Game-based Experiments on Human Visual Attention

Vero Vanden Abeele[1], Pieter Tierens[1], Bob De Schutter[3], Lee De-Wit[2] and Luc Geurts[1]
[1] KU Leuven, e-Media Lab, Campus Group T, A. Vesaliusstraat 13, 3000 Leuven, Belgium
[2] KU Leuven, Experimental Psychology, Tiensestraat 102, 3000 Leuven, Belgium
{vero.vandenabeele, pieter.tierens, lee.dewit, luc.geurts}@kuleuven.be
[3] Miami University, 800 E High Street, 45056 Oxford, OH, USA
b@bobdeschutter.be

ABSTRACT
Psychological experiments on visual attention are repetitive, and can become tedious for test participants. This is not only an unpleasant experience but can be harmful for the reliability of the results as well. In this study we investigated whether a game-based version of the Posner cueing test (measuring visual attention shifts) can provide a more engaging experience, while still reproducing scientifically valid results. Therefore, we created both a game-based version incorporating the Posner cueing test and recreated a version of the classic Posner cueing test. We tested both versions with 20 participants and compared the (game) experience and intrinsic motivation of both, as well as the scientific measurement of visual attention shifting. The results show that the game version provided a reliable cueing effect. Furthermore, the game version did generally provide a more motivating and enjoyable experience for the participants. However, the game-based assessment also gave players a lower feeling of perceived competence. Overall, we conclude that gamification of experiments on visual attention tests hold promise.

Keywords
Serious games, gamification, visual attention, experiments, Posner, game experience, game-based assessment

1. INTRODUCTION
In experimental psychology there is – as the name suggests – a tendency to do a lot of experimental tests with people, in order to find out more about the underlying mechanisms of the human psyche. Due to the scientific nature of these experiments, these tests often turn out to be dull and repetitive. This not only makes it an unpleasant experience for the test subject, it can also harm the reliability of the results, because of the decrease in motivation and the withering attention during the test. This is particularly harmful in experimental tests where attention is paramount. One possible solution is to use game-based assessments, where the motivational potential of games is harnessed to keep users engaged throughout the test [6,11].

1.1 Posner cueing test
Cognitive researchers have been trying to understand the characteristics of selective attention and in particular visio-spatial attention, to be understood as *"those attentional processes that select visual stimuli based on their spatial location"*[22]. Visio-spatial attention is perhaps the most widely studied variety of attention in normal populations and neurological populations [3].

For our research we looked at one test on visio-spatial attention, the Posner cueing test or simply the Posner test [15]. With this test researchers try to measure the cueing effect, which is a manifestation of the participant's ability to perform an attentional shift. An attentional shift is made when a signal in the visual field causes the centre of focus of visual attention to move towards that signal; it is the effect of this movement of the visual attention that the Posner cueing test aims to measure.

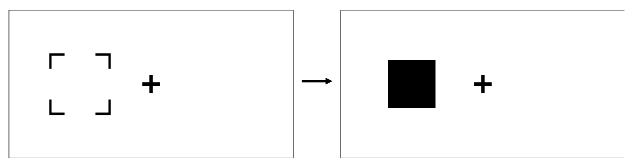


Figure 1. Valid trial in a classic Posner cueing test.

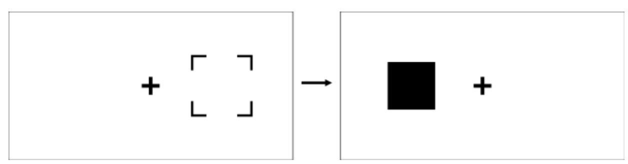


Figure 2. Invalid trial in a classic Posner cueing test.

There are several versions of the test. In the most basic version of the test, the test participant is asked to focus on the plus sign in the centre of the screen and to not move his eyes during the test (see Figure 1). At the start of a trial (of which there are typically one hundred or more in a test) a *cue* – in this case four corners of an empty square – flashes either left or right of the plus sign. After that, a *stimulus* – in this case a filled square – also flashes left or right of the plus sign on the screen. As soon as the participant detects the stimulus, s/he is asked to react as quickly as possible, e.g. by hitting a button. The time between the onset of the stimulus and the participant hitting the button is called the *reaction time* and is the main outcome measurement of the test. When the cue flashes on the same side as the stimulus we call it a valid trial (Figure 1) and when the cue and the stimulus appear on opposite sides we call it an invalid trial (Figure 2).

The cueing effect is then calculated by taking the difference between the average reaction time of all validly cued stimuli and the average of all invalidly cued stimuli. The results of such experiments show that on average, participants respond approximately 20 milliseconds faster to target stimuli in valid trials, as compared to response times without cue. Participants respond approximately 20 millisecond slower in the case of invalid trials compared to response times without a cue. Hence, (valid) spatial cueing leads to *faster response times*.

Moreover, it has been shown that (valid) spatial cueing effects lead to *lower error rates*. A slightly more extended version of the Posner test asks the participants not simply to press one button, but to hit a left button in case the stimulus was on the left side, and a right button, in case the stimulus was on the right side [15]. It has been shown that participants also make more mistakes in case of invalid cueing. Again, Posner theory explains this effect; the cue already shifts the visual attention towards that region in the visual field. In the case of a valid trial this results in a faster reaction time and less errors, in the case of invalid trials this results in slower reaction times and more errors.

Clinical relevance of the Posner test

When the effects of spatial cueing manifest differently, or when they are not present at all, this could be an indicator of mental illnesses, and as such, more recently, the Posner test has been used as a tool to support diagnoses. The Posner tests has been used to assess acute and chronic stroke [4,16], hepatic encephalopathy (decreased levels of consciousness as a result of liver failure) [2], Parkinson’s disease [12] and even psychosis [10]. It has also been suggested that children with attention deficit hyperactivity disorder have slower reaction times in both valid and invalid trials than do typically developed children [14]. Hence, while in origin a purely theoretical test, recently, research indicates clinical relevance of such tests on spatial attention [16].

1.2 Game-based assessment of the Posner test

In the past couple of years there has been a rise in attention for games to serve purposes beyond entertainment. Terms like serious games, applied gaming, gamification are quite popular, in particular in health [5] and cognition [19]. In this case we focus on game-based assessments [20]. As discussed before, psychological tests like the Posner cueing test have the tendency to be perceived as dull and repetitive. This makes it an unpleasant experience for the test subject and is potentially dangerous for the reliability of the test results. Game-based assessments could very well be of help here as it offers entertainment value to motivate test subjects in these tests [6]. In this study, we wanted to investigate whether embedding the Posner test in a game, could offer a more engaging experience for participants, but still offer correct and scientifically valid results.

Hypotheses and research questions

First of all, we expect that the game-based version of the Posner test version will generate the same results as the classic Posner test.

- ⇨ Hypothesis 1A. Participants will have faster response times when cued validly via the game-based assessment than when cued invalidly.
- ⇨ Hypothesis 1B. Participants will have lower error rates when cued validly via the game-based assessment than when cued invalidly.

Second, we expect participants to be more intrinsically motivated and to have an enjoying experience.

- ⇨ Hypothesis 2A: Participants will have a more enjoyable experience taking the game-based assessment than when taking the classic Posner test.

In addition, we hypothesize that participants will be more engaged by the game-based version than by the classic Posner cueing test. Hence, we expect that the participants will experience more flow in the game than in the classic version, i.e., that they will be captured by the game-based version, and to some extent forget what is happening around them. We also expect that the participants will feel more challenged by the game-based assessment, and find it more important to perform well. This would also increase their feeling of pressure and tension.

- ⇨ Hypothesis 2B: Participants will be more engaged when taking the game-based assessment than when taking the classic Posner test. They will also experience more pressure, more flow, more challenge and attach more importance to performing well.

Finally, the game is designed to be at least as intuitive as the classic Posner test and therefore we expect that both the Posner test and the game-based version will give the same feeling of competence and control to the participants.

- ⇨ Hypothesis 2C: Participants will experience the same amount of competence and control when taking the game-based assessment as when taking the classic Posner cueing test.

2. METHOD

We recreated both the classic Posner cueing test (as described in Figure 1 and Figure 2) as well as a novel game-based assessment that embeds the Posner cueing test. Both were created with the Unity game engine. Since it is well described how to recreate a classical test [15], in this paper we will in particular detail how we designed the game-based assessment.

2.1 Game –based version

A custom-build infinite runner game was developed in Unity 3D. The game situates itself in a space-like environment. The overall goal is to collect bananas in order to keep monkeys alive (see Figure 3). The game largely involves running on a path that is infinitely generated, on which the player has to turn left or right when another path appears, while at the same time collecting bananas. The graphics are kept simple as not to introduce possible inconsistencies. The black background in space offers high contrast and a virtually endless game world.

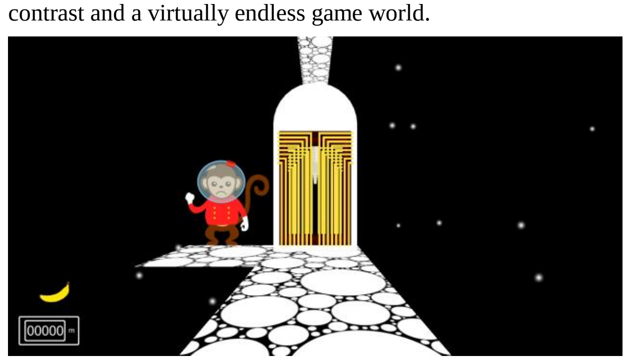
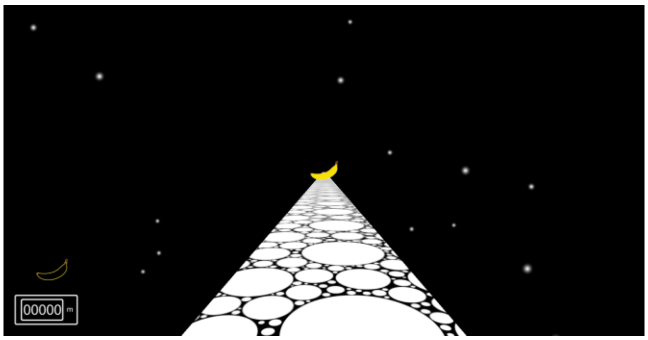
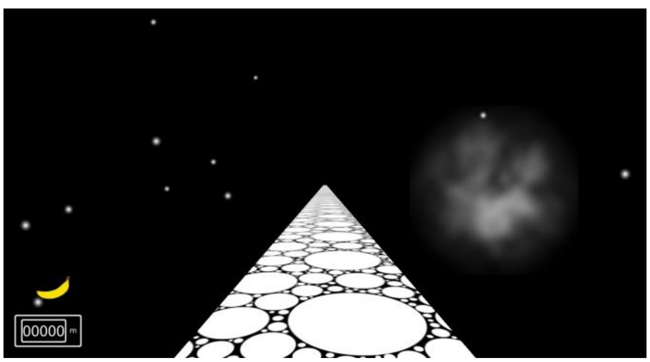


Figure 3: Bellboy monkey with elevator

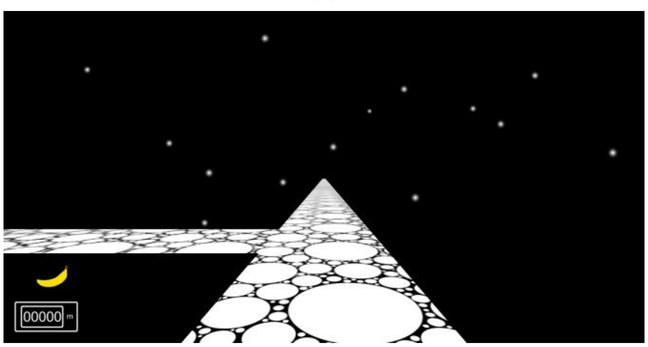
The general goal of the game is to survive (by turning onto the appropriate path) and to collect bananas, in order to go to the next level. These bananas appear briefly at the center of the screen and are caught by pressing the spacebar when they flash in the center of the screen. This mechanic ensures that players fixate the middle of the screen and that there is no saccadic eye movement, only a covert attentional shift (see Figure 4a). At specific times a cue is given in the shape of a flash at the left or right side of the path (see Figure 4b). This is followed by a stimulus in the shape of a turning path, 200 milliseconds later. Again, the stimulus will either be to the left or the right (see Figure 4c). When the cue (i.e., the flash) and the stimulus (i.e., the path) appear on the same side, this is a valid trial. When they appear on opposite sides, this constitutes an invalid trial.



(a)



(b)



(c)

Figure 4. An invalid trial of spatial cueing consists of the following steps: (a) Fixation: players fixate at the center of the screen in order to capture the bananas that appear briefly, (b) cue: a flash appears on the right side of the path to cue the player, (c) target stimulus: the path appears on the left (opposite side), on which the player needs to turn.

Players need to hit the left or right arrow key to make a turn. The time interval between the onset of the stimulus (the turning path) and the key press represents the reaction time of the user. With each mistake of the player (a wrong turn or a missed turn), the path crumbles a little. After a number of missed turns, the path will completely crumble and the player will fall down into space and need to start over from level 1.

The first level however is considered a tutorial level, where players get familiar with the control scheme. Therefore no data from the first level is included in the analysis. As the player progresses in the game, s/he needs to collect more and more bananas, in order to proceed to the next level. However, this increasing difficulty does not influence the response time of the player, who simply needs to turn as fast as possible after seeing the path appear on the left or right.

2.2 Experimental design

To investigate the research hypotheses, an experimental design was carried out with both the game and the classic Posner cueing test as within-subjects factor, and the order in which the two versions were played as a between-subjects factor. After each condition (i.e., having played the game or having performed the classic Posner cueing test), the participants filled out two questionnaires (see measurement instruments below) to measure the participants' experience. We counterbalanced the order, half of the participants first played the game and the other half started with the classic Posner cueing test.

Each participant completed one hundred trials in a standard Posner cueing test, which will act as reference, and also played ten minutes of the game, which resulted in a similar number of trials of the Posner cueing test within the game. Overall, participants completed the experiment in 30 to 45 minutes.

2.3 Participants and location

Twenty participants did the tests, consisting of volunteer students from both the Faculty of Engineering Technology as the Faculty of Psychology and Educational Sciences. The data from two participants were excluded due to anomalously long reaction times (above 1 second).

The tests were done in a Perception Lab, this is a closed dark room where external sound is filtered out so that there is no interference from outside elements or visual stimuli that could influence the results.

2.4 Measurement instruments

The motivation and player experience were measured by using the Intrinsic Motivation Inventory (IMI) and an adjusted version of the Game Experience Questionnaire (GEQ).

The IMI [13] treats the user experience as a multidimensional construct so the participant's subjective experience can be assessed. It assesses participants' interest/enjoyment (e.g., *"I would describe this activity as very interesting."*), perceived competence (e.g., *"I feel very capable and effective when playing."*), effort/importance (e.g., *"I put a lot of effort into this."*) and felt pressure/tension (e.g., *"I felt pressured while doing these."*). The different constructs are assessed by three to five different items for each construct that are evaluated by using a seven point scale, ranging from strongly disagree to strongly agree.

The GEQ we used was an adapted version from the GEQ developed by IJsselsteijn et al. [7,8]. It provides a wider range of constructs than can be assessed via the IMI. However, we adapted some items (that were worded awkwardly for Dutch speaking students), and removed some items from the subscale that proved invalid in previous experiments. The subscales used in our GEQ were the following: Annoyance (e.g., *"I felt irritable."*), Challenge (e.g., *"I had to put a lot of effort into it."*), Competence (e.g., *"I was good at it."*), Flow (e.g., *"I was fully occupied."*), Negative affect (e.g., *"I was bored."*) and Positive affect (e.g., *"I enjoyed it."*). Finally, we added four items to check for control (*"The controls were intuitive."*, *"It was easy to control the game."*, *"I knew how to react in the game."*, and *"Controlling the game was not a problem."*).

When checking the reliability, we decided to remove item 17 "This game did not hold my attention" from the construct Interest-Enjoyment of the IMI. By removing this item we increased the Cronbach's alpha from 0.509 to 0.790 for the classic version and from 0.840 to 0.892 for the game version.

Table 1. The Cronbach's alpha of the constructs that were researched of both the IMI and the GEQ.

Cronbach's alpha	Posner	Game
IMI Competence	0,873	0,897
IMI Effort/Importance	0,853	0,699
IMI Interest/Enjoyment	0,790	0,892
IMI Tension/Pressure	0,846	0,627
GEQ Annoyance	0,939	0,847
GEQ Challenge	0,880	0,580
GEQ Competence	0,944	0,905
GEQ Flow	0,924	0,847
GEQ Negative Affect	0,659	0,863
GEQ Positive Affect	0,761	0,766
Control	0,642	0,885

3. RESULTS

3.1 Posner test results

First the Classical Posner test was checked, to ensure that this reference version was able to pick up a cueing effect. When looking at the valid trials, the paired t-test statistic showed there is an influence on the reaction time (mean effect = 34 ms, 95% CI = 60-7 ms, $t(17) = 2.7$, $p = 0.015$) (see Figure 5).

We also found a robust effect of cue validity on accuracy (i.e. the absence of errors), mean effect = 3.5%, 95% CI = 5.6% - 1.2%, $t(17) = 3.16$, $p = 0.006$ (see Figure 6).

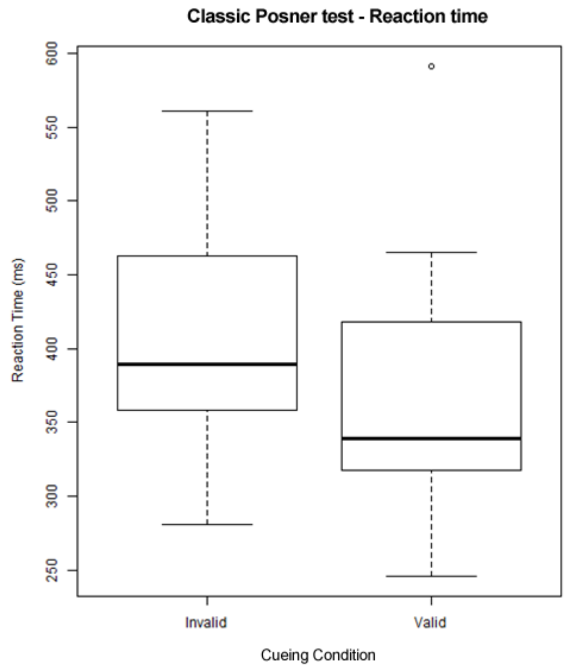


Figure 5. Reaction time for the classic Posner test, for both valid and invalid cueing.

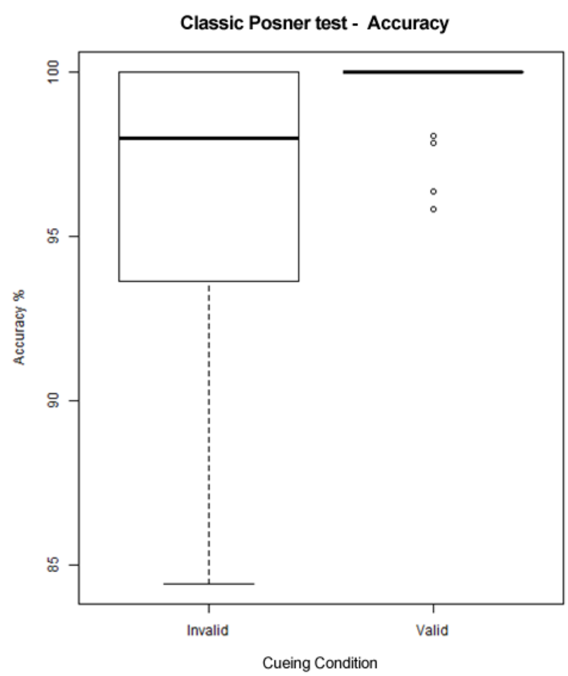


Figure 6. Accuracy (error rate) results for the classic Posner cueing test, for both valid and invalid cueing.

The same analysis was done for the results of the game-based version of the Posner cueing test. When looking at the valid trials, we found there is an influence on the reaction time (mean effect = 13 ms, 95% CI = 66-3 ms, t(17) = 2.33, p = 0.032) (Figure 7). This was again a significant effect. We also found a robust effect of cue validity on accuracy (i.e. absence of errors): mean effect = 5.5%, 95% CI = 7.9%-3%, t(17) = 4.78, p = 0.0001 (Figure 8).

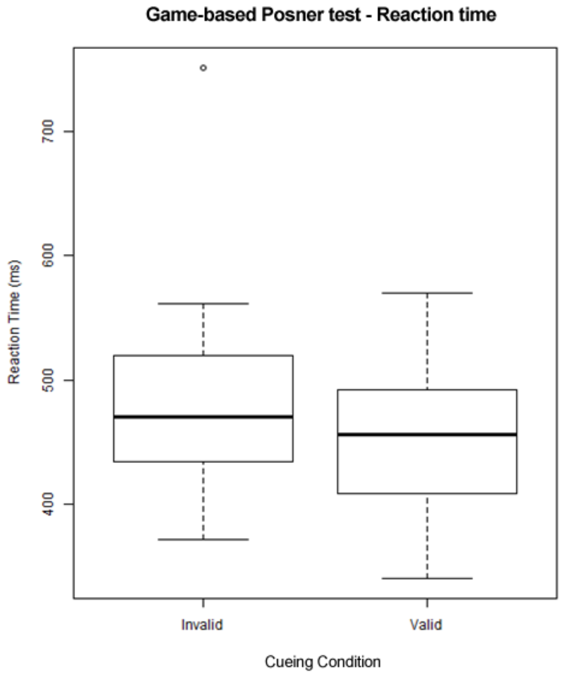


Figure 7. Reaction time for the game-based version of the Posner test for both valid and invalid cueing.

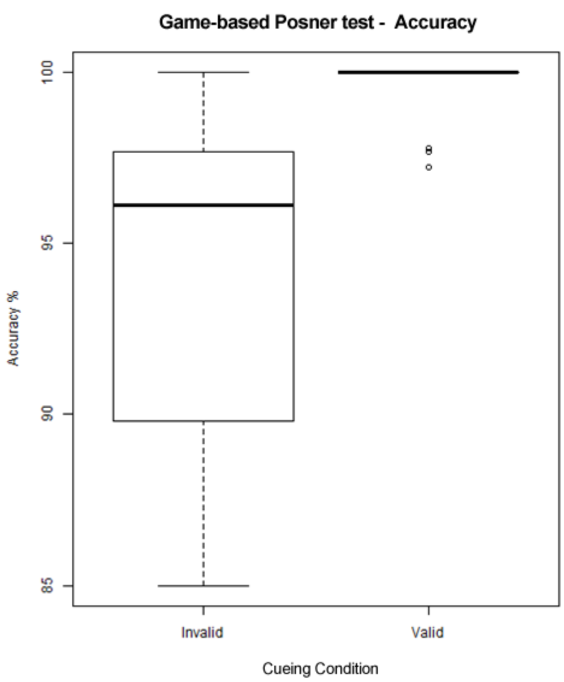


Figure 8. Accuracy (error rate) results for the game version of the Posner cueing test for both valid and invalid cueing.

3.2 Game experience

Perceived enjoyment

Regarding hypothesis 2A, the perceived enjoyment was determined by measuring the following dimensions: IMI Interest/Enjoyment, GEQ Positive Affect, GEQ Annoyance and GEQ Negative Affect.

IMI Interest/Enjoyment. For the dimension IMI Interest/Enjoyment we found a main effect (mean effect size = -0.7500, 90% CI = -1.1735 − -0.3266, F(1,5.625) = 11.258, p < 0.01). According to the IMI Interest/Enjoyment dimension, the player had a higher feeling of enjoyment while playing the game-based version. In addition the dimension had a significant order effect where participants who first completed the classic version of the Posner Cueing experienced less enjoyment than those who played the game version first (see Figure 9).

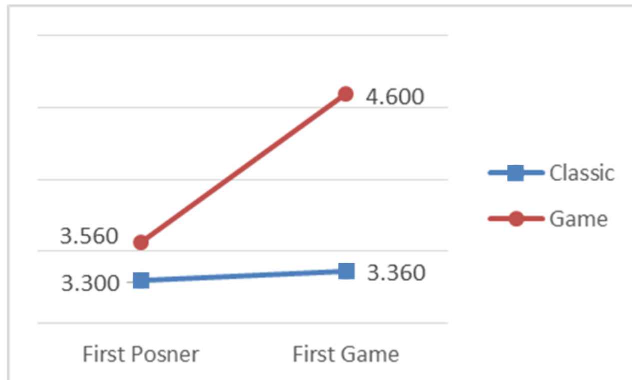


Figure 9. The results of the IMI Interest-Enjoyment.

GEQ Positive Affect. For the dimension GEQ Positive Affect we found a main effect (mean effect size= -0.5444, 90% CI= -0.7080 − -0.3809, F(1, 2.640) = 31.245, p < 0.01). The score shows that the player experienced a greater feeling of enjoyment while playing the game. There was no order effect found (see Figure 10).

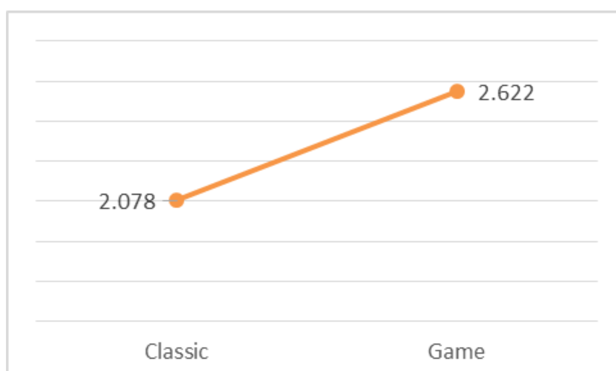


Figure 10. The results of the GEQ Positive Affect which show a significant higher score for the game version.

GEQ Annoyance. For the dimension GEQ Annoyance we found a main effect (mean effect size = 0.4400, 90% CI = 0.1415 − 0.7386, F(1, 1.936) = 6.373, p < 0.05). The player experienced less annoyance when playing the game than when doing the classic version of the Posner cueing test. The GEQ Annoyance dimension had no order effect (see Figure 11).

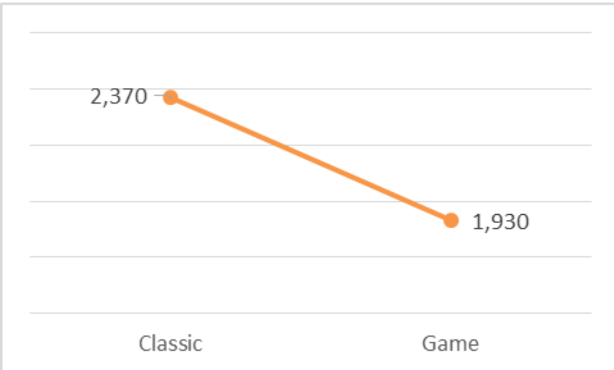


Figure 11. GEQ Annoyance construct.

GEQ Negative Affect. For the dimension GEQ Negative affect we found no main effect (mean effect size = -0.2000, 90% CI = -0.4684 − 0.6832, F(1, 0.321) = 1.439, p = n.s.).

Perceived Engagement

Regarding hypothesis 2B, the perceived engagement was determined by measuring the following dimensions: IMI Effort/Importance, IMI Tension/Pressure, GEQ Flow and GEQ Challenge.

IMI Effort/Importance. For the dimension IMI Effort/Importance we found no main effect (mean effect size = -0.3, 90% CI= -0.67171 – 0.07171, F(1, 0.900) = 2.089, p = n.s.).

IMI Tension/Pressure. For the dimension IMI Tension/Pressure we found a main effect (mean effect size = -0.4000, 90% CI = -0.6216 – -0.1785, F(1,1.600) = 10.056 , p < 0.01). According to the IMI Tension/Pressure dimension the player experienced more tension and pressure when playing the game-based version then when doing the classic version of the Posner cueing test. There was however influence by an order effect as can be seen in Figure 12. Players who played the game first had less pressure when doing the classic Posner cueing test than those who did the game-based version first.

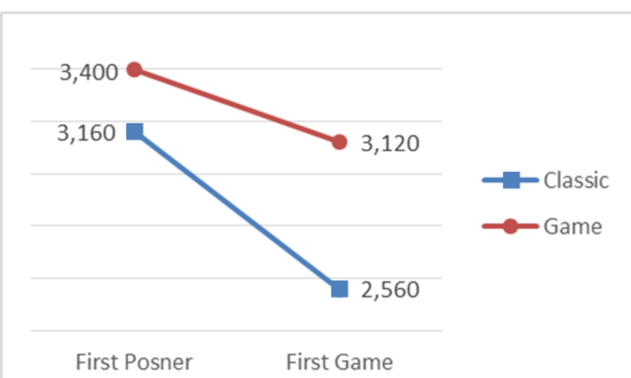


Figure 12. IMI Tension/Pressure

GEQ Flow. For the dimension GEQ Flow we found no main effect (mean effect size = -0.1400, 90% CI = -0.4021 – 0.1221, F(1,0.196) = 0.809, p > n.s.).

GEQ Challenge. For the dimension GEQ Challenge we found a main effect (mean effect size = -0.8800, 90% CI = -1.1928 – -0.5673, F(1, 7.744) = 26.847, p < 0.01). The player experienced a greater feeling of being challenged when playing the game-based version than when playing the classic Posner cueing test. An order effect was found where players who first played the game were less challenged by the classic version than those who played the classic version first (see Figure 13).

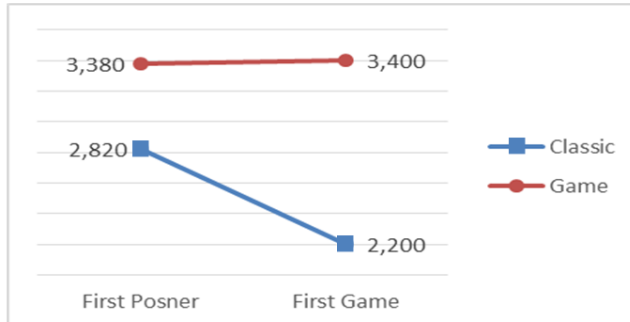


Figure 13. GEQ Challenge.

Perceived Competence and Control

Regarding hypothesis 2C, the perceived competence and control were determined by measuring the following dimensions: IMI Competence, GEQ Competence and GEQ Control.

IMI Competence. For the dimension IMI Competence we found a main effect (mean effect size= 1.2400, 90% CI= 0.7922–1.6880, F(1, 15.376) = 24.493, p < 0.01). The player felt to be more competent when doing the classic version than when playing the game version. No order effect was found (see Figure 14).

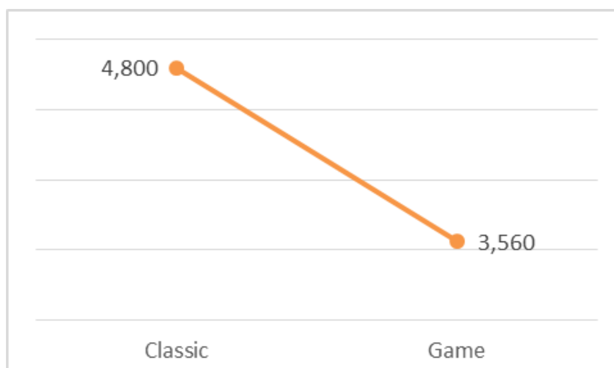


Figure 14. IMI Competence.

GEQ Competence. For the dimension GEQ Competence we found a main effect (mean effect size = 0.8800, 90% CI = 0.5468–1.2132, F(1, 70744) = 20.167, p < 0.01). The player had a greater feeling of being competent when doing the classic version than when playing the game version. There was no order effect found for the GEQ Competence dimension (see Figure 15).

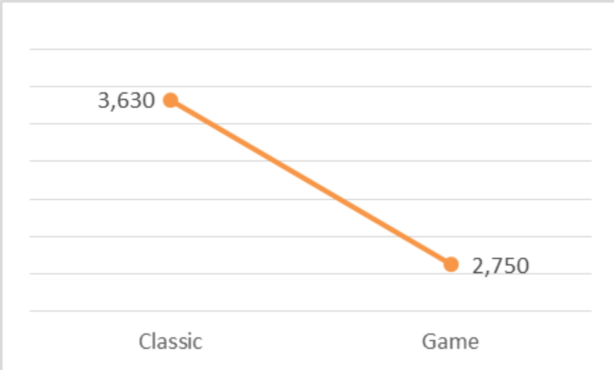


Figure 15 GEQ Competence.

GEQ Control. For the dimension GEQ Control we found a main effect (mean effect size = 0.8111, 90% CI = 0.3633– 1.2589, $F(1, 5.270) = 9.451$, $p < 0.01$). The player felt more in control when taking the classic version than when playing the game version (see Figure 16).

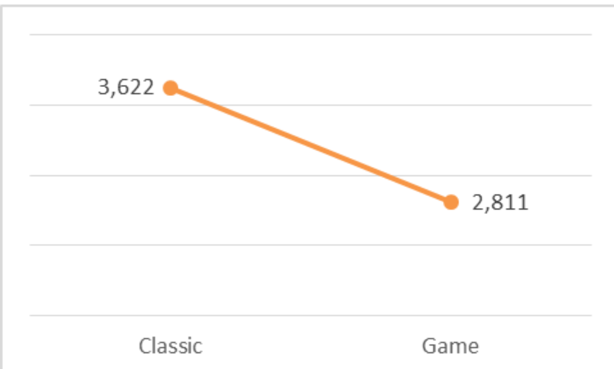


Figure 16. GEQ Control

4. DISCUSSION

Response times & error rates

The results show that the game-based version of the Posner cueing test is able to detect a cueing effect. Players do have faster response times when validly cued, and they have a higher accuracy (i.e., they make less errors). We can therefore confirm our hypotheses 1A and 1B, and state that the game is able to act as a Posner cueing test and will generate reliable results.

However, the results were less articulated than in case of the classical test. Possibly, this might be attributed to the visualization of the cue by means of a flash, this cue was somewhat less bright than the cue used in the classic test. This might also be due to the extra manipulation of having to press the bananas via the space bar. As discussed further, players felt less in control when taking the game-based assessment. However, this remains speculation, and further research is necessary to investigate the effect of this flash visualization and extra player manipulation on the measured response times and error rates.

Perceived enjoyment

The game-based version scored higher on IMI Interest/Enjoyment and GEQ Positive Affect while scoring lower on GEQ Annoyance. The results showed that the participants were less annoyed and bored by the game while having a greater feeling of enjoyment.

Perceived Engagement

Due to the player feeling more pressured and engaged he or she may produce more reliable results as they will do more their best. The game scored higher on IMI Tension/Pressure and on GEQ Challenge. For the IMI Effort/Importance we did not find any significant result but it was trending towards a greater effect for the game-based version than the classic version. We can therefore state that the players had a greater feeling of pressure and being challenged when playing the game. The lack of a significant effect in the GEQ Flow dimension can perhaps be explained by the fact that a flow experience is a difficult construct to be measured by a survey. To have a more reliable measurement of flow, a more elaborate survey should be used for flow alone [9].

Perceived Competence and Control

The classic version of the Posner cueing test scored higher on the IMI Competence, on the GEQ Competence and on the GEQ Control. Thus, we can state that player felt less in control and less competent while playing the game. This can possibly be explained by the fact that the game has a more complex structure which necessitates more complex controls to be able to navigate through the added elements. The bananas, for example need to be collected by pressing the space bar. This added an extra control which required the player to adjust more to the game. The game also needed a longer learning period due to the game rules that are introduced. There are multiple actions while playing the game that require training and thus may explain the lower feeling for competence when playing the game. Further qualitative research is necessary to better understand which parts made the player feel less competent, but arguably this remains hard to avoid in a game-based environment. We therefore have to state that the results refute our hypothesis 2C regarding the competence and control the player was expected to have when playing the game.

Do we need game-based experiments?

In most standard psychology procedures, it is essential to minimize any distraction. Often the goal is to have participants perform repetitive tasks, and it is not desirable to 'entertain' participants. Even simply framing the experimental task as a game might introduce confounds. Hence, the very idea of game-based experiments can be questioned.

In most Posner tests however, participants are *instructed* to focus at the center, and researchers have to trust participants to do so[1]. As often over a 100 trials are required, it is highly likely that participants' dedication may wane occasionally. In the case a trial is offered where a participant's focus is not on the center, this trial does not deliver reliable data. Therefore, by introducing this game-based variant we intrinsically motivated participants to sustain their focus on the center. In order to move to the next level, participants had to collect bananas, only briefly appearing in the center. If the participant's gaze would dwell, rewards

[1] More elaborate test set ups make use of eye tracking technology to check whether participants are staring at the center of the screen. However this demands more resources of the experimenter, both financially, as time wise. Eye tracking technology demands a calibration procedure. Hence, most current Posner experiments do not make use of eye tracking technology.

(bananas) were missed out. Hence, our aim was that gamifying the task of looking at the center would result in more reliable measurements.

We argue that many psychological tests equally demand sustained attention from participants. Such tests may equally benefit from a game-based version, ensuring that participants are intrinsically motivated to keep concentrating on the task. One such other domain is the testing of psychometric thresholds of participants. Ongoing research by our research lab on assessing phonological deficits in children has equally shown that game-based assessments offer the advantage of motivating and engaging participants, and as a result increase attention, and deliver more reliable measurements than the prior test procedures [6,21]. Hence, although a game-based assessment risk to introduce new confounds, it offers the opportunity to increases attention and engagement of participants, and as such it may increase reliability of measurements.

In addition, we have found that game-based versions lengthen the time participants are willing to spend on such a test. This is equally important as in many experiments, there is a trade-off between the length of the procedure that is still deemed acceptable for subjects to 'endure', and the number of data points that is necessary to collect enough data for statistical analysis. We acknowledge however that in this study, as testing was fixed to ten minutes per condition, this could not be assessed.

Finally, as aforementioned, it has been shown that the Posner test not only contributes to insight in human visual attention. Equally, it is becoming a tool to support diagnosis for several afflictions such as psychosis, stroke, Parkinson's and even ADHD [1,2,3,10,14]. In this case, it is highly likely that the Posner test has to be taken outside of a strictly controlled lab environment and into a clinical environment, i.e. at the doctor's office, in a care facility, a therapist's practice or maybe even at home. A researcher, necessary to instruct patients, might not be present during the testing. Moreover, typically, such assessments should be longitudinal (i.e. repeated several times over a longer period), in order to measure how the affliction progresses or mitigates. In such cases, a game-based version that keeps a participant intrinsically motivated to partake in the test is paramount. Hence, this evolution towards clinical relevance of tests on visuo-spatial attention, strengthens the argument for game-based experimental testing where there is no need for the presence of a researcher to ensure that patients are focused on the task.

To summarize, we argue that game-based assessments of experiments may be particularly beneficial in case: 1) test subjects need to be engaged and attentive over a lengthy period, 2) tests need to be taken repetitively, 3) tests need to be taken without the presence of the researcher.

5. FUTURE WORK

This study is only scratching the surface of visio-spatial effects. In this paper, we have not yet discussed other variants of spatial cueing such as endogenous cueing [18] and object-cueing [4,17] that might further influence response times of players. Therefore, this game-based experiment is only the first of a series of experiments in which we will test visuo-spatial cueing effects. A better understanding of the operation of spatial attention in neurologically normal observers (aka players) can also guide assessments in brain-damaged patients. Knowing which processes or sub-processes of spatial attention are afflicted may be helpful for developing assessment techniques or rehabilitation strategies.

6. CONCLUSION AND FUTURE WORK

In order to enhance the participant's experience during the Posner cueing test, and to safeguard the reliability of the results from growing boredom, we designed and developed a game-based version of the Posner cueing test. This game-based version, as well as a classic implementation of the Posner cueing test, were played by 20 participants. After each play session they were asked about their experience with both the Intrinsic Motivation Inventory (IMI) and an adjusted version of the Game Experience Questionnaire (GEQ).

We found that both versions showed a significant cueing effect. This means that the game-based version qualifies as a reliable Posner cueing test as well. Furthermore, the game did provide more enjoyment and less annoyance. Our hypothesis regarding the challenge and pressure that the game should provide was also confirmed. The game-based version was able to keep the participant more engaged. This can lead to more reliable Posner cueing test results in the long run. Our results however also showed that the game version was not as easy to control as the classic Posner test, and the threshold for participants to feel competent in the game version was generally higher. Our hypothesis regarding expected competence and control the player was therefore refuted. Nevertheless, the game-based version did provide both a reliable way of performing the Posner cueing test and a more enjoying experience to the participant.

Hopefully this opens the door to more game-based assessments in experimental psychology (and other fields for that matter), contributing to the reliability of scientific research. More generally this is a plea for game designers and experimental psychologists to exchange knowledge.

7. REFERENCES

1. Alberts, J.L., Tresilian, J.R., and Stelmach, G.E. The co-ordination and phasing of a bilateral prehension task. The influence of Parkinson's disease. *Brain: A Journal of Neurology 121 (Pt 4)*, (1998), 725–742.
2. Amodio, P., Marchetti, P., Del Piccolo, F., et al. Visual attention in cirrhotic patients: A study on covert visual attention orienting. *Hepatology 27*, 6 (1998), 1517–1523.
3. Desimone, R. and Duncan, J. Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience 18*, 1 (1995), 193–222.
4. Egly, R., Driver, J., and Rafal, R.D. Shifting visual attention between objects and locations: evidence from normal and parietal lesion subjects. *Journal of experimental psychology. General 123*, 2 (1994), 161–177.
5. Gamberini, L., Barresi, G., Maier, A., and Scarpetta, F. A game a day keeps the doctor away: a short review of computer games in mental healthcare. *Journal of CyberTherapy and Rehabilitation 1*, 2 (2008), 127–146.
6. Geurts, L., Abeele, V., Celis, V., et al. DIESEL-X: A game-based tool for early risk detection of dyslexia in preschoolers. In J. Torbeyns, E. Lehtinen and J. Elen, eds., *Developing competencies in learners: From ascertaining to intervening*. 2015.
7. IJsselsteijn, W., de Kort, Y.A.W., and Poels, K. The Game Experience Questionnaire:Development of a self-report

measure to assess the psychological impact of digital games. Manuscript in Preparation. (in preparation).
8. IJsselsteijn, W., de Kort, Y.A.W., Poels, K., Jurgelionis, A., and Bellotti, V. Characterising and Measuring User Experiences in Digital Games. ACM Press (2007).
9. Jackson, S.A. and Eklund, R.C. *Flow Scales Manual.* Fitness Information Technology, 2004.
10. Karper, L.P., Freeman, G.K., Grillon, C., Morgan III, C.A., Charney, D.S., and Krystal, J.H. Preliminary evidence of an association between sensorimotor gating and distractibility in psychosis. *The Journal of neuropsychiatry and clinical neurosciences*, (1996).
11. Laneau, J., Boets, B., Moonen, M., van Wieringen, A., and Wouters, J. A flexible auditory research platform using acoustic or electric stimuli for adults and young children. *Journal of Neuroscience Methods 142*, 1 (2005), 131–136.
12. Mari, M., Bennett, K.M.B., Scarpa, M., Brighetti, G., and Castiello, U. Processing efficiency of the orienting and the focusing of covert attention in relation to the level of disability in Parkinson's disease. *Parkinsonism & Related Disorders 3*, 1 (1997), 27–36.
13. McAuley, E., Duncan, T., and Tammen, V.V. Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research quarterly for exercise and sport 60*, 1 (1989), 48–58.
14. McDonald, S., Bennett, K.M.B., Chambers, H., and Castiello, U. Covert orienting and focusing of attention in children with attention deficit hyperactivity disorder. *Neuropsychologia 37*, 3 (1999), 345–356.
15. Posner, M.I. Orienting of attention. *Quarterly journal of experimental psychology 32*, 1 (1980), 3–25.
16. Rengachary, J., d' Avossa, G., Sapir, A., Shulman, G.L., and Corbetta, M. Is the posner reaction time test more accurate than clinical tests in detecting left neglect in acute and chronic stroke? *Archives of physical medicine and rehabilitation 90*, 12 (2009), 2081–2088.
17. Reppa, I., Schmidt, W.C., and Leek, E.C. Successes and failures in producing attentional object-based cueing effects. *Attention, Perception, & Psychophysics 74*, 1 (2012), 43–69.
18. Theeuwes, J. Exogenous and endogenous control of attention: the effect of visual onsets and offsets. *Perception & psychophysics 49*, 1 (1991), 83–90.
19. Tong, T. and Chignell, M. Developing a Serious Game for Cognitive Assessment: Choosing Settings and Measuring Performance. *Proceedings of the Second International Symposium of Chinese CHI*, ACM (2014), 70–79.
20. Tong, T. and Chignell, M. Developing a Serious Game for Cognitive Assessment: Choosing Settings and Measuring Performance. *Proceedings of the Second International Symposium of Chinese CHI*, ACM (2014), 70–79.
21. Vanden Abeele, V., Wouters, J., Ghesquière, P., Goeleven, A., and Geurts, L. Game-based Assessment of Psychoacoustic Thresholds: Not All Games Are Equal! 2015. https://lirias.kuleuven.be/handle/123456789/494491.
22. Vecera, S.P. and Rizzo, M. Spatial attention: normal processes and their breakdown. *Neurologic clinics 21*, 3 (2003), 575–607.