METHODOLOGICAL AND STATISTICAL COMPONENTS OF TECHNOLOGY-BASED INTERVENTION STUDIES TO IMPROVE READING SKILLS: A SYSTEMATIZED LITERATURE REVIEW

COMPONENTES METODOLÓGICOS Y ESTADÍSTICOS DE ESTUDIOS SOBRE INTERVENCIONES TECNOLÓGICAS PARA MEJORAR LA LECTURA: REVISIÓN DE LITERATURA SISTEMATIZADA

Recibido: 11 de abril de 2023 | Aceptado: 2 de mayo de 2024
DOI: https://doi.org/10.55611/reps.3501.10

Mario E. **Bermonti-Pérez** [1], Giselle **Cordero-Arroyo** [1], Christian E. **De León Casillas** [1], Vicmarie **Vargas Álvarez**
[1.] Ponce Health Sciences University, Ponce, Puerto Rico

ABSTRACT

In this review, we aimed to examine the methodological and statistical components of studies on the effects of technology-based interventions on reading skills. We conducted a systematized review and analyzed 21 studies to examine these components. The findings suggest that studies addressing the efficacy of technology-based reading interventions employ moderately strong methodological and statistical techniques, although improvements are necessary to determine their effect more clearly. Future studies should focus on employing random assignment of participants, counterbalancing testing procedures, and incorporating statistical methods that estimate unbiased effect sizes and uncertainty in the findings. Technology is flexible, powerful, and highly engaging for children, making it an ideal venue to explore and integrate with traditional intervention strategies.

KEYWORDS: Computerized interventions, literature review, reading, technology-based interventions.

RESUMEN

El propósito de nuestra revisión fue examinar los componentes metodológicos y estadísticos de estudios sobre los efectos de intervenciones basadas en la tecnología sobre las destrezas de lectura. Realizamos una revisión sistematizada y analizamos 21 estudios para examinar estos componentes. Nuestros resultados sugieren que los estudios que abordan la eficacia de las intervenciones de lectura basadas en la tecnología emplean técnicas metodológicas y estadísticas moderadamente rigurosas, pese a que hay necesidad de mejoras. Futuros estudios deben implementar asignación aleatoria de participantes, balancear el orden de los procedimientos de medición e incorporar métodos estadísticos que estimen el tamaño del efecto de forma no sesgada y la incertidumbre de los hallazgos. La tecnología es flexible, poderosa y altamente atractiva para los niños, lo que la convierte en una vía ideal para explorar su integración en las estrategias de intervención convencionales.

PALABRAS CLAVE: Intervenciones computarizadas, revisión de literatura, lectura, intervenciones basadas en tecnología.

Correspondencia de este artículo debe ser dirigida a Mario E. Bermonti-Pérez. E-mail: mbermonti@psm.edu

There has been an increasing use of technology-based interventions to improve the basic reading skills of struggling readers. Technology-based interventions incorporate technological devices (i.e., computers, tablets) to deliver interventions to improve basic reading skills. These technology-based interventions present a feasible alternative to traditional interventions. Nonetheless, evidence about the effectiveness of technology-based interventions remains inconclusive.

Several studies have found positive effects of technology-based interventions on reading skills. Technology-based interventions have improved several reading outcomes in children, such as reading accuracy (Kathryn-Horne, 2017), reduced reading delays, decoding, rapid naming, phonological short-term memory, and capacity of executive loaded working memory (Messer & Nash, 2018). Similarly, Schmitt and colleagues (2019) found significant effects of technology-based interventions to improve reading comprehension and accuracy in middle school students. These findings suggest a positive effect of technology-based interventions on basic reading skills.

Other studies, however, have not found a positive effect of technology-based interventions on reading skills. Slavin and colleagues (2011) argued in their review that computer-assisted instruction does not influence reading. As such, the evidence about the effectiveness of technology-based interventions for improving basic reading skills is inconsistent, which may suggest issues of reproducibility and replicability. In this review, we will examine the methodological and statistical practices of researchers as important factors that may impact reproducibility and replicability, although these issues are caused by a plethora of factors.

Fundamentally, the issues of reproducibility and replicability may be understood as unexplained variance and error inserted across different areas in a given study (TNASEM, 2019). Unexplained variance and error may be inserted through the methodological components such as non-probabilistic sampling techniques, not implementing group randomization, or "masking". Unexplained variance and error can also be inserted through the statistical analysis component, which may include uncontrolled confounding effects (Pourhoseingholi et al. 2012), misuse and misinterpretation of p-values (Nuzzo, 2014), omission of effect sizes, omission of confidence intervals (Cameron et al. 2021) and inadequate statistical tests. Taken together, these issues may aggregate in a given study and increase the probability of making type I and type II errors during inferences, which leads to erroneous conclusions. If a research field is dominated by studies influenced by unexplained variance and erroneous effects, such a field could present reproducibility and replicability issues that manifest as inconsistent findings (TNASEM, 2019).

Another phenomenon that contributes to issues of reproducibility and replicability occurs at the level of reporting (Montenegro-Montero & García-Basteiro, 2019). The lack of transparent methodological and statistical reporting in scientific journals reduces the quality of communication among researchers (Graf et al. 2007). Insufficient methodological and statistical details may lead the prospecting reproducing or replicating researcher to fill in the gaps with interpretations. These interpretations during reproducing or replicating attempts may result in a study with poor methodological fidelity. As such, good reporting practices are equally vital for promoting reproducible and replicable studies that produce consistent findings.

Examining the methodological and statistical components and reporting practices of researchers is crucial to developing adequate suggestions that improve the quality of studies in this research field. The purpose of our systematized literature review is to examine the methodological and statistical practices of researchers testing the effect of technology-based interventions on basic

reading skills. Specifically, we aim to: (a) identify common methodological practices (e.g., design, sampling procedures); (b) identify common statistical analysis practices (e.g., inferential statistics, effect sizes), and (c) evaluate the methodological and analytical practices employed in these studies. The contributions of this review are important for gaining awareness of the practices of researchers and thus provide constructive suggestions for future research.

METHOD

We conducted this review using the LR-Sys approach, which combines aspects from systematic and narrative reviews (De León-Casillas & Moreno-Torres, 2020; Snyder, 2019). Similar to systematic reviews, the LR-Sys uses a pre-established research protocol, but like narrative reviews, allows more flexibility to tailor the protocol to the requirements of the research question and synthesize a wide range of studies (De León-Casillas & Moreno-Torres, 2020). We deemed the LR-Sys approach adequate for this review because it provides sufficient flexibility to synthesize a wide range of studies while implementing rigor to minimize bias (De León-Casillas & Moreno-Torres, 2020). We considered the systematic approach inadequate because its focus on synthesizing randomized controlled trials (Higgins et al. 2008) would not have allowed a representative survey of the field and the narrative approach would bias this review to the reviewer's preselected studies (Green, 2006).

We developed a search protocol that included the eligibility criteria and search strategy following the Cochrane Collaboration (Higgins et al. 2008) and PRISMA (Institute of Medicine, 2011). To systematize the review process, our research team developed protocols to evaluate the quality of evidence and the synthesis of articles, but these protocols were not deemed sufficiently rigorous to classify both activities as systematic. Our review team consisted of two researchers and two psychology graduate students. The students were involved in the phases of identification and search of research articles, the selection of articles, and the collection of data, while researchers were involved in these phases and the synthesis of the articles as well. We standardized and documented the methods for this review to improve replicability. Our team developed protocols for searching the database providers using specific criteria, selecting articles, and analyzing the data (refer to the supplemental materials S1 for more details). We prepared this article following the PRISMA systematic review reporting standards (Liberati et al., 2009).

Eligibility Criteria

Studies had to meet the following characteristics to be included in this review: (a) examined the effects of technology-based interventions on reading skills; (b) the population of interest was children who experienced reading problems; (c) the intervention had to be administered through a computer, tablet, or other technological devices to be considered a technology-based intervention; (d) the outcomes of interest are related to basic reading skills (i.e., decoding, word identification, word reading, phonological awareness, phonemic awareness, and phonics), and (e) published in a peer-reviewed journal during the years 2009-2019.

Search Strategy

We searched for research articles electronically and did not search for printed research or directly contact authors of unpublished studies. The following electronic search strategy was copied and pasted into each database to identify the relevant literature:

(1) Elementary
(2) AND Computer* Reading Intervention
(3) AND Phonics OR (phonological awareness) OR (Word Recognition) OR Decoding OR (Phonemic Awareness).

We developed this search strategy following the guidelines of the Cochrane Collaboration (Higgins et al. 2008) and

PRISMA (Institute of Medicine, 2011). The first line of the illustrated search strategy refers to the population of interest, the second line refers to the topic of this review, and the third line is related to basic reading skills, which are the outcomes of interest. After copying and pasting this search strategy, the delimiters *apply related words* and *apply equivalent subjects* were also used when the options were available.

Study Selection

The procedures conducted to search, identify, and select the studies were the following: (1) the search strategy and their delimiters were applied in each database; (2) information regarding the date of the review and the initial results were recorded; (3) titles and abstracts were screened for eligibility; (4) the full-text of eligible articles were downloaded and saved in a master folder, and (5) four reviewers screened full-texts for eligibility. All the reviewers had to agree on the eligibility of an article for it to be included in the final analysis. Whenever there was disagreement between the reviewers about the eligibility of an article, reviewers had an informal discussion of the inconsistencies and reached a consensus among all reviewers.

The criteria for the eligibility of the articles were: (a) the article must describe an empirical and quantitative study, (b) the intervention must be administered through a computer, tablet, or another technological device, and (c) the intervention must address one or more of the basic reading skills.

The screening process was as follows: (1) 451 studies were identified during the initial search; (2) 29 studies remained after the initial screening process because many studies did not meet eligibility criteria due to these consisting of qualitative studies, interventions were not technology-based or not focused on reading skills (i.e., decoding, word identification, word reading, phonological awareness, phonemic awareness, and phonics) or participants suffered from severe neurodevelopmental disorders such as autism spectrum disorder; (3) 25 studies remained after duplicates were removed, and (4) 21 remained after the final selection by judges because some studies did not address basic reading skills and the same study was published with different titles and variations of their analytical approach (only one of the studies was included in this review). Table 1 presents the general characteristics of the studies we selected for this review.

TABLE 1.
Characteristics of the Studies.

Author	Purpose	Intervention name	Targeted reading skills
(Cazzell et al., 2016)	Evaluate the effects of a computer-based flashcard reading program with self-determined response intervals on sight-word acquisition in elementary-school children with intellectual disabilities.	Researcher-developed: Computer-based flashcard reading program	Word reading
(Chai, 2017)	Evaluate the effectiveness of using an iPad intervention to improve phonological awareness skills of young children with mild developmental delays in a rural elementary school.	Touch Sound	Phonological skills
(Comaskey et al., 2009)	Explore the effectiveness of a web-based literacy programme that delivered two distinct phonics' programmes.	A Balanced Reading Approach for Canadians Designed to Achieve Best Results for All	Phonological skills, reading skills (not specified)
(Ecalle et al., 2009)	Examine the effects of a computer-assisted learning program in which syllabic units were highlighted inside words in comparison with a program in which the words were not segmented.	Researcher-developed (name not provided)	Phonological skills, word reading

Author	Purpose	Intervention name	Targeted reading skills
(Ecalle et al., 2013)[a]	Compare the effectiveness of Grapho-syllabic training, Grapho-phonemic training, and a control group in French second-grade poor readers.	Grapho-syllabic training, Grapho-phonemic training	Word reading
(Ecalle et al., 2013)	Examine the long-term effects of Grapho-syllabic training with first grade French children.	Grapho-syllabic training, Grapho-phonemic training	Word reading, reading comprehension
(Fan et al., 2018)	Determine the efficacy of Phonoblocks for improving word reading skills and spelling accuracy among Mandarin-speaking English language learners.	Phonoblocks	Reading skills (not specified), spelling
(Gustafson et al., 2011)	Compare the efficacy of interventions focused on bottom-up processing, focused on top-down processing, and a traditional comprehension training on phonological abilities and word decoding skills.	COMPHOT, Omega-Interactive Sentences	Reading comprehension, word reading, phonological skills
(Karemaker et al., 2010)	Investigate if the whole-word multimedia software 'ORT for Clicker' facilitates developing literacy skills of first grade struggling readers.	Oxford Reading Tree for Clicker	Phonological skills
(Kleinsz et al., 2017)	Investigate the effects of two types of reading training administered in parallel to different subgroups of poor readers.	Grapho-syllabic training, Comprehension training	Word reading, reading comprehension, phonological skills, vocabulary
(Kyle et al., 2013)	Assess the efficacy of Graphogame as a supplementary Computer-Assisted Reading Instruction for students learning to read in English.	Graphogame	Vocabulary, word reading, spelling, phonological skills
(Messer & Nash, 2018)	Determine whether the use of a computer-assisted intervention that uses visual mnemonics as part of the tutorial process helps the development of reading abilities.	Trainertext	Phonological skills, rapid automatized naming, spelling
(Moser et al., 2017)	Examine the effectiveness of word structure practice using application software with fourth grade readers.	8 great word patters	Reading fluency, vocabulary, reading comprehension, word reading
(O'Callaghan et al., 2016)	Evaluate the effectiveness of the Lexia Reading Core 5 intervention with four- to six-year-old children in Northern Ireland.	Lexia Reading Core 5	Phonological skills
(Pindiprolu & Forbush, 2009)	Evaluate the effects of parent implemented Funnix and Headsprout reading programs on the acquisition of basic early literacy skills of students with reading difficulties.	Funnix, Headsprout	Word reading, reading fluency, reading comprehension, phonological skills, phonics, vocabulary
(Potocki et al., 2015)	Determine the effects of a computerized training program on the reading skills of normal readers, poor decoders, poor comprehenders, and general poor readers.	Chassymo, Locotex	Word reading, reading fluency, reading comprehension
(Rosas et al., 2017)	Evaluate the impact of an explicit, sustained, and direct intervention of the phonic aspects of reading in Chilean children enrolled in their first year of primary education from a low socioeconomic status and at risk of manifesting reading difficulties.	Graphogame	Word reading, phonological skills, phonics, rapid automatized naming
(Saine et al., 2010)	Compare the effectiveness of a remedial reading intervention, computer assisted remedial reading intervention, and mainstream instruction in children with different profiles of compromised pre-reading skills before school age.	Graphogame	Word reading

Author	Purpose	Intervention name	Targeted reading skills
(Schmitt et al., 2019)	Determine the effectiveness of a web-based game played at home on literacy development among low- and middle-socioeconomic status preschool and kindergarten students.	PBS KIDS Island	Phonics, phonological skills, word reading, vocabulary
(Solheim et al., 2018)	Investigate the efficacy of an early reading intervention delivered alongside formal reading instruction to Norwegian 6-year-old children at risk for reading difficulties with a two-year follow-up.	Graphogame, On track ABC	Phonics, phonological skills, word reading, rapid automatized naming, vocabulary
(Wood et al., 2013)	Evaluate the effects of a supplemental phonemic instruction program using computer-assisted reciprocal peer tutoring with embedded audio prompting.	Researcher-developed (name not provided)	Phonological skills

Note. [a]Both are reported in the same article but as different studies.

Data Collection Process

We developed an Excel Sheet with qualitative codes which was used by all reviewers to extract data from the articles. Two reviewers double-checked the data extracted to ensure its reliability. Disagreements were discussed informally and resolved when consensus was reached among all reviewers. Please refer to the supplemental materials S2 to access the raw data.

Data Items

We utilized a codebook with a priori qualitative codes to extract the data from the studies (refer to the supplemental materials S3 for more details). This codebook had two main categories: methodological component and statistical component, which referred to the objectives of this review. The methodological component category had five codes: study design, participants, intervention, instruments, and controlled variables. These helped determine the study's quality because they aid in evaluating how well error was minimized and how robust the methods employed were. The statistical component category had two subcategories: descriptive statistics and inferential statistics. The descriptive statistics subcategory had six codes: descriptive statistics, average/mean, standard deviation, mean confidence interval, others, and additional comments. The inferential statistics subcategory also had six codes: inferential statistics, covariates, p-value, effect size, confidence interval, and additional comments. The statistical codes selected helped readers evaluate the study results by showing how robust these were.

Synthesis of Results

The synthesis of results consisted of content analysis based on code frequency. We analyzed codes on the Excel sheet used for the data extraction process using the pandas' package (Reback et al., 2021) from the Python Programming Language. Frequency analysis is a useful technique for organizing data, identifying patterns, and providing a general description of commonality amongst individuals.

RESULTS

Methodological Component

Design

The design codes include the study designs' names, as reported by the authors. In cases where the study designs were omitted from the article, we inferred them from general methodological descriptions. The pretest-posttest design was the preferred approach of researchers. First, most studies used pretest-posttest designs to compare multiple experimental groups exposed to different technology-based interventions to improve basic reading skills (33%). Second, 23% of studies used pretest-posttest designs to compare technology-based interventions with single experimental and control groups. Thirdly, 19%

of studies used pretest-posttest designs with randomized controlled trials. Finally, 19% of studies did not report their designs. Their designs were inferred from general methodological descriptions (e.g., random assignment, administration of pretest and posttest).

Participants

The participants' codes include age, grade, primary language, sampling techniques, and sample sizes. Most studies examined technology-based interventions with children between four to eight years (67%). Similarly, most interventions were tested on younger children from kindergarten to fourth grade (67%). Most studies focused on the reading skills of English and French native speakers, which are considered opaque orthographies because the same sound (i.e., phoneme) can be associated with different visual representations (i.e., graphemes; Aro, 2013).

Only 10% of studies used probabilistic sampling techniques. The median number of participants across studies was 31 but varied greatly (minimum = 2, maximum = 744). Researchers showed a tendency across studies to omit the size of the groups (45%). Nonetheless, when reported, small groups were preferred (2 – 3 participants; 36%).

Intervention Implementation

The intervention implementation code refers to the name of the intervention, number of sessions, duration of sessions, frequency of sessions, modality, and whether they were supervised. A wide range of interventions were explored in the studies, but GraphoGame and its variants were the most common technology-based intervention (33%). Many studies did not report the number of intervention sessions (38%). There was significant variability among the studies that reported the number of sessions (i.e., less than 20 to more than 40 sessions). Furthermore, most studies reported the duration of sessions to be between 10 and 30 minutes (76%). Half of the studies reported a frequency of four weekly sessions (50%). The interventions were administered in individual and group modalities (individual = 52%, group = 48%). Finally, most interventions were administered in controlled settings (e.g., schools) under supervision (90%), but this supervision was mainly provided for technical difficulties.

Variable Control

This code includes reported details about procedures to increase variable control, specifically in variance and error, such as random assignment, group balancing, and counterbalancing test administration. Most studies used randomized assignment to allocate participants to each group in the study (67%). Almost half of the studies (47%) balanced the different groups based on essential characteristics. However, few studies counterbalanced the order of test administration (17%).

Statistical Component

Descriptive Statistics

This code captures the practices for reporting descriptive statistics of the main outcomes, central tendency measures, and other descriptive statistics. Most studies reported descriptive statistics for the main outcomes (90%). The common central tendency measures were the mean/median (90%) and standard deviation (75%). Other reported descriptive statistics were the total scores (33%) and accuracies (50%). Please refer to the supplemental materials for more details (S6).

Inferential Statistics

This code captures the practices for reporting inferential statistics and includes inferential analysis, adjusted variables, p-values, and effect sizes. The analysis of covariance (ANCOVA; 39%) and analysis of variance (ANOVA; 33%) were the two most common choices among studies for examining the impact of interventions. Baseline reading skills were the most common controlled variable across studies that controlled any variable (39%), although most

studies did not control other variables (44%). Only one study-controlled variables other than cognitive ability or academic skills (e.g., age, income, parent's educational level). Most studies reported the specific p-value for the conducted analysis (78%), particularly when statistical significance was not reached. Most studies provided the effect sizes, and Cohen's d was the most common (56%), but only a few studies reported confidence intervals for the means (12%) and the mean difference (14%).

DISCUSSION

Technology-based interventions for the basic reading skills of struggling readers have been increasingly proposed as complementary tools to traditional interventions (i.e., pencil-and-paper). However, the effectiveness of technology-based interventions has not been established due to inconsistent findings. Underlying methodological and statistical practices may be contributing to these inconsistencies. The purpose of our LR-Sys was to examine the methodological and statistical practices of researchers examining the effect of technology-based interventions on basic reading skills. Our review reveals strengths and weaknesses in these studies' methodological and statistical practices.

Methodological Component

Regarding the methodological component, researchers generally implemented adequate pretest-posttest designs with randomized group allocation to examine the effects of interventions. The pretest-posttest designs (Dimitrov & Rumrill, 2003) and the randomized group allocation (Suresh, 2011) are recommended practices for testing intervention effects. Pretest-posttest designs allow to contrast two time-point measures (e.g., before and after the intervention) of outcomes of interest to infer that the intervention was responsible for the change. A pretest-posttest design implemented with the appropriate control of confounding effects provides the methodological rigor required to produce reliable cause-effect findings.

Randomized group allocation is vital to balance groups according to individual variables with potential confounding effects (i.e., age, socioeconomic status, prior reading skills, language, IQ) and reduces confounding effects. Confounding effects are caused by other variables (i.e., observed, or unobserved) that obscure the direct relationship between the intervention and the outcome of interest (Pourhoseingholi et al., 2012) In this sense, confounding effects are other sources of influence that may explain the observed effect and thus weaken the claim that a specific intervention has an effect on a given outcome. As such, confounding variables must be controlled to observe the effect of an intervention accurately.

An alarming number of studies did not explicitly report their designs. Omitting the study design from the journals is a reporting practice that diminishes the transparency required for adequate reproducible and replicable studies in a given field (Graf et al. 2007; Montenegro-Montero & García-Basteiro, 2019). Omitted or unclear methodological details, such as the study design, may lead readers to infer the procedures from other general descriptions incorrectly, which may in turn lead to issues when attempting to reproduce and replicate studies (TNASEM, 2019). Transparent reporting is paramount for improving reproducibility and replicability in a given field and thus obtaining consistent results. The field testing of the effectiveness of technology-based interventions to improve basic reading skills should aim toward a higher level of transparent reporting to increase the likelihood of reproducibility and replicability.

A source of confounding found in most studies were carryover effects. A carryover effect occurs when the performance on a task is impacted by the performance on another task. This effect leads to patterns of results that may impact the results of the study, such as participants always being tired on the last task completed or participants learning from previous tasks and having improved performance in subsequent tasks that are

similar. This potential source of confounding can be easily addressed by systematically varying the order of conditions in a study to control bias related to the order of conditions (Allen, 2017). In this regard, the observed effect across the reviewed studies may not be caused solely by the intervention but could be influenced by the order of tasks. Carryover effects further contribute to reproducibility and replicability issues.

Most studies did not employ probabilistic sampling techniques. Probabilistic sampling is vital to obtain a representative sample of the population of interest (Elfil & Negida 2017). The lack of probabilistic sampling among the studies reduces the potential for generalizing the findings to the population. The observed effects of the intervention may not apply to the population of struggling readers—interventions without generalizability present issues of reliability and feasibility (TNASEM, 2019). To effectively assess technology-based interventions, it is essential to implement probabilistic sampling techniques to increase generalizability to the population of struggling readers.

Although the sample sizes of the studies varied significantly, small sample sizes were common. Small sample sizes compromise the statistical power required to reject the null hypothesis (Norton & Strube, 2001) and do not allow discarding if non-statistically significant effects may be the results of type II errors.

A final methodological component that needs attention relates to the primary language of the studied samples. Most studies tested the technology-based interventions on English or French speakers, which are considered opaque orthographies. It is important to assess the impact of technology-based interventions on more transparent orthographies (e.g., Spanish and Italian).

Statistical Component

Researchers followed good reporting practices for descriptive statistics by the Journal Articles Reporting Standards (JARS) of the American Psychological Association (APA, 2020). The studies reported descriptive statistics for main outcomes using mean/median and standard deviation indicators. Descriptive statistics are essential because they describe the central tendency (i.e., mean/median), dispersion (i.e., standard deviation, variance, range), and shape of the distribution of the dataset (Howell, 2010). This is important to examine the normal distribution assumption (Mishra et al. 2019). Statistical tests conducted upon violated assumptions of normality may lead to errors given to wider confidence intervals. Moreover, descriptive statistics help to examine potential sampling issues. For example, extremely low mean scores could suggest that a control group may not be a typical group, being an inappropriate control group.

Researchers also showed adequate inferential statistical practices. The ANCOVA and ANOVA were the preferred statistical techniques to assess the effect of the interventions. These statistical tests are considered widely used to examine intervention effects (Kirk, 2012). Both tests compare means between two or more groups by examining the variation between samples relative to the variation within each sample in a continuous variable. An important difference between the ANOVAs and ANCOVAs is that the latter provides more reliable estimates because it allows adjusting for potential confounding effects and reduces error, making it a preferred statistical technique for testing intervention effects (Howell, 2010).

Essential for inferential statistics are the p-values and effect sizes. Researchers exhibited adequate reporting practices for p-values. Most studies reported the p-values, and most reported the exact p-values when significance was not reached. It is important to report exact p-values because this is a probabilistic indicator, not a dichotomous indicator (i.e., significant or non-significant; Loannidis, 2005; Nuzzo, 2014). The p-value indicates the probability of observing a result

as extreme as the one found in the current study if one assumes that there is no real difference, and this interpretation can be valuable to determine if a non-statistically significant result is worth following up in a future study.

Reporting effect sizes independently of significance is also crucial. While p-values inform whether a hypothesis is worth re-testing (Loannidis, 2005), the effect size indicates the magnitude of the relationship between variables (Durlak, 2009). Cohen's d was the most common index for effect sizes and while Cohen's d is widely used and provides a generally adequate effect size, it is prone to overestimating effect sizes (Hedges, 1981; Turner & Bernard, 2006), particularly with small sample sizes. Hedges' g addresses this bias and provides a better effect size index (Hedges, 1981; Turner & Bernard, 2006).

Lastly, an alarming rate of studies did not report mean differences and mean confidence intervals. Mean differences and mean confidence intervals are essential to make accurate interpretations (Howell, 2010; Turner & Bernard, 2006). These values have important implications for practical applications (Cameron et al. 2021). The mean difference is the average difference between groups and this value is helpful to provide a single descriptive indicator of group differences. However, this value should be used in conjunction with mean confidence intervals which provide a range of mean differences that contain the true mean difference and quantify the uncertainty in the estimate (Cameron et al. 2021; Turner & Bernard, 2006). For example, a study may obtain a mean difference of 15 points with a CI from 2 – 28. This suggests that the real difference may be as small as 2 points or as large as 28 points. The same principles apply to reliably interpreting effect sizes. Confidence intervals help quantify the range of uncertainty of the effect size (Turner & Bernard, 2006). Studies should report mean differences, mean confidence intervals and effect sizes confidence intervals to provide reliable values assumed to reflect true effects.

The studies presented several strengths, but the field could benefit from several suggestions to increase methodological and statistical practices. Regarding research methodology, more studies should implement pretest-posttest designs with randomized group allocation. Additionally, studies should implement probabilistic sampling techniques, control confounding variables, and counter-balance the order of test administration. Studies should aim for larger sample sizes to achieve the statistical power required to reject the null hypothesis if there is a true effect and reduce the likelihood of type II error inferences. Studies should also test technology-based interventions with transparent orthographic languages such as Spanish and Italian. Finally, studies should aim for maximum clarity and transparency when describing their methodological procedures in their scientific reports (APA, 2020; Graf et al. 2007; Montenegro-Montero & García-Basteiro, 2019; TNASEM, 2019).

Regarding the statistical component, studies should implement statistical tests that allow for confounding control, such as ANCOVAs and multiple regression models. Studies should interpret mean differences, mean confidence intervals, and effect size confidence intervals in addition to p-values and effect sizes. Confidence intervals are crucial for accurate interpretations and translating statistical findings to real-world scenarios. Finally, studies should use an unbiased effect size index (e.g., Hedges' g).

The results of our review help to reveal the research practices that may be influencing the issue of inconsistent findings among the studies examining the effects of technology-based interventions for basic reading skills in struggling readers. Although the general scientific community is aware of the relevance of collective research practices for developing interventions, few studies are exploring the research practices in a given field. To our

knowledge, this is the only review exploring such methodological and statistical practices to offer suggestions that can improve collective research practices and move toward higher reproducibility and replicability.

Limitations

We did not assess the risk of bias of individual studies because the objectives of this review were to identify the tendencies of the researchers, not the outcome results. RL-Sys are suitable when the research question is aimed at exposing information for which there is little knowledge or exposing general information about a topic (De León-Casillas & Moreno-Torres, 2020). Research questions that focus on theoretical issues usually have minor immediate social implications, which leads to a lower degree of systematization in the literature review (De León-Casillas y Moreno-Torres, 2020).

De León-Casillas and colleagues (2020) suggest that there should be experts in the research topic of the LR-Sys and research methods among the reviewers. There were only two experts on the research topic in this LR-Sys and these were two of the main authors, which may increase the risk of bias. However, the databases were randomly distributed among the reviewers, decreasing this risk (De León-Casillas et al., 2020).

Other important limitations are related to a level of standardization that is less than ideal. For example, the procedure for resolving discrepancies between reviewers was an informal discussion and informal vote counting, and not based on the internal validity of the review. The team did not establish a general protocol with the procedures for each of the phases that make up an LR-Sys. Instead, we only developed protocols for the phases of searching the database providers using specific criteria, selecting articles, and analyzing the data. Finally, the search for research articles was done electronically and no manual or hand search was conducted, contrary to what is recommended by the PRISMA (Institute of Medicine, 2011) and the Cochrane Collaboration guidelines (Higgins et al. 2008).

Conclusion

The findings of our study suggest that the current evidence on the effectiveness of technology-based interventions to improve reading skills is moderate. Future studies should use random assignment, counterbalance testing procedures, and consider other important variables that may be obscuring the findings, such as baseline levels of reading skills and sociodemographic characteristics of the participants to develop more compelling evidence on the effect of these interventions. These changes should be accompanied by estimates of effect size that are unbiased and confidence intervals to understand the magnitude of the effects and the uncertainty in the estimates. We recommend that future studies adopt methodological, analytical, and reporting standards similar to those employed by Comaskey et al. (2009). Their study comprehensively reported on all the important components discussed in this paper and adhered to best practices. Technology is flexible, powerful, and highly engaging for children making it an ideal venue to explore.

Research Ethical Standards

Funding: This study did not receive any funds.

Declaration of Conflicting Interest: The authors declared no conflicts of interest.

Approval from the Institutional Review Board for Human Research: No human subjects participated in this research. Therefore, the study did not need approval from an Institutional Review Board.

Informed Consent/Assent: N/A

REFERENCES

Allen, M. (Ed). (2017). *The SAGE encyclopedia of communication research methods.* Sage Publications.

American Psychological Association. (2020). *Publication manual of the American Psychological Association.* (7th ed.). American Psychological Association.

Aro, M. (2013). Literacy acquisition from cross-linguistic perspectives. In R. Malatesha Joshi & P. G. Aaron (Eds.), *Handbook of orthography and literacy* (pp. 531–550). Routledge. https://doi.org/10.4324/9780203824719

Cameron, C., Turner, R. & Samaranayaka, A. (2021). Understanding confidence intervals and why they are so important. *The New Zealand Medical Student Journal, 33*(1), 42–43. https://doi.org/10.57129/AGAG5939

Cazzell, S., Skinner, C. H., Ciancio, D., Aspiranti, K., Watson, T., Taylor, K., McCurdy, M., & Skinner, A. (2016). Evaluating a computer flash-card sight-word recognition intervention with self-determined response intervals in elementary students with intellectual disability. *School Psychology Quarterly, 32*(3), 367–378. https://doi.org/10.1037/spq0000172

Chai, Z. (2017). Improving early reading skills in young children through an iPad app: small-group instruction and observational learning. *Rural Special Education Quarterly, 36*(2), 101–111. https://doi.org/10.1177/8756870517712491

Comaskey, E. M., Savage, R. S., & Abrami, P. (2009). A randomized efficacy study of web-based synthetic and analytic programmes among disadvantaged urban kindergarten children. *Journal of Research in Reading, 32*(1), 92–108. https://doi.org/10.1111/j.1467-9817.2008.01383.x

De León-Casillas, C. E., Bermonti-Pérez, M., & Moreno-Torres, M. A. (2020). Guía metodológica para una revisión de literatura. *Revista Salud y Conducta Humana, 7*(1), 24-38.

De León-Casillas, C. E., & Moreno-Torres, M. A. (2020). Marco teórico para una revisión de literatura sistemática. *Revista Salud y Conducta Humana, 7*(1), 10–23.

Dimitrov, D. M., & P Rumrill, P. D. Jr. (2003). Pretest-posttest designs and measurement of change. *Work, 20*(1), 159-165.

Durlak, J. A. (2009). How to Select, Calculate, and Interpret Effect Sizes. *Journal of Pediatric Psychology, 34*(9), 917–928. https://doi.org/10.1093/jpepsy/jsp004

Ecalle, J., Kleinsz, N., & Magnan, A. (2013). Computer-assisted learning in young poor readers: The effect of grapho-syllabic training on the development of word reading and reading comprehend-sion. *Computers in Human Behavior, 29*(4), 1368–1376. https://doi.org/10.1016/j.chb.2013.01.041

Ecalle, J., Magnan, A., & Calmus, C. (2009). Lasting effects on literacy skills with a computer-assisted learning using syllabic units in low-progress readers. *Computers and Education, 52*(3), 554–561. https://doi.org/10.1016/j.compedu.2008.10.010

Elfil, M. & Negida, A. (2017). Sampling methods in clinical research; an educational review. *Emergency (Tehran, Iran), 5*(1).

Fan, M., Antle, A. N., Hoskyn, M., & Neustaedter, C. (2018). A design case study of a tangible system supporting young English language learners. *International Journal of Child-Computer Interaction, 18*, 67–78. https://doi.org/10.1016/j.ijcci.2018.08.001

Graf, C., Wager, E., Alyson, B., Fiack, S., Diane Scott-Lichter, D. & Robinson, A. (2007). Best practice guidelines on publication ethics: A publisher's perspective. *International Journal of Clinical Practice, 61*(152), 1-26. https://doi.org/10.1111/j.1742-1241.2006.01230.x

Green, B. N., Johnson, C. D. & Adams, A. (2006). Writing narrative literature reviews for peer-reviewed journals: Secrets of the trade. *Clinical Update, 5*(6), 101-117. https://doi.org/10.1016/S0899-3467(07)60142-6

Gustafson, S., Fälth, L., Svensson, I., Tjus, T., & Heimann, M. (2011). Effects of three interventions on the reading skills of children with reading disabilities in grade 2. *Journal of Learning Disabilities, 44*(2), 123–135. https://doi.org/10.1177/0022219410391187

Hedges, L. V. (1981). Distribution theory for Glass' s estimator of effect size and related estimators. *Journal of Educational Statistics, 6*(2), 107–128. https://doi.org/10.2307/1164588

Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2008). Cochrane handbook for systematic reviews of interventions. En J. P. Higgins & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions*.

Howell, D. (2010). *Statistical methods for psychology*. (8th ed). Wadsworth Cengage Learning.

Institute of Medicine. (2011). Finding what works in health care: standards for systematic reviews. National Academies Press. https://doi.org/10.17226/13059

Karemaker, A., Pitchford, N. J., & O'Malley, C. (2010). Enhanced recognition of written words and enjoyment of reading in struggling beginner readers through whole-word multimedia software. *Computers and Education, 54*(1), 199–208. https://doi.org/10.1016/j.compedu.2009.07.018

Kathryn-Horne, J. (2017). Reading Comprehension: A computerized intervention with primary-age poor readers. *Dyslexia, 23*(1), 119-140.

Kirk, R. E. (2012). *Experimental design: Procedures for the behavioral sciences* (4th ed.). Sage Publications.

Kleinsz, N., Potocki, A., Ecalle, J., & Magnan, A. (2017). Profiles of French poor readers: Underlying difficulties and effects of computerized training programs. *Learning and Individual Differences, 57*, 45–57. https://doi.org/10.1016/j.lindif.2017.05.009

Kyle, F., Kujala, J., Richardson, U., Lyytinen, H., & Goswami, U. (2013). Assessing the effectiveness of two theoretically motivated computer-assisted reading interventions in the United Kingdom: GG Rime and GG Phoneme. *Reading Research Quarterly, 48*(1), 61–76. https://doi.org/10.1002/rrq.038

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Medicine, 6*(7). https://doi.org/10.1371/journal.pmed.1000100

Loannidis, J. P. A. (2005). Why most published research findings are false. *PloS Medicine, 2*(8), 696-701. https://doi.org/10.1371/journal.pmed.0020124

Messer, D., & Nash, G. (2018). An evaluation of the effectiveness of a computer-assisted reading intervention. *Journal of Research in Reading, 41*(1), 140–158. https://doi.org/10.1111/1467-9817.12107

Mishra, P., Pandey, C. M., Singh, S., Gupta, A. Sahu, C. & Keshri, A. (2019). Descriptive statistics and normality tests for statistical data. *Ann Card Anaesth, 22*(1), 67-72. https://doi.org/10.4103/aca.ACA_157_18

Montenegro-Montero, A. & García-Basteiro, A. L. (2019). Transparency and reproducibility: A step forward. *Health Sciences Reports, 2*(3). https://doi.org/10.1002/hsr2.117

Moser, G. P., Morrison, T. G., & Wilcox, B. (2017). Supporting fourth-grade students' word identification using application software. *Reading Psychology, 38*(4), 349–368. https://doi.org/10.1080/02702711.2016.1278414

Nuzzo, R. (2014). Statistical errors. *Nature, 506*(1), 150-153.

Norton, B. J. & Strube, M. J. (2001). Understanding Statistical Power. *Journal of Orthopaedic & Sports Physical Therapy, 31*(6), 307-515. https://doi.org/10.2519/jospt.2001.31.6.307

O'Callaghan, P., McIvor, A., McVeigh, C., & Rushe, T. (2016). A randomized controlled trial of an early-intervention, computer-based literacy program to boost phonological skills in 4- to 6-year-old children. *British Journal of Educational Psychology, 86*(4), 546–558. https://doi.org/10.1111/bjep.12122

Pindiprolu, S., & Forbush, D. (2009). Evaluating the promise of computer-based reading interventions with students with reading difficulties. *Journal on School Educational Technology, 4*(3), 41–49. https://doi.org/10.26634/jsch.4.3.588

Potocki, A., Magnan, A., & Ecalle, J. (2015). Computerized trainings in four groups of struggling readers: Specific effects on word reading and comprehension. *Research in Developmental Disabilities, 45–46*, 83–92. https://doi.org/10.1016/j.ridd.2015.07.016

Pourhoseingholi, A. M., Baghestani, A. R. & Vahedi, M. (2012). How to control confounding effects by statistical analysis. *Gastroenterology and Hepatology From Bed to Bench, 5*(2), 79-83.

Reback, J., jbrockmendel, McKinney, W., Bossche, J. Van den, Augspurger, T., Cloud, P., Hawkins, S., gfyoung, Roeschke, M., Sinhrks, Klein, A., Petersen, T., Tratner, J., She, C., Ayd, W., Hoefler, P., Naveh, S., Garcia, M., Schendel, J., … Seabold, S. (2021). *pandas-dev/pandas: Pandas 1.3.3*. https://doi.org/10.5281/ZENODO.5501881

Rosas, R., Escobar, J. P., Ramírez, M. P., Meneses, A., & Guajardo, A. (2017). Impact of a computer-based intervention in Chilean children at risk of manifesting reading difficulties / Impacto de una intervención basada en ordenador en niños chilenos con riesgo de manifestar dificultades lectoras. *Infancia y Aprendizaje, 40*(1), 158–188. https://doi.org/10.1080/02103702.2016.1263451

Saine, N. L., Lerkkanen, M. K., Ahonen, T., Tolvanen, A., & Lyytinen, H. (2010). Predicting word-level reading fluency outcomes in three contrastive groups: Remedial and computer-assisted remedial reading intervention, and mainstream instruction. *Learning and Individual Differences, 20*(5), 402–414. https://doi.org/10.1016/j.lindif.2010.06.004

Schmitt, K. L., Hurwitz, L. B., Sheridan Duel, L., & Nichols Linebarger, D. L. (2018). Learning through play: The impact of web-based games on early literacy development. *Computers in Human Behavior, 81*, 378–389. https://doi.org/10.1016/j.chb.2017.12.036

Schmitt, A. J., McCallum, E., Hawkins, R. O., Stephenson, E., & Vicencio, K. (2019). The effects of two assistive technologies on reading comprehension accuracy and rate. *Assistive Technology, 31*(4), 220–230. https://doi.org/10.1080/10400435.2018.1431974

Slavin, R. E., Lake, C., Davis, S. & Madden, N. A. (2011). Effective programs for struggling readers: A best-evidence synthesis. *Educational Research Review, 6*(1), 1-26.

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research, 104*(1), 333-339.

https://doi.org/10.1016/j.jbusres.2019.07.039

Solheim, O. J., Frijters, J. C., Lundetræ, K., & Uppstad, P. H. (2018). Effectiveness of an early reading intervention in a semi-transparent orthography: A group randomised controlled trial. *Learning and Instruction, 58*, 65–79. http://doi.org/10.1016/j.learninstruc.2018.05.004

Suresh, K. P. (2011). An overview of randomization techniques: An unbiased assessment of outcome in clinical research. *Journal of Human Reproductive Sciences, 4*(1), 8-11. http://doi.org/10.4103/0974-1208.82352

The National Academies of Sciences, Engineering and Medicine. (2019). *Reproducibility and replicability in science.* The National Academies Press.

Turner, H. M. & Bernard, R. M. (2006). Calculating and synthesizing effect sizes. *Contemporary Issues in Communication Science and Disorders, 33*(1), 42-55.

Wood, C. L., Mustian, A. L., & Lo, Y. yu. (2013). Effects of supplemental computer-assisted reciprocal peer tutoring on kindergarteners' phoneme segmentation fluency. *Education and Treatment of Children, 36*(1), 33–48. https://doi.org/10.1353/etc.2013.0004