The Reliability of Fitts's Law as a Movement Model for People with and without Limited Fine Motor Function

Ather Sharif
asharif@cs.washington.edu
Paul G. Allen School of Computer Science & Engineering,
University of Washington
Seattle, Washington

Victoria Pao
paov@uw.edu
Department of Psychology,
University of Washington
Seattle, Washington

Katharina Reinecke
reinecke@cs.washington.edu
Paul G. Allen School of Computer Science & Engineering,
University of Washington
Seattle, Washington

Jacob O. Wobbrock
wobbrock@uw.edu
The Information School,
University of Washington
Seattle, Washington

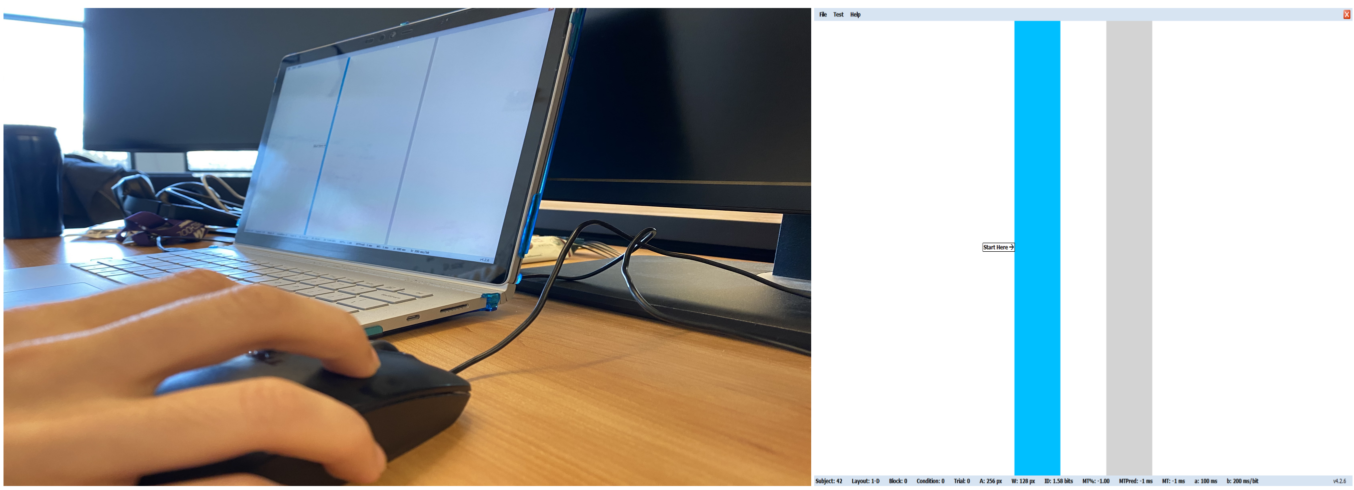

Figure 1: (left) Participant using a mouse to perform reciprocal pointing tasks. (right) Screen shot of a 1-D reciprocal pointing task from the *FittsStudy* program [68] showing two vertical ribbon targets. The starting target is highlighted in blue. A label with the text "Start Here" is shown indicating where to begin the series of pointing trials.

ABSTRACT

For over six decades, Fitts's law (1954) has been utilized by researchers to quantify human pointing performance in terms of "throughput," a combined speed-accuracy measure of aimed movement efficiency. Throughput measurements are commonly used to evaluate pointing techniques and devices, helping to inform software and hardware developments. Although Fitts's law has been used extensively in HCI and beyond, its test-retest reliability, both in terms of throughput and model fit, from one session to the next, is still unexplored. Additionally, despite the fact that prior work has shown that Fitts's law provides good model fits, with Pearson correlation coefficients commonly at r=.90 or above, the model fitness of Fitts's law has not been thoroughly investigated for people who exhibit limited fine motor function in their dominant hand. To fill these gaps, we conducted a study with 21 participants with limited fine motor function and 34 participants without such limitations. Each participant performed a classic reciprocal pointing task comprising vertical ribbons in a 1-D layout in two sessions, which were at least four hours and at most 48 hours apart. Our findings indicate that the throughput values between the two sessions were statistically significantly different, both for people with and without limited fine motor function, suggesting that Fitts's law provides low test-retest reliability. Importantly, the test-retest reliability of Fitts's throughput metric was 4.7% lower for people with limited fine motor function. Additionally, we found that the model fitness

ASSETS '20, October 26–28, 2020, Virtual Event, Greece

ACM ISBN 978-1-4503-7103-2/20/10…$15.00
https://doi.org/10.1145/3373625.3416999

of Fitts's law as measured by Pearson correlation coefficient, r, was .89 (SD=0.08) for people without limited fine motor function, and .81 (SD=0.09) for people with limited fine motor function. Taken together, these results indicate that Fitts's law should be used with caution and, if possible, over multiple sessions, especially when used in assistive technology evaluations.

CCS CONCEPTS

• **Human-centered computing** → **Pointing devices**; **Accessibility design and evaluation methods**; User interface design.

KEYWORDS

Fitts's law, test-retest reliability, models, throughput, model fitness, mouse

ACM Reference Format:

Ather Sharif, Victoria Pao, Katharina Reinecke, and Jacob O. Wobbrock. 2020. The Reliability of Fitts's Law as a Movement Model for People with and without Limited Fine Motor Function. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '20), October 26–28, 2020, Virtual Event, Greece.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3373625.3416999

1 INTRODUCTION

Fitts's law [17] was introduced in 1954 – 66 years ago. Since then, numerous published works in the field of human-computer interaction (HCI) and in several other domains including ergonomics and psychology have adopted Fitts's law to both describe and predict movement time in aimed pointing movements. Fitts's law combines speed and accuracy into a single metric for pointing efficiency called "throughput," [45, 58, 69] which, to date, continues to serve in the creation and evaluation of new and existing pointing techniques, hardware devices, and software systems [6, 10, 13, 55, 59, 60]. Notably, Fitts's law's utilization goes beyond customary desktops and touch screens, as it has also been employed to assess the performance of state-of-the-art devices (such as AR/VR and Leap Motion [2, 51]) as well as in different environments, including virtual environments [12, 37], underwater environments [16, 35], and even under a microscope [28, 38]). Such a broad range of usage indeed shows just how widely adopted Fitts's law has become.

Despite its extensive usage, the test-retest reliability (a metric widely used in the scientific community to measure the consistency of a metric [53, 64]) of Fitts's law's metrics remains unexplored. Most Fitts's law studies consist of a single session per participant, and thus never confront the question of how reliable Fitts's law's throughput metric or model fits actually are, and how their reliability could affect the device or technique being evaluated. Furthermore, Fitts's law has seen extensive use in assistive technology evaluations, particularly for assistive pointing software and devices [27, 29, 54, 66]. But whether Fitts's law is suitable for such evaluations, especially for people with limited fine motor function, is an unresolved question.

In addition to the test-retest reliability of Fitts's throughput metric, our work also investigates the *model fitness* of Fitts's law, expressed via Pearson correlation coefficient, r, which indicates how well Fitts's law applies to the observed experiment data. Fitts's law states that the time to perform a pointing task is a function of its difficulty. In other words, the more difficult a task is the longer it takes. Mathematically, this relationship between the task difficulty (referred as "index of difficulty", ID) and movement time (MT) is represented using a regression equation based on the "Shannon formulation" [42, 43] (considered as the preferred formulation as per prior work [58]). The equation is presented below:

$$MT = a + b \cdot ID \tag{1}$$

In Equation 1, MT is movement time, a and b are fitted regression coefficients, and ID is the index of difficulty shown in Equation 2, below:

$$ID = log_2 \left(\frac{A}{W} + 1 \right) \tag{2}$$

In Equation 2, ID is index of difficulty, in bits, A is movement amplitude (i.e., distance), and W is target width.

Numerous past studies [4, 21, 43, 44, 58] indicate that Fitts's law provides good model fits in predicting human pointing performance, with Pearson correlation coefficients (r) often at .90 and above. However, the model fitness over subsequent sessions, and for people with limited fine motor function, is still unknown. Given Fitts's law's wide applicability in the development of assistive technology, it is important to determine whether its suitability as a model extends beyond a single session and holds for people with limited fine motor function.

We analyzed the test-retest reliability and model fitness of Fitts's law, both for people with and without limited fine motor function. We conducted a study with 21 people with and 34 people without limited fine motor function. The participants performed the ISO 9241-9 [33] pointing tasks in a 1-D layout using the *FittsStudy* program [68] (Figure 1). For calculating throughput, we employ both the mean-of-means approach [58] and the slope-inverse approach [69] using the traditional $A \times W$ experiment design. We also use Guiard's [26] $Form \times Scale$ experiment design, which limits varying either A or W from Equation 1, but not both together.

Our results indicate that the test-retest reliability of Fitts's law's throughput metric is low for both fine motor function groups, and about 4.7% lower for people with limited fine motor function. Additionally, our findings show that the model fitness of Fitts's law as measured by Pearson correlation coefficient, r, was .89 (SD=0.08) for people without limited fine motor function, which is in line with the results from previous studies [4, 21, 43, 44, 58]. However, for people with limited fine motor function, the model fitness was 8.9% lower, at r=.81 (SD=0.09). In light of our findings, we urge caution when employing Fitts's law, especially in evaluations of assistive technology software and devices. Fitts's law's metrics should be calculated over multiple sessions, as opposed to a single session.

2 FITTS'S LAW IN HCI

Before reviewing work related to our current investigation, for context, we provide a brief overview of Fitts's law in HCI. What today is known as Fitts's law [17] was introduced in 1954 when Paul Fitts published his seminal work on modeling aimed pointing movements. His 1954 paper introduced a reciprocal pointing task between two vertical "ribbon" targets, which required participants to alternately tap on targets of different widths (W) and at different amplitudes (i.e., distances) from each other (*A*). Fitts published a follow-up to this work in 1964 [18], extending the applicability of Fitts's law from serial to discrete tasks. Fitts's law was used in HCI for the first time in 1978 by Stuart Card *et al.* [11] in an empirical study that compared the performance of a mouse and an isometric joystick.

From an experiment employing Fitts's law, two outcomes are of interest to our work: *throughput* and Pearson correlation coefficients (*r*). The throughput measure (*TP*, in bits per second) from Fitts's law represents the efficiency of aimed pointing movements. This metric is calculated in one of two ways, each of which have been defended in the literature. In one way [58], the mean throughput from each condition is divided by the mean movement time from those same conditions. A condition's index of difficulty is computed as shown in Equation 2.

In a competing approach to calculating throughput [69], the inverse of the slope of the regression line fitted from Equation 1, above, is used. When the y-intercept a in Equation 1 is zero, these two methods of calculation result in the same throughput. When a is non-zero, they differ. Some prior work has shown practical benefits to computing throughput under the first approach [68].

Pearson correlation coefficient, r, communicates model fitness, and describes the correlation between the movement time (*MT*) and the index of difficulty (*ID*). It is a crucial measure in determining how well Fitts's law models the data generated from an experiment. In the domain of Fitts's law, a good model fit is expected to have a Pearson correlation coefficients (*r*) at .9 or greater, as per prior work [4, 21, 43, 44, 58], and anything below .9 does not show a reliable correlation between task movement time (*MT*) and index of difficulty (*ID*). The threshold for what constitutes a "good model fit" (r>=.9) was determined based on researchers' experiences over decades of studying these models. After over 60 years of seeing results with r>.9, and often even r>.95, we know that r values that are considerably lower constitute a significant departure from typical model fits. For example, results from Hrezo's work [32] show an overall positive Pearson coefficient statistically, but the author decided that even with a positive correlation of r=0.72, the model was not well fit to the data. In our work, we therefore apply the same threshold for determining a good model fit.

Following Guiard [26], there has been further questioning about confounds in Fitts's law experiments. In particular, Gori *et al.* [25] have pointed out that pooling and averaging data belonging to different amplitude-width (*A*, *W*) pairs for the same *ID* can introduce confounds in Fitts's law experiments. In our work, we heeded Gori *et al.*'s advice by avoiding such pooling, and by utilizing the effective index of difficulty ID_e, which is also recommended in prior work [58]. We did not, however, consider an alternative to Pearson's r, when validating Fitts's law as a model, nor did we employ stochastic sampling, using Fitts's law in line with current practice, which is what we wished to scrutinize.

In a typical experiment involving Fitts's law, researchers invite participants to a controlled laboratory environment and present them with pointing tasks under different conditions – for example, pointing devices, interaction techniques, environments, body postures, target sizes, or target distances. Then, researchers calculate throughput and utilize it to compare and evaluate systems and devices, or to compare user groups or user environments. Traditionally, the calculation of the throughput value, the single quantitative measure of pointing efficiency, involves testing participants in a single session each, as opposed to over multiple sessions. Thus, the question of test-retest reliability rarely arises. In this work, we address this gap by exploring the test-retest reliability of Fitts's law's throughput metric and model fitness. Additionally, we investigate whether Fitts's law as a model is suitable for people with limited fine motor function.

3 RELATED WORK

Fitts's law has been used widely within HCI and beyond. As of the date of this writing, Google Scholar indicates that Paul Fitts's original 1954 article has been cited 8,570 times[1]. We highlight the most relevant prior work related to the test-retest reliability of Fitts's law's throughput and model fitness. Broader surveys of Fitts's law's use in HCI are available in prior work [43, 58].

3.1 Test-Retest Reliability of Fitts's Throughput Metric

To date, various prior work utilizing Fitts's law has made use of its throughput metric to evaluate and compare, existing and new, devices and techniques. However, none have, to the best of our knowledge, examined the test-retest reliability of Fitts's law's throughput metric from one session to the next.

Harada *et al.* [29] conducted a longitudinal study lasting 2.5 weeks with people with and without limited fine motor function. They introduced their participants to their voice-based user interface control system and used Fitts's law to measure the improvement in the participants' target acquisition performance across 10 sessions. They found that over the sessions the performance of both people with and without limited fine motor function improved by at least 20%. Specifically, people without limited fine motor function (ranging from 25% to 49%) showed more improvement as compared to people with limited fine motor function (ranging from 24% to 40%). While their work is closely related to ours in terms of calculating throughput values over multiple sessions, they specifically looked for the effects of their system on the performance of the participants; hence, the difference between the throughput values was attributed to their system rather than the test-retest reliability of Fitts's throughput metric.

Brogmus [9] utilized the Fitts's law data collected by the Baltimore Longitudinal Study of Aging[2] (BLSA) from 1,318 subjects from 1960-1981 and tested 121 unique formulas based on the prior modifications of Fitts's law to determine the best formula based on the standard error of estimate. Then, they examined the effect of age and gender using this formula. They found the age had a statistically significant main effect on movement time but gender did not. However, their analyses did not explore the test-retest reliability of Fitts's throughput metric, and their participant pool only included people without limited hand function. Looser *et al.* [41] evaluated 3-D gaming environments as a viable tool for Fitts's law experiments and carried out Fitts's law studies on 11 participants.

[1] https://scholar.google.com/scholar?cites=13463669318867480633
[2] https://www.blsa.nih.gov/

They found that the Fitts's law's throughput metric from the traditional 1-D "ribbon tasks" (TP=5.5) and the 3-D game interface (TP=5.3) was within 4% of each other. However, their analyses and participant pool were similarly limited as Brogmus's work [9]. Similarly, Wobbrock *et al.* [66] conducted Fitts's law experiments to evaluate the performance of their pointing technique *Angle Mouse*. They found that the Fitts's law's throughput metric was better for people without limited fine motor function (TP=4.26) than that of people with limited fine motor function (TP=3.03). However, they only calculated the throughput for each pointing technique over a single experiment. Numerous research work [6, 10, 13, 55, 59, 60] have followed similar protocols – utilizing Fitts's law to compare performance between different factors, but none have, to the best of our knowledge, explored the consistency of throughput values from Fitts's law experiments across multiple sessions with the same conditions and factors.

3.2 Model Fitness for People with Limited Fine Motor Function

Numerous research projects have reported Pearson correlation coefficients (r) as model fits for Fitts's law, including for participants with limited fine motor function. However, no conclusive agreement has been reached on whether Fitts's law is suitable for people with limited fine motor function. Findings are mixed.

Wobbrock *et al.* [67] found that Fitts's law, both for people with and without limited fine motor function, accurately modeled pointing performance ($R^2 = 0.993$ and $R^2 = 0.969$, respectively) and crossing performance ($R^2 = 0.996$ and $R^2 = 0.987$, respectively). However, they noted that their data excluded error trials and as a result, the data from people with limited fine motor function was not sufficient to delineate a normal distribution of hits, which is assumed for Fitts's law models. Additionally, they found models using Crossman's recommended correction for error normalization [15, 65] to be very poor for people with limited fine motor function. Rao *et al.* [54] found Fitts's law to be a good predictor of pointing tasks for people with Cerebral Palsy and people without any known neurological or physical limitations, although there is no indication of how many, if any, of their subjects with Cerebral Palsy had limited fine motor function in their dominant hand. Furthermore, they differed from our work in their choice of pointing devices used for the experiment.

Gajos *et al.* [22] developed an alternative method for modeling pointing movements and reported that Fitts's law was a poor fit for people with limited fine motor function, but only tested three such people, who used different pointing devices. Our work presents findings from 21 participants with limited fine motor function, all using a mouse as the pointing device. Gump *et al.* [27] showed that pointing data from people with Cerebral Palsy, who had moderate to severe spasticity, resulted in a statistically significant main effect of ID on movement time (MT), which they consider as the adherence to Fitts's law. However, they noted that such adherence was achieved only after the exclusion of error trials, similar to Wobbrock *et al.* [67], which reduced their sample size to only four participants. On the contrary, our work has sufficient data for Fitts's law modeling, including adequate error trial percentages, increasing the power of our analysis to detect significant effects.

Our work contributes findings showing that the test-retest reliability of Fitts's law's throughput metric and model fitness are 4.7% and 8.9% lower, respectively, for people with limited fine motor function, than for people without such limitations. We arrive at these findings using both the $A{\times}W$ and Guiard's $Form \times Scale$ [26] experiment designs.

4 EXPERIMENT DESIGN

To examine the test-retest reliability of Fitts's law's throughput and model fitness, we conducted a 2×2 mixed within-between-subjects experiment with people with and without limited fine motor function. The experiment was conducted in-person at our laboratory or online, depending on the participant's preference.

4.1 Participants

We recruited 55 total participants for our study. Thirty-four participants reported no fine motor function limitations in their dominant hand. Among this group of 34 participants, 24 identified as female, 9 as male, and 1 as genderqueer. A preliminary study with 4 participants suggested a medium-to-large effect; we therefore used a power analysis to estimate this sample size. All participants were right-handed. Their average age was 23.0 (SD=3.3). Of the 21 participants who reported limited fine motor function in their dominant hand, 14 identified as female, 5 as male, and 2 as genderqueer. Seventeen participants were right-handed. The average age in this group was 43.81 (SD=17.4). Note that the 80^{th} percentile for age in this group was 63.8 years old, which is slightly lesser than the age at which human pointing performance has been shown to significantly affect pointing performance (at an average of 64 years old) as reported in prior work [30, 36, 56, 62, 63]. As we report below, this is in keeping with our findings, where *Age* did not have a statistically significant effect on either the throughput or the Pearson correlation coefficients (r) under investigation.

Out of the total participants, 41 had their sessions run in our laboratory setting, while 14 had their sessions run remotely over Google Hangouts or Microsoft Skype. Four participants (without limited fine motor function) who produced errors greater than 8% (N=4) were excluded from the analysis. One participant (with limited fine motor function) who inadvertently used a touchpad instead of a mouse was also excluded. After these exclusions, there were 20 and 30 participants with and without limited fine motor function in the study. The online participants were screened for having similar apparatuses as that used in our laboratory. Participants were compensated with a $20 Amazon gift card for about an hour of their time over two sessions.

Before each of the two sessions in the study, information was recorded from the participants including their gender identification, age, dominant hand, current cursor speed setting, fatigue level, stress level, and daily usage of a computer. Participants reported their fatigue and stress levels using a Likert scale from 1 ("no/minimal") to 7 ("severe/maximal"). For participants who identified as having limited fine motor function, their diagnosis and functional limitations were also recorded (see Appendix A, Table 6). After the end of the second session, additional information was recorded including open-ended input of possible external events (such as medications, emotions) that may have caused a performance difference between the two sessions, and feedback on improving future studies.

We recruited participants using word-of-mouth, snowball sampling, and advertisements through channels including social media, such as Facebook and Twitter, and from flyers posted in rehabilitation facilities. We also advertised on email distribution lists for people with disabilities.

4.2 Apparatus

The study was conducted on computers running the Microsoft Windows operating system. For the laboratory setting, we used a Microsoft Surface Book 2 laptop measuring 13.1" by 9" set as 3000 x 2000 resolution running the Windows 10 operating system using 8 GB RAM. The *FittsStudy* program [68] was used as a software testbed, which is publicly available for download and runs in full-screen mode to avoid any potential distractions from other applications. Participants who preferred to participate online were guided through the downloading and installation of FittsStudy. A mouse (Dell Optical Mouse for the laboratory setting) was used as the pointing device in the study by all participants, with its speed set to the default cursor speed setting for Microsoft Windows. (This value is a 10 on a scale of 1-20 in the Windows mouse control panel. It corresponds to a control-display gain value of about 5.4 [66].) The study sessions were observed by at least one author, who took detailed notes during the sessions.

4.3 Procedure

Each participant took part in two sessions, which were at least four and at most 48 hours apart [49]. In each session, participants performed 10 target acquisitions for 5 target widths (W: 8, 16, 32, 64, 128 px) $\times$ 3 target distances (A: 256, 384, 512 px), for a total of 15 $A \times W$ conditions and $10 \times 15 = 150$ target acquisitions. Each acquisition was a single attempt to click a 1-D vertical "ribbon target" (Figure 1). The target sizes and distances were decided based on typical icon sizes and the distances between the elements present in conventional user interfaces such as Web pages. To account for learning effects, the order of the conditions was randomized across sessions and participants, as is standard practice in Fitts's law studies [58]. Participants were instructed to perform the tasks as quickly as possible while conforming to an error rate between 4-8% [43], equating to a total of 6-12 target misses per session. Participants were given the choice to practice the tasks before the start of the session. However, none of the participants chose that option and found the instructions sufficient. Additionally, participants were encouraged to take breaks in between the $A \times W$ conditions but not during the trials.

4.4 Design and Analysis

The experiment was a mixed factorial design with the following factors and levels:

- Limited Fine Motor Function (LFMF), *between-subjects*: yes, no
- Session, *within-subjects*: 1, 2

Our dependent variables were throughput (bits/s) and Fitts's law's model fitness expressed as a Pearson correlation coefficient, r. We analyzed these dependent variables using a mixed-effects model analysis of variance [20, 40], which included the above factors, their interaction, and a covariate to control for *Age*. We also included a random effect for *Subject* to account for repeated measures across two sessions.

In a second analysis, we used a binary representation of model fitness (i.e., fit or not for Pearson correlation coefficient, $r \geq .90$ as per prior work [4, 21, 43, 44, 58]). For this analysis, we used mixed-effects logistic regression [24] with the same model as described above.

Participants were tested over 15 $A \times W$ conditions in each of the two sessions, resulting in a total of $10 \times 15 \times 2 = 300$ trials per participant. With 50 participants, a total of $50 \times 300 = 15,000$ trials were produced and analyzed in this study.

4.5 Approach to Calculating Throughput

4.5.1 Mean-of-means approach vs. Slope-inverse approach. As discussed above, two approaches to calculating throughput are the mean-of-means approach (TP_{avg}) [58] and the slope-inverse approach (TP_{inv}) [69]. Mean-of-means (Eq. 3) approach calculates throughput as an average of indices of difficulty (ID_e) divided by movement times (MT), whereas the slope-inverse approach (Eq. 4) calculates throughput as the reciprocal of the Fitts's linear regression slope. While the literature contains several discussions on the choice of one over the other [58, 68, 69], there has been no resolution to the question of which one is "correct."

$$TP_{avg} = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{ID_{e_i}}{MT_i} \right), \text{ where } N = |A| \times |W|, \tag{3}$$

In Equation 3, TP_{avg} is the throughput in bits/s, ID is the index of difficulty from Equation 2, MT is the movement time as in Equation 1, and A and W have their same meanings as in Equation 1.

$$TP_{inv} = 1/b \tag{4}$$

In Equation 4, TP_{inv} is the throughput in bits/s, and b is the slope parameter from Equation 1.

We compared the difference between the averages of TP_{avg} and TP_{inv} from both study sessions. For people *with* and *without* limited fine motor function, the differences, in bits/s, between TP_{avg} and TP_{inv} were 2.88 $vs.$ 4.83 (with), and 4.77 $vs.$ 7.28 (without), respectively. Furthermore, the standard deviation differences were 1.06 $vs.$ 2.43 (with), and 0.45 $vs.$ 2.90 (without), respectively. Thus, there were much greater differences between the two groups with the slope-inverse calculation approach than with the mean-of-means calculation approach for average throughputs. Furthermore, the standard deviations are smaller with the mean-of-means approach, with findings consistent with prior work [68].

Additionally, the computed TP_{avg} mean for people without limited hand function was 4.70 bps for the first session, 4.83 bps for the second session, and 4.77 bps overall. These values were within the expected range of 3.7-4.9 bps for people using a mouse suggested by Soukoreff and MacKenzie [58] based on reviews of prior work [34, 46–48, 50]. Unfortunately, such a range for comparison is undocumented for TP_{inv} and for people with limited hand function.

Therefore, in light of these findings, and in keeping with pragmatic recommendations [68], we utilize the mean-of-means approach for calculating throughput in this paper.

4.5.2 Guiard's $Form \times Scale$ Design. Guiard [26] argues that manipulating *both* A and W in a Fitts's law experiment can introduce potential confounds [68]. According to Guiard's argument, the experiment design should be $Form \times Scale$ instead of the traditional $A \times W$, where *Form* is the ID and *Scale* is either A or W, but not both. To remove such a potential confound, Guiard's practical recommendation is to hold one of either A or W constant and manipulate the other during a Fitts's law experiment [68].

In this work, we selected five target widths (W) and three target amplitudes (A). Given that holding W constant yields only three data points for A, which is not enough to calculate a meaningful and trustworthy correlation, following Guiard's recommendation [26], we performed additional analyses by holding A constant while allowing W to vary.

5 RESULTS

In this section, we present the results of the experiment focusing on the test-retest reliability of Fitts's law's throughput metric and model fitness for participants with and without limited fine motor function.

5.1 Test-Retest Reliability of Fitts's Throughput

Our results, shown in Table 1, indicate that the test-retest reliability of Fitts's throughput metric is low for both participant groups, but considerably lower (by about 4.7%) for participants with limited fine motor function (see Figures 2 and 3). Below we discuss our findings from the traditional $A \times W$ experiment design, as well as from Guiard's $Form \times Scale$ design [26].

5.1.1 Using Traditional $A \times W$ Design. Our results indicate a significant main effect of *Session* on throughput overall ($F(1,48)=24.08$, $p<.001$, Cohen's $d=1.42$). Specifically, the average throughput for Session 1 was 3.93 bits/s ($SD=1.21$). For Session 2, it was 4.09 bits/s ($SD=1.19$). The mean absolute throughput difference between sessions was 0.23 bits/s ($SD=0.17$). This result indicates that from one session to the next, the throughput results were significantly different, casting doubt on the test-retest reliability of Fitts's throughput metric.

There was also a significant main effect of *Limited Fine Motor Function (LFMF)* on throughput overall ($F(1,47)=34.30$, $p<.001$, Cohen's $d=1.71$). Specifically, participants with limited fine motor function had an average throughput of 2.88 bits/s ($SD=1.06$), while participants without limited fine motor function had an average throughput of 4.77 bits/s ($SD=0.45$). This result indicates that our participants with limited fine motor function had an average throughput about 39.6% lower than those without.

We also examined whether changes in Fitts's law's throughput were proportionally similar or different from one session to the next for participants in each fine motor function group. To do so, we examined the *Session* × *LFMF* interaction, but found it to be non-significant ($F(1,48)=1.05$, *n.s.*). However, the change in the throughput values between sessions, as shown in Table 2, was higher for participants with limited fine motor function (9.16%) compared to participants without (2.94%). In combination, this result indicates that although throughput values changed significantly from one session to the next, that change was proportional for participants in each of the fine motor function groups, but numerically higher for those with limited fine motor function. Specifically, the test-retest reliability of Fitts's law's throughput metric was 4.7% lower for participants with limited fine motor function.

Additionally, *Age* did not have a significant effect on throughput ($F(1,47)=0.69$, *n.s.*). Furthermore, for both participants with and without limited fine motor function, we examined self-reported fatigue and stress level but found them to have no significant effects.

5.1.2 Using Guiard's $Form \times Scale$ Design. Following Guiard's recommendation [26] to hold either A or W constant in a Fitts's law throughput analysis, we recalculated TP_{avg} while separately holding each of the A values (256, 384, 512) constant while manipulating the W values. We re-ran our analyses after excluding participants for whom the recalculated error rates were greater than 8%. This process excluded 5, 7, and 8 participants from our original 50 for each value of A, respectively. Table 2 shows the summary results of our additional analyses.

Our results indicate that for both groups, *LFMF* had a significant effect on throughput for all values of A. Similarly, *Session* also had a significant effect except for A=384, suggesting that the throughput values varied significantly between sessions when the target distances were either small or large. As for the traditional $A \times W$ design, Guiard's design did not show a significant effect for the *Session* × *LFMF* interaction, *Age*, and *Fatigue* or *Stress*. These results were consistent for each value of A that was held constant.

5.2 Model Fitness

Anderson-Darling [1] goodness-of-fit tests of normality showed that model fits, as Pearson's r values, were non-normal. Close inspection revealed these values to be conditionally lognormal, and so a logarithmic transformation was applied prior to analysis, as is common practice [3, 31, 39]. Examination of this log-transformed response indicated that it was indeed normal for the traditional $A \times W$ design (p=.07) as well as for Guiard's design (p>.5 for all A). For ease of communication, plots of this dependent variable are shown using the original, non-transformed values.

In addition to examining model fitness using Pearson correlation coefficients (r) as a continuous dependent variable (Table 3), we also analyzed the Pearson correlation coefficients (r) as a binary measure, classifying the data as a "good fit" if r was .90 or greater, and a "poor fit" otherwise [58]. Overall, our findings, as shown in Tables 4 and 5, suggest that model fitness differs significantly between the two *LFMF* groups, and is worse for participants with limited fine motor function (see Figure 4). Here, we present our findings utilizing the Pearson correlation coefficient, r, both as a continuous and a binary measure, using the traditional $A \times W$ design as well as Guiard's $Form \times Scale$ design [26].

5.2.1 Using the Traditional $A \times W$ Design. For this analysis, we treated Pearson correlation coefficient, r, model fits as a continuous dependent variable. *Session* had a significant main effect on Pearson

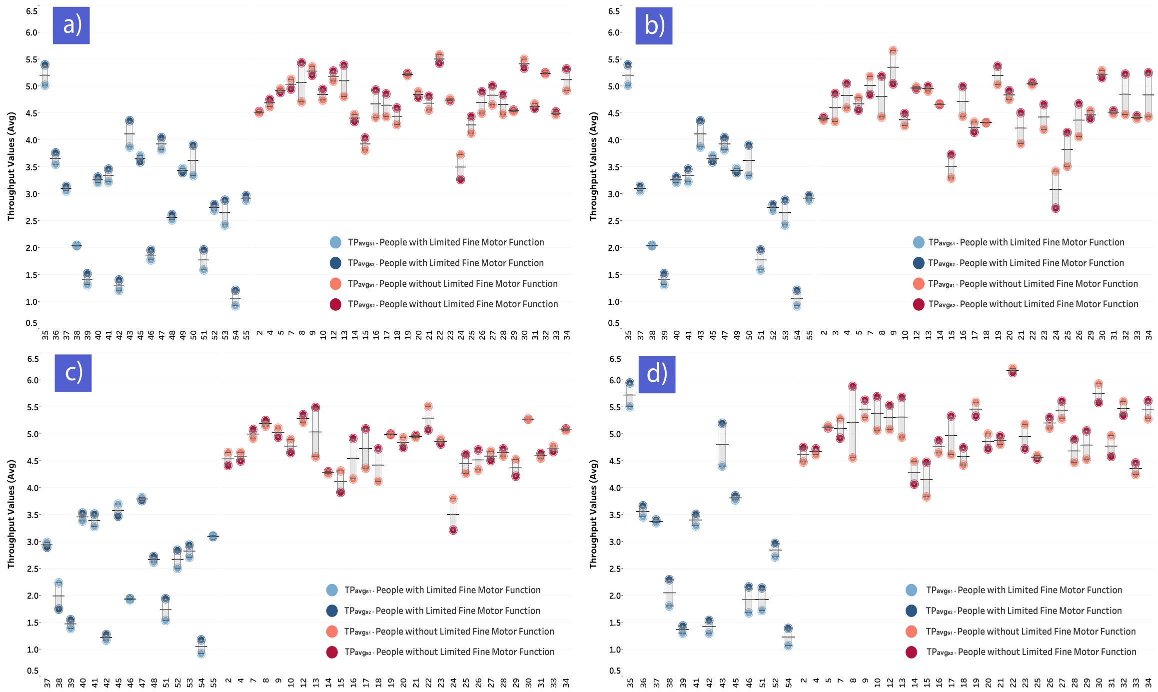

Figure 2: Differences in the TP_{avg} values across Session 1 and 2 for people with and without limited fine motor function using (a) Traditional $A{\times}W$ design, (b) Using Guiard's design (A=256), (c) Using Guiard's design (A=384), and (d) Using Guiard's design (A=512).

	Session					*LFMF*				
	df_n	df_d	F	p	Cohen's d	df_n	df_d	F	p	Cohen's d
Traditional $A \times W$	1	48	24.08	< .001	1.42	1	47	34.30	< .001	1.71
Guiard's $A = 256$	1	43	10.68	< .05	1.00	1	42	17.57	< .001	1.30
Guiard's $A = 384$	1	41	1.41	0.24	0.37	1	40	72.53	< .001	2.69
Guiard's $A = 512$	1	40	15.27	< .001	1.24	1	39	22.76	< .001	1.53

	Session $\times$ *LFMF*					*Age*				
	df_n	df_d	F	p	Cohen's d	df_n	df_d	F	p	Cohen's d
Traditional $A \times W$	1	48	1.05	0.31	0.30	1	47	0.69	0.41	0.24
Guiard's $A = 256$	1	43	0.11	0.74	0.10	1	42	1.94	0.17	0.43
Guiard's $A = 384$	1	41	0.05	0.83	0.07	1	40	0.75	0.39	0.27
Guiard's $A = 512$	1	40	1.33	0.26	0.37	1	39	2.05	0.16	0.46

Table 1: Summary results from 50 participants with and without limited fine motor function (LFMF) using a mixed-effects model analysis of variance [20, 40]. The statistical model was *TP = Session* $\times$ *LFMF + Age + Subject*, where *Subject* was modeled with a random intercept. Throughput values were calculated using both the traditional $A{\times}W$ design as well as Guiard's $Form \times Scale$ design [26]. Cohen's d is a measure of effect size [14].

correlation coefficients (F(1,48)=5.79, p<.05, Cohen's d=0.69). Specifically, the average Pearson correlation coefficient, r, for Session 1 was .85 (SD=0.09). For Session 2, it was .87 (SD=0.09). The mean absolute Pearson correlation coefficient, r, difference between sessions was 0.07 (SD=0.08). This result indicates a significant change in model fitness from one session to the next, calling the test-retest reliability of Fitts's law's model fitness into question.

Our results also present a significant main effect of *LFMF* on Pearson correlation coefficients (F(1,47)=8.76, p<.05, Cohen's d=1.35). Specifically, participants with limited fine motor function had an average r of 0.81 (SD=0.09), while participants without limited fine

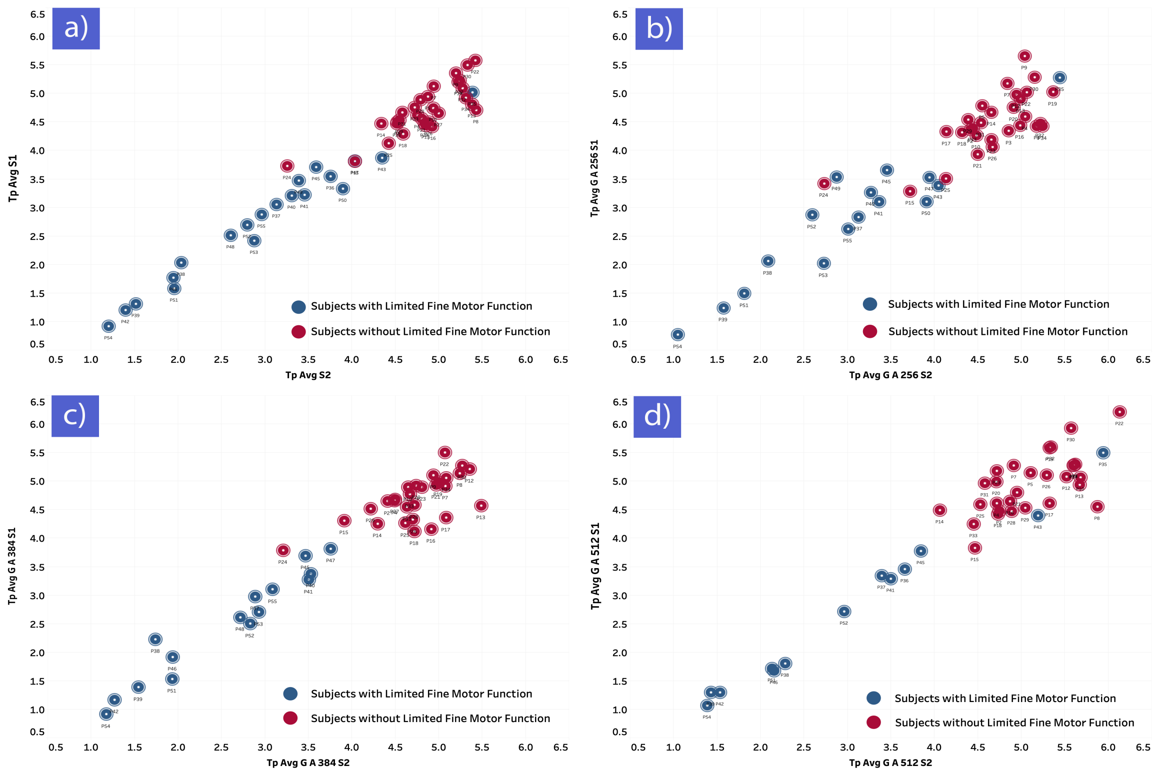

Figure 3: TP_{avg} values across Session 1 and 2 for people with and without limited fine motor function using (a) Traditional $A{\times}W$ design, (b) Using Guiard's design (A=256), (c) Using Guiard's design (A=384), and (d) Using Guiard's design (A=512).

	People *with* Limited Fine Motor Function (N=20)					
	N	TP_{s_1}	SD_{s_1}	TP_{s_2}	SD_{s_2}	ΔTP_{avg}
Traditional $A{\times}W$	20	2.78	1.06	2.98	1.08	0.22 (9.16%)
Guiard's $A = 256$	16	2.80	1.10	3.02	1.08	0.36 (11.30%)
Guiard's $A = 384$	15	2.48	0.93	2.56	0.88	0.19 (5.25%)
Guiard's $A = 512$	13	2.71	1.37	3.03	1.42	0.32 (14.60%)

	People *without* Limited Fine Motor Function (N=30)					
	N	TP_{s_1}	SD_{s_1}	TP_{s_2}	SD_{s_2}	ΔTP_{avg}
Traditional $A{\times}W$	30	4.70	0.44	4.83	0.46	0.23 (2.94%)
Guiard's $A = 256$	29	4.48	0.53	4.66	0.54	0.34 (4.38%)
Guiard's $A = 384$	28	4.69	0.40	4.74	0.47	0.27 (1.24%)
Guiard's $A = 512$	29	4.94	0.52	5.11	0.50	0.37 (3.91%)

Table 2: Summary results over (top) 20 subjects with limited fine motor function and (bottom) 30 subjects without limited fine motor function for the test-retest reliability of Fitts's law's throughput metric using 1-D reciprocal pointing tasks. ΔTP_{avg} shows the mean absolute and percentage difference between the throughput values obtained from Session 1 and 2 using the mean-of-means approach to throughput calculation [58].

motor function had an average r of 0.89 (SD=0.08). This result indicates that participants with limited fine motor function had an average Pearson correlation coefficient, r, that was 8.76% lower than those without limited fine motor function.

We examined the *Session* × *LFMF* interaction to investigate whether the Pearson's r was proportionally similar or different for each of the fine motor function groups over the two sessions. No statistically significant interaction effect was found (F(1,47)=1.37,

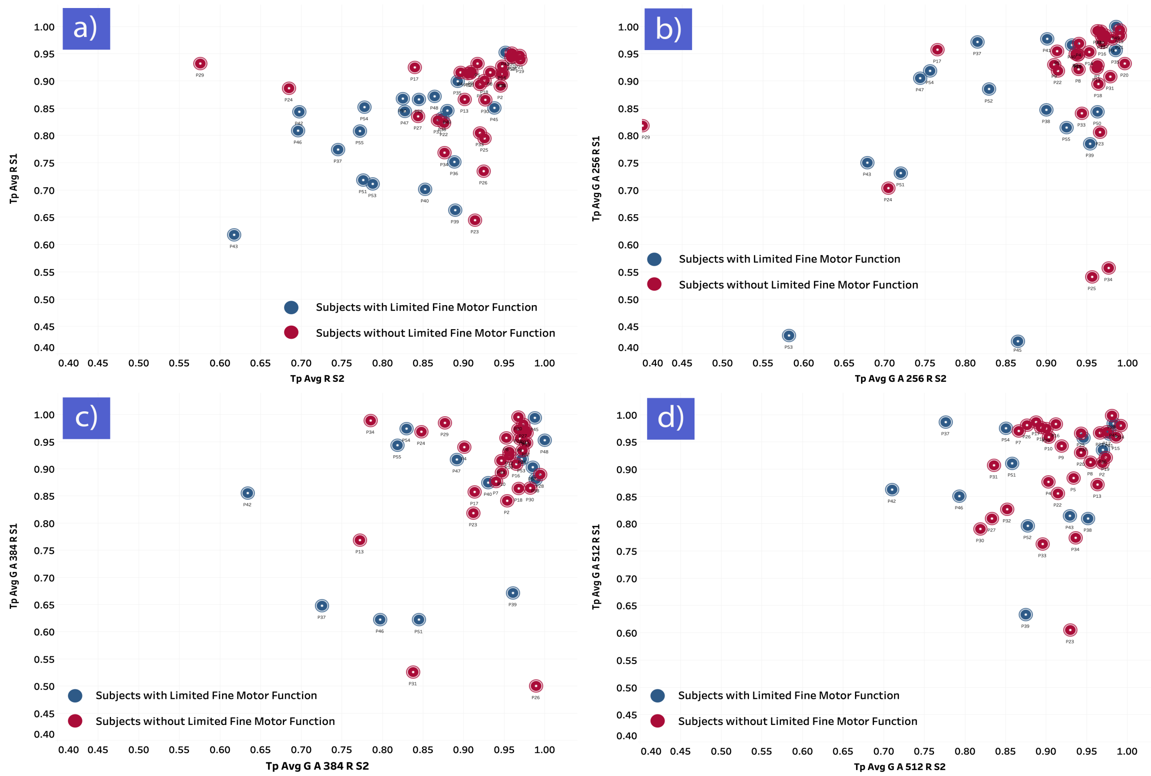

Figure 4: Pearson correlation coefficients (r) across Session 1 and 2 for people with and without limited fine motor function using (a) Traditional $A \times W$ design, (b) Using Guiard's design (A=256), (c) Using Guiard's design (A=384), and (d) Using Guiard's design (A=512).

n.s.). Similarly, *Age* did not have a significant effect (F(1,47)=3.58, *n.s.*).

Next, we performed a second analysis using a binary representation of Pearson's r. Prior work [4, 21, 43, 44, 58] generally indicates that r values at or above .90 indicate "good" model fits for Fitts's law. Out of the 20 participants with limited fine motor function, each of whom completed two sessions for a total of 40 sessions, 3 sessions (7.5%) had outcomes where r was at or above .90. Similarly, out of the 30 participants without limited fine motor function, each of whom also completed two sessions for a total of 60 sessions, 39 sessions (65%) had outcomes where Pearson's r was at or above .90. Our statistical tests show that only *LFMF* had a significant main effect on the Pearson correlation coefficients dichotomized as "good" or "poor" model fits (χ^2(1, N=50)=8.76, p<.05, Cohen's d=1.53). This result indicates that Fitts's law as a model fits poorly for participants with limited fine motor function, but fits well for participants without limited fine motor function. Tables 4 and 5 show the summary of these results.

In addition, for both analyses of model fitness and both groups, self-reported fatigue and stress level did not have any significant effect.

5.2.2 Using Guiard's Form × Scale Design. Similar to our calculations for the test-retest reliability of Fitts's law's throughput metric, following Guiard's recommendation [26], we regenerated our Pearson's r values, dependent variables and error rates separately for each of the target amplitudes (A). We re-ran our analyses after excluding participants for whom the recalculated error rates were greater than 8%. This process excluded 5, 7, and 8 participants from our original 50 for each value of A, respectively.

Using Guiard's design [26], the effect of our model terms was inconsistent across varying values of A, both when considering Pearson correlation coefficient, r, as a continuous measure and as a dichotomized "good" vs. "poor" fit measure. Our summary of these results can be seen in Tables 4 and 5.

6 DISCUSSION

Our experiment shows that Fitts's law's throughput metric and model fitness offer low test-retest reliability, both for people with and without limited fine motor function. Furthermore, test-retest reliability is lower for people with limited fine motor function. Additionally, Fitts's law's model fitness, as conveyed by Pearson correlation coefficients, is generally "good" ($r \geq .90$) for people without limited fine motor function (M=.89, SD=0.08), but generally "poor" ($r < .90$) for people with limited fine motor function (M=.81, SD=0.09). Taken together, these findings cast doubt on the suitability of Fitts's law as a movement model for evaluating assistive pointing devices and techniques.

	Session					*LFMF*				
	df_n	df_d	F	p	Cohen's d	df_n	df_d	F	p	Cohen's d
Traditional $A \times W$	1	48	5.79	$< .05$	0.69	1	47	21.40	$< .001$	1.35
Guiard's $A = 256$	1	43	0.01	0.94	0.02	1	42	4.08	0.05	0.62
Guiard's $A = 384$	1	41	4.30	$< .05$	0.65	1	40	7.89	$< .05$	0.89
Guiard's $A = 512$	1	40	0.23	0.64	0.15	1	39	0.45	0.51	0.22

	Session $\times$ *LFMF*					*Age*				
	df_n	df_d	F	p	Cohen's d	df_n	df_d	F	p	Cohen's d
Traditional $A \times W$	1	48	1.37	0.25	0.34	1	47	3.58	0.07	0.55
Guiard's $A = 256$	1	43	1.17	0.29	0.33	1	42	0.67	0.42	0.25
Guiard's $A = 384$	1	41	0.01	0.94	0.02	1	40	5.49	$< .05$	0.74
Guiard's $A = 512$	1	40	0	0.99	0	1	39	0.01	0.92	0.03

Table 3: Summary results from 50 participants with and without limited fine motor function (LFMF) using a mixed-effects model analysis of variance [20, 40]. The statistical model was *Pearson correlation coefficient,* r = *Session* × *LFMF* + *Age* + *Subject*, where *Subject* was modeled with a random intercept to account for repeated measures. Pearson correlation coefficients (r) were calculated using both the traditional $A{\times}W$ design as well as Guiard's $Form \times Scale$ design [26]. Cohen's d is a measure of effect size [14].

	Session				*LFMF*			
	N	χ^2	p	Cohen's d	N	χ^2	p	Cohen's d
Traditional $A \times W$	50	1.50	0.22	0.45	50	8.76	$< .05$	1.53
Guiard's $A = 256$	45	1.08	0.30	0.26	45	5.55	$< .05$	0.98
Guiard's $A = 384$	43	1.19	0.28	0.33	43	5.88	$< .05$	0.86
Guiard's $A = 512$	42	0.03	0.86	0.06	42	1.24	0.27	0.34

	Session $\times$ *LFMF*				*Age*			
	N	χ^2	p	Cohen's d	N	χ^2	p	Cohen's d
Traditional $A \times W$	50	0	0.96	0.24	50	2.06	0.15	0.33
Guiard's $A = 256$	45	1.08	0.30	0.26	45	0.58	0.45	0.25
Guiard's $A = 384$	43	1.19	0.28	0.33	43	3.54	0.06	0.68
Guiard's $A = 512$	42	0.26	0.61	0.17	42	0.01	0.93	0.03

Table 4: Summary results from 50 participants with and without limited fine motor function (LFMF) using mixed-model logistic regression [24]. The statistical model was *Pearson correlation coefficient,* r = *Session* × *LFMF* + *Age* + *Subject*, where *Subject* was modeled with a random intercept to account for repeated measures. Fitts's law as a model was considered to have a "good fit" when r was $\geq .90$. Pearson correlation coefficients (r) were calculated using both the traditional $A{\times}W$ design as well as Guiard's $Form \times Scale$ design [26]. Cohen's d is a measure of effect size [14].

	People *with* LFMF (N=20)			People *without* LFMF (N=30)		
	N	No. of Good Model-fits	%	N	No. of Good Model-fits	%
Traditional $A{\times}W$	40	3	7.50	60	39	65.00
Guiard's $A = 256$	34	14	41.18	58	48	82.76
Guiard's $A = 384$	33	18	54.55	56	40	71.43
Guiard's $A = 512$	33	17	51.52	58	39	67.24

Table 5: Results from 100 participant sessions – 20 participants with, and 30 participants without, limited fine motor function (LFMF) – from the Fitts's law experiment using 1-D reciprocal pointing tasks. Fitts's law models were considered to have "good" fitness when their Pearson correlation coefficient, r , was $\geq .90$. "No. Good Model Fits" columns represent the number of sessions out of total sessions for which this criterion was met.

Our analyses included the traditional $A{\times}W$ design as well as Guiard's $Form \times Scale$ design [26]. In our results, the $Form \times Scale$ design did not always produce consistent findings. For example, Guiard's design produced different main effects of *Session*, *Limited Fine Motor Function (LFMF)*, and *Age* for different target distances (A), as shown in Table 5. Given the inconsistency in findings from the various Guiard designs, we focus further discussion of our results based on the outcomes from the traditional $A{\times}W$ design, which mostly agreed with the Guiard designs.

It is important to emphasize that even though the test-retest reliability of Fitts's throughput metric is low for both participant groups, it is lower for people with limited fine motor function, as evident from the mean throughput differences between the two sessions for the two participant groups (2.94% $vs.$ 9.16%). Unsurprisingly, the throughput values for people with limited fine motor function are not only lower than those for people without limited fine motor function, but they also have greater variation in range (see Figures 2 and 3). This observation suggests that Fitts's law should be employed with caution when developing assistive pointing devices and techniques.

Interestingly, despite the significant overall differences between the throughput values across sessions, the differences for some participants in *both* groups were minimal (as depicted in Figure 2). In other words, for this subset of participants, Fitts's law's throughput metric was found to be reliable across the sessions. Similarly, the presence of participants in both participant groups for whom Fitts's law was indeed a good model (see Figure 4) suggests that while Pearson's r changes significantly across sessions for both groups, there are people for whom it is a "reliable" model. Future work could examine precisely what kinematic characteristics make for reliable Fitts's law models across sessions.

Our experiment confirms that Fitts's law does indeed produce good model fits for people without limited fine motor function (39 of 60 sessions produced good model fits), but not for people with limited fine motor function (only 3 of 40 sessions). In fact, the results from only one participant with limited fine motor function showed an r greater than .90 in both sessions, suggesting that Fitts's law is generally unsuitable for people with limited fine motor function. This result adds to the debate on the suitability of Fitts's law for people with limited fine motor function, in which some prior work has found good model fits for people with limited fine motor function [57, 66, 67], while other prior work has found the opposite [8, 22, 27]. It is worth remembering that in the context of Fitts's law, the Pearson correlation coefficient, r, represents the correlation between Fitts's law's effective indices of difficulty (ID_e) and participants' movement times (MT). Generally, the harder the pointing task, the higher the index of difficulty, and the higher the movement time. But for people with limited fine motor function, a low Pearson's r might be due to several other factors besides the index of difficulty of a pointing task. Spasticity, tremors, fatigue, limited range of motion, or the effects of medication all might make any given pointing task more difficult at any given time.

The fact that our results question the test-retest reliability of Fitts's law's throughtput metric and model fitness for people with and without limited fine motor function has important implications for past and future research. As of the date of this writing, a keyword search for "Fitts's law" on the ACM Digital Library (http://dl.acm.org/) shows that 1.5% of the total publications at ACM ASSETS[3], 2.5% at ACM TACCESS[4], and 2.4% of the total publications at ACM CHI[5] mention Fitts's law. Together, this constitutes a total of 511 publications in the ASSETS, TACCESS, and CHI proceedings that use or refer to Fitts's law. Depending on the capacity in which these publications employ Fitts's law, their claims could warrant further scrutiny based on our finding that Fitts's law provides a poor model for people with limited fine motor function. Additionally, given that our results indicate that the test-retest reliability of Fitts's law's throughput metric and model fitness is low for *both* fine motor function groups, the results from these publications could be called into question concerning their use of Fitts's law.

6.1 Recommendations

Based on our findings, we offer the following recommendations for future studies that intend to use Fitts's law to model human pointing performance:

(1) If time and resources allow, measure throughput over multiple sessions. Given that our throughput data were normally distributed, both for people with and without limited fine motor function, it is reasonable to calculate and utilize the mean throughput value for a given participant over multiple sessions.
(2) Similarly, calculate model fitness over multiple sessions. Unlike throughput, model fitness was lognormally distributed, so we recommend using the median Pearson correlation coefficient, r, for a given participant when judging the fitness of a Fitts's law model for that person.
(3) When quantifying human pointing performance for people with limited fine motor function, consider alternatives to Fitts's law. For example, speed and accuracy each can be reported individually. Custom regression equations can be fit to individual participants and involve interface-specific terms other than just target distance (A) and target size (W). For example, the Ability Modeler by Gajos et. al [22, 23] takes this approach.

7 FUTURE WORK

Our findings indicate that the test-retest reliability of Fitts's law's throughput metric and model fitness is low for both people with and without limited fine motor function. However, despite the overall statistically significant main effect of *Session* and *Limited Fine Motor Function* on throughput, there was the presence of participants in both fine motor function groups for whom the pointing performance, as measured via Fitts's law's throughput metric, *was* consistent over the two sessions. Similarly, when considering model fitness, 42 out of the 100 total participant sessions produced good model fits. Future work could therefore examine what specific kinematic characteristics of a person's movement make that movement suitable for modeling by Fitts's law. Prior kinematic analyses

[3]https://dl.acm.org/action/doSearch?AllField="fitts's law"&SpecifiedLevelConceptID=119685&expand=all
[4]https://dl.acm.org/action/doSearch?AllField="fitts's law"&ConceptID=118230&expand=all&SeriesKey=taccess
[5]https://dl.acm.org/action/doSearch?AllField="fitts's law"&SpecifiedLevelConceptID=119596&expand=all

[7, 19, 36, 52, 61] might be adapted to formulating individualized movement models for people with limited fine motor function. If the precise kinematic properties that make Fitts's law suitable can be discovered, then one could determine *a priori* which participants should and should not be modelable by Fitts's law, and why.

Additionally, our experiments relied on using a mouse as the pointing device, and calculated throughput values from 1-D reciprocal pointing tasks, as suggested by ISO 9241-9 [33]. While each participant used the same device and the same task dimensionality across both sessions, different settings for such factors are worth exploring. Hence, future work can examine Fitts's law's throughput metric and model fitness using a different pointing device, such as a trackball [5], and a different task dimensionality, such as the 2-D "ring of circles" task [44].

8 CONCLUSION

Fitts's law has been extensively utilized for decades to quantify human pointing performance, both within and outside the field of HCI. In this work, we conducted a Fitts's law experiment over two sessions with 50 participants with and without limited fine motor function. Specifically, we analyzed Fitts's law's throughput metric and model fitness for both populations using the traditional $A{\times}W$ experiment design, as well as Guiard's [26] $Form \times Scale$ design. Our results indicate that the test-retest reliability of Fitts's law's throughput metric and model fitness is low for both fine motor function groups, and lower for people with limited fine motor function. Furthermore, Fitts's law as a model fits relatively poorly for people with limited fine motor function. In light of our findings, we urge caution when employing Fitts's law in single sessions and for people with limited fine motor function. It is our hope that our findings and recommendations will help improve the measurement of human pointing performance, especially concerning the development and evaluation of assistive pointing devices and techniques.

ACKNOWLEDGMENTS

This work was supported in part by The Mani Charitable Foundation. We extend our gratitude to the AccessComputing staff for their support and assistance in the recruitment of participants. We would also like to thank the anonymous reviewers for their helpful comments and suggestions. Any opinions, findings, conclusions, or recommendations expressed in this work are those of the authors and do not necessarily reflect those of any supporter.

REFERENCES

[1] Theodore W Anderson and Donald A Darling. 1954. A test of goodness of fit. *Journal of the American statistical association* 49, 268 (1954), 765–769.

[2] Anil Ufuk Batmaz, Aunnoy K Mutasim, Morteza Malekmakan, Elham Sadr, and Wolfgang Stuerzlinger. 2020. Touch the wall: Comparison of virtual and augmented reality with conventional 2D screen eye-hand coordination training systems. In *The IEEE Conference on Virtual Reality and 3D User Interfaces (IEEE VR).*

[3] Donald A Berry. 1987. Logarithmic transformations in ANOVA. *Biometrics* (1987), 439–456.

[4] M Bertucco, P Cesari, and ML Latash. 2013. Fitts' Law in early postural adjustments. *Neuroscience* 231 (2013), 61–69.

[5] Barbara Bied Sperling and Thomas S Tullis. 1988. Are you a better" mouser" or" trackballer"? A comparison of cursor-positioning performance. *ACM SIGCHI Bulletin* 19, 3 (1988), 77–81.

[6] Renaud Blanch and Michael Ortega. 2011. Benchmarking pointing techniques with distractors: adding a density factor to Fitts' pointing paradigm. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 1629–1638.

[7] Reinoud J Bootsma, Laure Fernandez, and Denis Mottet. 2004. Behind Fitts' law: kinematic patterns in goal-directed movements. *International Journal of Human-Computer Studies* 61, 6 (2004), 811–821.

[8] Pedro E Bravo, Miriam LeGare, Albert M Cook, and Susan Hussey. 1993. A study of the application of Fitts' law to selected cerebral palsied adults. *Perceptual and motor skills* 77, 3_suppl (1993), 1107–1117.

[9] George Erich Brogmus. 1991. Effects of age and sex on speed and accuracy of hand movements: And the refinements they suggest for Fitts' Law. In *Proceedings of the Human Factors Society Annual Meeting*, Vol. 35. SAGE Publications Sage CA: Los Angeles, CA, 208–212.

[10] Jesse Burstyn, Juan Pablo Carrascal, and Roel Vertegaal. 2016. Fitts' Law and the Effects of Input Mapping and Stiffness on Flexible Display Interactions. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* 3649–3658.

[11] Stuart K Card, William K English, and Betty J Burr. 1978. Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT. *Ergonomics* 21, 8 (1978), 601–613.

[12] Logan D Clark, Aakash B Bhagat, and Sara L Riggs. 2020. Extending Fitts' law in three-dimensional virtual environments with current low-cost virtual reality technology. *International Journal of Human-Computer Studies* 139 (2020), 102413.

[13] Andy Cockburn and Andrew Firth. 2004. Improving the acquisition of small targets. In *People and Computers XVII—Designing for Society.* Springer, 181–196.

[14] Jacob Cohen. 1988. Statistical power analysis for the social sciences. (1988).

[15] ERFW Crossman. 1957. The speed and accuracy of simple hand movements. *The nature and acquisition of industrial skills* (1957).

[16] David Jud Dixon. 1985. *Underwater movement and Fitts' law.* Ph.D. Dissertation. Texas Tech University.

[17] Paul M Fitts. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology* 47, 6 (1954), 381.

[18] Paul M Fitts and James R Peterson. 1964. Information capacity of discrete motor responses. *Journal of experimental psychology* 67, 2 (1964), 103.

[19] Tamar Flash and Neville Hogan. 1985. The coordination of arm movements: an experimentally confirmed mathematical model. *Journal of neuroscience* 5, 7 (1985), 1688–1703.

[20] Brigitte N Frederick. 1999. Fixed-, Random-, and Mixed-Effects ANOVA Models: A User-Friendly Guide for Increasing the Generalizability of ANOVA Results. (1999).

[21] Naomi Friedlander, Kevin Schlueter, and Marilyn Mantei. 1998. Bullseye! when Fitts' law doesn't fit. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* 257–264.

[22] Krzysztof Z Gajos, Jacob O Wobbrock, and Daniel S Weld. 2007. Automatically generating user interfaces adapted to users' motor and vision capabilities. In *Proceedings of the 20th annual ACM symposium on User interface software and technology.* 231–240.

[23] Krzysztof Z Gajos, Jacob O Wobbrock, and Daniel S Weld. 2008. Improving the performance of motor-impaired users with automatically-generated, ability-based interfaces. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems.* 1257–1266.

[24] AR Gilmour, RD Anderson, and AL Rae. 1985. The analysis of binomial data by a generalized linear mixed model. *Biometrika* 72, 3 (1985), 593–599.

[25] Julien Gori, Olivier Rioul, Yves Guiard, and Michel Beaudouin-Lafon. 2018. The Perils of Confounding Factors: How Fitts' Law Experiments can Lead to False Conclusions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* 1–10.

[26] Yves Guiard. 2009. The problem of consistency in the design of Fitts' law experiments: Consider either target distance and width or movement form and scale. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 1809–1818.

[27] Alison Gump, Miriam Legare, and Deborah L Hunt. 2002. Application of Fitts' law to individuals with cerebral palsy. *Perceptual and motor skills* 94, 3 (2002), 883–895.

[28] Walton M Hancock, Gary Langolf, and Daniel O Clark. 1973. Development of standard data for stereoscopic microscope work. *AIIE Transactions* 5, 2 (1973), 113–118.

[29] Susumu Harada, Jacob O Wobbrock, Jonathan Malkin, Jeff A Bilmes, and James A Landay. 2009. Longitudinal study of people learning to use continuous voice-based cursor control. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 347–356.

[30] Morten Hertzum and Kasper Hornbæk. 2010. How age affects pointing with mouse and touchpad: A comparison of young, adult, and elderly users. *Intl. Journal of Human–Computer Interaction* 26, 7 (2010), 703–734.

[31] MH Hoyle. 1973. Transformations: An introduction and a bibliography. *International Statistical Review/Revue Internationale de Statistique* (1973), 203–223.

[32] Gary W Hrezo. 2001. Measuring pointing times of a non-visual haptic interface. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 45. SAGE Publications Sage CA: Los Angeles, CA, 654–656.

[33] ISO ISO. [n.d.]. 9241-9 Ergonomic requirements for office work with visual display terminals (VDTs)-Part 9: Requirements for non-keyboard input devices (FDIS-Final Draft International Standard), 2000. *International Organization for Standardization* ([n. d.]).

[34] Poika Isokoski and Roope Raisamo. 2002. Speed-accuracy measures in a population of six mice. In *Proc. APCHI2002: 5th Asia Pacific Conference on Computer Human Interaction*. Science Press, 765–777.

[35] Robert Kerr. 1973. Movement time in an underwater environment. *Journal of Motor Behavior* 5, 3 (1973), 175–178.

[36] Caroline J Ketcham, Rachael D Seidler, Arend WA Van Gemmert, and George E Stelmach. 2002. Age-related kinematic differences as influenced by task difficulty, target size, and movement amplitude. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 57, 1 (2002), P54–P64.

[37] Alexander Kulik, André Kunert, and Bernd Froehlich. 2020. On Motor Performance in Virtual 3D Object Manipulation. *IEEE Transactions on Visualization and Computer Graphics* 26, 5 (2020), 2041–2050.

[38] Gary Langolf and Walton M Hancock. 1975. Human performance times in microscope work. *AIIE Transactions* 7, 2 (1975), 110–117.

[39] Eckhard Limpert, Werner A Stahel, and Markus Abbt. 2001. Log-normal distributions across the sciences: keys and clues: on the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probability—normal or log-normal: that is the question. *BioScience* 51, 5 (2001), 341–352.

[40] Ramon C Littell, PR Henry, and Clarence B Ammerman. 1998. Statistical analysis of repeated measures data using SAS procedures. *Journal of animal science* 76, 4 (1998), 1216–1231.

[41] Julian Looser, Andy Cockburn, Joshua Savage, and N Christchurch. 2005. On the validity of using First-Person Shooters for Fitts' law studies. *People and Computers XIX* 2 (2005), 33–36.

[42] I Scott MacKenzie. 1989. A note on the information-theoretic basis for Fitts' law. *Journal of motor behavior* 21, 3 (1989), 323–330.

[43] I Scott MacKenzie. 1992. Fitts' law as a research and design tool in human-computer interaction. *Human-computer interaction* 7, 1 (1992), 91–139.

[44] I Scott MacKenzie and William Buxton. 1992. Extending Fitts' law to two-dimensional tasks. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 219–226.

[45] I Scott MacKenzie and Poika Isokoski. 2008. Fitts' throughput and the speed-accuracy tradeoff. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1633–1636.

[46] I Scott MacKenzie and Shaidah Jusoh. 2001. An evaluation of two input devices for remote pointing. In *IFIP International Conference on Engineering for Human-Computer Interaction*. Springer, 235–250.

[47] I Scott MacKenzie, Tatu Kauppinen, and Miika Silfverberg. 2001. Accuracy measures for evaluating computer pointing devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 9–16.

[48] I Scott MacKenzie and Aleks Oniszczak. 1998. A comparison of three selection techniques for touchpads. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 336–343.

[49] I Scott MacKenzie and Shawn X Zhang. 1999. The design and evaluation of a high-performance soft keyboard. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 25–31.

[50] Ji-Young Oh and Wolfgang Stuerzlinger. 2002. Laser pointers as collaborative pointing devices. In *Graphics Interface*, Vol. 2002. 141–149.

[51] Kyudong Park, Dohyeon Kim, and Sung H Han. 2020. Usability of the size, spacing, and operation method of virtual buttons with virtual hand on head-mounted displays. *International Journal of Industrial Ergonomics* 76 (2020), 102939.

[52] Réjean Plamondon and Adel M Alimi. 1997. Speed/accuracy trade-offs in target-directed movements. *Behavioral and brain sciences* 20, 2 (1997), 279–303.

[53] Paul C Price, Rajiv Jhangiani, I-Chant A Chiang, et al. 2015. *Research methods in psychology*. BCCampus.

[54] Raghavendra S Rao, Rami Seliktar, and Tariq Rahman. 2000. Evaluation of an isometric and a position joystick in a target acquisition task for individuals with cerebral palsy. *IEEE Transactions on Rehabilitation Engineering* 8, 1 (2000), 118–125.

[55] Maria Francesca Roig-Maimó, I Scott MacKenzie, Cristina Manresa-Yee, and Javier Varona. 2018. Head-tracking interfaces on mobile devices: Evaluation using Fitts' law and a new multi-directional corner task for small displays. *International Journal of Human-Computer Studies* 112 (2018), 1–15.

[56] Michael W Smith, Joseph Sharit, and Sara J Czaja. 1999. Aging, motor control, and the performance of computer mouse tasks. *Human factors* 41, 3 (1999), 389–396.

[57] BCM Smits-Engelsman, EAA Rameckers, and Jaak Duysens. 2007. Children with congenital spastic hemiplegia obey Fitts' Law in a visually guided tapping task. *Experimental Brain Research* 177, 4 (2007), 431–439.

[58] R William Soukoreff and I Scott MacKenzie. 2004. Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. *International journal of human-computer studies* 61, 6 (2004), 751–789.

[59] Christine Sutter and Martina Ziefle. 2004. Psychomotor efficiency in users of notebook input devices: Confirmation and restrictions of Fitts' law as an evaluative tool for user-friendly design. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 48. SAGE Publications Sage CA: Los Angeles, CA, 773–777.

[60] Zachary C Thumser, Andrew B Slifkin, Dylan T Beckler, and Paul D Marasco. 2018. Fitts' law in the control of isometric grip force with naturalistic targets. *Frontiers in psychology* 9 (2018), 560.

[61] Paolo Viviani and Tamar Flash. 1995. Minimum-jerk, two-thirds power law, and isochrony: converging approaches to movement planning. *Journal of Experimental Psychology: Human Perception and Performance* 21, 1 (1995), 32.

[62] Neff Walker, Jeff Millians, and Aileen Worden. 1996. Mouse accelerations and performance of older computer users. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 40. SAGE Publications Sage CA: Los Angeles, CA, 151–154.

[63] Neff Walker, David A Philbin, and Arthur D Fisk. 1997. Age-related differences in movement control: Adjusting submovement structure to optimize performance. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 52, 1 (1997), P40–P53.

[64] Joseph P Weir. 2005. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *The Journal of Strength & Conditioning Research* 19, 1 (2005), 231–240.

[65] Alan Traviss Welford. 1968. Fundamentals of skill. (1968).

[66] Jacob O Wobbrock, James Fogarty, Shih-Yen Liu, Shunichi Kimuro, and Susumu Harada. 2009. The angle mouse: target-agnostic dynamic gain adjustment based on angular deviation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1401–1410.

[67] Jacob O Wobbrock and Krzysztof Z Gajos. 2007. A comparison of area pointing and goal crossing for people with and without motor impairments. In *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility*. 3–10.

[68] Jacob O Wobbrock, Kristen Shinohara, and Alex Jansen. 2011. The effects of task dimensionality, endpoint deviation, throughput calculation, and experiment design on pointing measures and models. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1639–1648.

[69] Shumin Zhai. 2004. Characterizing computer input with Fitts' law parameters—the information and non-information aspects of pointing. *International Journal of Human-Computer Studies* 61, 6 (2004), 791–809.

A PARTICIPANTS WITH LIMITED FINE MOTOR FUNCTION

Table 6: Participants with limited fine motor function, their gender identification, age, diagnosis, and functional limitations. Under the *Gender* column, *M* = *Male*, *F* = *Female*, and *GQ* = *Genderqueer*.

Participant	Gender	Age	Diagnosis	Functional Limitations
P35	GQ	23	Congenital Amputee	Low strength, Stiffness, Numbness, Pain, Difficulty gripping, Difficulty holding, Difficulty forming hand postures
P36	F	51	Orsteoarthritis, De Quervain's	Poor coordination, Low strength, Stiffness, Pain, Difficulty gripping, Difficulty lifting, Difficulty holding, Difficulty forming hand postures, Difficulty controlling movement direction
P37	F	31	Muscular Dystrophy	Rapid fatigue, Poor coordination, Low strength, Slow movements, Pain, Difficulty gripping, Difficulty lifting, Difficulty holding, Difficulty forming hand postures, Difficulty controlling movement direction, Difficulty controlling movement distance
P38	M	69	Essential Tremor	Tremor, Difficulty holding still, Difficulty controlling movement direction, Difficulty controlling movement distance
P39	F	63	Lupus, Cerebral Palsy	Rapid fatigue, Poor coordination, Low strength, Slow movements, Tremor, Spasm, Stiffness, Numbness, Pain, Difficulty gripping, Difficulty lifting, Difficulty holding, Difficulty holding still, Difficulty forming hand postures, Difficulty controlling movement direction, Difficulty controlling movement distance
P40	F	69	Rheumatoid Arthritis	Rapid fatigue, Poor coordination, Low strength, Slow movements, Spasm, Stiffness, Pain, Difficulty gripping, Difficulty lifting, Difficulty holding, Difficulty forming hand postures
P41	F	67	Essential Tremor, Arthritis	Low strength, Tremor, Spasm, Stiffness, Pain, Difficulty gripping, Difficulty holding still, Difficulty controlling movement direction
P42	F	25	Spinal Cord Injury	Rapid fatigue, Low strength, Spasm, Stiffness, Pain, Difficulty lifting, Difficulty holding, Difficulty forming hand postures
P43	F	27	Spinal Cord Injury	Rapid fatigue, Poor coordination, Low strength, Stiffness, Difficulty gripping, Difficulty holding, Difficulty holding still, Difficulty forming hand postures
P44	F	37	Spinal Cord Injury	Rapid fatigue, Low strength, Slow movements, Tremor, Spasm, Stiffness, Pain, Difficulty gripping, Difficulty lifting, Difficulty holding, Difficulty forming hand postures, Difficulty controlling movement direction, Difficulty controlling movement distance
P45	F	31	Charcot Marie Tooth	Rapid fatigue, Poor coordination, Low strength, Slow movements, Tremor, Spasm, Stiffness, Numbness, Difficulty gripping, Difficulty lifting, Difficulty holding, Difficulty holding still, Difficulty forming hand postures

P46	F	27	Rheumatoid Arthritis	Rapid fatigue, Poor coordination, Low strength, Stiffness, Difficulty gripping, Difficulty lifting, Difficulty holding, Difficulty forming hand postures, Difficulty controlling movement direction, Difficulty controlling movement distance
P47	GQ	24	Symbrachydactyly	Rapid fatigue, Poor coordination, Pain, Difficulty gripping, Difficulty lifting, Difficulty holding, Difficulty forming hand postures
P48	F	70	Spinal Cord Injury	Poor coordination, Low strength, Slow movements, Tremor, Spasm, Stiffness, Numbness, Pain, Difficulty gripping, Difficulty lifting, Difficulty holding, Difficulty forming hand postures, Difficulty controlling movement direction, Difficulty controlling movement distance
P49	M	34	Spinal Cord Injury	Poor coordination, Low strength, Spasm, Stiffness, Pain, Difficulty gripping, Difficulty lifting, Difficulty holding, Difficulty forming hand postures
P50	F	54	Charcot Marie Tooth	Rapid fatigue, Poor coordination, Low strength, Slow movements, Stiffness, Difficulty gripping, Difficulty lifting, Difficulty holding, Difficulty holding still, Difficulty forming hand postures
P51	M	38	Spinal Cord Injury	Rapid fatigue, Poor coordination, Low strength, Slow movements, Spasm, Stiffness, Pain, Difficulty gripping, Difficulty lifting, Difficulty holding, Difficulty forming hand postures
P52	F	63	Bell's palsy	Poor coordination, Low strength, Slow movements, Difficulty holding
P53	M	32	Spinal Cord Injury	Low strength, Slow movements, Spasm, Stiffness, Difficulty gripping, Difficulty lifting, Difficulty holding, Difficulty holding still, Difficulty forming hand postures, Difficulty controlling movement direction, Difficulty controlling movement distance
P54	M	31	Spinal Cord Injury	Rapid fatigue, Poor coordination, Low strength, Slow movements, Spasm, Stiffness, Numbness, Pain, Difficulty gripping, Difficulty lifting, Difficulty holding, Difficulty holding still, Difficulty forming hand postures, Difficulty controlling movement direction, Difficulty controlling movement distance
P55	F	54	Spinal Cord Injury	Poor coordination, Low strength, Slow movements, Spasm, Stiffness, Numbness, Difficulty gripping, Difficulty lifting, Difficulty holding, Difficulty forming hand postures, Difficulty controlling movement direction, Difficulty controlling movement distance