EXAMINING THE ROLE OF REASONING AND WORKING MEMORY IN PREDICTING CASUAL GAME PERFORMANCE ACROSS EXTENDED GAMEPLAY

BY

MICHAEL BENJAMIN KRANZ

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Arts in Psychology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Master's Committee:

Professor Arthur Kramer, Chair
Professor Ranxiao Frances Wang

Abstract

The variety and availability of casual video games presents an exciting opportunity for applications such as cognitive training. Casual games have been associated with fluid abilities such as working memory (WM) and reasoning, but the importance of these cognitive constructs in predicting performance may change across extended gameplay and vary with game structure. The current investigation examined the relationship between cognitive abilities and casual game performance over time by analyzing first and final session performance over 4-5 weeks of game play. We focused on two groups of subjects who played different types of casual games previously shown to relate to WM and reasoning when played for a single session: 1) puzzle-based games played adaptively across sessions and 2) speeded switching games played non-adaptively across sessions. Reasoning uniquely predicted first session casual game scores for both groups and accounted for much of the relationship with WM. Furthermore, over time, WM became uniquely important for predicting casual game performance for the adaptive games but not for the non-adaptive games. These results extend the burgeoning literature on cognitive abilities involved in video games by elucidating the differential relationships of fluid abilities across game type and extended play. More broadly, the current study illustrates the usefulness of using multiple cognitive measures in predicting performance and provides potential directions for game-based cognitive training research.

Table of Contents

CHAPTER 1: INTRODUCTION

Video game websites (e.g., miniclip.com, addictinggames.com) offer hundreds of games across a variety of genres. These freely available, highly accessible, easy-to-learn games—often called casual games—provide a leisurely, yet cognitively engaging, activity even for people with limited video game experience. Whether maneuvering around obstacles to reach a door or exit, quickly collecting coins, or shooting down enemy ships, casual games challenge players' cognitive abilities in a variety of ways. Can we harness this potential for cognitive applications such as cognitive training? Addressing this question requires a deeper understanding of the cognitive processes involved in casual games over extended gameplay.

In a recent study, several of these freely available, web-based casual games were quantitatively evaluated in terms of their relationship with cognitive abilities (Baniqued et al., 2013). Specifically, participants completed a battery of neurocognitive/neuropsychological tests and played several casual games for one short period of time (i.e., 20 minutes per game) while instructed to achieve the highest score or level. Performance on several games correlated with tests of working memory (WM), which relates to actively maintaining and manipulating information in mind (Baddeley, 1992), and reasoning, which relates to solving novel problems (also called fluid intelligence; Cattell 1987). Although informative, this evaluation did not assess the relationships over a longer period of time such as extended gameplay over several sessions, which is common in both everyday use (http://www.casualgamesassociation.org) and in cognitive training research (Baniqued et al., 2014, Lee et al., 2015, Boot et al., 2010, Basak et al., 2008, Colom et al. 2012, Owen et al. 2010). Given that the relationship between cognitive abilities and games (or tasks) motivates cognitive training research design (Baniqued et al., 2013, Jaeggi et al., 2010), evaluating how these relationships change after extended play for different types of games is important.

In one recent framework of complex skill acquisition, it is thought that individuals first form strategies in an effortful and error-prone process, and that performance is largely associated with fluid abilities (Fleishman et al., 1972; Woltz, 1988; Ackerman, 1988; Ackerman et al., 2005a) such as working memory and reasoning. After initial learning, the relationship with fluid abilities becomes dependent on task consistency (Ackerman, 1988; Ackerman et al., 2005a). In consistent task environments, individuals tune and automatize strategies over time, leading to more efficient task performance; the association between task performance and fluid abilities decreases, while the association between task performance and the speed of strategy deployment (i.e., processing speed) increases. In contrast, in inconsistent or variable task environments, individuals must update strategies in response to changing task components, and task performance remains associated with fluid abilities over time. Thus, if the goal of a training program is to improve fluid abilities, inconsistent or variable cognitive training environments may be desirable or more optimal. This framework has been commonly applied to understand complex skill acquisition on a range of complex tasks from short-term learning (Zhang et al., 2007), computer programming (Schute and Kyllonen, 1990), to commercial brain training games (Quiroga et al., 2009,2011,2015; Ackerman et al., 2011). However, applying the framework to casual games is more complex because many casual games are adaptive, where difficulty increases as the game progresses. For some games, difficulty increases with more complex obstacles and unique relationships to learn on each new level. For these particular games, players often start on the highest level reached from the previous sessions. This game structure requires players to learn novel rules and skills across multiple sessions of gameplay (adaptive across sessions). In contrast, other casual games involve repeating the same levels, with each session of gameplay starting at the same difficulty level (non-adaptive across sessions), but with each attempt involving increases in difficulty until performance limits are reached (e.g., a "game over"). In this study, we examine whether performance in these different types of casual games is differentially predicted by fluid abilities.

To measure fluid abilities, training studies have primarily used reasoning tasks (e.g., Ackerman, 1988) and working memory (WM) tasks (e.g., Woltz, 1988, Kyllonen and Stevens, 1990, Schute and Kyllonen, 1990). Starting with Kyllonen and Christal (1990), research has consistently identified a strong association between WM and reasoning (e.g., Engle et al., 1999, Colom et al., 2003, Conway et al., 2003, Ackerman et al., 2005b). In one recent meta-analysis, WM and reasoning shared approximately 50% of their variance across studies (Kane et al., 2005). Due in part to this robust relationship, many cognitive training paradigms seek to improve fluid abilities, as measured with reasoning tasks, using tasks that tap working memory ability—as WM is thought to reflect a more basic and fundamental process underlying reasoning ability (see Jaeggi et al., 2010; Colom et al., 2010).

Despite their strong relationship, the two constructs are not considered isomorphic (Kane et al., 2005, Ackerman et al., 2005b). Indeed, several areas of research illustrate the importance of the non-overlapping variance of WM and reasoning. In one domain—academics—higher reasoning and WM abilities uniquely predicted literacy and mathematic achievement test scores in children (Alloway and Alloway, 2010; Dumontheil and Klingberg, 2012). In the laboratory, WM and reasoning uniquely predicted performance on problem solving tasks placing high demand on WM (3-8 disk Tower of Hanoi problems), but only reasoning predicted performance on problem solving tasks lower in WM demand (2-5 move Tower of London problems; Zook et al., 2004). Moreover, increasing the demand on WM tasks seemed to have no effect on the relationship between reasoning and working memory (Salthouse, 2008, Unsworth and Engle, 2005).

The current study assesses the relationship between abilities implicated in skill acquisition and casual game performance across extensive gameplay . This inquiry will shed light on the cognitive components of casual games and help us understand how a leisure time activity pursued by an increasing number of individuals is associated with aspects of cognition. A better understanding of these associations may ultimately lead to more informed use of casual games for cognitive training research and more generally, better informed applications of computer-based games for other interventions or real-world situations.

To this end, we leveraged data from two groups of participants who played casual games over multiple sessions as part of a cognitive training study (see Baniqued et al. 2014). Each game was selected based on their correlation with WM and reasoning abilities based on a single session of game play (Baniqued et al., 2012, 2014). Although the nature and magnitude of the game vs. WM and reasoning relationships did not differ between the two groups at baseline, the groups differed in game structure—a distinction that could become important when examining cognitive ability relationships across extended gameplay. In one group, players solved novel and increasingly challenging problems in order to progress to each new level. At each subsequent training session, players in this group started on the last level they had reached in the previous session (i.e., puzzle-based games played adaptively across sessions). In the second group, players quickly switched attention to different game components (e.g., falling coins or numbers) in order to reach the highest score possible, with increasing components or switching demands at each new level. However, unlike the first group, players started at the same level at each session and after each "game over" or failed attempt (i.e., speeded switching games played non-adaptively across sessions).

We used baseline cognitive assessments and game data from participants' first and final training sessions and to investigate how game performance-cognitive ability relationships depend on game type and time or experience. We then examined the unique predictive ability amongst the fluid abilities commonly used in skill acquisition (WM and reasoning) by running a series of step-wise regression models with both working memory and reasoning as predictors. Thirdly, we explored the unique predictive ability of perceptual speed, given its importance in later stages of complex skill acquisition (e.g., Ackerman 1988; see Ackerman and and Cianciolo 2000, Experiment 3 for similar regression analyses using first and final session complex task performance metrics).

CHAPTER 2: MATERIALS AND METHODS

In these analyses, we used a subset of participants from a cognitive training study that tested the effects of casual game training on cognitive performance (Baniqued et al., 2014). The data included in these analyses are from participants in the "WM-Reas 1" (non-adaptive) and "WM-Reas 2" (adaptive) training groups that played working memory and reasoning games. Brief descriptions of the procedure for this subset of participants will be presented subsequently. For a detailed description of the procedure and entire study details, see Baniqued et al., 2014 and the supplemental material located at http://lbc.beckman.illinois.edu/pdfs/CasualGames_ SuppMethods.pdf.

2.1. Participants

Participants were recruited from online postings, flyers and newspaper advertisements. Respondents were screened with several criteria including a prerequisite of 3 hours or less of video and board game play per week in the last 6 months. All participants signed a consent form approved by the University of Illinois Institutional Review Board. Participants who completed the study were paid \$15 an hour, and if they dropped out at any point during the study they were paid \$7.50 an hour for the time that they had completed. Table 1 shows the demographic information for participants included in these analyses.

Table 1. Participant demographics for two game groups

Group	*N* excluded due to video game play>3 hours per week	*N* excluded due to casual games played outside training	*N* included in analyses	Males	Age
Non-Adaptive	3	2	48	15	21.29 (2.20)
Adaptive	1	5	45	15	21.18 (2.53)

Note. N = Number of participants, gender and age values are for the participants included in analyses

2.2. Procedure

After group assignment, participants underwent four testing sessions consisting of three cognitive sessions and one magnetic resonance imaging (MRI) testing session. The neuroimaging data will not be discussed in this paper. The four testing sessions and tasks within these sessions were administered in a fixed order. After baseline testing, participants completed ten video game training sessions at a rate of two to three times per week. Each game was played for 20 minutes per session, with each session lasting around 1.5 hours. At the end of training, re-testing was completed to assess transfer of cognitive task skills; transfer analyses are reported elsewhere and not the focus of the current study (Baniqued et al., 2014). The current data analyses excluded participants based on video game play outside the laboratory based on the same criteria as Baniqued et al. 2014 (see Table 1).

The cognitive measures used as predictors in this study are from the tasks administered during the baseline testing sessions, while the casual game scores are derived from casual game performance in the first and final training sessions. As the focus was on beginning and end performance, the other training sessions (sessions 2-9) were not used but correlations with these sessions are included in the supplemental analysis section.

Figure 1 summarizes the measures used in the longitudinal design of the casual game project.

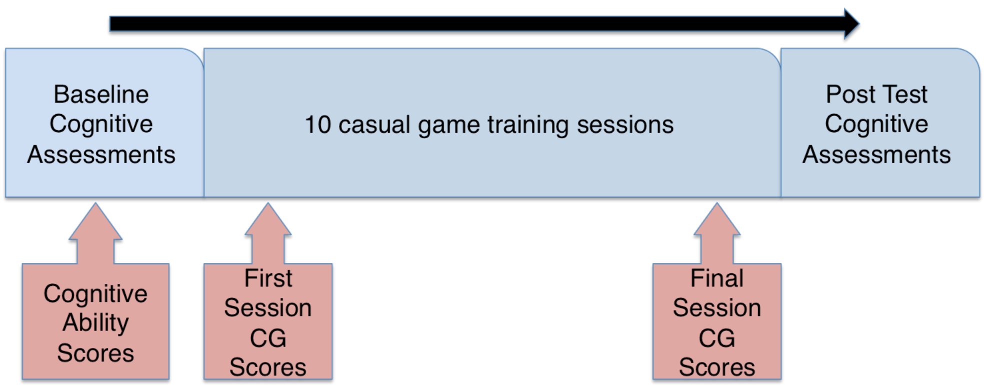


Figure 1. General procedure for the entire casual game project (in blue) and the metrics for the current primary analysis (in red). Note that the post testing cognitive assessments were not used since the focus of the current study was on predicting casual game performance across time using *pre-existing* cognitive abilities.

2.3. Baseline Cognitive Tasks

Tasks administered during baseline testing were divided into three categories: reasoning, working memory, and perceptual speed. Below are descriptions of each task administered, with more detailed task descriptions in the supplemental methods of the published training report (Baniqued et al., 2014).

2.3.1 Reasoning

All tasks except Matrix Reasoning were taken from the Virginia Cognitive Aging Project (VCAP; see Salthouse and Ferrer-Caja, 2003; Salthouse, 2004, 2005, 2008). Matrix reasoning was based on Crone et al., 2009 and performed within an MRI machine.

Furthermore, all tasks were completed on a computer with the exception of the Shipley Abstract test, which was administered with paper and pencil.

Form Boards (Ekstrom et al., 1976). Participants choose shapes to exactly fill the area of a bigger shape on a computer. Participants were given 8 minutes to complete as many of the problems as possible. The dependent measure was total correct problems.

Letter Sets (Ekstrom et al., 1976). Participants are given five patterns of letter strings and choose the string that does not match the other four strings. Participants were given 10 minutes to complete as many of the different letter sets as possible. The dependent measure was total correct problems.

Paper Folding (Ekstrom, French, Harman & Dermen, 1976). Participants identify the resulting pattern of holes from a sequence of folds and a punch through the folded sheet. Participants were given 10 minutes to complete as many of the problems as possible. The dependent measure was total correct problems.

Spatial Relations (Bennett, Seashore & Wesman, 1997). Participants choose a two dimensional unfolded object that will match a three-dimensional folded object. Participants were given 10 minutes to complete as many as possible. The dependent measure was total correct problems.

Shipley Abstract (Zachary & Shipley, 1986). Participants fill in a missing item(s) to complete progressive sequences of numbers, letters, and words written on one sheet of paper. Participants were instructed to attempt to complete all 20 sequences in 5 minutes and told that if you get stuck on one sequence, they may skip it and come back to it. The dependent measure was total correct problems.

Matrix Reasoning (Crone et al., 2009; Raven, 1962). Participants viewed a 3 x 3 matrix containing patterns along the rows and columns in all but once cell and chose an item that best completes the pattern. Trials were divided based on the amount of relational integration needed to solve the problem. There were 30 control trials in which no integration was required and 30 reasoning trials in which successful completion required integrated patterns across the cells. Participants had 12 seconds to solve each problem. The dependent measure was the mean accuracy of the reasoning trials.

2.3.2 Working Memory

Visual Short Term Memory (Luck & Vogel, 1997). A probe array of four shapes briefly appeared on the screen. After a delay, a target shape appeared and participants had to decide whether this stimulus was in the probe array. The experiment consisted of three blocks with stimuli varying only in color on the first block, only in shape on the second block, and the conjunctions of both color and shape on the third block. Each block consisted of 60 trials. The dependent measure was overall accuracy.

Spatial Working Memory (Erickson et al., 2011, Greenwood et al. 2005). Each trial consisted of a probe configuration of one, two, or three black dots on the screen. After a brief delay, a red target dot appeared, and participants were instructed to determine if the red dot was in the same position as one of the black probe dots in that trial. There were 40 trials (20 same and 20 different) per condition randomly varying in dot locations and condition. The dependent measure was overall accuracy.

N-Back (Kirchner, 1958; Kane et al., 2007). For three blocks of trials, participants viewed as sequence of centrally presented letters. For each letter, participants were instructed to determine if the current letter matched the previous letter (first block), two letters back (second block), or three letters back (third block). The most demanding condition, the 3 back condition, was used as a metric of working memory performance on this task. There were five 20 letter sequences per condition for a total of 100 trials (25 target trials for all conditions and 10 lure trials for the 2 and 3 back) per condition. The dependent measure was the combined accuracy across the two and three back conditions.

Running Span (Broadway and Engle 2010). For each trial, a sequence of letters were rapidly presented on the screen. After the list was presented, participants were told to recall the last n (ex. 2,3,or 4) items on the screen.

Symmetry span (Unsworth et al., 2005; Redick et al. 2012). A sequence of red squares within a matrix was presented while participants judged whether two figures were symmetrical in between presentation of these red squares. Participants were instructed to recall in order the locations of the previously presented sequence. 12 trials (3 of list lengths 2,3,4, and 5).

2.3.3 Perceptual Speed

All tasks are from VCAP and completed with paper and pencil.

Digit Symbol Coding (Wechsler, 1997a). Participants write the corresponding symbol for each digit using a coding system for reference. There were 9 symbols and corresponding digits. Participants had 2 minutes to fill in as many symbols for each indicated digit as possible. The total number of correct symbols written was used as the dependent measure.

Pattern Comparison (Salthouse & Babcock, 1991). Participants determine whether a pair of patterns is the same or different. Participants had 30 seconds to match as many pattern pairs as possible in one set and completed 2 sets of patterns. The average total correct across the two sets was used as the dependent measure.

Letter Comparison (Salthouse & Babcock, 1991). Same as pattern comparison except with letter strings. Participants had 30 seconds to match as many letter sequence pairs as possible in one set and completed 2 sets of this task. The average total correct across the two sets was used as the dependent measure.

2.4. Composite Cognitive Scores

Each baseline cognitive score was standardized and averaged together with the other cognitive scores in the same construct (based on the model-based grouping listed above for Reasoning, Working Memory, Perceptual Speed). Although WM and reasoning were the main focus of analyses—given that casual games were selected based on their associations with WM and reasoning, the relationship with perceptual speed scores were also analyzed given the construct's previous implications in skill acquisition (e.g., Ackerman 1988).

Although there is much consensus for reasoning and perceptual speed as general constructs (Salthouse and Ferrer-Caja, 2003; Salthouse, 2004, 2005, 2008), working memory has been operationalized quite differently. For example, researchers who emphasize WM capacity as attention control use a battery of span tasks to compile WM scores (e.g., Engle et al., 1999) while others have used visual change detection paradigms to define WM capacity as the focus of attention (e.g., Cowan, 2010; Luck and Vogel, 1997). Different types of WM tasks display unique variance when predicting other complex cognitive abilities such as reasoning/fluid intelligence (Kane et al., 2007; Unsworth et al., 2014). In the current study, we use a combination of different WM tasks in order to create a measure of general WM ability (see Wilhelm et al., 2013; Schmiedek et al., 2014).

2.5. Casual Games Used For Training

For both groups, four casual games previously associated with WM and reasoning (Baniqued et al., 2013, 2014) were each played for 20 minutes in a pseudo-random order for each of the 10 training sessions. The original training study did not explicitly manipulate game type and adaptive-ness for the two groups; their groupings for the purposes of this study are defined post-hoc. For brevity, we refer to the puzzle-based games played adaptively across sessions as the *adaptive games* and the speeded switching games played non-adaptively across sessions as the *non-adaptive games*.

For the puzzle-based adaptive group, common to each game was the goal to complete as many levels or stages as possible within the 20 minute session. Participants needed to complete one level before advancing to the next. At the end of the 20 minutes, the current level was recorded as the high level for that session and used as the performance metric. An experimenter recorded this level information and the corresponding game code to type in for the next session, which started at the previous session's high level. One game was left out of analyses as the majority of participants completed all the levels before the end of the training sessions (Aengie Quest). After all data was collected and game metrics were entered, video recordings for each game in each session were reviewed to ensure that the correct procedure was followed. That is, each participant must start on the level they were attempting from the previous session. If a subject did not start on the correct level (e.g., started on the first level instead of a higher level from a previous session), the data for that game was not included in calculating either CG score composite measure (see section 1.4.3 for composite CG score explanation).

For the non-adaptive group, one game was an adaptive game and was thus left out of analyses (Silversphere). For each session in the remaining three games, participants started over at the beginning of the game, or the game was structured such that within a session, a participant would have several attempts with each attempt starting from the first level of difficulty. To obtain the performance metric for each game, video recordings of each game in each session were watched and the score of each game attempt was collected. If no video recording was obtained for either the first or final training session of a game, that game was excluded for both first and final CG composite scores (see section 1.4.3 for composite CG score explanation).

Two subjects from both the adaptive and the non-adaptive groups were excluded from analyses because 2 out of 3 games had excluded or missing data for first or final session scores.

Below are descriptions of the games used in the current study. Figure 2 provides screen shots from the training games.

2.5.1. Adaptive Games

Silversphere (miniclip.com). Move a sphere from the starting position to a blue vortex while avoiding falling off the platform. On the platform, various objects block the path or serve to provide help in creating a path. These objects have different features and may be needed in various combinations to complete the goal of getting to the exit.

Blockdrop (miniclip.com). Move around a gem on three-dimensional blocks to remove all blocks except the checkered block. Unique block arrangements are presented in each level.

Gude Balls (bigfishgames.com). Explode all plates by filling a plate with four of the same colored balls and switching balls to other plates to complete the level and advance to the next level.

For all three games, the performance metric was the highest level reached at the end of the twenty minutes of gameplay.

2.5.2. Non-Adaptive Games

Digital Switch (miniclip.com). In the main game, participants must collect falling colored coins by lining up the colored digibot switches with the correct color. After a game was over (i.e., players do not reach a certain achievement level), players start on level 1. For each level, the number of coins to be collected increases by 5 coins. Players increase their score by collecting these coins. Highest score achieved was the metric used.

Two Three (armorgames.com). Participants play as a tank and must shoot down rapidly presented numbers by pointing their tank at them with the mouse and subtracting the presented numbers down to exactly 0 using units of 2 and 3 to earn points and increase in level. These subtractions are achieved by typing in 2 and 3 on the keyboard. If a number is not correctly subtracted down to exactly 0, it hits the player's tank, and the tank moves up the game screen. When the tank reaches the top of the game screen, the attempt is over. Participants then restart the game at the beginning (level 1 and 0 points). Highest score achieved was the metric used.

Sushi Go Round (miniclip.com). Participants must serve a certain number of customers and obtain a certain amount of money in the allotted time by learning and preparing different recipes correctly, cleaning tables, and ordering ingredients. If these goals are achieved within the allotted time, players keep their money and advance to the next day. If players do not achieve these goals, the game is restarted on the first day with no money. Highest amount of money collected was the metric used.

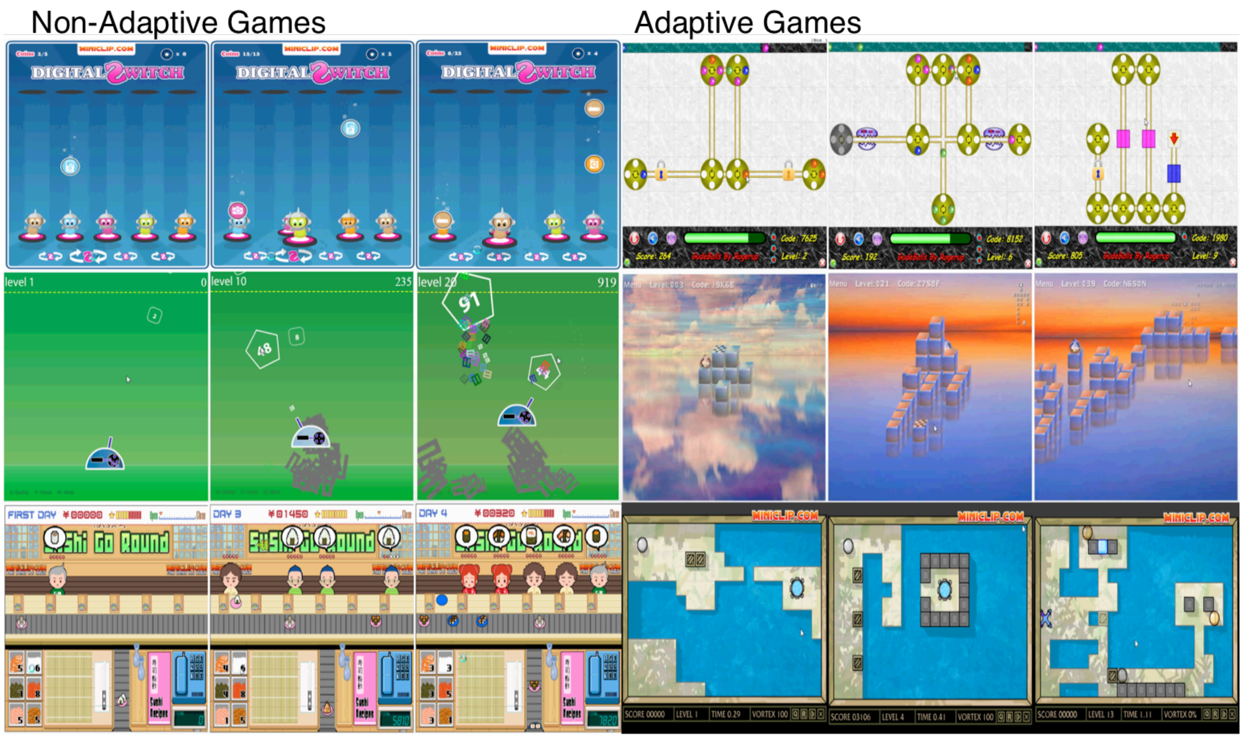


Figure 2. Screenshots illustrating the non-adaptive games on the left (from top to bottom: Digital Switch, Two Three, Sushi Go Round) and adaptive games on the right (from top to bottom: Gude Balls, Block Drop, Silversphere). Within each game, difficulty increases from left to right. For the adaptive group, difficulty increased throughout training with each new level. For the non-adaptive games, participants worked their way up difficulty levels within one attempt but started on the same difficulty levels across attempts and sessions.

2.6. Casual Game First and Final Session Composite Scores

Training session 1 game metrics were standardized and averaged together to create a composite first session CG score. Training session 10 game metrics were standardized and averaged together to create a composite score of final session CG score.

For the puzzle-based adaptive group, CG final session scores were created using the highest level reached after the 10th session of Gude Balls, Silversphere, and Block Drop.

CHAPTER 3: RESULTS

3.1 Descriptive Statistics

Descriptive statistics for each cognitive task measure and casual game measure are reported in Table 2 and Table 3, respectively. The cognitive measures used in these analyses are from the baseline testing sessions only, while the casual game measures are from the training sessions completed after baseline cognitive testing. The two groups did not significantly differ on any of the individual measures or composite scores, according to an independent samples t-test (all ps>.05). Participants achieved significantly higher scores after training (final session scores) compared to first session scores (all ps<.001).

Table 2. *Descriptive statistics for the pre-test cognitive assessment measures*

Cognitive Ability	Task	Measure	Adaptive *M* (*SD*)	Non-Adaptive *M* (*SD*)	Group Differences
Reasoning	Matrix Reasoning	% accuracy	78.59 (9.28)	79.86 (8.48)	$t(88) = 0.73$, $p = .47$
Reasoning	Form Boards	total correct	9.8 (3.93)	9.6 (4.35)	$t(88) = -0.42$, $p = .68$
Reasoning	Paper Folding	total correct	8.84 (1.97)	8.19 (2.36)	$t(88) = -1.65$, $p = .1$
Reasoning	Spatial Relations	total correct	12.36 (4.13)	11.77 (4.34)	$t(88) = -0.85$, $p = .39$
Reasoning	Letter Sets	total correct	12.56 (1.63)	12.35 (1.78)	$t(88) = -0.72$, $p = .47$
Reasoning	Shipley Abstract	total correct	15.33 (2.27)	15.81 (2.16)	$t(88) = 1.23$, $p = .22$
Working Memory	SPWM	% accuracy	.87 (.07)	.88 (.07)	$t(86) = 0.87$, $p = .39$
Working Memory	Nback	% accuracy	.86 (.09)	.88 (.06)	$t(86) = 1.06$, $p = .29$
Working Memory	VSTM	% accuracy	.81 (.06)	.8 (.06)	$t(88) = -0.8$, $p = .42$
Working Memory	Running Span	total correct	22.49 (5.57)	21.79 (5.36)	$t(87) = -0.22$, $p = .83$
Working Memory	Symmetry Span*	total correct	18.89 (6.71)	16.92 (8.76)	$t(65) = -0.8$, $p = .43$

Table 2. *(continued)*

Perceptual Speed	Pattern Comp	total correct	20.74 (3.4)	21.52 (4)	$t(88) = 0.96$, $p = .34$
Perceptual Speed	Letter Comp	total correct	12.42 (2.32)	13.22 (2.48)	$t(88) = 1.49$, $p = .14$
Perceptual Speed	Digit Symbol Coding	total correct	90.76 (13.28)	93.13 (14.07)	$t(88) = 0.89$, $p = .38$
Composite	Reasoning	std average	0.02 (0.65)	-0.02 (0.72)	$t(88) = -0.4$, $p = .69$
Composite	Working Memory	std average	0 (0.53)	0 (0.59)	$t(88) = 0.07$, $p = .95$
Composite	Perceptual Speed	std average	-0.12 (0.71)	0.11 (0.86)	$t(88) = 1.41$, $p = .16$

Note. M=Mean, *SD*=Standard Deviation; *Only 25 participants completed Symmetry Span for the Non-Adaptive group as this measure was added half way through data collection.

Table 3. *Casual Game Achievement Descriptive Statistics*

Group	Games	First Session *M* (*SD*)	Final Session *M* (*SD*)
Adaptive	Silversphere	8.96 (2)	19.41 (3.24)
	Block Drop	16.41 (3.38)	52.59 (8.42)
	Aengie Quest	8.26 (2.08)	19.24 (3.57)
	Gude Balls	4.46 (1.19)	14.76 (2.66)
Non-Adaptive	Two Three	535.38 (186.47)	1079.32 (225.23)
	Digi Switch	7078.3 (2279.85)	14410.21 (3782.55)
	Sushi Go Round	2807.5 (1110.53)	6637.29 (794.25)

Note. *M* = Mean, *SD* = Standard Deviation

3.2 Correlations between Casual Game and Cognitive Scores

First we calculated bivariate correlations of all baseline cognitive and session CG scores used in subsequent regression analyses (Table 4). Significance values shown in Table 4 are uncorrected for multiple comparisons. A significant relationship between both fluid abilities (WM and reasoning scores), and both first and final session CG scores were observed for both adaptive and non-adaptive groups. Importantly, there was no evidence that the relationships between first session CG scores and fluid abilities were different between groups (REAS: Z = .29, p = .77; WM: Z = .4, p = .69). Perceptual speed was significantly related to both first and final session scores for the non-*adaptive* group only. Consistent with previous studies (Kane et al., 2005, Ackerman et al., 2005b), there was a strong relationship between working memory and reasoning. Furthermore, for both groups, final and first session CG scores were highly related.

Supplemental table 1 contains the bivariate correlations between cognitive ability (task and composite measures) and each individual game measure included in the composite score while Supplemental table 2 contains the bivariate correlations for sessions not included in the main analyses.

Table 4. *Correlation matrices of composite scores for both groups (Adaptive/Non-Adaptive)*

Composite Score	Reasoning	Working Memory	Perceptual Speed	Final Session CG Score
Reasoning				
Working Memory	.62***/ .47***			
Perceptual Speed	.06 / .13	.15 / .35*		
Final Session CG Score	.72***/ .43**	.62***/ .37*	.09 / .37*	
First Session CG Score	.58***/ .62***	.42** / .49***	-.05 / .39**	.79***/ .66***

*Note. ***p < .001, **p < .01,*p < .05*

3.3 Dynamics of cognitive ability casual game scores across game sessions

To assess how these relationships between cognitive abilities and casual game scores changed across time, and if this change differed between the two groups, we created three linear mixed effects models for each cognitive composite score (i.e., reasoning, working memory, and perceptual speed). In each of these models, we included a random effect of subject. These models were implemented with the "lmerTest" package in R (Kuznetsova, A, Brockhoff P.B., & Christensen R.H.B., 2015; R Core Team). Fixed effects parameters included main effects of cognitive score, game session (final vs. first), and group (adaptive vs. non-adaptive) with all interaction terms included (i.e., cognitive score by session, cognitive score by group, group by session, and group by session by cognitive score). The three-way interaction of group, time and condition (Table 5) showed that both reasoning and working memory became more related to casual game scores from first to final session for the adaptive group compared to the non-adaptive group.

Table 5. *Model Summaries Predicting CG Game Scores*

Cognitive Predictor	Reasoning	Working Memory	Perceptual Speed
Fixed Effects			
Cognitive	0.649***	0.613***	0.341**
	(0.133)	(0.171)	(0.135)
Group	-0.058	-0.013	0.014
	(0.136)	(0.148)	(0.167)
Session	-0.011	-0.005	-0.004
	(0.085)	(0.085)	(0.087)
Cognitive*Group	0.098	0.021	-0.401*
	(0.204)	(0.260)	(0.216)
Cognitive*Session	-0.199*	-0.157	-0.017
	(0.118)	(0.140)	(0.101)
Group*Session	0.004	0.006	0.024
	(0.122)	(0.121)	(0.125)
Cognitive*Group*Session	0.424**	0.503**	0.191
	(0.182)	(0.213)	(0.163)
Intercept	0.019	-0.002	-0.039
	(0.095)	(0.104)	(0.116)
Model Goodness of Fit Measures			
Log Likelihood	-163.094	-171.424	-184.731
Akaike Inf. Crit.	346.188	362.847	389.462
Bayesian Inf. Crit.	378.117	394.777	421.392

Note: *$p < .05$, **$p < .01$, ***$p < .001$

3.4 Predicting first and final session casual game scores

To examine the extent to which working memory and reasoning predicted casual game performance, multiple sets of stepwise regression analyses were performed. Summary statistics for each added variable are reported in Table 6. All models and model comparisons were generated in R (R Core Team 2015) with bootstrapped confidence intervals generated with the "boot" package (Canty and Ripley 2015). The variance inflation factor for added variables in the final models were close to 1 and never above 10 (adaptive group: M = 1.53, Range = 1.26-1.84; non-adaptive group: M = 1.64,Range = 1.43-2.31), suggesting that multicollinearity was not a concern (Field, 2012). Added variable plots were created for each final model in each analysis to illustrate the unique effects of each predictor as well as to identify possible outliers and/or influential points (Supplemental Figure 2; Fox and Weisberg, 2011).

In a first set of stepwise regression analyses, first session CG scores were used as the outcome variable and cognitive scores were used as predictors. For the adaptive game group, reasoning emerged as the only significant predictor of first session CG scores (Table 6). For the non-adaptive game group, although some evidence for WM improving model fit existed in terms of a significant change in F (p = .05), this inference should be drawn cautiously as the bootstrapped confidence interval included 0. For the adaptive group full model, one possible outlier with a studentized residual value of -3.6 was identified. When excluding this subject from analyses, WM (β = .07, p > .05, *BCA* 95% *CI* [-.24, 0.4]) remained insignificant and reasoning (β = .61, p < .001, *BCA* 95% *CI* [.26, .84]) remained significant. Figure 3 (left) shows each group's predicted first session CG scores from the full models (p < .001; including the previously investigated outliers).

In a second set of stepwise regression analyses, WM and reasoning were used as predictors of final session CG scores. Both WM and reasoning uniquely predicted final session CG scores (Table 6) in the adaptive group. Adding reasoning to a model with WM significantly improved the model fit, $F(1, 41) = 30.72$, $p < .001$, while adding WM to a model with reasoning, $F(1, 40) = 4.45$, $p < .05$ also improved model fit. One possible outlier with a studentized residual value of 4.0 was identified in the adaptive group full model. When excluding this subject from analyses, the WM ($\beta = .27$, $p < .05$, *BCA* 95% *CI* [.03, .53]) and reasoning ($\beta = .63$, $p < .001$, *BCA* 95% *CI* [.35, .82]) parameters remained significant. For the non-adaptive group, only reasoning significantly predicted final session scores (Table 6). Figure 3 (right) shows each group's predicted first session CG scores from the full models ($p < .001$; including the previously investigated outliers).

In another stepwise regression analysis, we examined the unique predictive ability of reasoning and WM over and above first session performance. First session CG scores were added to the models before the cognitive predictors. For the adaptive group, WM and reasoning uniquely predicted final session CG scores above and beyond first session CG scores (Table 6). However, only WM emerged as a unique predictor when all three variables of WM, reasoning, and first session CG scores were added in the regression model (Table 6). We identified one potential influential point with a studentized residual of 3.2 and a cook's distance of 1.0. When excluding this subject from analyses, both the WM ($\beta = .16$, $p = .08$, *BCA* 95% *CI* [-.02, .36]) and reasoning ($\beta = .15$, $p = .15$, *BCA* 95% *CI* [-.05, .36]) were only marginally significant. This suggests some caution should be taken in interpreting these results. For the non-adaptive group, neither reasoning nor WM uniquely predicted final session CG scores above and beyond first session CG scores (Table 6).

Table 6. *Summary of hierarchical regression analyses predicting casual game achievement using working memory and reasoning*

	Adaptive			Non-Adaptive		
Variable added	β [*BCA* 95% *CI*]	Adj R^2	$p(\Delta F)$	β [*BCA* 95% *CI*]	Adj R^2	$p(\Delta F)$
Predicting first session						
Step 1: REAS	0.58 [0.18, 0.78]	.32		0.62 [0.44, 0.73]	.37	
Step 2: WM	0.1 [-0.23, 0.41]	.31	.53	0.26 [-0.03, 0.55]	.41	.05
Predicting first session						
Step 1: WM	0.42 [0.04, 0.64]	.16		0.49 [0.26, 0.67]	.23	
Step 2: REAS	0.52 [0.13, 0.79]	.31	.00	0.5 [0.22, 0.71]	.41	.00
Predicting final session						
Step 1: REAS	0.72 [0.41, 0.84]	.51		0.43 [0.22, 0.62]	.17	
Step 2: WM	0.29 [0.07, 0.53]	.55	.03	0.21 [-0.08, 0.53]	.18	.17
Predicting final session						
Step 1: WM	0.62 [0.42, 0.79]	.38		0.37 [0.14, 0.6]	.12	
Step 2: REAS	0.54 [0.24, 0.76]	.55	.00	0.33 [-0.01, 0.54]	.18	.04
Predicting final session						
Step 1: first	0.79 [0.48, 0.9]	.62		0.66 [0.47, 0.81]	.43	
Step 2: REAS	0.4 [0.12, 0.72]	.72	.00	0.03 [-0.22, 0.37]	.41	.83
Step 3: WM	0.23 [0.04, 0.46]	.74	.02	0.05 [-0.25, 0.3]	.40	.72
Predicting final session						
Step 1: first	0.79 [0.5, 0.9]	.62		0.66 [0.45, 0.8]	.43	
Step 2: WM	0.35 [0.14, 0.65]	.71	.00	0.05 [-0.22, 0.34]	.41	.68
Step 3: REAS	0.26 [0.03, 0.6]	.74	.02	0.02 [-0.26, 0.32]	.40	.90

Note. REAS=Reasoning, WM=Working Memory, first= first session CG game scores.

As a follow up analysis in the non-adaptive group, we examined the unique relationship of perceptual speed (independent of fluid ability scores), given that these non-adaptive games place greater demand on speed and accuracy of motor responses and some evidence for a relationship was found in the previous correlation analysis (Table 4). Indeed, we found that perceptual speed predicted final session CG scores above and beyond reasoning and working memory (β = .33, *BCA* 95% *CI* [.08, .55], $p < .05$). However, perceptual speed scores did not significantly predict final session scores over and above first session CG scores (β = .15, p = .15, *BCA* 95% *CI* [-.05, .33]).

For comparison, we also performed these previous analyses with the adaptive group. Perceptual speed scores did not significantly predict CG scores in any model (ps > .05).

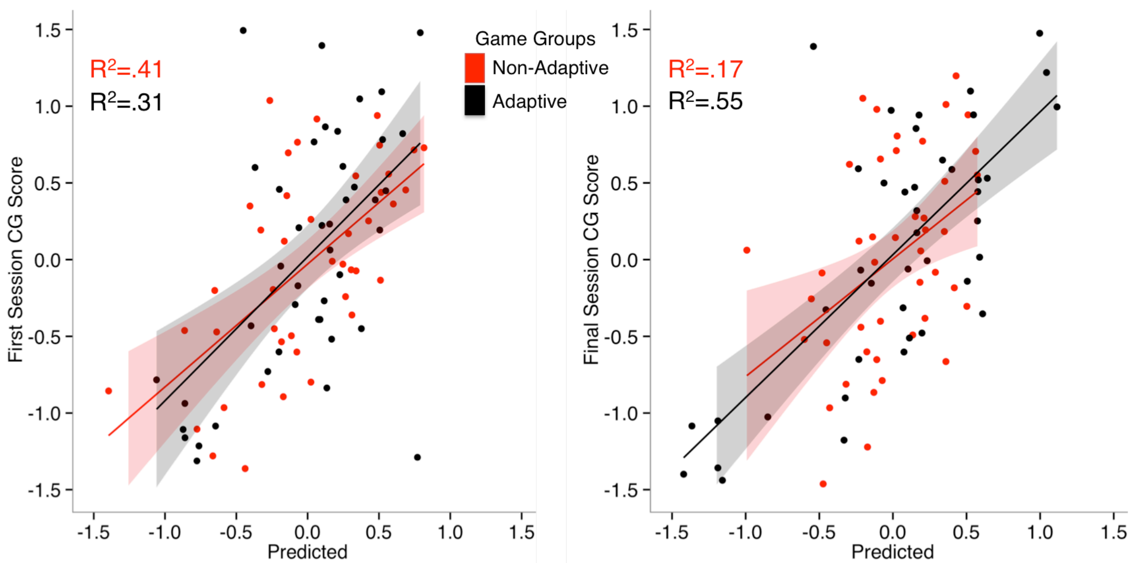


Figure 3. Multiple regression plots showing the predicted values for first (left) and final (right) session casual game performance for non-adaptive and adaptive group games derived from regression models using WM and reasoning as predictors. R^2= adjusted R^2. The shaded area represents the 95% confidence region for each predictor in the model.

CHAPTER 4: DISCUSSION

Casual video games provide an exciting resource for cognitive training research. Cognitive training research typically involves selecting games or tasks based on putative associations with specific cognitive abilities. Although informative, this approach overlooks changes in these relationships with extended gameplay—changes likely to have implications for the effectiveness of training targeted abilities. To shed light on this issue, we investigated the relationship between fluid abilities (WM, reasoning) and casual game performance over time. In line with our previous study's findings (Baniqued et al. 2013), initial CG scores were robustly associated with WM and reasoning scores. The current analysis took a closer look at these relationships and found that reasoning and WM predict relatively distinct aspects of performance over time. Specifically, reasoning uniquely predicted first session CG scores for both the adaptive and non-adaptive games and accounted for the relationship with WM. WM and reasoning uniquely predicted final CG scores (i.e., performance after multiple hours and sessions) for the adaptive game group—above and beyond first session CG scores—while reasoning remained the only unique predictor of CG scores for the non-adaptive group.

Although WM and reasoning have both been used to measure fluid abilities, they are rarely used together to understand the involvement of cognitive abilities in complex skill acquisition, despite more recent evidence for their unique relationships with some complex tasks (Alloway and Alloway, 2010; Dumontheil and Klingberg, 2012; Zook et al., 2004). In the current study, using both WM and reasoning provided a deeper understanding on the role of fluid abilities in CG performance over time. Specifically, reasoning, and much of the overlapping variance of WM, may be important for processes involved in novel task learning common to both the adaptive and non-adaptive games: finding solutions for novel game problems, integrating task instructions, and forming overall game strategies. In contrast, the unique predictive ability of WM was only evident in the adaptive games, when levels encountered later in training presumably placed greater demand on WM.

The current analyses show some support for the framework of complex skill acquisition, where task consistency moderates the relationship with cognitive abilities (Ackerman 1988). For complex tasks with varied processing demands, the relationship to fluid abilities is thought to remain stable or increase over time as individuals remain in an effortful, cognitive stage of skill acquisition—a pattern exhibited by the adaptive game group that encountered novel rules and problems at each level. However, as stated previously, the emergence of WM as a unique predictor after several adaptive game sessions shows that different aspects of these fluid abilities may become more important across tasks with varied processing demands. This distinction would not have been captured by use of a single measure of fluid ability or general cognitive ability (i.e., the common variance of all tasks).

Meanwhile, the relationships between cognitive abilities and CG scores in the non-adaptive group are more akin to tasks with consistent processing demands (also called consistent mapping; Schneider and Shiffrin 1977a, Schneider and Shiffrin 1977b). Despite a decrease in both the overall sample variance for game performance and in the relationship to fluid abilities over time, CG performance remained related to perceptual speed abilities across gameplay over and above both fluid abilities. These are patterns exhibited by consistent mapping tasks that emphasize speeded motor responses in Ackerman's (1988) framework (and see Ackerman 2000). Consistent processing demands likely emerge because the same strategies can be deployed over many instances (in this case, repeated levels), leading to automaticity (Logan, 1988). Participants may have initially learned a strategy, which they were then able to directly use for higher difficulty levels, without attempting alternative strategies. The speed in which these strategies were deployed may have determined individual differences marked by the stable association with perceptual speed observed in the current study. For example, in the game TwoThree, participants reported using a strategy of shooting down larger numbers as fast as possible until the numbers were reduced to smaller, more manageable numbers (http://lbc.beckman.illinois.edu/pdfs/CasualGames_SuppAnalyses.pdf). Therefore, higher difficulty levels (e.g., larger numbers to subtract on Two Three) may not necessarily mean that the relationship with fluid abilities will remain stable or increase if the same strategy is deployed at every level. This highlights an important distinction between task consistency and difficulty in adaptive tasks.

Future studies would confirm and expand these results. For example, examining different functions (e.g., Miyake et al., 2000; Unsworth et al., 2014) or content specific processes (e.g., verbal and non-verbal WM specific storage and strategies compared to a general WM capacity; Kane et al., 2004) may pinpoint the unique and common aspects of cognitive abilities important for predicting casual game performance over time.

Similarly, although games in each group were classified by common game elements (e.g., all the adaptive games involved reaching an end state to advance to the next problem), some of these puzzle-based games had speed and attentional switching demands similar to the non-adaptive games. For example, speeded responses and frequent shifts of attention are required in Gude Balls and Silversphere with demanding time constraints. On the other hand, Block Drop had no such time constraints. Future research should identify other components that may affect the involvement of specific cognitive abilities.

In terms of alternative explanations, one possibility is that the changing relationships between fluid abilities and game performance across sessions may not necessarily result from increased working memory load or inconsistency per se, but rather, changing levels of engagement and motivation (Csikszentmihalyi et al. 2005). That is, by the final game session, the adaptive games may keep participants with higher fluid abilities more engaged while the more consistent non-adaptive tasks may result in the opposite effect (Payne et al., 2012). Therefore, future research should investigate how these factors mediate the cognitive ability-game relationships across videogame play.

The current study may also provide information on future cognitive training designs with casual games. Motivated by the strong relationship between WM and reasoning found in previous individual difference studies, many recent cognitive training studies have sought to train WM abilities thought to be central to reasoning (i.e., fluid intelligence; e.g., Jaeggi et al., 2008, 2011; Harrison et al., 2013). However, this logic rests on the assumption that the WM mechanisms or skills improved through training are the same processes driving the common variance of WM and reasoning. A large motivation for video games, whether casual games (Baniqued et al., 2014) or other off-the-shelf video games (e.g., Basak et al., 2008) in cognitive research is to train cognitive abilities highly related to video game performance. In the current study, the common variance of WM and reasoning was most evident early in gameplay as shown by the strong relationship of reasoning and first session CG scores. Future cognitive training studies using casual games or similar tasks may consider introducing more casual games throughout training to maximally engage fluid abilities common to WM and reasoning tasks, where successful performance requires learning completely novel task environments. This may lead to developing more general skills important for forming strategies in novel tasks —a skill termed "learning to learn" hypothesized to underlie the effects of action video games on attention (Bavelier et al., 2012). Casual games associated with fluid abilities may be a valuable resource for this endeavor, given the influx of new and creative games introduced each year (http://www.casualgamesassociation.org) and the large collections of games already freely available. Maximizing training in this way may lead to a broader set of learned skills, addressing a drawback in cognitive training studies which primarily train on one paradigm, or a small subset of tasks or games with little to no improvement in fluid abilities such as reasoning (WM tasks: Shipstead et al 2012; Melby-Lervåg et al. 2013; Dougherty et al 2015; Melby-Lervåg et al. 2015; casual games: Baniqued et al. 2014; see Boot and Kramer 2014).

In conclusion, the current study sheds light on how the relationship of different types of casual games and cognitive abilities change after prolonged gameplay. More generally, we provide information on how the importance of certain cognitive abilities in video game performance may change depending on game structure, adding to a relatively scarce amount of literature on the relationship between cognitive abilities and video games (see Quiroga 2009, 2011, 2015; Ackerman 2011). Most importantly, this study illustrates how game structure should be considered when designing a study using video games or complex tasks to improve or measure cognitive abilities.

REFERENCES

Ackerman, P. L. (1988). Determinants of Individual Differences in Skill Acquisition: Cognitive Abilities and Information Processing. *Journal of Experimental Psychology. General*, *117*(3), 288–318.

Ackerman, P. L. (2005). Ability determinants of individual differences in skilled performance. In R. J. Sternberg & J. E. Pretz (Eds.), *Cognition and Intelligence: Identifying the Mechanisms of the Mind* (pp. 142–159). New York: Cambridge University Press.

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: the same or different constructs? *Psychological Bulletin*, *131*(1), 30–60. doi:10.1037/0033-2909.131.1.30

Ackerman, P. L., & Cianciolo, A. T. (2000). Cognitive, Perceptual-Speed, and Psychomotor Determinants of Individual Differences During Skill Acquisition, *6*(4), 259–290.

Ackerman, P. L., Kanfer, R., & Calderwood, C. (2010). Use it or Lose it? Wii Brain Exercise Practice and Reading for Domain Knowledge. *Psychology and Aging*, *25*(4), 753–766. doi:10.1037/a0019277.Use

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology*, *106*(1), 20–9. doi:10.1016/j.jecp.2009.11.003

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556–559.

Baniqued, P. L., Kranz, M. B., Voss, M. W., Lee, H., Cosman, J. D., Severson, J., & Kramer, A. F. (2014). Cognitive training with casual video games: points to consider. *Frontiers in Psychology*, *4*(January), 1010. doi:10.3389/fpsyg.2013.01010

Baniqued, P. L., Lee, H., Voss, M. W., Basak, C., Cosman, J. D., Desouza, S., … Kramer, A. F. (2013). Selling points: What cognitive abilities are tapped by casual video games? *Acta Psychologica*, *142*(1), 74–86. doi:10.1016/j.actpsy.2012.11.009

Bavelier, D., Green, C. S., Pouget, A., & Schrater, P. (2012). Brain plasticity through the life span: learning to learn and action video games. *Annual Review of Neuroscience*, *35*, 391–416. doi:10.1146/annurev-neuro-060909-152832

Boot, W. R., & Kramer, A. F. (2014). The Brain-Games Conundrum : Does Cognitive Training Really Sharpen the Mind ? *Cerebrum*, (November), 1–15.

Boot, W. R., Basak, C., Erickson, K. I., Neider, M., Simons, D. J., Fabiani, M., … Kramer, A. F. (2010). Transfer of skill engendered by complex task training under conditions of variable priority. *Acta Psychologica*, *135*(3), 349–357. doi:10.1016/j.actpsy.2010.09.005

Broadway, J. M., & Engle, R. W. (2010). Validating running memory span: measurement of working memory capacity and links with fluid intelligence. *Behavior Research Methods*, *42*(2), 563–70. doi:10.3758/BRM.42.2.563

Canty A. and Ripley B. (2015). boot: Bootstrap R (S-Plus) Functions. R package version 1.3-17.

Cattell, R.B. (1987). *Intelligence: Its structure, growth, and action*. New York: Elsevier.

Csikszentmihalyi, M., Abuhamdeh, S., & Nakamura, J. (2005). Flow. In Elliot, A. J., & Dweck, C. S. (Eds.), *Handbook of Competence and Motivation,* 598–608. New York, NY: Guilford Press.

Colom, R., Flores-Mendoza, C., & Rebollo, I. (2003). Working memory and intelligence. *Personality and Individual Differences*, *34*(1), 33–39. doi:10.1016/S0191-8869(02)00023-5

Colom, R., Quiroga, M. Á., Shih, P. C., Martínez, K., Burgaleta, M., Martínez-Molina, A., … Ramírez, I. (2010). Improvement in working memory is not related to increased intelligence scores. *Intelligence*, *38*(5), 497–505. doi:10.1016/j.intell.2010.06.008

Colom, R., Quiroga, M. Angeles, Solana, A. B., Burgaleta, M., Roman, F. J., Privado, J., … Karama, S. (2012). Structural changes after videogame practice related to a brain network associated with intelligence. *Intelligence*, *40*(5), 479–489. doi:10.1016/j.intell.2012.05.004

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, *7*, 547–552. doi:http://dx.doi.org/10.1016/j.tics.2003.10.005

Cowan, N. (2010). The Magical Mystery Four: How is Working Memory Capacity Limited, and Why? *Current Directions in Psychological Science*, *19*(1), 51–57. doi:10.1177/0963721409359277

Dougherty, Michael R., Toby Hamovitz, and Joe W. Tidwell, 'Reevaluating the Effectiveness of N-Back Training on Transfer through the Bayesian Lens: Support for the Null', *Psychonomic Bulletin & Review*, 2015 <http://dx.doi.org/10.3758/s13423-015-0865-9>

Dumontheil, I., & Klingberg, T. (2012). Brain activity during a visuospatial working memory task predicts arithmetical performance 2 years later. *Cerebral Cortex (New York, N.Y. : 1991)*, *22*(5), 1078–85. doi:10.1093/cercor/bhr175

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of Experimental Psychology. General*, *128*, 309–331. doi:10.1037/0096-3445.128.3.309

Engle, R., Kane, M., & Tuholski, S. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* (pp. 102–134). London: Cambridge Press.

Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics using R* (p. 992). London: Sage publications.

Fleishman, E. A. (1972). On the relation between abilities, learning, and human performance 1, (November), 1017–1032.

Fox, J., & Weisberg, S. (2011). *An R Companion to Applied Regression* (Second., p. 472). Thousand Oaks CA: Sage publications.

Fry, A. F., & Hale, S. (2000). Relationships among processing speed, working memory, and fluid intelligence in children. *Biological Psychology*, *54*(1-3), 1–34. doi:10.1016/S0301-0511(00)00051-X

Greenwood, P. M., Lambert, C., Sunderland, T., & Parasuraman, R. (2005). Effects of Apolipoprotein E Genotype on Spatial Attention, Working Memory, and Their Interaction in Healthy, Middle-Aged Adults: Results From the National Institute of Mental Health's BIOCARD Study. *Neuropsychology*, *19*(2), 199–211. doi:10.1016/j.micinf.2011.07.011.Innate

Harrison, T. L., Shipstead, Z., Hicks, K. L., Hambrick, D. Z., Redick, T. S., & Engle, R. W. (2013). Working memory training may increase working memory capacity but not fluid intelligence. *Psychological Science*, *24*(12), 2409–19. doi:10.1177/0956797613492984

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(19), 6829–33. doi:10.1073/pnas.0801268105

Jaeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y.-F., Jonides, J., & Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning — implications for training and transfer. *Intelligence*, *38*(6), 625–635. doi:10.1016/j.intell.2010.09.001

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Shah, P. (2011). Short- and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(25), 10081–6. doi:10.1073/pnas.1103228108

Jensen, A. R. (1998). *The Science of Mental Ability* (p. 648). Westport, CT: Praeger.

Kane, M. J., Conway, A. R. a, Miura, T. K., & Colflesh, G. J. H. (2007). Working memory, attention control, and the N-back task: a question of construct validity. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *33*(3), 615–22. doi:10.1037/0278-7393.33.3.615

Kane, M. J., Hambrick, D. Z., & Conway, A. R. a. (2005). Working memory capacity and fluid intelligence are strongly related constructs: comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*(1), 66–71; author reply 72–5. doi:10.1037/0033-2909.131.1.66

Kuznetsova, A, Brockhoff P.B., & Christensen R.H.B. (2015).

lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0-29. http://CRAN.R-project.org/package=lmerTest

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, *14*(4), 389–433. doi:10.1016/S0160-2896(05)80012-1

Kyllonen, P. C., & L. Stephens, D. (1990). Cognitive abilities as determinants of success in acquiring logic skill. *Learning and Individual Differences*, *2*(2), 129–160. doi:10.1016/1041-6080(90)90020-H

Lee, H., Boot, W. R., Baniqued, P. L., Voss, M. W., Prakash, R. S., Basak, C., & Kramer, A. F. (2015). The Relationship between Intelligence and Training Gains Is Moderated by Training Strategy. *PloS One*, *10*(4), e0123259. doi:10.1371/journal.pone.0123259

Logan, G. D. (1988). Toward an Instance Theory of Automatization. *Psychological Review*, *95*(4), 492–527.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions, *193*(1996), 1996–1998.

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*(2), 270–91. doi:10.1037/a0028228

Melby-Lervåg, Monica, and Charles Hulme, 'There Is No Convincing Evidence That Working Memory Training Is Effective: A Reply to Au et Al. (2014) and Karbach and Verhaeghen (2014), *Psychonomic Bulletin & Review*, 2015 http://dx.doi.org/10.3758/s13423-015-0862-z

Miyake, a, Friedman, N. P., Emerson, M. J., Witzki, a H., Howerter, a, & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex

"Frontal Lobe" tasks: a latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. doi:10.1006/cogp.1999.0734

Owen, A. M., Hampshire, A., Grahn, J. a, Stenton, R., Dajani, S., Burns, A. S., … Ballard, C. G. (2010). Putting brain training to the test. *Nature*, *465*(7299), 775–778. doi:10.1038/nature09042

Primi, R., Ferrão, M. E., & Almeida, L. S. (2010). Fluid intelligence as a predictor of learning: A longitudinal multilevel approach applied to math. *Learning and Individual Differences*, *20*(5), 446–451. doi:10.1016/j.lindif.2010.05.001

Quiroga, M. Á., Herranz, M., Gómez-Abad, M., Kebir, M., Ruiz, J., & Colom, R. (2009). Video-games: Do they require general intelligence? *Computers & Education*, *53*(2), 414–418. doi:10.1016/j.compedu.2009.02.017

Quiroga, M. Á., Román, F. J., Catalán, A., Rodríguez, H., Ruiz, J., Herranz, M., … Colom, R. (2011). Videogame Performance (Not Always) Requires Intelligence. *International Journal of Online Pedagogy and Course Design*, *1*(3), 18–32. doi:10.4018/ijopcd.2011070102

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring Working Memory Capacity With Automated Complex Span Tasks. *European Journal of Psychological Assessment*, *28*(3), 164–171. doi:10.1027/1015-5759/a000123

Redick, T. S., Unsworth, N., Kelly, A. J., & Engle, R. W. (2012). Faster, smarter? Working memory capacity and perceptual speed in relation to fluid intelligence. *Journal of Cognitive Psychology*, *24*(7), 844–854. doi:10.1080/20445911.2012.704359

Salthouse, T. a, Pink, J. E., & Tucker-Drob, E. M. (2008). Contextual analysis of fluid intelligence. *Intelligence*, *36*(5), 464–486. doi:10.1016/j.intell.2007.10.003

Salthouse, T. a. (2004). Localizing age-related individual differences in a hierarchical structure. *Intelligence*, *32*(6), 541–561. doi:10.1016/j.intell.2004.07.003

Salthouse, T. a. (2005). Relations between cognitive abilities and measures of executive functioning. *Neuropsychology*, *19*(4), 532–45. doi:10.1037/0894-4105.19.4.532

Salthouse, T. a. (2005). Relations between cognitive abilities and measures of executive functioning. *Neuropsychology*, *19*(4), 532–45. doi:10.1037/0894-4105.19.4.532

Salthouse, T. a., & Ferrer-Caja, E. (2003). What needs to be explained to account for age-related effects on multiple cognitive variables? *Psychology and Aging*, *18*(1), 91–110. doi:10.1037/0882-7974.18.1.91

Salthouse, T. a., & Pink, J. E. (2013). Why is working memory related to fluid intelligence? *Psychonomic Bulletin & Review*, *15*(2), 364–371. doi:10.3758/PBR.15.2.364

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2014). A task is a task is a task: putting complex span, n-back, and other working memory indicators in psychometric context. *Frontiers in Psychology*, *5*(December), 1–8. doi:10.3389/fpsyg.2014.01475

Schneider, W., & Shiffrin, R. M. (1977). Controlled and Automatic Human Information Processing: I. Detection, Search, and Attention. *Psychological Review*, *84*(1), 1–66.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and Automatic Human Information Processing: II. Perceptual Learning, Automatic Attending, and a General Theory. *Psychological Review*, *84*(2), 127–190.

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin*, *138*(4), 628–54. doi:10.1037/a0027473

Shute, V. J., & Kyllonen, P. C. (1990). *Modeling Individual Differences in programming skill acquisition*.

Unsworth, N., & Engle, R. (2005). Working memory capacity and fluid abilities: Examining the correlation between Operation Span and Raven. *Intelligence*, *33*(1), 67–81. doi:10.1016/j.intell.2004.08.003

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: capacity, attention control, and secondary memory retrieval. *Cognitive Psychology*, *71*, 1–26. doi:10.1016/j.cogpsych.2014.01.003

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505. doi:10.3758/BF03192720

Voelkle, M. C., Wittmann, W. W., & Ackerman, P. L. (2006). Abilities and skill acquisition: A latent growth curve approach. *Learning and Individual Differences*, *16*(4), 303–319. doi:10.1016/j.lindif.2006.01.001

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology*, *4*(July), 433. doi:10.3389/fpsyg.2013.00433

Woltz, D. J. (1987). An Investigation of the Role of Working Memory in Procedural Skill Acquisition, (3), 319–331.

Zhang, Z., Davis, H. P., Salthouse, T. a., & Tucker-Drob, E. M. (2007). Correlates of individual, and age-related, differences in short-term learning. *Learning and Individual Differences*, *17*(3), 231–240. doi:10.1016/j.lindif.2007.01.004

Zook, N. a, Davalos, D. B., Delosh, E. L., & Davis, H. P. (2004). Working memory, inhibition, and fluid intelligence as predictors of performance on Tower of Hanoi and London tasks. *Brain and Cognition*, *56*(3), 286–92. doi:10.1016/j.bandc.2004.07.003

SUPPLEMENTAL TABLES AND FIGURES

Supplemental Table 1. Bivariate correlation coefficients of individual tasks and games for the adaptive group games

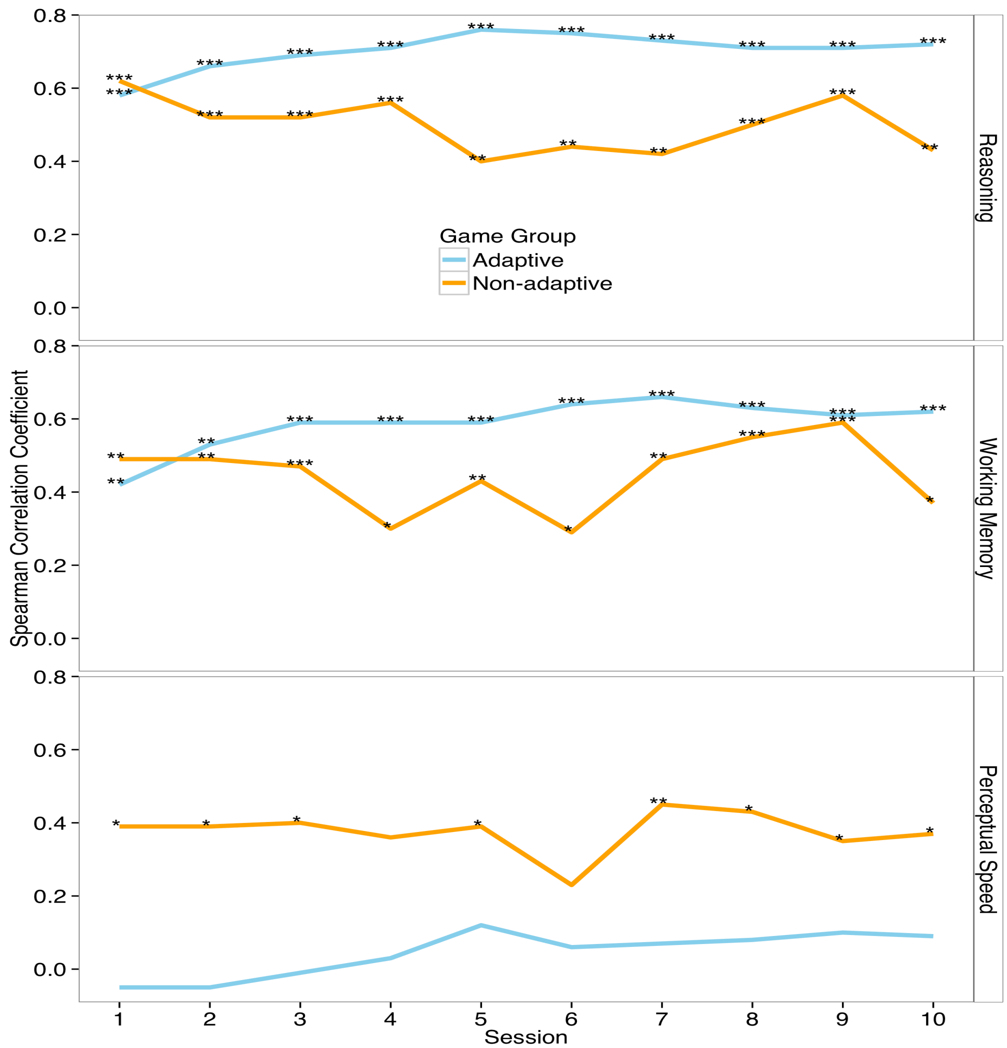
	Game Scores	Composite		Silversphere		Blockdrop		Gudeballs	
Cognitive Ability	Measure	First	Final	1	10	1	10	1	10
Reasoning	MatrixReasoning	0.27	.45**	.3*	.37*	0.24	.43**	0.17	0.17
Reasoning	FormBoards	.47**	.56***	.5***	.44**	0.29	.53***	0.27	.46**
Reasoning	PaperFolding	.39**	.42**	.39**	.4**	0.3	.37*	.35*	.35*
Reasoning	SpatialRelations	.53***	.63***	.63***	.6***	0.26	.63***	.36*	.37*
Reasoning	LetterSets	.54***	.47**	.33*	0.29	.48**	.49***	.49**	.35*
Reasoning	ShipleyAbstract	0.13	.33*	0.04	0.29	0.13	0.25	0.14	0.3
WM	SPWM	0.2	.38*	0.18	.42**	0.04	0.27	0.32	0.29
WM	Nback 3back	.43**	.52***	.5***	.38*	0.11	.43**	.5**	.53***
WM	VSTM	0.1	0.25	0.1	0.08	0.11	0.21	0.23	0.32
WM	Run Span	0.05	0.13	0.06	0.01	0.03	0.03	0.01	.44**
WM	Symmetry Span	.49***	.42**	0.29	0.3	.37*	.34*	.52***	.46**
Perceptual Speed	PatternComp	0.07	0.01	0.1	0.12	0.1	0.09	0	0.03
Perceptual Speed	LetterComp	0.04	0.07	0.21	0.03	0.02	0.06	0.17	0.21
Perceptual Speed	DSST	0.01	0.15	0.09	0.07	0.04	0.02	0.21	.42**
Composite	REAS	.58***	.72***	.54***	.6***	.42**	.68***	.37*	.51**
Composite	WM	.42**	.62***	.37*	.43**	0.12	.47**	.55***	.72***
Composite	PSpeed	0.05	0.09	0.18	0.07	0.05	0	0.16	0.26

Note. ***p<.001,**p<.01,*p<.05,First= first training session casual game performance, Final=final training session casual game performance, DSST=Digit Symbol Coding REAS=Reasoning, WM=Working Memory, PSpeed=Perceptual Speed.

Supplemental Table 2. Bivariate correlation coefficients of individual tasks and games for the non-adaptive group games

Cognitive Ability	Game Scores	Composite		TwoThree		DigiSwitch		Sushi	
	Measure	First	Final	1	10	1	10	1	10
Reasoning	MatrixReasoning	.43**	.31*	0.28	0.19	.46**	.43**	0.28	0.01
Reasoning	FormBoards	.57***	.38*	.32*	0.22	.62***	.4**	.38*	0.19
Reasoning	PaperFolding	.46**	.34*	.32*	0.24	.38*	0.27	.36*	0.26
Reasoning	SpatialRelations	.43**	.35*	0.13	0.19	.43**	.34*	.44**	0.19
Reasoning	LetterSets	0.23	0.2	0.15	0.12	0.27	0.15	0.14	0.14
Reasoning	ShipleyAbstract	.51***	0.24	0.25	0.09	.5***	0.27	.43**	0.21
WM	SPWM	0.21	0.21	0.07	0.15	.36*	0.24	0.08	0.03
WM	Nback 3back	0.27	.3*	0.11	0.17	0.23	0.2	0.29	.31*
WM	VSTM	0.28	.34*	.3*	0.15	0.2	0.26	0.18	.34*
WM	Run Span	0.17	0.08	0.1	0.18	0.11	0.15	0.2	0.13
WM	Symmetry Span^	.84***	0.22	0.4	0.24	.72***	0.12	.77***	0.07
Perceptual Speed	PatternComp	.53***	.43**	.44**	.31*	.39**	.48***	.39**	0.24
Perceptual Speed	LetterComp	0.02	0.1	0.06	0.11	0.11	0.1	0.11	0.09
Perceptual Speed	DSST	.4**	.37*	.35*	0.22	.47**	0.29	0.14	.37*
Composite	REAS	.62***	.43**	.34*	0.25	.62***	.44**	.48**	0.24
Composite	WM	.49***	.37*	.29*	0.26	.45**	.31*	.43**	0.24
Composite	PSpeed	.39**	.37*	.35*	0.26	.39**	.35*	0.17	0.28

Note. ***p<.001,**p<.01,*p<.05,First= first training session casual game performance, Final=final training session casual game performance, DSST=Digit Symbol Coding REAS=Reasoning, WM=Working Memory, PSpeed=Perceptual Speed. ^Only 24 participants completed Symmetry Span for the Non-Adaptive group as this measure was added half way through data collection.

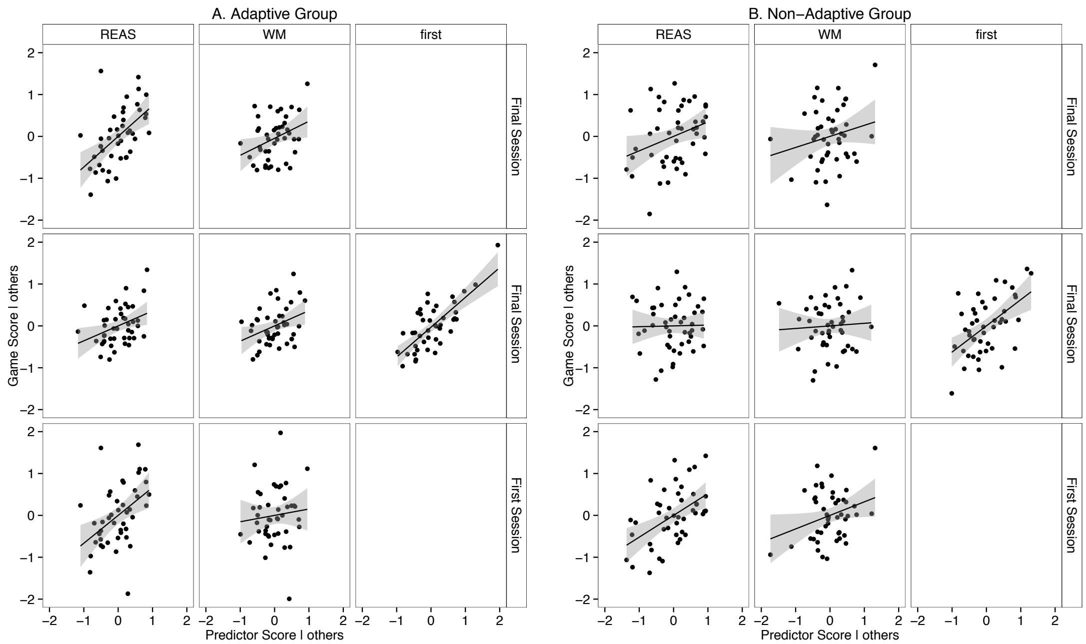


Supplemental Figure 1. *Spearman correlations across all sessions.* Note: *$p < .05$, **$p < .01$, ***$p < .001$

Supplemental Table 3. *Linear mixed models predicting standardized game performance across all 10 sessions.*

Parameters	Reasoning	Working Memory	Perceptual Speed
Fixed Effects			
Cognitive	0.591***	0.567***	0.355***
	(-0.119)	(-0.155)	(-0.13)
Group	-0.065	-0.017	0.009
	(-0.123)	(-0.135)	(-0.161)
Session	-0.001	-0.0002	0.0002
	(-0.006)	(-0.006)	(-0.006)
Cognitive*Group	0.286	0.233	-0.418**
	(-0.183)	(-0.236)	(-0.209)
Cognitive*Session	-0.014*	-0.001	-0.003
	(-0.008)	(-0.009)	(-0.007)
Group*Session	0.001	0.001	0.003
	(-0.008)	(-0.008)	(-0.008)
Cognitive*Group*Session	0.029**	0.029**	0.025**
	(-0.012)	(-0.014)	(-0.011)
Intercept	0.014	-0.005	-0.043
	(-0.086)	(-0.094)	(-0.112)
Model Goodness of Fit Measures			
Log Likelihood	-487.175	-494.865	-510.042
Akaike Inf. Crit.	994.35	1009.73	1040.084
Bayesian Inf. Crit.	1042.352	1057.731	1088.085

Note. *$p < .05$, **$p < .01$, ***$p < .001$



Supplemental Figure 2. Added variable plots for puzzle-based adaptive (A) and non-adaptive (B) game group full regression models. Along the x-axis is predictor score and the y-axis is standardized, composite game score after adjusting for the effects of the other predictors. Each panel column represents a predictor (REAS=reasoning, WM=working memory, first=first session casual game scores). Each panel row represents the dependent casual game variable (top row: predicting first session game performance with REAS and WM, middle row: predicting final session game performance with REAS, WM and first session predictors, bottom row: predicting final session game performance with first session game performance, REAS, and WM predictors) The shaded area represents the 95% confidence region for each predictor in the model.