Applying Cost-Benefit Analysis to Inform the Selection of Academic Interventions

by

Andrea Antoniuk

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Education

in

School and Clinical Child Psychology

Department of Educational Psychology

University of Alberta

Abstract

More than 40% of working-age Canadians live with low literacy skills. Furthermore, there is no foreseeable decline in this striking percentage, with nearly 20% of Canadians ages 16-19 demonstrating low literacy skills. To address this literacy deficit, numerous interventions are available that address the fundamental building blocks of literacy. Interventions often vary in terms of costs (i.e., materials, training for teachers, time), and choosing the most effective interventions from the abundance of options can be a difficult algorithm to navigate. Further, choosing cost-effective interventions can benefit students while efficiently using limited resource funds. My aims in the present study were to develop a novel process for analyzing the cost-benefit of academic interventions, and to apply this process to fluency (i.e., word reading speed) interventions. The cost-benefit process compares interventions by studying the effect size benefit per unit of cost. In applying this process to fluency interventions, I reviewed 614 publications, and was able to successfully compare two interventions (i.e., QuickReads, Helping Early Literacy with Practice Strategies; HELPS). The HELPS intervention was seven times more cost-effective than QuickReads, which suggests that HELPS may increase fluency (i.e., word reading speed) gains more quickly and use fewer resources than QuickReads. The cost-benefit methodology developed in the present study provided useful comparisons and is potentially adaptable for other academic interventions.

Keywords: fluency, cost-benefit, Early Years, reading interventions, academic interventions

Preface

This thesis is an original work by Andrea Antoniuk.

Acknowledgements

I would like to thank my supervisor, Dr. Damien Cormier, for his guidance and mentorship for this thesis and throughout my studies at the University of Alberta. I would also like to thank Dr. George Georgiou for giving me opportunities to work in his lab and learn from his mentorship. I wish special thanks to my examining committee (Drs. Carbonaro, Bulut, Georgiou, and Cormier) for their time and insights into the present study. I am also grateful for Jude, Adrian, James, Hawk, Kenneth, Lorne, Paul, and our book club for their continued and incredible support. I thank my parents and brother, who never needed a thesis to be proud of me. I would also like to thank the SSHRC, the Government of Alberta, the University of Alberta, and Brandon University for their generous funding throughout my master's studies. I respectfully acknowledge that the University of Alberta is situated on Treaty 6 territory, traditional lands of First Nations and Métis peoples.

Table of Contents

Introduction 1

Method 12

Results 19

Discussion 23

References 29

Appendix A 36

Appendix B 37

Appendix C 38

List of Figures

Figure 1: Data Collected 32

List of Tables

Table 1: Excluded Publications 33

Table 2: Cost-Benefit Analysis...34

Applying Cost-Benefit Analysis to Inform the Selection of Academic Interventions

Introduction

More than 40% of working-age Canadians have low literacy skills (Statistics Canada, 2011). Furthermore, there is no foreseeable decline in this striking percentage with nearly 20% of Canadians ages 16-19 demonstrating low literacy skills (Rubenson, Desjardins, & Yoon, 2007). The negative impact of poor literacy is widespread. People with weak literacy skills are being left behind in Canada's increasingly knowledge-based job market, resulting in lower incomes and fewer job opportunities than those with strong literacy skills (Chiswick & Miller, 1990; Frank, Phythian, Walters, & Anisef, 2013; Kirsh, Jenkins, Jungeblut, & Kolstad, 1993; Rubenson et al., 2007; Smith & Fernandez, 2017). Poor literacy skills are associated with many negative health outcomes, including increased hospital service utilization, reduced compliance with medication, and early mortality (Berkman, Sheridan, Donahue, Halpern, & Crotty, 2011; DeWalt, Berkman, Sheridan, Lohr, & Pignone, 2004; Rubenson et al., 2007). In addition, people with low literacy skills are also less likely to participate in civic duties, including voting, in comparison to people with higher literacy skills (Kaplan & Venezky, 1994; Rubenson et al., 2007).

Literacy is colloquially understood as the ability to read and write. Taking a functional approach, other literacy definitions emphasize the ability to use printed and written materials to achieve goals, acquire knowledge, and participate in society (UNESCO, 2004). Using printed materials for those complex functions require higher-order skills such as interpretation, creation, communication, computation, and identification (UNESCO, 2004). Literacy-related objectives may also require computer and numeracy skills.

These definitions emphasize that literacy is not restricted to reading and writing; rather, it is a complex domain that is enmeshed with other skills. Moreover, literacy weaknesses can affect widespread achievement. It is not uncommon for students with literacy weaknesses to exhibit low achievement in additional areas (e.g., mathematics, science), as literacy skills may cause or exacerbate academic difficulties.

Academic Learning

Academic skill development. Haring and Eaton's (1978) instructional hierarchy describes learning as a sequential progression through three stages: acquisition, fluency building, generalization and adaptation. The authors argued that this sequence applies to all academic domains (i.e., reading, writing, spelling, mathematics). In the acquisition stage, the student works to understand the lesson or skill being taught, and their responses to questions and tasks may be slow and/or inaccurate (Haring & Eaton, 1978). In the fluency building stage, the student's responding increases in speed (Haring & Eaton, 1978). Haring and Eaton's (1978) final stage is generalization and adaptation, where the student develops their ability to apply and flex the lesson or skill to novel settings.

Haring and Eaton's (1978) instructional hierarchy remains current in educational psychology, with citations extending into 2019 (e.g., Erion & Hardy, 2019). This hierarchy is theoretically and practically useful; for instance, interventionists such as Burns, VanDerHeyden, and Riley-Tillman (2012) included Haring and Eaton's hierarchy as part of evidence-based intervention practices. The authors argued that targeting interventions to hierarchy stages (i.e., acquisition, fluency, generalization and adaptation) increases the likeliness of intervention success and promotes academic development (Burns et al., 2012).

Reading skills. It is well-known that the fundamental building blocks of reading are phonemic awareness, phonics, vocabulary, fluency, and comprehension (e.g., Kelly, 2011). Theorists posit that reading skills are built by first developing accuracy and then moving to automaticity (Biancarosa & Shanley, 2016; LaBerge & Samuels, 1974). In the accuracy stage, attention is focused on decoding—converting visual representations of the letters and words on the page into phonological sounds (Biancarosa & Shanley, 2016). As accuracy increases, less attention is required, and reading becomes increasingly automatic (Biancarosa & Shanley, 2016; LaBerge & Samuels, 1974; Perfetti, 1977, 1985). Attention is freed for other purposes, including building vocabulary and comprehending text (Biancarosa & Shanley, 2016, LaBerge & Samuels, 1974).

Reading fluency, the ability to read texts with grade-level accuracy, speed, and prosody, also facilitates reading comprehension (Biancarosa & Shanley, 2016; Hudson, Lane, & Pullen, 2005; Kuhn & Stahl, 2000, Stanovich, 1991). Not only does reading accuracy inform understanding, but reading at a proficient rate suggests decoding automaticity, and frees up attentional resources necessary for comprehension and interpretation (Biancarosa & Shanley, 2016; Kuhn & Stahl, 2000; LaBerge & Samuels, 1974).

Dual-route model (DRM). According to the DRM, verbal word reading develops along lexical and non-lexical routes (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Law & Cupples, 2017; Marshall and Newcombe, 1973). Speech production along the lexical route is a word recognition process (Coltheart et al., 2001; Law & Cupples, 2017; Marshall and Newcombe, 1973). This route involves an orthographic lexicon (i.e., knowledge of spelling conventions), phonological lexicon (i.e., grapheme-phoneme correspondence), and semantic knowledge (Castles, Bates, Coltheart, Luciano, & Martin, 2006; Law & Cupples, 2017). Interactions between route nodes enable students to encode words into memory and recall or recognize those previously-encoded words (Coltheart et al., 2001). When a reader views an unfamiliar word, the non-lexical route is engaged as the reader attempts to sound out the parts of the words (Coltheart et al., 2001). In doing so, the reader uses their knowledge of grapheme-phoneme correspondences to decode the word (Coltheart et al., 2001).

Effective use of the lexical and non-lexical routes is associated with proficient reading (Coltheart et al., 2001; Law & Cupples, 2017; Marshall and Newcombe, 1973). The lexical route yields faster speech production than the non-lexical route (Law & Cupples, 2017). This is because word recognition is a parallel process; in contrast, sounding parts of words is sequential (Law & Cupples, 2017). The lexical route also facilitates decoding among irregular words such as *yacht* (Law & Cupples, 2017). Readers who develop both routes are able to switch between them as needed to accomplish literacy tasks (Law & Cupples, 2017). Impairment in either or both routes is often associated with reading disabilities such as dyslexia (Coltheart et al., 2001; Law & Cupples, 2017; Marshall and Newcombe, 1973).

Connectionist dual process model (CDP+). Critics argued that the DRM inadequately explains word reading for the non-lexical route (Law & Cupples, 2017; Perry, Ziegler, & Zorzi, 2007). As a result, researchers proposed the CDP+ (Law & Cupples, 2017; Perry et al., 2007; Zorzi, 2010). In this model, the lexical route remains the same as in the DRM, but the non-lexical route supplanted the notion of explicit grapheme-phoneme rules with probabilistic pattern identification (Law & Cupples, 2017; Perry et al., 2007; Perry, Ziegler, Braun, & Zorzi, 2010). Researchers contended that repeated experience with letter-sound connections cannot feasibly produce memory of complex rules; rather, that repeated experience creates a statistical distribution of possible values (Perry et al., 2007; 2010).

Reading development and differences. Reading involves interactions between domains (Biancarosa & Shanley, 2016; LaBerge & Samuels, 1974). Similar to Haring and Eaton (1978), Chall, Jacobs, and Baldwin (1990) emphasized that reading develops in hierarchical progressions, and underdeveloped early stages can impair development in further stages. Students learn words (i.e., semantics, pronunciation, spelling) at vastly different rates; for instance, Lyon (1997) argued that students require 4-14 exposures before they can automatically recognize a word. A student who is struggling to decode individual words is likely going to also struggle with understanding sentences and interpreting text (Biancarosa & Shanley, 2016; LaBerge & Samuels, 1974). These challenges often cumulate–creating obstacles to content mastery and completing academic tasks within manageable timeframes. When a student is learning to read, it is difficult to read to learn (Chall et al., 1990; Kuhn & Stahl, 2000).

Academic Interventions

Academic interventions are educational strategies and programs that address academic difficulties (Burns et al., 2012). Academic interventions typically involve several steps: problem identification, intervention selection, intervention implementation, and progress monitoring (Burns et al., 2012).

Academic difficulties tend to be identified by educators when a student is displaying lower academic achievement in comparison to their same-grade peers. Educators may further investigate with their own assessments or refer the student to other specialists (e.g., reading specialists, learning support specialists, school psychologists), depending on available school resources and the student's presenting concerns.

Once the academic difficulty or difficulties are investigated, an intervention based on the student's identified learning needs is selected and implemented. The student's progress is monitored so that the intervention can be adjusted as needed and discontinued when the academic difficulty is resolved. The use of academic interventions in schools has been influenced by several trends; namely, evidence-based practice, early intervention, and Response to Intervention.

Evidence-based practice. Adopted from medicine, evidence-based practice (EBP) is the use of empirical research findings to inform one's decision-making and clinical actions, rather than relying on one's own experience (APA Presidential Task Force on Evidence-Based

Practice, 2006). For academic interventions, evidence-based measures are used to identify academic difficulties, and academic interventions are evaluated in terms of previous research reporting their efficacy (APA Presidential Task Force on Evidence-Based Practice, 2006). As such, EBP enables clinicians to use assessments and interventions that have the greatest probability of positive effect and lowest probability of harm (APA Presidential Task Force on Evidence-Based Practice, 2006; Burns et al., 2012). This is important given the widespread negative impact of academic difficulties among students. Data-driven approaches increase the likeliness that academic difficulties will be correctly assessed and addressed; they incorporate the results of other clinicians in larger populations, which reduces the risk that decisions may be made based on the potential confines of a clinician's individual experience.

Early intervention. Although reading interventions are appropriate for all ages, it is well-known that helping struggling readers is most effective at early ages when they are building foundational reading skills (e.g., Foorman, Francis, S. Shaywitz, B. Shaywitz, & Fletcher, 1997). Reading interventions that are implemented before a student completes Grade 3 are more effective than beyond. Furthermore, the positive outcomes associated with early intervention tend to decrease the likelihood that they will drop out of high school (McIntosh, Chard, Boland & Horner, 2006; Good, Simmons, Kame'enui, 2011; Walker & Sprague, 1999).

Response to Intervention. Response to Intervention (RTI) is a framework developed to use EBP to provide high quality direct instruction, identify academic difficulties, and support academic difficulties through intervention and/or accommodation (Burns et al., 2012). As described by Burns and colleagues (2012), teachers use class-wide curriculum-based assessments to identify academic difficulties among students as early as possible. According to the RTI framework, students who display academic weaknesses receive small-group instruction targeted at those specific weaknesses (Burns et al., 2012). Students receiving the small group intervention are tested after a specific time interval (i.e., the end of the small group intervention or a planned amount of intervention time), and students whose scores suggest that the intervention was effective (i.e., they no longer exhibit that weakness) are able to benefit from regular classroom instruction (Burns et al., 2012). On the other hand, students whose scores suggest poor response to the intervention (i.e., persisting academic weakness) will continue to receive individual intervention in that area (Burns et al., 2012). As stated by Burns and colleagues (2012), students whose scores suggest persisting academic weakness are referred for psychoeducational assessment to diagnose learning disability and plan interventions and accommodations.

Academic interventions that are informed by evidence, implemented early, and occur within an RTI framework provide opportunities for students to receive timely assessment and effective interventions. Taken together, these increase the likeliness that students will thrive in educational settings, rather than experience persisting struggles that may compound to greater concerns (e.g., early exit). To this end, we may progress towards reducing Canada's persisting literacy deficit, and develop effective approaches to addressing academic difficulties for all.

Targeted Fluency Instruction and Monitoring

Fluency Interventions. Targeted fluency interventions fall into three non-mutually exclusive categories–assisted reading, repeated reading, and performance reading (Padeliadu & Giazitzidou, 2018). According to Padeliadu and Giazitzidou (2018), repeated reading interventions fall into five sub-categories: a) independent repeated reading; b) assisted repeated reading; c) cues (i.e., goal-setting, self-monitoring); d) repeated reading with pre-teaching key concepts; e) performance reading.

Evidence-based recommendations. In Stern and Piper's (2018) international review of primary education interventions, fluency interventions were associated with a mean effect size of .44. Stern and Piper's (2018) result suggests that fluency interventions, in general, are effective in building fluency. Hudson and colleagues' (2005) review of intervention studies found that for all readers (i.e., average, reading difficulty, LD), fluency is effectively developed when teachers model fluent oral reading, provide direct instruction and corrective feedback, and provide opportunities to repeatedly read text. The researchers also recommended that students be given level-appropriate texts to read independently (Hudson et al., 2005).

Chard, Vaughn, and Tyler's (2002) synthesis of fluency interventions for elementary students with learning disabilities (LD) investigated features of successful fluency interventions. Chard and colleagues (2002) found that repeated reading-based interventions were generally effective, and that repeated reading with a model was more effective than without a model. The effectiveness of repeated reading interventions was enhanced with additional elements; specifically, when interventionists provided corrective feedback and advanced students using pre-established performance criteria (Chard et al., 2002). Peer tutoring was associated with uncertain effects on reading fluency, as results were conflicted (Chard et al., 2002).

More current syntheses agree with these erstwhile findings. Stevens, Walker, and Vaughn's (2017) synthesis indicated that repeated readings-based interventions are effective in building fluency among elementary students with LD. Padeliadu and Giazitzidou (2018) found that repeated readings interventions are effective among readers of various abilities, and that effectiveness is bolstered with the addition of strategies (i.e., self-monitoring, goal-setting, and model reading).

Assessment and progress monitoring. Fluency is often measured in terms of oral word reading speed (Hudson et al., 2005). Word reading speed typically includes the number of correctly-read words per minute or the time a reader takes to read a complete passage of text (Hudson et al., 2005). To measure fluency progress, interventionists complete a baseline assessment with the student and then monitor progress at regular intervals (e.g., Hudson et al., 2005). Hudson and colleagues (2005) indicated that a number of fluency assessments are available, including AIMSweb Standard Reading Assessment Passages (RAPs), Dynamic Indicators of Basic Early Literacy Skills (DIBELS), Gray Oral Reading Test, Fourth Edition (GORT-4). As a psychologist-in-training, I have measured fluency using several of the assessments mentioned above, as well as measures in the Test of Word Reading Efficiency-Second Edition, Woodcock-Johnson IV Tests of Achievement and the Wechsler Individual Achievement Test – Third Edition.

Intervention Selection

Decision makers. Intervention selection often depends on school policies and available resources. Teachers, principals, learning support coordinators, school psychologists, and other professionals may be involved in deciding which interventions to purchase and when to use specific interventions with students. Many schools take a team approach to these decisions; however, some schools lack the resources to do so, and teachers may make these decisions without the ability to consult with other specialists (e.g., school psychologists).

Rationales. Interventions are matched to students based on presenting concerns. When similar interventions are available for the same presenting concern, the final selection is made for a number of different reasons. Many practitioners choose academic interventions because of familiarity (i.e., either through prior use or a positive word-of-mouth reputation), or perceptions that a particular intervention will be easier or faster to learn how to implement (Burns et al., 2012). Interventions are sometimes chosen because they are new, appear to be more up-to-date and efficacious than other interventions, or are associated with new technology and appear appealing to students. An EBP approach matches students with interventions that have the strongest evidence supporting their effectiveness (APA Presidential Task Force on Evidence-Based Practice, 2006). However, EBP is not without its challenges, as practitioners may struggle to remain knowledgeable about the evidentiary status of interventions (e.g., when new supporting or refuting evidence is published), and rely on old information.

Abundance of choice. School principals, teachers, and reading specialists are often presented with a number of academic interventions to choose from, and left to navigate this decision on their own. For instance, one interventionist website stated that more than seven evidence-based reading interventions are available for each of the five major domains in reading (e.g., phonemic awareness, phonics, fluency, comprehension, and vocabulary), leaving decision-makers to choose between 40 interventions (Kelly, 2011). Academic interventions often vary in cost, content, training for teachers, time required, and student-to-teacher ratio. As a result, administrators are presented with a challenge—to juggle all of the attributes of each intervention and use limited funds to select a few programs, or perhaps a single program, that they hope will have the greatest influence on educational outcomes. Unfortunately, this complex algorithm seems to be very difficult to navigate effectively, which may leave educational professionals without a straightforward way of selecting the best ways to address the needs of their struggling learners.

Opportunities. Choosing the most effective interventions for students increases the probability that positive outcomes will occur. Time-effective interventions can be particularly helpful for students who may have lower school attendance, or who may not receive long-term interventions (i.e., without long-term funding, changing schools). Benefits extend to teachers and educational assistants: maximizing intervention delivery times reduces hours wasted on ineffective teaching methods and opportunity costs associated with teaching other students. For psychologists who consider RTI in assessing LD, time-effective interventions result in more efficient diagnoses in comparison to unnecessarily lengthy interventions with a lower probability of success. For schools that operate with limited budgets, interventions that offer the highest benefits with limited resources can maximize an organization's effectiveness in addressing academic concerns. This may result in more students receiving effective interventions without being waitlisted or denied.

Relying on status quo approaches to student-intervention matching can certainly be effective in addressing academic difficulties. Yet, in schools where funding, time, and teaching is limited, cost-effective selections can maximize intervention efficacy at the student and organizational levels. In this sense, the gap between research and practice is feasibility. A logical next step in reducing this gap is to support decision makers–by investigating how to apply evidence-based methods in the most feasible ways.

Purpose

Previous research has studied the effectiveness of academic interventions; however, little is known about the cost per effect ratio of these programs. In the present study, I developed a process of analyzing the cost-effect of academic interventions. In addition, I applied this analytical process to Early Years fluency interventions, by comparing fluency interventions in terms of effectiveness, and when taking resources into account (i.e., cost, time, staff, and training).

Research Questions

Published literature does not provide theory or research findings to enable me to develop hypotheses. As such, two research questions will be explored in this study, as stated below:

a) Can a methodology be developed to determine the cost-effectiveness of academic interventions?

b) What is the cost-effect ratio of evidence-based fluency interventions?

Method

Procedure

Cost-benefit methodology. I developed a process for comparing the costs and benefits of academic interventions. This process was informed by Siegel, Laska & Meisner's (1996) cost-benefit analysis for medical interventions, and loosely informed by the general notion of return on investment. Costs refer the intervention's expected costs, while benefits indicate the effects of an intervention (i.e., treatment) in comparison to a baseline or control group (Briggs et al., 2002; Siegel et al., 1996). Expected costs and benefits are compared using a ratio of expected costs to expected effects, excluding outliers (Briggs et al., 2002; Siegel et al., 1996). Ratios are often-used measurement tools for identifying the relationship between two variables; for instance, the proportion of effect in relation to cost outlay of an intervention, or the return on an investment. Ratios enable researchers to quantitatively compare the cost-benefit of various interventions and rank interventions by their effect per unit of cost.

To determine intervention benefits, I selected meta-analyses to pool effect sizes using Cohen's *d*. Meta-analysis data are ideal for the present study, as they are often associated with systematic review procedures to select studies (Cleophas & Zwinderman, 2017; Field & Gillett, 2010). A meta-analysis enables researchers to determine effect sizes based not only on the results of a single study, but one that considers multiple studies. The benefit of this approach is determining effect sizes that are more likely to be generalizable to the general population, rather than limited to a single population or cohort (Cleophas & Zwinderman, 2017). In other words, systematically synthesizing findings from multiple sources yields a more robust effect size than using a single study (Borenstein, Hedges, Higgins, & Rothstein, 2009).

After reviewing literature on meta-analysis, I chose Basu (2017) and Field and Gillett (2010) because they offered clear guidance on meta-analytic procedures. I chose Basu's (2017) and Field and Gillett's (2010) recommended steps for meta-analysis: systematic literature search, select studies, code data, and analyze data. This sequence is well-accepted by many other experts in quantitative methods, and it offered a straightforward approach to finding effect sizes for studies implemented with rigor.

Cost, also interpreted as *investment*, is derived by adding fixed and variable costs (e.g. Krugman, 1979). Fixed costs represent singular cost outlays (e.g., computer software), while variable costs are incurred per student (e.g., workbooks, staff salaries). Total investment was calculated with the following equation:

$$Total\ Investment = Fixed\ Costs + Variable\ Costs\ (t) \quad (1)$$

The above equation describes total investment as the sum of fixed costs and variable costs. Variable costs are calculated by multiplying their rate (e.g., wage) and units (e.g., time).

I identified several steps for cost-benefit analysis: target identification, intervention selection, article selection, coding intervention and cost data, and comparisons. These steps are elucidated below.

Target identification. The first step in cost-benefit analysis was to determine what specific parameters to address; namely, what intervention to analyze and for what specific population. The objectives for a particular cost-benefit analysis will inform the parameters, for instance, to advance knowledge in a domain, compare effectiveness among specific populations, or compile a technical report for a school division.

Determining the target requires the user to identify the specific academic domain under study (e.g., fluency, spelling, computation) as larger domains (e.g., reading, writing, mathematics) would be too vague to produce informative data. Identifying the population requires some planning in terms of how the cost-benefit will be used. If, for example, the analysis is intended to compare cost-benefit for the general classroom, a grade level could be useful. If the analysis is intended to make comparisons for students who may have more individualized concerns (i.e., gender, cognitive abilities, English language learners), then the target population should be adjusted. In this step, it may be helpful to perform some preliminary database searches to pre-test the target population and adjust it if necessary. If insufficient publications are available for the analysis, then the target population might need to be adjusted by including fewer specific characteristics. A large number of articles may suggest that the target population could be more specific.

Intervention selection. To select interventions, my supervisor guided me towards popular interventionist websites to help compile a list of interventions. This ensures relevance of the cost-benefit model, as it would presumably be evaluating interventions used by educators. We also discussed using keyword searches to identify additional interventions. Basu (2017) recommends using multiple databases and keyword searches to reduce biases associated with using only one. After compiling a list of potential interventions, the next step was to evaluate the interventions in terms of inclusion criteria. Included interventions would be those that were currently available for purchase and resultantly able to compare costs and benefits. I also planned to evaluate interventions based on their generalizability, as confounding factors may impact positive outcomes. For instance, teacher-centred interventions may be too dependent on the teacher.

Article selection. I chose systematic review procedures because they entail strategically searching for publications and selecting ones that meet specific criteria. This approach enabled me to code information from selected studies and pool effect sizes. I planned systematic review to occur in two rounds; first, an abstract review was planned for a time-efficient way to scan articles and eliminate articles that did not meet criteria. Articles that met criteria were recorded in a Microsoft Excel file, downloaded, and named a unique number. The unique number was also recorded in the Excel file for easy access. I visually scanned each downloaded article to verify whether it met inclusion criteria. Articles that met inclusion criteria were subsequently coded into SPSS.

Next, I developed inclusion criteria for evaluating articles. Articles would be eliminated if they were not peer-reviewed journal articles, theses, or dissertations, did not directly answer the target intervention and population noted above. I also excluded studies more than ten years old, as older studies may be subject to cohort effects such as the Flynn Effect, where some cohorts' fluency may develop at a different rate than others due to differences in intelligence (e.g. Teasdale & Owen, 2005). Participants in a more intelligent cohort may respond to intervention differently than another cohort, because higher intelligence may be associated with the ability to learn and apply information more quickly. I excluded studies that did not include quantitative fluency measures with effect sizes or statistics and did not report quantified data that can be converted into effect sizes. For instance, means, standard deviations, p-values, *F* tests, and *t* tests enable researchers to calculate effect sizes (Swanson, 1999). I excluded studies of participants who are learning English as a second or additional language, as fluency may develop differently for this group in comparison to Canadian-born readers. I also excluded studies of participants with clinical concerns and/or intellectual disabilities as fluency may develop at a different rate, which may otherwise introduce confounding elements into the present study. I excluded studies where interventions had been modified for a specific culture, as I felt that results would be difficult to generalize beyond that specific culture. Studies not written in English were also excluded, as I did not have enough time to seek a qualified translator. I also excluded intervention studies that used classroom formats, as class size may confound effect sizes with group dynamics and teacher-student ratio factors.

Code data. I considered both effect size data and cost data when planning what data to code into SPSS, and used Basu (2017) and Field and Gillett (2010) for guidance. Because I was uncertain if studies would include within groups or between groups effect sizes, I included columns for both. I also included columns to indicate whether the study was associated with a significant or non-significant treatment effect. I coded sample sizes for treatment and control groups for the meta-analysis process (Basu, 2017; Field & Gillett, 2010). I was also interested in benefits of interventions over time, so I coded the mean duration of intervention sessions, rate of sessions per week, number of intervention weeks, and planned to use these numbers to calculate intervention minutes per student. I also coded the highest level of education for interventionists. This would help me to determine the cost of delivering the intervention (i.e., teaching time and salary).

Pool effect sizes. Cohen's *d* is typically reported in meta-analyses—it represents the difference between means divided by the pooled standard deviation (Schmidt & Hunter, 2015; Swanson, 1999). Effect size measures the magnitude of treatment effects between a treatment group versus a control group, or a post-test versus a pre-test. Researchers (e.g., Field & Gillett, 2010) described various methods for pooling effect sizes, such as the Hedges and colleagues' method, and the Hunter and Schmidt method. Choosing the specific method for pooling effect sizes and determining confidence intervals seemed dependent on the nature of the data collected. As such, I opted to choose the method after examining the data.

Code cost data. I decided to use current cost information, as the intent for the cost-benefit model is to be used currently and adjusted as costs change over time. I gathered cost information from intervention developers' websites, and salary information from Canadian averages, and transferred them to a Microsoft Excel spreadsheet.

Comparisons. I chose Siegel and colleagues' (1996) process of combining costs and benefits into ratios for further comparison. For this step, Microsoft Excel would be sufficient.

See Appendix A for a visual summary of data analyses used in the proposed study.

Cost-Effectiveness of Fluency Interventions

Target identification. I considered my purposes for the present analysis; namely, to test the cost-benefit methodology, and to contribute to reading research. I initially planned to evaluate interventions for all domains of literacy, but realized quickly from pre-testing targets that it would not be a feasible methodological test given the limited timeframe of my master's degree. I narrowed the target to fluency interventions for one-on-one student support. Because reading fluency includes speed, accuracy, and prosody (i.e.., Biancarosa & Shanley, 2016), I further narrowed the target to word reading speed). In addition, I initially planned to study interventions for Grade 3 students, but the pre-test identified a lack of publications for that grade. I widened the target to Early Years students, as in my experience, reading growth among these grades is similar enough for comparisons in the present study.

Intervention selection. I created a list of potential interventions from Reading Rockets, which is an online resource of interventions and research that many educators and psychologists use to support students' literacy needs using research-informed practises. I also used database searches to identify more potential interventions. Next, I reviewed each intervention on this list and excluded interventions that would not be usable because cost data was unavailable. I also scanned each intervention to assess whether the results might be generalizable to future students, or whether the intervention's results may be constrained by situation-specific factors.

Article selection and coding. For each intervention included in the present study, I searched for fluency intervention articles using multiple databases. Specifically, I used Google Scholar, the Educational Resources Information Center (ERIC), British Education Index, ScienceDirect, Literary Reference Center, MasterFILE Premier, ProjectMuse, Airiti Library eBooks & Journals, JSTR, ScienceDirect, Education Research Complete, Journals @ OVID, and the Networked Digital Library of Theses and Dissertations for these searches. Similarly, I used multiple search terms (i.e. [Intervention Name] + [READING FLUENCY]; [Intervention Name] + INTERVENTION OR EFFECTIVENESS OR EFFICACY]) to avoid biasing searches to publications with one type of keyword. I listed each publication in a Microsoft Excel spreadsheet and coded the included publications into SPSS.

Pooled effect sizes. I visually inspected the effect sizes for the remaining studies and formulated a plan to proceed. Effect sizes were available between subjects (i.e, treatment vs. control, treatment vs. treatment), and within subjects (i.e., pre-test vs. post-test). I focused on treatment vs. control effect sizes because they compare the effect of the fluency intervention against no treatment, and resultantly isolate the benefit of the fluency intervention itself. Comparing effect sizes within subjects was less desirable because the fluency intervention's results may be confounded by participant factors, such as quality fluency teachers or normative fluency development during the school year. I also noted that pooled effect sizes were not feasible for all interventions; as a result, singular effect sizes would be used when pooling was not possible.

Cost data. I obtained cost data for the remaining interventions from developers' websites and from national salary averages, and transferred this information into an Excel spreadsheet. I considered what costs could be held constant due to being very small costs, or whether certain supplies costs would already be available at school. I evaluated which costs would be fixed (e.g., one-time materials purchase) or variable (e.g., teaching time). I determined variable costs such as teaching time using time reported in the relevant included publications and national averages.

Comparisons. I compiled costs and effect sizes for each remaining intervention and compared the cost per unit of effect.

Results

Cost-Benefit Methodology

The methodology developed in the present study enabled me to identify a feasible target for cost-benefit analysis, short-list and evaluate available interventions for that target, and systematically review efficacy studies for those interventions. I was able to establish effect sizes for several studies that met inclusion criteria based on results generalizability and rigor, and combine this information with costs. Overall, the methodology was effective in performing cost-benefit analysis for fluency interventions.

Cost-Effectiveness of Fluency Interventions

Target selection. I considered my objectives for the present study; namely, to develop a cost-benefit methodology and apply it. I chose fluency (i.e., word reading speed) interventions because of fluency's significant role in reading development (e.g., Biancarosa & Shanley, 2016). Taking a pragmatic approach, I selected Early Years (i.e., Kindergarten to Grade 3) readers due to the benefits of early intervention. I planned initially to focus on one grade level (e.g., Grade 3), but my preliminary publication search revealed a lack of data for this group. Expanding the target to Early Years seemed logical, as this group represents emergent readers, so outcomes would likely be similar for students in these grades.

Intervention selection. Using a resource from Reading Rockets, I identified eight Early Years fluency interventions to investigate in the proposed study: Read Naturally, Great Leaps, Quick Reads, Corrective Reading Level A, Open Court, Read Well, Voyager Universal Literacy System, and Horizons Fast Track (Kelly, 2011).

I included more fluency intervention packages that I found when searching for efficacy studies for the abovementioned interventions (e.g., other interventions identified while reading articles). I also searched databases for Early Years interventions using keywords: [READING

FLUENCY] + [KINDERGARTEN OR GRADE 1 OR GRADE 2 OR GRADE 3] and identified several other fluency interventions. These interventions were Readers Theatre, Rock and Read, Repeated Reading, Helping Early Literacy with Practice Strategies (HELPS), Read Two Impress, The Two Minute Solution, Earobics, Reading Together (i.e., Read 2 Impress), Multiple Exemplars, Peer Assisted Learning Strategies (PALS).

I excluded Open Court and Read Well because cost data (i.e. purchase price, materials cost) were unavailable at the time of this thesis' writing. I also excluded Multiple Exemplars because it appeared to be more of a general teaching strategy than a specific intervention. As a result, this strategy's effectiveness seemed too dependent on variables such as the teacher's ability to choose exemplars that would increase fluency.

Article selection and coding. I found 614 peer-reviewed publications between 2008 and 2018 and excluded 596 from further study. After exclusions, 18 publications qualified for cost-benefit analysis. See Appendix B for a table displaying reasons for exclusions.

Visual inspection of the SPSS dataset revealed that the majority of studies measured fluency using a single type of assessment, and that there was little agreement on which type of assessment was used (e.g., GORT, WJ, AIMSweb, DIBELS, ORF, WCPM, TOWRE). The results of one of the few studies to include multiple fluency measures (Mitchell, 2010) revealed a marked variability between measures, suggesting that using different fluency measures in a meta-analysis may introduce a significant validity threat. At the time of this thesis' writing, interchangeability between fluency measures had not been well-published, leaving uncertainty about whether using effect sizes from different validity measures would lead to problematic comparisons. I speculated that different fluency measures involve presenting different text to students, and results could be confounded by these differences (e.g., frequency of high usage words, similarities to intervention words, content). I opted to use the fluency measure used most often in the dataset, which was the GORT. Unfortunately, the cost of this decision was eliminating many interventions (n = 6), leaving HELPS, Quick Reads, and Great Leaps for meta-analysis.

Pooled effect sizes. I pooled effect sizes for each remaining category of fluency intervention (i.e., Great Leaps, HELPS, Quick Reads). A weighted average effect size was possible for HELPS. Quick Reads contained one study; as such, its effect size was not pooled. Great Leaps' interventions did not meet inclusion criteria, except one study, which contained non-significant results and was excluded from further comparisons.

To pool effect sizes for the HELPS interventions, I used the Hunter-Schmidt method because of its weighted approach to pooling effect sizes, and ability to measure variability using credible intervals. Using this approach, I focused on between subject effect sizes, as within subject effect sizes may be likely confounded by variables such as students' typical reading development. The treatment groups for the four HELPS studies included in effects size analyses were weighted by the number of participants in their treatment condition. The weighted mean effect size of HELPS was large ($\bar{r}$ = .80, N = 4), with 95% credibility intervals between .26 and 1.34. QuickReads' effect size from Vadasy and Sanders (2008) fell in the medium range (r = .31, N = 1).

Cost data. I held supplies costs (e.g., binders, pencils) constant for both interventions, assuming that they were fairly small and too insubstantial for the present analysis. Neither intervention required certification, classroom training, or a minimal level of educational expertise (HELPS, 2013; Pearson, 2019). According to the HELPS website (2013), intervention materials (e.g., books, instructions, training) has no cost, with only a suggested donation of $55 to its associated non-profit organization. QuickReads' intervention materials for Early Years students costs $422 for 12 students. The website was unclear about whether student packages could be re-used for new students, so I assumed that this was possible. To assess teaching costs for QuickReads, I multiplied the mean intervention hours from Vadasy and Sanders (2008) hourly teacher cost. On average, students in Vadasy and Sanders (2008) received 30 hours of intervention time. Interventionists for Vadasy and Sanders (2008) were educational assistants, so I used the Canadian national average hourly wage of $22 (indeed.com, n.d.) to derive the teaching cost of 30 hours of QuickReads intervention time. Interventionists in HELPS included certified teachers, undergraduate education students, and graduate research assistants. I simplified teaching costs to using the average hourly wage for elementary school teachers, which was estimated by Living in Canada (2019) to be $38.62. For teaching time, I determined the mean teaching hours (per student) for the included HELPS intervention studies. On average, students in the HELPS intervention received 8.78 hours of teaching time. Overall, total costs for QuickReads were $1,082. Total costs of HELPS, including the suggested donation of $55, was $406.

Comparisons. QuickReads' effect size from Vadasy and Sanders (2008) was .31, while its total cost was $1,082. As such, QuickReads' cost per unit of effect was $3,490. The HELPS pooled effect size was .80, and its total cost was $394. Its cost per unit of effect was $493. Cost and effect size comparisons are displayed in Appendix C.

Discussion

Cost-Benefit Methodology

Canada's literacy deficit remains persistent despite decades of reading research (e.g., Rubenson et al., 2007; Statistics Canada, 2011). One possible explanation for this trend is that limited resources may be invested in interventions that are selected for non-evidence-based reasons including convenience and familiarity (Burns et al., 2012). The most effective interventions may be difficult to identify; as a result, these difficulties may increase reliance on choices based on convenience and familiarity.

The cost-benefit analysis developed in the present study effectively compared resource costs (e.g., time, training, materials) per unit of effect size for fluency interventions. This methodology demonstrates the usefulness of thinking beyond convenience when choosing among interventions, as a popular or familiar intervention may be less effective and more costly than lesser-known alternatives.

The steps developed in the present study are not confined to fluency interventions. The cost-benefit methodology appears to be adaptable for academic interventions as a method to systematically review intervention studies and compare the return on an intervention's resource investment. This methodology can provide useful information in EBP when practitioners match interventions with students (APA Presidential Task Force on Evidence-Based Practice, 2006; Burns et al., 2012). The cost-benefit methodology fits into the RTI framework by helping practitioners choose interventions that have the highest likeliness of success; if the student does not respond adequately to the intervention, practitioners receive time-effective and data-driven support that the student may require further assessment. Choosing cost-effective interventions may also enable practitioners to work effectively within limited budgets and reduce the likeliness that a student may be waitlisted. Given that early intervention supports student achievement and reduces the likeliness of early exit (McIntosh, Chard, Boland & Horner, 2006; Good, Simmons, Kame'enui, 2011; Walker & Sprague, 1999), the cost-benefit methodology developed in the present study may facilitate this trend.

The cost-benefit methodology developed in the present study is not without its limitations; specifically, that the quality of the inputs affect the cost-benefit method's ability to make comparisons. The number of interventions excluded was high due to a lack of publications demonstrating quantitative effectiveness, and the variability of fluency measures (Appelbaum et al., 2018). This lack of data (i.e., or reporting) limited my ability to pool effect sizes for all interventions, establish confidence intervals, and perform regression analyses. These factors would have increased the quality and predictive value of the present analysis.

Cost-Benefit of Fluency Interventions

Using cost-benefit analysis, the HELPS intervention was seven times more cost-effective than QuickReads at increating word reading speed. Pooled effect size comparisons indicated that HELPS was associated with larger effect sizes than QuickReads. The HELPS intervention was also less costly than QuickReads in terms of intervention materials and teaching time. This result suggests that HELPS may accomplish fluency gains faster and at a lower price point than QuickReads. For organizations with limited resources, or students with limited funding, HELPS may be the more strategic choice.

The difference between HELPS and QuickReads' cost-benefit ratios may relate to differences in their methods. QuickReads reports to be an intervention focused on repeatedly reading text, and the instructor models fluent reading (Pearson, 2019). The HELPS intervention includes instructor modelling with repeated readings, and incorporates strategies such as goal-setting, rewards, and memory cues (Begeny, Mann, Cunningham, & Tsuen, 2009). This result aligns with Padeliadu and Giazitzidou's (2018) finding that repeated readings-based interventions are more effective with self-monitoring, goal-setting, and modelled reading than without these strategies.

Evidence-based practices evaluate the causal evidence for an intervention's efficacy (APA Presidential Task Force on Evidence-Based Practice, 2006). A relevant concern that applies in the present study is that the HELPS studies share a common author and perhaps originated from the same lab. It is currently unknown whether the fluency gains offered by HELPS can be repeated by other researchers and interventionists, which slightly weakens HELPS' efficacy evidence. The promising findings from HELPS studies are a reason, however, to pursue future research using this intervention.

Effect size comparisons. Meta-analysis enables researchers and practitioners to evaluate an intervention's efficacy by pooling the results of multiple studies (APA Presidential Task Force on Evidence-Based Practice, 2006). Effect sizes in the present study are constrained by the paucity of available research and/or lack of reporting (i.e., Appelbaum et al., 2018). QuickReads' single usable effect size means that pooled effect size analysis was not possible, thus, it is difficult to assess the generalizability of its effect size to other populations. Because QuickReads' study focused on paraeducator interventionists, it is uncertain whether higher effect sizes could have been achieved if certified teachers or pre-service teachers delivered these interventions. Although HELPS interventions contained more usable effect sizes, these studies were completed by the same researcher, which may impact our ability to generalize these findings beyond their lab and to other populations of struggling readers (i.e., APA Presidential Task Force on Evidence-Based Practice, 2006). Thus, the variability of fluency measures and a lack of research on their validity also limited the findings. My conservative approach (i.e., including studies with the same fluency measure) resulted in the elimination of fluency interventions, which constricted my cost-effect analysis to only two interventions. Although the cost-effect comparisons in the present study are limited, I argue that these results contribute to our understanding of Early Years fluency interventions by directly comparing two interventions and highlighting the need for more research in this area.

Cost stability. I assumed that fluency intervention costs are stable over time and that discounted prices occur infrequently. Cost-effect comparisons, as a result, must be interpreted in relation to current market prices. The cost-benefit methodology developed in the present study enabled fluency interventions to be compared, but lack of rigorous research studies and variability among measures limited the findings. It is apparent that the analytical quality of this cost-benefit methodology depends on the availability and accuracy of cost and time (i.e., intervention hours) information.

Cost-benefit comparisons. Overall, the methodology developed in the present study appears to be adaptable for other academic interventions. In some cases, meta-analysis may be published, which can save researchers time when performing cost-benefit comparisons. If a targeted intervention presents with a larger set of included studies (i.e., in comparison to the present study), more statistical options should be considered, including pooled effect sizes with confidence intervals, and regression analysis to further compare the relationship between cost and effect sizes.

Future Directions

The present study highlights the need for more controlled studies and meta-analyses on the effectiveness of fluency interventions for Early Years readers. Quantitative measures using multiple assessments would increase our ability to compare effect sizes; in addition, more research on the validity of different fluency measures would also progress our understanding in this area. In doing so, we can strengthen the quality of EBP (APA Presidential Task Force on Evidence-Based Practice, 2006). Data-driven clinical approaches increase the likeliness of positive outcomes (APA Presidential Task Force on Evidence-Based Practice, 2006). Replicated studies enable meta-analysis, which bolsters evidence supporting intervention efficacy (APA Presidential Task Force on Evidence-Based Practice, 2006). To enhance our ability to examine cost-benefit of interventions, it would be helpful for researchers to be transparent about intervention times and associated costs.

Despite its limitations, the results of my study remain to be a useful investigation into the relationship of cost and benefit for fluency interventions, and I am considering adapting this method to cost-benefit comparisons for other academic interventions. When data is available, I would include regression analysis in the cost-benefit analysis to further study the relationship between intervention costs and benefits. The cost-benefit process can also be expedited if current effect sizes are available for relevant interventions.

Results for HELPS are promising. Future research is warranted to assess the effectiveness of this intervention, particularly because its publications appear to be produced from the same author and presumably the same lab. Replicated HELPS intervention research would help us to assess whether the positive gains were due to the intervention itself, or due to confounding variables.

Practical Implications

The statistical cost-benefit model developed in this study is adaptable for similar meta-analytic studies that compare cost-benefits for other building blocks of fluency, other age groups, and adult literacy programs. The general notion of the cost of effective interventions can facilitate decision-making by explaining how various investments in fluency interventions can inform different outcomes, and that strategically planning these investments can lead to more effective interventions. Choosing cost-effective reading interventions will facilitate strategic planning for schools and school districts, which may increase intervention accessibility and reduce wait times.

References

American Psychological Association Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist, 61,* 271-285.

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wison, E., Nezu, A. M., Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA publications and communications board task force report. *American Psychologist, 73*(1), 3-25.

Basu, A. (2017) How to conduct meta-analysis: A basic tutorial [PDF file]. Retrieved from https://peerj.com/preprints/2978.pdf

Begeny, J. C., Mann, C. M., Cunningham, M. D., & Tsuen, H. Y. (2009). The Helping Early Literacy with Practice Strategies (HELPS) curriculum: Instructional passages developed for use with the HELPS program. Raleigh, NC: The HELPS Education Fund. Retrieved from http://www.helpsprogram.org

Berkman, N. D., Sheridan, S.L., Donahue, K. E., Halpern, D. J., & Crotty, K. (2011). Low health literacy and health outcomes: An updated systematic review. *Annals of Internal Medicine, 155*(2)*,* 97-107. Doi: 10.7326/0003-4819-155-2-201107190-00005

Biancarosa, G. & Shanley, L. (2016). What is fluency? In K. D. Cummings and Y. Petscher (Eds.) *The Fluency Construct: Curriculum-Based Measurement Concepts and Applications.* New York, NY: Springer. doi: https://doi.org/10.1007/978-1-4939-2803-3_1

Boreinstein, M., Hedges, L. V., Higgins, J. P. T., Rothstein, H. R. (2009). *Introduction to meta-analysis.* West Sussex, United Kingdom: John Wiley & Sons, Ltd.

Briggs, A. H., O’Brien, B. J., & Blackhouse, G. (2002). Thinking outside the box: recent advances in the analysis and presentation of uncertainty in cost-effectiveness studies. *Annual review of public health*, *23*(1), 377-401. Doi: https://doi.org/10.1146/annurev.publhealth.23.100901.140534

Burns, M. K., VanDerHeyden, A. M., & Riley-Tillman, T. C. (2012). *RTI applications, volume 1: Academic and behavioral interventions.* New York: The Guilford Press.

Chard, D. J., Vaughn, S., & Tyler, B. (2002). A synthesis of research on effective interventions for building reading fluency with elementary students with learning disabilities. *Journal of Learning Disabilities, 35*, 386-406. doi: 10.1080/15377903.2017.1328628

Chall, J. S., Jacobs, V., & Baldwin, L. (1990). *The reading crisis.* Cambridge, MA: Harvard University Press.

Chiswick, B. R. & P.W. Miller. (1990). *Language in the labor market: The immigrant experience in Canada and the United States.* Paper presented at the Conference on Immigration, Language, and Ethnic Issues: Public Policy in Canada and the United States, Washington, DC. Retrieved from: http://files.eric.ed.gov.login.ezproxy.library.ualberta.ca/fulltext/ED341233.pdf

Cleophas, T.J. & Zwinderman, A.H. (2017). *Modern meta-analysis: Review and update of methodologies.* Sliedrecht, The Netherlands: Springer International Publishing. https://doi-org.login.ezproxy.library.ualberta.ca/10.1007/978-3-319-55895-0

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204-256. doi:10.1037/0033-295X.108.1.204

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches.* Los Angeles: SAGE Publications, Inc.

DeWalt, D. A., Berkman, N.D., Sheridan, S., Lohr, K. N., Pignone, M.D. (2004). Literacy and health outcomes: A systematic review of the literature. *Journal of General Internal Medicine, 19*(12), 1228-1239. Doi: https://doi.org/10.1111/j.1525-1497.2004.40153.x

Dozois, D. J., Mikail, S. F., Alden, L. E., Bieling, P. J., Bourgon, G., Clark, D. A., ... & Johnston, C. (2014). The CPA Presidential Task Force on Evidence-Based Practice of Psychological Treatments. *Canadian Psychology/Psychologie Canadienne*, *55*(3), 153. Retrieved from https://open.library.ubc.ca/cIRcle/collections/facultyresearchandpublications/52383/items/1.0224125

Erion, J., & Hardy, J. (2019). Parent tutoring, instructional hierarchy, and reading: A case study. *Preventing School Failure: Alternative Education for Children and Youth,* 1-11.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). London, UK: SAGE Publications, Inc.

Field, A. P., & Gillett, R. (2010). How to do a meta-analysis. *British Journal of Mathematical and Statistical Psychology*, *63*(3), 665-694. doi: https://doiorg.login.ezproxy.library.ualberta.ca/10.1348/000711010X502733

Foorman, B. R., Francis, D. J., Shaywitz, S. E., Shaywitz, B. A., & Fletcher, J. M. (1997). The case for early reading intervention. In B. A. Blachman (Ed.), *Foundations of reading acquisition and dyslexia: Implications for early intervention* (pp. 243-264). Mahwah, NJ: Lawrence Erlbaum Associates.

Frank, K., Phythian, K., Walters, D., & Anisef, P. (2013). Understanding the economic integration of immigrants: A wage decomposition of the earnings disparities between native-born Canadians and recent immigrant cohorts. *Social Sciences*, *2*(2): 40-61. doi: https://doi.org/10.3390/socsci2020040

Good III, R. H., Simmons, D. C., & Kame'enui, E.J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading, 5*(3), 257-288. doi:dx.doi.org/10.1207/S1532799XSSR0503_4

Haring, N. G., & Eton, M. D. (1978). Systematic instructional technology: An instructional hierarchy. In N. G. Haring, T. C. Lovitt, M. D. Eaton, & C. L. Hansen (Eds.), *The fourth R: Research in the classroom* (pp. 23-40). Columbus, OH: Merrill.

Hedges, L. V. & Olkin, I. (1985). *Statistical methods for meta-analysis.* San Diego, CA: Academic Press.

HELPS (2013). HELPS Program Materials. Retrieved from: http://www.helpsprogram.org/materials.php

Hudson, R. F., Lane, H. B., & Pullen, P. C. (2011). Reading fluency assessment and instruction: What, why and how? *The Reading Teacher, 58*, 702-714.

Kaplan, D., & Venezky, R.L. (1994). Literacy and voting behaviour: A bivariate probit model with sample selection. *Social Science Research, 23*(4), 350-367 doi: https://doi.org/10.1006/ssre.1994.1014

Kelly, C. (2011). *Reading intervention programs: A comparative chart* [PDF file]. Retrieved from http://www.readingrockets.org/pdfs/Reading-intervention-programs-chart.pdf

Kirsh, I. S., Jenkins, L., Jungeblut, A., & Kolstad, A. (1993). Adult literacy in America: A first look at the results of the National Adult Literacy Survey [PDF file]. National Center for Education Statistics. Retrieved from: https://nces.ed.gov/pubs93/93275.pdf

Kuhn, M. R., & Stahl, S. A. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology, 95*(1), 3-21. http://dx.doi.org/10.1037/0022-0663.95.1.3

Krugman, P. R. (1979). Increasing returns, monopolistic competition, and international trade. *Journal of International Economics*, *9*(4), 469-479. doi: https://doi.org/10.1016/0022-1996(79)90017-5

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive psychology*, *6*(2), 293-323. doi: https://doi.org/10.1016/0010-0285(74)90015-2

Law, C., & Cupples, L. (2017). Thinking outside the boxes: Using current reading models to assess and treat developmental surface dyslexia. *Neuropsychological Rehabilitation, 27*(2), 149-195. https://doi.org/10.1080/09602011.2015.1064453

Lyon, G. R. (July 10, 1997). *Report on Learning Disabilities Research.* Testimony before the Committee on Education and the Workforce, U.S. House of Representatives.

McIntosh, K., Chard, D. J., Boland, J.B., & Horner, R. H. (2006). Demonstration of combined efforts in school-wide academic and behavioral systems and incidence of reading and behavior challenges in early elementary grades. *Journal of Positive Behavior Interventions, 8*(3), 146-154. doi: https://doi.org/10.1177/10983007060080030301

OECD, Statistics Canada (2011). *Literacy for Life: Further Results from the Adult Literacy and Life Skills Survey.* OECD Publishing. doi: http://dx.doi.org/9789264091269–en

Padeliadu, S., & Giazitzidou, S. (2018). A synthesis of research on reading fluency development: Study of eight meta-analyses. *European Journal of Special Education Research, 3,* 232-256. doi: 10.5281/zenodo.1477124

Pearson (2019). Quickreads. Retrieved from https://tinyurl.com/y3sy7ant

Perfetti, C. A. (1977). Language comprehension and fast decoding: Some psycholinguistic prerequisites for skilled reading comprehension. In J. T. Guthrie (Ed.), *Cognition, curriculum, and comprehension* (pp. 20–41). Newark, DE: International Reading Association.

Perfetti, C. A. (1985). *Reading ability,* New York: Oxford University Press.

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. Psychological Review, 114, 273315

Perry, C., Ziegler, J. C., Braun, M., & Zorzi, M. (2010). Rules versus statistics in reading aloud: New evidence on an old debate. *European Journal of Cognitive Psychology*, *22*(5), 798-812.

Rubenson, K., Desjardins, R., & Yoon, E. (2007). *Adult learning in Canada: A comparative perspective. Results from the Adult Literacy and Life Skills Survey* [PDF file). Retrieved from http://odesi2.scholarsportal.info/documentation/ALL/89-552-m2007017-eng.pdf

Schmidt, F. L. & Hunter, J. E. (2015). *Methods of meta-Analysis: Correcting error and bias in research findings (3rd ed.)* doi: http://dx.doi.org.login.ezproxy.library.ualberta.ca/10.4135/9781483398105

Siegel, C., Laska, E., & Meisner, M. (1996). Statistical methods for cost-effectiveness analyses. *Contemporary Clinical Trials*, *17*(5), 387-406. doi: https://doi-org.login.ezproxy.library.ualberta.ca/10.1002/hec.1477

Smith, W. C., & Fernandez, F. (2017). Education, skills, and wage gaps in Canada and the United States. *International Migration, 55*(3), 57-73. doi: 10.1111/imig.12328

Stanovich, K. E. (1991). Word recognition: Changing perspectives. In R. Barr, M.L. Kamil, P. Mosenthal, & P.D. Pearson (Eds.), *Handbook of reading research* (Vol. 2., pp. 418-452). New York: Longman.

Stevens, E. A., Walker, M. A., & Vaughn, S. (2017). The effects of reading fluency interventions on the reading fluency and reading comprehension performance of elementary students with learning disabilities: A synthesis of the research from 2001 to 2014. *Journal of Learning Disabilities, 50*, 576-590. doi:10.1177/0022219416638028

Stern, J. M. B., & Piper, B. (2018). Resetting targets : Examining large effect sizes and disappointing benchmark progress. *RTI Press.* Retrieved from https://tinyurl.com/y397t3b7

Song, D. (2010). Retrospective study. In N. J. Salking (Ed.) *Encyclopedia of research design*. (pp.1283-1284). Thousand Oaks, U.S.A: SAGE Publications, Inc.

Swanson, H. L. (1999). Reading research for students with LD: A meta-analysis of intervention outcomes. *Journal of learning disabilities*, *32*(6), 504-532. doi: https://doi.org/10.1177/002221949903200605

Teasdale, T. W., & Owen, D. R. (2005). A long-term rise and recent decline in intelligence test performance: The Flynn Effect in reverse. *Personality and Individual Differences*, *39*(4), 837-843. doi: 10.1016/j.paid.2005.01.029

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2004). The plurality of literacy and its implications for policies and programs. UNSECO Education Sector Position Paper. Retrieved from: https://unesdoc.unesco.org/ark:/48223/pf0000136246

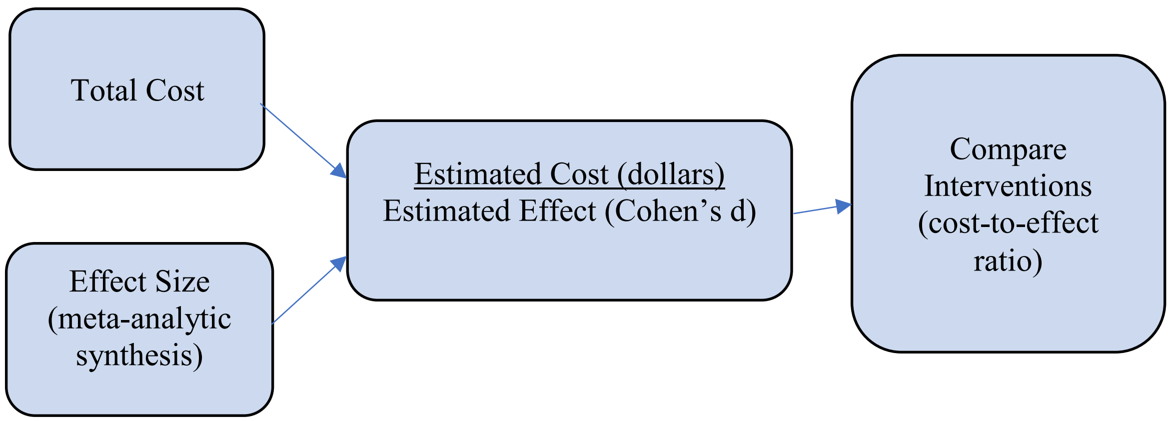
Walker, H. M., & Sprague, J. R. (1999). The path to school failure, delinquency and violence: Causal factors and some potential solutions. *Intervention in School and Clinic, 35,* 67-73. doi: https://doi.org/10.1177/105345129903500201

What Works Clearinghouse (2010). Voyager Reading Program. Retrieved from: https://ies.ed.gov/ncee/wwc/EvidenceSnapshot/537

Appendix A

Figure 1

Data Collected



Note. Intervention-level analyses performed separate analyses for each fluency intervention included in the study. Intervention-level cost-effectiveness ratios were calculated for each intervention. These were compared and ranked in the final stage of the analysis.

Appendix B

Table 1

Excluded Publications (n = 596)

Reason	N
Interventions	
Group format	20
Modified for a specific culture	1
Combined with another intervention, fluency data unable to be separated	446
Population	
Participants in Grade 4 or older, Early Years data unable to be separated	60
Participants included English Language Learners	30
Participants belonged to specific clinical population	8
Publications	
Duplicates	22
Publication not written in English	4
Statistics	
Effect sizes indeterminable	5

Appendix C

Table 2

Cost-Benefit Analysis

		Intervention Type	
		HELPS	QuickReads
Fixed Costs ($)	Materials	55	422
Variable Costs ($)	Teaching	339	660
	Costs per Student	0	0
	Training	0	0
Total Costs ($)		394	1082
Effect Size		.8	.31
Cost/Effect Ratio		493	3490

Note. Teaching time for HELPS was calculated by multiplying intervention hours (8.78) and estimated wage ($38.62/hour). Teaching time for QuickReads was calculated by multiplying intervention hours (30) and wage ($22/hour).