Designing for Response Tasks: An Empirical Evaluation of Response Tasks in a Video Game Context

Naia Kamawar-MacLeod
Carleton University
Ottawa, Canada
naiakamawarmacleod@cmail.carleton.ca

Robert J Teather
Carleton University
Ottawa, Canada
robteather@carleton.ca

Abstract

Numerous studies have examined reaction time differences due to stimulus or response methods but few have examined the relationship between these factors, and even fewer in a video game context. We present a formal experiment comparing the impact of four display types across four input methods, and three indicator speeds across three target area sizes on accuracy and response time in game-like displays. Our main findings indicate a number of significant main effects across different display types, indicator speeds, and target area sizes, and many interaction effects between pairs of speeds and sizes. The vertical meter and horizontal meter were found to have a significantly faster response time than the circular meter and pulsating ring. The vertical meter, horizontal meter, and circular meter were all found to have higher average accuracy than the pulsating ring, but participants generally reported finding the circular meter the easiest to successfully respond to due to its cyclical nature. Overall, larger target sizes and faster speeds led to lower response times, and larger target sizes and slower speeds led to higher accuracy.

CCS Concepts

• **Human-centered computing → Empirical studies in HCI**; **Graphical user interfaces**; **User interface design**; **Interaction techniques**.

Keywords

Video Game Design, Reaction Time, Display Type, Response Method, Visual Stimulus

ACM Reference Format:
Naia Kamawar-MacLeod and Robert J Teather. 2025. Designing for Response Tasks: An Empirical Evaluation of Response Tasks in a Video Game Context. In *Proceedings of Graphics Interface (GI '25)*. ACM, New York, NY, USA, 10 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Reaction time — the time it takes to observe and respond to a stimulus — is something that affects daily life, from playing a sport, to driving a car, to playing video games. Each require the person to register an external stimulus (e.g., a soccer ball coming towards them) and react to it (e.g., by kicking it). The time this takes is *reaction time* [17]. Reaction time tasks are common in HCI. Many computer-based tasks involve time-sensitive responses from the user when a stimulus is presented (e.g., a blinking notification icon) requiring a user response before proceeding (e.g., clicking the icon to dismiss it). Such tasks are especially common in video games, which often rely on "time-pressure" to impose game challenges [50].

Variations of these basic tasks also exist, within the broader category of what we refer to as *response time* (RT) tasks. Response time task complexity varies based on the nature of the task, such as whether only certain stimuli require a response (recognition reaction time) or if different stimuli require different responses (choice reaction time) [17, 26]. Target selection tasks are similar to standard response time tasks with the added complexity of a moving target and/or cursor: the user moves a cursor to a target, and must react quickly when the cursor (visual indicator) enters the target . Many other factors influence response time tasks, such as stimulus colour [49], contrast [9], size [24], movement speed [20], and response input method [1]. It is thus important to quantify relationships between these factors and response speed to understand the RT in diverse situations.

Response time tasks are especially common in video games, and accuracy is the primary metric measured [51]. Many games involve fast-paced actions (e.g., combat, sports), where a quick response to a stimulus is crucial to gameplay, with user RT often directly affecting the game outcome [10, 20, 52]. The tight feedback loop and speedy responses required by games make them an interesting case for study. Yet, game-like RT tasks are relatively understudied compared to RT in non-game contexts [17, 30, 36, 37]. Thus while a multitude of factors influence these interactions there is little empirical research to guide game developers in how to scale difficulty of such tasks in their games [10, 19, 52].

We present a formal experiment comparing various tasks that rely on user response time viz hitting a target zone in a moving meter, common in recent video games. We compared response time and accuracy in game-like information displays across four factors: *display type* (vertical meter, circular meter, pulsating ring, and horizontal meter; see Figure 1), *response method* (keypress, mouse, touchpad, and game controller button), *indicator movement speed* (three speeds), and *target area size* (three sizes). Full details of each factors are detailed in section 4. To our knowledge this is the first experimental investigation of factors influencing tasks commonly found in video games which rely on response time.

Our study was guided by the following research questions:

(1) Which of display types facilitates the fastest response?;
(2) Which of the input methods has the lowest RT?;
(3) Do these two factors interact?;
(4) How are the RT and accuracy of these factors influenced by the indicator speed and target area size?

And more broadly: How can these factors be leveraged to influence RT and accuracy when designing a video game with RT tests?

GI '25, 2025,
2025. ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/XXXXXXX.XXXXXXX

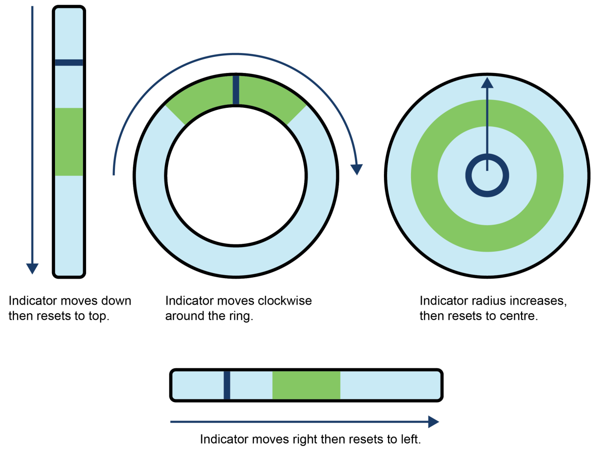
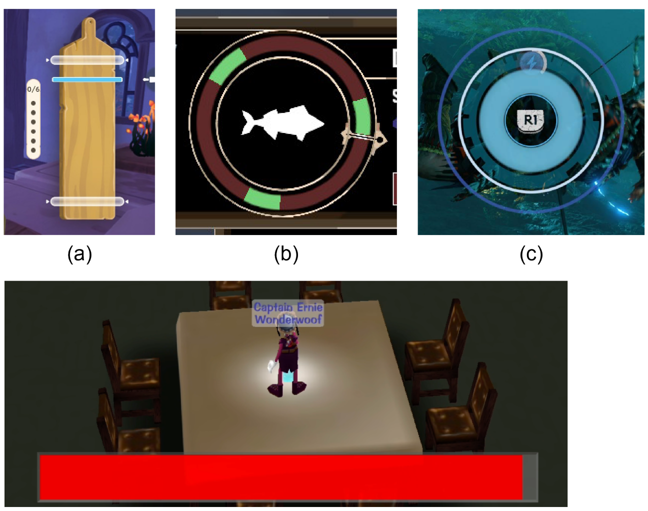


Figure 1: Display types used in the study. From left to right, top to bottom: vertical meter; circular meter; pulsating ring; and horizontal meter



(a) (b) (c)

(d)

Figure 2: Commercial game displays. (a) Vertical meter from *Palia* [43]. (b) Circular meter from *Dredge* [47]. (c) Pulsating ring from *Horizon Forbidden West: Burning Shores* [44]. (d) Horizontal meter from *Toontown Rewritten* [48].

Players generally prefer games with a balanced effort/reward ratio; too little effort yields boredom, while too much effort is frustrating [2, 4]. Players enjoy games that they can do well at [38], while also requiring some uncertainty of success to avoid tediousness [8]. This balance is crucial for enjoyability. Our experiment provides designers a starting point in determining parameters in game RT tasks, so they can adjust to provide compelling challenges to their players. Our analysis consists of comparing display type crossed with input method, and indicator speed crossed with target area size to examine how accuracy and response time are affected. This allows us to determine which levels and which combinations are best suited for specific video game design goals.

2 Related Work

Many studies examined response time and factors influencing it. Typically such experiments have participants observe a stimulus (often but not always on a screen), with the goal of responding as quickly and as accurately as possible [10, 13, 19, 28, 49, 52].

2.1 Comparing Reaction Time Test Type

With two exceptions, we focused our literature review on studies looking at only one type of response time test. Otaki and Shibata [36] compared simple, recognition, and choice reaction times of young and middle-aged people, across visual stimuli (e.g., directional arrows versus a dot or cross in choice tasks). Participants scored similarly in simple tests regardless of age, but younger people were faster in both recognition and choice reaction time tests. Mudric et al. [30] examined how participant karate skill level influenced RT in both simple and choice tests. Participants with a higher karate skill had faster RTs in both simple and choice tests, with simple RT on average lower across participants, compared to choice RT. These results suggest that different RT test parameters impact performance, as does prior skill level in tasks involving fast responses.

2.2 Comparing Reaction Time Test Stimuli

Many past studies varied stimulus factors (e.g., visual, auditory, or somatosensory) between conditions in RT tasks. Most studies [13, 24, 28, 49] constrained this to different stimuli in a single modality (usually visual), but two [17, 37] varied stimuli across modalities. Both Jain et al. [17] and Pascual-Leone et al. [37] compared auditory stimuli to visual stimuli and report that auditory stimuli response time is faster. Pascual-Leone et al. further compared somatosensory stimuli RT, and found it varied depending on stimuli location.

Other researchers focused on varying visual stimuli, such as colour variations [24, 49] in simple reaction tests. Vishteh et al. [49] and Lyons [24] varied stimulus colour across their conditions; Lyons further varied stimulus size and blink rate. Both studies found statistically significant differences across the various conditions. Visheh et al. report differences in RT across different colour stimuli, while Lyons found little difference, but found that both stimulus size and blinking speed impacted RT. The difference in findings could be attributed to the substantial difference in trial length: Lyons measured RT in the scale of seconds while Vishteh et al. measured in the scale of milliseconds. Stimulus colour effects seem to be most pronounced on smaller scales.

Hasbroucq and Guiard [13] and Miller et al. [28] conducted similar studies varying visual stimulus parameters in different reaction time tests. Hasbroucq and Guiard studied choice reaction time tasks where the response and stimulus colour matched (or not), and their physical positions lined up (or did not). RTs were faster for colour- and position-matched trials than non-matched trials. Miller et al. added a visual warning (e.g., an outline appearing) prior to the stimulus' appearance in a recognition task. While the visual warning decreased RT for stimuli that require a response, it also increased incorrect responses to stimuli that did not require a response. These studies further illustrate how numerous visual stimuli elements impact human reaction times.

2.3 Comparing Input Methods

Input method — the action participants perform to end a response time task — can also influence performance [1, 23]. Ahmetovic et al. [1] compared standard and alternative response methods for mobile games towards improving accessibility. Lin and Wu [23] examined differences in RT and accuracy between digital and physical keyboards. Both studies revealed significant differences between the average RTs due to these varying response methods.

Several studies of related tasks (e.g., target selection) also examined differing input methods. Schwind et al. [41], Roig-Maimó et al. [39], and Sanchez et al. [40] compared input methods in selection tasks, the former looking at effects of avatar appearance and the latter two looking at using head movement as input. Hild et al. [14] and Huang et al. [16] both examined varying input methods as well, namely through comparing gaze-based input to standard mouse input and comparing mouse, stylus, and touch inputs, respectively. Overall, these studies reveal that input method significantly effects performance in related tasks; this is likely the case in response time as well.

2.4 Comparing Speed and Target Size

Target selection - both moving and stationary - research reveals that target size, distance, and speed of a moving indicator/target influence response time [5, 15, 22, 40]. Chiu et al. [5], Lee and Hong [22], Huang et al. [15], and Sanchez et al. [40] all compared speed and target size in terms of accuracy and/or reaction time in their work. Their findings were all generally consistent with Fitts' Law [27] and found that higher speeds and smaller targets increased both response time and error rates. Of the studies detailed above, only Lee et al. [20] examined speed without an accuracy requirement, examining varying target speed and cue viewing time, defined as the time from trial start to the time the indicator enters the target area. They report a significant interaction between target speed and cue viewing time. These studies overall show that varying target/indicator speed and size can influence performance in a variety of RT tasks.

2.5 Response Time in Video Games

While the above studies reveal several factors affecting RT, only Ahmetovic et al.'s [1] study relates to video games. Both Huang et al. [15] and Lee et al. [21] further studied response times specifically in video games. Huang et al. found a significant difference in error rates between a standard method and a predictive method of processing user selection in a moving grid game [15]. Lee et al. had participants stack objects by responding when a moving object is centered over their tower. They found that results with the game were consistent with their earlier model [20]. Duinkharjav et al. [9] studied how saccade latency (eye movements) and varying colour and contrast impacted RT across multiple game types (e.g., sports, shooters), and report that higher contrast decreased RT. Lee et al. [21] modeled latency estimation and compensation in reaction time tasks common in video games, specifically in the mobile game *Flappy Bird*. Additionally, three studies using standard reaction time test methodologies investigated RTs in games. Kim et al. [19] found differences in response time across two distinct timing-based video games. Two studies, by Dye et al. [10] and Ziv et al. [52], compared RTs between video game players and non-video game players, reporting differences between the two groups. Overall, there are few formal experiments specifically about RT in games; designers must interpret results of peripherally related research to guide future game designs.

3 Example Games

We surveyed commercial games with response time tasks to identify common displays, focusing on tasks in the game's user interface (UI) rather than its environment (e.g., 3D world). We omitted tasks which did not include a persistent indicator and target area. We identified eighteen games which include tasks across four major display type categories, with four or more games in each category. In all cases, task difficulty stems from the limited time a moving indicator spends in a target area, and is impacted by both the indicator's speed and target area size.

While many other types of response time tasks exist in video games, such as timed button sequences often found in rhythm games or tasks that require multiple buttons to be pressed to successfully respond, we focused our research specifically on single button press tasks with moving indicators and target areas. We chose to focus on this task type specifically as in our experience playing video games, these types of tasks are ubiquitous across a wider variety of game genres, including role-playing, adventure, casual, and party games.

3.1 Vertical Meter

Vertical meter displays consist of a vertical path that the indicator moves along and that the target area is situated in. When the indicator reaches the end of the path, it generally resets to its original position, reverses direction, or ends the trial unsuccessfully. We identified four games that featured vertical meters. *Palia*, by Singularity 6, included all four of the target selection test types we include in our experiment. The vertical meter styled cooking stimulus is shown in Figure 2a. Other games including vertical meter RT tests were: the *Spin N' Sauce* minigame in *Papa's Sushiria To Go!* by Flipline Studios, where the player must tap the "Spin" button when the topmost light bulb is lit up on an oscillating power meter to win a prize [12]; the *Barrel Daredevil* minigame in Nintendo's *Wii Party*, where players respond to a falling barrel as close to the target line as possible [32]; and the *Study Fall* minigame in *Mario Party DS*, also by Nintendo, where players halt their character's vertical fall as close to the target line as possible [31].

3.2 Horizontal Meter

The horizontal meter differs from the vertical meter only in that it is rotated 90 degrees. We consider this rotation sufficiently different to warrant a separate display type, based on evidence that horizontal and vertical processing speeds differ [7]. *Toontown Rewritten* by the Toontown Rewritten Team includes a horizontal meter RT test during the CEO boss battle. The meter increases from left to right as the player holds a key, with the target area at the right end of the meter, with some room for error. Releasing the key in the target area increases attack damage against the boss [48]. See Figure 2d. Other horizontal meters include: the loom mechanic in *Spiritfarer* by Thunder Lotus Games; the drink station in *Papa's Sushiria To Go!* by Flipline Studios; *Thirsty Suitors* by Annapurna Interactive [3], where one of the main interaction mechanics of the game consists of reaction time tasks; and *Palia*, where the horizontal meter is another of the tests given to players when cooking [43].

3.3 Circular Meter

Like vertical meters, circular meters have a 1D path which contains the target area and that the indicator moves within. Unlike a vertical meter, the path is a circle, and when the indicator reaches the end position, it simply continues along the continuous path. We found five games with circular meters, including *Dredge*, a fishing game by Team17, which includes a circular meter (among others) where players catch fish by pressing a button while the indicator is within a target area. See Figure 2b. The other games with circular meters include: *Pop the Lock* by Simple Machine LLC, where the circular meter simulates a combination lock [42]; the *Ticktock Hop* minigame in Nintendo's *Mario Party Superstars*, where players must press a button when the hands of a clockface reach their character [35]; *Palia*'s cooking mechanic again [43]; and the sauce station in *Papa's Wingeria To Go!* by Flipline Studios, where the player must react at indicated times to correctly distribute the sauce on the wings they prepare [11].

3.4 Pulsating Ring

The pulsating ring differs from the other displays in that instead of varying an indicator position, the indicator radius varies. The user responds when the indicator is within the target area. Like the vertical meter, when the indicator reaches the meter's edge, it resets to its original radius, reverses scaling direction, or ends the trial unsuccessfully. *Horizon Forbidden West*'s "Burning Shores" expansion, by Sony Interactive Entertainment, added an example of a pulsating ring. When players attempt a particular attack, a ring representing the target area appears, along with an indicator ring that decreases in size. The player must press the attack button when the indicator is in the target area [44]. See Figure 2c. Other pulsating ring examples include: the mechanic for apprehending a stolen vehicle in *Marvel's Spider-Man 2*, also by Sony Interactive Entertainment [45]; the fishing mechanic in *Fire Emblem: Three Houses* by Nintendo [34]; the salvaging mechanic in *Xenoblade Chronicles 2*, also by Nintendo [33]; and *Thirsty Suitors* by Annapurna Interactive, where another of the main interaction mechanics includes a pulsating ring target selection test [3].

4 Methodology

There is little empirical work investigating target selection or response time with the common video game meters described above. Prior studies have not looked at stimulus factors or input methods, despite evidence that a link is likely present [1, 23, 24, 49]. We developed software to present response tasks across varying display types and with different input methods to respond to meters similar to those described above. Since prior work [5, 15, 22] indicates that indicator speed and target area size matter, our testbed also supports systematic variation of these factors.

4.1 Participants

We recruited 25 participants (aged 19-43, mean age of 24.9 years, *SD* = 5.78, 18 women, 6 men, 1 non-binary), all with (corrected-to-)normal vision. They spent an average of 7.0 hours (*SD* = 9.0) per week playing video games, with 3.7 hours (*SD* = 5.0) per week playing reaction- or timing-based video games. We assessed their dominant hand using the Edinburgh Handedness Inventory, and prior experience with keyboards, mice, touchpads, and game controllers. All participants were right-handed or ambidextrous, and were familiar with all input methods used in the study.

4.2 Apparatus

4.2.1 Hardware. We used a desktop PC with an Intel Core i7-7700K CPU @ 4.20GHz and 32GB RAM running Windows 10, along with four input devices used to vary the input method during trials. The input devices included a Logitech G810 Orion Spectrum keyboard for keypress trials, a Logitech G400s optical mouse for the mouse click trials, an Apple Magic Trackpad for the touchpad tap trials, and an Xbox 360 game controller for the controller button trials.

4.2.2 Software. We developed custom software using Unity and C# to run trials and record performance metrics. The software would display one the four stimulus types seen in Figure 1 depending on the experiment condition, each representing one of the major classes of common tasks seen in video games outlined above. Each set of trials began with the software displaying instructions to the participant, indicating which input method and display type was next. See Figure 3a. Participants could take a break if needed at this time. Upon responding with the appropriate input for the current condition, the first trial would begin. When a trial began, the associated display type appeared on the screen. A trial ended when the participant responded with the current input method, regardless of accuracy (i.e., whether the marker was in the target area or not). The software displayed the current indicator beginning in the default position. The target area position was randomized, as seen in Figure 3b. The indicator started moving as soon as the trial began. Participants were asked to press the input method control while the indicator was within the target area. See Figure 3.

Each trial set was characterized by the input method and display type condition. The software randomized indicator speeds and target area sizes ensuring all combinations were presented. Prior to beginning recorded trials, the software presented two practice trials for each display type and input method combination. For each display type, we selected indicator speeds such that a full indicator cycle (e.g., from top to bottom for the vertical meter or a complete rotation for the circular meter) took the same amount of time across all stimulus styles, for a given indicator speed. The slow speed took four seconds per cycle for each display type, medium speed took two seconds per cycle, while fast speed took one second per cycle. Similarly, for each display type, we selected target area sizes to ensure that they took up the same proportion of the meter. This way, the time the indicator spent in the target region was consistent across all display types for a given speed-size combination. The large target size took up a quarter of the meter, while medium took up an eighth, and small took up a sixteenth.

The software counterbalanced display types according to a balanced Latin square within each input method to mitigate switching

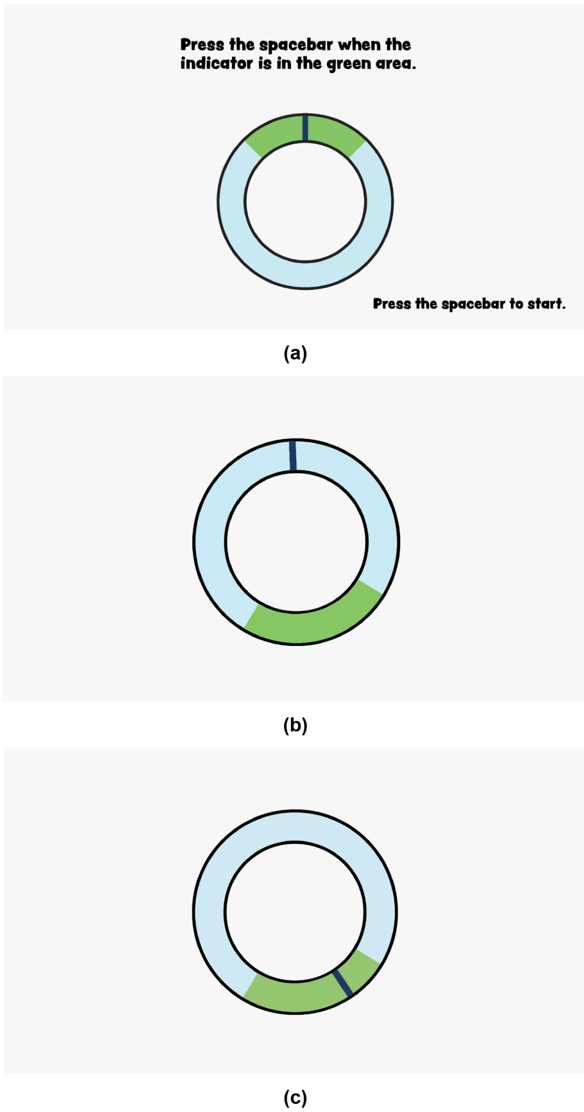


Figure 3: Software used in the study. (a) Sample instruction screen shown before trial set begins. (b) Circular meter trial with randomized target position and indicator at start position. (c) A correct response, with indicator within target area.

cost between devices [46]. So participants could not anticipate timing, each trial (10 per condition) had the target area randomly positioned within the meter. While these random positions introduced variety in best possible accurate RT for individual trials, this was averaged across the total number of trials that were measured per condition. This variety was caused by response time being bounded by the placement of the target, as even if the input were to be given the first instant it would be valid, enough time for the indicator to reach the target area would have to have passed already. However, across the high number of individual trials per independent variable for each participant, the risk of potential inconsistencies in target placement that may have occurred at a lower trial count and influenced results was mitigated. As such, we do not anticipate the randomization of the target placement skewing the results for any of the conditions. The software recorded the participant's unique anonymous ID, along with each trial's combination, and the participant's response time and accuracy.

4.3 Procedure

We first welcomed participants and asked them to read and complete the informed consent form. They then filled out a preliminary questionnaire to confirm eligibility, and completed a demographic questionnaire. Next, participants were seated at the desk, positioned 65cm from the screen to control for distance. We then demonstrated the task and instructed participants to respond as quickly and accurately as possible to the visual stimulus by pressing the indicated control while the indicator was within the target region. We also described the different input methods, and gave participants a chance to ask any clarifying questions. They then completed practice trials for each level of input method crossed with display type.

After this briefing, the first set of recorded trials began. Recorded trials proceeded as described above, with each combination of input method and display type (in counterbalanced order) and each target area and speed combination (in randomized order). Participants completed 10 trials for each combination of conditions and could take breaks after each set of trials. This process was repeated for each indicator speed, within each display type, within each input method yielding a total of 1440 trials per participant.

Upon the completion of each set of ten trials, the software recorded participant ID, condition information, speed, and accuracy. Overall, each participant took around thirty minutes to complete all trials. We then thanked them for their time and provided $10 (CAD) as compensation.

To determine if these findings are supported by users' subjective experiences, we recruited an additional five participants to complete the same trial set followed by a qualitative questionnaire. They were asked to rank the display types and input methods from easiest to hardest to respond to both quickly and accurately, and provide a brief explanation why. Although this feedback came from a different participant pool, the trials were administered in the same way and so we do not anticipate this having an effect on the findings. Supporting this, these responses were consistent with spontaneous feedback from the original 25 participants.

4.4 Design

Our study employed a within-subjects design with the following independent variables and levels:

- *Display Type*: Vertical meter, circular meter, pulsating ring, and horizontal meter
- *Input Method*: Keypress, mouse click, touchpad tap, and controller button press
- *Indicator Speed*: Slow, medium, fast
- *Target Area Size*: Small, medium, large

This yielded a total of 144 conditions. The dependent variables included accuracy and response time. Accuracy indicated whether a trial was accurate or not, based on whether the participant's response occurred while the indicator was within the target region

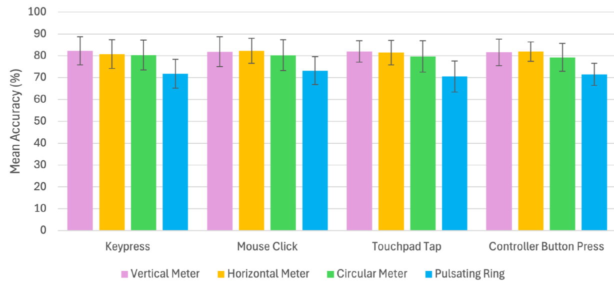


Figure 4: Mean accuracy by input method and display type. Error bars show ±1 standard deviations.

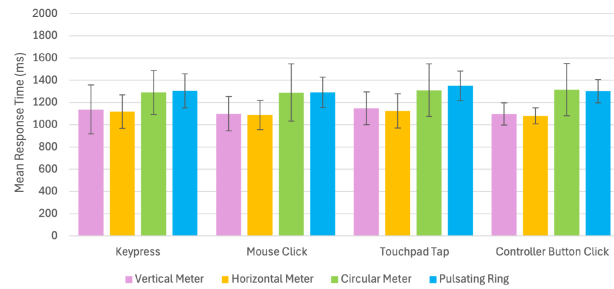


Figure 6: Mean response time by input method and display type. Error bars show ±1 standard deviations.

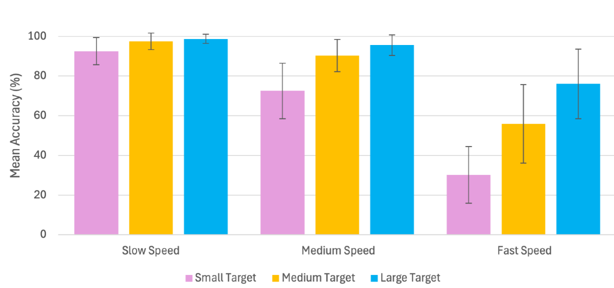


Figure 5: Mean accuracy by indicator speed and target area. Error bars show ±1 standard deviations.

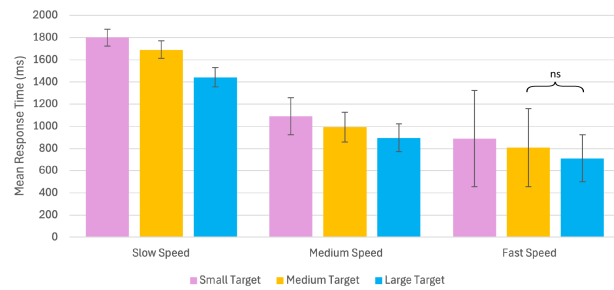


Figure 7: Mean response time by indicator speed and target area size. Error bars show ±1 standard deviations.

(accurate) or outside (inaccurate). Response time was measured in milliseconds, with the timer beginning when the visual indicator appeared on the screen and stopped when the participant's response was registered (regardless of accuracy). Our analysis only used response time scores for accurate trials. For each participant, a mean RT score was calculated per combination. These RT scores were used in the analyses. Similarly, a mean accuracy score was calculated for each participant per combination for analyses. Input methods were counterbalanced according to a balanced Latin square; we further counterbalanced display type order within each input method. Indicator speeds and target area sizes were randomized. Order effects (for learning, fatigue, etc.) were not examined.

As a four-way ANOVA is difficult to interpret and dividing our data to that level of granularity would yield only 250 trials per condition (reducing reliability of results), we opted to instead analyze the input method and display type independently of the indicator speed and target area size. This allowed us to obtain more reliable and interpretable results from each of the variable pairings and ultimately draw stronger conclusions.

5 Results

Across all participants, we ran a total of 36000 trials. Accuracy is the percentage of trials that ended with the indicator inside the target region. The grand mean accuracy was 79% and mean accuracy by combination is seen in Figure 4 and Figure 5.

Accuracy violated Mauchly's test of sphericity, so we employed Greenhouse-Geisser correction to a repeated measures ANOVA. There was a significant main effect of display type on accuracy, ($F_{1.988,47.714} = 51.312$, $p < .05$, $\eta_p^2 = .681$). The effect of input method was not significant ($F_{3,72} = 0.555$, ns, $\eta_p^2 = .023$), nor was the interaction effect ($F_{9,\,216} = 0.707$, ns, $\eta_p^2 = .029$). There were also significant main effects for both indicator speed ($F_{1.238,29.700} = 523.614$, $p < .05$, $\eta_p^2 = .956$) and target area size ($F_{1.502,36.057} = 591.629$, $p < .05$, $\eta_p^2 = .961$). Their interaction effect was also significant ($F_{2.250,53.990} = 115.409$, $p < .05$, $\eta_p^2 = .828$). Bonferroni post-hoc testing revealed that all speed-size pairs were significantly different from all other speed-size pairs in terms of average accuracy. See Figure 5.

As noted above, response time included only trials that ended successfully (i.e., accurate trials). The grand mean response time was 1202ms, (SD = 742ms). See Figure 6 and Figure 7 for mean response times by combination.

Since response time also violated Mauchly's test of sphericity, we again used a Greenhouse-Geisser corrected ANOVA. There was a significant main effect of display type on response time, ($F_{1.873,44.959} = 59.028$, $p < .05$, $\eta_p^2 = .711$). The effect of input method was not significant ($F_{3,72} = 2.305$, ns, $\eta_p^2 = .088$), nor was there a significant interaction effect ($F_{9,\,216} = 0.723$, ns, $\eta_p^2 = .029$). There were significant main effects of both indicator speed ($F_{1.057,23.380} = 218.460$, $p < .05$, $\eta_p^2 = .901$) and target area size ($F_{1.280,30.710} = 106.470$, $p < .05$, $\eta_p^2 = .816$) on response time. The interaction effect between the two variables was also significant ($F_{1.641,39.386} = 8.947$, $p < .05$, $\eta_p^2 = .272$). Bonferroni post-hoc testing revealed that all speed-size pairs were significantly different from all other speed-size pairs in terms of average RT, with the exception of fast speed/medium target and fast speed/large target. See Figure 7.

In terms of qualitative results, participants who received the questionnaire were asked to rank the display types from easiest to hardest to respond to both quickly and accurately. 80% reported the circular meter stimulus was easiest, and all participants reported the pulsating ring was hardest. They were asked the same question regarding the input methods, and two each identified the keyboard and the controller as easiest, while the remaining participant identified the trackpad as easiest. Three participants identified the trackpad as hardest, while the remaining two indicated the mouse was hardest. Commonalities were identified in participant answers when asked to explain these choices, which are discussed below. These results were also supported by spontaneous feedback given by participants after trials.

6 Discussion

Our results contribute novel findings on simple game-like response tasks, and provide guidance to game developers towards improving user experience with in-game RT tasks.

6.1 Revisiting Research Questions

6.1.1 RQ1. Our results reveal that response time was significantly faster for both the vertical and horizontal meters compared to the circular meter and the pulsating ring. A potential explanation is that the vertical and horizontal meters are comparatively simpler than the circular meter and the pulsating ring since they are linear. Both meters only use translational indicator movement, rather than more visually complex operations (e.g., rotating or scaling). Having only the indicator's position change could help users more easily detect and respond when the indicator is in the target area. Both meters are also quite similar to one another, so it is not surprising that both would offer similar performance. Further systematic experimentation on variations on linear versus non-linear meters (e.g., varying the degree of curve from straight, to slightly curved, to substantially curved, etc.) could better determine if this is the source of the difference we found.

6.1.2 RQ2. While we aimed to examine how input method would effect RT, we failed to detect a statistically significant difference in response time between the four input methods. The RT scores across input methods are very similar (see Figure 6) suggesting little difference. We thus cannot conclude whether any of the input methods studied offer a faster response time than the others. This similarity may be because, although the input methods are distinct, all consisted of moving a finger using very similar motor actions. More different input methods (e.g., swinging or twisting a controller, common in party games relying on tasks such as those we studied), would likely reveal larger differences in RT.

6.1.3 RQ3. Our results failed to detect interaction effects across the display types and input methods. However, considering the accuracy of these factors individually, the pulsating ring yielded significantly worse accuracy than the other three stimuli. This may be because pulsating ring changed the indicator's scale rather than position or orientation, which may be less intuitive. All participants who completed the questionnaire reported that the pulsating ring was the hardest visual to respond to quickly and accurately. Several attributed this difficulty to the fact that they found this stimulus harder to track due to its unusual shape, with the indicator consisting of a ring with varying sizes rather than a moving bar. Specifically, one participant said "[...] because it was a "circle" line instead of a straight line, it felt a little harder to keep track of." Interestingly, most participants identified the circular meter as the easiest visual, with the shared reasoning of this meter's indicator being easier to follow, as it circles continuously and does not reset. As a result, it does not disappear from view when the next cycle begins. In particular, one participant said "I was able to always keep track of where the meter was [...] because the meter was always on screen." Despite this, a significant difference was not identified in this meter's accuracy. For input method, no significant differences were found between methods, and so a conclusion cannot be drawn. Similarly, there was no consensus across the qualitative questionnaires collected. No more than two of the five participants named any of the input methods as the easiest, and no more than three named any of the input methods the hardest.

6.1.4 RQ4. The fast indicator speed (1s/cycle), had a significantly lower average RT than the medium (2s/cycle), and both fast and medium were significantly lower than slow (4s/cycle). Similarly, the large target area size (1/4 of the meter) yielded a lower RT than the medium (1/8 of the meter), and the RT of both large and medium were lower than that of the small (1/16 of the meter) size. While intuitive, these results provide valuable insight to how these factors can affect response time. When a target takes up more of the meter, the indicator spends less time outside of it, and enters the target area sooner on any given trial. The user can simply respond sooner in trials with larger target areas than in those with smaller ones. Similarly, since the indicator spends more time in the target area, there is also more time for users to respond after detecting the indicator entering the area. With smaller targets, the indicator leaves the target area more quickly, and users must wait for the next cycle to try again. Finally, an interaction effect showed differences between the average RT of every speed-size pair. See Figure 7. Considering individual pairings provides insight into specific cases when designing RT tasks. For example, in general, faster speed and larger target size yields faster responses. The converse is also true.

Target size and indicator speed also influenced accuracy. Accuracy with the slow speed was significantly higher than both medium and fast, and medium was more accurate than fast. Accuracy was also higher with large target area compared to the medium and small; accuracy with medium was higher than small. Both effects were likely caused by the fact that the slower the speed or the larger the target, the more time the indicator spends in the target area, and so the longer the chance the user has to respond accurately within a cycle. In general, slower speeds with larger targets yielded higher accuracy; the more time the indicator spends in the target area per cycle, the higher the accuracy. Another observation that can be made by comparing the average accuracies of speed-size pairings is that the differences between the accuracies of the target area sizes become more pronounced the faster the indicator speed gets. Within each speed level, the small target always has the lowest accuracy, followed by the medium target, however the absolute value of these differences increases from the slow speed to the medium speed, and again from the medium speed to the fast speed. Along with this increase in speed and decrease in accuracy, also comes an increase in the size of the error bars, which depict the standard deviation. This shows that at faster speeds, and especially with smaller targets, the variability in player performance generally increases.

Through analyzing our results, we determined display type, indicator speed, and target size to each have an effect on both response time and accuracy. However, we found no significant effects for either in relation to the input method.

6.2 Application in Video Games

A broader goal of our study was to determine how various response task display types, input methods, indicator speeds, and target sizes could help game designers predict and control difficulty faced by players in these commonly occurring tasks. Our results suggest that if a designer wants to decrease game difficulty, then they should not use pulsating rings and instead consider the vertical, horizontal, or circular meters. Similarly, if the designer wishes to offer faster response times, then the vertical or horizontal meters are better suited than circular meter or pulsating ring. In terms of indicator speed and target area size, broader rules such as larger targets and slower speeds are generally more accurate, and larger targets and higher speeds generally offer lower RT can help guide game design.

6.2.1 Sample Application. To help illustrate how our findings could be applied, we describe a sample game design. Consider a combat game where the main mechanics for attacking enemies are response tasks similar to our experiment. Since vertical and horizontal meters generally offer faster responses, these stimuli could be used for weapons that lend themselves to faster attacks, such as daggers. Conversely, the circular meter and pulsating ring could be used with weapons that take longer to attack with, such as a magic staff. Knowing how different display types influence RT in these tasks can help designers ensure that they select the appropriate display type to match the intended gameplay. Similarly, indicator speed and target area size offer developers the ability to fine-tune game difficulty. For example, indicator speed or target area size could be adjusted so that attacks with low-level weapons are easier to succeed with, but deal less damage, while task difficulty and damage scale with weapon level. If a level one weapon has a slow indicator speed and a medium target size, and the designer intends to decrease accuracy for level two, our results show that the same test with a medium speed but a larger target would be less accurate. Understanding the parameters influencing response tasks can help designers make evidence-based decisions for their games.

6.3 Broader Applications in HCI

Beyond video games, the findings of this work can also be applied to other fields in HCI. With an increasing number of users relying on accessibility features for interacting with software [6], it is important to understand how diverse factors affect these features. An interaction method that is becoming more prevalent in HCI for accessibility is one button interaction [18, 29]. This method relies on the user interacting with the software with a single button, with other techniques being leveraged to give this single button access to more actions in the software. A common example is the use of single key scanning keyboards for text entry [25]. The response task structure we examine in our work could be applied here, with an indicator moving along a meter and the user responding when it is in the range of the target they want to select. Our study provides an analysis of variations in these tasks that can affect how long it takes and how accurately the user responds to such an interface, which can form a foundation for designing accessibility features.

6.4 Limitations and Future Work

When considering our findings, it is also important to note some limitations. Our participants were all healthy young adults currently enrolled in a post-secondary education, and so they are likely not representative of the broader population. The average age of participants was mid-twenties with a low standard deviation. This limits the application of our findings to the broader group of all video game players, specifically children and older adults, and invites further research to be done with a more representative sample. Future studies could include more diverse participant demographics, or a comparative study examining if age affects performance on these types of common video game response tasks.

Emerging technologies are another avenue in which future research following up on our results could be done. While the visual indicators and input methods we examined are reflective of standard game interfaces (such as on PCs or standard game consoles), they are not necessarily representative of more recent developments in the field, such as VR and AR technologies. Future works could examine such emerging technologies, looking specifically at new input or output methods. This could include examining the impacts of observing the stimuli through a VR headset, or responding through controls that rely on gestures or eye-tracking. Further exploration into the impacts of these newer technologies would help to expand on the findings of this work.

It is worth noting that while we counterbalanced our trial orders, we did not examine order effects, such as learning effects or fatigue. Future work could examine the potential impact of order effects on a similar study. Given the numerous significant effects we detected, future work could also look at expanding on any of these findings, such as determining why vertical and horizontal meters have a lower RT than the circular meter and pulsating ring, as mentioned previously. More systematic exploration of speed-size pairs could yield more precise thresholds for these changes, and potentially facilitate development of performance models similar to Fitts' law.

7 Conclusion

The findings we presented here outline numerous significant differences between the accuracy and RT of variations in game-like response tasks, especially when it comes to varying indicator speed and target area size. Through answering our research questions, we aimed to provide insight into the broader goal of leveraging the effects of different response methods, display types, indicator speeds, and target area sizes towards making more informed decisions when designing these tasks for video games. Overall, we determined and discussed multiple aspects of a response task's design that should be considered if aiming for a target difficulty in terms of response time, accuracy, or both, while inviting future work to be done to examine specific findings in more detail.

References

[1] Dragan Ahmetovic, Daniele Riboli, Cristian Bernareggi, and Sergio Mascetti. 2021. RePlay: Touchscreen Interaction Substitution Method for Accessible Gaming. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction* (Toulouse & Virtual, France) *(MobileHCI '21)*. Association for Computing Machinery, New York, NY, USA, Article 27, 12 pages. https://doi.org/10.1145/3447526.3472044

[2] Dennis Ang and Alex Mitchell. 2017. Comparing Effects of Dynamic Difficulty Adjustment Systems on Video Game Experience. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (Amsterdam, The Netherlands) *(CHI PLAY '17)*. Association for Computing Machinery, New York, NY, USA, 317–327. https://doi.org/10.1145/3116595.3116623

[3] Annapurna Interactive. 2023. Thirsty Suitors.

[4] Nicholas David Bowman, Justin Keene, and Christina Jimenez Najera. 2021. Flow Encourages Task Focus, but Frustration Drives Task Switching: How Reward and Effort Combine to Influence Player Engagement in a Simple Video Game. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 119, 8 pages. https://doi.org/10.1145/3411764.3445678

[5] Tien-Ting Chiu, Kuu-Young Young, Shang-Hwa Hsu, Chun-Ling Lin, Chin-Teng Lin, Bing-Shiang Yang, and Zong-Ren Huang. 2011. A Study of Fitts' Law on Goal-Directed Aiming Task with Moving Targets. *Perceptual and Motor Skills* 113, 1 (2011), 339–352. https://doi.org/10.2466/05.06.25.PMS.113.4.339-352 arXiv:https://doi.org/10.2466/05.06.25.PMS.113.4.339-352 PMID: 21987931.

[6] Jason Colman and Paul Gnanayutham. 2010. *One-button" brain-computer interfaces.* University of Trás-os-Montes and Alto Douro, 133–140.

[7] Xiaoyan Deng, Barbara E. Kahn, H. Rao Unnava, and Hyojin Lee. 2016. A "Wide" Variety: Effects of Horizontal versus Vertical Display on Assortment Processing, Perceived Variety, and Choice. *Journal of Marketing Research* 53, 5 (2016), 682–698. https://doi.org/10.1509/jmr.13.0151 arXiv:https://doi.org/10.1509/jmr.13.0151

[8] Sebastian Deterding, Marc Malmdorf Andersen, Julian Kiverstein, and Mark Miller. 2022. Mastering uncertainty: A predictive processing account of enjoying uncertain success in video game play. *Frontiers in Psychology* 13 (2022). https://doi.org/10.3389/fpsyg.2022.924953

[9] Budmonde Duinkharjav, Praneeth Chakravarthula, Rachel Brown, Anjul Patney, and Qi Sun. 2022. Image Features Influence Reaction Time: A Learned Probabilistic Perceptual Model for Saccade Latency. *ACM Trans. Graph.* 41, 4, Article 144 (jul 2022), 15 pages. https://doi.org/10.1145/3528223.3530055

[10] 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 (2009), 321–326. https://doi.org/10.1111/j.1467-8721.2009.01660.x arXiv:https://doi.org/10.1111/j.1467-8721.2009.01660.x PMID: 20485453.

[11] Flipline Studios. 2019. Papa's Wingeria To Go!

[12] Flipline Studios. 2020. Papa's Sushiria To Go!

[13] Thierry Hasbroucq and Yves Guiard. 1991. Stimulus-response compatibility and the Simon effect: Toward a conceptual clarification. *Journal of Experimental Psychology: Human Perception and Performance* 17, 1 (1991), 246–266. https://doi.org/10.1037/0096-1523.17.1.246

[14] Jutta Hild, Christian Kühnle, and Jürgen Beyerer. 2016. Gaze-based moving target acquisition in real-time full motion video. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications* (Charleston, South Carolina) *(ETRA '16)*. Association for Computing Machinery, New York, NY, USA, 241–244. https://doi.org/10.1145/2857491.2857525

[15] Jin Huang, Feng Tian, Xiangmin Fan, Xiaolong (Luke) Zhang, and Shumin Zhai. 2018. Understanding the Uncertainty in 1D Unidirectional Moving Target Selection. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173811

[16] Jin Huang, Feng Tian, Nianlong Li, and Xiangmin Fan. 2019. Modeling the Uncertainty in 2D Moving Target Selection. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 1031–1043. https://doi.org/10.1145/3332165.3347880

[17] Aditya Jain, Ramta Bansal, Avnish Kumar, and K D Singh. 2015. A comparative study of visual and auditory reaction times on the basis of gender and physical activity levels of medical first year students. *International Journal of Applied and Basic Medical Research* 5, 2 (May-Aug 2015), 124–127. https://doi.org/10.4103/2229-516X.157168

[18] Manisha Varma Kamarushi, Stacey L. Watson, Garreth W. Tigwell, and Roshan L. Peiris. 2022. OneButtonPIN: A Single Button Authentication Method for Blind or Low Vision Users to Improve Accessibility and Prevent Eavesdropping. *Proc. ACM Hum.-Comput. Interact.* 6, MHCI, Article 212 (Sept. 2022), 22 pages. https://doi.org/10.1145/3546747

[19] Hyunchul Kim, Kasper Hornbæk, and Byungjoo Lee. 2022. Quantifying Proactive and Reactive Button Input. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 40, 18 pages. https://doi.org/10.1145/3491102.3501913

[20] Byungjoo Lee, Sunjun Kim, Antti Oulasvirta, Jong-In Lee, and Eunji Park. 2018. Moving Target Selection: A Cue Integration Model. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173804

[21] Injung Lee, Sunjun Kim, and Byungjoo Lee. 2019. Geometrically Compensating Effect of End-to-End Latency in Moving-Target Selection Games. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300790

[22] Jongsung Lee and Seung-Kweon Hong. 2022. Time to Capture a Moving Target Travelling along a Circular Trajectory. *Applied Sciences* 12, 4 (2022). https://doi.org/10.3390/app12041911

[23] Cheng-Jhe Lin and Changxu Wu. 2013. Reactions, accuracy and response complexity of numerical typing on touch screens. *Ergonomics* 56, 5 (2013), 818–831. https://doi.org/10.1080/00140139.2013.767384

[24] Kent Lyons. 2016. Visual Parameters Impacting Reaction Times on Smartwatches. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Florence, Italy) *(MobileHCI '16)*. Association for Computing Machinery, New York, NY, USA, 190–194. https://doi.org/10.1145/2935334.2935344

[25] I Scott MacKenzie. 2009. The one-key challenge: searching for a fast one-key text entry method. In *Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility*. 91–98.

[26] I. Scott Mackenzie. 2013. *2.7 Human Performance.* Morgan Kaufman, 54–67.

[27] I. Scott Mackenzie. 2013. *5 Designing HCI Experiments.* Morgan Kaufman, 157–189.

[28] Jeff Miller, Roland Schäffer, and Steven A. Hackley. 1991. Effects of preliminary information in a Go versus No-go task. *Acta Psychologica* 76, 3 (1991), 241–292. https://doi.org/10.1016/0001-6918(91)90022-R

[29] Takahiro Miura, Masatsugu Sakajiri, Murtada Eljailani, Haruo Matsuzaka, Junji Onishi, and Tsukasa Ono. 2014. Accessible Single Button Characteristics of Touchscreen Interfaces under Screen Readers in People with Visual Impairments. In *Computers Helping People with Special Needs*, Klaus Miesenberger, Deborah Fels, Dominique Archambault, Petr Peňáz, and Wolfgang Zagler (Eds.). Springer International Publishing, Cham, 369–376.

[30] Milos Mudric, Ivan Cuk, Aleksandar Nedeljkovic, Srecko Jovanovic, and Slobodan Jaric. 2015. Evaluation of Video-based method for the measurement of reaction time in specific sport situation. *International Journal of Performance Analysis in Sport* 15, 3 (2015), 1077–1089. https://doi.org/10.1080/24748668.2015.11868852 arXiv:https://doi.org/10.1080/24748668.2015.11868852

[31] Nintendo. 2007. Mario Party DS.

[32] Nintendo. 2010. Wii Party.

[33] Nintendo. 2017. Xenoblade Chronicles 2.

[34] Nintendo. 2019. Fire Emblem: Three Houses.

[35] Nintendo. 2021. Mario Party Superstars.

[36] Makoto Otaki and Katsuyuki Shibata. 2019. The effect of different visual stimuli on reaction times: a performance comparison of young and middle-aged people. *Journal of Physical Therapy Science* 31, 3 (Mar 2019), 250–254. https://doi.org/10.1589/jpts.31.250

[37] Alvero Pascuale-Leone, Joaquim P. Brasil-Neto, Josep Valls-Solé, Leonardo G. Cohen, and Mark Hallett. 1992. Simple reaction time to focal transcranial magnetic stimulation: Comparison with reaction time to acoustic, visual and somatosensory stimuli. *Brain* 115, 1 (02 1992), 109–122. https://doi.org/10.1093/brain/115.1.109 arXiv:https://academic.oup.com/brain/article-pdf/115/1/109/836407/115-1-109.pdf

[38] Diana Rieger, Tim Wulf, Julia Kneer, Lena Frischlich, and Gary Bente. 2014. The winner takes it all: The effect of in-game success and need satisfaction on mood repair and enjoyment. *Computers in Human Behavior* 39 (2014), 281–286. https://doi.org/10.1016/j.chb.2014.07.037

[39] Maria Francesca Roig-Maimó, I. Scott MacKenzie, Cristina Manresa-Yee, and Javier Varona. 2017. Evaluating fitts' law performance with a non-ISO task. In *Proceedings of the XVIII International Conference on Human Computer Interaction* (Cancun, Mexico) *(Interacción '17)*. Association for Computing Machinery, New York, NY, USA, Article 5, 8 pages. https://doi.org/10.1145/3123818.3123827

[40] Cristina Sanchez, Vanina Costa, Rodrigo Garcia-Carmona, Eloy Urendes, Javier Tejedor, and Rafael Raya. 2021. Evaluation of Child–Computer Interaction Using Fitts' Law: A Comparison between a Standard Computer Mouse and a Head Mouse. *Sensors* 21, 11 (2021). https://doi.org/10.3390/s21113826

[41] Valentin Schwind, Jan Leusmann, and Niels Henze. 2019. Understanding Visual-Haptic Integration of Avatar Hands Using a Fitts' Law Task in Virtual Reality. In *Proceedings of Mensch Und Computer 2019* (Hamburg, Germany) *(MuC '19)*. Association for Computing Machinery, New York, NY, USA, 211–222. https://doi.org/10.1145/3340764.3340769

[42] Simple Machine, LLC. 2015. Pop the Lock.

[43] Singularity 6. 2024. Palia.

[44] Sony Interactive Entertainment. 2022. Horizon Forbidden West: Burning Shores.

[45] Sony Interactive Entertainment. 2023. Marvel's Spider-Man 2.
[46] Rachel Swainson, Laura Prosser, Kostadin Karavasilev, and Aleksandra Romanczuk. 2021. The effect of performing versus preparing a task on the subsequent switch cost. *Psychol Res* 85, 1 (Feb 2021), 364–383. https://doi.org/10.1007/s00426-019-01254-7
[47] Team 17. 2023. Dredge.
[48] The Toontown Rewritten Team. 2014. Toontown Rewritten.
[49] Rasoul Amini Vishteh, Ali Mirzajani, Ebrahim Jafarzadehpour, and Samireh Darvishpour. 2019. Evaluation of Simple Visual Reaction Time of Different Colored Light Stimuli in Visually Normal Students. *Clinical Optometry* 2019:11 (December 2019), 167–171. https://doi.org/10.2147/OPTO.S236328
[50] Irem Gökçe Yildirim. 2016. Time Pressure as Video Game Design Element and Basic Need Satisfaction. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI EA '16)*. Association for Computing Machinery, New York, NY, USA, 2005–2011. https://doi.org/10.1145/2851581.2892298
[51] Yawen Zheng, Jin Huang, Juan Liu, Chenglei Yang, and Feng Tian. 2021. A Scenario Adaptive Model for Predicting Error Rates in Moving Target Selection on Smartphones. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction* (Toulouse & Virtual, France) *(MobileHCI '21)*. Association for Computing Machinery, New York, NY, USA, Article 10, 15 pages. https://doi.org/10.1145/3447526.3472049
[52] Gal Ziv, Ronnie Lidor, and Oron Levin. 2022. Reaction time and working memory in gamers and non-gamers. *Scientific Reports* 12, 1 (2022), 6798. https://doi.org/10.1038/s41598-022-10986-3

Received 7 March 2025