Game Player Response Times versus Task Dexterity and Decision Complexity

Chunzhen Jiang, Aritra Kundu, and Mark Claypool
cjiang4,akundu,claypool@wpi.edu
Worcester Polytechnic Institute, Worcester, MA, USA

ABSTRACT

Game tasks commonly require dexterous actions and varying amounts of decision-making. People that play games may perform better for basic reaction and decision making tasks in comparison to people that do not play games. This paper presents results from two user studies that evaluate the relationship between self-rated gamer ability and reaction time for two tasks: 1) a task with varying decision complexity, and 2) a task with varying dexterity requirements. Analysis of data from over 150 users shows small effects of self-rated gamer ability on task, but substantial effects of the decision parameters (choices) and dexterity parameters (size and distance) on performance.

ACM Reference Format:
Chunzhen Jiang, Aritra Kundu, and Mark Claypool. 2020. Game Player Response Times versus Task Dexterity and Decision Complexity. In *Extended Abstracts of the 2020 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '20 EA), November 2–4, 2020, Virtual Event, Canada.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3383668.3419891

1 INTRODUCTION

Computer games are the world's most popular form of entertainment, with global sales increasing at an annual rate of 10% or more [15]. Moreover, the impacts of gaming are not only fiscal – fast-paced gaming has been shown to improve rapid-response decision making [5], with gamers able to respond to visual stimuli faster than non-gamers [8]. Our own research has shown self-rating of gamer ability correlates with performance for some game-specific tasks [10].

What has not been explored is how self-rating of gamer ability correlates with performance along different dimensions of interaction during play. In particular, we are interested in how self-rated gamer ability predicts task performance along two dimension common to many computer games: 1) dexterity and 2) decision complexity. For decision complexity, we evaluate how self-rated gamer ability impacts a rapid-response task with different numbers of choices. For task dexterity, we evaluate self-rated gamer ability in regards to Fitts' law, which governs the time to select a target based on the target size and target distance.

CHI PLAY '20 EA, November 2–4, 2020, Virtual Event, Canada

ACM ISBN 978-1-4503-7587-0/20/11...$15.00
https://doi.org/10.1145/3383668.3419891

We design and conduct two users studies deployed via Javascript through the Web. Both studies measure participants' reaction times. The decision complexity study has participants respond as fast as possible to visual stimuli with 1, 2 or 3 choices. The dexterity study has participants select circles of different sizes and distances as quickly as possible.

Analysis of data from over 150 participants shows a modest reduction in reaction time versus gamer ability – about a 350 milliseconds reaction time mean for low skill players versus about 325 milliseconds for high skill players. For decisions, these same trends hold at all three levels of complexity. Adding a second choice increases response times by about 50% over a single choice reaction time test, and having three choices approximately doubles response times over a single choice reaction time test. For dexterity, low and medium skill players have a similar linear fit for response time versus Fitts' index of difficulty, but high skill players have a lower y-intercept, indicating faster response times.

2 RELATED WORK

Table 1 summarizes findings from previous work on reaction times for gamers [2, 3, 5, 8, 14, 16, 18, 20]. Elite athletes have reaction times of about 150 ms, and elite gamers are probably close to that, if perhaps slightly slower. Good gamers have reaction times of about 250 ms, and average gamers about 300 ms. Non-gamers have reaction times of about 350 ms.

Table 1: Summary of typical reaction times.

Group	Reaction Time	Ref.
Human minimum	109 ms	[14]
Elite athlete	150 ms	[18]
Elite gamer	150-200 ms	[2]
Good gamer	250 ms	[8]
Average gamer	300 ms	[16]
Non-gamer	350 ms	[16]

Fitts' law [4] describes the time (T) to select a stationary target based on an index of difficulty (ID): $T = k_1 + k_2 \cdot ID$, where k_1 and k_2 are constants specific to the users and task at hand. The index of difficulty (ID) is proportional to the distance (D) from the source to the target and the width of the target (W): $ID = \log_2\left(\frac{2D}{W}\right)$. Fitts' law has been shown to be applicable to a variety of conditions (e.g., underwater [7]) and input devices (e.g., eye tracking [19] and computer mice [13, 17]). Fitts' law has been applied to games, too, assessing game controllers [9] and first person shooters [11].

3 METHODOLOGY

To assess gamer reaction times for tasks with different decision complexities and task dexterity difficulties, we conduct two different user studies. Both users studies deploy tasks via Javascript applications run through a Web browser. While absolute reaction times are difficult to assess through a Web interface, our intent is to compare the relative performance for different tasks and to compare gamers based on their self-rated skills. Both user studies followed the same procedure:

(1) Users answered demographic questions: age, gender, and self-rating as a gamer.

(2) Users navigated with a computer and mouse to our Web page with our Javascript applications.

(3) Users did a small set of tasks and submitted their results.

Participation was voluntary and approved by the university Institutional Review Board (IRB). Participants were solicited through campus mailing lists and the authors' online social networks (e.g., Facebook, Weibo).

3.1 Decision Complexity

Based on previous studies that had users react to visual changes [12], we developed a Javascript application similar to [6], but with additional cases for 2 and 3 choice tests.

For task 1, users click the "Ready" button above a white box to start the test. The box changes color from white to green after a random time interval of up to 3 seconds, chosen based on pilot studies confirming this provides a sufficiently random interval without being too long. When the box changes to green, the user must press the Z key as quickly as possible. The time from the color change until the Z is pressed is recorded as the reaction time. This cycle is repeated 5 times. Anytime the user presses Z after "Ready' but before the box changes color, the test must be re-done.

Task 2 is the same as task 1 except that instead of 1 choice, the user has 2. If the white box changes color to green, the user still presses Z. However, if the white box changes to yellow, the user must press X.

Task 3 is the same as task 2, but with 3 choices – green requires a Z-press, yellow requires an X-press, and blue requires a C-press.

For tasks 2 and 3, mistakes (i.e., pressing the wrong key) are recorded, but users need to complete the tests successfully 5 times. Once all 5 tests are complete for a task, the responses are displayed on the screen so the users can copy and paste the data into our survey for our later analysis. The presentation order for the 3 tasks is randomized.

3.2 Task Dexterity

The demographic questionnaire asks for the user's display size and then a pop-up window that instructs the user to use a mouse and play in full-screen mode.

We developed a Javascript application that had users react to input with a mouse click. Users click the "Ready" button to start the round, whereupon the screen changes color after a random time interval between 0 and 2 seconds (duration chosen based on pilot studies). The user subsequently clicks the mouse as quickly as possible. If the user clicks before the screen changes color, an alert warning is displayed and the round must be re-done. After 5 successful rounds, the responses are displayed on the screen so participants can copy and paste the response data into our survey for our later analysis.

Based on principles in previous work [13], we developed another Javascript application where users click the "Ready", thus centering the mouse on the screen. After between 0 and 2 seconds, a circle of one of two sizes appears a random distance from the center of the screen. The user subsequently clicks on the circle as fast as possible. After 5 successful rounds, the responses are displayed on the screen so the participants can copy and paste the data into our survey for our later analysis.

4 RESULTS

This section analyzes the results (see our technical report [1] for more details).

Our decision complexity user study had 66 participants and our task dexterity study had 88 participants. Table 2 summarizes the main demographics, with standard deviations in parentheses. "Gamer" is a self-rating from 1-low to 5-high.

Table 2: Demographic summary.

Study	N	Age	Gender	Gamer
Decision	66	23.6 (8.7)	49 ♂ 14 ♀ 1 ?	3.6 (1.1)
Dexterity	88	20.9 (4.0)	66 ♂ 20 ♀ 2 ?	3.8 (1.0)
Total	154	22.1 (6.6)	115 ♂ 34 ♀ 3 ?	3.7 (1.0)

Table 3 shows the breakdown of self-rated gamer ability for each user study, with the mean and standard deviation in the last two columns. The bottom row shows the breakdown of both studies combined into one. Both studies have a slight skew towards high self-rated skill (mean self-rated skill is slightly above 3 and the mode is 4 for each dataset), but there are players of all self-rated skill levels in each dataset.

Table 3: Breakdown of self-rated skill

	Self-rated skill						
Study	1	2	3	4	5	$\bar{x}$	s
Decision	3	8	21	19	15	13.2	7.56
Dexterity	2	6	21	35	24	17.6	13.5
Total	5	14	42	54	39	30.8	20.4

4.1 Reaction Time

Both of our user studies have a reaction time task where users respond as quickly as possible to a change in color on the screen. We combine the data from both tests and analyze the relationship between self-rated skill as a gamer and reaction time. Based on previous work [10], we combine users in skill groups 1-2 and skill groups 4-5 to obtain 3 skill groups: low, medium and high.

Given the nature of the study, users not paying attention results in unusually high response times. To account for this, we remove all outliers that are higher than $1.5 \times IQR$. In total, 20 points were removed (out of 770).

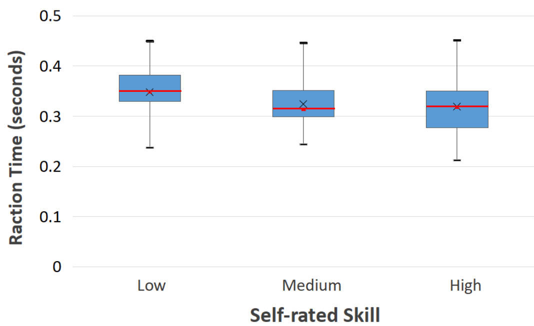

Figure 1: Reaction time versus skill group.

Figure 1 depicts a boxplot of the reaction times. The x-axis is the self-rated skill and the y-axis is the reaction time. The boxes denote the quartiles, with the middle line the median and the 'x' the mean. From the figure, there is a slight visual trend downward (faster) with higher gamer skill. However, post-hoc pairwise tests in Table 4 show the differences are not statistically significant, although the effect sizes (Cohen's d) for medium and high skill versus low skill are moderate.

Table 4: T test for Reaction Time task.

Comparison	t-Statistic	p value	d
Low - Med	t(26) = 1.72	0.10	0.63
Med - High	t(48) = -0.07	0.95	0.02
Low - High	t(40) = 1.45	0.16	0.51

4.2 Decision Complexity

Figure 2 depicts the response time on the x-axis versus the self-rated skill on the y-axis. The three decision tasks are indicated by different colors and shapes. Each point is the response time averaged across all users for a given skill group and task, shown with a 90% confidence interval. From the figure, vertically there is a clear separation of confidence intervals at all skill levels for all tasks, with the higher decision complexity tasks taking more time on average than the lower complexity tasks. Horizontally, there is a visible downward trend (lower is better) in average response time with higher self-rated skill. However, there is overlap in the confidence intervals for each adjacent pair-wise comparisons.

T test results ($\alpha = 0.1$) are shown in Table 5. The top half of the table compares self-rated skill groups and the bottom half compares decision task groups. The last column is the effect size (Cohen's d). From the table, differences in skill group are not statistically significant although the effect sizes are medium, but differences in task decision complexity are significant for each pair, even using a Bonferroni correction, with large effect sizes.

4.3 Dexterity

Figure 3 shows a scatter plot of the response time versus Fitts' Index of Difficulty (ID) [4, 13]. The x-axis is the ID score and the y-axis is the response time. Each point is the time to select a target with the ID computed based on the target distance from the center

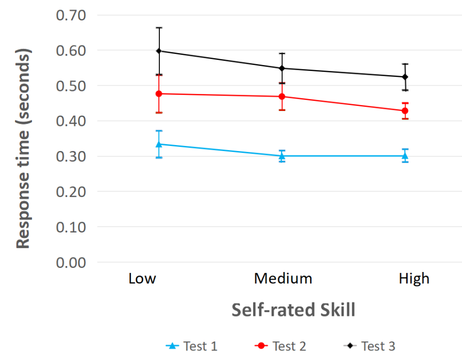

Figure 2: Average response time for decision tasks.

Table 5: T test for decision complexity tasks.

Comparison	t-statistic	p value	d
Low - Med.	t(23) = 0.6	0.55	0.28
Med - High	t(42) = 1.2	0.24	0.38
Low - High	t(31) = 1.5	0.15	0.64
Task 1 - Task 2	t(111)= -10.8	< .001	2.0
Task 2 - Task 3	t(105)= -5.0	< .001	2.8
Task 1 - Task 3	t(112)=-14.8	< .001	1.0

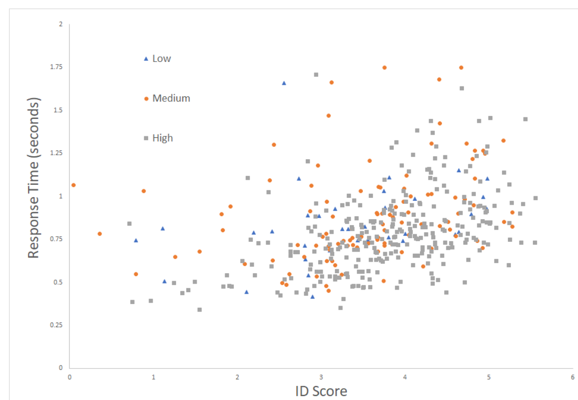

Figure 3: Response time versus ID score.

and the target width. The point shapes and colors differentiate the self-rated skill groups for each user: low, medium and high. From the figure, there is a visual upward trend, left to right, in that as ID scores increase so does response time, confirming Fitts' law [4]. There is considerable variation, however, in that response times vary even for the same ID score. This is to be expected given the natural variation in reaction times, demonstrated above. Visually, it is difficult to discern differences in response times for the different skills groups.

Since Fitts' law states there is a linear relationship between response time and ID, we do a linear regression for each skill group. Table 6 shows the line fit slope, y-intercept, and R^2. The first three rows are for each self-rated skill group and the last row is for the combined dataset. From the table, Fitts' law explains about 12%

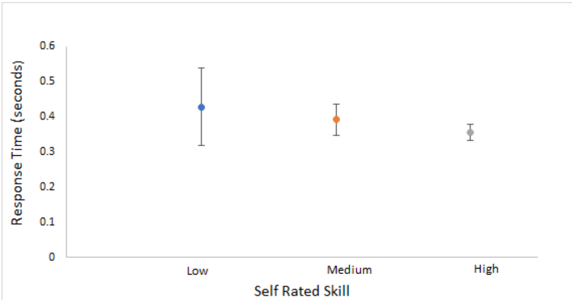

Figure 4: Average response time for dexterity task.

of the variation in response time overall, about 10% for low- and medium self-rated skill players and up to 20% for high self-rated skill players.

Table 6: Linear regression for dexterity task.

Skill	Slope	Y Intercept	R^2
Low	0.07	0.62	0.10
Medium	0.08	0.62	0.08
High	0.11	0.38	0.20
All	0.08	0.49	0.12

Figure 4 depicts response time versus skill for the dexterity task. The x-axis is the self-rated skill and the y-axis is the response time. Each data point is the average response time for all users in that category shown with a 90% confidence interval. From the figure, there is a visual downward trend in mean response time as skill group increases. However, there is some overlap in the confidence intervals.

A between-subjects ANOVA test (0.1 significance) for the 3 skill groups shows a significant effect of self-rated gamer skill on response time for the dexterity task (F(2, 83) = 3.39, p = 0.038). Post-hoc t test results ($\alpha = 0.1$) are shown in Table 7. From the Table, the low-skill group is not statistically different than either medium or high, owing to the large variance and fewer users in this group, but the difference in skill group medium to high is significant even using a Bonferroni correction. The effect sizes are moderate.

Table 7: T test for dexterity task.

Comparison	t-Statistic	p value	d
Low - Med.	t(26) = -0.62	0.268	0.231
Med. - High	t(77) = 2.57	0.006	0.582
Low - High	t(63) = 0.92	0.180	0.407

5 CONCLUSION

This paper presents results from two users studies evaluating the reaction times versus self-rated gamer skill along two dimensions: decision complexity, in the form of 1, 2 and 3 choices, and dexterity, in the form of targets of varying sizes and distance as in Fitts' law. Analysis of results from over 150 participants shows small effects of self-rated gamer skill on reaction time, but moderate effects on task decision complexity for 1, 2 and 3 choices and dexterity in selecting targets of varying distance and size.

Future work includes exploring more axes for decision complexity, as is the case for many games. Gamer skill can also be assessed through multiple questions, differentiating skills for different types of games (e.g., fast-paced vs. strategy) and correlated with response time. Users studies could also include a wider range of demographics (e.g., age) to more broadly apply to the gamer population at large. Future work could also analyze *why* the observed relationships between self-rated gamer skill and response times occur.

REFERENCES

[1] Aritra Kundu Chunzhen Jiang and Mark Claypool. 2020. *Game Player Response Times versus Task Dexterity and Decision Complexity*. Technical Report WPI-CS-TR-20-04. Computer Science Department at Worcester Polytechnic Institute.

[2] CSGO Demo Reviews. 2017. Pro CSGO Players Reaction Times. Online at: https://youtu.be/y2PwyBTbCcQ?t=435.

[3] Matthew W. G. Dye, C. Shawn Green, and Daphne Bavelier. 2009. Increasing Speed of Processing With Action Video Games. *Current Directions in Psychological Science* 18, 6 (Dec. 2009), 321 – 326.

[4] Paul M. Fitts. 1954. The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement. *Journal of Experimental Psychology* 47, 6 (June 1954), 381–391.

[5] C. Shawn Green, Alexandre Pouget, and Daphne Bavelier. 2010. Improved Probabilistic Inference as a General Learning Mechanism with Action Video Games. *Current Biology* 20 (Sept. 2010), 1573 – 1579.

[6] Human Benchmark. [n.d.]. Reaction Time Test. https://humanbenchmark.com/tests/reactiontime. Accessed: 2010-09-30.

[7] R. Kerr. 1973. Movement Time in an Underwater Environment. *Journal of Motor Behavior* 5, 3 (1973), 175 – 178.

[8] Andrew Latham, Lucy Patston, Christine Westermann, Ian Kirk, and Lynette Tippett. 2013. Earlier Visual N1 Latencies in Expert Video-Game Players: A Temporal Basis of Enhanced Visuospatial Performance? *PLoS ONE* 8, 9 (2013).

[9] A. Lee, K. Song, H. B. Ryu, J. Kim, and G. Kwon. 2015. Fingerstroke Time Estimates For Touchscreen-Based Mobile Gaming Interaction. *Human Movement Science* 44 (Dec. 2015), 211–224.

[10] Shengmei Liu, Mark Claypool, Bhuvana Devigere, Atsuo Kuwahara, and Jamie Sherman. 2020. 'Git Gud!' - Evaluation of Self-Rated Player Skill Compared to Actual Player Performance. In *Proceedings of the ACM Annual Symposium on Computer-Human Interaction in Play (CHI PLAY)*. Online (Virtual Conference). Work In Progress.

[11] Julian Looser, Andy Cockburn, and Joshua Savage. 2005. On the Validity of Using First-Person Shooters for Fitts' Law Studies. *People and Computers XIX* 2 (2005), 33–36.

[12] C. S. Green M. W. G. Dye and D. Bavelier. 2009. Increasing Speed of Processing With Action Video Games. *Current Directions in Psychological Science* 18, 6 (Dec. 2009), 321–326.

[13] I. S. MacKenzie and W. Buxton. 1992. Extending Fitts' Law to Two-Dimensional Tasks. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. Monterey, CA, USA, 219 – 226.

[14] Reference.com. 2011. What Is the Average Human Reaction Time? https://tinyurl.com/yc5wx6af, Last accessed: July 2020.

[15] Reuters.com. 2018. Investing in the Soaring Popularity of Gaming. https://www.reuters.com/sponsored/article/popularity-of-gaming.

[16] B. Richardson, D. Ellis, A. Celori, J. Cherry, and C. Meador. 2014. Reaction Times Differences in Video Game and Non Video Game Players. In *Symposium of University Research and Creative Expression (SOURCE)*. Department of Psychology, Central Washington University, Ellensburg, WA, USA. https://digitalcommons.cwu.edu/source/2014/posters/175/.

[17] R. W. Soukoreff and I. S. MacKenzie. 2004. Towards a Standard for Pointing Device Evaluation – Perspectives on 27 Years of Fitts' Law Research in HCI. *Elsevier International Journal of Human-Computer Studies* 61, 6 (2004), 751 – 789.

[18] E. Tonnessen, T. Haugen, and S. A. I. Shalfawi. 2013. Reaction Time Aspects of Elite Sprinters in Athletic World Championships. *Strength and Conditioning Research* 27, 4 (April 2013), 885 – 892.

[19] C. Ware and H. Mikaelian. 1987. An Evaluation of an Eye Tracker as a Device for Computer Input. In *Proceedings of ACM Human Factors in Comp. Systems and Graphics Interface*. Toronto, Canada.

[20] David L. Woods, John M. Wyma, E. William Yund, Timothy J. Herron, and Bruce Reed. 2015. Factors Influencing the Latency of Simple Reaction Time. *Frontiers in Human Neuroscience* 9 (2015), 131.