Computers & Education 75 (2014) 136–148

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

A multilevel analysis of the effects of external rewards on elementary students' motivation, engagement and learning in an educational game

Michael Filsecker[a,*], Daniel Thomas Hickey[b]

[a] Duisburg-Essen University, Berliner Platz 6-8, Essen 45127, Germany
[b] Indiana University, 201 North Rose Avenue, Suite 4000, Bloomington, IN 47405-1006, USA

A R T I C L E I N F O

Article history:
Received 9 November 2013
Received in revised form
8 February 2014
Accepted 11 February 2014
Available online 26 February 2014

Keywords:
Educational games
External rewards
Digital badges
Gamification
Motivation
Engagement

A B S T R A C T

This study investigated the effects of external rewards on fifth graders' motivation, engagement and learning while playing an educational game. We were interested in exploring whether the feedback-rich environment of the game could mitigate the predicted negative effects of external rewards. Data of students' engagement and learning were collected and analyzed at multiple levels. A quasi-experimental design was used to examine the effect of external rewards in one group ($n = 50$) compared to a control group without such rewards ($n = 56$). According to the results, the external rewards did not undermine students' motivation (e.g., at proximal and distal levels), however they did not foster disciplinary engagement. On the other hand, students in the reward condition showed significantly larger gains in conceptual understanding (proximal) and non-significantly larger gains in achievement (distal). These results suggest that the predicted negative consequences of external rewards may be addressed in this new generation of learning environments. Future research and contributions of the study are provided.

1. Introduction

Despite the controversy concerning the use of external rewards in education (Cameron, Pierce, Banko, & Gear, 2005), they have remained a common practice for supporting achievement, motivation and appropriate behavior (Hoffmann, Huff, Patterson, & Nietfeld, 2009). Furthermore, today such rewards have rapidly expanded alongside increasingly popular educational innovations such as *educational games* and *digital badges* as well as other *gamification* strategies. However, apparently little research has explored both the potential negative and positive consequences of such reward systems on individuals' motivation, engagement and learning[1] in these newer technological contexts.

The past decade has seen an interest in the design of *educational* games at the national and international level (Federation of American Scientists, 2006; Honey & Hilton, 2011; Kirriemuir & McFarlane, 2004). Well-designed educational games offer continual opportunities for player improvement, massive amounts of feedback, tasks too complex for any one individual to solve alone, and environments that change in response to learners' actions (Barab, Gresalfi, & Ingram-Goble, 2010; Garris, Ahlers, & Driskell, 2002). However, research has consistently shown that learners often do not take advantage of the resources they could use to improve and learn (e.g., Hickey, Ingram-Goble, & Jameson, 2009; Nelson, 2007). All this translates into a poor engagement with the learning content, which may be at the core of the modest empirical evidence of games as learning tools (cf. Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012; Hay, 2005; Randel, Morris, Wetzel, & Whitehill, 1992). In terms of educational game design, these resources usually take the form of facts and content embedded in the game narrative, which may be experienced as disruptive or disconnected (Filsecker and Kerres, in press, 2013) instead of a more "endogenous" sense in Malone (1981). Therefore, an extra motivational support might be needed. This situation was also the case in Taiga (Hickey et al., 2009) – the learning environment used in the present study and which is part of the larger project *Quest Atlantis* (Barab et al., 2010). As Taiga

* Corresponding author. Tel.: +49 0201 183 6398.
E-mail addresses: michael.filsecker@uni-due.de, michael.filsecker@gmail.com (M. Filsecker), dthickey@indiana.edu (D.T. Hickey).

[1] Motivation and engagement are usually employed interchangeable. However, we follow the motivation–volition distinction proposed by Kuhl (1987) and consider engagement as describing an ongoing, volitional process (Filsecker & Kerres, in press). Therefore, in this paper we discuss motivation and engagement in separate sections.

http://dx.doi.org/10.1016/j.compedu.2014.02.008
0360-1315/

has usually been a two weeks implementation, other motivational features in Quest Atlantis, which require more time, have not been fully implemented. To overcome this problem we sought quicker motivational strategies. One of these strategies is the inclusion of reward systems. Emulating the reward systems built around videogames (Tobias & Fletcher, 2011) we included *externals rewards* in order to induce individuals' deeper engagement with the resources and learning content embedded in educational games. However, while commercial videogames offer players some form of external reward (such as points or "levels") to motivate their progress, rewards remain controversial in education due to their possible negative side effects on individuals' motivation.

More recently, these rewards have rapidly expanded in educational settings in the form of digital badges and other gamification practices (e.g., Deterding, Dixon, Khaled, & Nacke, 2011; Dominguez et al., 2013; Hickey & Rehak, 2013; Lee & Hammer, 2011; Landers & Callan, 2011; Muntean, 2011). These digital tokens appear as icons or logos on web pages or other online venues. Numerous schools, organizations, and programs have begun offering them to signify individual accomplishments, such as completion of a project, mastery of a skill, or gaining of an experience. Although they are implemented in order to provide credentials, they are also used to support individuals' motivation to learn and engagement. This last use has raised some concerns about the unintended negative effects of badges on scholars such as Mitchell Resnick[2] and Henry Jenkins,[3] supported also by preliminary empirical evidence (see Abramovich, Schunn, & Higashi, 2013). Their concerns echo earlier ones on rewards and motivation (see Section 1.1), namely that such incentives may undermine individuals' intrinsic motivation and interest in the activity, by focusing on the acquisition of incentives and losing interest in the material itself. As with educational games, proponents of such systems appear to be ignoring these concerns. In contrast, this study aims at empirically examining such concerns in the context of a specific educational game.

Previous research on rewards and gamification has presented the following limitations. First, research on rewards has been conducted mainly in laboratory conditions rather than in more ecologically valid settings (e.g., Pierce, Cameron, Banko, & So, 2003). Second, rewards are usually given by test performance and not during the performance in a learning activity (e.g., drafting a quest). Third, studies seldom examine whether intrinsic motivation for a task transfers to broader activities (e.g., from solving an ecological problem to general interest in scientific issues) (e.g., Cameron et al., 2005). Finally, although some studies have proposed mediator variables to understand the process by which rewards affect motivation (e.g., Harackiewicz & Manderlink, 1984), most studies do not provide evidence related to the differential quality of individuals' engagement during specific learning activities (e.g., quality of drafted quests). Concerning digital badges and gamification, research is still in its infancy and only a few empirical studies have addressed the effects of such motivational strategies (e.g., Abramovich et al., 2013; Domínguez et al., 2013; Landers & Callan, 2011). A central weakness of this research is that it usually acknowledges the role of gamification as fostering engagement, but without providing a definition and operationalization of the concept. Without an explicit definition of the desired outcome (i.e., engagement), it makes it difficult to empirically address the effectiveness of gamification and similar design strategies. Finally, these studies tend to replicate the tendency of awarding badges for test performance, instead of awarding them during an ongoing learning activity.

The aim of this paper is to examine the consequences of external rewards within an educational game on students' motivation, engagement and learning science. In the context of Quest Atlantis, a quasi-experimental study compared students ($n = 106$) playing in two conditions. In one condition, students were rewarded with a badge affixed to their in-game virtual avatar. They also were invited to move a paper version of their avatar up and across a physical "leader board" that was prominently placed in the room. Therefore, this reward condition was called Public Recognition (PR) condition. In the control or Non Public Recognition (NPR) condition, students were not offered badges nor given the opportunity to display their progress via the leader board. To gain a deeper understanding of the effects of this reward system on engagement and learning, we used a multilevel model of assessment (Hickey & Anderson, 2007) that captures at different levels increasingly removed from the curriculum (i.e., individuals' actions, discourse, understanding and achievement) motivational and learning variables. Thus, fine-grained analyses were conducted of individuals' actions, discourse and understanding within an ecologically valid setting.[4]

1.1. External rewards and motivation

Rewards represent a positive external influence. As Cameron and Pierce explain: "external rewards are those that come from outside the person and are usually arranged by other people" (Cameron & Pierce, 2002, p. 27). Cognitive theorists have suggested that rewards are detrimental for individuals' intrinsic motivation and subsequent engagement by undermining their perception of competence and autonomy and/or by deviating the perceived source of motivation to external causes. The first mechanism is accounted for by Cognitive Evaluation Theory or CET (Deci, Koestner, & Ryan, 1999), while the second mechanism is accounted for by attribution theory and the "overjustification" hypothesis (Lepper, Greene, & Nisbett, 1973; Tang & Hall, 1995). According to Deci et al., rewards have both controlling and informational qualities. If tied to performance standards, rewards are likely to be perceived as controlling, undermining the perception of autonomy and therefore reducing individuals' intrinsic motivation. On the other hand, the overjustification hypothesis suggests that when an external reward is introduced for an activity which was previously interesting, the individual subsequently attributes the basis for the activity to the external reward. Hundreds of studies have shown that "extrinsic" rewards direct attention away from intrinsically motivated learning, leading to diminished engagement once rewards are no longer offered (Tang & Hall, 1995).

Cameron and Pierce (2002), after reviewing 30 years of research on rewards, provided a summary of the conditions that could lead to the negative effects of rewards: (1) task is of high initial interest; (2) use of tangible/material rewards; (3) rewards is offered beforehand (expected); (4) rewards are delivered regardless of the person performance; (5) intrinsic motivation is indexed as free-choice behavior or self-reported task interest following the withdrawal of the reward; and, (6) intrinsic motivation measured with a single assessment. Therefore, from the initial controversy of whether or not external rewards were negative per se, the scientific discussion has focused on the

[2] See http://www.hastac.org/blogs/mres/2012/02/27/still-badge-skeptic.
[3] See http://henryjenkins.org/2012/03/how_to_earn_your_skeptic_badge.html.
[4] By "ecologically valid setting" we mean that the results of our study were produced (in contrast to more controlled studies typical in external reward research) in the context of real classrooms.

factors that might contribute to the effects of rewards on individuals' performance and motivation. As Cameron et al. explain: "the effects of reward depend on the types of rewards used, the reward contingency, how rewards are allocated, and the context in which rewards are administered" (Cameron et al., 2005, p. 641).

By its focus on the *context* as a source of rich "informative value" for individuals, newer sociocultural theories provide a different approach to the debate over external rewards and motivation. This perspective suggests that individuals' participation in practices within a community is motivated by a set of collectively negotiated standards and values that define "successful" participation. In this sense, individuals are motivated to participate in such practices to the extent that the constraints and affordances imposed by the collective standards and values are perceived as consonant with their identities. As Hickey and Zuiker explain: "efforts to motivate learning from a sociocultural perspective should be judged according to the extent they support meaningful participation in authentic practices of the targeted knowledge domain" (Hickey & Zuiker, 2005, p. 283). Among such efforts to motivate learning is the use of external rewards. As discussed above, for cognitive views of motivation this possibility has seemed to be problematic. In contrast, for a sociocultural perspective focused on supporting a context with rich informative value, the negative impact of external rewards is less likely (Hickey, 1997, 2003). As a case in point, in their groundbreaking paper on *cognitive apprenticeship*, Collins, Brown, and Newman (1989) suggested that the corrosive educational effects of competition (which is typically fostered by external rewards) may be more the result of impoverished learning environments that lacked opportunities to improve and the formative feedback needed to do so. Given that most of the prior studies of rewards were conducted in highly structured laboratory settings or very traditional classrooms, this suggests to us that in the newest generation of educational videogames and other more "participatory" networked learning environments rewards might have many positive consequences that outweigh or even eliminate any negative ones.

From these sociocultural considerations, the present article assumes that the potential negative consequences of rewards and its likely associated competition are less likely to occur in rich learning environments (e.g., Quest Atlantis) with considerable useful feedback and opportunities to improve. Second, the rich interactive narratives in the latest generation of immersive videogames and the participatory culture of many networked learning environments might counter or even reverse the overjustification effect by providing individuals with valued standards of practice (Hickey, 2003). Finally, a third assumption is that some of the activities that are likely to be necessary to give games or learning networks sufficient academic content to allow in-school-use may not be sufficiently interesting or appealing to motivate all learners toward sustained engagement.

1.2. Educational games

It has been claimed in the last 30 years that the engaging nature of games may facilitate involvement, motivation and interest, and the retention of learned skills (Greenblat, 1981; Greenfield, 1984; Loftus & Loftus, 1983; Malone, 1981). More recently, it has also been suggested that players of commercial games are developing problem solving and literacy skills and that good commercial games represent good learning principles that provide opportunities for gamers to engage actively and reflectively during gameplay (e.g., Gee, 2003). Others have proposed that games involve individuals in a virtual cycle of action–feedback–reflection (Hickey et al., 2009) or user judgments–feedback–behavior (Garris et al., 2002), and have situated engagement as the core of gameplay (Dickey, 2005). Furthermore, the "immersive" nature of modern educational games is suggested to have the potential to transform the knowledge outlined in school content standards into "just-in-time" knowledge to solve meaningful problems within interactive narratives (e.g., Barab & Dede, 2007).

However, from an empirical perspective, more tempered conclusions seem more appropriate concerning games' effectiveness. First, the evidence of effectiveness of games on cognitive and conative outcomes is modest (Hays, 2005). Although games may keep some promise for supporting learning, until today rhetoric over empirical studies have prevailed (Tobias & Fletcher, 2011). Second, studies show inconsistent patterns of results concerning motivation and learning. These studies can be essentially classified into three groups: 1) studies showing no motivational effect with positive learning outcomes (e.g., Kebritchi, Hirumi, & Bai, 2010), 2) studies showing a motivational effect but no learning outcomes reported (e.g., Egenfeldt-Nielsen, 2005), and 3) studies showing both motivational and learning effects, but without reporting the relationship between the two (e.g., Cordova & Lepper, 1996). Finally, the effectiveness of games depends to a high degree on how individuals use it and the goals they set for themselves (e.g., Elliot & Harackiewicz, 1994), for example, the kinds of cognitive strategies used might well be what distinguish motivated and unmotivated learners (Renkl, 1997).

On the other hand, the controversy over external rewards is also represented in the history of educational computer games design. External rewards were central to the first generation of games, ranging from PLATO[5] (Tobias & Fletcher, 2011) to commercial drill and practice games (e.g., *Math Blaster*) and "scavenger hunts" (e.g., *Where is Carmen San Diego?*). The rewards in these games are extrinsic because they are arbitrarily related to the educational content (what Rieber, 1996, deemed "exogenous" rewards). In response to their findings of the aforementioned overjustification effect, Lepper and Malone (1987) advanced an influential set of guidelines for using more intrinsic sources of motivation such as curiosity and interest (what Rieber, deemed "endogenous" rewards). These guidelines were manifested in a whole new generation of educational computer games such as LOGO and *Freddy Fish*. While the distinction between intrinsic and extrinsic rewards was not always straightforward, the debate raged as increased computing power allowed for increasingly sophisticated educational computer games. Perhaps the most important compromise in all of these considerations is Lepper and Malone's concession that rewards should only be considered when (a) the initial level of interest in the activity is low, and (b) the activity's attractiveness becomes apparent once the learning is engaged.

1.3. Gamification and digital badges

Gamification is a design strategy attempting to reproduce the engagement power of games by emulating key game mechanics without actually designing a *full* game and implementing them in a non-gaming context (e.g., industry, education, etc.). Some of their proponents define gamification as "the use of game design elements in non-game contexts" (Deterding et al., 2011, p. 10). Others, following this

[5] For more information about PLATO, see http://faculty.coe.uh.edu/smcneil/cuin6373/idhistory/plato.html.

definition, have emphasized the fact that gamification is implemented through technology. For example, Dominguez et al. (2013) defined gamification as "incorporating game elements into a non-gaming software application to increase user experience and engagement" (p. 381). This latter appears more specific and includes also the purpose of gamification as fostering engagement, and it is the one used in this study.

Among the game elements usually found in the gamification literature are rewards systems and, in particular, the use of badges for displaying achievement (e.g., Landers & Callan, 2011). Abramovich et al. (2013) characterized educational badges as (1) normally offered to certificate learning outside formal institutions, (2) viewable by others on the learners' profile, and (3) awarded for either incidental activity (i.e., participation) or mastery of skills/demonstration of knowledge. In other words, badges represent a reward and recognition for the learner's participation and performance (e.g., Simões, Díaz, & Fernández, 2013).

As mentioned before, research on gamification is still in its infancy. There are frameworks to "gamify" social platforms such as *Schoooools.com* for K-6 (Simões et al., 2013), and other practical attempts in the context of online and blended learning (Hickey & Rehak, 2013; Landers & Callan, 2011; Muntean, 2011) and with intelligent-tutor systems (Abramovich et al., 2013). For example, Simões et al. deploy a social gamification framework for a wiki-like online social platform, where students, parents and teachers collaborate and communicate with each other while developing specific projects. The social gamification features of this project can be summarized *wikifolio* as making it possible to give feedback, reward and recognize students' participation and performance by teachers and parents. Similarly, Hickey reported the design of a "Big Open Online Course" (BOOC),[6] for teaching classroom assessment. The design of the course represents a mix of e-portfolio, wikis and digital badges to support disciplinary social engagement and individual understanding. From a sociocultural approach (Hickey & Zuiker, 2005) and an educational assessment perspective (Filsecker & Kerres, 2012), the course is structured to encourage participants to problematize and contextualize the content of the course in the form of a, while at the same time encouraging social interaction by asking participants to discuss, pose questions, and select an exemplary wikifolio while providing reasons for such decision. Although these applications of gamification are promising, no empirical evidence of its effects on motivation and engagement is provided. On the other hand, Landers & Callan also used badges in a platform called *socialpsych*. College students would answer small quiz and according to their results they would be ranked and provided with a ribbon which indicates their performance level. The platform considered five performance levels: "Newbie", "Novice", "Intermediate", "Expert" and "Master". These levels were displayed publicly in the student's profile. A preliminary evaluation of the project in terms of the motivational value of the "gamified" platform showed that a considerable number of students used the gamification features and that by doing so taking the quiz seemed to them a more enjoyable and rewarding activity. Abramovich et al. introduced badges in the context of an intelligent tutor system. The badges were awarded based both on individuals mastery of a skill and on their participation in the system. The authors expected, contrary to the literature on external rewards (e.g., Cameron & Pierce, 2002), that the participation badges, regardless of the actual performance of individuals, would increase their motivation. Although they did not provide data to support nor to reject this hypothesis, they found that after using the intelligent tutor system – which included the badge system – individuals showed more interest in learning mathematics and were less concerned about poor performance as compared to their peers. More specifically, the authors provided some evidence that earning badges was associated with higher expectation for success, but also with higher "counter-productive educational goals" (p. 229). In brief, the study showed that positive and negative effects can be expected from the introduction of educational badges, warranting a more careful study of such mixed effects.

In summary, research on badges has just begun and questions related to their impact on learning, motivation and engagement needs a systematic exploration. This study represents one of the first empirical explorations of the relationship between motivation, engagement and learning with an emphasis on the potential negative consequences of badges as reward systems.

1.4. Engagement

A complete review of the concept of engagement is beyond the scope of this article and is provided elsewhere (Filsecker and Kerres, in press). The meaning of educational engagement is bound to views of learning. Prior scholars have advanced notions such as *mindfulness* (Salomon & Globerson, 1987), *intentional learning* (Bereiter & Scardamalia, 1989) and *committed learning* (diSessa, 2000). As Dewey put it a century ago "…the educational significance of effort, its value for an educative growth, resides in its connection with a stimulation of greater *thoughtfulness* not in the greater strain it imposes" (Dewey, 1913, p. 58). Sociocultural approaches highlight Dewey's thoughtfulness as the process by which students engage in an activity, interact with each other and use resources and tools purposefully. In this characterization, the role of discourse is key to supporting any claim concerning engagement. Finally, as suggested in Section 1.1, engagement is about negotiating our own identities within a particular community of practice. Such engagement depends on the extent at which individuals practices are in line or attuned to the constrains and affordances imposed by the standards and values of a community (Hickey & Zuiker, 2005). In this direction, Engel and Conant's (2002) notion of *productive disciplinary engagement* highlights (a) the number of students making substantive disciplinary contributions, (b) the number of disciplinary contributions made in coordination with each other, (c) students attending to each other and making emotional displays, and (d) students spontaneously reengaging. For the purpose of this study we will focus our attention on the idea of disciplinary engagement defined as the extent at which students "make contact" with both general scholarly discourse and ecological discourse in particular.

1.5. Research questions and hypotheses

In sum, a common postulation is that external rewards, although they can foster individual performance, have negative consequences on individuals' motivation. However, more recent research has shown that under specific circumstances and when rewards have a more informational than controlling value, the negative consequences could be avoided. As suggested before, the richness of the information embedded in a reward system depends on the support and context around which the rewards are implemented, and hence the negative consequences of external rewards may be more indicative of impoverished learning environments and the lack of feedback and opportunity to improve, than of a fundamental consequence of the rewards themselves. One expectation is, therefore, that the immersive educational

[6] http://remediatingassessment.blogspot.de/2013/12/the-varied-functions-of-digital-badges.html.

game used here with its supporting features should protect individuals from the negative consequences on their motivation and interest. Consequently, the study searched for evidence of the predicted negative consequences for motivation during the game and interest in solving ecological-related problems in the future. Our second expectation is that the introduction of external rewards should impact individuals' disciplinary engagement with the content knowledge embedded in the educational game used here and therefore have a positive effect on learning. In particular, the external rewards were expected to result in (a) more access to the resources embedded in the game, (b) higher-quality submissions, (c) increased understanding of the scientific concepts in the game, and (d) increased achievement of targeted science standards.

To test our expectations, a quasi-experimental design was conducted to examine the effect of providing external rewards on students' motivation, engagement and learning science. For two of the classrooms in this study (Public Recognition condition or PR), the teacher's acceptance of a written quest at one of three increasingly accomplished levels (proficient, expert, or wise) was rewarded with a corresponding badge that players could affix to their in-game virtual avatar and a paper version of their avatar they could add to a physical "leader board" (see Sections 2.3 and 2.5 for more details). In two other classrooms students were not offered badges or a ready means to communicate their level of progress to the other students (Non Public Recognition condition or NPR).

Drawing on the theoretical considerations and empirical evidence synthesized in the previous sections, this study was guided by the following research questions and hypotheses:

(1) Does the Taiga environment protect individuals' intrinsic motivation from the negative consequences of the external reward introduced?

Hypothesis 1a
There will be no difference between the PR and NPR conditions in self-reported intrinsic motivation during the Quest 2 essay.

Hypothesis 1b
There will be no differences between the PR and NPR conditions in self-reported personal interest in learning to solve ecological-related problems in the future.

(2) Do individuals who receive an external reward show a deeper disciplinary engagement with the learning activities embedded in Taiga?

Hypothesis 2a
Students in the PR condition will access more feedback pages in the process of drafting their quests than students in the NPR condition.

Hypothesis 2b
Students in the PR condition will use more relevant scientific concepts accurately than students in the NPR condition.

(3) Do individuals who receive an external reward show larger gains in understanding and achievement of the targeted science concepts in Taiga?

Hypothesis 3a
Students in the PR condition will exhibit significantly larger scores on the Quest 2 essay about the ecological process and their indicators than students in the NPR condition.

Hypothesis 3b
Students in the PR condition will exhibit significantly larger gains in conceptual understanding of the targeted science concepts than students in the NPR condition.

Hypothesis 3c
Students in the PR condition will exhibit significantly larger gains in achievement of the targeted science standards than students in the NPR condition.

2. Method

2.1. Participants

This research was conducted at a public elementary school in a medium-sized city in the Midwestern US. As is typical of university communities, the students were predominantly Euro American and most came from well-educated professional families. In this study, average grades from prior work were used to identify pairs of similar achieving classes, and one class in each pair was assigned to the Public Recognition (PR) and the Non Public Recognition (NPR) condition. Consent to participate in the study was obtained from almost every student, resulting in 106 participants (56 females and 50 males).

2.2. Material

The game-based learning environment used in this study was Quest Atlantis (QA). Building directly on the technology and practices of commercial videogames, QA immerses learners in a virtual environment where they can experience complex social situations. "Questers" move through the 3D virtual space interacting with other players via text-based chat and with non-player characters (NPCs) via structured dialogs. The various worlds in QA allow learners to take on roles and make decisions that have consequences in the virtual world. QA provides a rich narrative in which players inform themselves regarding a particular scientific problem, while showing their understanding through written "quests." These quests are drafted with the help of different in-game resources, including graphs and tables, the various non-player characters, peers, and feedback from the teacher (who inhabits the role of the park ranger when reviewing the submissions).

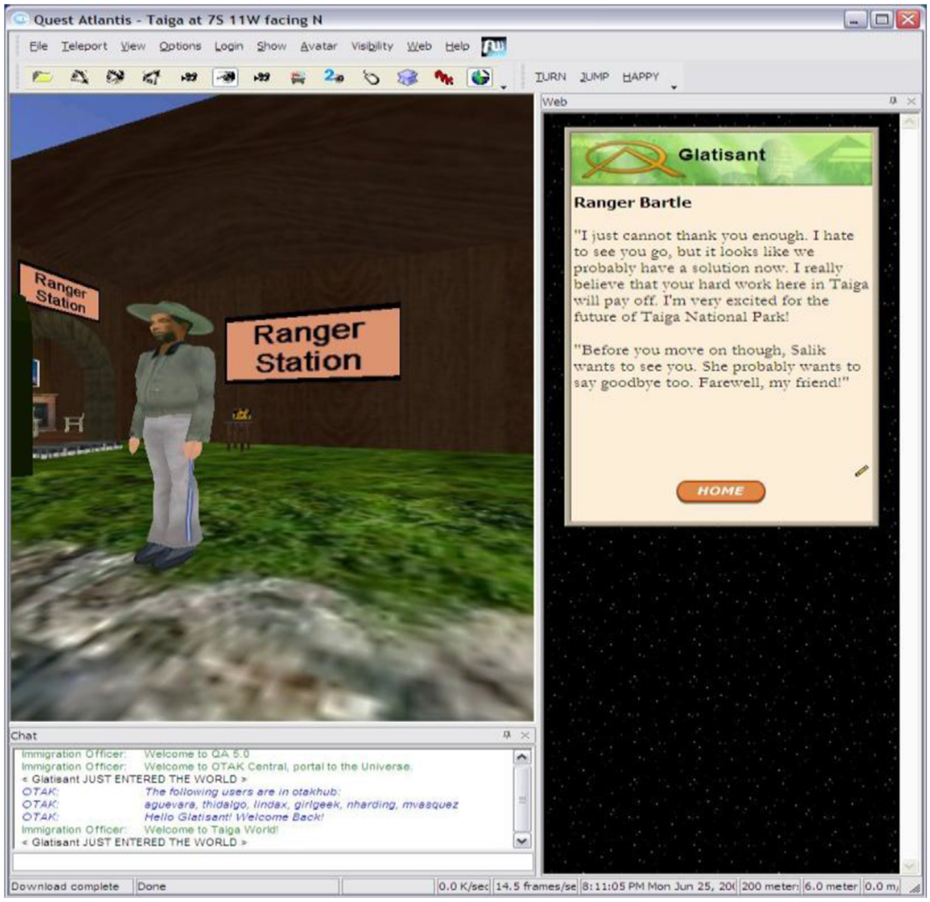


Fig. 1. The Quest Atlantis (Taiga) environment and the NPC Ranger Bartle.

This scenario supposes a feedback- rich space for reflection and the improvement of understanding of the fictional problem embedded in the narrative of the game.

One of these scenarios/worlds in Quest Atlantis is Taiga. Taiga is a park located along a river, and is populated by loggers, tourists, an indigenous farming community, a fishing resort, and a park administration. This world was designed to engage students in complex socio-scientific inquiry (Sadler, 2004), while also helping them learn ecological science concepts like erosion, eutrophication, and hypothesis testing. In Taiga students play the role of an apprentice field investigator and submit their quests to Ranger Bartle (Fig. 1). By means of a detailed letter, Ranger Bartle asks these field investigators to help him figure out why the fish population is declining in his park.[7]

Students then complete a series of missions, with each mission culminating with writing and submitting a 50–100 word quest. Quest 1 requires students to develop an initial hypothesis about the cause(s) of the fish decline. After submitting their hypotheses, students collect and analyze water samples. They evaluate their hypotheses and speculate on alternative causes of the problem in Quest 2 (shown in Fig. 2). In order to warrant their hypotheses about the cause(s) of the fish decline, students must enlist a number of scientific concepts in Quest 2. These include (1) the various water quality indicators such as pH and dissolved oxygen, (2) the processes of erosion and eutrophication, and (3) the dynamic relationship between indicators and processes in the simulated landscape. Quest 2 was a crucial event in the Taiga learning trajectory.

2.3. Measures

In order to address the issues of assessment sensitivity (Annetta, Minogue, Holmes, & Cheng, 2009) and construct-irrelevant variance (Messick, 1994), the measures were collected within a model to balance formative and summative functions of assessment by means of considering different levels of learning outcomes (Ruiz-Primo, Shavelson, Hamilton, & Klein, 2002). The details of the model used here can be found elsewhere (Hickey & Anderson, 2007; Hickey, Zuiker, Taasobshirazi, Schafer, & Michael, 2006).

The model distinguishes four levels: immediate, close, proximal and distal. The immediate-level refers to the enactment of sequences of inquiry-oriented game activities; the close-level refers to the participation among the player, teacher, and non-player characters in writing and revising written "quests" after those activities. The proximal-level refers to the conceptual understanding assessed with a curriculum-oriented performance assessment. And, finally, the distal-level refers to individuals' achievement measured with a multiple-choice test. In the present study, we extended this multilevel model to the motivation and engagement measures as well.

[7] For more details on Taiga see http://www.youtube.com/watch?v=SZoT3pilNPI and www.questatlantis.org.

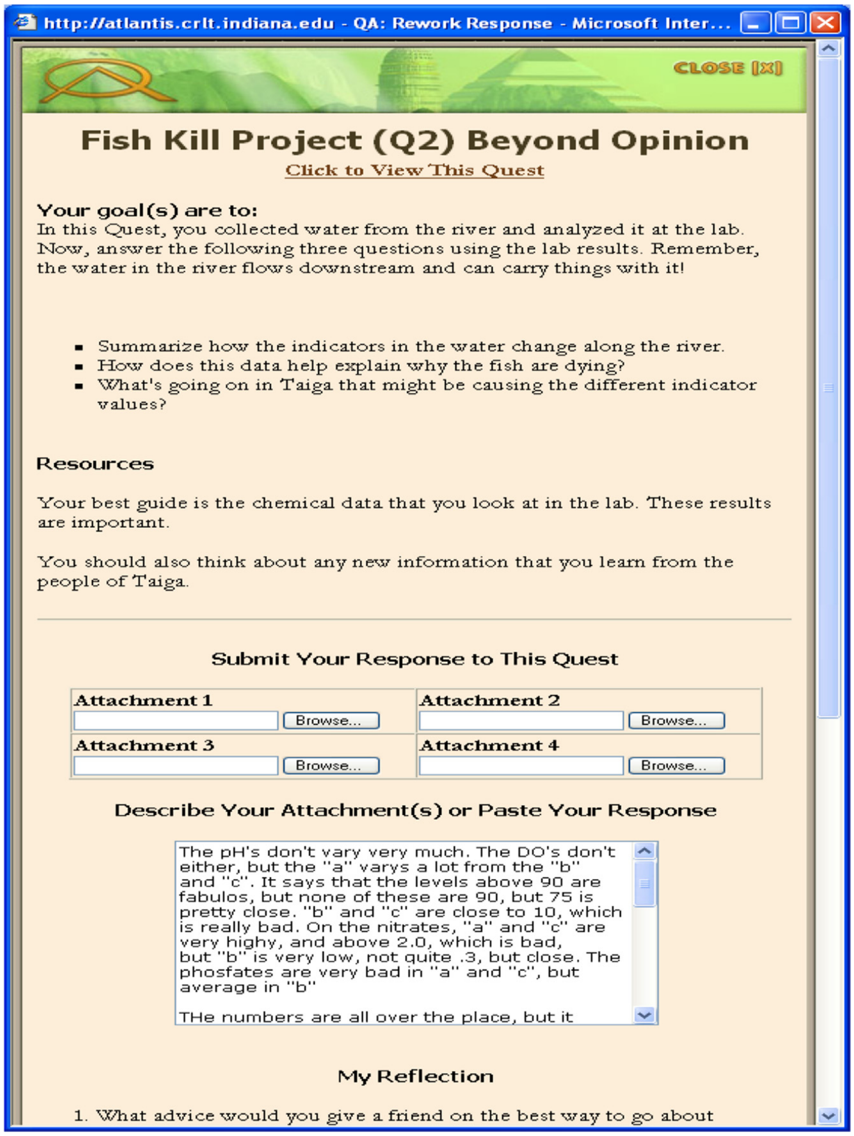


Fig. 2. Quest 2 *Beyond Opinion*.

Motivation and Interest. At the proximal level we examined students' motivation by assessing players' situational motivation regarding the Quest 2 activity. The scale consisted of 4 or 5 Likert-type items (strongly disagree, disagree, neutral, agree, or strongly agree) for each of the following subscales of the motivational states that prior research has shown to be diminished by rewards: *interest* in the activity, *value* for completing the activity, *perceived competence* during the activity, and *effort* completing the activity. Concerning the reliability of the scales, the Cronbach's Alpha values were .82, .81, .84, .78, respectively. So long as the individual scores for each set of items are internally reliable, scores on each scale are presumed to be indicative of various aspects of students' cognitive engagement during the tasks (see Fredricks, Blumenfeld, & Paris, 2004). At the distal level, we measured changes in personal interest in solving the types of problems students were learning to solve in Taiga. Many studies on external rewards have also examined self-reported interest in the activities (and sometimes instead of) free choice engagement. To this end, we measured students' self-reported personal interest in the three types of problems that they were learning to solve in Taiga: *water ecology* problems, *complex scientific* problems, and *controversial socio-scientific* problems. An 18-item survey was created consisting of six Likert-scale items for each type of problem and was administered before and after students played the game. Concerning the reliability of the scales, the Cronbach's Alpha values on the pretest were .76 for water ecology problems, .76 for complex scientific problems, and .77 for controversial problems. The Cronbach's Alpha values for the posttest were .79, .80, .84, respectively.

Disciplinary engagement. At the immediate level we analyzed the number of screens of formative feedback that students accessed by analyzing the log files generated during gameplay as in previous studies (Hickey et al., 2009). At the close level, we analyzed the extent at which students made "contact" with the scientific and ecological discourse (Engel & Conant, 2002). For that purpose, with the help of NVivo 8 software, we quantified the verbal data in Quest 2 for the whole sample (Chi, 1997) to capture this domain-specific or *disciplinary* discourse around students' Quest 2 submissions. Initial and final submissions in Quest 2 were coded in terms of the meaningful appropriation of nine relevant scientific concepts. We were interested in capturing students' engagement with the content in a progressive knowledgeable way as a result of the reward manipulation, instead of students' actual scientific argumentation (e.g., Kelly, Drucker, & Chen, 1998).

Table 1 provides examples of actual students' answers (misspellings in original responses) and the category in which they were coded. The meaningful appropriation of the concepts has to do with (1) identifying the right level of the indicators displayed in the charts, (2) the concept being used to establish a valid relationship with other concepts, (3) relating the concept to a relevant activity or event, and (4) identifying the concept as being the cause/effect of another concept or associated activity or event. A non-meaningful appropriation situation has to do with (1) establishing the concepts in an invalid relationship with one another (2) the concept being used to explain the wrong ecological process (e.g., erosion or eutrophication), or (3) the concept being used incorrectly as a cause or effect of an event or another concept. The category "other" was used when the student's response was too ambiguous to discern among the above categories.

Understanding and achievement. At the close level we analyzed the quality of the initial and final submissions of crucial Quest 2 (scored by two researchers, inter-rater reliability = .85) using a 14-point scale rubric which assigned six points for summarizing the water quality indicators, four points for explaining what the processes were (i.e., erosion and eutrophication), and four points for describing the dynamic relationship between indicators and processes. To examine learning gains at the proximal level, we used the *Lee River* performance assessment developed in the prior design cycles. The assessment was "curriculum-oriented" in that it asked students to solve similar problems as in Taiga but in a somewhat different context. The assessment had been created alongside extensive refinements to Taiga the previous year and was designed to be highly sensitive to different enactments of the curriculum. It involved another fictional watershed and a range of stakeholders who had similar (but not identical) effects on the ecosystem. For example, both Taiga and Lee River involve stakeholders with different land use practices who are arranged along a river. The stakeholders from both scenarios impact their ecosystems by doing things that cause erosion and eutrophication; however, erosion is caused by loggers in Taiga and by construction in Lee River. To capture a range of understanding at the pretest and the posttest, the items covered a broad range of difficulty. The assessment included several multi-part items that started out with simple tasks that most students would be able to answer without instruction, and proceeded to a few complex items that focused on the nuances of scientific hypotheses, the relationship between social issues and scientific inquiry, and the relationship between water quality indicators such as dissolved oxygen and processes like eutrophication. A 21-point scoring rubric was used to score completed assessments, with a subset of assessments scored by two researchers (inter-rater reliability = .85). Finally, at the distal level we examined students' achievement through a 20-item test that had been created the previous year by random sampling from pools of items aligned to the four targeted content standards, but independent of the Taiga curriculum. Such standards-oriented tests are necessary to support claims of impact on externally-developed achievement measures and to compare the impact of different curricula that target those standards. Such tests are not particularly sensitive to specific interventions and represent a relatively ambitious target for innovative curricula like Taiga.

2.4. Procedures

The teacher used Taiga in all four of his fifth-grade science classes in a daily basis during two weeks. Each day students would use Taiga for about 1.2 h. Before starting the Taiga curriculum, students were asked to fill out the pretests which included the questionnaire on future interest, the performance assessment and the achievement test. In this implementation, the teacher was asked to deem each of the quest submissions as indicative of being a beginner, knowledgeable, expert, or wise field investigator (cf. Landers & Callan, 2011), according to specially designed rubrics. Only wise submissions were immediately accepted; otherwise, the teacher provided in-game formative feedback (by pasting and revising text from the rubric in the reviewed quests) and encouraged students to revise and resubmit to attain higher status. On the crucial Quest 2, students were encouraged to visit the lab technician and read through the 30-screen tutorial before resubmitting.

Once students' Quest 2 submission was accepted, they were asked to complete the brief survey on situational motivation. The survey asked students, "How did you feel while completing Quest 2?" The survey also encouraged students to respond honestly and assured students that their responses were confidential. Then students returned to Taiga. By means of an email, students received a notification concerning their submitted quest. Once students' quest were accepted (or an accepted quest revised and improved), students were rewarded with a corresponding badge that they could affix to their in-game virtual avatar (Fig. 3a). Additionally, students were invited to move a paper version of their avatar up and across a physical "leader board" that was prominently placed in the room (Fig. 3b). These students belonged to the Public Recognition condition or PR. In two other classrooms taught by the same teacher in the same semester, students in the Non Public Recognition (NPR) condition were not offered badges or a ready means to communicate their level of progress to the other students and in-game information on rewards was replaced by messages encouraging players to work hard to save the park and become more capable apprentices, following the advice of Lepper and Malone (1987). After approximately two weeks, the curricular unit was completed and date for the posttest was scheduled. The posttest included the scale on personal interest in solving the types of problems presented in Taiga, the performance assessment and the achievement test.

2.5. Data analysis

In order to test the hypotheses, this study compared learning and engagement between the two groups. Although there are different aspects of groups that can be compared (e.g., measures of location or variation) we were interested in the *typical* behavior between the two

Table 1
Examples of a few scientific concepts coded as accurate, non-accurate or other.

Concept	Examples of evidence for coding		
	Accurately	Non-accurately	Other
Dissolved oxygen	Warm temperature takes out the DO in the water so the fish suffocate… (Student ID 118412)	The Temperature is effected by Do level in the river (Student ID 118517)	There is too much of every thing except DO and it is way to hot. (Student ID 118306)
Turbidity	Turbidity is caused some by erosion. (Student ID 118221)	Without turbidity the sun will get through the water and then the plants can't grow. (Student ID 118509)	At site A and C, the Turbidity is in between. (Student ID 118504)

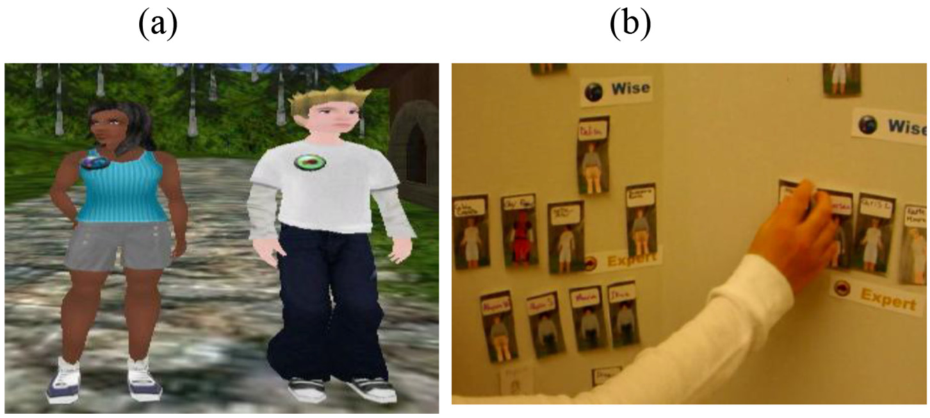


Fig. 3. Avatar badges and leader board displaying public recognition of proficiency.

conditions (i.e., PR and NPR) and the degree at which the independent variable affected that typical behavior in terms of motivation, engagement and learning. Therefore, the mean was employed to compare the two groups (Wilcox, 2012) and different types of analysis of variance (ANOVA) were conducted. In order to examine the effects of the rewards on motivation, we conducted a one way ANOVA on the scores of the situational motivation scale. For the interest scale, which was administered as pre and posttest, a gain score – based on the pre and post scores – was calculated in a repeated measures ANOVA. Secondly, as for the disciplinary engagement, a MANOVA was conducted on the number of appropriate uses of scientific concepts as reflected on Quest 2, in which the different concepts represent the different dependent variables. Thirdly, the quality of Quest 2 was also assessed by a 14-points scale rubric and the final scores were compared by an ANOVA. Likewise, the performance assessment was analyzed using a 21-point scoring rubric. As the performance assessment was administered as pre and posttest, a repeated measures ANOVA was conducted. Finally, scores on the 20-item achievement test, administered as pre and posttest, were analyzed using also a repeated measures ANOVA. Statistical significance for all tests was set at $p < .05$.

3. Results

3.1. Motivation

In order to test our hypothesis of the effects of external rewards on students' motivation and interest, we conducted two ANOVAs. First, at the proximal level, the ANOVA on perceived interest, value, competence, and effort were all slightly higher for the PR condition, but the difference did not reach statistical significance ($F<1$). As for the scales of interest (distal level) in solving the three different types of problems, a one-way repeated measures ANOVA was conducted to compare the effects of rewards on the three indices of interest. While all three types of interest increased slightly in the PR condition relative to the NPR condition, none of the three group $\times$ time interactions were statistically unlikely ($F < 1$). Altogether, these results indicate that students in both conditions reported similar levels of motivation while doing the quest and the same level of interest in solving ecological-related problems.

3.2. Engagement

For disciplinary engagement at the immediate level, analysis of the log files did not show significant differences between the PR and NPR conditions. This indicates that students in both conditions who were asked to revise their Quest 2 submissions accessed nearly the same number of feedback pages before their final submission (averages of 9.8 and 9.7, respectively). To test the hypothesis of the quality of the quests, a one-way MANOVA was conducted to compare the effects between conditions on the meaningful appropriation of the scientific concepts as enlisted during the drafting of Quest 2. The analysis revealed higher levels in the PR condition, but they did not reach statistical significance [Wilks' Lambda $= .973$, $F(1,102) = 2.797$, $p = .097$], indicating that students in both conditions accurately employed a similar number of scientific concepts.

3.3. Learning

Finally, in order to analyze the influence of the manipulation on students' conceptual understanding (i.e., learning at the close and proximal levels) two analyses were carried out. First, an ANOVA was conducted on the initial and final Quest 2 submissions. Resubmitted Quest 2 scores went from an average of 1.5–4.2 points in the PR condition and 1.7 to 4.9 in the NPR condition. These differences were not statistically significant, which indicates that students showed a similar understanding of the causes of the fish dying in the park as related to the water quality indicators and the ecological processes involved. Second, concerning proximal learning, the two groups started out at similar levels. Scores on the problem solving performance assessment increased 9.7 points (of 21) in the PR condition, compared to 7.3 points for the NPR condition. Repeated measures ANOVA revealed that the difference between the two sets of gains was very unlikely to have occurred by chance [$F = 5.6$, $p = .02$]. As shown in Fig. 4, this represented the difference between 1.4 and 1.1 SD gain, given the pooled standard deviations across the score points. Importantly, the differences in gains between the two classes in each condition were not statistically significant ($F < 1$). This result confirms that the understanding of the ecological processes and their indicators was deeper in the PR condition than in the NPR condition.

For distal learning, the achievement tests revealed strong internal consistency, and showed that students in the PR condition gained 5.44 points compared to 4.02 points for the other students. Given the variance within the scores, this was a difference between gains of 1.1 and .8 SD (Fig. 4). A one-way repeated measures ANOVA revealed that this difference in gains did not reach conventional criteria for statistical significance [$F = 3.2$, $p = .075$, gains between classes within groups was again $F < 1$]. However, such a gain seems highly unlikely to have occurred by chance given the corresponding significant difference in gains in proximal understanding. Thus, the students in the PR condition developed significantly greater understandings of the concepts, topics and processes associated with solving scientific and socio-scientific problems involving water quality in the performance assessment.

4. Discussion

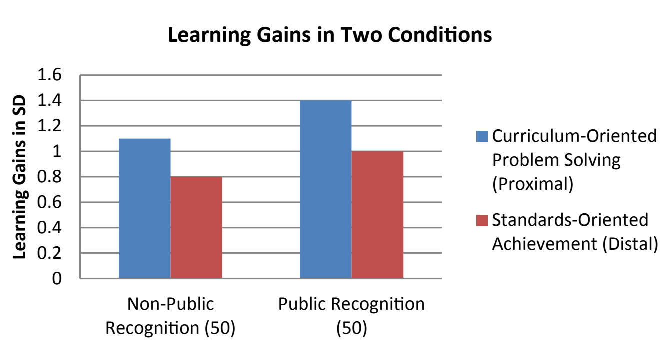
In this study we examined whether or not the introduction of external rewards had negative effects on students' motivation and whether or not they had positive effects on students' disciplinary engagement and learning. As hypothesized, the introduction of the external rewards did not undermine students' motivation (Hypothesis 1a) and interest (Hypothesis 1b). Contrary to our expectations, the external rewards did not show an effect on students' disciplinary engagement (Hypotheses 2a and 2b). In terms of learning, the positive effects of the rewards were only partially confirmed. As hypothesized, students in the PR condition show a deeper understanding of the scientific concepts underlying the narrative in Taiga (Hypothesis 3b). However, the impact of the rewards on the quality of the quest drafted (Hypothesis 3a) and on the achievement test (Hypothesis 3c) could not be established.

4.1. Rewards and motivation

Our findings concerning Hypotheses 1a and 1b indicate that the external rewards did not have negative effects on students' motivation while learning and that future interest was not undermined by the introduction of the badges in Taiga. This suggests that the negative consequences of external rewards are less likely to occur in learning environments such as Taiga. This is consistent with the argument advanced by Collins et al. (1989) that the negative consequences of competition may be more indicative of impoverished learning environments and lack of feedback and opportunity to improve, than of any fundamental consequence of competition. These findings are also consistent with the notion of the informational value of external rewards (Deci et al., 1999). In Taiga, the narrative and the way in which the feedback and the badges were given, emphasized much more the informational value focused on the problem and the narrative, highlighting the several opportunities to improve, and deploying better the role of "field investigators". Arguably, this could have more likely led to perceptions of competence, than to any controlling feature of the badge system as implemented in the present study. In other words, individuals may have made more internal attributions for doing the activities within Taiga (Lepper et al., 1973). In this sense, the study provides some initial empirical support for the speculations about sociocultural theories of engagement in Hickey (2003), and empirically matches with latest studies of the effects of badges on motivation (e.g., Abramovich et al., 2013; Landers & Callan, 2011). Finally, these findings confirm the suggestions of Cameron and Pierce (2002) as to the frequent conditions that lead to the negative outcomes of external rewards. In our case, Taiga and the badges as implemented here fulfilled all the requirements suggested by Cameron and Pierce. In other words, we implemented our system of badges for a task that we assumed was not highly interesting for all students (i.e., drafting a short essay or "Quest"), used intangible, unexpected at first, rewards, which were awarded in relation to the students' performance in drafting the quest. A future study could examine the extent at which the narrative of Taiga contributes in protecting students' motivation. This could be achieved by considering a mixed method approach with questionnaires and interviews exploring Taiga's specific design features. Another issue, that brings together the cognitive and sociocultural approach, is to determine to what extent students make attributions related to themselves (i.e., internal) or related to the narrative of Taiga (i.e., identity and role taking). Finally, it would be important to ask the extent at which students perceived their activities in Taiga as meaningful (Hickey & Zuiker, 2005) and the role that external rewards may play in such perceptions.

4.2. Rewards and engagement

Given the above findings, we had also expected to find a deeper disciplinary engagement (Hypotheses 2a and 2b) with the learning activities (i.e., drafting/revising Quest 2 and using the resources embedded in the game) in the PR condition. However, the study did not find statistically significant differences between the two conditions. For both, PR and NPR, the number of feedback pages accessed while drafting



Note. * = p<.05; *** = p<.01.

Fig. 4. Proximal and Distal Learning Outcomes by Condition. Note. * = $p < .05$; *** = $p < .01$.

Quest 2 and the accurate use of relevant scientific concepts were not statistically different. This suggests that our implementation of the external rewards was not enough for fostering a more appropriate disciplinary engagement. One explanation may be that students in the PR condition did actually used the resources more wisely by selecting more appropriately the feedback screens embedded in the game, and by doing so they actually did acquire the curricular information as intended – which could explain the difference in understanding discussed later. However, technical reasons precluded us from carrying out deeper analysis of the log files to search for patterns of access to the pages related to the content knowledge embedded in Taiga. Second, although Taiga provided opportunities to improve, it is known that to take such opportunities requires having set a particular learning goal. For example, the playful elements of Taiga may have represented a distractor from the more academic related activities – the contradictory "play" versus "learn" modes inherent in educational games (e.g., Filsecker & Kerres, in press). More broadly, it may well be that Taiga and its interactive narrative, which is designed to provide students with valued standards of practice, might have not been enough to help students set valuable learning goals that would keep them in a "learning mode" (e.g., try to understand what Turbidity means and where it comes from so that one can help the community in Taiga and save the fish as the narrative suggests). Future research should consider the use of data mining techniques (e.g., Romero, Ventura, Pechenizky, & Baker, 2010) to better understand the trajectories of students in the game and explore how they relate to students self-reported goals and purposes and to outcomes such as learning. In this way we could have a better understanding of what students actually do in such environments and, more importantly, why they do what they do. A phenomenological approach seems central also to understand how students perceived their role in Taiga (e.g., "field investigator") and how students interpret the different activities they are supposed to carry out in Taiga. Finally, a future study should examine what factors might external rewards effective for supporting students' disciplinary engagement. As mentioned at the beginning of this article, motivation and engagement are two different constructs and therefore the recommendations of Cameron et al. (2005) about introducing external rewards with minimum negative unintended effects might not applied to fostering disciplinary engagement, hence the need to address this issue in future studies.

4.3. Rewards and learning

As for the effect on learning, Hypothesis 3b suggests that students in the PR condition had a deeper understanding of the scientific inquiry and its relation to broader social issues (e.g., dying fish in a national park). They also showed deeper understanding of the relationships between water quality indicators and key ecological processes such as erosion and eutrophication and how human activities can activate such processes. This finding indicates that the use of external rewards on the context of such technology-enhanced environments does have a positive effect on learning, without, as mentioned above, the negative consequences for motivation predicted by Cognitive Evaluation Theory and by the "overjustification" hypothesis. These theories proposed mechanisms to explain how detrimental external rewards can be for motivation, but not mechanisms of how external rewards can lead to positive learning outcomes such as the ones obtained in this study. Given that we expected motivation to be the same between the two conditions, our proposed mechanism for the effect of external rewards focused on the role of students' deeper disciplinary engagement. However, the finding of no effects on students' disciplinary engagement (Hypotheses 2a and 2b) indicates that such engagement did not apparently play a role in the PR students' higher level of understanding of the curricular content embedded in Taiga. As mentioned above, one reason may be related to the differential quality of students access to the information embedded in Taiga that was not captured by our present analysis of the log files. It is also possible that this pattern of results may be related to the fact that our measures of engagement focused only in one central activity: the drafting, revision and resubmission of Quest 2, and in doing so, it might have lost other aspects of students' behavior such as teacher–students, students–students, and students–technology interactions. Concerning the measure of quality of Quest 2, it may well be that our analysis (i.e., counting the appropriate use of a concept) did not capture more sophisticated, deeper or complex knowledge representation, that in turn could have shown a difference between the two conditions. This may be due to the intensive writing that Quest 2 requires. From the literature on writing-to-learn, it has been acknowledged the difficulties and complexities of the writing process (Bereiter & Scardamalia, 1987). Therefore, it seems plausible that even though students might have acquired the knowledge and understanding concerning water quality indicators and their relationship to the ecological processes of erosion and eutrophication, they were more prepared to express such understanding in the context of a performance assessment than in the writing intensive Quest 2. Consequently, future research should consider expanding the measures of engagement across quests and examining other strategies to explore students' understanding, for example, through the development of conceptual maps. It also seem valuable to pursue studies aiming at designing supports or scaffolds that help students draft their quest. Finally, it may well be that other variables not considered in this study had also played a role. For example, exploring how external rewards in Taiga can lead to a stronger sense of Flow (see Pavlas, 2010) while drafting, revising, and consulting information during the quests in Taiga. Addressing these questions can help understand how motivation affects cognitive processing and knowledge acquisition (e.g., Renkl, 1997).

4.4. Limitations

Although so far the limitations of the present study may have become evident for the reader, we would like to summarize the most relevant ones. First, the information in the log files was analyzed without incorporating the newer data mining techniques, so that more nuanced analyses could have been conducted that would have helped us better understand the pattern of results obtained in this study. Second, the focus of our measuring effort to capture students' engagement may have missed important aspects of the broader social and material context in which students interacted and learned. Third, a more ambitious analysis of the students products (i.e., quest 2 essays) in terms of understanding or knowledge representation would have been more revealing, in particular in terms of the pattern of results obtained in this study. Fourth, some relevant individual differences were not controlled. For example, any measure of self-regulation is important given that technology-enhanced learning environments such as Taiga are very demanding in terms of self-regulated skills (Fredricks et al., 2004). Fourth, as we suggested that Taiga protected students' motivation by providing opportunities to improve and massive amounts of feedback, a detailed exploration of the in-game feedback exchange between students and the teacher would have provided with additional evidence to support our assumption. Finally, in relation to the design of Taiga, although it does a better job than many other educational games, it also suffers from the common limitation of such learning environments: its relative simplicity and limited trajectories and choices available to students. An approach to overcome this problem is the gradual incorporation of the game design patterns (Björk & Holopainen, 2005) in the repertoire of educational game designers. How to go about doing this is an open issue, however some suggestions have been already proposed (Filsecker & Kerres, 2013).

4.5. Conclusion

All in all, our study represents a pioneer effort and an important contribution to the field of educational games. First, we have empirically shown how badges in the context of a technology-based innovation in an elementary school can enhance learning without undermining motivation. Second, in contrast to previous research, by collecting data from different levels, we could clearly operationalize the concepts of motivation, engagement and learning. In doing so, we were able to shed more light into the traditional "black box" approach to research on games for learning (Honey & Hilton, 2011). Third, this fundamental distinction may well inform educational game designers about the trade-offs when designing for supporting motivation or engagement (see Filsecker & Kerres, in press). This seems like a promising way around the obvious dilemma facing many motivational interventions: program that focus directly on changing behavior may deliver behavioral change, but fail to impact cognition, while programs that focus directly on cognition may indeed impact cognition but fail to deliver enduring changes in behavior. We hope that future researchers in the field may find the distinction theoretically sound and practically useful for understanding the central processes in educational games, namely, motivation, engagement and learning.

References

Abramovich, S., Schunn, C., & Higashi, R. (2013). Are badges useful in education?: it depends upon the type of badge and expertise of learner. *Educational Technology Research and Development, 61*, 217–232.

Annetta, L. A., Minogue, J., Holmes, S. Y., & Cheng, M. T. (2009). Investigating the impact of video games on high school students' engagement and learning about genetics. *Computers & Education, 53*(1), 74–85.

Barab, S. A., & Dede, C. (2007). Games and immersive participatory simulations for science education: an emerging type of curricula. *Journal of Science Education and Technology, 16*(1), 1–3. http://dx.doi.org/10.1007/s10956-007-9043-9.

Barab, S. A., Gresalfi, M., & Ingram-Goble, A. (2010). Transformational play using games to position person, content, and context. *Educational Researcher, 39*(7), 525–536.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Erlbaum.

Bereiter, C., & Scardamalia, M. (1989). Intentional learning as a goal of instruction. In L. Resnick (Ed.), *Cognition and instruction: Issues and agendas* (pp. 361–379). Hillsdale, NJ: Lawrence Erlbaum Associates.

Björk, S., & Holopainen, J. (2005). *Patterns in game design*. Boston, MA: Charles River Media.

Cameron, J., & Pierce, W. D. (2002). *Rewards and intrinsic motivation: Resolving the controversy*. Westport, CT: Bergin & Garvey.

Cameron, J., Pierce, W. D., Banko, K., & Gear, A. (2005). Achievement-based rewards and intrinsic motivation: a test of cognitive mediators. *Journal of Educational Psychology, 97*(4), 641–655.

Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: a practical guide. *Journal of the Learning Sciences, 6*(3), 271–315.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453–494). Hillsdale, NJ: Erlbaum.

Connolly, T., Boyle, E., MacArthur, E., Hainey, T., & Boyle, J. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education, 59*, 661–686. http://dx.doi.org/10.1016/j.compedu.2012.03.004.

Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology, 88*(4), 715–730. http://dx.doi.org/10.1037/0022-0663.88.4.715.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125*(6), 627–668.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th international academic MindTrek conference* (pp. 9–15).

Dewey, J. (1913). *Interest and effort in education*. Boston, MA: Riverside Press.

Dickey, M. D. (2005). Engaging by design: how engagement strategies in popular computer and video games can inform instructional design. *Education Training Research and Development, 53*(2), 67–83. http://dx.doi.org/10.1007/BF02504866.

diSessa, A. A. (2000). *Changing minds: Computers, learning and literacy*. Cambridge, MA: MIT Press.

Domínguez, A., Saenz-de-Navarrete, J., de-Marcos, L., Fernández-Sanz, L., Pagés, C., & Martínez-Herráiz, J. (2013). Gamifying learning experiences: practical implications and outcomes. *Computers & Education, 63*, 380–392.

Egenfeldt-Nielsen, S. (2005). *Beyond edutainment: Exploring the educational potential of computer games*. Unpublished doctoral dissertation. Copenhagen: IT-University Copenhagen.

Elliot, A. J., & Harackiewicz, J. M. (1994). Goal setting, achievement orientation, and intrinsic motivation: a mediational analysis. *Journal of Personality and Social Psychology, 66*(5), 968–980. http://dx.doi.org/10.1037/0022-3514.66.5.968.

Engel, R. A., & Conant, F. (2002). Guiding principles for fostering productive disciplinary engagement: explaining an emergent argument in a community of learners classroom. *Cognition and Instruction, 20*(4), 399–483.

Federation of American Scientists. (2006). *Summit on educational games: Harnessing the power of video games for learning*. Washington, DC: Author. Retrieved from http://www.fas.org/programs/ltp/policy_and_publications/summit/Summit%20on%20Educational%20Games.pdf.

Filsecker, M., & Kerres, M. Engagement as a volitional construct. A conceptual framework for theory, research, and instructional design of educational games. *Simulation & Gaming*, in press.

Filsecker, M., & Kerres, M. (2012). Repositioning formative assessment from an educational assessment perspective: a response to Dunn & Mulvenon (2009). *Practical Assessment, Research & Evaluation, 17*(16). Retrieved from http://pareonline.net/getvn.asp?v=17&n=16.

Filsecker, M., & Kerres, M. (2013). Designing and studying educational games: limitations of current design and research approaches in game-based learning. In B. Bigl, & S. Stoppe (Eds.), *Playing with virtuality, Theories and methods of computer game studies* (pp. 349–368). Frankfurt: Peter Lang.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59–109.

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: a research and practice model. *Simulation & Gaming, 33*(4), 441–467. http://dx.doi.org/10.1177/1046878102238607.

Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. New York: Palgrave Macmillan.

Greenblat, C. S. (1981). Teaching with simulation games: a review of claims and evidence. In C. S. Greenblat, & R. D. Duke (Eds.), *Principles and practices of gaming-simulation* (pp. 139–153). Beverly Hills, CA: Sage Publications.

Greenfield, P. M. (1984). *Mind and media: The effects of television, video games, and computers*. Cambridge, MA: Harvard University Press.

Harackiewicz, J., & Manderlink, G. (1984). A process analysis of the effects of performance-contigent rewards on intrinsic motivation. *Journal of Experimental Social Psychology, 20*(6), 531–551.

Hays, R. T. (2005). *The effectiveness of instructional games: A literature review and discussion*. Orlando, FL: Naval Air Warfare Center Training Systems Division.

Hickey, D. T. (1997). Motivation and contemporary socio-constructivist instructional perspectives. *Educational Psychologist, 32*(3), 175–193.

Hickey, D. T. (2003). Engaged participation vs. marginal non-participation: a stridently sociocultural model of achievement motivation. *Elementary School Journal, 103*(4), 401–429.

Hickey, D. T., & Anderson, K. (2007). Situative approaches to student assessment: contextualizing evidence to transform practice. In P. Moss (Ed.), *Evidence and decision making, the 106th yearbook of the national society for the study of education (Pt. I)* (pp. 264–287). Chicago: National Society for the Study of Education.

Hickey, D. T., Ingram-Goble, A., & Jameson, E. M. (2009). Designing assessments and assessing designs in virtual educational environments. *Journal of Science Education and Technology, 18*(2), 187–208.
Hickey, D. T., & Rehak, A. (2013). Wikifolios and participatory assessment for engagement, understanding, and achievement in online courses. *Journal of Educational Media and Hypermedia, 22*(4), 407–441.
Hickey, D. T., & Zuiker, S. J. (2005). Engaged participation: a sociocultural model of motivation with implications for educational assessment. *Educational Assessment, 10*(3), 277–305.
Hickey, D. T., Zuiker, S. J., Taasobshirazi, G., Schafer, N. J., & Michael, M. A. (2006). Three is the magic number: a design-based framework for balancing formative and summative functions of assessment. *Studies in Educational Evaluation, 32*, 180–201.
Hoffmann, K., Huff, J., Patterson, A., & Nietfeld, J. (2009). Elementary teachers' use and perception of rewards in the classroom. *Teaching and Teacher Education, 25*, 843–849.
Honey, M. A., & Hilton, M. (2011). *Learning science through computer games and simulations*. Washington, DC: National Academies Press.
Kebritchi, M., Hirumi, A., & Bai, H. (2010). The effects of modern mathematics computer games on mathematics achievement and class motivation. *Computers & Education, 55*(2), 427–443. http://dx.doi.org/10.1016/j.compedu.2010.02.007.
Kelly, G. J., Drucker, S., & Chen, K. (1998). Students' reasoning about electricity: combining performance assessment with argumentation analysis. *International Journal of Science Education, 20*(7), 849–871.
Kirriemuir, J., & McFarlane, A. (2004). *Literature review in games and learning* (Report No. 8). Bristol: Nesta Futurelab.
Kuhl, J. (1987). Action control: the maintenance of motivational states. In F. Halisch, & J. Kuhl (Eds.), *Motivation, intention, and volition* (pp. 279–307). Berlin: Springer-Verlag.
Landers, R. N., & Callan, R. C. (2011). Casual social games as serious games: the psychology of gamification in undergraduate education and employee training. In M. Ma, A. Oikonomou, & L. C. Jain (Eds.), *Serious games and edutainment applications* (pp. 399–424). Surrey, UK: Springer.
Lee, J. J., & Hammer, J. (2011). Gamification in education: what, how, why bother? Definitions and uses. *Exchange Organizational Behavior Teaching Journal, 15*(2), 1–5.
Lepper, M. R., Greene, D., & Nisbett, R. E. (1973). Undermining children's intrinsic interest with extrinsic rewards: a test of the "overjustification" hypothesis. *Journal of Personality and Social Psychology, 28*(1), 129–137.
Lepper, M. R., & Malone, T. W. (1987). Intrinsic motivation and instructional effectiveness in computer-based education. *Aptitude, Learning, and Instruction, 3*, 255–286.
Loftus, G. R., & Loftus, E. F. (1983). *Mind at play: The psychology of video games*. New York: Basic Books.
Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science, 5*(4), 333–369. http://dx.doi.org/10.1016/S0364-0213(81)80017-1.
Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Education Researcher, 32*(2), 13–23. http://dx.doi.org/10.3102/0013189X023002013.
Muntean, C. I. (2011). *Raising engagement in e-learning through gamification*. Paper presented at the 6th International Conference on Virtual Learning ICVL, Romania.
Nelson, B. (2007). Exploring the use of individualized, reflective guidance in an educational multi-user virtual environment. *Journal of Science Education, 16*(1), 83–97. http://dx.doi.org/10.1007/s10956-006-9039-x.
Pavlas, D. (2010). *A model of flow and play in game-based learning: The impact of game characteristics, player traits, and player states*. Unpublished doctoral dissertation. Florida: Florida University.
Pierce, W. D., Cameron, J., Banko, K. M., & So, S. (2003). Positive effects of rewards and performance standards on intrinsic motivation. *Psychological Record, 53*, 561–579.
Randel, J. M., Morris, B. A., Wetzel, C. D., & Whitehill, B. V. (1992). The effectiveness of games for educational purposes: a review of recent research. *Simulation & Gaming, 23*(3), 261–276. http://dx.doi.org/10.1177/1046878192233001.
Renkl, A. (1997, April). *Intrinsic motivation, self-explanations, and transfer*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.
Rieber, L. P. (1996). Seriously considering play: designing interactive learning environments based on the blending of microworlds, simulations, and games. *Educational Technology Research and Development, 44*, 43–58. http://dx.doi.org/10.1007/BF02300540.
Romero, C., Ventura, S., Pechenizky, M., & Baker, R. (2010). *Handbook of educational data mining*. In *Data mining and knowledge discovery series*. Boca Raton, FL: Chapman and Hall/CRC Press.
Ruiz-Primo, M. A., Shavelson, R. J., Hamilton, L., & Klein, S. (2002). On the evaluation of systemic science education reform: searching for instructional sensitivity. *Journal of Research in Science Teaching, 39*(5), 369–393.
Sadler, T. D. (2004). Informal reasoning regarding socioscientific issues: a critical review of research. *Journal of Research in Science Teaching, 41*, 513–536.
Salomon, G., & Globerson, T. (1987). Skill is not enough: the role of mindfulness in learning and transfer. *International Journal of Educational Research, 11*, 623–637.
Simões, J., Díaz, R., & Fernández, A. (2013). A social gamification framework for a K-6 learning platform. *Computers in Human Behavior, 29*, 345–353.
Tang, S. H., & Hall, V. C. (1995). The overjustification effect: a meta-analysis. *Applied Cognitive Psychology, 9*(5), 365–404.
Tobias, S., & Fletcher, J. D. (Eds.). (2011). *Computer games and instruction*. Charlotte, NC: Information Age Publishing.
Wilcox, R. R. (2012). *Modern statistics for the social and behavioral sciences: A practical introduction*. Boca Raton, FL: CRC Press.