Separating the medium from the message: Effects of web- versus pencil and paper- delivery of the ABRACADABRA intervention on literacy, motivation and self-esteem

Abstract

This study reports a randomized control trial intervention investigating the impact of delivery format (computer versus paper) on students' reading and spelling skills, reading motivation and self-esteem using a web-based early literacy tool, A Balanced Approach for Children Designed to Achieve Best Results for All (ABRACADABRA) alongside a paper version of this tool. Based on critiques of technology by Clark (1983) and the Time-Displacement Hypothesis of technology (Vandewater, Bickham, & Lee, 2006), we predicted negative effects of technology on reading, spelling, and reading-related motivation, and self-esteem at post-test. The ABRACADABRA intervention was supplemental, delivered in three weekly 15-minute supplemental reading sessions for eight weeks. Results first showed no difference in the pace and depth of delivery across format and also showed comparable improvements in participants' reading and spelling at post-test in both the computer-based and paper ABRACADABRA instruction conditions and little evidence of difference by medium of intervention delivery on reading motivation, self-esteem, and enjoyment. It was concluded that the computer-based intervention does not have negative effects over its paper counterpart on students' literacy skills, and related literacy percepts, and provide no support for the Clark or Time-Displacement Hypotheses in this context.

Keywords: ABRACADABRA, randomized control trial, spelling, reading motivation, reading self-esteem

Separating the medium from the message: Effects of web- versus pencil and paper-delivery of the ABRACADABRA intervention on literacy, motivation, and self-esteem

Does the medium of instruction influence learning? An extended discourse on this issue that has become known as the Clark-Kozma technology/media and learning debate (Clark 1983, Kozma, 1994) explored contentions concerning the effectiveness of technology as a medium of instruction over traditional methods (Bavelier, Green, & Dye, 2010). Recently McNally, Ruiz-Valenzuela and Rolfe (2016) and Johnson et al. (2019) used ABRACADABRA (A Balanced Approach for Children Designed to Achieve Best Results for All), a web-based early literacy tool, to contrast the mode of delivery (computer versus paper) in teaching reading to lower elementary school students in the United Kingdom. McNally et al. (2016) and Johnson et al. (2019) found both forms of intervention to be effective compared to an untaught control group but also found the paper delivery to be somewhat more effective than the computer. The present study seeks to replicate the McNally et al. (2016) and Johnson et al. (2019) studies and expands it to explore spelling, reading motivation and reading self-esteem and also seeks to explore whether the pace and depth of delivery of the intervention affected outcomes. Effective early reading interventions are first reviewed. The role of computer- and paper-based programs in shaping elementary students' reading motivation, reading self-esteem and reading enjoyment is then discussed. This is followed by an overview of online computer-based reading programs and an overview of ABRACADABRA before the current study is described.

1.1 Reading Instruction and Effective Early Interventions

Evidence from systematic reviews and meta-analytic studies of 'what works' to improve reading provide a reasonable platform for understanding how to improve literacy. Savage and

Cloutier (2017) reviewed meta-analyses and systematic reviews of well designed (randomized control trial (RCT) and quasi-experimental design (QED) studies early reading interventions for phonics, reading comprehension and fluency. For phonics, some nine previous meta-analyses demonstrate robust support for phonics instruction facilitating reading growth. Phonics instruction is most effective when it involves letter-sound knowledge and phonological awareness training together, though more intervention studies that systematically contrast length and content are needed to understand what optimal practice might be. Effective reading comprehension involves instruction of key language sub-components, and explicitly teaching strategies of previewing, predicting, summarizing, and comprehension monitoring. Repeated reading interventions generally led to improved reading fluency at least for familiar, practiced passages. Savage and Cloutier (2017) argue that the case for using educational technology to support literacy has a less secure evidence base currently. This issue is considered below.

1.2. Reading Instruction and Educational Technologies

There is significant debate about the utility of educational technologies. A key question in such debates concerns whether the medium of delivery impacts the effectiveness of instruction. This question is captured in what has become known as the Clark-Kozma technology/media and learning debate. Clark’s (1983) hypothesis is that the medium and the instructional methods have been conflated, and that it is instructional methods that facilitate learning. Clark emphasized the need to examine whether distinct attributes of media have a special cognitive effect on learning or contribute meaningfully to other processes crucial to learning. In response, Kozma (1994) emphasized the importance of looking at wider components of learning (e.g. cognitive/social/emotional factors), in particular how the fundamental structure and functions of media might influence these processes. Kozma (1994) emphasized employing theories, frameworks and interventions that allow for effects of media on cognitive and social processes to be explored.

The Clark-Kozma debate still remains relevant in contemporary research. For example, a collection of recent studies on children and youth's reading of traditional books versus electronic activities reveals that evidence supporting the Time-Displacement Hypothesis that technology displaces time devoted to quality reading is limited (Thomson, Barzillai, van den Broek, & Schroeder, 2018). The present study seeks to inform this unresolved debate in the specific case of literacy, where, as we show below, there is currently a lack of relevant data addressing this issue.

Educational technology can be defined as "a variety of electronic tools and applications that help deliver learning materials and support learning process in classrooms" (Cheung & Slavin, 2012, p.201). These instructional mediums can include computer-assisted instruction, integrated learning systems and the use of multimedia such as videos (Cheung & Slavin, 2012). A series of meta-analyses (Becker, 1992; Blok, Oostdam, Otter, & Overmatt, 2002; Cheung & Slavin, 2012; Kulik & Kulik, 1991; Kulik, 2003; Ouyang, 1993; Soe, Koki, & Chang, 2000), have found effects ranging from what they describe as being (very) small (+0.06), to medium, (+0.43), for the use of educational technology on improving Grades K to 12 students' reading performance (Cheung & Slavin, 2012).

In considering the small overall impact of educational technology on reading instruction, Savage and Cloutier (2017) drew attention to three methodological issues: implementation of studies, quality of technology used, and the theoretical and pedagogical alignment of technologies. Where training quality and treatment integrity are high, educational technology for reading instruction can consistently produce medium effect sizes of 0.60, based on Cohen's established standards for assessing effect sizes (Cohen, 1988), (Archer, Savage, Sanghera-Sidhu, Wood, Gottardo, & Chen, 2014; Savage & Cloutier, 2017), suggesting the effects of attainment on literacy are of practical utility. Secondly, of the commercially available early literacy programs, only 15% of such programs provided targeted instruction for synthetic phonics, none provide instruction in other key reading skills such as phoneme segmentation or print-based knowledge and none teach comprehension strategies (Grant, Wood, Gottardo, Evans, Phillips, & Savage, 2017). The majority of reading programs are not based upon theories of technological literacy.

1. 3. Writing Instruction and Technology Media

There exists a modest literature of studies directly comparing the effects of digital and pen/pencil writing instruction on young children and elementary school students' writing abilities. A review by Wollscheid, Sjaastad and Tømte (2016) highlighted the mixed nature of findings. Wollscheid et al. (2016) note that studies are routinely based on different theoretical perspectives and associated methodologies (e.g. cognitive-theoretical, neuroscience and learning, socio-cultural) with very few RCTs. There also appeared to be an association between the year in which studies were conducted and preference for traditional writing tools (found in earlier studies) or digital writing tools (found in later studies).

1.4 Reading Motivation

Reading motivation refers to an individual's beliefs, values and goals with respect to reading (Guthrie & Wigfield, 2000). In the context of reading, some motivations can be termed intrinsic reading motivation and others, extrinsic reading motivation (Deci, Koestner, & Ryan, 1999). Intrinsic reading motivation is shown in enjoying reading, valuing books, attaching personal importance to reading and having interest in the topic contained in the reading material

(Becker, McElvany, & Kortenbruck, 2010), and can be linked to a learning goal orientation. Extrinsic reading motivation includes receiving recognition, rewards or incentives from others (Wang & Guthrie, 2004), and meeting others' expectations or avoiding negative consequences (Hidi, 2000). Extrinsic reading motivation is associated with a performance goal orientation. Research on modality of delivery and children's motivation to read is modest and mixed in outcome. Only two studies of motivation that used experimental pre- and post-test designs with treatment and control groups and which also matched the content of the intervention across modality were identified. In Greenlee-Moore and Smith (1996) students' outcomes favoured the computer modality of delivery, whereas in Aydemir and Oeztuerk (2012) paper delivery was advantaged.

1.5 Reading Self-esteem

The multi-dimensional model of self-concept (Harter, 2012; Shavelson, Hubner, & Stanton, 1976) describes self-esteem as a global, hierarchical construct that has many sub-dimensions such as academic self-concept and non-academic self-concepts. In this model, global self-concept otherwise known as self esteem, is at the highest point of the model, followed by academic self-concepts (verbal and mathematical), and non-academic self-concepts (social, emotional and physical). Reading self-esteem refers to an individual's beliefs about their ability to complete tasks within the domain of reading (Bandura, 1977) and additionally encompasses behaviours, attitudes and motivation related to reading (Gross, 2004). Thorough systematic searches of the entire literature suggest that there exists no empirical research comparing the effects of computer-based and paper-based intervention programs on students' reading self-esteem.

1.6 ABRACADABRA

ABRACADABRA (hereafter, ABRA) is a free web-based literacy program for educators, pre-literate and beginner reader students and parents. ABRA was developed by the Centre for the Study of Learning and Performance at Concordia. ABRA provides a well-balanced literacy curriculum that is highly interactive in nature, containing a broad selection of engaging activities and digital stories for students. ABRA contains visually engaging activities for phonological awareness, decoding, word reading, fluency, comprehension and writing, most with multiple progressive levels of task difficulty, all linked to 21 stories. ABRA provides feedback for learners. Activities include multi-level prompts to support the child in giving the correct answer. Correct answers receive a confirmatory response (e.g. "Cat. That's the word!"), followed by a positive comment (e.g. "Awesome job!") commending the child's knowledge and effort. After two consecutive incorrect answers, the system will provide visual-auditory representations (e.g. showing and saying the word "cat") of the correct answer. ABRA thus potentially aids motivation and related reader self-perceptions. ABRA has been intensively researched through a series of randomized control trial and quasi-experimental studies (for a meta-analysis see Abrami, Borohkovski, & Lysenko, 2015), and brought to scale in global contexts such as Northern Australia (Wolgemuth et al., 2013), Kenya (Abrami, Wade, Lysenko, Marsh, & Gioko, 2016) and Hong Kong (Cheung, Mak, Abrami, Wade, & Lysenko, 2016). For these reasons, ABRA is selected as the literacy tool to investigate the role of the technology medium on learning in this study.

McNally, Ruiz-Valenzuela and Rolfe (2016) and Johnson et al. (2019) addressed these questions in a rigorous randomized control study that incorporated blinded allocation to condition and independent treatment evaluation and analysis. Johnson et al.'s (2019) work is a direct extension of McNally et al.'s (2016) study incorporating a one-year delayed post-test whereby students' literacy skills were examined one year after the end of the intervention.

Trained teaching assistants (*n* = 51) delivered ABRA programs to typical Year 1 students (*n* = 2,241) in the United Kingdom. Schools (*n* = 51) were first randomly allocated as either 1) treatment school (computer or paper intervention) or 2) control school (regular instruction). There was one TA per school. In treatment schools, students were then randomly allocated to either 1) computer-based ABRA, 2) paper ABRA or 3) standard literacy instruction. Within the standard literacy instruction control group, there were two sub-groups: students in strictly control schools and students belonging to control groups within treatment schools. This design component was implemented to assess 'spillover effects' (from students in intervention groups to those in control groups) in treatment schools. The fifteen-minute intervention sessions were held four times per week. Teaching assistants delivered sessions to small groups of three to five students. Treatment integrity evaluations were high overall and assessed as equivalent across both conditions.

Regression analyses that also accounted for the nestedness of data within schools were conducted to produce standardized regression coefficients. The following standardized beta weights were reported by McNally et al.'s (2016) (page 31 Table 10) in analyses of heterogeneous effects of the interventions on the primary outcome variable – The Progress in Reading Assessment (PIRA) test. Results showed both ABRA interventions led to positive gains in literacy, although the paper condition (β = .23) was somewhat higher than the computer-based (β = .138) one. Students who received either medium of ABRA instruction made two to three months' worth of literacy gains compared to their peers who received standard literacy instruction, based on students' post-test results from the PIRA test. Thirdly, the positive effects of the intervention were more significantly pronounced in students who were from low socio-economic backgrounds for both the computer-based (β = .368) and paper (β = .396) ABRA conditions, and also for those who were below average readers at pre-test for both the computer-based (β =.215) and paper (β = .230) conditions. Overall, these results suggest that curriculum content carried more weight in facilitating overall literacy growth compared to medium of instruction.

McNally et al. (2016) and Johnson et al. (2019) also explored possible reasons for why the paper intervention had a marginally larger effect on students' post-test literacy achievement compared to the computer-based intervention. From the process evaluation findings, teaching assistants felt that the paper intervention could be more flexibly adapted to meet students with various ability levels. Arguably the pace of lessons and amount of content coverage across conditions is a possible confound that must also be empirically considered and stringently controlled for in future work. For example, it is possible greater pace and content delivery occurred in the pencil-and –paper intervention as it was not affected by the content buffering that sometimes impacted the technology-based version of ABRA. This potential confound was not formally controlled in McNally et al. (2016) and Johnson et al. (2019) but is here.

It is also possible however that observed patterns reflect a Time-Displacement Hypothesis of technology. The Time-Displacement Hypothesis states that assuming time spent on literacy is zero-sum and that technology has sub-optimal intrinsic value and may thus displace other non-technology based activities that are arguably of greater educational value (Vandewater et al., 2006; Weis & Cerankosky, 2010). Such hypotheses are consistent with views of other broader critics of educational technology (Bavelier et al., 2010; Clark 1983) noted earlier. Empirically, Weis and Cerankosky (2010) provide evidence from an RCT with mediation analyses that the introduction of video games led directly to reduced academic functioning and poorer behaviour in boys. By extension here, a modified Time-Displacement Hypothesis might suggest that the ‘game’ aspects of ABRA may displace more valuable paper-based instructional experiences.

The significance of McNally et al.’s (2016) study is two-fold. The study was commissioned by the Education Endowment Foundation (EEF), a British government-designated What Works Centre responsible for enhancing school-based learning outcomes for children. As such, the study is an independently evaluated randomized control trial with multiple quality design features (e.g. design and outcome pre-publication, power specification, blind independent school-level randomisation, and fully independent assessment, analysis, and reporting of results by a team unrelated to intervention design and delivery team). The subsequent report was then evaluated and approved by the EEF prior to dissemination to schools and the national public as part of recommended best practices in reading instruction, and with additions, published in a quality peer reviewed journal (Johnson et al., 2019).

The second significant contribution of McNally et al.’s (2016) and Johnson et al.’s (2019) studies is that they are also two of the very few studies directly contrasting computer- and paper-based reading instruction on students’ literacy skills using identical curriculum content. A comprehensive literature search was conducted here on two major education and psychology empirical databases ERIC and PsycINFO using the search term: “Educational technology” OR “computer based training” OR “computer assisted instruction” OR “computer assisted language learning” AND “reading achievement” OR “reading outcomes” OR “reading performance” OR “reading readiness” OR “reading ability” OR “reading skills” AND “Spelling” OR “orthography” OR “spelling development” OR “spelling components” OR “spelling skills” AND "elementary school students" OR "primary school students" OR "grade 1 students" OR "grade 2 students".

Based on search results, a total of seven meta-analyses investigating various types of educational technology on students' reading were found, including computer-assisted instruction (Blok et al., 2002; Fletcher-Finn & Gravatt, 1995; Ouyang, 1993; Soe, Koki, & Chang, 2000), computer-based instruction (Kulik & Kulik, 1991), integrated learning systems (Becker, 1992), or a comprehensive overview of all these technology types (Cheung & Slavin, 2012). None of these reviewed works have specifically investigated educational technologies where equivalent content was delivered via both computer and paper mediums. According to search results, only one randomized control trial and one quasi-experimental design studies examining the effects of digital and pen/pencil writing instruction on students' writing skills were found (Ouellette & Tims, 2014; Wollscheid, Siaastad, Tømte, & Løver, 2016). None of the reviewed studies were randomized control trial designs employing reading instruction interventions that examined the overarching literacy skills (i.e., reading and spelling) of Grade 1 and 2 students.

Similarly, there are limited studies in terms of number, design, or scope of investigation, exploring medium of instruction on social-emotional literacy components. Comprehensive literature searches on ERIC and PsycINFO using these groups of search terms: "Reading motivation" OR "academic achievement motivation" OR "intrinsic motivation" OR "extrinsic motivation" AND "elementary school students" OR "primary school students" OR "grade 1 students" OR "grade 2 students"; "Reading self-concept" OR "academic self-concept" OR "reading self-confidence" OR "reading self-esteem" OR "reading self-efficacy" AND "elementary school students" OR "primary school students" OR "grade 1 students" OR "grade 2 students"; "Reading enjoyment" OR "reading contentment" OR "reading euphoria" OR "reading happiness" OR "reading engagement" AND "elementary school students" OR "primary school students" OR "grade 1 students" OR "grade 2 students" were conducted on reading motivation, reading self-concept, and reading enjoyment, respectively.

For reading motivation, only two studies applied randomized control trial designs and provided equivalent reading content across computer and paper delivery (Greenlee-Moore & Smith, 1996; Aydemir & Oeztuerk, 2012). However, these studies were non-instruction based and focused on upper elementary Grade 4 and 5 students. In terms of reading self-esteem, a combined total of 107 search items were found from ERIC and PsycINFO. All items were reviewed, but none were randomized control trials exploring equivalent computer and paper-based reading delivery in relation to self-esteem. As for reading enjoyment, three studies used randomized control trial designs and provided identical reading content via computer and paper delivery (Grimshaw, Dungworth, McKnight, & Morris, 2007; Moody, Justice, & Cabell, 2010; Chaudhry, 2014). However, these studies were non-instruction based and participant population varied from preschool, Grade 1 to Grade 4.

None of the studies reviewed here fit the overarching criteria of a reading instruction-based randomized control trial, with equivalent computer and paper content delivery and its effects on Grade 1 and 2 students' literacy skills and related social-emotional outcome. Given this research gap, this led to an investigation of medium of reading instruction delivery and social-emotional literacy components in the present study.

1.7 The Present Study

The present study thus aims to address whether the medium of ABRA instruction carries any differential effects on overall literacy skills and social-emotional literacy components. A key part of this study will explore whether the computer-based intervention has general negative carryover effects on spelling skills over its paper counterpart. This study extends the work of McNally et al. (2016) and Johnson et al. (2019) in three areas: effects of ABRA interventions on spelling skills, exploring reading motivation, reading self-esteem and reading enjoyment and ensuring equal pace of both ABRA interventions. The research questions explored in this study are: 1) What are the effects of the computer-based vs. paper ABRA instruction on students' reading skills (listening comprehension, word reading) post-intervention? 2) What are the effects of the computer-based vs. paper ABRA instruction on students' spelling skills (letter writing and word spelling) post-intervention?, and 3) What are the effects of computer-based vs. paper ABRA reading activities on students' reading motivation and reading self-esteem post-intervention?

2.0 Method

2.1 Design

An experimental research design using a pre-test post-test RCT was applied to compare the effectiveness of a computer-based version of ABRA with its paper version on literacy, reading motivation, and reading self-esteem outcomes. The independent variable in this study was the delivery type of ABRA (computer vs. paper). The dependent variables were the literacy outcomes (reading and spelling attainment), reading motivation and reading self-esteem. The control variables that were kept constant throughout the experiment included the pace of lessons and treatment integrity. Power was calculated prior to the beginning of the study for the purpose of participant recruitment. A power analysis was conducted using GPower on this study's design, with N = 34 and with alpha = 0.05 and assuming a correlation of r = .7 between pre- and post-test reading measures. This design is powered at 0.8 to detect d = 0.4, which are small to medium effect sizes. It should be noted that the 0.8 power and d = 0.4 are based on actual observed sample size and are used to attain a sufficiently large sample in the first instance. All measures were administered at pre-test and post-test. Randomization of participants involved two parts: 1) generating matched within-class participant pairings and 2) randomly allocating each participant within pairings to either the computer-based or paper ABRA condition. Within-class randomization process was undertaken to avoid confounding effects of teacher on outcomes.

2.2 Participants

Participants were all Grade 1 and Grade 2 students (*N* = 34) from two elementary English language schools within the X school board, Quebec from whom formal parent consent to participate were received. All elementary English language schools within the X School Board (*N* = 27) were initially contacted with invitations to participate in this study. School principals were provided with the research study proposal, and a university and school board ethics approval certificate. Of these schools, two were willing to participate. Both schools have established working relationships with the ABRA research team from participating in previous ABRA studies. For participants in the first elementary school, the pre-test took place in mid-November 2017 and the post-test in early February 2018. For participants in the second elementary school, the pre-test took place in early January 2018 and the post-test in mid-April 2018.

The participant sampling is detailed in a CONSORT flow diagram in Figure 1. Initial recruitment yielded (*N* = 38) participants, however a subsequent total of four students withdrew from the study. During the pre-test process, one Grade 1 male student was unable to complete the pre-test measures due to behavioural difficulties. This student was therefore not included in the randomization process. Once the study began two Grade 1 male students declined to participate after attending the first session in week 1. In week two, one Grade 1 male student withdrew from the study, after his classroom teacher and parents had a discussion about his level of focus and compliance during reading sessions. Accordingly, all four students' parent questionnaire data and pre-test data were omitted from further data analyses. Overall 76% of parent questionnaires were returned. Students from both schools had typical reading and spelling. Students were unselected in that no exclusion criteria were applied in the selection of participants.

2.3 Apparatus and Materials

Wide Range Achievement Test (WRAT, Wilkinson & Robertson, 2006) Version 4 measures basic academic skills such as reading, spelling and math in individuals ages 5 to 94. The WRAT is norm-referenced with percentiles, standard scores, age equivalents and grade equivalents. The word reading and spelling sub-tests were individually administered to Grade 1 and Grade 2 students. It took approximately 15 to 20 minutes to complete. The word reading sub-test consists of two parts: letter reading (15 alphabet letters in total) and word reading (55 words in total). The McDonald's omega reliability in this sample was $\omega = 0.93$. The spelling sub-test consists of two parts: letter writing (writing one's name followed by writing 13 alphabet letters) and word spelling (42 words in total). The McDonald's omega reliability in this sample was $\omega = 0.92$.

Group Reading Assessment Diagnostic Evaluation (GRADE, Williams, 2001) is a diagnostic reading assessment that measures the developmental skills of students from Pre-K to Grade 12. Measurement of students' reading skills includes the areas of: word reading, vocabulary, listening comprehension and reading comprehension. It determines what developmental skills have been mastered and where instruction or intervention is needed. The GRADE is a norm-referenced and developmentally based assessment tool that is typically administered in groups. In this study only the listening comprehension sub-test was administered.

This sub-test provides a broad understanding of students' receptive language skills. The listening comprehension sub-test at Levels 1-2 appropriate for this age group consists of 17 questions where students listened to statements read aloud by the researcher and then responded by colouring in a small bubble for one of four picture options. It took approximately 15 to 20 minutes to complete. The McDonald's omega reliability in this sample was $\omega = 0.63$.

Reading Self-Concept Scale (RSCS, Chapman & Tunmer, 1995) is an assessment tool that measures the reading sub-component of academic self-concept. The tool includes three main areas of reading self-concept: 10 competence in reading items, 10 perceptions of difficulty with readings items and 10 attitudes towards reading items. Examples from each of these sub-scales include: "Can you work out hard words by yourself when you read?" (competence item), "Is reading to the class hard for you?" (difficulty item), and "Do you like word games in class?" (attitude item). The Reading Self-Concept Scale was administered to participants individually, whereby the researcher verbally delivered the questions, the participant provided verbal answers yes/no followed by a frequency rating on a Likert scale (1 – never and 5 – always). It takes approximately 15 minutes to complete. The McDonald's omega reliability in this sample for competence in reading items was $\omega = 0.73$, perceptions of difficulty with reading items was $\omega = 0.87$, and attitudes towards reading items was $\omega = 0.63$.

2.4 Procedure

The 8-week long reading intervention took place in the first half of the school year with the Grade 1 students in the first school, and then in the second half of the school year with the Grade 1 and Grade 2 students in the second school. Participants were first randomly assigned within classroom to the experimental condition (computer-based ABRA) or the control condition (paper ABRA) using a matched pairs random assignment procedure. This process involved inspecting the initial assessments of students on their baseline reading (WRAT 4 word reading sub-test), spelling (WRAT 4 spelling sub-test) and broader language abilities (GRADE listening comprehension sub-test) and then sorting students into pairs based on similar ability levels on these tests. Matching of all pairs was undertaken within the same classrooms as this provides a control for overall classroom effects. Each one of the pair of matched students was then randomly allocated to paper or computer ABRA condition using a random number generator (www.random.org). Students were divided into small instructional groups consisting of three to four students. All instructional groups within each condition (computer vs. paper) had a balance of male and female students. The gender distribution was 33.3% female and 66.7% male or 33.3% male and 66.7% female for most instructional groups, which consisted of three students each.

2.4.1 Reading Interventions

Prior to the intervention, all students provided verbal assent of their participation in the study. The reading intervention was conducted outside of the classroom in a quiet instructional space. Each group participated in the program for 15 minutes during a typical school day. Three sessions were held per week, for eight weeks, making a total of 24 sessions. For both groups, each session began with the experimenter defining one learning goal with each student. Then the experimenter instructed and supported the students in completing the learning activities, either on the computer or on paper. The computer-based and paper ABRA interventions were identical, apart from the delivery format.

2.4.2 Shared Components of the Computer-based and the Paper ABRA Conditions

Delivery format was the only difference between both conditions. For example, children’s stories were either in a digital format (e.g. e-book) or a paper format (e.g. paper book).

The curriculum in both cases was taken from the McNally et al. (2016) and Johnson et al. (2019) studies. The curriculum structure, organization and content were identical across the computer-based and paper ABRA conditions. There were twenty weekly session plans in the McNally et al. (2016) and Johnson et al. (2019) studies, with bi-weekly session plans for each children's story. In this study, six out of twenty weeks' content was covered, some 24 sessions carried out three times per week over a total of eight weeks (see Appendices A and B for the first 3 weeks of the computer-based and paper-based interventions respectively).

Daily session plans included a minimum of three to a maximum of five activities. These activities were either decoding (D), comprehension (C) or reading (R) tasks. These activities occurred in different combinations (e.g. DDRD, DDCD and CDRD). For each of these tasks, there was a specified length of time for completion. The total amount of time for completion of all these activities was fifteen minutes. Students' reading progress was closely monitored via regular session tracking forms. When students responded correctly 80 to 90% of the time as a group to an activity after three consecutive entries, they had reached the mastery level for that level and then moved on to the next level.

The pace of all of the ABRA intervention sessions was carefully controlled for both conditions (computer and paper). Each session was 15 minutes in length. At the start of each session, the researcher recorded the start time on the student progress tracking form. At the end of each session, the researcher recorded the end time on the same tracking form. During the intervention sessions, the researcher carefully timed each lesson component (e.g. alphabet, blending, tracking). The researcher also ensured participants were exposed to equivalent content (e.g. learning same number of letters/words, having equal opportunities to read parts of a story) and recorded the number of words taught during the blending and decoding activities, as per the lesson protocols. These steps were taken so that sufficient time was set aside to effectively instruct and fully engage participants throughout the lesson. The aim of controlling the pace of lessons including the number of trials delivered in activities delivered within each session was to ensure that the delivery and instruction of the computer-based and paper ABRA interventions were equal in terms of amount of content covered, as this potentially directly influences the effectiveness of the intervention and the quality of student learning.

Timing of lessons over the eight-week ABRA intervention by computer or paper condition was analyzed for each school to investigate whether lesson pacing was equivalent across both conditions. For the first school, there was no significant difference in overall pace of lessons between the computer condition ($M = 15.79$, $SD = 1.14$) and paper condition ($M = 15.76$, $SD = 1.08$), $t\ (14) = 0.05$, $p = 0.97$. For the second school, there was also no significant difference in overall pace of lessons between the computer condition ($M = 16.00$, $SD = 1.11$) and paper condition ($M = 16.54$, $SD = 1.67$), $t\ (14) = -0.76$, $p = 0.46$. This means that, overall, timing of lessons was effectively controlled for in the computer and paper conditions. This suggests that lessons were carried out for comparable lengths of time and equal amount of lesson content was covered in both conditions.

The overall number of words correctly recalled over the eight-week ABRA intervention by computer or paper was analyzed for each condition to investigate whether word exposure and word recall were equivalent across both conditions. The total number of words taught to participants was equivalent between the computer and paper ABRA conditions ($N = 50$). Between the computer-based and paper ABRA conditions, there were no significant differences in the number of taught words recalled $F(1, 98) = 2.29$, $p = 0.13$. This suggests that word exposure and learning was equivalent between both conditions, and confirming that the two conditions were not confounded on this feature of programme implementation.

2.4.3 Computer-based ABRA Condition

In the computer-based ABRA condition, the researcher facilitated sessions using a Macbook laptop. The ABRA program is located at the website address: https://grover.concordia.ca/abra/en/. The session plans were centred on a children's digital story. The digital stories were organized in a bi-weekly manner, where students spend two weeks reading and consolidating their knowledge and reading skills for each story. Bi-weekly session plans were structured in a way that students progressively build upon their storybook knowledge and reading skills through consistent exposure to a variety of reading activities. Small groups of students took turns interacting with the ABRA program on the computer. Examples of students' interactions included answering a question, identifying a sound or reading a sentence. Every effort was made to ensure that all students in each group had equal opportunities to interact with the ABRA program. If there was not enough time for students to all have a turn at an activity, then the researcher ensured that they had a turn next session.

2.4.4 Paper ABRA Condition

In the paper ABRA condition, the researcher facilitated sessions using a range of instructional materials, according to the paper ABRA curriculum based around the same bi-weekly story structure as in the computer-delivered ABRA intervention. Examples of paper instructional materials included: letter cards, mini whiteboards and stuffed toy animals. At the beginning of each session, the researcher set up the lesson's materials on a table ready for the students to begin their reading session. In all other respects it was identical to the computer-based ABRA condition in terms of content, progression, time allocated, and delivery structure.

2.5 Treatment Integrity

In this study, four measures were put in place to ensure treatment integrity: 1) keeping to 15 minute session timings, 2) keeping records of student progress, 3) moving students to higher levels when they achieved 80% response accuracy, and 4) scheduling school visits to conduct treatment integrity observations. In monitoring student progress, the researcher kept daily session logs of students' performance on blending and decoding activities. When students, as a group, answered correctly 80% of the time, then the researcher moved on to a higher difficulty level. In scheduling the school-based treatment integrity observations, the researcher followed these steps. The researcher first selected two university undergraduate or graduate students studying psychology or education to act as external observers. These two student observers were provided with an introduction to my study and training on how to observe and score the intervention sessions.

TI was evaluated using a TI observation checklist (McNally et al., 2016; Johnson et al., 2019). The TI observation checklist included 9 areas of assessment: 'exposure', 'planning', 'instructional guidance', 'opportunities to succeed (levels/differentiation)', 'group cohesion', 'pacing and efficient use of time', 'behaviour', 'adaptation/extension' and 'overall rating and comments'. 'Exposure' refers to the amount of instruction and amount of engagement students have with the intervention activities. 'Planning' is the extent to which preparation the intervention facilitator put in prior to the intervention session was visible. 'Instructional guidance' describes the quality and amount of facilitation provided for instructional activities. 'Opportunities to succeed (levels/differentiation)' requires the researcher to have an awareness of students' ability levels and demonstrate appropriate use of instructional activities to match students' abilities. 'Group cohesion' speaks to group dynamics and how well students work together on instructional activities. 'Pacing and efficient use of time' refers to whether the researcher planned the timing of the session and the instructional activities within the session accordingly. 'Behaviour' concerns how well students behaved during the intervention session (e.g. the presence of off-task or disruptive behaviours) and whether the researcher used effective behavioural management techniques to address disruptive behaviours. 'Adaptation/extension' describes whether the researcher delivered instructional activities with appropriate context or extension activities to reinforce the learning objectives of the session.

The TI evaluation form provides a number rating from 0 to 4 (0 being the lowest and 4 being the highest) for each of the 9 categories. TI observations were conducted on a randomly selected 20% of all intervention sessions by two observers. The overall inter-rater reliability (IRR) rate was 77%. Further analyses on the nine treatment integrity components were conducted to investigate whether treatment integrity outcomes were equal across both conditions. In all cases these contrasts were non-significant ($p < .05$) confirming that both interventions were equally well implemented.

3.0 Results

Preliminary data analyses were used to check all data for normality, skewness and kurtosis. These confirmed there were no deviations from normality, no floor and ceiling effects and no significant outliers in the data. Levene's test for homogeneity of variance was also non-significant in all cases for all variables reported below.

3.1 Results of the Group Matching Process

In order to confirm that the two groups were matched on relevant extraneous variables, independent samples *t*-tests were also conducted on participants' parental questionnaire data. Variables explored include the gender ratio of participants, participants' chronological age (years), parent-reported learning difficulties of their child participants' English reading frequency, participants' French reading frequency, mother's education level, mother's native language, mother-child home language, father's native language, and father-child home language. Results showed that the computer ABRA and paper ABRA groups were not significantly different from each other with regards to gender ratio, t (32) = -0.67, p = 0.51, chronological age (years), t (24) = 1.00, p = 0.33, parent-reported learning difficulties – the t statistic could not be computed because both groups are 0, English reading frequency, t (24) = -0.82, p = 0.42, French reading frequency, t (22) = 0.38, p = 0.71, mother's education level, t (24) = -0.58, p = 0.57, mother's native language, t (21) = 0.11, p = 0.91, mother-child home language, t (22) = -0.88, p = 0.39, father's native language, t (21) = -0.86, p = 0.40, and father-child home language, t (22) = -1.52, p = 0.14, which indicates that both groups were comparable on all these variables, and that there was low risk of selection bias in this study. The means and *SD*s for the pre-test scores and participants' parental questionnaire data are shown in Table 1 along with effect sizes (Cohen's *d*). Based on Cohen's established standards, effect sizes range from small (0.2), medium (0.5) to large (0.8) (Cohen, 1988). These effect sizes represent the variation in the magnitude of the ABRA intervention on improving students' literacy skills.

3.2 Descriptive Data Analyses

Table 2. Means, Standard Deviations and Cohen's *d* Effect Sizes for the Pretest and Posttest Literacy Measures by Intervention Group.

	Computer-based ABRACADABRA			Paper ABRACADABRA		
Measure	Pretest	Posttest	Effect Size	Pretest	Posttest	Effect Size
GRADE – Listening Comprehension[a]	13.59 (3.10)	15.41 (1.18)	0.69	14.00 (2.12)	15.53 (1.18)	0.58
WRAT – Reading[b]	99.76 (16.55)	111.00 (15.68)	0.66	96.29 (17.81)	108.76 (13.39)	0.73
WRAT – Spelling[b]	104.18 (17.29)	112.12 (15.03)	0.48	99.65 (15.81)	110.06 (12.07)	0.63
Reading Self-Concept Scale[a]	3.66 (0.38)	3.61 (0.45)	-0.14	3.57 (0.37)	3.46 (0.37)	-0.30
RSCS Difficulty	2.70 (0.87)	2.36 (0.78)	-0.12	2.82 (0.68)	2.51 (0.45)	-0.11
RSCS Competence	3.87 (0.76)	4.01 (0.78)	0.04	3.64 (0.71)	3.72 (0.55)	0.02
RSCS Attitude	4.42 (0.55)	4.44 (0.84)	0.00	4.24 (0.58)	4.17 (0.78)	-0.02

Note:

GRADE = Group Reading Assessment and Diagnostic Evaluation; WRAT = Wide Range Achievement Test IIII; RSCS Difficulty = Reading Self-Concept Difficulty Sub-scale; RSCS Competence = Reading Self-Concept Competence Sub-scale; RSCS Attitude = Concept Attitude Sub-scale;

[a]Values are represented by raw scores. [b]Values are represented by standard scores.

The means and standard deviations of ABRA intervention at pre- and post-test by modality of delivery are presented in Table 2 (please see above). Visual inspection of the means in Table 2 suggests substantial change in literacy outcomes between pre- and post-test that appears to be equivalent across both intervention conditions. More modest changes were evident between pre- and post-test for the motivation and self-esteem outcomes.

3.3 Inferential Analyses

First, independent samples *t*-tests were carried out to analyze participants' pre-test scores. The dependent variables included participants' pre-test scores on the GRADE – Listening

Comprehension sub-test, WRAT – Word Reading sub-test, WRAT – Spelling sub-test, Children's Reading Self-Concept Scale and Children's Author Recognition test. The independent variable was the ABRACADABRA intervention Condition (computer vs. paper). Results revealed that the computer ABRA and paper ABRA groups were not significantly different from each other at pre-test on listening comprehension, t (32) = -0.45, p = 0.66, reading, t (32) = 0.59, p = 0.56, spelling, t (32) = 0.98, p = 0.43, reading self-concept, t (32) = 0.56, p = 0.46, and reading enjoyment, t (32) = 0.43, p = 0.78. These results confirm that the matching process undertaken prior to randomization was successful in controlling for selection bias in primary outcome measures. A series of 2 Condition (computer-based ABRA vs. paper ABRA) x 2 Test (pre-test and post-test) mixed model ANOVAs were then conducted to compare participants' outcomes on a series of measures at pre-test and post-test. These analyses looked at the impact of intervention condition on listening comprehension, reading, spelling, reading self-concept (including reading motivation and reading self-esteem), and reading enjoyment.

In most cases there was a main effect of Test: listening comprehension, $\Lambda = 0.62$, F (1, 32) = 19.77, $p < 0.05$; reading, $\Lambda = 0.25$, F (1, 32) = 92.91, $p < 0.05$; spelling, $\Lambda = 0.37$, F (1, 32) = 53.92, $p < 0.05$; RSCS reading difficulty $\Lambda = 0.83$, F (1, 32) = 6.58, $p < 0.05$; RSCS competence, $\Lambda = 0.98$, F (1, 32) = 0.56, $p <.05$; Reading enjoyment, $\Lambda = 0.84$, F (1, 32) = 6.33, $p < 0.05$; though there was no main effect for RSCS overall, $\Lambda = 0.97$, F (1, 32) = 0.87, $p = 0.36$; or for RSCS attitude, $\Lambda = 1.00$, F (1, 32) = 0.02, $p = 0.89$. In no cases, however, were there any significant Test X Condition interactions: Listening comprehension, $\Lambda = 1.00$, F (1, 32) = 0.15, $p = 0.70$; reading, $\Lambda = 0.99$, F (1, 32) = 0.25, $p = 0.62$; spelling, $\Lambda = 0.97$, F (1, 32) = 0.98, $p = 0.33$; RSCS overall, $\Lambda = 1.00$, F (1, 32) = 0.71, $p = 0.79$; RSCS difficulty, $\Lambda = 1.00$, F (1, 32) = 0.01, $p = 0.93$; RSCS competence, $\Lambda = 1.00$, F (1, 32) = 0.05, $p = 0.83$; RSCS attitude, $\Lambda = 1.00$, *F* (1, 32) = 0.10, *p* = 0.76; and reading enjoyment Λ = 0.96, *F* (1, 32) = 1.49, *p* = 0.23. This latter finding suggests that there was no appreciable difference in the impact of different interventions. Finally, inspection of effect sizes in Table 2 also provided no reasons for concluding that there exists any differences across the per and computer-based modalities- all effect sizes were within .25 of their corresponding measure in the other condition, and generally were extremely modest.

In summary, this study presents a theory-driven intervention contrasting two otherwise matched computer- and pencil- based format within a well-matched sample that did not differ on a host of extraneous variables: gender and age; mother's education and mother and father's native language, and languages spoken in the family home. This sample also did not differ on a series of pre-test measures assessing participants' reading and spelling skills, reading motivation, reading self-concept and reading enjoyment. At post-test, participants showed improvement on measures including reading and spelling skills, and modest changes on measures of reading motivation and self-esteem. The data consistently suggested equivalent growth on these measures across both the computer-based and paper ABRA conditions.

4.0 Discussion

This pre- and post-test randomized control trial study aimed to investigate the following research questions: 1) What are the effects of the computer-based vs. paper ABRA instruction on students' reading skills (listening comprehension, letter reading and word reading) pre- and post-intervention, 2) What are the effects of the computer-based vs. paper ABRA instruction on students' spelling skills (letter writing and word spelling) pre- and post-intervention, and 3) What are the changes in students' reading motivation, reading self-esteem and reading enjoyment towards computer-based vs. paper ABRA reading activities pre- and post-intervention. Results showed that there was significant change between pre- and post-test in all attainment measures, but no effect of Condition suggesting change was equivalent across the computer-based and paper ABRA conditions. This pattern was also borne out in effect size analyses.

The non-significant differences between the computer-based and paper delivery of ABRA aligns with McNally et al.'s (2016) and Johnson et al.'s (2019) studies, which also found comparable positive effects of the computer-based (r = .183) and paper ABRA (r = .231) instruction on students' literacy skills. The difference between this study's findings and that of McNally et al.'s (2016) and Johnson et al.'s (2019) is that neither the computer-based nor the paper delivery were advantaged over the other in this study, whereas in those two studies, the paper delivery showed a somewhat larger size of effect and therefore was somewhat more effective in raising students' literacy outcomes.

Our findings alongside those of McNally et al. (2016) and Johnson et al. (2019) may indicate that when considered as group contrast, the medium of instruction in the ABRA program is secondary in importance compared to its curriculum content in this sample and age group. One methodological contribution of the present study was to carefully match content coverage within both modes of ABRA (computer-based and paper) was equivalent, ruling out this potential confound as a source of uncontrolled difference between the interventions. The broad finding of improvement after ABRA intervention is consistent with the pattern of the well-designed and well-balanced ABRA curriculum effectively promoting lower elementary students' literacy skills in Canada (Abrami et al., 2015; Savage et al., 2013). However, these findings cannot be attributed to ABRA specifically here in the absence of an unseen or alternate condition control. It should be noted however that children made greater than expected average progress (e.g. mean standard scores on reading and spelling at approximately 100 at pre-test and approximately 110 at post-test). This greater than expected growth might potentially be due to the value added by exposure to evidence-based ABRA content.

In terms of the social-emotional components of reading, there were very small negative effects of the computer-based and paper ABRA on Grade 1 and 2 students' reading self-concept (reading motivation and reading self-esteem). These patterns were consistent across both interventions. While reading motivation, reading self-esteem and reading attitude can be meaningfully related to reading (Chen & Savage, 2014; Hussein, 1999; Mihandoost, 2012), in the present context, however, there appears to be no particular advantage or disadvantage for technology in developing self-perceptions even when these students showed significant growth in attainment between pre- and post-test. Arguably, if these processes were significant drivers of development in one of the two interventions, then they would have shown up as meaningful differences in the results.

It should be noted however that these results do not rule out the possibility that for individual children that one modality of ABRA is more engaging or interesting than another, even within the age and ability levels sampled here. There are also arguably general unmeasured advantages for children that come with the computer-based ABRA such as basic proficiency with computer navigation and its use in group contexts, and some meta-learning about the nature of the structure of educational technologies. Arguably the web-based ABRA tool can be brought to contexts where the paper form - which is quite compendious, might be unwieldy. Computer-delivered ABRA may under some circumstances be more cost effective than color printing and resource maintenance of paper ABRA.

How might these results be interpreted in theoretical terms? According to the Time-Displacement Hypothesis of technology, time spent on literacy is zero-sum. On the assumption that technology has sub-optimal intrinsic value (Clark, 1983), it may displace other activities of greater educational value (Vandewater et al., 2006; Weis & Cerankosky, 2010). A modified Time-Displacement Hypothesis in the context of ABRA might suggest that the ‘game’ aspects of ABRA may displace more valuable paper-based instructional experiences evident in the content of a paper version of ABRA built as it is on evidence of ‘what works’ to improve literacy such as that reported in National Reading Panel (2000). We found no evidence of a negative effect of computer-based ABRA compared to the paper-delivered version, and as such these results provide no support for Clark’s position on technology or the Time-Displacement Hypothesis in this context.

4.1 Limitations

This study has a few limitations. Firstly, this study only had two groups (computer-based ABRA and paper ABRA). With a third regular classroom instruction group, this would serve as a control and provide additional data on the effectiveness of ABRA’s two delivery modes (computer and paper) in contrast to business as usual classroom instruction. Secondly, this study arguably had a modest sample size of 34 participants. In addition, this sample sustained a small loss to participants before post-test. With a larger sample size of participants, effect sizes resulting from participation in the different modes of ABRA would be more sensitive to the extent of ABRA’s effectiveness in facilitating literacy growth. Against this view, formal prospective power calculations showed that the study here was sufficiently powered to detect small-to-medium effects of intervention, and results of ANOVA analyses were supplemented by effect size analyses that were highly consistent with the findings of inferential analyses. Thirdly, this study was run for a total of eight weeks. This is not an untypical length of time for many published reading interventions, but to further enhance the effectiveness of ABRA on students’ literacy outcomes, extending the intervention period to twenty weeks, as in McNally et al.'s (2016) and Johnson et al.'s (2019) studies, could potentially be advantageous. Adding a delayed post-test would allow evaluation of longer-term effects of modality of delivery on relevant outcomes.

4.2 Conclusions

In conclusion, this study used a well-designed RCT assessing impacts of modality of intervention delivery contrasting paper- and computer-based delivery, carefully controlling the pace and scope of content delivery across formats. Across a range of academic and related socio-emotional and behaviour measures, there was no evidence of negative effects of computer-based versus paper ABRA modality on outcomes. We thus conclude that technology-based instruction does not necessarily have negative carryover effects on students' literacy skills, their motivation or self-concept.

5.0 References

Abrami, P. C., Borohkovski, E., & Lysenko, L. (2015). The effects of ABRACADABRA on reading outcomes: A meta-analysis of applied field research. *Journal of Interactive Learning Research*, *26*(4), 337-367.

Abrami, P. C., Wade, C. A., Lysenko, L., Marsh, J., & Gioko, A. (2016). Using educational technology to develop early literacy skills in Sub-Saharan Africa. *Education and Information Technologies*, *21*(4), 945-964.

Archer, K., Savage, R., Sanghera-Sidhu, S., Wood, E., Gottardo, A., & Chen, V. (2014). Examining the effectiveness of technology use in classrooms: A tertiary meta-analysis. *Computers & Education*, *78*, 140-149.

Aydemir, Z., & Oeztuerk, E. (2012). The effects of reading from the screen on the reading motivation levels of elementary 5th graders. *TOJET: The Turkish Online Journal of Educational Technology*, *11*(3).

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191.

Bavelier, D., Green, C. S., & Dye, M. W. (2010). Children, wired: For better and for worse. *Neuron*, *67*(5), 692-701.

Becker, H. J. (1992). Mindless or mindful use of integrated learning systems. *International Journal of Educational Research*, *21*(1), 65–79.

Becker, M., McElvany, N., & Kortenbruck, M. (2010). Intrinsic and extrinsic reading motivation as predictors of reading literacy: A longitudinal study. *Journal of Educational Psychology*, *102*(4), 773.

Blok, H., Oostdam, R., Otter, M. E., & Overmatt, M. (2002). Computer-assisted instruction in support of beginning reading instruction: A review. *Review of Educational Research*, *72*(1), 101–130.

Chapman, J. W., & Tunmer, W. E. (1995). Development of young children's reading self-concepts: An examination of emerging subcomponents and their relationship with reading achievement. *Journal of Educational Psychology*, *87*(1), 154.

Chaudhry, A. S. (2014). Student response to e-books: study of attitude toward reading among elementary school children in Kuwait. *The Electronic Library*, *32*(4), 458-472.

Chen, V., & Savage, R. S. (2014). Evidence for a simplicity principle: teaching common complex grapheme‐to‐phonemes improves reading and motivation in at‐risk readers. *Journal of Research in Reading*, *37*(2), 196-214.

Cheung, A., Mak, B., Abrami, P. C., Wade, A., & Lysenko, L. (2016). The effectiveness of the ABRACADABRA (ABRA) web-based literacy program on primary school students in Hong Kong. *Journal of Interactive Learning Research*, *27*(3), 219-245.

Cheung, A. C., & Slavin, R. E. (2012). How features of educational technology applications affect student reading outcomes: A meta-analysis. *Educational Research Review*, *7*(3), 198-215.

Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research*, *53*, 445-459.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). New York, NY: Lawrence Erlbaum Associates, Publishers. doi: http://www.utstat.toronto.edu/~brunner/oldclass/378f16/readings/CohenPower.pdf

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, *125*(6), 627.

Fletcher-Finn, C., & Gravatt, B. (1995). The efficacy of computer-assisted instruction (CAI): A meta-analysis. *Journal of Educational Computing Research*, *12*(3), 219–241.

Grant, A., Wood, E., Gottardo, A., Evans, M. A., Phillips, L., & Savage, R. S. (2012). Assessing the content and quality of commercially available reading software programs: Do they have the fundamental structures to promote the development of early reading skills in children?. *NHSA Dialog*, *15*(4), 319-342.

Greenlee- Moore, M. E., & Smith, L. L. (1996). Interactive computer software: The effects on young children's reading achievement. *Reading Psychology: An International Quarterly*, *17*(1), 43-64.

Grimshaw, S., Dungworth, N., McKnight, C., & Morris, A. (2007). Electronic books: Children's reading and comprehension. *British Journal of Educational Technology*, *38*(4), 583-599.

Gross, S. (2004). *Pathways to lasting self-esteem*. Bloomington, IN: House.

Guthrie, J. T., & Wigfield, A. (2000). Engagement and motivation in reading. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research, Volume III* (pp. 403-422). London: Lawrence Erlbaum Associates.

Harter, S. (2012). *The construction of the self: Developmental and sociocultural foundations*. New York, NT: Guilford Press.

Hidi, S. (2000). An interest researcher's perspective: The effects of extrinsic and intrinsic factors on motivation. In C. Sansone & J. M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance* (pp. 309–339). San Diego, CA: Academic Press.

Hussein, M. G. (1999). *The relationship between motivation to read and reading achievement in grades kindergarten through third* (Unpublished master's thesis). Walden University, United States of America.

Johnson, H., McNally, S., Rolfe, H., Ruiz-Valenzuela, J., Savage, R., Vousden, J., & Wood, C. (2019). Teaching assistants, computers and classroom management. *Labour Economics, 58*, 21-36. https://doi.org/10.1016/j.labeco.2019.02.006

Kozma, R. B. (1994). Will media influence learning? Reframing the debate. *Educational Technology Research and Development*, *42*(2), 7-19.

Kulik, J. A. (2003). *Effects of using instructional technology in elementary and secondary schools: What controlled evaluation studies say*. Arlington, VA: SRI International.

Kulik, C. L. C., & Kulik, J. A. (1991). Effectiveness of computer-based instruction: An update analysis. *Computers in Human Behavior*, *7*(1–2), 75–94.

McNally, S., Ruiz-Valenzuela, J., & Rolfe, H. (2016, October). ABRA: Online reading support (Rep.). doi:https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/ABRA_with_addendum.pdf

Mihandoost, Z. (2012). Relationship between motivation, self-concept, attitude and fluency of elementary school students. *Life Science Journal*, *9*(4), 9.

Moody, A. K., Justice, L. M., & Cabell, S. Q. (2010). Electronic versus traditional storybooks: Relative influence on preschool children's engagement and communication. *Journal of Early Childhood Literacy*, *10*(3), 294-313.

National Reading Panel (US), National Institute of Child Health, & Human Development (US).

(2000). *Report of the national reading panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. National Institute of Child Health and Human Development, National Institutes of Health.

Ouellette, G., & Tims, T. (2014). The write way to spell: Printing vs. typing effects on orthographic learning. *Frontiers in Psychology*, *5*, 117.

Ouyang, J. (1993). *Meta-analysis: CAI at the level of elementary education*. Paper presented at the World Conference on Education Multimedia and Hypermedia.

Savage, R. S., Abrami, P. C., Piquette, N. A., Wood, E., Deleveaux, G., Sanghera-Sidhu, S., & Burgos, G. (2013). A (Pan-Canadian) cluster randomized control effectiveness trial of the ABRACADABRA web-based literacy program. *Journal of Educational Psychology*, *105*(2), 310.

Savage, R. S., & Cloutier, E. (2017). Early reading interventions: The state of the practice and new directions in building causal theoretical models. In K. Cain, D. Compton, & R. Parrila (Eds.), *Theories of reading development* (pp. 409–436). Amsterdam, the Netherlands: John Benjamins.

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research*, *46*(3), 407-441.

Soe, K., Koki, S., & Chang, J. M. (2000). *Effect of computer assisted instruction (CAI) on reading achievement: A meta-analysis*. Washington, DC: OERI.

Thomson, J., Barzillai, M., van den Broek, P., & Schroeder, S. (2018). Learning to read in a digital world. *Learning to Read in a Digital World*, *17*, 225.

Vandewater, E. A., Bickham, D. S., & Lee, J. H. (2006). Time well spent? Relating television use to children's free-time activities. *Pediatrics, 117*(2), e181-e191.

Wang, J. H. Y., & Guthrie, J. T. (2004). Modeling the effects of intrinsic motivation, extrinsic motivation, amount of reading, and past reading achievement on text comprehension between US and Chinese students. *Reading Research Quarterly, 39*(2), 162-186.

Weis, R., & Cerankosky, B. C. (2010). Effects of video-game ownership on young boys academic and behavioural functioning: A randomized controlled study. *Psychological Science, 21*, 463-470. doi:10.1177?0956797610362670

Wilkinson, G. S., & Robertson, G. J. (2006). *WRAT 4: Wide Range Achievement Test*. Lutz, FL: Psychological Assessment Resources.

Williams, K. T. (2001). *Group reading assessment and diagnostic evaluation*. American Guidance Service.

Wolgemuth, J. R., Savage, R. S., Helmer, J., Harper, H., Lea, T., Abrami, P. C., ... & Louden, W. (2013). ABRACADABRA aids Indigenous and non-Indigenous early literacy in Australia: Evidence from a multisite randomized controlled trial. *Computers & Education*, *67*, 250-264.

Wollscheid, S., Sjaastad, J., & Tømte, C. (2016). The impact of digital devices vs. Pen (cil) and paper on primary school students' writing skills–A research review. *Computers & Education*, *95*, 19-35.

Wollscheid, S., Sjaastad, J., Tømte, C., & Løver, N. (2016). The effect of pen and paper or tablet computer on early writing–a pilot study. *Computers & Education*, *98*, 70-80.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. We would like to thank the staff, parents, and children of our participating schools.

Table 1. Matching Characteristics of the Intervention Sample by Condition.

Condition	Computer-based ABRACADABRA	Paper ABRACADABRA	Significance
Gender (% male)[a]	53	41	0.51 *ns*
Gender (% female)[a]	47	59	0.51 *ns*
Chronological age in years[a]	6.42	6.21	0.33 *ns*
Parent-reported learning difficulties[c]	0	0	0.00[d]
English reading frequency[a]	2.08 (0.95)	2.38 (0.96)	0.42 *ns*
French reading frequency[a]	3.09 (1.22)	2.92 (0.95)	0.71 *ns*
Mother's education[a]	3.15 (1.07)	3.38 (0.96)	0.57 *ns*
Mother's native language	2.33 (1.07)	2.27 (1.55)	0.91 *ns*
Mother-child language[a]	2.08 (0.90)	2.50 (1.38)	0.39 *ns*
Father's native language	2.17 (0.94)	2.64 (1.63)	0.40 *ns*
Father-child language[a]	1.83 (0.58)	2.33 (0.99)	0.14 *ns*
GRADE – Listening Comprehension[a]	13.59 (3.10)	14.00 (2.12)	0.66 *ns*
WRAT – Reading[b]	99.76 (16.55)	96.29 (17.81)	0.56 *ns*
WRAT – Spelling[b]	104.18 (17.29)	99.65 (15.81)	0.43 *ns*
Reading Self-Concept Scale[a]	3.66 (0.38)	3.57 (0.37)	0.46 *ns*
RSCS Difficulty	2.70 (0.87)	2.82 (0.68)	0.26 *ns*
RSCS Competence	3.87 (0.76)	3.64 (0.71)	0.89 *ns*
RSCS Attitude	4.42 (0.55)	4.24 (0.58)	0.32 *ns*

Note.
GRADE = Group Reading Assessment and Diagnostic Evaluation; WRAT = Wide Range Achievement Test IIII; RSCS Difficulty = Reading Self-Concept Difficulty Sub-scale; RSCS Competence = Reading Self-Concept Competence Sub-scale; RSCS Attitude = Reading Self-Concept Attitude Sub-scale; Recognition Test[a] = Children's Author Recognition Test
[a]Values are represented by raw scores. [b]Values are represented by standard scores.
[c]Values are represented by percentage. [d]Significance unable to be computed due to percentages being a value of 0.

Table 2. Means, Standard Deviations and Cohen's *d* Effect Sizes for the Pretest and Posttest Literacy Measures by Intervention Group.

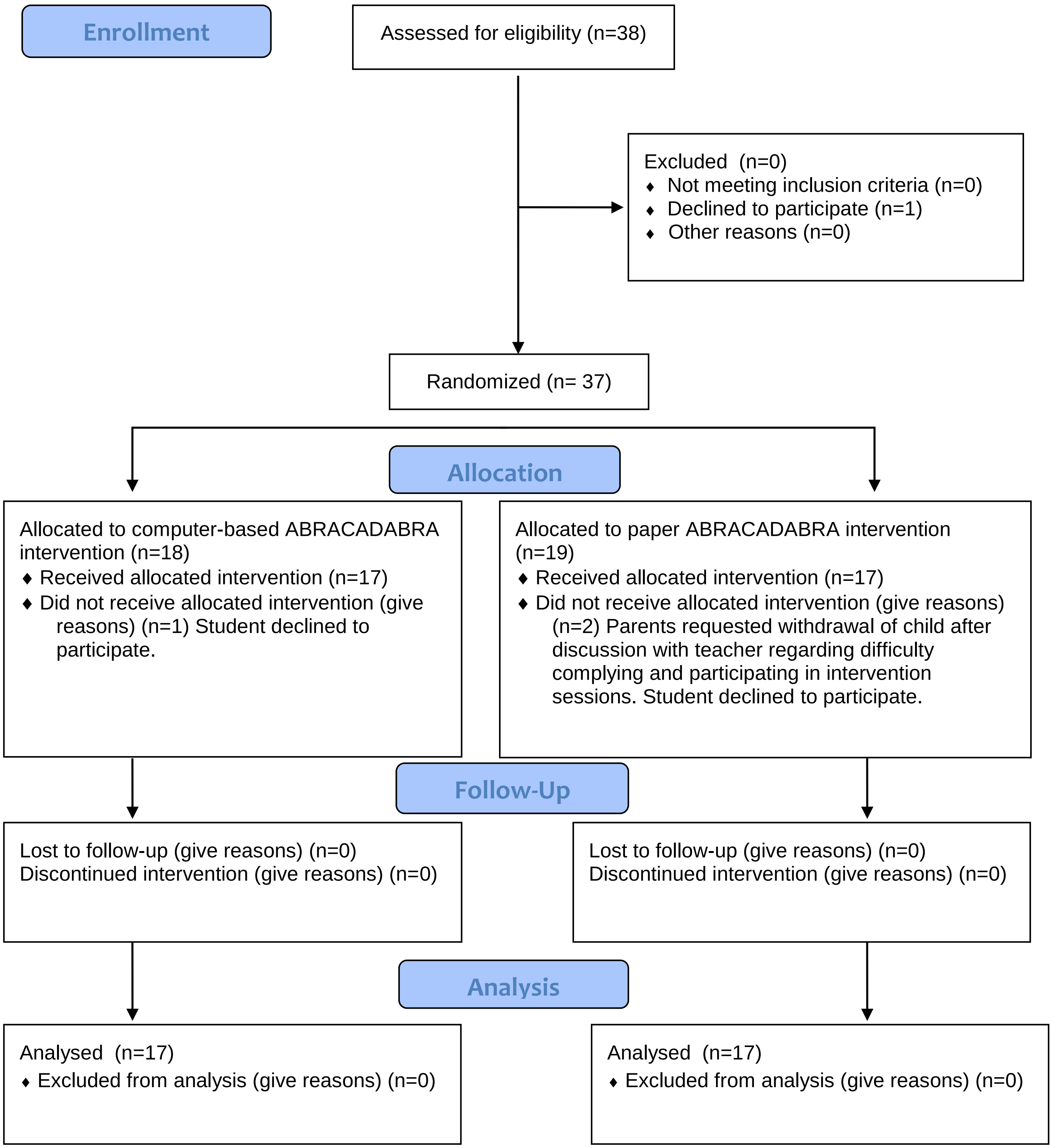
	Computer-based ABRACADABRA			Paper ABRACADABRA		
Measure	Pretest	Posttest	Effect Size	Pretest	Posttest	Effect Size
GRADE – Listening Comprehension[a]	13.59 (3.10)	15.41 (1.18)	0.69	14.00 (2.12)	15.53 (1.18)	0.58
WRAT – Reading[b]	99.76 (16.55)	111.00 (15.68)	0.66	96.29 (17.81)	108.76 (13.39)	0.73
WRAT – Spelling[b]	104.18 (17.29)	112.12 (15.03)	0.48	99.65 (15.81)	110.06 (12.07)	0.63
Reading Self-Concept Scale[a]	3.66 (0.38)	3.61 (0.45)	-0.14	3.57 (0.37)	3.46 (0.37)	-0.30
RSCS Difficulty	2.70 (0.87)	2.36 (0.78)	-0.12	2.82 (0.68)	2.51 (0.45)	-0.11
RSCS Competence	3.87 (0.76)	4.01 (0.78)	0.04	3.64 (0.71)	3.72 (0.55)	0.02
RSCS Attitude	4.42 (0.55)	4.44 (0.84)	0.00	4.24 (0.58)	4.17 (0.78)	-0.02

Note:

GRADE = Group Reading Assessment and Diagnostic Evaluation; WRAT = Wide Range Achievement Test IIII; RSCS Difficulty = Reading Self-Concept Difficulty Sub-scale; RSCS Competence = Reading Self-Concept Competence Sub-scale; RSCS Attitude = Concept Attitude Sub-scale;

[a]Values are represented by raw scores. [b]Values are represented by standard scores.

Figure 1. Consort flow diagram of participants.



Appendix A
Computer-based ABRACADABRA Curriculum (Weeks 1-3)

WEEK 1: Text – I Can Move Like A….

DAY 1		
Intro	Navigation	3 minutes
Intro	Group Rules (S*T*A*R)	5 minutes
Intro	Intro to ABRA (characters, stories)	7 minutes
DAY 2		
Intro	Recap Group Rules	2 minutes
Intro	Navigation Strategy	4 minutes
D	Letter Bingo	9 minutes
DAY 3		
D	Animated Alphabet – S and P	3 minutes
D	Auditory Blending – Level 1 (2 phonemes)	2 minutes
R	Tracking – First half of book	8 minutes
D	Auditory Blending – Level 1 (2 phonemes)	2 minutes
DAY 4		
D	Animated Alphabet – S and P	1 minute
D	Auditory Blending – Level 1	2 minutes
R	Tracking – Second half of the book	8 minutes
D	Auditory Blending – Level 1	2 minutes
D	Matching Sounds – Level 1	2 minutes

WEEK 2: Text – I Can Move Like A….

DAY 5		
D	Animated Alphabet – S,P,M	3 minutes
D	Auditory Blending (Level 2 if 80-90% accuracy achieved at Level 1)	2 minutes
R	Tracking – First half of book	8 minutes
D	Auditory Blending (Level 2 if 80-90% accuracy achieved at Level 1)	2 minutes
DAY 6		
D	Animated Alphabet – S and T	3 minutes
D	Blending Train – Level 1 or 2	2 minutes
R	Tracking – Second half of book	8 minutes
D	Blending Train – Level 1 or 2	2 minutes
DAY 7		
D	Animated Alphabet – P and T	3 minutes
D	Auditory Blending – Level 1 or 2	2 minutes
C	Vocabulary	8 minutes
D	Auditory Blending – Level 1 or 2	2 minutes
DAY 8		
C	Vocabulary (Finish)	6 minutes
D	Blending Train (Level 2 if 80-90% accuracy achieved at Level 1)	2 minutes
R	Tracking (Computer reads, children join in whole story)	5 minutes
D	Blending Train (Level 2 if 80-90% accuracy achieved at Level 1)	2 minutes

WEEK 3: Text – How A Bean Sprouts

DAY 9		
D	Animated Alphabet - S,P,M,T	3 minutes
D	Blending Train – Level 1/2	3 minutes
R	Tracking (Computer reads child choral co-reads, first 3-4 pages)	6 minutes
D	Blending Train – Level 1/2	3 minutes
DAY 10		
D	Animated Alphabet - L,R,M	3 minutes
D	Auditory Blending – Level 2	3 minutes
R	Tracking (Computer read child shares as above, last 3-4 pages)	6 minutes
D	Auditory Blending – Level 2	3 minutes
DAY 11		
D	Animated Alphabet – L,R,M	3 minutes
D	Auditory Blending – Level 2/3	3 minutes
C	Vocabulary	7 minutes
D	Auditory Blending – Level 3	2 minutes
DAY 12		
D	Animated Alphabet – L,R,M,I	3 minute
D	Auditory Blending – Level 3	2 minutes
C	Vocabulary	8 minutes
D	Auditory Blending – Level 3	2 minutes

Appendix B

Paper ABRACADABRA Curriculum (Weeks 1-3)

OVERVIEW WEEK 1

Book: I Can Move Like A……….

	DAY 1	
Intro	**Introduction to the Group**	**3 minutes**
Intro	**Group Rules**	**6 minutes**
Intro	**Letters in Name Activity**	**6 minutes**
	DAY 2	
Intro	**Recap Group Rules**	**2 minutes**
Intro	**Letters in Name Activity**	**4 minutes**
D	**Letter Bingo**	**9 minutes**
	DAY 3	
D	**Alphabet**	**3 minutes**
D	**Blending**	**2 minutes**
R	**Tracking**	**8 minutes**
D	**Blending**	**2 minutes**
	DAY 4	
D	**Alphabet**	**1 minute**
D	**Blending**	**2 minutes**
R	**Tracking**	**8 minutes**
D	**Blending**	**2 minutes**
D	**Same Sounds**	**2 minutes**

D = decoding tasks
C = comprehension tasks
R = reading tasks

OVERVIEW WEEK 2

Book: I Can Move Like A……….

		DAY 5	
D	Alphabet		3 minutes
D	Blending		2 minutes
R	Tracking		8 minutes
D	Blending		2 minutes
		DAY 6	
D	Alphabet		3 minutes
D	Blending		2 minutes
R	Tracking		8 minutes
D	Blending		2 minutes
		DAY 7	
D	Alphabet		3 minutes
D	Blending		2 minutes
C	Vocabulary		8 minutes
D	Blending		2 minutes
		DAY 8	
C	Vocabulary		6 minutes
D	Blending		2 minutes
R	Tracking		5 minutes
D	Blending		2 minutes

D = decoding tasks
C = comprehension tasks
R = reading tasks

OVERVIEW WEEK 3

Book: How a Bean Sprouts

		DAY 9	
D	Alphabet		3 minutes
D	Blending		3 minutes
R	Tracking		6 minutes
D	Blending		3 minutes
		DAY 10	
D	Alphabet		3 minutes
D	Blending		3 minutes
R	Tracking		6 minutes
D	Blending		3 minutes
		DAY 11	
D	Alphabet		3 minutes
D	Blending		3 minutes
C	Vocabulary		7 minutes
D	Blending		2 minutes
		DAY 12	
D	Alphabet		3 minute
D	Blending		2 minutes
C	Vocabulary		8 minutes
D	Blending		2 minutes

D = decoding tasks
C = comprehension tasks
R = reading tasks