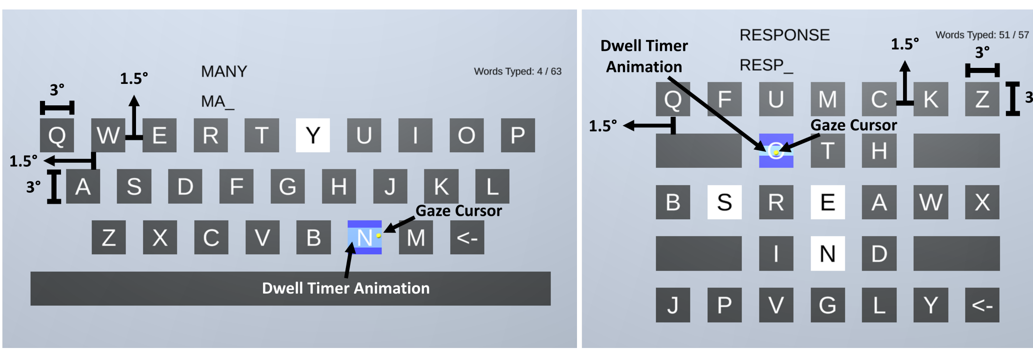
The Influence of Eye Gaze Interaction Technique Expertise and the Guided Evaluation Method on Text Entry Performance Evaluations

AUNNOY K MUTASIM*, School of Interactive Arts + Technology (SIAT), Simon Fraser University, Canada
ANIL UFUK BATMAZ, Computer Science & Software Engineering Department, Concordia University, Canada
MOAAZ HUDHUD MUGHRABI, Mechatronics Engineering Department, Kadir Has University, Turkey
WOLFGANG STUERZLINGER, School of Interactive Arts + Technology (SIAT), Simon Fraser University, Canada



(a) QWERTY (b) OPTI

Fig. 1. The two investigated keyboard layouts with the Guided Evaluation Method (GEM), which highlights all target word keys that have not been typed yet in white. In the GEM, participants first plan the typing of the target word before starting to type. In Figure 1b, with the target word "RESPONSE," and where the user has typed "RESP" so far, the letter 'R' is no longer highlighted as it is not present in the remaining portion of the target word. In contrast, 'E' is still highlighted since one more 'E' needs to be typed.

Any investigation of learning unfamiliar text entry systems is affected by the need to train participants on multiple new components simultaneously, such as novel interaction techniques and layouts. The Guided Evaluation Method (GEM) addresses this challenge by bypassing the need to learn layout-specific skills for text entry. However, a gap remains as the GEM's performance has not been assessed in situations where

*Corresponding Author: Aunnoy K Mutasim (amutasim@sfu.ca)

Authors' Contact Information: Aunnoy K Mutasim, amutasim@sfu.ca, School of Interactive Arts + Technology (SIAT), Simon Fraser University, Vancouver, BC, Canada; Anil Ufuk Batmaz, ufuk.batmaz@concordia.ca, Computer Science & Software Engineering Department, Concordia University, Montreal, QC, Canada; Moaaz Hudhud Mughrabi, moaaz.mazen@stu.khas.edu.tr, Mechatronics Engineering Department, Kadir Has University, Istanbul, Turkey; Wolfgang Stuerzlinger, w.s@sfu.ca, School of Interactive Arts + Technology (SIAT), Simon Fraser University, Vancouver, BC, Canada.

ACM 2573-0142/2025/5-ARTETRA17
https://doi.org/10.1145/3725842

users are unfamiliar with the interaction technique involved, here eye-gaze-based dwell. To address this, we trained participants on only the eye-gaze-based interaction technique over eight days with QWERTY and then evaluated their performance on the OPTI layout with the GEM. Results showed that the unfamiliar OPTI layout outperformed QWERTY, with QWERTY's speed aligning with previous findings, suggesting that interaction technique expertise significantly impacts performance outcomes. Importantly, we also identified that for scenarios where the familiarity with the involved interaction technique(s) is the same, the GEM analyzes the performance of keyboard layouts effectively and quickly identifies the best option.

CCS Concepts: • **Human-centered computing** → **Text input**; **Keyboards**; *Interaction techniques*; Virtual reality.

Additional Key Words and Phrases: Text Entry, Eye Tracking, Dwell, Soft Keyboards, OPTI, QWERTY, Virtual Reality

ACM Reference Format:
Aunnoy K Mutasim, Anil Ufuk Batmaz, Moaaz Hudhud Mughrabi, and Wolfgang Stuerzlinger. 2025. The Influence of Eye Gaze Interaction Technique Expertise and the Guided Evaluation Method on Text Entry Performance Evaluations. *Proc. ACM Hum.-Comput. Interact.* 9, 3, Article ETRA17 (May 2025), 19 pages.
https://doi.org/10.1145/3725842

1 Introduction

Evaluating a text entry system typically means estimating how fast trained/expert users can type using that system. The most reliable and valid approach to get to this estimate is to conduct longitudinal studies, i.e., training participants over multiple days [48]. Beyond estimating trained/expert user performance, researchers also focused on analyzing the learnability of the system, i.e., how quickly and easily users can learn a (novel) text entry system [6, 27, 37]. Yet, prior studies confounded the analysis of learning of such systems by training participants on multiple aspects simultaneously, e.g., novel/unfamiliar interaction techniques and layouts [5, 6, 37, 44].

It is well-documented in the literature that learning a novel/unfamiliar layout requires substantial commitment and training effort [37]. Thus, while more optimal layouts can offer typing speed advantages [61], convincing people to switch from QWERTY is challenging [48, 69]. For text entry research, however, and as demonstrated by Mutasim et al. [48], such layout training can be empirically simulated with the recently proposed Guided Evaluation Method (GEM) to quickly evaluate text entry systems.

The GEM involves **highlighting** all the keys in a target word but also *de-highlighting* the keys that have already been typed. Unlike *traditional training* that requires typing entire phrases, GEM presents **one word** at a time and encourages participants to **plan** the typing sequence for each word before they start typing (see Figure 1). As Mutasim et al. [48] showed, these combined features allow novice users to perform at a level similar to that of trained users.

Yet, while Mutasim et al. [48] established that the GEM can estimate trained user performance well, their findings are limited only to touchscreen-based input. Touch-based devices are quite common in our everyday lives and we are all well accustomed to touch-based interaction techniques. Thus it is reasonable to expect that the participants in their study (and all studies in the last 15 years) were well-trained in touch interaction. In contrast, interaction with eye gaze poses unique challenges, e.g., jitter or inaccurate calibration [46], and the general population has little to no experience/training with this interaction technique.

Thus, we investigate here *how much interaction technique expertise, particularly for eye gaze with dwell, influences the Guided Evaluation Method (GEM) in accurately estimating trained user text entry performance*, while still reducing the evaluation effort. In the process, we implicitly also investigate *whether the GEM approach is robust enough to be applied to unfamiliar interaction techniques.*

We conducted two user studies toward this goal, with the unfamiliar-to-our-participants eye gaze dwell-based interaction technique. First, we compared the OPTI and QWERTY layouts with complete novices using the GEM. Although we confirmed that OPTI outperformed QWERTY, the typing speeds were far from previous work [40, 59]. We speculated the difference to be due to participants' lack of experience with the interaction technique. To verify this, we separated the training associated with learning a new layout from that of the interaction technique. For this, Mutasim et al. [47] first trained the participants through an 8-day traditional longitudinal study *without* the GEM on a static dwell keyboard with the familiar QWERTY layout. We extended their study [47] by asking participants to perform another session on the QWERTY keyboard on the last/eighth day, but this time *with* the GEM. On the ninth day, these interaction-trained participants were exposed to two sessions with the unfamiliar OPTI layout, i.e., with and without the GEM. The outcomes from the single GEM session again demonstrated that OPTI significantly outperforms QWERTY. Also, QWERTY's performance matched previous literature [40, 59], supporting external validity.

The results for the interaction technique training, i.e., the first eight traditional training sessions with the QWERTY layout, had already been reported in separate work (see [47]). Specifically, they focused on analyzing the learnability of the different components of the (static) dwell technique. In contrast, we focus on how the interaction technique training affects the GEM. Thus, this work is primarily based on the (GEM) sessions after the training (see Figure 3).

Our contributions here include 1) showing that interaction technique expertise is a key factor in deriving accurate estimates of trained user text entry performance with the GEM and 2) demonstrating that the GEM is an efficient tool to identify which text entry layout is better when the confound of interaction technique expertise is removed.

For clarity, we also define several key terms here. Eye gaze, including head movements to support the visual system [23], is referred to as gaze. Participants trained via a typical longitudinal study, e.g., [40, 45], are referred to as "*trained users*". For any layout, a key difference between novices and trained users (or true experts [12]) is that trained/expert users can find keys faster, incurring much less visual search time [38, 65]. We refer to such users as "*layout-trained users*". Another difference focuses on whether users have sufficient training with the specific interaction technique (termed "*interaction-trained users*"), e.g., gaze and dwell. Finally, we call users who are both layout- and interaction-trained "*dual-trained users*."

2 Literature Review

2.1 Evaluation Methods for Text Entry Systems

Text entry studies typically train novices between 0.5-7.5 hours per participant per layout to evaluate the performance of a text entry system/layout. Researchers conducted this training either on the same day, e.g., [10, 13, 32, 33, 44, 77] or spread over several days within a longitudinal study, e.g., [6, 7, 19, 27, 31, 35, 37, 40, 45, 71]. Such longitudinal studies are more externally valid as they yield better approximations of trained user behaviors.

Yet, longitudinal studies are logistically expensive [48]. For example, in previous work [37], participants required about 4 hours of training on an unfamiliar OPTI layout to match their corresponding QWERTY typing speed. After 7 training hours, OPTI exhibited a speed of 44.3 words per minute (WPM) achieving better performance than QWERTY (40 WPM). Theoretical projections suggest that OPTI (60.7 WPM) could be ~15 WPM faster than QWERTY (44.8 WPM). Yet, demonstrating this would require at least 17 hours of training.

To reduce the logistical challenges of a longitudinal study, researchers explored an alternative approach that only involves participants repeatedly typing the same word(s) or phrase(s) [4, 5, 16, 72, 77]. Yet, Mutasim et al. [48, 51] recently showed that both repeated word/phrase typing approaches have limitations and do not estimate trained user performance well. Mathematical models (e.g., [27, 61]) and employing a system that simulates a perfect recognizer [28] have also been utilized in the past to estimate text entry performance.

To speed up layout learning, a significant challenge for novice users [37, 38, 65], assistance through visual clues has also been explored [17, 39]. The GEM [48] also builds on the idea of providing such visual clues. For touch-based systems, Mutasim et al. [48] showed that in just a few minutes, the GEM demonstrated quite similar performance to that observed in a longitudinal study. Thus, the GEM could be a great tool for text entry research. However, it is still an open question whether the GEM can accurately estimate text entry performance for non-touch-based interaction, e.g., with unfamiliar, more challenging, and/or noisy interaction techniques like eye tracking.

2.2 Gaze and Dwell Interaction Technique Training

Compared to button clicks, dwell selection is slower than other alternatives because dwell demands users to (unnaturally) wait on the target for the entirety of the *dwell time* [34, 66]. Thus, to improve the performance of dwell, researchers experimented with reducing the dwell time by adjusting it based on various criteria [20, 25, 45, 52, 53, 56, 57, 74]. One approach explicitly lets the user decide what dwell threshold they are comfortable with while simultaneously trying and training to reduce their dwell time as much as possible.

Majaranta et al. [40] conducted one such (longitudinal) study, where participants adjusted the dwell time to their preference. Eleven participants typed 15 minutes per day on a QWERTY keyboard for ten days. This training significantly improved the typing speed and dwell time from the first (6.9 WPM and 876 ms) to the last day (19.9 WPM and 282 ms). Räihä and Ovaska [59] conducted a longer study with 19 sessions, each 15 minutes, and allowed participants to adjust the dwell time only during the first 10 sessions. Results from both studies aligned, with participants reaching an average typing speed of 20 WPM and dwell times set between 240–340 ms. However, the reduced dwell times increased the rate of inadvertent selection of non-targets, i.e., increased the Midas Touch problem, [26], raising the keystrokes per character (KSPC) metric. KSPC rose from 1.09 to 1.18 in Majaranta et al.'s [40] study, while Räihä and Ovaska [59] reported an even higher KSPC of about 1.25 in the final session.

The above studies investigated adjusting the dwell time in depth. Yet, no studies analyzed how much performance can improve by training users on a *static* dwell-based gaze interaction system. While Mutasim et al. [47] recently investigated this in more detail, we primarily focus here on how much training on the interaction technique influences the ability of the GEM to derive accurate trained user text entry performance estimates.

2.3 Gaze-Only Text Entry

Here, we briefly review text entry systems that involve only one's (eye) gaze, i.e., which are *not* multimodal. One category of gaze-only text entry is dwell-free typing. Examples include Quikwriting (7.8 WPM) [2], pEYEwrite (13.47 WPM) [73], Context Switching QWERTY (13.42 WPM) [44], the popular Dasher (14.2 WPM) [62], and others [54, 63]. Recent research on gaze-based text entry centers on gaze-based gesture/swipe dwell-free typing. While such systems are predicted to reach 46 WPM [28], current gaze-only systems like Filteryedping (15.95 WPM) [55], EyeSwipe (11.7 WPM) [30], Swipe&Switch (13.7 WPM) [31], and GlanceWriter (10.89 WPM) [8] seem to be far from reaching that goal.

As for dwell-based approaches, the dwell time is typically set between 450–1000 ms for novices, resulting in typing speeds between 5-10 WPM [33]. Dynamically adjusting the dwell time proportional to the likelihood of a given key resulted in a speed of 13.7 WPM [45]. Using the information of how long it takes a user to exit a key after dwell selection, Špakov and Miniotas [74] demonstrated 12.1 WPM. Diaz-Tula and Morimoto [9] proposed a dwell-based text entry system called AugKey that utilizes word predictions. The authors showed that AugKey was faster (16.7 WPM) while requiring substantially (~0.8) fewer KSPC than conventional dwell-based systems.

The focus of our work is not to come up with a better text entry system. Instead, we investigate whether the GEM approach is robust enough to be used in non-touchscreen-based text entry studies, i.e., irrespective of the specific interaction/pointing/selection technique. To do this, we chose dwell-based selection alongside eye gaze pointing. Still, we note that, in principle, dwell-based input is similar to single-finger- or stylus-based tapping input [37, 48], where the target word/phrase is typed one letter at a time, i.e., there are no simultaneous movements as in two-hand touch typing. While investigating the influence of interaction technique expertise on GEM, and since OPTI was proven to be superior to QWERTY for single-tap-based input [37, 48], we also investigate whether OPTI's superiority holds for gaze-based input as well.

3 Guided Evaluation Method (GEM)

The GEM [48] is designed to be more resource-efficient than traditional evaluation methods, particularly longitudinal studies. To alleviate the visual search effort of novice users when navigating an unfamiliar keyboard [65], the GEM builds upon the concept of providing visual cues [39] and provides a form of "*guidance*" to the user by **highlighting** all the target **word** letters. Highlighting keys to be typed allows de-highlighting all keys that have been typed – as long as they do not exist elsewhere in the target word. For example, with a target word of "RESPONSE" and when "RESP" has already been typed, the letter 'R' is no longer highlighted as it does not appear elsewhere in the remaining portion of the target word. In contrast, 'E' is still highlighted since one more 'E' needs to be typed (see Figure 1). Also, if participants make a mistake, keys are not de-highlighted from the mistake onward. This approach aids participants in recognizing mistakes early in the process, eliminating the need to constantly verify whether the typed and target text match. This (de-)highlighting substantially reduces the search space while having to type only a single word at a time also demands less memorization than phrase-based approaches.

Another important feature of the GEM is its requirement for participants to first **plan** the typing pattern for each word before actually typing it. This step essentially simulates practice rounds without actual typing. All these features combined substantially reduce layout learning effects and thus, eliminate/reduce the need for a longitudinal study.

The GEM also requires participants to type an extra "space" on the virtual keyboard to mark the end of the target word. To generate the target words for our studies, we randomly chose complete phrases from MacKenzie and Soukoreff's [36] set comprising 500 phrases with minimum/maximum/average phrase lengths of 16/43/28.6 characters, respectively. However, as demanded by the GEM, the words in each phrase were presented in a one-word-at-a-time manner.

4 User Study 1

As shown by Mutasim et al. [48], the GEM is promisingly accurate in estimating dual-trained user performance for touch-based text entry. However, their findings could be skewed towards touch-based interfaces as the general population is already very familiar with touch interaction. Yet, the same is not true for more challenging and unfamiliar interaction techniques like gaze. Thus, we first investigate the validity of the GEM approach for the gaze-based dwell interaction technique. To do this, we compared the performance of OPTI and QWERTY keyboard layouts with the GEM in Virtual Reality (VR). In the process, we also investigate whether OPTI's superiority for stylus/touch-based input, as found in previous work [37, 48, 61], also holds for gaze-based input.

We decided to conduct the study in a VR system, as the eye trackers in VR headsets are easy to use and typically perform fairly well. Moreover, Rajanna and Hansen [60] showed that gaze-based text entry is a viable input method for VR, achieving results matching several non-VR-based previous studies [33].

4.1 Participants

12 novice participants (8 male, 4 female), aged 22.8 ± 4.71 years, took part in this study. They all had either no visual impairments or corrected-to-normal vision. 10 of the 12 participants had nine or more years of experience typing on a QWERTY keyboard, one 5-7, and the other had 7-9 years of experience. None of the participants had prior experience with the OPTI keyboard or with eye tracking. The participants were paid the equivalent of US $10.

4.2 Keyboard Designs

Following previous work [21, 44, 58, 64], we set the width and height of each key to 3°. Also, based on pilot studies, we decided to add a gap of 1.5° between keys. This was done to avoid inadvertently triggering neighboring keys when the eye-tracker exhibits low(er) tracking accuracy [60]. Based on our pilots, and again following a similar procedure as previous work [60], we also set the dwell time to 450 ms. Further, the dwell timer for each word was only started when participants hit the spacebar of a physical keyboard placed in front of them. We did this to avoid unintended selections of keys during the typing pattern planning phase of the task [43]. Whenever the user's gaze cursor came in contact with a key, that key was highlighted in blue. After the planning, i.e., when the physical spacebar was hit, and the cursor hovered over a key, an animation showing the progress of the dwell timer was also started [41]. When 450 ms dwell time had passed, the key was selected and users were given confirmatory auditory and visual feedback, which highlighted the key in green for 100 ms [70]. Continuing to dwell over the same key resulted in repeated selections every 450 ms.

Like Mutasim et al. [48], the backspace key was added at the bottom-right corner of the OPTI layout to facilitate error correction. For QWERTY, the backspace key was placed next to the 'M' key, following smartphone keyboards. Participants' task progress was shown at the top-right of the keyboard. Both keyboards were world-fixed in VR, and their center was placed at the eye level of the participants. Based on exploratory pilots, we put the keyboard two meters away from the participant in the virtual environment. Figure 1 presents the design of both keyboards.

4.3 Apparatus

We ran the study on a computer with an i7-11700F processor, 32 GB RAM, and an RTX 3070 graphics card, with Unity 2019.2.14f1. We used an HTC VIVE Pro Eye VR headset with a resolution of 2880×1600 pixels, 110° (diagonal) FOV, and 90 Hz refresh rate. The embedded Tobii eye-tracker in the VR headset transmits data at 120 Hz. Before each condition, the eye-tracker was calibrated using Tobii's 5-point calibration method.

4.4 Procedure

For comparability, we followed a procedure similar to Mutasim et al. [48]. First, participants were asked to fill out a consent form and a demographic questionnaire, which included questions regarding their experience using OPTI and QWERTY layouts, age, and gender. Then, in a within-subjects experimental design with one independent variable – the two keyboards – they typed eleven phrases using both the OPTI and QWERTY layouts, presented in counterbalanced order. For the planning phase of the GEM, we encouraged participants to take as long as needed. We also instructed them to memorize each target word before planning/typing. Further, they were asked to correct any mistakes immediately when they noticed them [44]. Also, to speed up typing, participants were advised not to look back at the (partially) typed word; instead, they were instructed to remember what they had typed and what to type next. Participants were not given practice trials. Instead, we regarded only the first phrase as practice and excluded it from the analysis [77]. At the end, participants completed another short questionnaire where they shared their preferences and provided feedback on aspects like ease of interaction, mental and physical fatigue, frustration, and perceived precision and speed for the two keyboard layouts using 7-point Likert scales. Each keyboard condition lasted approximately 10-12 minutes, and the entire experiment, including the two questionnaires, took about 40 minutes.

4.5 Performance Metrics

We chose the following metrics to evaluate the performance: **Words per minute (WPM)**, which is the average number of words typed every minute, where a word is defined as the sequence of any five characters [1], i.e., includes spaces and special characters. Thus, "I ATE" is considered one word, and "THAT" is 0.8. **Keystrokes per character (KSPC)** represents the average number of keys selected to (correctly) type a single character. More precisely, KSPC is the ratio of the total number of selected keys (including backspaces) over the length of the typed text [67]. For example, if three letters are typed ("THR"), then one is deleted ("TH"), and another letter is added ("THE"), the KSPC is 5/3 = 1.67. **Minimum String Distance Error Rate (MSD ER)**, where MSD is the minimum amount of changes required – insertions, deletions, and substitutions – to convert one string to another, i.e., from the incorrect to the correct text. Here, we use the MSD ER metric formulation by Soukoreff and MacKenzie [68]. **Planning Time** is how long it takes the participant to plan the typing pattern of a word. Simply, it is the time taken from the presentation of a word to the first key selection for entering said word [48].

4.6 Results

We analyzed the data using dependent t-tests with $\alpha = 0.05$ in SPSS 29. We considered data normally distributed when Skewness and Kurtosis values were within ±1.5 [18, 42]. For dependent variables that did not have a normal/log-normal distribution, data was transformed using Aligned Rank Transform (ART) [76]. For brevity, we detail only statistically significant results.

4.6.1 WPM, KSPC, MSD ER, and Planning Time. The average typing speed for OPTI was 12.7 ± 2.30 WPM (mean ± SD), which was significantly faster ($t_{11} = 3.37, p < 0.01, d = 0.97$) than the 11.7 ± 2.22 WPM with QWERTY (see Figure 2a). As per Figure 2b and 2c, no significant differences were observed for KSPC (OPTI: 1.023 ± 0.060, QWERTY: 1.036 ± 0.110) and MSD ER (OPTI: 0.41 ± 1.18%, QWERTY: 0.34 ± 1.23%). The average planning time with OPTI (22.9 ± 9.31 seconds) was significantly longer ($t_{11} = 2.07, p < 0.05, d = 0.60$) than with QWERTY (20.0 ± 8.95 seconds; see Figure 2d).

4.6.2 Subjective Measures. 7 of 12 participants preferred QWERTY, 3 chose OPTI, while 2 participants mentioned both were the same. Participants who preferred QWERTY mentioned "*I am used to it*", "*easier*", and "*Easier to plan*". Example reasons for the choice of OPTI were "*the letters [were] closer and easier to find with my peripheral vision*", "*Because of the positioning of the letters, it was easier to reach most of the letters*", and "*The buttons were close together and eye-tracking accuracy was [perceived to be] better. Thus, [it] felt faster and easier.*" Participants who did not have a preference stated that "*both showed the same performance*" and "*the difference between my performance in each keyboard was very minor.*" When asked about the ease of interaction, frustration, mental and physical fatigue, and perceived speed and precision on a 7-point Likert scale (with 7 signifying

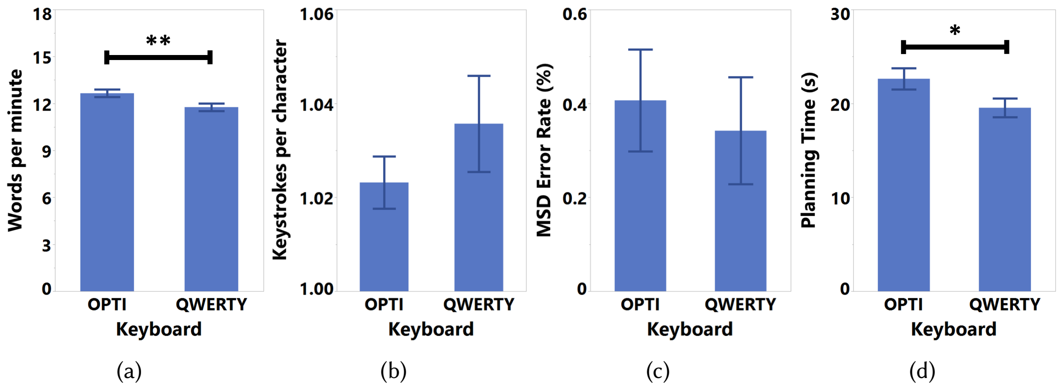


Fig. 2. Study 1 results for (a) words per minute (WPM), (b) keystrokes per character (KSPC), (c) Minimum String Distance Error Rate (MSD ER), and (d) Planning Time for OPTI and QWERTY. Significance levels are shown as *** for $p < 0.001$, ** for $p < 0.01$, and * for $p < 0.05$. The error bars show the standard error of means.

the highest preference), participants consistently rated OPTI lower than QWERTY for all these subjective measures. However, none of these differences were significant.

4.7 Study 1 Discussion

In Study 1, we investigated whether the GEM can estimate trained user performance for text entry methods that use an unfamiliar interaction technique like gaze-based dwell. Thus, we compared OPTI and QWERTY with the GEM here and found **OPTI significantly outperforms QWERTY** (see Figures 2a-2c).

At about 12 WPM, our participants typed slightly faster with the QWERTY layout than the previously reported speed of 5-10 WPM on dwell-based QWERTY keyboards [33]. We attribute this difference to the GEM tasking participants to type one word at a time, which requires less effort and memory, compared to the traditional approach of typing whole phrases. Comparing our results with Majaranta et al.'s [40] work is not straightforward, as they did not report the exact values of dwell time and typing speed found in each session. Thus, we reverse-engineered the data from their figures and observed that participants set their average dwell time slightly lower than 450 ms in the third session, which yielded about 13.5 WPM. Our results for the QWERTY keyboard with a 450 ms dwell time of 11.7 WPM are thus potentially somewhat different compared to the speed reported by Majaranta et al. [40].

We further calculated an estimated typing speed based on the results of Majaranta et al. [40], i.e., a dwell time of 282 ms and an additional 1.18 KSPC. We did this by subtracting the dwell time difference from the total selection time of each key press and multiplying it by 1.18. With that computation, we predict an average typing speed of 15.8 ± 4.57 WPM and 14.2 ± 4.25 WPM for OPTI and QWERTY, respectively. This estimate of a trained user typing speed on QWERTY (14.2 WPM) still differs substantially from the 20 WPM reported previously [40, 59]. We believe the reason for this difference is that, based on our definitions, participants of previous studies [40, 59] were dual-trained users. In other words, and compared to interaction-trained users, our novice participants struggled, e.g., to adapt to the continuous gaze cursor jitter [24] and/or eye-tracker accuracy limitations [49, 50].

As for our results for the subjective measures, and although participants typed faster with the OPTI layout, a larger number of participants preferred QWERTY over OPTI (7 out of 12). As mentioned by the participants, this is because they are used to QWERTY, and thus, it was easier for them to plan the typing pattern. This is also evident in the significantly shorter planning time results for QWERTY (Figure 2d) and matches the findings of Mutasim et al. [48] that increased planning time corresponds directly to increased effort. These results also suggest that the planning time should get shorter as layout familiarity improves over time. However, future work is required to verify this.

5 User Study 2

In Study 1, we speculated that participants' lack of interaction technique expertise is the main cause for the performance difference between our results and those of previous studies [40, 59]. Thus, to better understand the influence of gaze-based dwell selection in isolation, participants were first trained to use dwell interaction on a layout very familiar to them, i.e., *only* QWERTY, and without the Guided Evaluation Method (GEM) in work we build directly upon [47]. We then extended Mutasim et al.'s study [47] by exposing these interaction-trained users to the GEM, one session each with OPTI and QWERTY (see Figure 3).

5.1 Participants

A different set of 9 novice participants (3 females), aged 28.2 ± 4.99 years, took part in this longitudinal study. They all had no visual impairments or corrected-to-normal vision, over 9 years of average experience typing on a QWERTY keyboard, and no prior experience with eye tracking and OPTI. In total, they were paid US $60 for the study.

5.2 Keyboard Designs

We used the same QWERTY keyboard design as Study 1 and kept the dwell time static at 450 ms throughout the experiment. This choice enabled us to compare with the results of Study 1 directly. Still, we made two changes to the screen and functionality for these traditional training sessions, i.e., sessions without the GEM. First, we replaced the task progress in the top-right corner with their achieved typing speed in the last typed phrase. The intention was to gamify the task and encourage participants to aim to try to match or potentially even beat their "score" every time (i.e., every phrase/session). The other change was that we did not enable the use of the physical spacebar for the non-GEM training sessions, as a traditional evaluation method does not require an explicit planning phase.

5.3 Apparatus

A Unity 2021.3.24f1 and Meta Quest Pro VR headset-capable computer with an i7-12700H processor, 16 GB RAM, and an RTX 4060 graphics card, was used. This VR headset has a resolution of 1800×1920 pixels per eye, 95.57° (diagonal) FOV, and a 90 Hz refresh rate. Meta's built-in method was used to calibrate the 30 Hz eye tracker before every session. Moreover, we used the headset in wireless mode for more flexible data collection options (also reducing the potential for cable failures), ensuring a consistent hardware setup throughout the 9-day study.

5.4 Procedure

In Mutasim et al.'s work [47], participants initially signed a consent form and filled out a demographic questionnaire, where they answered questions regarding their age, gender, and experience using the QWERTY layout. Then participants performed their first 15-minute session of traditional text entry training with complete phrases chosen randomly from the set of MacKenzie and Soukoreff [36]. This 15-minute traditional training continued for seven more sessions. For this, participants

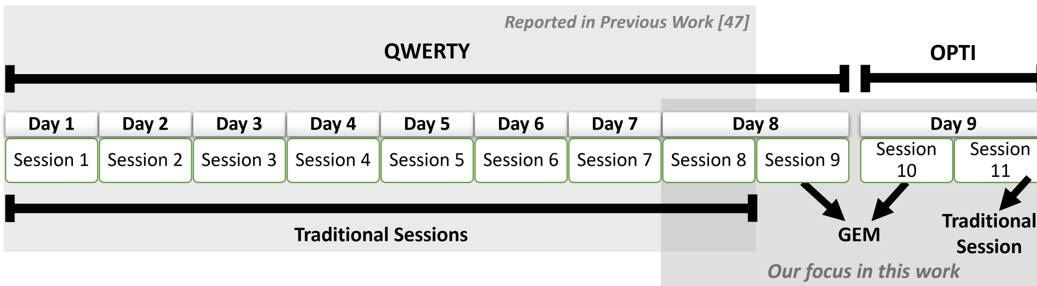


Fig. 3. Longitudinal study experiment procedure design used in Study 2.

came to the lab on eight separate days, within a window of 12 days and a gap of up to two days between sessions, to account for weekends.

We extended Mutasim et al.'s study [47], by asking the same participants to perform a GEM session on the QWERTY layout following a 5-minute break after completing the last traditional session on the eighth day (see Figure 3).

Subsequently, all participants came to the lab on the following day, i.e., the ninth. This time, they performed a GEM session on the unfamiliar OPTI layout followed by a single 15-minute traditional session with a 5-minute break in between. Since the GEM is intended to be used with novice users, we intentionally did not counterbalance the order of the two evaluation methods. We did this to ensure no layout learning effects were carried forward when using the GEM. A graphical representation of the experimental procedure design is presented in Figure 3.

For the traditional sessions, if the participants were in the middle of typing a phrase when the 15-minute timer expired, the software ended the session only upon completion of that phrase. Participants ended every phrase by typing an extra "space" after the end of the last word. Everything else remained the same as in Study 1: typing eleven phrases but discarding the first phrase from the analysis on the GEM sessions, correcting mistakes immediately and only within the current word, and being discouraged from visually matching the typed and target text.

5.5 Experimental Design and Performance Metrics

To analyze the effects of training on GEM, we *only compare here the last traditional and GEM sessions of both OPTI and QWERTY*. More specifically, we analyze the effect of the two **Evaluation Methods**, i.e., *traditional* and *GEM*, on the two **Keyboards**, *OPTI* and *QWERTY*, in a 2×2 within-subjects design. As dependent variables, we again used words per minute (WPM), keystrokes per character (KSPC), Minimum String Distance Error Rate (MSD ER), and Planning Time.

5.6 Results

The data was analyzed using repeated measures (RM) ANOVA with $\alpha = 0.05$ in SPSS 29. Tests for normality and transformations of the data following the failure of normality were done using the same procedure as in Study 1. Huynh-Feldt correction was applied upon violation of Mauchly's sphericity test, where $\epsilon < 0.75$. Post-hoc analyses were conducted using the Bonferroni method. Only statistically significant results are reported here.

5.6.1 WPM. The average typing speed for QWERTY in the traditional training sessions started at 11.6 ± 1.77 WPM in session 1 and reached 14.8 ± 1.63 WPM in session 8. A more detailed analysis of the performance in the 8 traditional sessions is presented in a separate work (see [47]). In session 9, and after this QWERTY training, participants typed at 15.0 ± 1.73 WPM with the GEM on QWERTY.

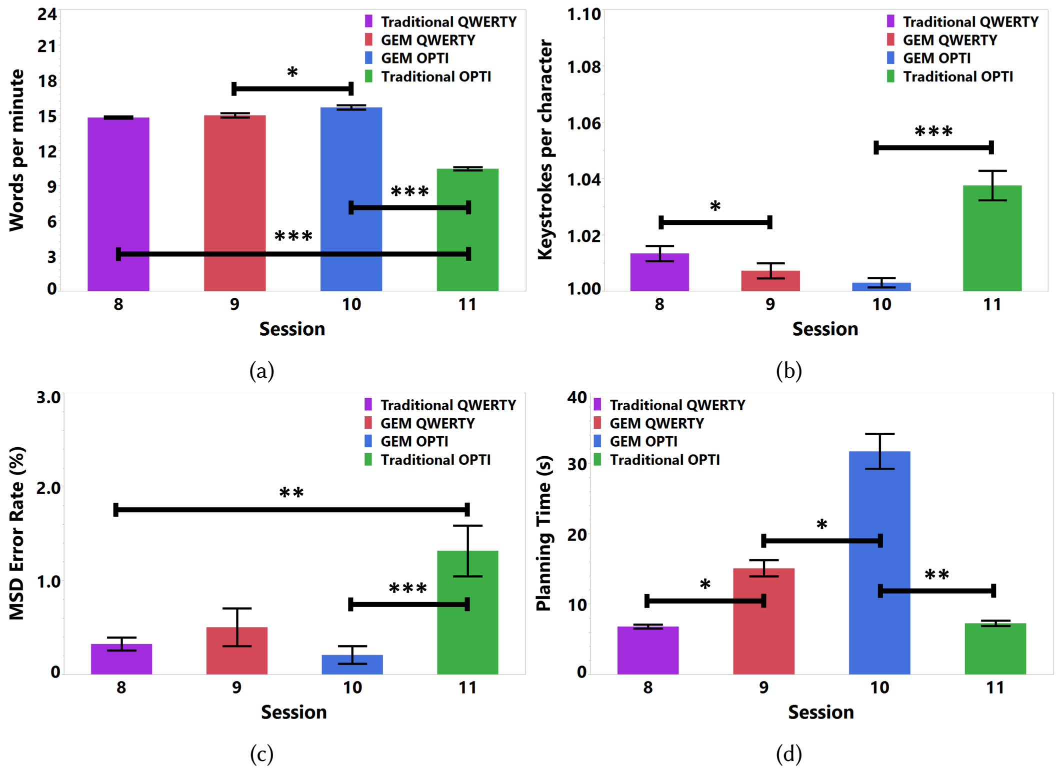


Fig. 4. Study 2 results for (a) words per minute (WPM), (b) keystrokes per character (KSPC), (c) Minimum String Distance Error Rate (MSD ER), and (d) Planning Time for the GEM and Traditional Evaluation Methods and OPTI and QWERTY Keyboards. Significance levels are shown as *** for $p < 0.001$, ** for $p < 0.01$, and * for $p < 0.05$. The error bars show the standard error of means.

In session 10, participants could type 15.7 ± 1.54 WPM with the OPTI layout with the GEM, and in session 11, 10.2 ± 1.51 WPM with the traditional approach (without the GEM).

The two-way RM ANOVA results for the analysis across the two *Evaluation Methods* and *Keyboards* are presented in Table 1 and Figure 4a. For traditional training, participants typed significantly faster with QWERTY (session 8) than with OPTI (session 11). Yet, with the GEM, i.e., sessions 9 and 10, **OPTI significantly outperformed QWERTY**. Moreover, GEM-OPTI (session 10) exhibited significantly faster speeds than traditional-OPTI (session 11).

5.6.2 KSPC and MSD ER. The results for KSPC and MSD ER are presented in Figures 4b and 4c. For QWERTY, participants' average KSPCs in the first and last sessions with traditional training, i.e., sessions 1 and 8, were 1.043 ± 0.082 and 1.013 ± 0.045, respectively [47]. With the GEM, the average KSPC was 1.007 ± 0.026 in the 9th session. For the MSD ER and traditional training, the averages for the first and last sessions were 1.32 ± 3.30% and 0.32 ± 1.14%, respectively [47]. For the GEM, the average MSD ER was 0.50 ± 1.91% in session 9, i.e., after the training.

For OPTI, the average KSPC was 1.003 ± 0.016 with the GEM (session 10) and 1.038 ± 0.073 without it (session 11). 0.21 ± 0.88% and 1.32 ± 3.81% were the average MSD ER for the GEM and traditional approaches, respectively.

Table 1. RM ANOVA results for the two Evaluation Methods and Keyboards, with significant results in bold.

	Evaluation Method	Keyboard	Evaluation Method × Keyboard
WPM	$\boldsymbol{F_{1,8} = 93.1, p < 0.001, \eta^2 = 0.921}$	$\boldsymbol{F_{1,8} = 23.8, p < 0.001, \eta^2 = 0.748}$	$\boldsymbol{F_{1,8} = 78.4, p < 0.001, \eta^2 = 0.907}$
KSPC	$\boldsymbol{F_{1,8} = 26.0, p < 0.001, \eta^2 = 0.891}$	$F_{1,8} = 0.40$, *n.s.*, $\eta^2 = 0.048$	$\boldsymbol{F_{1,8} = 7.43, p < 0.05, \eta^2 = 0.481}$
MSD ER	$\boldsymbol{F_{1,8} = 6.83, p < 0.05, \eta^2 = 0.460}$	$F_{1,8} = 4.08$, *n.s.*, $\eta^2 = 0.338$	$\boldsymbol{F_{1,8} = 14.0, p < 0.01, \eta^2 = 0.636}$
Planning Time	$\boldsymbol{F_{1,8} = 15.1, p < 0.01, \eta^2 = 0.653}$	$\boldsymbol{F_{1,8} = 8.07, p < 0.05, \eta^2 = 0.502}$	$\boldsymbol{F_{1,8} = 8.35, p < 0.05, \eta^2 = 0.511}$

We did not observe significant differences across the layouts for KSPC. However, significant differences were observed within layouts, with the traditional evaluation approach requiring more KSPC than GEM for both OPTI and QWERTY. As for MSD ER, traditional-OPTI exhibited significantly higher MSD ER than GEM-OPTI and traditional-QWERTY.

5.6.3 Planning Time. Both GEM sessions required significantly longer planning time compared to their traditional counterparts. Like Study 1, participants again needed significantly more time to plan with GEM-OPTI than GEM-QWERTY (see Figure 4d).

5.7 Study 2 Discussion

In this study, we investigated the influence of the gaze-based dwell interaction technique expertise on the GEM with Mutasim et al.'s [47] participants who were first trained only on the interaction technique itself and using the familiar QWERTY layout without the GEM, i.e., following the traditional training approach of typing different phrases. Subsequently, we extended Mutasim et al.'s study [47] by asking the same participants to perform single GEM sessions with both QWERTY and the unfamiliar OPTI layout. At the end, participants also performed one more traditional session with the OPTI layout.

Through the GEM, we again found OPTI (15.7 WPM) to significantly outperform QWERTY (15.0 WPM) after participants had trained with the gaze-based dwell interaction technique. This performance, however, came at the cost of higher planning time for OPTI (see Figure 4d) matching the findings of Mutasim et al. [48] and Study 1.

To be able to compare our results with the findings of previous work [40, 59] and similar to Study 1, we again calculated an estimated typing speed based on a dwell time of 282 ms and an additional 1.18 KSPC. This computation predicted an average typing speed of 21.3 ± 3.95 WPM for OPTI with the GEM, an average last-QWERTY-day typing speed of 19.5 ± 3.17 WPM for the traditional approach, and 19.9 WPM ± 3.56 WPM for QWERTY with the GEM. The predicted results for QWERTY are very close to the 20 WPM reported by previous work [40, 59], suggesting that our findings match real-world outcomes. These predictions suggest that the GEM does not directly help users learn to use an interaction technique (nor that the GEM accounts for the corresponding effects), thus presenting a very likely explanation for the difference between our Study 1 results and those of previous work [40, 59].

However, contrary to the findings of Majaranta et al. [40], our observation of 1.007 KSPC did not come close to the 1.18 KSPC reported previously. We believe the reason for this is that previous work [40, 59] allowed their participants to adjust, or more specifically, reduce the dwell time, which in turn increased the unintentional activation of non-targets, a.k.a. the Midas Touch problem [26].

In contrast, we stayed with a static dwell time of 450 ms for eight days. Thus, the 450 ms dwell time, which we deemed suitable for novice users, was eventually mastered by our interaction-trained users as time progressed, leading to a very small KSPC of 1.007.

Our results also revealed that with the GEM, participants could indeed exhibit significantly better performance for the unfamiliar OPTI layout in terms of increased typing speed, lower KSPC, and fewer errors (see Figures 4a-4c), compared to the traditional approach with the same layout. Note that GEM outperformed the traditional approach even though we did not counterbalance the two conditions, and our sequencing (intentionally) introduced a learning bias against the GEM. Still, our results show that with the GEM, our interaction-trained but layout-*novice* users performed to some extent like layout- and therefore dual-trained users. However, the same was not true for QWERTY, i.e., no significant differences were found between the GEM and traditional sessions on the eighth day, except for KSPC (see Figure 4b). In our pilots, we also found no significant differences between the GEM and the traditional approach for QWERTY with interaction-*novice* users. Moreover, QWERTY's 11.7 ± 2.22 WPM with the GEM in Study 1 is very similar to the speed of 11.6 ± 1.77 WPM in the first traditional session of Study 2. This suggests that the GEM approach does not affect participants' (years of) experience with a familiar layout much [48].

GEM-OPTI (15.7 WPM) also demonstrated a 54% increase in typing speed compared to GEM-traditional (10.2 WPM). This finding is in some contrast to Grüneis et al.'s [17] results for touchscreen-based text entry, which reported only a 36% increase in typing speed with visual clues compared to no clues. We believe the reason for this difference, other than the difference in layout and visual clues, is that the GEM requires participants to plan to type the word before entering it, which is an implicit behavior typical for trained/expert users [48].

6 General Discussion and Future Work

In this paper, we investigated the influence of gaze-based dwell interaction expertise and the Guided Evaluation Method (GEM) on text entry performance evaluations. Results from our user studies revealed that the GEM can mitigate the effects of layout learning by simulating the typing behavior of a layout-trained user. However, the GEM is strongly influenced by interaction technique expertise when aiming to derive accurate estimates of dual-trained user performance. Yet, according to our findings from Studies 1 and 2, **when participants' familiarity with the interaction technique is similar across conditions, the GEM is powerful enough to identify in a single 11-phrase session whether one text entry system/layout is better than another**. Thus, we believe that the similarities of our findings across both studies strengthen the arguments for the validity and generality of the GEM. This convinces us that the GEM has strong potential to reduce the need for longitudinal studies in (gaze-based) text entry research. Moreover, due to the existence of the GEM, researchers can now focus directly on training users with an interaction technique over multiple sessions (e.g., touch, dwell, saccades, hand-tracking, or pinch), without simultaneously focusing on layout learning.

Still, as per Figures 2d and 4d, GEM's good performance as an evaluation method comes at the cost of increased planning time and, therefore, higher cognitive effort. Yet, unlike previous work (e.g., [11, 19, 44, 72]), we were able to demonstrate in both studies that **an unfamiliar keyboard layout, i.e., OPTI, was able to outperform QWERTY with absolute layout-novices in a single 11-phrase session with the GEM**. This outweighs the cost of longer planning since the other viable alternative is the traditional approach, which inevitably demands a longitudinal study, e.g., [37]. However, it is yet to be investigated exactly which component(s) of the GEM – typing a word, planning before typing, and de-highlighting keys – contribute to making the GEM so effective that it enabled us to easily demonstrate in a single session that OPTI can "beat" QWERTY. Nonetheless, we reemphasize that *for GEM to correctly identify how an unfamiliar layout performs* *relative to a known one (typically QWERTY), the confounding factor of the interaction technique expertise must be well-controlled in experiments, i.e., interaction expertise should be somewhat similar across conditions.*

The observed improvement in typing speed with the OPTI layout relative to QWERTY in both Studies 1 (OPTI: 12.7 WPM, QWERTY: 11.7 WPM) and 2 (OPTI: 15.7 WPM, QWERTY: 15.0 WPM), although significantly different, was not that high in absolute terms. Even the predicted speeds of 21.3 WPM for OPTI and 19.9 WPM for QWERTY with 282 ms dwell time are, relatively speaking, not a big improvement. Although both the OPTI and QWERTY layouts outperformed Swipe&Switch's 13.7 WPM [31] and CS QWERTY's 13.4 WPM [44], we encourage future research to use the GEM to explore other, theoretically more optimal, layouts [61].

Still, we caution that our typing speed estimates also match the prediction made by Kristensson and Vertanen [28], who identified that 20 WPM could be the maximum reachable on a dwell QWERTY layout. We see swipe typing (e.g., [13, 78]), which is predicted to perform better than tapping for touch-based text entry [61], to be a better direction for further work in gaze-based text entry systems. Gaze-based swipe typing can potentially reach 46 WPM on a QWERTY layout [28]. Thus, we recommend that, with training users first on the interaction technique over multiple sessions, future studies use the GEM to investigate gaze-based swipe typing on better layouts, e.g., [61]. This approach will potentially yield better estimates of dual-trained performance and, answer whether an unfamiliar layout is worth learning [48].

Another area to explore is the applicability and effectiveness of the GEM on dynamic keyboard layouts. A challenge in adapting the GEM to dynamic layouts like Dasher [75] would be the design and implementation of the planning phase. One solution could be to reveal all the correct depth layers of Dasher side-by-side by default for each word, and for participants to plan/practice with this exposed layout until they feel ready. For layouts like SliceType [3] that changes in place by dynamically increasing the size of the most probable keys after each key press, GEM could highlight a larger area, i.e., beyond the size of the key, to depict that more area will be eventually assigned to this key, thus leveraging flexible and (somewhat) realistic planning. As for systems that support auto-completion via word predictions, e.g., [9], highlighting keys can be limited to only those letters that need to be typed before the system correctly predicts the target. Similarly, the target word in the list of predictions could also be highlighted beforehand, i.e., from the planning phase.

As for the societal risks of our work, we do not foresee significant risks beyond minor headaches and eye fatigue from extended VR headset and eye tracker use. In other words, the potential risks are minimal. The ethics committee of our organization also approved both studies on such grounds.

7 Limitations

Although our interaction-trained OPTI-layout-novice users performed similarly to dual-trained users, it is unknown what level of traditional training is required to reach that performance. More studies are also needed to explore how a diverse participant pool, e.g., different types of interaction-trained users who did *not* only train on a specific text entry system, would affect the outcome of a GEM-based evaluation with different layouts and interaction techniques.

We acknowledge that using a gaze cursor instead of highlighting the gazed-upon key may have limited our results [14, 15]. We also recognize that the 11-phrase GEM sessions required participants to type fewer phrases than the 15-minute traditional sessions. We used only eleven phrases for the GEM as we wanted to keep the sessions short while also keeping confounds due to a limited variety of phrases, e.g., in terms of length and character sequence, at its minimum. This approach still maintains comparability with the literature, as several studies defined a single session to comprise even fewer phrases, e.g., five [22, 29, 44]. Also, we do not believe that this approach significantly affected the results, as our findings are consistent across our and previous studies, e.g., [40, 59].

Also, although the learning bias in Study 2 was effectively *against* the GEM for the OPTI layout, the order of OPTI and QWERTY was not counterbalanced. While an extra day's training with the interaction technique was carried forward on the ninth day, i.e., the OPTI day, this should not affect the results substantially. As shown in other work [47], the performance in sessions 6-8 in Study 2 already stopped showing significant improvements. Moreover, our Study 2 results were similar to that of Study 1. Finally, while participants were trained on the gaze dwell interaction technique, analyzing the learnability of the interaction technique was not the focus of this work. Thus, these findings were reported in a separate work [47].

8 Conclusion

In this paper, we evaluated the robustness of a recently proposed evaluation method, the GEM, in terms of its applicability to an unfamiliar interaction technique – eye gaze dwell. Results from our user studies revealed that the GEM can mitigate the effects of layout learning by simulating the typing behavior of a layout-trained user very well. However, interaction technique expertise heavily influenced accurate estimates of dual-trained (i.e., both layout- and interaction-trained) user performance. Still, when the level of interaction technique expertise was similar, the GEM was effective in identifying in a very short period, specifically the time required to type eleven phrases, whether one text entry system/layout is better than the other, even if one of the layouts is very well-known to participants (e.g., QWERTY). This feature of the GEM greatly reduces the need for longitudinal studies, and thus, this approach offers great potential for text entry research.

References

[1] A. S. Arif and W. Stuerzlinger. 2009. Analysis of text entry performance metrics. In *2009 IEEE Toronto International Conference Science and Technology for Humanity (TIC-STH)*. IEEE, New York, USA, 100–105. doi:10.1109/TIC-STH.2009.5444533

[2] Nikolaus Bee and Elisabeth André. 2008. Writing with Your Eye: A Dwell Time Free Writing System Adapted to the Nature of Human Eye Gaze. In *Perception in Multimodal Dialogue Systems*, Elisabeth André, Laila Dybkjær, Wolfgang Minker, Heiko Neumann, Roberto Pieraccini, and Michael Weber (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 111–122.

[3] Burak Benligiray, Cihan Topal, and Cuneyt Akinlar. 2019. Slicetype: fast gaze typing with a merging keyboard. *Journal on Multimodal User Interfaces* 13, 4 (2019), 321–334.

[4] Xiaojun Bi, Barton A. Smith, and Shumin Zhai. 2010. Quasi-Qwerty Soft Keyboard Optimization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 283–286. doi:10.1145/1753326.1753367

[5] Xiaojun Bi and Shumin Zhai. 2016. IJQwerty: What Difference Does One Key Change Make? Gesture Typing Keyboard Optimization Bounded by One Key Position Change from Qwerty. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 49–58. doi:10.1145/2858036.2858421

[6] Steven J. Castellucci, I. Scott MacKenzie, Mudit Misra, Laxmi Pandey, and Ahmed Sabbir Arif. 2019. TiltWriter: Design and Evaluation of a No-Touch Tilt-Based Text Entry Method for Handheld Devices. In *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia* (Pisa, Italy) *(MUM '19)*. Association for Computing Machinery, New York, NY, USA, Article 7, 8 pages. doi:10.1145/3365610.3365629

[7] Edward Clarkson, James Clawson, Kent Lyons, and Thad Starner. 2005. An Empirical Study of Typing Rates on Mini-QWERTY Keyboards. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems* (Portland, OR, USA) *(CHI EA '05)*. Association for Computing Machinery, New York, NY, USA, 1288–1291. doi:10.1145/1056808.1056898

[8] Wenzhe Cui, Rui Liu, Zhi Li, Yifan Wang, Andrew Wang, Xia Zhao, Sina Rashidian, Furqan Baig, IV Ramakrishnan, Fusheng Wang, and Xiaojun Bi. 2023. GlanceWriter: Writing Text by Glancing Over Letters with Gaze. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 719, 13 pages. doi:10.1145/3544548.3581269

[9] Antonio Diaz-Tula and Carlos H. Morimoto. 2016. AugKey: Increasing Foveal Throughput in Eye Typing with Augmented Keys. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 3533–3544. doi:10.1145/2858036.2858517

[10] Tafadzwa Joseph Dube and Ahmed Sabbir Arif. 2019. Text Entry in Virtual Reality: A Comprehensive Review of the Literature. In *Human-Computer Interaction. Recognition and Interaction Technologies*, Masaaki Kurosu (Ed.). Springer International Publishing, Cham, 419–437.

[11] Mark Dunlop and John Levine. 2012. Multidimensional Pareto Optimization of Touchscreen Keyboards for Speed, Familiarity and Improved Spell Checking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 2669–2678. doi:10.1145/2207676.2208659

[12] K Anders Ericsson, Ralf T Krampe, and Clemens Tesch-Römer. 1993. The role of deliberate practice in the acquisition of expert performance. *Psychological review* 100, 3 (1993), 363.

[13] Wenxin Feng, Jiangnan Zou, Andrew Kurauchi, Carlos H Morimoto, and Margrit Betke. 2021. HGaze Typing: Head-Gesture Assisted Gaze Typing. In *ACM Symposium on Eye Tracking Research and Applications*. Association for Computing Machinery, New York, NY, USA, Article 11, 11 pages. https://doi.org/10.1145/3448017.3457379

[14] Ajoy S. Fernandes, T. Scott Murdison, and Michael J. Proulx. 2023. Leveling the Playing Field: A Comparative Reevaluation of Unmodified Eye Tracking as an Input and Interaction Modality for VR. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (2023), 2269–2279. doi:10.1109/TVCG.2023.3247058

[15] Ajoy S Fernandes, Immo Schütz, T Scott Murdison, and Michael J Proulx. 2025. Gaze Inputs for Targeting: The Eyes Have It, Not With a Cursor. *International Journal of Human–Computer Interaction* 0, 0 (2025), 1–19. doi:10.1080/10447318.2025.2453966 arXiv:https://doi.org/10.1080/10447318.2025.2453966

[16] Sanjay Ghosh, Anirudha Joshi, Manjiri Joshi, Nagraj Emmadi, Girish Dalvi, Shashank Ahire, and Swati Rangale. 2017. Shift+Tap or Tap+LongPress? The Upper Bound of Typing Speed on InScript. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 2059–2063. doi:10.1145/3025453.3025944

[17] Dominik Grüneis, Marc Kurz, and Erik Sonnleitner. 2023. Let Me Help You: Improving the Novice Experience of High-Performance Keyboard Layouts with Visual Clues. *Applied Sciences* 13, 16 (2023), 16 pages. doi:10.3390/app13169391

[18] Joseph F Hair Jr, William C Black, Barry J Babin, and Rolph E. Anderson. 2014. Multivariate data analysis.

[19] Hiroyuki Hakoda, Buntarou Shizuki, and Jiro Tanaka. 2016. QAZ Keyboard: QWERTY Based Portrait Soft Keyboard. In *Design, User Experience, and Usability: Technological Contexts*, Aaron Marcus (Ed.). Springer International Publishing, Cham, 24–35.

[20] John Paulin Hansen, Anders Sewerin Johansen, Dan Witzner Hansen, Kenji Ito, and Satoru Mashino. 2003. Command Without a Click: Dwell Time Typing by Mouse and Gaze Selections. In *Human-Computer Interaction INTERACT '03: IFIP TC13 International Conference on Human-Computer Interaction, 1st-5th September 2003, Zurich, Switzerland*, Matthias Rauterberg, Marino Menozzi, and Janet Wesson (Eds.). IOS Press, Zurich, Switzerland.

[21] John Paulin Hansen, Vijay Rajanna, I. Scott MacKenzie, and Per Bækgaard. 2018. A Fitts' Law Study of Click and Dwell Interaction by Gaze, Head and Mouse with a Head-Mounted Display. In *Proceedings of the Workshop on Communication by Gaze Interaction* (Warsaw, Poland) *(COGAIN '18)*. Association for Computing Machinery, New York, NY, USA, Article 7, 5 pages. doi:10.1145/3206343.3206344

[22] Ramin Hedeshy, Chandan Kumar, Raphael Menges, and Steffen Staab. 2021. Hummer: Text Entry by Gaze and Hum. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, Article 741, 11 pages. https://doi.org/10.1145/3411764.3445501

[23] Baosheng James Hou, Joshua Newn, Ludwig Sidenmark, Anam Ahmad Khan, Per Bækgaard, and Hans Gellersen. 2023. Classifying Head Movements to Separate Head-Gaze and Head Gestures as Distinct Modes of Input. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>Hamburg</city>, <country>Germany</country>, </conf-loc>) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 253, 14 pages. doi:10.1145/3544548.3581201

[24] Moaaz Hudhud Mughrabi, Aunnoy K Mutasim, Wolfgang Stuerzlinger, and Anil Ufuk Batmaz. 2022. My Eyes Hurt: Effects of Jitter in 3D Gaze Tracking. 310-315 pages. doi:10.1109/VRW55335.2022.00070

[25] Toshiya Isomoto, Toshiyuki Ando, Buntarou Shizuki, and Shin Takahashi. 2018. Dwell Time Reduction Technique Using Fitts' Law for Gaze-Based Target Acquisition. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research and Applications* (Warsaw, Poland) *(ETRA '18)*. Association for Computing Machinery, New York, NY, USA, Article 26, 7 pages. doi:10.1145/3204493.3204532

[26] Robert JK Jacob. 1995. Eye tracking in advanced interface design. *Virtual environments and advanced interface design* 258, 288 (1995), 2.

[27] Jussi P. P. Jokinen, Sayan Sarcar, Antti Oulasvirta, Chaklam Silpasuwanchai, Zhenxin Wang, and Xiangshi Ren. 2017. Modelling Learning of New Keyboard Layouts. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 4203–4215. doi:10.1145/3025453.3025580

[28] Per Ola Kristensson and Keith Vertanen. 2012. The Potential of Dwell-Free Eye-Typing for Fast Assistive Gaze Communication. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (Santa Barbara, California) *(ETRA '12)*. Association for Computing Machinery, New York, NY, USA, 241–244. doi:10.1145/2168556.2168605

[29] Chandan Kumar, Ramin Hedeshy, I. Scott MacKenzie, and Steffen Staab. 2020. TAGSwipe: Touch Assisted Gaze Swipe for Text Entry. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3313831.3376317

[30] Andrew Kurauchi, Wenxin Feng, Ajjen Joshi, Carlos Morimoto, and Margrit Betke. 2016. EyeSwipe: Dwell-Free Text Entry Using Gaze Paths. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1952–1956. doi:10.1145/2858036.2858335

[31] Andrew Kurauchi, Wenxin Feng, Ajjen Joshi, Carlos H. Morimoto, and Margrit Betke. 2020. Swipe&Switch: Text Entry Using Gaze Paths and Context Switching. In *Adjunct Publication of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '20 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 84–86. doi:10.1145/3379350.3416193

[32] Frank Chun Yat Li, Richard T. Guy, Koji Yatani, and Khai N. Truong. 2011. The 1line Keyboard: A QWERTY Layout in a Single Line. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology* (Santa Barbara, California, USA) *(UIST '11)*. Association for Computing Machinery, New York, NY, USA, 461–470. doi:10.1145/2047196.2047257

[33] X. Lu, D. Yu, H. N. Liang, W. Xu, Y. Chen, X. Li, and K. Hasan. 2020. Exploration of Hands-free Text Entry Techniques For Virtual Reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, New York, USA, 344–349. doi:10.1109/ISMAR50242.2020.00061

[34] Christof Lutteroth, Moiz Penkar, and Gerald Weber. 2015. Gaze vs. Mouse: A Fast and Accurate Gaze-Only Click Alternative. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (Charlotte, NC, USA) *(UIST '15)*. Association for Computing Machinery, New York, NY, USA, 385–394. doi:10.1145/2807442.2807461

[35] Kent Lyons, Thad Starner, Daniel Plaisted, James Fusia, Amanda Lyons, Aaron Drew, and E. W. Looney. 2004. Twiddler Typing: One-Handed Chording Text Entry for Mobile Phones. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vienna, Austria) *(CHI '04)*. Association for Computing Machinery, New York, NY, USA, 671–678. doi:10.1145/985692.985777

[36] I. Scott MacKenzie and R. William Soukoreff. 2003. Phrase Sets for Evaluating Text Entry Techniques. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI EA '03)*. Association for Computing Machinery, New York, NY, USA, 754–755. doi:10.1145/765891.765971

[37] 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* (Pittsburgh, Pennsylvania, USA) *(CHI '99)*. Association for Computing Machinery, New York, NY, USA, 25–31. doi:10.1145/302979.302983

[38] I. Scott MacKenzie and Shawn X. Zhang. 2001. An empirical investigation of the novice experience with soft keyboards. *Behaviour & Information Technology* 20, 6 (2001), 411–418. doi:10.1080/01449290110089561 arXiv:https://doi.org/10.1080/01449290110089561

[39] Laurent Magnien, Jean Léon Bouraoui, and Nadine Vigouroux. 2004. Mobile Text Input with Soft Keyboards: Optimization by Means of Visual Clues. In *Mobile Human-Computer Interaction - MobileHCI 2004*, Stephen Brewster and Mark Dunlop (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 337–341.

[40] Päivi Majaranta, Ulla-Kaija Ahola, and Oleg Špakov. 2009. Fast Gaze Typing with an Adjustable Dwell Time. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 357–360. doi:10.1145/1518701.1518758

[41] Päivi Majaranta, I Scott MacKenzie, Anne Aula, and Kari-Jouko Räihä. 2006. Effects of feedback and dwell time on eye typing speed and accuracy. *Universal Access in the Information Society* 5, 2 (2006), 199–208.

[42] Paul Mallery and Darren George. 2003. *SPSS for Windows step by step: a simple guide and reference.* Pearson, London, UK. 1–402 pages.

[43] Carlos H. Morimoto and Arnon Amir. 2010. Context Switching for Fast Key Selection in Text Entry Applications. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (Austin, Texas) *(ETRA '10)*. Association for Computing Machinery, New York, NY, USA, 271–274. doi:10.1145/1743666.1743730

[44] Carlos H. Morimoto, Jose A. T. Leyva, and Antonio Diaz-Tula. 2018. Context Switching Eye Typing Using Dynamic Expanding Targets. In *Proceedings of the Workshop on Communication by Gaze Interaction* (Warsaw, Poland) *(COGAIN '18)*. Association for Computing Machinery, New York, NY, USA, Article 6, 9 pages. doi:10.1145/3206343.3206347

[45] Martez E. Mott, Shane Williams, Jacob O. Wobbrock, and Meredith Ringel Morris. 2017. Improving Dwell-Based Gaze Typing with Dynamic, Cascading Dwell Times. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 2558–2570. doi:10.1145/3025453.3025517

[46] Aunnoy Mutasim, Anil Ufuk Batmaz, Moaaz Hudhud Mughrabi, and Wolfgang Stuerzlinger. 2022. Performance Analysis of Saccades for Primary and Confirmatory Target Selection. In *Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology* (Tsukuba, Japan) *(VRST '22)*. Association for Computing Machinery, New York, NY, USA, Article 18, 12 pages. doi:10.1145/3562939.3565619

[47] Aunnoy K Mutasim, Mohammad Raihanul Bashar, Christof Lutteroth, Anil Ufuk Batmaz, and Wolfgang Stuerzlinger. 2025. There Is More to Dwell Than Meets the Eye: Toward Better Gaze-Based Text Entry Systems With Multi-Threshold Dwell. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '25)*. Association for Computing Machinery, New York, NY, USA, 25. doi:10.1145/3706598.3713781

[48] Aunnoy K Mutasim, Anil Ufuk Batmaz, Moaaz Hudhud Mughrabi, and Wolfgang Stuerzlinger. 2024. The Guided Evaluation Method: An easier way to empirically estimate trained user performance for unfamiliar keyboard layouts. *International Journal of Human-Computer Studies* 190 (2024), 103317. doi:10.1016/j.ijhcs.2024.103317

[49] Aunnoy K Mutasim, Anil Ufuk Batmaz, and Wolfgang Stuerzlinger. 2020. Gaze Tracking for Eye-Hand Coordination Training Systems in Virtual Reality. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–9. doi:10.1145/3334480.3382924

[50] Aunnoy K Mutasim, Anil Ufuk Batmaz, and Wolfgang Stuerzlinger. 2021. Pinch, Click, or Dwell: Comparing Different Selection Techniques for Eye-Gaze-Based Pointing in Virtual Reality. In *ACM Symposium on Eye Tracking Research and Applications* (Virtual Event, Germany) *(ETRA '21 Short Papers)*. Association for Computing Machinery, New York, NY, USA, Article 15, 7 pages. doi:10.1145/3448018.3457998

[51] Aunnoy K Mutasim, Moaaz Hudhud Mughrabi, Anil Ufuk Batmaz, and Wolfgang Stuerzlinger. 2023. Does Repeatedly Typing the Same Phrase Provide a Good Estimate of Expert Text Entry Performance?. Extended Abstract. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, 8 pages. doi:10.1145/3544549.3585647

[52] Aanand Nayyar, Utkarsh Dwivedi, Karan Ahuja, Nitendra Rajput, Seema Nagar, and Kuntal Dey. 2017. OptiDwell: Intelligent Adjustment of Dwell Click Time. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces* (Limassol, Cyprus) *(IUI '17)*. Association for Computing Machinery, New York, NY, USA, 193–204. doi:10.1145/3025171.3025202

[53] Prateek Panwar, Sayan Sarcar, and Debasis Samanta. 2012. EyeBoard: A fast and accurate eye gaze-based text entry system. In *2012 4th International Conference on Intelligent Human Computer Interaction (IHCI)*. IEEE, Kharagpur, India, 1–8. doi:10.1109/IHCI.2012.6481793

[54] Pawan Patidar, Himanshu Raghuvanshi, and Sayan Sarcar. 2014. Quickpie: An Interface for Fast and Accurate Eye Gazed based Text Entry. arXiv:1407.7313 [cs.HC]

[55] Diogo Pedrosa, Maria Da Graça Pimentel, Amy Wright, and Khai N. Truong. 2015. Filteryedping: Design Challenges and User Performance of Dwell-Free Eye Typing. *ACM Trans. Access. Comput.* 6, 1, Article 3 (mar 2015), 37 pages. doi:10.1145/2724728

[56] Abdul Moiz Penkar, Christof Lutteroth, and Gerald Weber. 2012. Designing for the eye: design parameters for dwell in gaze interaction. In *Proceedings of the 24th Australian Computer-Human Interaction Conference* (Melbourne, Australia) *(OzCHI '12)*. Association for Computing Machinery, New York, NY, USA, 479–488. doi:10.1145/2414536.2414609

[57] Jimin Pi and Bertram E. Shi. 2017. Probabilistic adjustment of dwell time for eye typing. In *2017 10th International Conference on Human System Interactions (HSI)*. IEEE, Ulsan, Korea (South), 251–257. doi:10.1109/HSI.2017.8005041

[58] Yuan Yuan Qian and Robert J. Teather. 2017. The Eyes Don't Have It: An Empirical Comparison of Head-Based and Eye-Based Selection in Virtual Reality. In *Proceedings of the 5th Symposium on Spatial User Interaction* (Brighton, United Kingdom) *(SUI '17)*. Association for Computing Machinery, New York, NY, USA, 91–98. doi:10.1145/3131277.3132182

[59] Kari-Jouko Räihä and Saila Ovaska. 2012. An Exploratory Study of Eye Typing Fundamentals: Dwell Time, Text Entry Rate, Errors, and Workload. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 3001–3010. doi:10.1145/2207676.2208711

[60] Vijay Rajanna and John Paulin Hansen. 2018. Gaze Typing in Virtual Reality: Impact of Keyboard Design, Selection Method, and Motion. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research and Applications* (Warsaw, Poland) *(ETRA '18)*. Association for Computing Machinery, New York, NY, USA, Article 15, 10 pages. doi:10.1145/3204493.3204541

[61] Jochen Rick. 2010. Performance Optimizations of Virtual Keyboards for Stroke-Based Text Entry on a Touch-Based Tabletop. In *Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology* (New York, New York, USA) *(UIST '10)*. Association for Computing Machinery, New York, NY, USA, 77–86. doi:10.1145/1866029.1866043

[62] Daniel Rough, Keith Vertanen, and Per Ola Kristensson. 2014. An Evaluation of Dasher with a High-Performance Language Model as a Gaze Communication Method. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces* (Como, Italy) *(AVI '14)*. Association for Computing Machinery, New York, NY, USA, 169–176.

doi:10.1145/2598153.2598157
[63] Sayan Sarcar, Prateek Panwar, and Tuhin Chakraborty. 2013. EyeK: An Efficient Dwell-Free Eye Gaze-Based Text Entry System. In *Proceedings of the 11th Asia Pacific Conference on Computer Human Interaction* (Bangalore, India) *(APCHI '13)*. Association for Computing Machinery, New York, NY, USA, 215–220. doi:10.1145/2525194.2525288
[64] Immo Schuetz, T. Scott Murdison, Kevin J. MacKenzie, and Marina Zannoli. 2019. An Explanation of Fitts' Law-like Performance in Gaze-Based Selection Tasks Using a Psychophysics Approach. 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, 1–13. doi:10.1145/3290605.3300765
[65] Andrew Sears, Julie A. Jacko, Josey Chu, and Francisco Moro. 2001. The role of visual search in the design of effective soft keyboards. *Behaviour & Information Technology* 20, 3 (2001), 159–166. doi:10.1080/01449290110049790 arXiv:https://doi.org/10.1080/01449290110049790
[66] Ludwig Sidenmark and Hans Gellersen. 2019. Eye&Head: Synergetic Eye and Head Movement for Gaze Pointing and 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, 1161–1174. doi:10.1145/3332165.3347921
[67] R. William Soukoreff and I. Scott MacKenzie. 2001. Measuring Errors in Text Entry Tasks: An Application of the Levenshtein String Distance Statistic. In *CHI '01 Extended Abstracts on Human Factors in Computing Systems* (Seattle, Washington) *(CHI EA '01)*. Association for Computing Machinery, New York, NY, USA, 319–320. doi:10.1145/634067.634256
[68] R. William Soukoreff and I. Scott MacKenzie. 2003. Metrics for Text Entry Research: An Evaluation of MSD and KSPC, and a New Unified Error Metric. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI '03)*. Association for Computing Machinery, New York, NY, USA, 113–120. doi:10.1145/642611.642632
[69] Marco Speicher, Anna Maria Feit, Pascal Ziegler, and Antonio Krüger. 2018. Selection-Based Text Entry in Virtual Reality. 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, 1–13. doi:10.1145/3173574.3174221
[70] Robert J. Teather and Wolfgang Stuerzlinger. 2014. Visual Aids in 3D Point Selection Experiments. In *Proceedings of the 2nd ACM Symposium on Spatial User Interaction* (Honolulu, Hawaii, USA) *(SUI '14)*. Association for Computing Machinery, New York, NY, USA, 127–136. doi:10.1145/2659766.2659770
[71] Outi Tuisku, Päivi Majaranta, Poika Isokoski, and Kari-Jouko Räihä. 2008. Now Dasher! Dash Away! Longitudinal Study of Fast Text Entry by Eye Gaze. In *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications* (Savannah, Georgia) *(ETRA '08)*. Association for Computing Machinery, New York, NY, USA, 19–26. doi:10.1145/1344471.1344476
[72] Mario H. Urbina and Anke Huckauf. 2007. Dwell time free eye typing approaches. In *Proceedings of the 3rd Conference on Communication by Gaze Interaction* (Leicester, UK) *(COGAIN '07)*. Association for Computing Machinery, New York, NY, USA, 65–70.
[73] Mario H. Urbina and Anke Huckauf. 2010. Alternatives to single character entry and dwell time selection on eye typing. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (Austin, Texas) *(ETRA '10)*. Association for Computing Machinery, New York, NY, USA, 315–322. doi:10.1145/1743666.1743738
[74] Oleg Špakov and Darius Miniotas. 2004. On-line adjustment of dwell time for target selection by gaze. In *Proceedings of the Third Nordic Conference on Human-Computer Interaction* (Tampere, Finland) *(NordiCHI '04)*. Association for Computing Machinery, New York, NY, USA, 203–206. doi:10.1145/1028014.1028045
[75] David J Ward, Alan F Blackwell, and David JC MacKay. 2002. Dasher: A gesture-driven data entry interface for mobile computing. *Human–Computer Interaction* 17, 2-3 (2002), 199–228.
[76] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. ACM, New York, NY, USA, 143–146. doi:10.1145/1978942.1978963
[77] Chun Yu, Ke Sun, Mingyuan Zhong, Xincheng Li, Peijun Zhao, and Yuanchun Shi. 2016. One-Dimensional Handwriting: Inputting Letters and Words on Smart Glasses. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 71–82. doi:10.1145/2858036.2858542
[78] Maozheng Zhao, Alec M Pierce, Ran Tan, Ting Zhang, Tianyi Wang, Tanya R. Jonker, Hrvoje Benko, and Aakar Gupta. 2023. Gaze Speedup: Eye Gaze Assisted Gesture Typing in Virtual Reality. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (Sydney, NSW, Australia) *(IUI '23)*. Association for Computing Machinery, New York, NY, USA, 595–606. doi:10.1145/3581641.3584072

Received November 2024; revised February 2025; accepted March 2025