Developing a Serious Game for Cognitive Assessment: Choosing Settings and Measuring Performance

Tiffany Tong
University of Toronto
5 King’s College Road
tiffany.tong@mail.utoronto.ca

Mark Chignell
University of Toronto
5 King’s College Road
chignell@mie.utoronto.ca

ABSTRACT
Gamification and serious games are becoming increasingly important for training, wellness, and other applications. How can games be developed for non-traditional gaming populations such as the elderly, and how can gaming be applied in non-traditional areas such as cognitive assessment? The application that we were interested in is detection of cognitive impairment in the elderly. Example use cases where gamified cognitive assessment might be useful are: prediction of delirium onset risk in emergency departments and post-operative hospital wards; evaluation of recovery from stroke in neuro-rehabilitation; monitoring of transitions from mild cognitive impairment to dementia in long-term care. With the rapid increase in cognitive disorders in many countries, inexpensive methods of measuring cognitive status on an ongoing basis, and to large numbers of people, are needed. In order to address this challenge we have developed a novel game-based method of cognitive assessment. In this paper, we present findings from a usability study conducted on the game that we developed for measuring changes in cognitive status. We report on the game’s ability to predict cognitive status under varying game parameters, and we introduce a method to calibrate the game that takes into account differences in speed and accuracy, and in motor coordination. Recommendations concerning the development of serious games for cognitive assessment are made, and detailed recommendations concerning future development of the whack-a-mole game are also provided.

Author Keywords
Calibration; digital game design; Fitts’ Law, game usability; gamification; human-computer interaction; interface design; serious games; speed-accuracy tradeoff; user experience; user-centered design.

ACM Classification Keywords
H.5.2 [User Interfaces]: User-centered design, prototyping, evaluation/methodology; K.8 [Personal computing]: Games.

INTRODUCTION
As healthcare shifts from an emphasis on treatment of acute conditions (such as infectious diseases) to an emphasis on chronic conditions (such as congestive heart failure and diabetes), there has been a corresponding increase in continuous measurement of patient status through variables such as body weight, blood pressure, and blood sugar level. In addition to these physiological measures, measures of cognitive status are important in elderly populations and are particularly relevant in cases of head trauma, addictions, and mental disease. Ideally, methods of cognitive assessment should be easy and inexpensive to administer, and they should support repeated, and reliable, measurement of cognitive status. In this paper, we report on research that takes an existing game (Figure 1) and uses it as a platform for creating cognitive measurement opportunities by modifying game parameters appropriately. A method is developed for tuning the game parameters so that they can be adapted to fit the motor dexterity, and cognitive speed, of users.

BACKGROUND
Cognitive impairment [8] and changes in cognitive status [23] can easily go unnoticed in clinical settings. Cognitive assessments are used to evaluate a patient, but are often time-consuming to complete, and require administration by a trained healthcare specialist.

A number of cognitive assessment techniques have been developed for use in assessing the elderly in clinical settings. They are often designed to be administered by neurologists or gerontologists [24]. The Montreal Cognitive Assessment (MoCA) [24] is a relatively brief paper and pencil test that can be administered by non-clinicians. However, the MoCA is not designed for repeated use, and, like most clinical tests, it looks for evidence of impairment rather than differences in cognitive status in the normal range which may be indicative of decline or improvement.

New methods of cognitive assessment are needed that are inexpensive and that can be carried out on many people over extended periods of time in order to detect sudden changes in cognitive state that may be due to conditions such as delirium [15] or to transitions from mild cognitive impairment to states of dementia. Large scale, efficient and effective cognitive assessment could revolutionize detection of neurological conditions in the elderly.

Existing tools for cognitive assessment tend to be paper-based, reflecting technologies and scientific knowledge that
are decades, and sometimes more than a century, old. New technologies and platforms such as computer tablets offer many opportunities for creating tasks and interactive experiences from which cognitive status could potentially be inferred. One approach for utilizing new technology in cognitive assessment takes existing paper and pencil tests such as the symbol-digit matching test and converts them to an equivalent computer-based task, which can be self-administered [20]. For example, there are cognitive assessment software suites available such as CogTest [2], and the Cambridge Neuropsychological Test Automated Battery [18], which offer computerized versions of traditional tests. Problems with these existing tools include a need to re-validate the tests for the computer medium, and a potential lack of motivation when performing somewhat uninteresting tasks on a computer. Games have also been promoted as a way to stimulate cognitive activity in elderly users [7] or to improve brain fitness or to preserve cognitive status using specially designed training games, many of which are intended to enhance or preserve working memory. However, a recent study found no evidence that training on such games improved broader measures of intelligence [9].

In our research, we are addressing the lack of simple, and enjoyable methods of self-administered cognitive assessment by developing a game that assesses cognitive status using a touch-based device. The game is based on the classic carnival game called ‘whack-a-mole’. The goal of the carnival version of the game is to whack a mole using a mallet. On a mobile device, hitting a mole with a mallet can be replaced by hitting the image of a mole with one’s finger. The findings from a usability study on the game are presented below, and methods are presented for analyzing individual performance data, and for calibrating performance for users with varying abilities.

GAME CONCEPT

The ‘whack-a-mole’ game was adapted to emphasize the central executive component of working memory [1] and the inhibitory function in particular. A condition was added where a second character (a distractor such as a picture of a butterfly) was sometimes shown, which the player was not supposed to hit. We wanted to assess ability to inhibit the whacking response on distractor trials, since inhibition ability has been shown to decline with age, independent of processing speed [4].

We propose the whack-a-mole game as a useful platform for assessing abilities on central executive functions within working memory [4]. While the basic whack-a-mole game seems most closely related to inhibition, it can be modified in various ways to assess different central executive functions. For instance, moles can have numbers on them and the game rules can change so that a mole should only be whacked if it appears in the right numerical order (adding a requirement to use the updating executive function).

Central executive functions (EFs) are regulated within the pre-frontal cortex of the human brain, and they incorporate complex cognitive processes such as working memory, problem solving, and reasoning [6], [19]. One model of central EFs proposed by [13], suggests that the three EFs of inhibition (ability to prevent an action or behaviour), shifting (ability to switch between tasks), and updating (ability to update one’s working memory) ability can be used to understand complex cognitive tasks and task performance. In [11]’s book “Thinking, Fast and Slow”, EFs are associated with what he calls “system 2”, the effortful processing associated with complex tasks and with controlling attention. [5] stated that “executive functions make possible mentally playing with ideas; taking the time to think before acting; meeting novel, unanticipated challenges; resisting temptations; and staying focused”. The “core EFs” that she listed were inhibition, interference control, working memory - which likely overlaps with [13]’s conception of updating and [12]’s notion of monitoring - and cognitive flexibility. There is general agreement that EFs are important indicators of cognitive status, and declines in these abilities have been used to detect adverse changes in cognitive status such as post-operative delirium [13], [22]. Mechanisms of inhibition and selective attention are key constructs of executive functioning which are included in almost all discussions of EFs. Impaired executive functioning is associated with ageing adults and functional decline, thereby impacting quality of life and ability to complete activities of daily living such as maintaining mobility, and bathing [10].

Using the CogMed system, [3], found evidence that training on EFs is beneficial, although transfer tends to be to closely related functions, e.g. “computer training on spatial working memory transfers to other measures of spatial working memory”. [5] claimed that EFs can be improved across the life
cycle and especially in the elderly. She cited research showing that EFs improve with physical fitness and there have also been promising results from computerized EF training with the elderly (e.g. [17]).

METHODOLOGY
Requirements Analysis
To gain an understanding of the needs of elderly adults a series of informal meetings were held with physicians, psychologists, researchers, and occupational therapists at two Toronto hospitals.

We identified two key requirements in these meetings. The first was a need to use a touch-based device that would be lightweight, portable, and easy to sanitize after each use. The second requirement was for a self-administered assessment task that would be enjoyable to use and that could be adapted to the capabilities of the user.

Game Design Process
An iterative design approach was utilized. First, paper-and-pencil wireframes were created, and the architecture of the game was mapped out. The resulting prototype was then subjected to initial usability review. A version of the game was then developed on the Android platform and used in the experiment reported below.

Game Features
In the whack-a-mole game designed for this research, there are five adjustable settings: game duration, target size, grid size, distractor style, and feedback style (Figure 2). Each of these settings is customizable in order to provide a gaming experience that is the appropriate level of difficulty for varying levels of users. Each adjustable parameter is described below:

- **Game duration (seconds)**: This parameter allows the user to select the length of the game.
- **Target size (150 pixel, 175 pixel, 200 pixel)**: This feature sets the size of the targets (i.e. moles and butterflies). Note that distractors were constrained to be the same size as targets since having distractors be of a different size than the targets, when playing the game, seemed unnatural.
- **Grid size (2x2, 3x3)**: This setting adjusts the size of the game board, and thus the number of potential ‘holes’ a character can appear in.
- **Distractor style (absent, present)**: This setting allows users to select whether they want a game with moles only (distractor absent setting) or with both moles and butterflies (distractor present setting).
- **Feedback style (absent, present)**: This parameter allows users to select if they want feedback after hitting a target. When a mole is correctly hit, a ‘checkmark’ appears over the mole. In contrast, when a butterfly is a hit, an ‘x’ appears on the butterfly. Note that the size of the checkmark/x character was scaled to the size of the target/distractor, so that as the target/distractor got bigger, so did the feedback character.

Figure 2. Screen capture of the game settings menu displaying the five adjustable parameters.

Usability Study
A usability study was conducted at the Interactive Media Lab at the University of Toronto according to the requirements of a research protocol approved by an Ethics Review Board at the University of Toronto.

Participants
The study was conducted with 24 healthy volunteers. All participants were recruited from the University of Toronto. Participants received $20.00 CAD as compensation. The sample population included 7 females and 17 males between the ages of 21 to 51 years.

Research Questions
The following research questions were addressed:

- Can game performance data gathered on a touch-based device be modelled using Fitts’ Law and a speed-accuracy tradeoff function?
- Can an individual game performance metric be created that considers both speed and accuracy in an effective way?
- Can the whack-a-mole game predict cognitive ability?

Overview
The study was comprised of three parts: (i) a background questionnaire on demographic information and technology use, (ii) cognitive ability tests administered on a computer, and (iii) the game played on a tablet followed by an exit-survey on the usability of the game and tablet experience.
**Background Questionnaires**
Background questionnaires were asked participants about demographics, and about their familiarity with touch-based technologies.

**Computer-Based Cognitive Tests**
In this phase participants were asked to complete a set of three tests assessing their cognitive abilities on a computer: (i) the Stroop Test was used to test inhibition, (ii) the Wisconsin-Card Sorting Test was used to evaluate shifting, and (iii) a Color Monitoring Test was used to examine updating ability. Description of these tests are available in [14].

**Tablet-Based Whack-A-Mole Game**
Following the cognitive ability tests, participants were presented with a Google Nexus 7 tablet (Figure 3) and asked to play the whack-a-mole game.

The usability experiment was designed as a within-subjects repeated-measures design. The independent variables were the four adjustable game parameters: target size, grid size, distractor style, and feedback style. There were eight blocks in total, and six trials within each block. Each of the three target sizes appeared twice within each block, and the order of the target sizes was randomized subject to that constraint. Each trial lasted 20 seconds. The four distinct combinations of grid size and distractor style (i.e. 2x2 grid with moles only, 2x2 grid with moles and butterflies, 3x3 grid with moles, and 3x3 grid with moles and butterflies) were presented in the first four blocks in an order intended to facilitate learning (smaller to larger grid size, and no distractor to with-distractor). The second four blocks were then presented in the reverse order to provide a measure of counterbalancing. In the first half of the experiment, no feedback was provided to users, and feedback was then provided for the second half of the experiment. The design of the experiment is summarized in Table 1.

<table>
<thead>
<tr>
<th>Block</th>
<th>Grid Size</th>
<th>Distractor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2x2</td>
<td>M</td>
</tr>
<tr>
<td>2</td>
<td>2x2</td>
<td>MB</td>
</tr>
<tr>
<td>3</td>
<td>3x3</td>
<td>M</td>
</tr>
<tr>
<td>4</td>
<td>3x3</td>
<td>MB</td>
</tr>
<tr>
<td>5</td>
<td>3x3</td>
<td>MB</td>
</tr>
<tr>
<td>6</td>
<td>3x3</td>
<td>M</td>
</tr>
<tr>
<td>7</td>
<td>2x2</td>
<td>MB</td>
</tr>
<tr>
<td>8</td>
<td>2x2</td>
<td>M</td>
</tr>
</tbody>
</table>

Table 1. An example of the experimental design for a participant. Blocks 1 through 4 are games with feedback absent, and blocks 5 through 8 are games with feedback present. M stands for mole only games, and MB stands for games with both moles and butterflies.

2 distractor styles (distractor absent, distractor present) x 2 feedback styles (feedback absent, feedback present) x 2 repetitions = 1152 trials in total

The dependent variables were response time, and accuracy. Response time (RT) was defined as the time taken to hit a target after it appeared. Response accuracy was defined as the pixel deviation distance from the user’s hit to the centre of the target.

The experimenter also observed participants while they completed the study to examine how they interacted with the tablet. The experiment took approximately 60 minutes to complete. Participants were encouraged to take breaks between trials.

**Data Analysis**
Data was analyzed using Microsoft Excel, MathWorks MatLab, and IBM SPSS.

**RESULTS**
In this section, the results from the usability study will be presented. First, analyses of the game performance data will be examined using Fitts’ Law and speed-accuracy tradeoff approaches. Next, the ability of the game performance data to predict cognitive ability will be discussed.

**Fitts’ Law and Speed-Accuracy Tradeoff**
The following sections will analyse game performance in terms of Fitts’ Law, and in terms of speed-accuracy tradeoffs.

**Modelling Fitts’ Law Using Game Performance**
The game performance data was analysed using Fitts’ Law, which models the time needed to move to a target based on its distance and size [21]. This law is shown in Equation 1, where A is the distance from the user’s last hit to the target (pixels), and W is the width of the target (pixels). The constants a, and b are determined by plotting the user’s MT against the index of difficulty (ID) of the task (Equation 2).
Movement Time = a + b \log_2 \left( \frac{2A}{W} \right) \quad (1)

Index of Difficulty = \log_2 \left( \frac{2A}{W} \right) \quad (2)

The median RT for hitting a mole from all participants was fitted against the ID of the task using linear regression. There was no sign of any relationship (the correlation coefficient was around zero). This suggests that game performance on the tablet, as implemented in our study, does not constitute a Fitts’ Law task at the single trial level. One possible explanation is that it is difficult to assess whether users moved their finger from their last touch position, which causes the problem of an indefinite start point. For trials where distractors were potentially present it was clearly not a pure Fitts’ Law task since the overall RT included the decision time for deciding whether or not a target was present prior to moving to the target. Even for trials with targets only, the participant first had to determine which hole the mole had appeared in (positional uncertainty) prior to moving to it. Thus it seems likely that some combination of positional uncertainty of the target and ambiguity as to the starting position of the finger to be used in hitting the target (i.e. no use of a pointing device such as a mouse or stylus) meant that this did not qualify as a Fitts’ Law task. However, failure to follow Fitts’ Law cannot be attributed to the fact that the task was implemented on a tablet since in other research with soft keyboards, [21] showed that text entry with a stylus in their task did follow Fitts’ Law.

To illustrate that Fitts’ Law works on a general level, a repeated measures analysis of variance (RM ANOVA) was carried out with target size and grid size as the factors and median RT as the dependent measure. Significant main effects of both grid size ($F[1,23] = 4.667, p < 0.05$) and target size ($F[2,46] = 5.528, p < 0.05$) were found, and there was a significant interaction between target size and grid size, ($F[2,46], = 3.3031, p < 0.05$) (Figure 4). Response times were faster with the 2x2 grid and with larger target sizes, as would be predicted by Fitts’ Law. As seen in Figure 4, the significant interaction was due to participants hitting the medium target size (175 pixels) faster than they did for the other two target sizes (150 and 200 pixels) in the smaller 2x2 (but not the 3x3) grid. One possible explanation for this result may be that the middle sized target was easier to detect when it appeared in a hole in the 2x2 grid, although further research would be needed to demonstrate that this was in fact the case.

Modelling a Speed-Accuracy Tradeoff Using Game Performance

The game performance data was then analysed using a speed-accuracy tradeoff function, which examines the relationship between reaction time and accuracy [16]. In the context of the game, RT was measured as the time taken to hit a target after it has appeared, and accuracy was measured as the distance from the user’s touch to the centre of the target in pixels. This relationship was plotted for the entire data set (pooled across participants and conditions) and there was a linear fit ($R^2 = 0.335$), indicating a tradeoff between median RT and accuracy (Figure 5). As users took longer to respond, they tended to respond more accurately. When the data were pooled within participants so that median RT and accuracy were compared between the 24 participants, the correlation between median RT and accuracy (i.e. the speed-accuracy tradeoff) was 0.822. Thus, there was a strong tendency for some participants to be faster than others, but at the expense of accuracy as measured by distance in pixels between the point of touch and the centre of the target.

**Figure 4.** Graph depicting interaction effect between target size and grid size on RT.

**Figure 5.** Speed-accuracy tradeoff.

Standardized Performance Metrics

Accuracy and RT scores can be standardized across the entire sample by calculating z-scores for the data pooled across participants and conditions. The advantage of using this approach is that all the scores are then on a common scale. An
were significant effects of the covariates involving inhibition, back characters as the target size increased may have been more distracting for participants with lower inhibition ability scores.

Table 2. Standardized performance metric correlated with cognitive ability scores.

<table>
<thead>
<tr>
<th></th>
<th>Inhibition</th>
<th>Shifting</th>
<th>Updating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z(time)</td>
<td>-.257</td>
<td>.136</td>
<td>.275</td>
</tr>
<tr>
<td>-Z(accuracy)</td>
<td>.011</td>
<td>-.113</td>
<td>.220</td>
</tr>
<tr>
<td>-Z(accuracy)-Z(time)</td>
<td>-.600**</td>
<td>-.399*</td>
<td>-.354*</td>
</tr>
</tbody>
</table>

Absence of an asterisk next to a value in the table indicates that the value was not statistically significant (p > .05), **p < .001, *p < .05.

Overall Performance Score = \(-Z(\text{accuracy}) - Z(\text{time})\)  

The z-scores in Equation 3 are subtracted (reversed) because lower pixel deviations, and lower response times, indicate better performance. When plotting the Z(accuracy) scores against Z(time), a linear fit (R² = 0.335) was observed. This is the same linear fit observed in the speed-accuracy tradeoff for the unstandardised data, which is expected since these quantities are mathematically equivalent.

Next, to determine whether this new standardized metric is related to cognitive ability, a correlation analysis was performed (Table 2). There was a significant relationship (r = 0.60, r = 0.40, and r = 0.35 respectively) between the overall performance score, -Z(accuracy)-Z(time), and all three EFs. No similar relationships were found with Z(time) or Z(accuracy) each considered separately. This suggests that the overall performance score is more sensitive as a predictor of cognitive ability than either accuracy or time alone for this task. To illustrate this relationship, a Venn diagram depicts (approximately to scale) the variance shared between game performance scores for each participant and their corresponding inhibition and updating scores (Figure 6). The values were obtained using partial correlations. Shifting is not shown in Figure 6 because it had a low partial correlation with game performance (less than one percent). Thus the apparent correlation between game performance and shifting (Table 2) was due to the fact that shifting was correlated with inhibition, which in turn was correlated with game performance.

Game Performance Results

Next, to gain an understanding of the relationship between game performance and cognitive abilities, we carried out a RM ANOVA. The four game parameters constituted the four factors in three analyses, each with a different one of the three EFs as the single covariate. The median -Z(accuracy)-Z(time) transformation for the response of hitting a mole (i.e., the overall performance score) was used as the dependent variable for these analyses.

In the interaction between grid size and feedback, there were significant effects of the covariates involving inhibition, (F[1,20] = 9.68, r = 0.57, and shifting, (F[1,20] = 6.29, r = 0.49). Contrasts revealed that -Z(accuracy)-Z(time) values with a smaller grid size of 2x2 were significantly higher than a larger grid size of 3x3 as can be seen in figures 7, and 8, where the impact of inhibition, and shifting, ability, respectively is also shown.

Relationship Between EF and Game Performance

To determine how the different combinations of game parameter are related to inhibition scores, correlations were calculated between the performance metric for hitting a mole and inhibition ability as measured in the Stroop task. Correlations were assessed across the twenty-four combinations of target size, grid size, feedback style, and distractor style. The correlation between inhibition ability and the performance metric was statistically significant for 22 of the 24 combinations of game parameters that were assessed.

The correlations between inhibition and the performance metric for the four combinations of distractor style X feedback style have been depicted in Figures 9, 10, 11, and 12 (the error bars represent one standard error). In Figure 12, when both distractors and feedback are present, it is apparent that correlation increased with target size in the smaller grid size (2x2) condition. In contrast, when the distractor was present but feedback was absent, the correlation decreased with target size in games with a 2x2 grid (Figure 12). In games with distractors present and the 2x2 grid, the increasing size of feedback characters as the target size increased may have been more distracting for participants with lower inhibition ability. Overall, the largest correlation was when distractors were
present, and feedback was absent, in the 2x2 grid with 150 px
targets (r = -0.69) (Figure 12).

The relatively strong correlations between inhibition ability and performance metric indicate that the whack-a-mole game is requiring inhibition ability. The relationship occurs broadly across all the game parameters, and when the analysis is restricted to particular game parameters some significant correlations still occurred. While the strongest correlation occurred with the 2x2 grid, the medium target size (175 pixels) and in the presence of distractors, significant correlations were also observed for the 3x3 grid (medium and large targets) when there were no distractors. Thus it seems that inhibitory ability may have been required not only to avoid hitting butterflies but also to maintain general task focus (consistent with [13]'s view of inhibition as being part of a common or shared EF).
DISCUSSION

Our overall finding was that the current version of the game had a moderate relationship with inhibition ability, with a correlation of 0.60 with the overall performance score. There were also significant, but smaller, correlations with the other two cognitive abilities that we examined (shifting and updating). Thus, the whack-a-mole game was measuring all three cognitive abilities that we examined to some extent, although it was not a pure measure of any of them.

Our first research question concerned the relevance of Fitts’ Law and the speed-accuracy tradeoff to the game. We found no evidence that the whack-a-mole game, as implemented, could be modelled as a Fitts’ Law task at the individual trial level.

However, median RT was faster with the smaller grid size and with larger targets. Thus, at a general level, the expected Fitts’ Law relationship between RT and task difficulty held. For speed-accuracy, we found a strong correlation between response speed and accuracy between the participants, $r > 0.8$. There was also a significant correlation between speed and accuracy over the entire data set ($r = 0.58$), showing that the speed-accuracy tradeoff occurred within, as well as between, individuals.

Our second research question was answered in the affirmative. In addition to showing a strong effect of a speed-accuracy tradeoff in game performance, we found that an overall performance score that combined z-scores for speed and accuracy was more sensitive to differences in the three cognitive abilities evaluated in this study. The overall performance score was significantly correlated with all three cognitive abilities, whereas neither RT nor accuracy alone were significantly correlated with any of the three cognitive abilities. The implication of this result is that the novel overall performance score (which was introduced in this paper) can be used in the design of the next iteration of the whack-a-mole game to adjust the difficulty level based on a user’s performance and ability. For instance, if a user takes longer to respond in hitting a mole, and is not very accurate in doing so, the game can be calibrated to automatically increase the target size area to minimize the difficulty and increase the probability of the user hitting the mole. This adaptive gameplay mechanism based on the user’s overall performance metric can also reduce associated learning curves and may motivate the user to continue playing the game to keep cognitively active.

Finally, our third research question was focused on the ability of the whack-a-mole game to predict cognitive ability. We found a strong relationship between EF and gameplay using the overall performance score that we developed. Thus, our third hypothesis was answered in the affirmative as we have demonstrated that the game is related to inhibition ability, which is an important component of cognitive ability. An implication of this finding is that the whack-a-mole game has the potential to assist clinicians in assessing a patient’s cognitive status. In [13]’s latest model of central EFs, they propose that the EF of inhibition is not a separate entity, but is instead part of a “common EF”, which is an underlying EF required for all EFs (i.e. shifting and updating abilities). Since significant correlations between overall game performance and cognitive ability were found for all three of the EFs tested (inhibition, shifting, and updating), whack-a-mole game performance may be related to the common EF rather than to a specific EF such as inhibition.

Through the analysis of the game performance data, we have gained a better understanding of how different game parameters and metrics are influencing RT and accuracy. In order to provide a uniform gaming experience to all users to account for considerations such as their cognitive ability, dexterity, and motor skills, we propose adjusting the size of the target and allotted time to hit each target based on the user’s performance. Users would probably not want to play the game if they weren’t able to hit any of the moles nor would they want to if the game were so easy that they could hit all the moles (and avoid butterflies) without expending any effort. In future there could be an additional setting in the game where users could adjust the difficulty of game play. For instance, if the setting was moved to an easier position, the target area in which a touch was registered as a hit could be expanded, or the time over which a mole can be hit could be lengthened. Another strategy would be to automatically adjust the target tolerance in time and space (i.e. game difficulty) based on the z-scores for RT and accuracy. A person with a high z-score for RT (i.e. someone responding relatively slowly) might be given slower moving moles so that there was more time to respond, while a person with high z-score for error (pixel deviation) could be given a wider target area to work with (Figure 13).

By applying this adaptive game tolerance mechanism, we hope to provide an enriched gaming experience for all users while gaining insight into their cognitive well-being. This will also provide a unique experience for each user based on their individual abilities, thereby encouraging them to play the game frequently.

Figure 13. Diagram displaying the tolerance in pixels for hitting the target. This tolerance area can increase or decrease depending on the user’s performance. The colored squares of increasing size represent increased areas of target tolerance.
Recommendations
Based on the findings from this research, we have proposed the following guidelines for the design of game-based cognitive assessments for the elderly. These guidelines should apply to game-based cognitive assessment in general, and not just the whack-a-mole game that is the focus of this paper.

1. Conduct a user requirements analysis with your end-users to determine the needs and constraints of your user population to inform the prototyping process.

2. Design game components to reflect psychometric properties of existing neuro-psychological tasks. For instance, in designing a game that intends to measure updating ability, researchers should look to validated updating tasks such as the Trail Making Task, and should incorporate relevant features of those tasks [19], consistent with the goals and aesthetics of the game.

3. Adjust game characteristics based on the capabilities of the user so that a satisfying level of game play can be achieved.

4. If the game involves speed and accuracy use the overall performance metric presented in this paper to provide a global level of performance.

5. Carry out usability testing to identify problems with the game that may be a threat to its validity as a measure of the cognitive ability (or abilities) of interest.

In the usability testing that we carried out with the whack-a-mole game, its validity as a measure of inhibition ability varied based on the particular game parameters that were used. Tentative recommendations derived from the usability findings are presented below:

1. Use a distractor if the person can handle the added complexity of dealing with the distractor.

2. If a person can handle playing the game with distractors and without feedback, then grid size and target size can be adjusted to suit their visual and motor capabilities.

3. If no distractor is used then it would be preferable to use no feedback, and a 3x3 grid.

4. If a person can handle the added complexity of distractors, but needs feedback as well, then use the 3x3 grid.

The design of the size and shape of feedback, when it is provided, is a question for future research.

Limitations
This study consisted of participants recruited from a university environment. Our goal in this study was to evaluate the basic properties of the game that we had developed and thus assessment of the impact of the game on elderly users was left as a task for future research.

CONCLUSION
This paper advocates the use of games to assess cognitive status in order to address current issues with conventional assessments. We illustrated this approach with the whack-a-mole game that we have developed, and we demonstrated the impact of various cognitive abilities and game parameters on game performance. Results were reported from an initial usability study completed with a healthy sample population. The results from this study will be used as a baseline for game-based cognitive assessment against which corresponding performance by an elderly population may be compared.

This paper has demonstrated that the performance data from the whack-a-game can be used in the calculation of a standardized performance metric based on summation/subtraction of standardized speed and accuracy measures. This overall performance score was shown to be more predictive of inhibition ability than either speed or accuracy by itself. While game performance was most related to inhibition ability, significant relationships were also found with shifting and updating ability. Thus we succeeded in developing a game that is predictive of inhibition ability, and that has the potential to be a useful measure of executive functioning.

In future, we plan to modify the game through further programmatic testing in order to develop versions that can accurately measure common EF as well as specific EFs such as shifting and updating. We hypothesize that the whack-a-mole game may be a useful platform on which to build assessments of different types of EF by varying game characteristics in appropriate ways, while still retaining the essential spirit of the whack-a-mole task.

As an extension of this current pilot study, we plan to conduct another usability study with healthy, elderly adults, and eventually with groups of elderly adults in hospital settings. From this research, we hope to learn how the elderly interact with touch-based devices and to determine if the standardized performance metric obtained with the whack-a-mole game is still a good predictor of inhibition ability.

Our ultimate goal is to create a generalized and modified version of the whack-a-mole game that can provide a sensitive test of executive functioning for a wide range of people in both clinical and community settings. We hope that the methodologies introduced in this paper may also be applicable to other serious gaming applications where it is desirable to integrate speed and accuracy into overall measures of performance and where the impact of game parameters needs to be calibrated and aligned to different levels of player ability.

Our ultimate goal is to create a game that can provide a sensitive test of executive functioning for a wide range of people in both clinical and community settings. Future research should compare the whack-a-mole game with other games to see if it is a particularly useful game for executive function assessment, or if in fact executive functioning is required in a wide range of games.

We hope that the methodologies introduced in this paper may also be applicable to other serious gaming applications where it is desirable to integrate speed and accuracy into overall measures of performance and where the impact of game parameters needs to be calibrated and aligned to different levels of player ability.
REFERENCES