Recommendations for the Design, Conduct, and Reporting of Research on Text Entry with Alternative Access Interfaces

Heidi Koester[1] and Sajay Arthanat[2]

[1]Koester Performance Research, [2]University of New Hampshire

ABSTRACT

Based on our systematic review of research on computer text entry by people with physical disabilities [1,2], we have identified some inconsistencies and gaps in the design and reporting of text entry studies. This paper presents some recommendations for improvement, in order to strengthen the utility and replicability of future studies. Key recommendations include: provide enough information to allow replication of the study; include participants who are members of the target user population; specify thoroughly the intervention, data collection task, and definition of dependent variables; and report sufficient descriptive statistics to allow for future meta-analysis across studies. Following these recommendations can help ensure that the considerable time and effort put in to conducting a study will provide lasting impact to the field.

INTRODUCTION

Providing effective access interventions to people with physical impairments often includes the important task of facilitating an easy and productive means of text entry. We have been working to organize the available published evidence in this area, to help answer questions such as: What is the typical typing speed when an individual with a C6 spinal cord injury uses typing splints on a standard keyboard? Our goal is to build foundational knowledge that can inform decision-making for device selection and configuration and provide rough expectations for learning and long-term performance. In our review of published evidence, we identified several inconsistencies and gaps in the design and description of text entry studies. The purpose of this paper is to provide methodological recommendations to strengthen utility and replicability of future research studies.

BACKGROUND

We recently completed a systematic review on computer text entry by people with physical disabilities. We found 39 studies that met all inclusion criteria, dating back to 1986, and have used the data to report on the text entry rates associated with different interfaces, diagnoses, and body sites [1,2].

Studies were included if individuals with physical impairments were in the study population, typing speed was reported in words per minute or equivalent, and the access interface was available for public use. Studies were excluded if the method of measuring typing speed did not follow conventional, or equally valid, techniques, and if the results reported were anecdotal or unclear. Additionally, to quantitatively combine results across different studies, studies needed to provide sample size, average, and standard deviation for text entry rate or to give enough information for us to calculate those summary statistics.

While the systematic review yielded useful results, it also revealed that many additional studies are needed, and that those studies that have been conducted do not always report important details or use consistent methods. Study authors have an opportunity to design methods that not only address the specific questions of that individual study, but also allow use of the data for meta-analysis and cumulation across multiple studies. By reporting key details for the methods and results, a study's contribution can be leveraged for greater impact on the field.

GOALS

The immediate goal of this paper is to contribute to the conduct of strong and useful studies in the access domain. Based on examining dozens of studies in our systematic review, as well as our own experience conducting similar studies, we offer the following recommendations for the design, conduct, and reporting of research on text entry with alternative access interfaces.

In the longer term, we hope to refine these recommendations, with input from others, to define a common structure for performing text entry studies. Such a common structure will provide a stronger platform for cumulating results across studies over time and building a strong knowledge base to inform decision-making.

RECOMMENDATIONS

The following recommendations include specific suggestions for a few key areas. Additionally, we recommend that authors report the choices made in each of these areas, whether they follow our specific suggestions or not. Authors should strive to provide enough detail to allow another person to replicate the study; if space is limited when reporting on the study, a link to the full protocol and results could be provided.

Participants

This defines who is going to enter text and be measured in the study. We strongly recommend including real-world users as participants, i.e., people with physical impairments who already use the interface or who fit the characteristics of the interface's target user group. Users without impairments may have a role in some studies, such as providing a convenient sampling pool for early stage, proof of concept work, or serving as a complementary group in a study that also includes people with impairments. But unless real-world users who are potential users of the interfaces in terms of their diagnosis and motor deficits are also included, the external validity of the study is likely to be limited.

Participant Characteristics

Provide information on key participant demographics, such as age, gender, as well as education, socioeconomic status, and employment, if available. Also include information about each participant's disabling condition. This is commonly presented as the individual's medical diagnosis, which is useful but can be insufficient to understand the degree of impairment in diagnoses such as cerebral palsy (CP). Reporting functional measures such as the Gross Motor Function Classification System (for people with CP) can provide a much stronger basis for identifying and grouping similar participants, both within and between studies. Functional measures of any kind were reported in only a handful of the 39 studies in our review, so we were forced to rely on diagnosis alone in our analyses. Descriptions of participants' strengths and limitations are better than nothing but can be difficult to use those to group similar participants across studies.

Intervention Specification

This umbrella term is used by Lenker et al. [3] to cover assistive technology devices and the services associated with their provision and implementation. In the context of text entry studies, this includes a full description of the interface used for text entry, how the user interacts with the interface, and the context of its use. The intervention may be specified for the participant group as a whole when homogeneous across the entire group. Otherwise, report for each participant. Some specific aspects to consider include:

1. Interface used: general type of interface, specific setup (e.g., use of word prediction, letter layout), availability of the interface (prototype, commercially available)
2. Body site: how exactly the participant interacts with the interface
3. Participant experience: both with this specific interface and with text entry in any form. Report the amount of experience with the interface prior to the study (in months or years of regular use), as well as during the study period itself (hours of use during study sessions and estimated hours of use outside of study sessions, where applicable).
4. Training: report the training protocol including length of time, frequency, duration, goals set, etc.
5. Other service delivery features: For those studies conducted in a clinical service provision context, provide information about the practice setting, practitioner knowledge and skills, assessment and recommendation process, participant context (where interface was used or considered for), goals set, etc.

See our reviews [1,2] for category definitions for interfaces, body sites, and participant experience. Training and other service delivery features may not have pre-defined categories, but providing the information will allow others to put the data into categories, even if those categories are defined in an ad hoc manner.

Procedure

This defines exactly what each participant was asked to do in the study. Ensuring that each participant received the same instructions and reporting the procedure in sufficient detail to be replicable helps demonstrate treatment fidelity [3].

Typing Task

A key part of the procedure in a text entry study is the typing task itself. Use a transcription task unless the goals of the study make that inappropriate [4,5]. A composition task can also be included if that is of interest.

Instructions and subsequent report on the study should be explicit about how errors are handled; we typically allow participants to correct errors that they notice during text entry. The length of the task can be defined by the length of text (e.g., 3 sentences) or amount of time (e.g., 5 minutes). Either way, it should represent at least a few minutes' worth of text entry. The text itself should be representative of basic English (or the language used in the study), and if participants enter text in multiple sessions over time, the text in each session should be similar in characteristics (word length, reading level) but not identical.

Measuring Dependent Variables

In a text entry study, the most commonly reported, and perhaps most relevant, dependent variables are text entry rate (TER) and error rate. Clear operational definitions of each of these variables (and any others that are measured) are necessary.

For TER, the measure should be correct characters per minute. It can be useful to report TER in correct words per minute in English by dividing by 5 [4]. The key point is that incorrect characters do not count toward TER, and this needs to be explicitly reported. In our systematic review of TER, we were fairly liberal in accepting results based on vague descriptions of how TER was measured, since measuring correct WPM is a common approach. It would have provided more confidence if all studies had explicitly specified how errors were considered in the TER measurement.

Error rate reflects the number of incorrect characters entered during the task and can seem intuitive to measure. But the specifics of measuring error rate can be tricky [6]. If circumstances require a manual measure of error rate, it can be approximated by counting the incorrect characters remaining in the transcription text divided by the total number of characters entered.

An easier and more accurate approach to measuring text entry performance is to use a software tool designed for that purpose. These tools help ensure the validity and replicability of the text entry task by providing consistent task presentation, data collection, and data analysis. MacKenzie and Wobbrock each have made available the software they use in their text entry research [7,8]. Compass software for access assessment can also be used, and this tool provides additional flexibility in test presentation and setup (available at kpronline.com). The usability, measurement accuracy, and psychometric validity and reliability of Compass has also been demonstrated [9,10,11].

Statistics and Presentation of Results

A study's statistics need to first address the specific goals of the study, but we urge authors to also consider presenting enough information to allow others to include the data in future meta-analyses or systematic reviews. To pool text entry rate across studies, the following descriptive statistics must be reported for a participant group: sample size (number of participants), average TER, and standard deviation of TER. The range for the group, as well as outliers or quartile ranges, can also be useful. The most flexible approach is to report results for each individual participant, including key characteristics of that participant (e.g., diagnosis, body site, demographics, functional measures, interface setup). This gives the most possibilities for post hoc analysis by other researchers.

A second key consideration comes into play when there are multiple measures for a given interface over time. It can be helpful to define clearly which of the measures is most representative and note the extent to which this measure might represent typical long-term performance for this individual or participant group. For example, is the TER in the last session the "definitive" one, or is it better to take the average across sessions for this study? Any systematic review or other post hoc analysis will need to make this decision, so it is useful to have the authors' perspective clearly stated in the original paper.

DISCUSSION

These recommendations can help address the clear need to develop and implement standard and replicable methods with reliable outcome tools for examining text entry performance. Incorporating these recommendations helps ensure a well-designed study, one that can address its own goals in a valid and robust way as well as support a powerful synthesis of information across studies. These recommendations can also be used by practitioners wishing to collect text entry data with their clients for clinical purposes. Text entry studies should strive to use appropriate and consistent methods in order to facilitate effective replication, evidence synthesis, and knowledge translation. This helps ensure that the considerable time and effort put in to conducting the study will provide lasting impact to the field.

CONTACT

Heidi Koester, hhk@kpronline.com

REFERENCES

[1] Koester HH, Arthanat S. (2017). Effect of diagnosis, body site, and experience on text entry rate of individuals with physical disabilities: A systematic review. *Disability & Rehabilitation: Assistive Technology*, ePub ahead of print. http://www.tandfonline.com/eprint/MKjjiM9rX8EiRbBR5sZe/full

[2] Koester HH, Arthanat S. (2017). Text Entry Rate of Access Interfaces Used by People with Physical Disabilities: A Systematic Review. *Assistive Technology*, ePub ahead of print. http://www.tandfonline.com/eprint/4A2Gu8hJRmY8cUVsbdxk/full

[3] Lenker JA, Fuhrer MJ, Jutai JW, Demers L, Scherer MJ, & DeRuyter F. (2010). Treatment theory, intervention specification, and treatment fidelity in assistive technology outcomes research. *Assistive Technology*, *22*(3), 129–138 10p. https://doi.org/10.1080/10400430903519910.

[4] MacKenzie IS. (2007). Evaluation of text entry techniques. In MacKenzie iS, Tanaki-Ishii K (Eds.) *Text entry systems: Mobility, accessibility, universality*, pp 75-101. San Francisco CA: Morgan Kaufmann.

[5] Koester HH. (1993). Methodological decisions in the study of augmentative communication systems. *Proceedings of RESNA 1993 Annual Conference*, Las Vegas, NV. Arlington, VA: RESNA Press. Available at http://kpronline.com/pubs.php.

[6] Soukoreff RW, MacKenzie IS. (2003). Metrics for text entry research: An evaluation of MSD and KSPC, and a new unified error metric. *Proceedings of the ACM Conference on Human Factors in Computing Systems - CHI 2003*, pp. 113-120. New York: ACM.

[7] MacKenzie IS. TypingTestExperiment software. Available at http://www.yorku.ca/mack/ExperimentSoftware/.

[8] Wobbrock J. TextTest software. Available at http://depts.washington.edu/madlab/proj/texttest/.

[9] Koester HH, LoPresti EF, Simpson RC. (2006). Measurement validity for Compass assessment software. *Proceedings of RESNA 2006 Annual Conference*. Arlington, VA: RESNA Press. Available at http://kpronline.com/pubs.php.

[10] Koester HH, Simpson RC, Spaeth D, LoPresti EF. (2007). Reliability and validity of Compass software for access assessment. *Proceedings of RESNA 2007 Annual Conference*. Arlington, VA: RESNA Press. Available at http://kpronline.com/pubs.php

[11] Ashlock G, Koester HH, LoPresti E, McMillan W, Simpson R. (2003). User-centered design of software for assessing computer usage skills. *Proceedings of RESNA 2003 Annual Conference*, Atlanta, GA. Arlington, VA: RESNA Press. Available at http://kpronline.com/pubs.php.