included and examined RCF in an existing DSG on the topic of IT administration for secondary vocational education. Like in real professional life, students receive an IT problem to work on during a number of fixed phases and steps of game play. (Virtual) clients and stakeholders have to be consulted during these phases. The RCF provided is based on the information exchange and questions asked during these consultations, and contains information about intended products and processes (eg, about the number of relevant questions asked), categorised

on four dimensions. Our study on RCF effects is restricted to the first phase, which can be completed within one classroom session (4 hours), for methodological reasons. Students in that phase learn to draw up a needs report with taking into account client's wishes for their IT solutions. In a randomized control study, one group received a treatment version of the game with RCF while the control group received the game without RCF. We assumed the group with RCF to achieve higher learning results than the group without RCF.

The next, second section of this paper will argue for the importance of RCF and describe its implications for DSG. It describes why and which specific cognitive feedback guidelines have been applied in the design of the RCF and DSG we studied. The third section then describes the method of our study and the fourth section presents its results. The last, fifth section discusses main findings and limitations and offers some suggestions for future research.

### **RCF guidelines for serious games**

This second section presents salient research findings on cognitive feedback and implications for designing DSG. We divide the section in four subsections that: define feedback and describe how technological and societal changes have influenced the landscape of feedback research (first subsection); argue why DSG hold potential for complex skills learning and what then are specific challenges and possibilities for providing RCF (second subsection); summarize relevant cognitive feedback research findings as guidelines (third subsection); and describe how these were included and applied in our RCF for the game under study (fourth subsection).

#### *Feedback and current influences*

It is widely accepted that feedback is needed to foster learning because it helps to shape the perception, cognition or action of the learner (Mayer & Johnson, 2010; Moreno, 2004). Feedback is defined as all post-response information about the actual state of learning or performance in order to regulate the process of learning towards specified standards (Narciss, 2013; Shute, 2008). So, feedback comes after instruction, it feeds back on instruction received. Our research will focus on cognitive feedback that is formative, that contains follow up instructions (Shute, 2008). We use simple rule-based reasoning upon learners' interaction data to provide such *formative cognitive feedback* for learners to monitor their progress.

*Technological developments* have increased the range of possible feedback strategies in online digital learning (Dysthe, Lilleford, Vines & Wasson, 2010; Gikandi, Morrow & Davis, 2011). The use of digital media types, the increased transparency of the learning process via "activity traces," the permanent availability of feedback, and other developments offer new possibilities (Ifenthaler, Eseryel & Ge, 2012; Shute, 2011).

*Societal developments* have increased the awareness that curricula need to more adequately reflect the twenty-first century skills needed for nowadays professional life (Redecker & Johannessen, 2013). Tasks become increasingly complex and (electronic) learning environments have to cater for more complex skills acquisition. In this context, DSG containing RCF cater for achieving twenty-first century skills because of their technological (learning progress model, logging) and conceptual affordances (active participation, motivation).

#### *RCF and serious games*

In DSG for acquiring complex professional skills, the context of gameplay has to resemble the context in which students apply what has been learned. Learning activities will then lead to active participation, will be motivating through their real-world relevance, and will increase the likelihood of transfer to real-world situations (Herrington, Oliver & Reeves, 2003). Studies indicate that well designed DSG lead to better learning results when compared to traditional learning

(Connolly, Boyle, MacArthur, Hainey & Boyle, 2012; Tsai, Tsai & Lin, 2015; Wouters & van Oostendorp, 2013).

Several game-based learning scholars stress the need for immediate feedback (Gee, 2003; Kickmeier-Rust & Albert, 2010), but there is a downside as promptness may distort game experiences and lead to superficial (instead of deep) learning. It is therefore better to consider “timely” cognitive feedback provided just-in-time after some accomplishment, when learners can still associate this with their efforts. Furthermore, experts argue why cognitive feedback needs to be based on students’ changing abilities to keep them “in their right flow of engagement” (Csikszentmihalyi, 1990; Gee, 2003), and needs to be unobtrusive in order not to hamper this flow (Bente & Breuer, 2009). A main challenge in designing DSG is to provide customized cognitive feedback to learners while maintaining this delicate balance between learning and gaming (eg, Harteveld, 2011; Schell, 2008). RCF (derived from user-specific interaction data) can contain such customized information about learner’s progress towards objectives.

Current approaches for including customized feedback in DSG by unobtrusively gathering data about learners’ progress are technically grounded in learner modeling (Khenissi, Essalmi, Jemni & Kinshuk, 2015). Creating such a learner model requires many observations of learner-game interactions and interpreting them in terms of both learning progress and learner characteristics. For example, Shute and colleagues used the framework of Evidence Centered Design (ECD) to develop conceptual assessment models, which in turn support the design of valid assessments (Shute & Kee, 2012). Although the ECD and similar frameworks are powerful, their practical use for skill-based DSG is complex and their top-down approach can be too rigorous, they are always time-consuming, and often need many DSG-users. Part of the complexity stems from the open-ended nature that is inherent to skill-based DSG where it often is difficult to attach meaning to collected data (logging and traces) and to demonstrate that what students have learned is similar to intended learning outcomes. We need a more feasible development and exploitation approach for developing RCF in skill-based DSG. For this, we propose to restrict feedback customization to learning progress and not to include learner characteristics as these require more obtrusive collection of user data, induce more complexity or might even be not feasible in case of small amounts of DSG-users. Only few studies have centred on the delivery of RCF and indeed appear labor intensive and restricted to problems with clear solutions (Kickmeier-Rust & Albert, 2010; Narciss, 2013; Tsai *et al.* 2015), or simply limited to conceptual considerations (Westera, 2015).

No specific research uses more feasible development and exploitation approaches, which we consider key to larger uptake of DSG in daily educational practices, and many issues remain on how to design such feedback (both conceptually and practically). In the last two subsections, the findings from feedback research will be used to derive guidelines that inform RCF design for DSG in general, and to describe the RCF for our DSG and study in particular.

### *Design guidelines for RCF in DSG*

Based on most salient feedback research findings (mainly from classroom learning, some from digital learning) in review studies (Azevedo & Bernard, 1995; Bangert-Drowns *et al.* 1991; Hattie & Gan, 2011; Hattie & Timberley, 2007; Kluger & DeNisi, 1996; Kulhavy & Wager, 1993; Shute, 2008; Thurlings *et al.* 2013), we derive six general guidelines for feedback design: (1) Feedback should be both formative and informative; (2) Feedback should be customized to learner’s progress; (3) Feedback should be neutral and informative towards improving learning; (4) Feedback should be both process- and product-oriented; (5) Check if feedback is used and understood; and (6) Feedback should be presented just-in-time.

Regarding (1), formative assessment triggers students to take responsibility for their own learning and gain a sense of self-agency (Hudson & Bristow, 2006), encourages adoption of active learning

strategies and helps achieve deeper learning (Rolfe & McPherson, 1995), and contributes to learning more than feedback from summative assessment will (Shute, 2008). Formative feedback not only provides information about “How did I do?” but also about “Where am I going (goals)?,” “How am I going (progress towards these goals)?,” and “Where should I go next (what activities need to be undertaken in which order to progress)?” (Hattie & Timperley, 2007).

Regarding (2), findings show that detailed feedback is only useful for novice performers (Davis, Carson, Ammeter & Treadway, 2005), and can even hinder transfer of sustainable knowledge (Goodman, Wood & Chen, 2011). Feedback effectiveness is related to learning experiences (Davis *et al.* 2005; Goodman *et al.* 2011). All studies suggest that feedback should adapt to the learner’s level of mastery (Kalyuga, Ayres, Chandler & Sweller, 2003), and powerful feedback manages to engage the learner with learning tasks on adequate levels of functioning (Hattie & Gan, 2011).

Regarding (3), feedback should be specific and clear, task- or goal-directed, informative and neutral (Thurlings *et al.* 2013; Bangert-Drowns *et al.* 1991; Van der Kleij, Timmers & Eggen, 2011). Kulhavy and Wager (1993) suggest to focus on the instructional content and not to undermine intrinsic motivation. Although serious games are undeniably based on preservation of motivation, this guideline stipulates that motivation depends more on other game characteristics (like authenticity and storyline).

Regarding (4), cognitive feedback should focus on both the learning process and the learning product (Gibbs & Simpson, 2004; Rochera & Espasa, 2013).

Regarding (5), it is important that learners recognize the value of cognitive feedback and use it timely. Feedback receivers should be able to interact with feedback providers (Thurlings *et al.* 2013). Hepplestone, Holden, Irwin, Parkin and Thorpe (2011) underline that feedback is often discarded by learners or consulted too late (Hepplestone *et al.* 2011).

Regarding (6), where only behaviorists claim that feedback should be given immediate, most other learning theories agree that feedback should be provided frequently and timely, that is when the content is still relevant for the learner (Gibbs & Simpson, 2004; Thurlings *et al.* 2013). The closer feedback is to actual performance (just-in-time), the more powerful its impact will be on subsequent performance and learner motivation (Tsai *et al.* 2015; Van der Kleij *et al.* 2011, Wouters & van Oostendorp, 2013).

Guidelines 1, 3 and 4 refer to the content of the RCF-message, guidelines 2 and 6 to the RCF-mechanics (rules), whereas guideline 5 is more generic though also applicable for RCF. Although we seem to know much about the power of cognitive feedback, thanks to several decades of feedback research, we still seem to know little about how to make it work effectively and efficiently in DSG, so we will now describe our approach for using RCF in this study.

#### *RCF as affordable approach*

We followed a rather simple rule-based approach for developing RCF and incorporating it in an existing game on IT administration for secondary vocational education. Based on individual learner progress scores on four assessment dimensions (systematic, adequacy, inefficiency and ineffectiveness), populated by data on interviewing in-game (virtual) characters during game play, customized RCF messages were generated (guideline 2), that are both process- and product oriented (guideline 4). The first dimension (“*systematic*”) is *process-related* and keeps track if learners ask questions on time (and to the right person) according to a phase-based method for solving problems in IT administration. The second dimension “*adequacy*” is *product-related* and keeps track if learners have asked questions to the right persons. The third dimension “*inefficiency*” is *process-related* and keeps track of unnecessary questions. The fourth dimension “*ineffectiveness*” is *product-related* and keeps track of not-asked but necessary questions. Game developers incorporated RCF-



rules into the game script, and applied threshold values to calculate (under) performance scores (eg, when over 40% of essential questions had been missed, the RCF message would contain a “+” evaluation for “ineffectiveness”, that is under performance).

RCF for this study was presented just before the learner started to write a needs-analysis report at the end of the first phase.

A concrete example of such a RCF message could be: *“You seem to adhere to the systematic approach in your client-contacts [systematic +]. It appears that you have missed essential questions (and therefore information) during your interviews [ineffectiveness +]. You could have worked more efficient as you did ask some unnecessary questions now and then [inefficiency –]. A couple of times you asked people questions they cannot answer within their job [adequacy –]. All in all, it seems helpful to interview some persons again and think in advance about which person you are going to ask which questions. Always carefully listen to their answers and make some notes that help you to write your report.”* (Text between square brackets [] is not shown to students).

Such RCF contains rather global information (is not that elaborated) and is not constantly present, but does allow learners to reflect on their previous actions and decide if additional work is needed. We assume that learners will only decide to do additional work if RCF indicates underperformance on dimensions, and otherwise will start writing the needs-analysis report. The RCF feedback message example contains formative and informative information (guideline 1) and is formulated neutral and informative towards improving learning (guideline 3). Snippets of formative information are for example “missed essential questions”; “asked unnecessary questions,” whereas informative information is reflected by snippets like “interview again”; “think in advance”; “listen careful”; or “make some notes.” In our research we investigated whether RCF was used and understood (guideline 5), assuming that RCF was presented just-in-time (guideline 6), to be reflected by learning increase.

## Method

In a randomized control study one group also received retrospective cognitive feedback (RCF condition), whereas the control group only received storyline-guidance (no-RCF condition). We studied whether RCF yielded higher learning outcomes than no-RCF.

### Participants

After excluding those with incomplete datasets and those refusing RCF, data of 65 secondary vocational education students in IT administration participated from four Dutch Regional Centres for Vocational Education (ROCs), with RCF ( $n = 31$ ) and no-RCF ( $n = 34$ ). These students (1 female, 64 male; mean age = 19.0 years,  $SD = 2.0$ ) were all in their second year of study. Comparability of their prior knowledge on the subject was checked by using the pre-test questionnaire which will be described under “instruments.” Comparability of their initial motivation was also checked and described under “instruments.”

### Materials

The study was conducted in the first half of 2015 and used an existing IT-client consultation game (in Dutch), that had been developed (in 2014) by the Dutch Foundation for Practice-based Learning (Stichting Praktijkleren). Game developers used the Emergo method and toolkit (Nadolski *et al.* 2008). The investment in developing and incorporating RCF took less than 3% additional time.

The IT-client consultation game helps students acquire communication skills. Main task is to improve the IT system of a luxury holiday resort. When designing an improved IT-system it is important that the IT-administrator goes through a process of five phases in close cooperation with his client, customers and other stakeholders. Of less importance here are the technical skills,

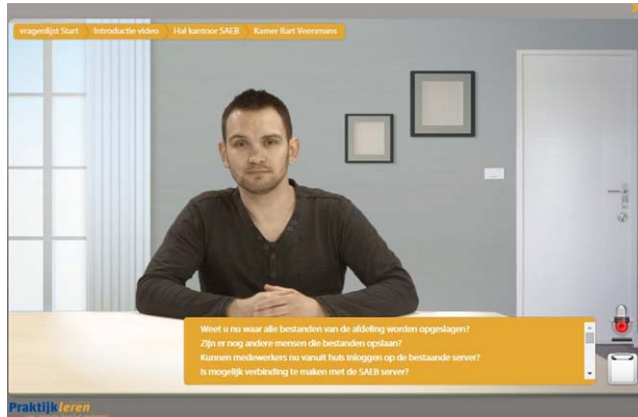


Figure 1: Questioning game-character Bart

but more important the soft skills to clarify clients' needs and desires and stay tuned with clients' expectations throughout development of the IT system. This has to be achieved via good listening skills, clarifying when necessary, and above all by maintaining a flexible and client-oriented attitude.

Most important and frequent student-actions therefore involve questioning game-characters, with at the end of every phase a report that needs to be written and submitted to the (virtual) client. All game-characters (4) have a fixed set of eighty questions that can be asked throughout the game (see Figure 1). In each phase the same set of questions (in writing) can be asked, but the answer (on video) on a specific question depends on the game-character and may depend on the phase in which this question is asked. Examples of questions are: "Should employees be able to log in from home on an existing server?," "Can you help me with the preparation of the technical design?," "Are there sufficient connections to the network?" Students in the RCF group receive RCF from their client when they are invited to write a report (ie, needs-analysis report). The control group only receives the invitation to start writing this report, after other game-characters have been visited with at least one question asked. These five phases of the development process constitute a generally accepted approach for solving IT-problems, and include a functional design, a technical design and a test plan. Both versions of the game included the same storyline-guidance. Guidance is provided by game-characters that indicate the start, end and sequence of phases and when students need to perform tasks. Task instructions are provided within a phase, together with game-navigation hints.

### Instruments

Information on prior knowledge and on motivation was collected by means of two *online questionnaires* (Google Forms): the first, pre-test one, before being enabled to start the first phase; and the second, post-test one, before being enabled to continue with the second phase.

Participants' *prior knowledge* (pre-test) with respect to the subject (client-consultation for IT-system design) was measured by seven MC-items (contained in the first questionnaire).

Participants' *motivation* (pre- and post-test) was measured by including items from an existing instrument (IMI) in the second questionnaire, that were scored on a 7-point Likert-scale. These items represented two subscales (interest/enjoyment (seven items); and value/usefulness (seven items), and two individual items on effort and perceived choice. The internal consistency for these 16 items (with Median = 64) appeared high (Cronbach's  $\alpha = 0.961$ ).

We *logged* all game-behavior through the Emergo-engine, and globally checked data to establish whether participants actually worked with the game and to determine time spent before submitting needs-analysis reports. First author checked for each receiving RCF if the RCF-recommendation was followed up or refused (latter were excluded from the population).

An *assessment instrument* was developed to establish the quality of the phase reports as learning outcomes, developed by two content experts of the Foundation for Practice-based Learning (SPL). For this study, the quality of the needs analysis report was assessed on five aspects/criteria (description of the client's order, description of the reason for this order, mentioning key issues involved in this reason, "must haves" mentioned in correct categories, and "should and won't haves" mentioned in correct categories). These criteria were based on performance indicators that are nationally accredited as part of so called Qualification Dossiers. Each criterion for task-execution receives a performance score of 0 (insufficient), 1 (poor) or 2 (good), leading to a total score between 0 and 10 for the report. Assessment forms with criteria and examples were available during scoring.

### *Procedure*

All students enrolled in the course ( $n = 110$ ) were informed by their teacher that they could voluntary participate in this study about 2 weeks before the start date, and were provided with information on research-specific activities, their time-investment, and the time-schedule. Students signed an informed consent form before participating. They were informed that their data would be anonymized for scientific purposes and that they could withdraw at any time without giving reasons. From each ROC-institute, participants were randomly assigned to one of the two conditions (RCF, no-RCF). Students were told the research investigated the usability and usefulness of the game, but were unaware that two different game-versions were examined. All five game-sessions (about 4 hours each) took place in a classroom. Students were required to work individually, which was checked by teachers being present. After completing the first phase/game-session (focus of this study), another four phases/sessions followed. We limited our study on the first phase for two methodological reasons: (1) prevent interdependency between phases; and (2) exclude the opportunity for students to discuss/cheat between game-sessions (about 20 hours are needed to complete the complete DSG). Also for reasons of experimental control, students could only study the web-based game during these fixed time slots.

At the start of the first session, participants filled in a first electronic questionnaire (Google-form). After submitting their responses they were automatically directed to the game. Students were required to conduct a needs-analysis in this first game-session. This session was finished by submitting their needs-analysis report by using the in-game mail-function. As was already mentioned under "Materials," participants in the RCF-condition received RCF before writing their report. At the end of this phase, the second electronic questionnaire (Google-form) had to be filled in and submitted before being able to continue with the next phase. Then four subsequent game-sessions followed a similar approach, with the game being finished after the fifth session.

About a week after students had finished the game, focused interviews were conducted by the second author at each institution with groups of students and teachers involved. Students could reflect on game's usability and usefulness in a more qualitative sense. The electronic questionnaires mainly gathered quantitative information. After the interview, students received a small present (memory stick) for their participation.

### *Data analysis and scoring*

A total of 110 students started the course and were randomly assigned to a condition. Upon data collection we excluded participants with incomplete data sets ( $n = 34$ ), either because they did not finish the game-session, just submitted an empty report, or did not submit both



questionnaires (non-respons 31%). Furthermore, we decided to also exclude those participants who did receive RCF but refused to follow up the recommendations ( $n = 11$ ). For example, the recommendation might be to do some additional interviews, but from the logging it appeared that no additional interviews were conducted before submitting the needs-analysis report. Those participants technically did not take the RCF treatment so could not benefit from it neither. The RCF-refusers did not differ in their motivation as compared to included RCF-participants.

Differences reported for both conditions are therefore for the remaining 65 students (59% of all students that took the course). Two domain experts (teachers from different education institutes involved) independently and blindly scored all reports that had been submitted after phase 1 (needs analysis) on five phase-specific criteria with a total score on a 10 point-scale. To establish interrater reliability of the assessment instrument, we calculated a weighted Kappa that can be considered as “fair” ( $K_w = 0.363$ ,  $SD = 0.063$ ). Leaving out one or more assessment criteria did not substantially change the Kappa value for this assessment instrument.

## Results

This section first describes our findings on performance and main control variables (motivation, prior knowledge and time-on-task). Then we present some more qualitative findings from focused interviews.

### *Performance*

The mean performance results for needs-analysis reports of individual respondents, based upon averaging values from both raters’ assessments, are summarized in Table 1.

An independent samples *T*-test revealed a significant performance difference between both conditions ( $t(63) = 2.11$ ,  $p < .05$ ,  $d = 0.52$ ). As predicted, the RCF-group showed higher performance than the no-RCF-group. The effect size (by interpreting the value of Cohen’s  $d$ ) could be considered as “moderate” ( $0.3 < d < 0.6$ ) towards “large” ( $d > 0.6$ ).

### *Motivation, prior knowledge, time-on-task*

An independent samples *T*-test revealed that both groups did not differ on prior knowledge, initial motivation and time spent when submitting their needs analysis report. The post-test motivation results for both groups are summarized in Table 2. An independent samples *T*-test revealed that both groups did not significantly differ on motivation.

### *Qualitative findings*

From questionnaires and interviews some relevant, more qualitative impressions could be derived. Both teachers and students stressed that the game was not yet well embedded in the

Table 1: Performance results

RCF (n = 31)		no-RCF (n = 34)	
M	SD	M	SD
3.90	1.68	3.03	1.66

Table 2: Motivation results

RCF (n = 31)		no-RCF (n = 34)	
M	SD	M	SD
62.03	20.52	60.62	22.01

curricular organisation, with the consequence that they started without being adequately prepared. It also appeared that students had expected a more dynamic game (containing a high-end environment, game mechanics like scoring and levelling, and more interactive), with more specific and detailed feedback that make expected behavior more evident to students. Although the navigation and in-game guidance was generally evaluated as positive, students were not always certain which specific actions were needed when recommended to continue questioning game-characters, in the absence of more detailed feedback ("Which exact question(s) do I need to ask then, and to whom?").

## Discussion

This study demonstrated a positive contribution of RCF within a DSG for IT administration. The RCF was designed according to cognitive feedback guidelines derived from feedback research literature. The RCF was implemented using a small set of simple reasoning rules as efficient, feasible and lean approach for customizing feedback. Such an approach could stimulate adoption of DSG in formal education, and better exploit new affordances of DSG for learning and assessment (Gee, 2003). The results of this study show that RCF can indeed increase learning outcomes without at the same time sacrificing motivation, provided that RCF is actually followed up (Hattie & Gan, 2011).

However, the findings in this study should be considered with some care. Both game-versions on average yield rather disappointing scores on both performance and motivation. Partly due to students' false expectations as already mentioned under "qualitative findings." In addition, some students might not have taken the needs-analysis ("just" an intermediate result) as that important because they are used to be evaluated on end results. Finally, some organizational flaws might have hampered the findings.

Furthermore, we have come to feel that our approach should better address the need for more detailed feedback (eg, include exact information on which essential questions were not asked). Interviews indicate that some students would have liked intermediate feedback from their teachers as well, which points to feedback as mechanism in developing trustworthy teacher-student relationships (Hattie & Gan, 2011). Carrying out this study in actual practice made it ecologically valid, but at the same time implied a rather limited number of participants and less controllable circumstances which complicated researching the effect of RCF. A better interrater reliability of the assessment instrument might be achieved if raters are trained in its usage. Also, the generalizability of findings might be limited because of students' characteristics and IT-domain features. Finally, as our study focused on cognitive feedback *within* serious games, it stands the risk that metacognitive feedback and motivational feedback are overlooked when exploiting serious games. The third guideline stipulates leaving motivational feedback out of the game. We think that several other game characteristics (like authenticity, storyline) already cater for prolonged motivation. In this study, only sessions after the game explicitly addressed metacognitive functioning. However, it is open for discussion whether metacognitive feedback could and should also be a natural part of the storyline within the game. Narciss (2013) offers an excellent but extended model for feedback which encompasses the three main functions of feedback and that could eliminate the risk of overlooking.

Our approach for customized feedback seems promising since it is both effective and efficient (less costly and less complex). It might therefore be a more feasible approach in the context of formal education. A possible disadvantage of the approach might be that rather global and infrequent RCF does not sufficiently match with more specific learning objectives. The improvement of the approach towards generating more content-valid RCF would therefore be needed when we use DSG for summative assessment.

## Acknowledgements

Authors would like to thank all students and teachers that participated, in particular Tom Hogendoorn (IT teacher) and Ron Grieken (IT coordinator) at Nova College. We thank Stichting Praktijkleren for collaborating in the development and research of the game, more in particular Martin van Kollenberg (IT coordinator) and Ton Remeeus (director). We thank Marc Hector (content) and Hub Kurvers (programming) for game development and Wim Westera of the Welten Institute at the Open University of the Netherlands for his inspiring ideas as project leader of SLEM. This publication has been produced in the context of the SLEM project which has received funding from the Dutch NRO under grant agreement No. 405-14-504.

## Statements on open data, ethics and conflict of interest

At the time of writing, the data are stored on a local server. In the future, the data will be transferred to DANS Data Archiving and Networked Services. More information is available on <http://www.dans.knaw.nl/en>.

The cETO (institutional ethics committee at the Open University of the Netherlands) approved this study (U2015/01613/HVM). More info about cETO (in Dutch) is available on <http://www.ou.nl/web/onderzoek/ethiek-wetenschappelijke-integriteit-en-wetgeving>.

There is no conflict of interest.

## References

- Azevedo, R. & Bernard, R. M. (1995). A meta-analysis of the effects of feedback in computer-based instruction. *Journal of Educational Computing Research*, 13, 111–127.
- Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A. & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, 61, 213–238.
- Bente, G. & Breuer, J. (2009). Making the implicit explicit: embedded measurement in serious games. In U. Ritterfeld, M. Cody & M. Vorderer (Eds.), *Serious games: mechanisms and effects* (pp. 322–343). New York: Routledge.
- Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T. & Boyle, J. M. (2012). A systemic literature review of empirical evidence on computer games and serious games. *Computers & Education*, 59, 661–686.
- Csikszentmihalyi, M. (1990). *Flow: the psychology of optimal experience*. New York: Harper Perennial.
- Davis, W. D., Carson, C. M., Ammeter, A. P. & Treadway, D. C. (2005). The interactive effects of goal orientation and feedback specificity on task performance. *Human Performance*, 18, 409–426.
- Dysthe, O., Lilleford, S., Vines, A. & Wasson, B. (2010). Productive E-feedback in higher education. Some critical issues. In S. Ludvigsen, A. Lund, I. Rasmussen & R. Säljö (Eds.), *Learning across sites: new tools, infrastructures and practices* (pp. 243–258). Oxford, UK: Pergamon Press.
- Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. New York: Palgrave Macmillan.
- Gibbs, G. & Simpson, C. (2004). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education*, 1, 3–31.
- Gikandi, J. W., Morrow, D. & Davis, N. E. (2011). Online formative assessment in higher education: a review of the literature. *Computers & Education*, 57, 2333–2351.
- Goodman, J. S., Wood, R. E. & Chen, Z. (2011). Feedback specificity, information processing, and transfer of learning. *Organizational Behavior and Human Decision Processes*, 115, 253–267.
- Harteveld, C. (2011). *Triadic game design: balancing reality, meaning and play*. London: Springer-Verlag
- Hattie, J. A. & Gan, M. (2011). Instruction based on feedback. In R. Mayer & P. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 249–271). New York: Routledge
- Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81–112.
- Hepplestone, S., Holden, G., Irwin, B., Parkin, H. P. & Thorpe, L. (2011). Using technology to encourage student engagement with feedback: a literature review. *Research in Learning Technology*, 19, 117–127.
- Herrington, J., Oliver, R. & Reeves, T. C. (2003). Patterns of engagement in authentic on line learning environments. *Australian Journal of Educational Technology*, 19, 59–71.

- Hudson, J. N. & Bristow, D. R. (2006). Formative assessment can be fun as well as educational. *Advances in Physiology Education*, 30, 33–37.
- Ifenthaler, D., Eseryel, D. & Ge, X. (2012). *Assessment in game-based learning: foundations, innovations and perspectives*. New York: Springer.
- Kalyuga, S., Ayres, P., Chandler, P. & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38, 23–31.
- Kapp, K. M. (2012). *The gamification of learning and instruction: game-based methods and strategies for training and education*. San Francisco: Pfeiffer.
- Khenissi, M. A., Essalmi, F., Jemni, M. & Kinshuk, (2015). Learner modeling using educational games: a review of the literature. *Smart Learning Environments*, 2. doi:10.1186/s40561-015-0014-y. Retrieved June 6, 2015, from <http://www.slejournal.com/content/2/1/6>
- Kickmeier-Rust, M. D. & Albert, D. (2010). Micro-adaptivity: protecting immersion in didactically adaptive digital educational games. *Journal of Computer Assisted Learning*, 26, 95–105.
- Kluger, A. N. & DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119, 254–284.
- Koster, R. (2005). *A theory of fun for game design*. Scottsdale, AZ: Paraplygh Press.
- Kulhavy, R. W. & Wager, W. (1993). Feedback in programmed instruction: historical context and implications for practice. In J. Dempsey & G. Ales (Eds.), *Interactive instruction and feedback* (pp. 3–20). Englewood Cliffs, NJ: Educational Technology Publications.
- Mayer, R. E. & Johnson, C. I. (2010). Adding instructional features that promote learning in a game-like environment. *Journal of Educational Computing Research*, 42, 241–265.
- Moreno, R. (2004). Decreasing cognitive load for novice students: effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional Science*, 32, 99–113.
- Nadolski, R. J., Hummel, H. G. K., Van den Brink, H. J., Hoefakker, R., Slootmaker, A., Kurvers, H., *et al* (2008). EMERGO: methodology and toolkit for efficient development of serious games in higher education. *Simulations & Gaming*, 39, 338–352.
- Narciss, S. (2013). Designing and evaluating tutoring feedback strategies for digital learning environments on the basis of the interactive tutoring feedback Model. *Digital Education Review*, 23, 7–26.
- Redecker, C. & Johannessen, O. (2013). Changing assessment: towards a new assessment paradigm using ICT. *European Journal of Education*, 48, 79–96.
- Rochera, M. J. & Espasa, A. (2013). Guest editors' introduction formative feedback in digital learning environments. *Digital Education Review*, 23, 1–6.
- Rolle, I. & McPherson, J. (1995). Formative assessment: how am I doing? *Lancet*, 345, 837–839.
- Schell, J. (2008). *The art of game design: a book of lenses*. Burlington, MA: Morgan Kaufmann Publishers.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78, 153–189.
- 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: IAP Information Age Publishing.
- Shute, V. J. & Kee, F. (2012). Games, Learning, and Assessment. In D. Ifenthaler, D. Eseryel & X. Gee (Eds.), *Assessment in game-based learning: foundations, innovations and perspectives* (pp. 43–58). New York: Springer.
- Thurlings, M., Vermeulen, M., Bastiaens, T. & Stijnen, S. (2013). Understanding feedback: a learning theory perspective. *Educational Research Review*, 9, 1–15.
- Tsai, F. H., Tsai, C. C. & Lin, K. Y. (2015). The evaluation of different gaming modes and feedback types on game-based formative assessment in an online learning environment. *Computers & Education*, 81, 259–269.
- Van der Kleij, F. M., Timmers, C. F. & Eggen, T. J. H. M. (2011). The effectiveness of methods for providing written feedback through a computer-based assessment for learning: a systematic review. *CADMO*, 19, 21–39.
- Westera, W. (2015). On the cybernetic arrangement of feedback in serious games: a systems-theoretical perspective. *Education and Information Technologies*, 20, 57–73.
- Wouters, P. & van Oostendorp, H. (2013). A meta-analytic review of the role of instructional support in game-based learning. *Computers & Education*, 60, 412–425.