Computers & Education 83 (2015) 57–63

Contents lists available at ScienceDirect

Computers & Education

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

Gamification in assessment: Do points affect test performance?

Yigal Attali [a, *], Meirav Arieli-Attali [b, 1]

[a] Educational Testing Service, Rosedale Rd. MS-10-R, Princeton, NJ 08541, United States
[b] Educational Testing Service, Rosedale Rd. MS-16-R, Princeton, NJ 08541, United States

ARTICLE INFO

Article history:
Received 24 October 2014
Received in revised form
16 December 2014
Accepted 18 December 2014
Available online 27 December 2014

Keywords:
Gamification
Assessment
Performance
Engagement

ABSTRACT

Gamification, applying game mechanics to nongame contexts, has recently become a hot topic across a wide range of industries, and has been presented as a potential disruptive force in education. It is based on the premise that it can promote motivation and engagement and thus contribute to the learning process. However, research examining this assumption is scarce. In a set of studies we examined the effects of points, a basic element of gamification, on performance in a computerized assessment of mastery and fluency of basic mathematics concepts. The first study, with adult participants, found no effect of the point manipulation on accuracy of responses, although the speed of responses increased. In a second study, with 6–8 grade middle school participants, we found the same results for the two aspects of performance. In addition, middle school participants' reactions to the test revealed higher likeability ratings for the test under the points condition, but only in the first of the two sessions, and perceived effort during the test was higher in the points condition, but only for eighth grade students.

1. Introduction

Increasingly, web-based and mobile applications look to gamification, the use of game design elements (e.g., points, leaderboards, and badges) in nongame contexts to promote user engagement (Deterding, Sicart, Nacke, O'Hara, & Dixon, 2011). Gamification relies on the argument that many traditional activities (including school activities and traditional learning) are not inherently interesting, that games, especially computer-games, are “fun,” and therefore introducing game-like features into these otherwise dull activities would make them more attractive (McGonigal, 2011; Zichermann & Linder, 2010).

A typical example of embedding educational activities within a game-like environment is the DimensionU™ series of math games. In one of these games (TowerStorm), a virtual character (avatar) immersed within a 3-D environment “retrieves” a multiple-choice question from the “fountain of knowledge,” answers the question to collect a colored ball (each answer option has a different color), runs to a tower that generates stacked colored rings, and shoots the ball at a ring on the tower. Feedback about the correctness of the answer is provided when the ball hits the tower. If the answer was correct, the character will earn points.

Although it is intuitively clear that games are a strong motivating factor for students (Gee, 2003; Shaffer, 2006), there remains a dearth of research in which the effectiveness of the gaming environment has been directly compared with a traditional computerized application (Jackson & McNamara, 2013). For example, Kebritchi, Hirumi, and Bai (2010) investigated the effects of using the DimensionU™ series of mathematics games on high school students by assigning a treatment group to play the games for 30 min each week for 18 weeks. The treatment group showed higher gains in scores on a standardized mathematics achievement test. Unfortunately, the control group was not assigned to perform the same educational activities in a nongame environment, making it difficult to disentangle the gaming effect from the practice effect (Ericsson, Krampe, & Tesch-Römer, 1993).

In contrast, Papastergiou's (2009) control (nongame-like) group of high school students was exposed to the same content (2 h of instruction on computer memory) as the treatment (game-like) group. Nevertheless, the treatment group showed higher gains in performance from a pretest to the same posttest measuring knowledge of the content covered in instruction. In another recent study, Jackson,

* Corresponding author. Tel.: +1 609 734 1747.
E-mail addresses: yattali@ets.org (Y. Attali), mattali@ets.org (M. Arieli-Attali).
[1] Tel.: +1 609 734 5256.

http://dx.doi.org/10.1016/j.compedu.2014.12.012
0360-1315/

Dempsey, and McNamara (2012) compared a traditional tutoring system environment for reading strategy training with a counterpart game-based system which used the same task. Results indicated that performance of participants (college students) in the game-like condition was significantly lower than in the traditional environment, although participants in the game-like environment rated it as more engaging.

The purpose of this study was to test the effects of one particular game design element, by experimentally manipulating whether or not students were accumulating points while they were completing a mathematics assessment.

From a theoretical perspective, points provide feedback to the student. Providing feedback regarding task performance is one of the most frequently applied psychological interventions (Kluger & DeNisi, 1996). To have a positive effect on learning, feedback needs to provide information related to the task or process of learning (Sadler, 1989). It can do so through a number of different cognitive processes, including restructuring understandings, confirming to students that they are correct or incorrect, and/or indicating alternative strategies to understand particular information. Alternatively, it can operate through affective processes, such as increased effort, motivation, or engagement (Hattie & Timperley, 2007). However, despite a huge literature (Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Mory, 2004), the specific mechanisms relating feedback to performance are still not well understood. Historical reviews and meta-analyses on the subject describe the findings as "inconsistent," "contradictory," and "highly variable" (Azevedo & Bernard, 1995; Kluger & DeNisi, 1996). Moreover, in a comprehensive meta-analysis, Kluger and DeNisi found that although feedback interventions improve performance on average, they reduce performance in more than one third of the cases.

In and of themselves, points and other game-like elements provide information about success in the task. However, the gamification argument emphasizes the motivational aspect of game design elements over the possible cognitive or informational aspects. There are a number of motivational theoretical constructs that have been shown to mediate the effect of feedback and could be relevant to the points manipulation. In particular, the concept of locus of attention is of interest in this respect (Butler, 1987). Properties of feedback can direct attention to the self or to the task, and attention to self has been shown to attenuate or even reverse the effects of feedback (Butler, 1987) because it interferes with task performance. It is not clear whether points would be perceived by students as information about task performance (and then attention is more likely to be directed to the task) or as a form of prize for good performance (and then attention may be directed to the self).

The difference between intrinsic and extrinsic motivators is another theoretical distinction that may shed light on the effect of game-like features. In other words, points and badges can be seen as extrinsic rewards for performing the task. However, although early research has demonstrated the power of extrinsic rewards in controlling behavior (Skinner, 1953), later research has shown that for intrinsically motivated activities, tangible rewards may undermine this intrinsic motivation (Deci, Koestner, & Ryan, 1999). That is, rewarding students for performing well in an educational setting may be counterproductive in the long run. However, even in the short run, research on changing incentives (either by increasing personal stakes or providing external rewards) as a way to increase student performance is inconclusive.

On the one hand, past research supports the assumption that tests with no personal consequences, that is, low-stakes tests, are associated with a decrease in motivation and performance. Wise and DeMars (2005) reviewed 12 studies that included 25 comparisons between the performance of motivated and less-motivated groups of test takers. In most comparisons, the performance of the motivated group was significantly higher than that of the less-motivated group, and the average effect size was around .6. The manipulations used to motivate test takers varied. In most cases, the students were told the scores would count towards course or school grades. In a recent study, Liu, Bridgeman, and Adler (2012) found that increasing the personal stakes of a test by telling test takers that their scores could be released to faculty in their college or potential employers significantly increased test performance, with an effect size of .41.

On the other hand, studies that used monetary incentives to motivate students in tests have generally found weak effects on performance. O'Neil, Sugrue, and Baker (1996) offered 8th and 12th grade students taking the National Assessment of Educational Progress (NAEP) 1 dollar for each item they answered correctly, and their performance was compared to control groups that received standard NAEP instructions. A significant effect was found only for a subsample of 8th graders (those who correctly remembered their treatment condition), and only on the easy items. More recently, this study was replicated (O'Neil, Abedi, Miyoshi, & Mastergeorge, 2005) with 10 dollars per item, but again no effect on scores was found. Baumert and Demmrich (2001) offered to one group of students a flat payment if they answered more items than expected based on their school grade, but failed to find an effect on scores. Finally, Braun, Kirsch, and Yamamoto (2011) administered a NAEP assessment to 12th grade students and found weak effects (effect size of .08–.25) for monetary incentives (either a fixed amount of 20 dollars or variable according to performance) compared to a control group.

Taking into account the scarcity of research on the effectiveness of gamification and the inconclusive results in related lines of research, in the studies described below we sought to evaluate the effects of one particular gamification feature, points, on different aspects of performance in the context of an educational assessment. The assessment that was used in the two studies focuses on mastery and fluency of basic mathematical concepts and is itself part of the *CBAL*™ (Cognitively Based Assessment *of*, *for*, and *as* Learning) research initiative[2] to develop assessments that maximize the potential for positive effects on teaching and learning (Arieli-Attali & Cayton-Hodges, 2014; Bennett, 2011). The two studies reported in this paper were part of the development and piloting of this assessment. At the heart of the mathematics curriculum in the early years of elementary and middle school are concepts of the number system and operations with numbers. Research has shown that proficiency with numbers and numerical operations is an important foundation for further education in mathematics (National Mathematics Advisory Panel, 2008). Constraints on cognitive capacity provide one explanation for this connection. The more automatically a procedure is executed, the less mental effort is required, and this enables complex tasks to be carried out more efficiently (Case, 1985). Inefficient estimation and mental computations often lead to declarative and procedural errors, which carry on and impede subsequent problem solving (Cumming & Elkins, 1999). There are also important conceptual continuities between whole and rational number concepts and concepts of algebra (Empson, Levi, & Carpenter, 2011).

Since performance on an assessment of basic mathematical concepts could be determined by both accuracy and speed of the response, the instructions and point systems that were used in these studies reflected these two aspects with different emphases. All participants in

[2] See also CBAL website http://www.ets.org/research/topics/cbal/initiative.