How Points and Theme Affect Performance and Experience in a Gamified Cognitive Task

Katelyn Wiley, Sarah Vedress, Regan L. Mandryk
Department of Computer Science, University of Saskatchewan
Saskatoon, Canada
{firstname.lastname@usask.ca}

ABSTRACT
Cognitive tasks are increasingly being gamified in an attempt to leverage the motivational power of games; however, they are sensitive to manipulation and literature is divided on how adding game elements affects participant performance and experience. We applied two popular gamification approaches (points/feedback and theme/narrative) to a typical cognitive task (the dot probe) and measured performance and experience in two studies (N1=287, N2=321). Similar to prior work, we confirm in Study1 that points increase reaction time and error rate, and positive affect. We replicated these results in Study2, and expanded our analysis to investigate participant experience. Our findings suggest that theme creates expectations of an interesting game, which gamified tasks fail to deliver, whereas points maintain enjoyment better throughout the task itself. Important for the development of gamified cognitive tasks, our findings suggest that novel approaches to gameful assessment may be better than the status quo.

Author Keywords
cognitive tasks; dot probe; games; gamification; assessment

CSS Concepts
• Human-centered computing—Empirical studies in HCI
• Applied computing—Computer games

INTRODUCTION
Cognitive tasks are widely used in science, for collecting information about population norms, for conducting research in understanding cognition, and for informing mental health assessment [69]. High-quality, high-quantity data from these tasks, comprised of information from large cohorts over time, is crucial for developments in psychology and related fields [3]. Traditionally, capturing large-scale datasets was challenging, time-consuming, and expensive; however, advancements in technology have led to digital strategies in which tasks can be deployed remotely and automatically scored and evaluated [38,71].

While it is now possible to collect large amounts of data, it can still be difficult to recruit participants due to the boring and repetitive nature of cognitive tasks. Further, recruited participants may not be fully engaged with the task, due to lack of interest. People often exert suboptimal effort on cognitive tasks [22,36], making their collected data unreliable for interpretation [31]. As well, cognitive tasks are associated with high attrition rates [40], making data difficult to collect on a large-scale or over the long-term.

In attempts to facilitate engagement with cognitive tasks and collect quality data on a large-scale, researchers have turned to games. Over 164 million Americans of all ages, genders, and ethnicities play video games [72]—an attractive sample for psychological research. This broad demographic of gamers extends participant sampling to those who generally have not been involved in research (e.g., moving beyond the convenience of undergraduate student samples). Further, games are highly motivational to play and have been shown to satisfy human needs [58]. This motivational power of games has been leveraged in many contexts [29], such as in education [28] and health [37], giving rise to the phenomenon of gamification, in which game elements are used in non-game contexts [23]. Making cognitive tasks more interesting through gamification holds the potential to increase attention in the moment (leading to better data) and facilitate the long-term engagement of many people (leading to more data) [45].

However, cognitive tasks are sensitive to manipulation; even basic tasks are heavily studied to understand the effects of making small changes [54]. With such sensitive tasks, gamification could affect performance and experience in unexpected ways. Though many tasks have been gamified [39], few studies isolate individual game elements to understand their effects on data quality; rather, tasks are often gamified by applying several elements at once. Further, studies that do isolate individual game elements have shown mixed results [39]; there is little agreement on how typical gamification approaches affect performance on, and engagement with, cognitive tasks.

In this paper, we investigate how two popular gamification approaches (points/feedback and theme/narrative) affect performance on, and enjoyment of, a typical cognitive task (the dot probe). We created four versions of the dot probe task—the basic task plus versions that included points and theme individually and in combination. In two studies, we
found that points enhance experience, but that theme harms enjoyment and player experience. We further explored why theme decreases enjoyment, and provide evidence suggesting that experiencing a theme may raise player expectations, on which the gamified task fails to deliver.

Our results send a cautionary message about gamifying cognitive tasks. It is not always enough to layer game elements on an existing task and expect more engagement. We recommend using reward-based game elements (e.g., points and feedback) to increase engagement with existing cognitive tasks, but also recommend developing new methods of meaningful gameful assessment. Engaging games can be created if they are based on the underlying principles of existing cognitive tasks, but perhaps not on the tasks themselves, allowing for the true potential of games to be leveraged in large-scale cognitive assessment.

BACKGROUND

We discuss theories of gamification, gamification of cognitive assessment tasks, and the effects of gamification. We also justify our choice of cognitive task and game.

Gamification

According to Deterding et al. [23], gamification is “the use of game design elements in non-game contexts”. They also suggest ‘gamefulness’, a term that refers to play structured by rules, goals, and competition, and ‘gameful design’. Gamification is then the design strategy of using game elements, whereas gameful design is the goal of designing for gamefulness. These concepts do not always go together; a task can be gamified but not lead to gameful behavior.

Nicholson [50] proposes two types of gamification: reward-based gamification, which targets extrinsic motivation (i.e., the motivation to engage in an activity for a desirable but separate benefit [57]), and meaningful gamification, which targets intrinsic motivation (i.e., the motivation to engage in an activity because it is inherently interesting [57]). Reward-based gamification is implemented by adding elements like badges, achievements, and points. In contrast meaningful gamification uses six types of game elements to target intrinsic motivation: play, exposition (i.e., narrative), choice, information, engagement, and reflection.

These suggestions for meaningful gamification are based on Self-Determination Theory, which states that competence—the experience of mastery over challenges, autonomy—the experience of volition and choice, and relatedness—the experience of connectedness and camaraderie, all contribute to build intrinsic motivation [59]. Games have been shown to offer rich environments for building mastery, autonomy and relatedness, which helps explain why playing them is intrinsically motivating [58]. Due to these motivational features, gamification has been used in a variety of contexts, such as learning (e.g., [9]), behavior change (e.g., [8,35]), persuasion to change beliefs (e.g., [27]), and cognitive assessment. Cognitive tasks for assessment have been gamified with the intention of leveraging the power of games to capture better data through increased engagement in the moment and to capture more data through improved adherence, facilitated by task enjoyment [40].

Gamification of Cognitive Tasks for Assessment

Although there is a large body of work on gamifying cognitive tasks for training and treatment (e.g., [10,26]), we focus our review on cognitive tasks for assessment, which measure cognitive functions and are not considered diagnoses. People are required to trace patterns, memorize sequences, and respond to visual cues, among other tasks. Data from these tasks are used to approximate a measure of a cognitive domain, such as memory or attention [41]. The performance of an individual is then compared to the results from a normative sample group (norms). Differences between an individual’s performance and norms may inform research or indicate mental health issues [24].

However, there are problems associated with cognitive tasks. The interpretation of cognitive task data depends on the assumption that individuals are putting forth their best effort and are focused on the task, but cognitive tasks are often repetitive and boring, and suboptimal effort on these tasks is a common problem [22,36]. Data from an individual who is not fully engaged will not represent their true ability. This discrepancy can lead to inaccurate interpretations [31]. To improve individuals’ engagement, researchers have looked to games [65,66], with Aebberhard et al. [3] stating that “leveraging gamification to repeatedly obtain behavioral samples paves the way for next-generation high-throughput psychometric toolset”. For example, the game ‘Sea Hero Quest’, designed to assess spatial navigation and develop a large normative data set [49], has been played by over 4.3 million people, collecting far more data than would be possible with traditional methods [73]. Furthermore, as identified by Lumsden et al. [39], gamified tasks may be used to increase motivation, usability for various demographics, long-term engagement, ecological validity, and suitability for different disorders.

Previous work has typically taken a cognitive task, digitized it, and then gamified it by layering points and graphics over the task. In a systematic review of cognitive task gamification, Lumsden et al. [39] identified 2D graphics, sound effects, score and theme as prominent tactics for assessment gamification (the gamification of training tasks often uses more complex elements). These gamified tasks may create the appearance of a game through the application of game elements in a non-game activity, but they do not always lead to gamefulness [23]. Further, the inclusion of game-based elements has not always shown an improvement in data quality for cognitive assessment. We next present prior research on different game elements and their effects, limiting our review to papers that explicitly investigated specific game elements for assessment-based cognitive tasks; although gamification is an often-mentioned design intention, the effects of individual game elements are not always extracted or reported.
Points and Feedback
Many gamified cognitive tasks include ‘points’, which provide feedback and rewards to a player. Feedback can also be non-point-based, but points always give feedback. Often, adding points increases enjoyment [40,41,48]. Points may lead to the experience of competence and thus intrinsic motivation [33,55], but other research echoes Nicholson’s [50] idea that reward-based gamification (i.e., points) is less effective than meaningful gamification, suggesting that points increase motivation due to extrinsic incentives [47], rather than feelings of competence or intrinsic motivation.

Adding points to a task may also lead to improved performance (e.g., faster reactions times) [39]. In some cases, this improvement in performance can be detrimental to the task’s purpose. While gamification can help increase engagement for a target demographic, it can also pose difficulties for certain populations. For example, gamified tasks can normalize the performance of individuals with ADHD, meaning that the gamified cognitive task no longer differentiates between people with and without ADHD [20].

Sound Effects
Research on the use of sound effects in gamification is limited; however, one study found that sound effects (e.g. a xylophone note) resulted in attentional capture and shortened reaction times of a visual search task. Sound effects alone did not increase enjoyment [48].

Graphics
Graphics may make a task seem more game-like, but are also used as stimuli in cognitive tasks, and changing them can affect performance. Lumsden et al. [41] used cowboy characters in a Go/No-Go task and saw performance worsen relative to a control condition in which people differentiated green and red objects, possibly because the stimuli were not as easy to discriminate. Birk et al. [11] saw performance drop in a Go/No-Go task with graphic characters, again likely because it is harder to discriminate characters than shapes. Even in a Stop-Signal task in which the stimuli were similar (coloured fruit versus coloured circles), the graphics version resulted in lower performance [40].

Graphics may increase enjoyment of a task. In the cowboy game, while the graphics were detrimental to performance, they increased enjoyment [41]. However, it is difficult to separate out the effect of graphics-based stimuli from the presence of a narrative or theme. Lumsden et al.’s cowboy game not only had cartoon characters, but also a narrative that contextualized the graphics. In another study, however, graphics without narrative actually decreased enjoyment of a task relative to a control condition [40].

It should be noted that graphics may not always harm performance. It is difficult to draw conclusions as there are only a few studies that examine the isolated effect of graphics, but there are many studies in which cognitive tasks are gamified in multiple ways, including graphics, and which find acceptable performance scores [39].

Thematic Elements
Graphics are often added to allow for premise, backstory, character, and narrative—dramatic elements in games that can be described as ‘theme’. Research on including theme shows mixed results for performance and enjoyment.

Participants did enjoy the theme of Lumsden et al.’s [41] cowboy Go/No-Go game, in which they played a sheriff rescuing hostages in a saloon. They had to ‘shoot’ criminals but inhibit response to hostages. The narrative was well-received and raised enjoyment, but as noted, the graphics worsened performance. Birk et al. [11] also found that theme increased the number of hostile words used in an ambiguous word completion task. In the same study, which added premise and backstory to cognitive tasks, participants unexpectedly reported greater enjoyment, relatedness, autonomy and immersion in control tasks compared to tasks with theme. It is unclear why these differences occurred; however, the authors speculated that the packaging of a task as a game may have raised expectations that were then unmet by a thematic version of a cognitive task.

Again, it is difficult to draw conclusions on the isolated effects produced by theme, as most studies combine theme with other game elements, such as points [39].

Taken together, the effects on performance and enjoyment of applying game elements to cognitive tasks are mixed, and are further muddied by the dearth of studies that isolate and analyze individual gamification elements, rather than confounding them into a single ‘gamified’ design.

On Choosing a Cognitive Task and Game Elements
In addition to the choice of which gamification elements to apply, research findings are complicated by the use of different cognitive tasks, such as the Go/No-Go [11,41], stop-signal task [40], or an N-back task [11]. When choosing a task, researchers generally are most interested in what the task is meant to assess. In our research, we chose to study how gamification elements, when applied in isolation, affect performance and enjoyment of a standard, well-known, and highly-studied task—the dot-probe task.

Choosing a Cognitive Task
In the dot-probe task, individuals are presented with two sets of stimuli, one relevant to an area of interest, and one neutral (see Methods for additional details). Individuals may demonstrate an attentional bias to certain stimuli [18]; for example, individuals with social anxiety respond preferentially to images of threatening faces, in that they are both faster to engage with the threatening stimuli and slower to disengage with it [5,63]. The dot probe task has been used to study individuals with anxiety, PTSD, insomnia, and other disorders [5,25,44,62], and research has validated the use of a remote dot probe task delivered via the internet [43]. As a typical cognitive task that is already digital, validated for online use and heavily studied, the dot probe task is suitable for us to explore the effects of gamification on participant engagement and data quality.
Gamification of the Dot Probe Task
The dot probe task has not previously been gamified for the purposes of assessment. However, it has been gamified for training. Attentional-Bias Modification (ABM) uses the dot probe task to train users to attend towards preferential stimuli and/or away from undesirable stimuli [6,16,30]. For example, individuals with alcohol addiction problems demonstrate an attentional bias towards alcohol-related stimuli. An ABM dot probe task can train this bias away by presenting two stimuli, one alcohol-related and one neutral stimulus, and then requiring individuals to respond to the neutral stimulus. While at first alcoholic individuals will be distracted by the alcohol-related stimuli, over time they become faster at responding to the neutral stimuli [60]. ABM tasks have also been used for other addictions [4,19], anxiety [16,42], depression [15], and more [34,61].

Dennis and O’Toole [21] created a gamified version of an ABM dot probe task, designed for use as a mobile app. They incorporated animated characters, points, sound effects, and a swiping motion into a typical dot probe task. When played for over 45 minutes, this game reduced threat bias in anxious individuals. Another gamified version of an ABM task, called ‘Shots’ was designed by Boendermaker et al. [13] to influence heavy-drinking undergraduate students. This game, while functionally identical to the control ABM dot probe task, included rewards (in-game coins and points), a graphical theme, sound effects, and animations. While participants’ motivation to train decreased in all conditions, it decreased even more in the game condition. The authors suggest this effect could be due to participants’ expectations; they expect a game to be more than just a point-system with graphics, and thus experience disappointment and decreased motivation when the game fails to be fun. As well, only the regular training condition was successful in reducing attentional bias. Boendermaker et al. [13] conclude that the game elements distracted participants, hence the lack of improvement.

Game elements can also be added outside of the actual task; for example, Birk and Mandryk [12] added avatar customization to an ABM dot probe task. Participants were either assigned an avatar, or created one themselves, and these avatars were then used as stimuli by giving them angry or neutral expressions. Avatar customization increased both engagement and training efficacy, even though the task itself was identical in both conditions.

Choosing Game Elements
For assessment-based cognitive tasks, typical gamification strategies can be grouped into two categories: reward/feedback features, and narrative/theme features. Reward/feedback features include points and rewarding feedback (e.g., visual effects, bonus points, high scores). Narrative/theme features involve cohesive graphics, premise, characters, and story. We chose to focus on these two groups of game elements due to their popularity as gamification strategies in our review of previous literature.

The Present Research
We add to previous work on gamified cognitive assessments by examining the effects of reward/feedback features and narrative/theme features on participant performance and enjoyment in a typical cognitive assessment tool—the dot probe task. We developed several versions of the dot probe task, which individually layered different game elements, and in two studies, we measured their effects on reaction time, error rates, and various measures of enjoyment and experience.

STUDY1
The goal was to understand the effects of points and theme on performance and explore effects on experience.

Methods
We manipulated the inclusion of points and feedback (which we call points) and narrative and thematic elements (which we call theme) to develop four versions of a dot probe task: 1) a basic task as a control (points: off; theme: off); 2) a task with points/feedback (points: on; theme: off); 3) a task with narrative/theme game elements (points: off; theme: on); and 4) a task with both points/feedback and narrative/theme elements (points: on; theme: on).

Tasks
See Figure 1 for images of the tasks.

Basic Control: This task was designed as a basic dot probe task. Following standard protocols, we presented a fixation point (+) in the center of the screen for 500ms, followed by two images simultaneously displayed on the left and right side of the screen, also for 500ms. After this, a probe stimulus (a large black dot) was displayed on either the left or right location, and remained until participants made a response of either the left or right arrow keys. Participants were instructed to select the arrow key that corresponded with the probe’s location.

Theme: This version of the task added narrative and theme game elements to the basic control task. Participants played as a racoon who was recruited by another racoon, ‘Bijou’ to find their family’s stolen paintings. The instructions for the task describe the story of an art thief, ‘Zabat’, who had stolen the racoon family’s art collection and hidden them in a cave. To mask his trail, Zabat mixed the authentic paintings with inauthentic ones, and the only way to distinguish them was to look for a red seal underneath the paintings. After an introduction to this story, participants then proceeded to the dot probe task, with the images as the paintings and the probe as the red seal.

Points: This version of the task added reward and feedback game elements to the basic control task. The dot probe design was the same, though participants now earned and lost points, were given a score, and could see their reaction times to each trial. Faster responses were rewarded with more points. When participants indicated the dot location, colorful confetti animated on the screen for a correct response, and they lost points for an incorrect response.
Points and Theme: Finally, this version of the task combined the Points task and the Theme task. For this task, participants were presented with the same narrative about Zabat, and also were awarded points and given feedback. When participants correctly indicated the seal location, in addition to colorful confetti, a smiling Bijou appeared (and a frowning Bijou for incorrect responses).

Stimuli
The stimuli shown in all conditions were images of faces expressing different emotions; either non-threatening stimuli (neutral or happy expressions) or threatening stimuli (angry or sad expressions). Images were shown in non-threatening/threatening pairs (i.e., non-threatening on the left side and threatening on the right, or vice versa).

We took pictures of 14 student actors (9=male, 5=female). The actors made four expressions (angry, sad, neutral, happy) after being coached on how to do so (e.g., “purse your lips and furrow your eyebrows”). We then added art filters (to match the story of our theme conditions) to each of the images, as shown in Figure 1. We conducted a pre-study to validate the images’ emotional valence when filtered, along with filtered images from the Ekman set of faces [74]. Raters scored each image using the Self-Assessment Manikin (SAM), a pictorial scale used to assess the arousal and valence of a stimulus [14]. Participants also rated each image for happiness, anger, and sadness on a 7-point Likert scale from low (1) to high (7). In general, the art filters increased the arousal of images, but maintained the same valence when compared to the unfiltered image. We eliminated images that had too great a deviation in valence from their corresponding baseline image, resulting in 14 images for the study (3 from our actors, 11 from the Ekman set of faces). We used the filtered pictures for all the conditions to maintain the consistency of the stimuli and reduce any problems with graphic elements changing the nature of the task, as seen in previous studies [41].

Procedure
We used Amazon Mechanical Turk (MTurk) to recruit participants. MTurk is commonly used in human-computer interaction research to conduct studies [52], and has been shown to yield reliable data when precautionary methods around data gathering and analysis are used (e.g., [17,46]).

The study used a between-subjects design, with each participant being shown one of the four versions. Prior to the task, participants were surveyed on demographic information and game-playing habits. Participants also completed the Positive and Negative Affect Schedule (PANAS) before and after the task, in order to explore experience. PANAS measures the two primary dimensions of mood valence by asking participants to rate their agreement with different sentences (e.g., “I feel attentive”, “I feel irritable”) on a five-point scale, which translate into a measure of positive affect and a measure of negative affect [68]. The experiment took 15 minutes to complete on average and we compensated participants with $3.00 USD.

Analysis
We received 349 responses. Participants who did not meet threshold quality criteria were removed, according to best practices [17,46]. We removed participants who completed the questionnaires too quickly (faster than 1.5 seconds per
item) (n=40), those who showed zero variance within the questionnaires (e.g., answered every question with ‘disagree’) (n=9), and finally, those who incorrectly responded to the dot probe too frequently (total incorrect responses > 1SD above the mean number of incorrect responses) (n=13). Our final sample consisted of 287 people (122 female, 161 male, 4 non-binary, mean age=37.5, SD=12.03).

We conducted two-way ANOVAs with points (on/off) and theme (on/off) as between-subjects factors on reaction time (RT) and error rate data, and two-way ANCOVAs on post-play positive and negative affect, controlling for pre-play affect. To ensure a normally distributed data set for RT (which tends to be positively skewed), we took the log value of RT, as suggested when working with reaction time data [18,70]. Our required level of significance was p<.05.

Results
See Table 1 for results of the statistical tests.

Performance
We found significant main effects of points on RT and error rates. There were no significant differences in RT or error rate as a result of including theme. There was further no interaction between points and theme on RT or error rate.

Experience
In terms of experienced negative affect, there were no significant effect of points or theme, and no significant interaction. In terms of positive affect, participants rated their mood more positively when points were included. There was no significant effect of theme and no significant interaction; see Table 1 for results of statistical tests.

Discussion of Study1
Our results show that the inclusion of points affects performance by increasing reaction time and error rates. Points also increase affect; however, this same increase is not seen for theme. These findings echo previous research on thematic elements. Typically, researchers have included theme expecting increases in enjoyment; however, in studies that isolate the effects of theme, enjoyment has not increased [11,40]. In fact, Birk et al. [11] even found that participants experienced higher immersion, enjoyment, and relatedness after playing a basic task compared to a task with thematic elements (premise, backstory, and corresponding graphics). Our results, similar to these results in the literature, seem counter-intuitive. Nicholson [50] suggests using exposition (i.e., story and theme) as part of meaningful gamification. Villani et al. [67] suggest that games can improve emotional regulation and mood, in part because narrative elements immerse the player in the game and allow them to take on another identity. Following these guidelines, we included theme elements that created a story and allowed participants to take on a new identity of a wronged raccoon, eager to recover their stolen art. We anticipated increased affect from these elements, but instead saw results that mirror what others have seen—that thematic elements do not necessarily increase experience. Prior work, and our results that suggest how theme is not helpful, prompted us to conduct Study2.

We designed Study2 to confirm that theme does not improve experience and further investigate why it is detrimental when applied.

STUDY2
The goal of Study2 was to replicate the performance findings observed in Study1; specifically, that including points decreases RT and increases error rate, and to understand why theme is detrimental to experience.

Methods
Study2 used the same design, tasks and stimuli as Study1, with additional questionnaires to measure enjoyment.

Procedure
In Study2, we followed the same procedure as in Study1, but also included player experience measures. We added an enjoyment questionnaire that was given before, during (after the introduction sequence but before play), and after the task. Comprised of five statements—“I am enjoying playing this game”; “Playing this game is fun”; “This game is interesting”; “I am thinking about how much I am enjoying playing”; “This game is not holding my attention” (reverse-coded)—participants rated their agreement on a 7-pt scale. Items were worded for each time point, e.g., “I will enjoy playing this game”, “I am enjoying playing this game”, and...
“I enjoyed this game”. After the task, they completed the Player Experience Inventory (PXI), which measures ten constructs: meaning, mastery, immersion, autonomy, curiosity, ease of control, challenge, progress feedback, audiovisual appeal, and goals and rules [1]. Finally, they were optionally asked an open-ended question: “Please explain why you did or did not enjoy this game.”

Data Analyses
We followed the same analyses as Study1. We received 391 responses. After filtering out participants to ensure data quality, our final sample consisted of 321 people (143 female, 177 male, 1 other, mean age=37.3, SD=10.98). Again, we used the log value of RT [18,70].

We conducted two-way ANOVAs on RT and error rate data, and two-way ANCOVAs on post-play positive and negative affect (controlling for pre-play affect), identical to Study1. For enjoyment ratings, we conducted a repeated-measures ANOVA with time (before, during, after) as the within-subjects factor and points and theme as the two between-subjects factors. As the PXI has 10 subscales, which share variance, we conducted a MANOVA with points and theme as between-subjects factors and the 10 subscales as dependent measures. We did not conduct a qualitative analysis of the open-ended question; however, we include some quotes in our discussion as examples.

Hypotheses and Research Questions
We made the following hypotheses and research questions:

H1. RT will be faster when points are included. Based on the results of Study1, we expected the same result.

H2. Error rate will be higher when points are included. We expected the same result as seen in Study1.

H3. When controlling for incoming affect, post-play positive affect will be higher when points are included. Based on Study1, we expected the same result.

H4. Immersion in the game will be lower when theme is included. We expected this due to previous work, which found lower immersion with thematic elements similar in style and narrative to our own [11].

H5. Enjoyment of the game will be lower when theme is included. We expected this due to the results of Study1 on affect, combined with previous work which shows lowered enjoyment with thematic elements [11,40].

RQ1. If enjoyment is lower when theme is included, why? If the drop in enjoyment is due to raising expectations with theme and not meeting them, then we should see higher ratings following the introduction of the theme, and a significant drop post-game. However, if the drop in enjoyment is due to disliking the game’s design, then we should see lower ratings following the introduction of the theme, and a smaller drop post-game.

RQ2. How is player experience changed by the inclusion of points and theme? We included the PXI to understand how the inclusion of points and theme affect the experience of the task along constructs of player experience.

Results
See Table 1 for means and standard deviations.

Performance
As expected, we found significant main effects of points on RT and error rates. There were no significant differences in RT or error rate as a result of including theme. There was further no interaction between points and theme on RT or error rate. See Table 1 for results of the statistical tests.

Experience
PANAS: In terms of experienced negative affect, there were no significant effect of points or theme, and no significant interaction. In terms of positive affect, participants rated their mood more positively when points were included. There was no significant effect of theme and no significant interaction; see Table 1.

Enjoyment: There were no main effects of points (F_{1,317}=2.155, p=.143), theme (F_{1,317}=.001, p=.975), or interaction on enjoyment (F_{1,317}=.75, p=.784). There was a main effect of time (F_{1,317}=304.87, p<.001), with interactions between points and time (F_{1,317}=19.114, p<.001) and theme and time (F_{1,317}=12.686, p<.001). The main effect of time showed decreases in enjoyment at each time point (all p<.001). The interaction of points and time show that there was no difference in enjoyment before play (p=.964) or during play (after the game introduction but before the actual task/game; p=.269), but that points increased enjoyment ratings after play had completed (p<.001). The interaction of time and theme showed that there was no difference in enjoyment before play (p=.906), but theme increases enjoyment during play (p=.005). After play, enjoyment ratings were marginally lower (p=.057) as a result of theme. There was no three-way interaction (F_{1,317}=2.232, p=.108). See Figure 2. Thus, we provide evidence for H5 and answer RQ1.

Player Experience Inventory (PXI)
See Table 1 for means and standard deviations. As the MANOVA results were significant for points (F_{10,30}=50.69, p<.001, η^2=0.62), theme (F_{10,30}=6.12, p<.001, η^2=0.17), and...
the interaction ($F_{0.308}=2.02$, $p=.031$, $\eta^2=0.06$), we report the univariate results (see Table 1).

**Meaning:** There was a significant effect of both points and theme, with points yielding higher meaning and theme yielding lower meaning. There was also a significant interaction. Pairwise comparisons revealed that the loss of meaning from adding theme is only present when points are not included ($p=0.003$). With points, the negative effect of adding theme disappears ($p=.999$); i.e., points seem to protect against the harm to meaning from adding theme.

**Immersion:** There was a significant effect of theme, in which theme lowered immersion. There were no significant effects of points or interaction. The results from this construct provide evidence to accept H4.

**Autonomy:** There was a significant effect of points, with points associated with higher autonomy. There were no significant effects of theme or interaction.

**Ease of control:** There was a significant effect of theme, with theme associated with higher ease of control. There were no significant effects of points or interaction.

**Challenge:** There was a significant effect of both points and theme, with points associated with higher challenge and theme associated with lower challenge. There was no significant interaction effect.

**Progress and feedback:** There was a significant effect of points, with points yielding higher progress and feedback. There were no significant effects of theme or interaction.

**Audiovisual appeal:** There was a significant effect of points, with points yielding higher audiovisual appeal. There were no significant effects of theme or interaction.

**Goals and rules:** There was a significant effect of both points and theme, with both points and theme yielding higher goals and rules. There was no significant interaction. There were no significant effects for Mastery or Curiosity.

**Discussion of Study2**

Like Study1, Study2 found an increase in RT and error rate with points. Theme did not affect performance. Points increased positive affect, and while enjoyment suffered across time points for all conditions, the decrease was smallest when points were included. Enjoyment was lowest when theme was used alone, even lower than the basic task. The P3 further found that theme was associated with lower meaning, immersion, and challenge, though higher ease of control and goals and rules. In contrast, points increased meaning, autonomy, challenge, progress and feedback, audiovisual appeal, and goals and rules.

The results confirmed our hypotheses. As expected, the addition of points decreased participant reaction time (H1), and increased error rates (H2), and positive affect (H3), whereas immersion (H4) and enjoyment (H5) were lower with the addition of theme. We also addressed why theme lowered enjoyment. Enjoyment ratings collected during the game (after the introduction sequence) were higher when theme was included; however, after game play, the theme-only condition had the lowest rating of all. These results suggest that the lower ratings were not because participants disliked the theme itself, but rather that the game play was dissatisfying after having been introduced to the theme. This is likely due to the heightened expectations created by the introductory exposition, and then disappointment when the game play failed to meet those expectations.

**DISCUSSION**

We summarize our results, make design recommendations, and discuss limitations and future work.

**Summary of Results**

Our objective was to evaluate how adding game elements to a dot probe task affect task performance and player experience. Gamification of cognitive tasks often conflates many game elements into a single ‘gamified’ design that is compared to a control task [39], leading to little knowledge on how single elements affect performance and experience.

Together, our studies make several important findings:

1. Points decreased reaction time and increased error rates (Study1, Study2)
2. Points increased positive affect (Study1, Study2)
3. Theme lowered enjoyment and immersion (Study2)
4. Enjoyment was actually higher after a theme-based introduction to the task, but dropped after play when the game failed to deliver on player expectations (Study2)
5. Enjoyment of the game was higher when points were included (Study2)
6. Points increased challenge and meaning, whereas theme lowered challenge and meaning (Study2)
7. Both points and theme increased the clarity of goals and rules (Study2)

**Explanation of Results**

We explain our results in regard to the effects of points and theme on performance and experience.

**Performance**

**Points:** Our analysis of reaction time data found that when points and feedback are embedded into the dot probe task, participants are faster. This increase in speed suggests that adding points may manipulate attention. In this case, participants may have been motivated to concentrate on the location of the probe so that they could score points, see animations, and feel a sense of accomplishment as they got faster. Without points, there were no incentives to pay attention. Cognitive tasks rely on engaged participants. If a participant is daydreaming, distracted, and failing to pay attention, the data does not represent true cognitive ability. Increasing motivation is one way to increase data quality.

It should be noted that while our participants were faster when points were included, they were also more error prone. Higher engagement may not be worth the loss of accuracy. A speed-accuracy trade-off is generally expected and has been found with other gamified tasks [64]. The increased RT and error rates may interfere with a task’s interpretation and ability to discriminate.

**Theme:** Theme did not affect performance in our studies. Previous research has found thematic elements detrimental to
performance when the graphics interfere with the task. Given these results, for our studies we were careful to maintain consistent stimuli across conditions, to better isolate the effect of theme (versus the effect of graphics).

**Experience**

**Points:** The inclusion of points increased positive affect in both studies. Additionally, points yielded higher meaning, autonomy, challenge, progress and feedback, audiovisual appeal, and goals and rules. These findings reflect research on the mood repairing effects of games. In-game success is linked to higher moods and need satisfaction [56,67]. While participants were equally ‘successful’ in all versions of the task, the addition of points made the task seem more challenging, and success more meaningful. Correct trials corresponded to positive feedback, experienced as higher score communicated in visual effects. Self-Determination Theory suggests that challenges and positive feedback contribute to feelings of competence, and thus help satisfy needs [55]. Further, points anecdotally led to gameful behavior [23] as some participants described trying to beat a high score: “I enjoyed it because I was able to challenge myself to get better reaction times, so I was able to immerse myself in the game and had fun doing it” [points/theme].

Even when participants acknowledged the limitations of the game, they were able to find ways to engage with the game: “There wasn’t a whole lot to keep me focused besides wanting to beat my score. Luckily I’m a sucker when it comes to silly goals” [points/no theme]

While these quotes are anecdotal, they reflect potential reasons for why RT was faster when points were included, and link the speed increase to the experience of challenge, feedback, and enjoyment.

**Theme:** We anticipated that adding theme would enhance post-play affect; however, we did not see this in Study1 and confirmed in Study2 that theme was detrimental to enjoyment ratings post-play, even though enjoyment ratings were higher after the theme-based introduction. As others [13] have suggested, this may be due to expectations. Expectancy-disconfirmation theory suggests that individuals’ satisfaction with an experience can be explained by differences between their expectations and what actually happens [53]. Digital games are supposed to be fun and are often elaborate, with incredible graphics, deep stories, and interesting mechanics. In our study, the theme elements outlined a story and introduced characters. When that depth failed to carry through to the game play, participants may have felt more disappointment and disconnection from the game than if they were never introduced to the theme. When we call a task a game but then simply add a basic story, it fails to meet player expectations, leading to disappointment. For example, one participant in the theme-only condition said, “I thought it was cute and fun at first, but then there were too many trials and it became monotonous. I would have liked more interaction during the game.”

This theory may also explain why feelings of immersion were lower when theme was included. Jennett et al. [32] describe immersion as “a lack of awareness of time, a loss of awareness of the real world, involvement and a sense of being in the task environment”. The discrepancy between participants’ expectations and the reality of the game play may have interfered with their engagement in the task.

**Implications for Design**

Our results point to several important considerations when designing gamified cognitive tasks.

**Considerations about Performance**

First, there are trade-offs with gamification: adding reward and feedback game elements increased engagement and RT. However, these elements also resulted in higher error rates.

Any speed-accuracy trade-off needs to be investigated to ensure data from the task can be interpreted correctly. In our study, the interpretation of the dot probe task relies on the difference in RT between types of stimuli (a ‘bias’ score), rather than the overall RT. In order to fully examine the effects of theme, we obscured our stimuli with art filters, so we did not expect to gather interpretable data and cannot draw conclusions about how increased RT and error rates affect interpretation. Other studies have managed these changes in performance; for example, Tong and Chignell’s [64] gamified task led to higher RT and error rates, leading them to derive an overall performance score that better correlated with cognitive abilities than assessing speed and accuracy separately. The speed-accuracy trade-off could also be manipulated in a game’s design. Miranda and Palmer [48] suggest that point schemes can be allocated based on the goals; for example, in prioritizing speed over accuracy, and thus shaping behavior.

Performance may also suffer due to game elements. Boendermaker et al. [13] suggest that the game elements in their ABM game distracted participants from the goals of the game. In our study, if we had added many more reward and feedback game elements, our results may have been too error-prone to provide quality data. Each gamified task should be validated to ensure changes in performance can be accounted for when interpreting the data.

**Considerations about Experience**

Our results suggest that it may be important to design high-quality games that meet expectations. Our simple story, cartoon characters and graphics were insufficient to increase enjoyment. Vanden Abeel et al. [2] discuss the importance of game quality. They compared an existing game designed to measure psychoacoustic thresholds in preschoolers, with a new game which was designed by dyslexia researchers and game designers. The addition of game researchers to the development team resulted in a game that children vastly preferred and enjoyed, and was able to measure lower thresholds than the original game.

Sometimes, the inherent properties of cognitive tasks may not lend the task to meaningful gamification, even for the
most skilled game designer. Current cognitive tasks rely on uninterrupted, repeated trials, in distraction-free contexts. Slight changes to a task can affect data quality. Even non-gamified versions of tasks must be carefully controlled. For example, a difference of 500ms in exposure to a stimulus can affect results [7]. Tasks this sensitive to manipulation may be difficult to modify in meaningful ways.

Tong and Chignell [64] suggest that designers should “design game components to reflect psychometric properties of existing neuropsychological tasks” by looking to validated tasks and incorporating relevant features. Due to the nature of many cognitive tasks, designing high-quality games for assessment may require moving away from glorified digital copies of already-existing tasks, and instead design games that target the underlying principles of cognitive tasks in a new way. We should still look to validated tasks, but then also beyond them to the new capabilities that games and digitization offer.

Some research has started to move in this direction. For example, while Boendermaker et al.’s [13] gamified dot probe task did not work for ABM, Notebaert et al. [51] were able to create a successful ABM game, by targeting the underlying principles of a dot probe task. Their game displayed several stimuli/faces at once, moving around the screen, with instructions to track the face with a certain emotional expression. At times, the current target face would take on a new expression, while another face simultaneously took on the target expression. Participants then had to make the switch to tracking the new face as quickly as possible. At the end of each block, participants were given various scores based on how well they did and were encouraged to beat their current high score.

Another approach is to start from existing successful games. Commercial games already have mechanisms to collect vast amounts of data on measures like RT, working memory and more. What we do not yet understand is how to interpret that data given the game context. It may also be possible to embed specific tasks, meant for assessment, within an enjoyable game. For example, many Triple-A games include embedded mini games that do not interact or interfere with the overall performance or narrative of the story (e.g., lockpicking in Skyrim, hacking puzzles in Bioshock). Some games are made almost entirely of mini games in which players complete memory, reaction time, and gambling tasks (e.g., Mario Party), which are similarly simple and repetitive as cognitive tasks. Tasks embedded as mini-games in complex high-quality games may be less likely to yield lower enjoyment through raised expectations.

Adhering strictly to copies of cognitive tasks overlaid with game elements may be useful at times, but also limits the potential of games. Based on our results and synthesis of the literature, we recommend when to consider employing the two approaches to gameful assessment, described by Nicholson [50]. Reward-based gamification: Fairly simple game elements like points seem to improve enjoyment without drastically changing the underlying task. Theme often complicates stimuli and obscures tasks, and even if it does not, it hinders enjoyment when the task fails to match players’ expectations. Meaningful gamification: In order to fully leverage the power of games for assessment, designers must explore new methods that move beyond existing tasks. Elements like theme and narrative that are associated with meaningful gamification could be useful when paired with game play that adheres to players’ expectations of a well-designed game. This kind of design likely means creating new methods of assessment that use the underlying principles of cognitive tasks in a new way. New methods require rigorous validation, but the potential for enjoyment and more meaningful assessment is higher.

Limitations and Future Work
While we used a typical cognitive task and two popular gamification techniques, there are many tasks and many game elements that may interact differently. Our research provides some general guidelines about potential effects of points and theme, but each task and game iteration will be different. As tasks are gamified, each task needs to be validated to a rigorous standard before it can be used to collect meaningful data. Similarly, because we were focused on investigating isolated game elements, we did not design our dot probe tasks to generate interpretable data (i.e., bias scores). As well, individual differences will further affect how gamified tasks work. For example, gamified elements can normalize the performance of individuals with ADHD [20]. Any dot probe game meant for collecting data would need to accurately assess attentional bias and ensure that the game elements did not differentially affect individuals. Finally, gamified tasks are not always appropriate for research. While they can assist in collecting large quantities of data, they may not be suitable for more specialized questions; for example, when looking at specific effects in the brain.

Future work should also investigate new methods of gameful assessment, such as using data from existing games, or embedding tasks into more complex games.

CONCLUSION
Using gamification has the potential to increase the quality and quantity of data that can be collected from cognitive tasks. However, game elements cannot be used with the expectation that they will always lead to better performance and experience. Research has shown gamification to have a variety of effects on cognitive tasks, especially in regard to theme. Our study investigates why this occurs, and suggests that gamified tasks should align with players’ expectations of a game. In order to achieve this, we encourage the development of new methods for gameful assessment, which use the underlying principles of cognitive tasks while leveraging the full capabilities of digital games.

ACKNOWLEDGMENTS
We thank NSERC and SWaGUR for funding. Thanks also to Amy Fulmer and Camille Jensen for their contributions.
REFERENCES


