Impulsivity test implemented with game elements – can validity be maintained?

Olle Hilborn, Henrik Cederholm, Jeanette Eriksson, Craig Lindley, Charlotte Sennersten

Blekinge Institute of Technology
Blekinge Tekniska Högskola
SE – 371 79 Karlskrona
<firstname.lastname>@bth.se

Abstract. Four variants of the go/no task, which is an impulsivity test, have been implemented as computer games. A total of forty subjects tested the variants. The different implementations were compared against the Barrat Impulsiveness Scale (another impulsivity test) in order to see if all of the games target the same impulsivity constructs. Results showed that the different games actually measured different constructs, hence validity was not maintained. However, the games have more focused individual validity, since they correlate with other constructs, just not the same constructs as the original go/nogo task.

Keywords: Impulsivity, assessment tool, behavior game.

1 Introduction

Much research and discussion in the serious games community focuses on the development of learning and subsequent assessment of learning effectiveness (e.g. [1], [4], [9]). Here a different approach was taken that started with a cognitive task that was tested through several variants of game implementations in order to assess if the task's diagnostic value can be maintained through these variations. The aim of this research is to be able to build small games, or game components, based on cognitive assessment tests. If successful, flash games could be created that players enjoy playing that at the same time provide validated information about various behavior tendencies for use in treatment and education programs. The benefits of using games for this purpose include facilitating cognitive tests in ways that can be distributed by the internet or packed with application software, and to provide built-in motivation and rewards for engaging with the tests when they are distributed electronically in this way. However, validation is critical, especially to ensure that gameplay does not interfere with the assessment functions of the cognitive tests.

The EU-project xDELIA (Xcellence in Decision-making through Enhanced Learning in Immersive Applications, Contract No. 231830) seeks to improve financial capabilities. One target group for this was young adults since this group prone to making financial mistakes [6]. In order to help this group, xDELIA has identified impulsiveness as a construct that is associated with negative financial behavior. Games that diagnose impulsiveness could thus increase awareness of how it bears upon life experiences, especially financial ones, and hopefully contribute to preventing unwanted financial behavior.

For impulsiveness, there is a commonly used self-report tool, the Barratt Impulsiveness Scale (BIS) [7]. BIS consists of six first-order factors: *Attention*, focusing on a task at hand; *Motor impulsiveness*, acting on the spur of the moment; *Self-control*, planning and thinking carefully; *Cognitive complexity*, enjoying challenging mental tasks; *Perseverance*, the ability to stay focused on a task; *Cognitive instability*, thought insertions and racing thoughts.

The go/nogo task is a response-inhibition task, which means that the player is supposed to react as fast as possible in the response condition but must avoid reacting during the inhibition condition. The original task of the prototypes reported in this paper is modeled after the go/nogo test by Vocat et al.[11]. Measurements taken include reaction time (RT), omissions on go trials, and false alarms (FA) on nogo trials, where FA is seen as a measurement of impulsivity. The work presented here implemented go/nogo tasks with the added game elements of music and graphics (a moving robot character), based upon the hypothesis that: *The assessment of impulsivity through the go/nogo task will remain the same even if specific game elements are added*. This means that game elements, separately and in combination, had to be controlled and tested against the original impulsivity measures.

A recent model of work and performance is presented by [3], which characterises peoples' effectiveness on tasks in terms of how much cognitive resource they can allocate to them. A completely effective person engaged in a single task will allocate 100% of their cognitive resources to the task. This depends on the task attentional pull, off-task attentional demands (which can have an affective character, such as a feeling of sadness making thoughts wander) and the self-regulation and cognitive resources of the person (mediating the other two factors).

Posner et al. [8] made four propositions about visual attention compared with attention to other sensory modalities: *Proposition 1:* Visual stimuli are not as automatically alerting as stimuli in other modalities. *Proposition 2:* In order for a visual event to serve as an effective alerting stimulus, the subject must first process the visual event by active attention. *Proposition 3:* The consequence of active attention towards any one modality is a reduction in the availability of the attentive mechanisms available for input to other modalities. *Proposition 4:* To compensate for the low alerting capability of visual signals, subjects exhibit a general attentional bias toward the visual modality whenever they are likely to receive reliable input from that modality.

In the study reported in this paper, Posner et al.'s propositions mean that in order for a player to react appropriately, the player needs to focus attention away from music (in the condition where it applies) towards visual input via a conscious action, since music should have a higher off-task attentional demand than vision, according to proposition 1. The player must keep a constant high attention level directed towards the game, focused on the visual modality, since the task attentional pull can be presumed to be relatively low in this case.

The music used in two of the prototypes was "The Murder" by Bernard Hermann, which has been shown to induce fear during guided visual imagery [7]. Since emotions can affect performance, emotions were controlled for with the profile of mood states short form (SF-POMS [10]), which tests for depression, vigor, confusion, tension, anger, and fatigue.

2 Design

Four implementations of the go/nogo task were created: 1. as described in [11] called *standard*, 2. as the first implementation but with the music playing in the background (*Psycho*), 3. with an animated robot character running on a platform (*LineRacer*), and 4. with both the music and the animated character (*LineRacer Psycho*) (see table 1). In the LineRacer versions the players received feedback on their progress in the form of the speed of the robot character: a correct answer led to the character running (8 m/s), incorrect led to the character stopping (0 m/s), and no input led to the character walking at a medium pace (4 m/s).

Table 1. Overview of independent variable value combinations for the different prototypes.

	No Music	Music
No Character	Standard	Psycho
Character	LineRacer	LineRacer Psycho

The players had to react to a certain stimulus; first, a black arrow was shown pointing either up or down, and then another arrow was shown. If the second arrow was green AND pointing in the same direction as the first arrow, the player was supposed to press the button "left ctrl" as fast as possible (go condition). If not, the player was supposed to not press (nogo condition). Gameplay consisted of two phases, the first one including 14 consecutive turns, and the second 33 consecutive turns. Both of these phases together completed one experiment cycle. The first phase was a "calm" phase in which the player reacted as soon as possible, but without any other stressors (except for music in the relevant prototypes).

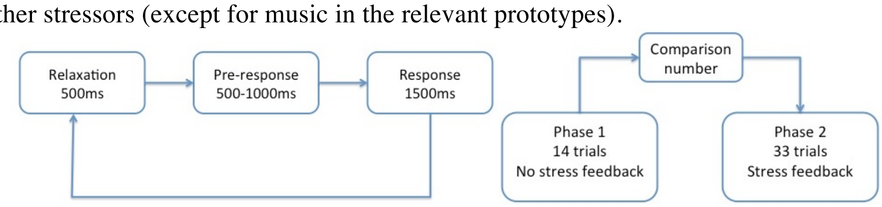

Fig. 1. The parts of a turn (left), and a cycle (right)

Each turn was divided into a relaxation period (500ms), a pre-response period (black arrow, random between 500-1000ms long), and a response period (green or blue arrow, 1500ms) (see figure 1, left). The animated character had a constant walking speed of 4.0 m/s. When a faulty answer i was s given, the character stopped for the rest of the response period. When a correct answer was given, the character started running (with a speed of 8.0 m/s) for the rest of the response period. The mean reaction time (on go-trials) was calculated for the 14 first turns in every other phase, then multiplied with 0.9, as in [11]. This number, comparison number, was then compared with every new go-answer that was given. If the reaction time was higher than the comparison number, stress feedback was displayed telling the player that he or she can react faster. After 33 trials with stress feedback, a cycle was completed (see figure 1, right) and a new cycle began. Each cycle contained four nogo trials during phase one and eight during phase two. There were three cycles throughout the game, resulting in 30 go trials per participant without any stress and 75 potential go trials with stress feedback.

The score obtained by the player in the game was the distance the character travelled during the game. The score was calculated exactly the same way in the version where there is no animated character.

3 Method

Forty participants were recruited for the experiment, including students and people employed at the faculty. All participants were first asked to fill out a web questionnaire consisting of the BIS, the SF-POMS, and demographic data (age, gender, years of education, education program, and origin). Each participant was then randomly assigned to play only one of the four prototypes. After playing, all participants were again given the SF-POMS to detect if any mood changes had occurred during the experiment.

4 Results

Participants had a mean age of 28.7 (SD = 7.2), with 22 males and 18 females. Mean time spent in education was 15.75 years (SD = 2.3). Two people did not understand the task and have therefore been excluded from the results. Additionally, one person was feeling ill and sleepy and nodded off during the experiment and was thus also excluded from the results.

Analysis of variances (ANOVA) was performed, but no significant results between groups were found among the game factors of total score and reaction time. However, there was a difference in FA (false alarms, $p = 0.028$) between the groups. The Bonferroni analysis of FA did not reveal which of the groups had a big difference between them, so they were grouped based on character and no character, but also with music and non-music for further analysis (see table 1, grouping based on rows and columns). A t-test between the LineRacer (character) group ($N = 17$) and the non-LineRacer group ($N = 20$) showed a significantly larger amount of FA in the LineRacer group ($p = 0.002$), but no difference in reaction time or game score. Regarding the music group ($N = 18$) and non-music group ($N = 19$), no significant differences were found regarding reaction time, game score, and FA.

No differences were found between the groups (Standard, Psycho, LineRacer, and LineRacer psycho) regarding results from the BIS according to the ANOVA statistics. Another ANOVA showed no significant difference between the SF-POMS results before or after the game was played. T-tests within each group were performed in order to control for mood changes during the experiment; the only significant change was that LineRacer psycho felt less vigorous after the test ($p = 0.017$).

Table 2. Overview of Pearson's correlations (Spearman's rho marked where applicable) for total score, false alarms (FA), reaction time (RT), demographic data, BIS measures, and SF-POMS with regard to the groups LineRacer, non-Lineracer, Music, and Non-music. Significant correlations marked in **bold**.

	LineRacer	Non-LineRacer	Music	Non-music
Attention - FA	**-0.564**	0.170	0.202	-0.368
Attention Total Score	-0.028	**-0.499**	-0.316	-0.211
Attention RT (Spearmans)	0.091	**0.470**	0.181	0.215
Motor - FA	**-0.495**	0.136	-0.41	-0.332
Motor FA (Spearmans)	**-0.606**	-0.009	-0.156	-0.258
Cognitive Instability RT (Spearmans)	**-0.508**	-0.093	-0.288	0.115
Age Total Score	-0.453	-0.22	0.016	**-0.458**
Age Total Score (Spearmans)	-0.400	-0.096	-0.041	**-0.590**
RT-FA	-0.101	-0.411	**-0.616**	-0.84

Correlation data for reaction time, total score, FA, demographic data, BIS, and SF-POMS are seen in table 2. No correlations were found for the regular groups (except interrelations within SF-POMS, which are already shown in [2]). Another finding (t-test) was that RT significantly differed (p = 0.006) between people who have an origin in western cultures (Sweden, Britain, and Australia) and in eastern cultures (Iran, Bangladesh, and Pakistan), where the westerners were faster.

5 Discussion and Conclusion

The hypothesis "*The assessment of impulsivity through the go/nogo task will remain the same even if specific game elements are added.*" is not supported by the results from the LineRacer and non-LineRacer comparison, since there was a very clear difference in the number of FA between these variants. Emotional reporting suggested no difference between these groups, suggesting that emotions did not increase off-task attention. Posner et al.'s [8] second proposition that a visual event must be actively attended in order to be properly processed is a plausible explanation model; when a player has the running character, since it is moving and is the only way of getting feedback on how the player is doing in the game, it might be more interesting for the player to look at the character (constituting off-task attention pull), thus loosing task focus.

The BIS constructs *attention* and *motor impulsiveness* were strongly negatively correlated with FA within the LineRacer group, so those games are a good candidate for further implementation as tests of impulsiveness. However, the task would no longer have the support of the current literature since the non-LineRacer group, which can be considered a more "pure" version of the go/nogo task, has opposite correlations on each impulsiveness factor in table 2 compared with the LineRacer group. As people scored higher in the attention construct, their RT increased (which explains the negative correlation between game score and attention, since a good score is dependent on low RT). This is a very peculiar relationship, since better attention is expected to speed up reaction, not slow it down.

There were no findings of a tradeoff between RT and FA, except for in the music group. This tradeoff is interesting since it was not found in the other group, nor in the literature on which these games were built [11]. It could be that the music induces this tradeoff.

The data presented above suggests that since variations of the task measured different constructs, the variations were no longer the same tasks. Even though most task characteristics were constant, the added animations and music recruit other cognitive systems or resources than the ones used in the standard task.

A concluding reflection regarding different go/nogo tasks is that it would be possible to try to implement a completely new task with much higher attentional pull right from the beginning, in order to prevent a standard example of the task from being polluted by distractors (game features). This way the task could be cross-validated with existing impulsiveness measures to accomplish a game assessment tool for impulsiveness. The reason for not taking this approach in this study was that a gradual increase of game elements was of interest in order to asses a derivation of the go/nogo task as an assessment tool for impulsivity in game form, and if the game elements did not have any influence upon cognitive processing, then it, in the game form, may serve no purpose.

References

1. Anetta, L.A.: The "Is" Have It: A Framework for Serious Educational Game Design. Review of General Psychology. 14, 105-112 (2010)
2. Baker, F., Denniston, M., Zabora, J., Polland, A., Dudley, W.N.: A POMS short form for cancer patients: psychometric and structural evaluation. Psycho-oncology. 11, 273-281 (2002)
3. Beal, D.J., Weiss, H.M., Barros, E., MacDermid, S.M.: An Episodic Process Model of Affective Influences on Performance. Journal of Applied Psychology. 90, 1054-1068 (2005)
4. de Freitas, S., Jarvis, S.: Serious games – engaging training solutions: A research and development project for supporting training needs. British Journal of Educational Technology. 38, 523-525 (2007)
5. Mayer, J.D., Allen, J.P., Beauregard, K.: Mood Inductions for Four Specific Moods: A Procedure Employing Guided Imagery Vignettes With Music. Journal of Mental Imagery, vol. 19, 133-150 (1995)
6. NSG-FC. Improving Financial Capability – a multi-stakeholder approach. Report of the National Steering Group on Financial Education, Dublin, Ireland. 2009.
7. Patton, J., Stanford, M., Barratt, E.: Factor Structure of the Barratt Impulsiveness Scale. Journal of Clinical Psychology. 51, 768-774 (1995)
8. Posner, M.I., Nissen, M.J., and Klein, R.M.: Visual Dominance: An Information-Processing Account of Its Origins and Significance. Psychological Review. 83, 157-171 (1976)
9. Prensky, M.: The Digital Game-Based Learning Revolution. Digital Game-Based Learning. McGraw-Hill, New York 2001.
10. Shacham S.: A shortened version of the Profile of Mood States. Journal of Personality Assessment. 47, 305-306 (1983)
11. Vocat, R., Pourtois, G., Vuilleumier, P.: Unavoidable errors: A spatio-temporal analysis time-course and neural sources of evoked potentials associated with error processing in a speeded task. Neuropsychologia. 46, 2545-2555 (2008).