

# A Meta-analysis of Multimedia Applications: How Effective Are Interventions with e-Books, Computer-Assisted Instruction and TV/Video on Literacy Learning?



Victor H. P. van Daal, Jenny Miglis Sandvik, and Herman J. Adèr

**Abstract** We examined how effectively multimedia applications (computer-assisted instruction, e-books, and TV/Video) benefit the literacy development of at-risk and not-at-risk children. Blok et al. (*Rev Educ Res* 72:101–130, 2002) analysed computer-assisted instruction studies undertaken in the 1990–2000 period and found an effect size of 0.254. Due to improvements in software and hardware over the 2000–2010 period, it was expected that the efficacy of multimedia applications had increased. Thirty-seven studies covering altogether 42 different treatments/experimental groups, in which in total 2525 children participated, were analysed. Eligibility criteria included quantitative research with participants up to 8 years old, which was published in English. An average effect size across all outcomes of .645 was found. Effects were larger when more time was spent on the task, and for pre-schoolers and kindergartners in comparison to first and second graders. Implications for the future development of smart phone and tablet applications are discussed.

**Keywords** Computer-assisted instruction · e-Books · TV/Video · Literacy · Meta-analysis

---

V. H. P. van Daal (✉)  
Faculty of Education, Edge Hill University, Ormskirk, UK  
e-mail: [vandaalv@edgehill.ac.uk](mailto:vandaalv@edgehill.ac.uk)

J. M. Sandvik  
Sandvik AS, Stavanger, Norway  
e-mail: [jenny@sandviks.com](mailto:jenny@sandviks.com)

H. J. Adèr  
Johannes van Kessel Advising, Huizen, The Netherlands  
e-mail: [herman.ader@kpnplanet.nl](mailto:herman.ader@kpnplanet.nl)

In this meta-analysis the efficiency of multimedia applications on literacy skills in developing young children was examined. In particular, we looked at computer-assisted instruction (CAI), picture storybooks presented on a computer with audio and video animations (e-books), and conventional TV/Video applications. The review is restricted to the 2000–2010 period in order to compare the results with a study that covered 1990–2000, an equally long period (Blok et al. 2002). Studies on learning to read and write in alphabetic languages were eligible.

## 1 Becoming Literate

Initially, all written words with the exception of a few words recognised from environmental print are completely unfamiliar to beginning readers. At school, children first learn how letters are pronounced, and then learn to read words by consecutively translating each letter (grapheme) into a sound (phoneme) and blending the sounds into a whole-word sound, a process called phonological recoding. Alternatively, look and say methods or a mix of decoding and whole word strategies are used for words such as *yacht*, the sixteenth century Dutch spelling. Thus, two processes are involved in word recognition: (1) phonological recoding, and (2) visual-orthographic look-up, coined by Coltheart (1984) as the dual route model of reading. Share (1995) speaks to the developmental aspects of the dual route model. He proposes that phonological recoding serves as a self-teaching mechanism for visual-orthographic look-up, enabling the beginning reader to proceed from slow deciphering through decoding to fast retrieval of word pronunciations through visual-orthographic look-up. The self-teaching hypothesis (Share 1995) contends that with every phonological recoding attempt, both the phonological (how the word is pronounced) and the orthographic (how the word is written) specifications will be strengthened in the lexicon.

The psychological process underpinning reading comprehension, the ultimate goal of reading, seems to be even more complicated. However, the assumption that reading comprehension builds on listening comprehension has proven to be a good starting point (Kintsch and Rawson 2005). According to these authors, comprehension largely depends on automatic processes that help us build up a representation of the text at hand. Automatic processes are processes that do not require conscious effort to execute them, such as listening comprehension (in one's native language). Another process that needs to be automatic is word recognition. Word identification processes need to be automatic in order to have resources available for understanding what the text is about. A text is represented at several levels, including a linguistic structure, a semantic representation, and a so-called situation model, that is, a mental model of what the text is about. Perfetti et al. (2005) suggest that the essential skills children should acquire include the following: (1) The parsing of meaning and form of sentences into a text representation; (2) Building up a situation model on the basis of the text representation; and (3) Drawing inferences, that is, making

the text coherent, because no text is completely explicit. Finally, the model developed by Perfetti and colleagues assumes that the real “bottleneck” in reading for meaning is decoding skill, that is, quick word recognition (see also Perfetti 1985).

## 2 Individual Differences in Reading Development

In the following section we describe where and how multimedia might benefit literacy learning, having first looked into developmental and behaviour-genetic studies of reading.

**Longitudinal Studies of the Development of Reading Skills** Stanovich (1986, 2000) conducted a series of studies to explain the ‘fan-spread’ effect on the variability of reading skill. He observed that students who start at a relatively high level of initial reading skill developed their skills much quicker than students who were less able when they started learning to read. He coined the term for this difference the ‘Matthew-effect’, from the biblical reference of the rich getting richer, the poor getting poorer. From recent research we know that the driving factor behind the increasing differences in reading skill is leisure time reading. More precisely, leisure-time reading activities were related to differences in the size of the vocabulary, and, in turn, vocabulary size promotes reading comprehension (Bast and Reitsma 1998).

Differences between students already exist when formal reading instruction starts, usually at the time they become 5, 6 or 7 years of age. It is clear that general cognitive skill is a powerful predictor of reading ability, as long as no specific skills for the effective processing of print are learnt, that is, when measured in kindergarten (Bowey 1995). Bowey (1995) and De Jong and Van der Leij (1999) explained with an assessment of vocabulary in kindergarten between 15% and 22% of the variance in reading in the first grade. Most probably, general cognitive ability contributes to reading success through efficient perceptual processes, such as being able to discriminate letters and sounds. Within normally developing children it is verbal ability at preschool age, rather than general cognitive ability, which determines later success at learning to read (Stanovich 2000). Subsequent studies have examined which specific aspects of verbal ability predict early reading achievement. Vocabulary predicts about 25% of the variance in end-of-first grade readers (Bowey 1995), whereas grammatical skills predict about 17% (Scarborough 1990). Phonological memory, commonly measured with a nonword repetition task (Baddeley and Gathercole 1992), predicts reading development in both deep (English) and relatively shallow orthographies, like Dutch (De Jong and Van der Leij 1999) and German (Naslund and Schneider 1996). Most of the research concentrating on speech perception and speech production has been carried out by Scarborough (1990) who found that errors in spontaneous speech in 30-month-old children predicted reading attainment in the second grade, and by Elbro et al. (1998)

who observed that the distinctness with which Danish children pronounced phonologically complex words predicted later reading success, even when effects of letter knowledge and other factors were controlled for.

Phonological sensitivity is perhaps the factor most researched. The initial finding that kindergartners' ability to count and manipulate phonemes and syllables in spoken words predicts later reading achievement (Mann and Liberman 1984) has led to an enormous amount of research not only in normally developing children, but also in children with dyslexia. The tasks typically require children to select a rhyming word with a given word, to say a word leaving out the last sound, or similar. Phonological skills play a relatively large role in learning to read in a deep orthography such as English, but are developmentally limited in shallow orthographies (Wimmer et al. 2000), that is, they are only relevant during a limited period (in the beginning of the year in which children start learning to read). Letter-name knowledge appears to be a very strong predictor of later reading achievement, explaining up to 36% of the variance in word identification at the end of the first year of reading instruction, especially when phonics reading programmes were used (Bowey 2005).

Finally, rapid automatised naming (RAN) has been a factor of much research interest. In RAN tasks a subject has to name as quickly as possible a continuous series of stimuli such as digits, common objects, colours, letters or words. There is still a debate over whether RAN is an independent contributor to early reading achievement over and above phonological skills. When assessed with digits and letters, it is likely that the effects are mediated through letter knowledge (Wagner et al. 1997).

**Behaviour-Genetic Studies of Reading** The power of behaviour-genetic studies in which monozygotic twins (MZ), who share 100% of their genes, are compared with dizygotic twins (DZ), who share about 50% of their genes, is that it facilitates an assessment of the genetic, shared environment, and non-shared environmental influences. An example of a shared environmental factor is, for example, the school, the teacher, and the reading method used. If one of the twins breaks a leg and misses school for some time is an example of a non-shared or unique environmental factor. If the correlation in DZ twins is more than half the MZ correlation, then there is an influence of the shared environment.<sup>1</sup> If the correlation is smaller, genetic factors play a relatively more important role. In short, behaviour-genetic studies can inform us of where teachers have the best chances to make a difference for their students and of where best to use technology, that is, where influences of the shared environment are relatively large. Behavioural genetics can also help us to find those components of reading skill that are only moderately or less heritable. These components depend much more on the environment and are sensitive to changes in the environment, for example, to teaching, training or intervention (with multimedia).

---

<sup>1</sup>Unique environment refers to the situation in which twins experience difference things, like attending different classes, one twin having an accident, etc. The unique environment term also comprises of measurement error, and is therefore hard to interpret.

***Behavioural-Genetic Studies of Decoding Skill*** With a genetically sensitive design in three different countries (U.S., Australia, and Norway and Sweden together), Samuelsson et al. (2007) looked at the contributions of phonological awareness (PA), rapid automatized naming (RAN), verbal memory, vocabulary, knowledge of grammar and morphology, and, knowledge of and experience with print to reading and spelling at the end of kindergarten. PA, RAN, and verbal memory showed substantial heritability, whereas knowledge of and experience with print and vocabulary showed strong influences of shared environment. Oliver et al. (2005) found similar results in a study conducted with a larger sample of twins in the UK.

***Behavioural-Genetic Studies of Reading Comprehension*** Byrne et al. (2009) replicated earlier findings that reading comprehension is substantially heritable and mostly determined by vocabulary, which has both substantial heritability and shared environment components in Grade 2. Keenan et al. (2006), working with older students in which the assessment of reading comprehension is less confounded with decoding skill, found that listening comprehension and word recognition (decoding) were the most important variables that independently drive reading comprehension.

### 3 Multimedia

Multimedia in the context of this meta-analysis refers to the integration of text, images, and sound presented electronically. Children, even very young children, are increasingly exposed to electronic media in the form of television, video, DVDs, computer programmes, electronic books, talking books, the internet, video games, tablet and smart phone applications, and interactive toys, to name a few.

As long as nearly 30 years ago, researchers called into question the efficacy of the prevailing teaching paradigm of one-dimensional, primarily verbal delivery of instruction (Clark and Paivio 1991) and recognised the potential for multimedia technologies to facilitate interactive learning opportunities. The National Association for the Education of Young People (NAEYC) issued a position statement acknowledging that “used appropriately, technology can enhance children’s cognitive and social abilities” and recommended that “computers should be integrated into early childhood practice physically, functionally, and philosophically” (NAEYC 1996, p. 2). An update was published in collaboration with the Fred Rogers Centre in January 2012 (<http://www.fredrogerscenter.org>). However, while some recognise the potential for multimedia to enhance learning, others debate the desirability of technology in early childhood education settings (Buckingham 2000; Lankshear and Knobel 2003; Stephen and Plowman 2003). Some argue that the use of technology in early childhood may not be developmentally appropriate, particularly in terms of cognitive overload (Kirschner 2002). Conversely, proponents of dual-coding theory maintain that the combination of visual with auditory stimuli results

in enhanced comprehension (Sadoski and Paivio 2007). Some reference teacher resistance to incorporating technology into lessons (Turbill 2001), while others argue that the cost of integrating technologies into classrooms, particularly those of young children, costs much and produces little in measurable educational gains (Yelland 2005). Still others go so far as to contend that the use of technology undermines the very nature of childhood (Buckingham 2000). Whether or not young children *should* engage with multimedia has been long debated. Nonetheless, it is clear that children *are*, in fact, *doing so* on a daily basis (Etta, [this volume](#); Rideout and Hamel 2006; Rideout 2014). Depending on which side of the debate one hails from, those who view technology as a powerful resource for early literacy enhancement, supporting ‘children of the digital age’ (Marsh 2005) or, alternatively, those who criticize technology as ‘the death of childhood’ (Buckingham 2000), a meta-analysis can tell us how effective multimedia applications are.

More importantly, it needs to be considered how technology and multimedia applications in particular might work, that is, how they actually might benefit literacy learning. Cheung and Slavin (2012) suggest that (new) technology might improve (1) the quality of instruction, because “content can be presented in a visual, varied, well-designed, and compelling way”; (2) the appropriate level of instruction because of the capacity to adapt the pace and level of the instruction to individual needs. Also, (3) the incentives to learn can be increased, as well as (4) the time on task and providing feedback.

**Reviews of Multimedia** Several literature reviews have attempted to provide an overview of the existing research on the topic (Courage [this volume](#); Hisrich and Blanchard 2009; Kamil et al. 2000; Lankshear and Knobel 2003; Plowman and Stephen 2003; Bus et al. [this volume](#); Yelland 2005; Zucker et al. 2009). See also recent reviews on the topic, Courage’s chapter and Bus, Sari, and Takacs’s chapter in this book. Kamil et al. (2000) undertook a comprehensive review of 350 articles including empirical studies and research reviews on the effects of multimedia on literacy. It was suggested that the use of multimedia facilitates children’s comprehension through ‘mental model building’, hypothesized to be a result of information presented as animation. Similarly, Lankshear and Knobel (2003) provided a synthesis of the research on the use of technology in promoting early literacy, focusing on young children. They found only 22 published articles that were relevant for review. In their quantitative assessment of the literature it was found that the research literature was unevenly distributed, with most focusing on the conventional aspects of reading such as decoding, rather than comprehension, or generating texts. Most significantly, they concluded that the effects of technology on early literacy development were “radically under-researched”. Likewise, Burnett’s (2009) literature review on literacy and technology in primary classrooms also noted a lack of research on the topic. A review of 38 studies published between 2000 and 2006, 22 quantitative and 16 qualitative, was conducted. It was concluded that the studies reviewed were limited in scope, as technology was used to support literacy in the same ways as print literacy, “assimilating technology by grafting it onto existing practices”, and therefore rendering the differential impact of multimedia on literacy development difficult to ascertain.

Recognising the need for research evidence on the topic, Zucker et al. (2009) provided a synthesis of studies published between 1997 and 2007 on the effects of electronic books (e-books) on the literacy outcomes of children from preschool through fifth grade. Seven randomized-trial studies and 20 quasi-experimental narrative studies met the selection criteria for their review. The aim of the study was to examine effects of e-books on children's comprehension and decoding-related skills, specifically in relation to emergent and beginning readers and children with reading disabilities or at risk of reading failure. Of the seven randomized-trial studies included, results of their meta-analysis showed small to medium effect sizes for comprehension. The effect on decoding was inconclusive, as only two studies that met the inclusion criteria examined it. The 20 studies included in the narrative review indicated mixed results. While it was found that e-books overall supported comprehension, they could, under some circumstances, actually undermine it (De Jong and Bus 2002). More recently, Cheung and Slavin (2012) found effect sizes of .37 for low-ability children, .27 for middle-ability children, and .08 for high-ability children, respectively, when reviewing 84 studies conducted in K12 over the period 1980–2010. Although these effects are small, it clearly indicates that those who need it most, benefit most: an indication that Matthew effects can be reversed!

## 4 Computer-Assisted Instruction

Since the late 60s computers have been used to assist in the teaching of reading and in the remediation of reading problems. Some computer programmes aim at practising a specific subskill of reading. Other programmes have been designed to combine the training of various subskills. An example of a combination of repeated reading, phonological awareness, and decoding is the WordBuild programme (McCandliss et al. 2003; Harm et al. 2003). The following two categories still seem to describe CAI for reading adequately: (1) computerised versions of basal reader programmes, and (2) tools that have especially been developed for (older) struggling readers.

**Computerised Versions of Basal Reader Programmes** These programmes come with a standard reading method and may differ from each other in several ways. In some reading methods the accompanying computer programme offers additional practice for struggling readers, in others all children go through the same programme, more or less in the same pace. The main characteristic is that these programmes contain several types of practice, usually from training phonological skills to text reading. More recently, reading and math programmes have been developed that keep motivation levels high by providing tasks that are not too easy nor too difficult for the individual learner (e.g., Klinkenberg et al. 2011).

***Tools Especially Developed for Struggling Readers and Older Persons with Dyslexia*** These programmes serve the purpose of supporting the user in reading, by reading aloud texts, such as Kurzweil 3000 (<http://www.kurzweilededu.com/>). Kurzweil 3000 offers also the possibility of scanning books while keeping the

original layout, including pictures, drawings, and tables. The spoken text can be exported as a MP3 file and then can be listened to everywhere, without the need to take a computer with you.

**Reviews of CAI** The Stanford project, aimed at a complete replacement of the teacher by a computer, was the first project to be evaluated. It did, however, not live up to the expectations (Fletcher and Atkinson 1972). The main reason that these reading programmes never would have become cost-efficient is because they ran on very expensive mainframe computers. Slavin (1991) evaluated IBM's Writing to Read programme in a meta-analysis study by looking at 29 studies and concluded that the efficiency of the programme was very low, that is, the costs in comparison to the learning effects were too high, a conclusion that is in line with other reviews (Krendl and Williams 1990).

Seven reviews that evaluate the use of CAI and beginning reading were published since 1990 (as far as we are aware). Two used a meta-analytic techniques and found effect sizes of 0.25 ( $SE = 0.07$ ) and 0.16 ( $SE = 0.08$ ), Kulik and Kulik (1991) and Ouyang (1993), respectively. Qualitative reviews were conducted by Torgesen and Horen (1992), Van der Leij (1994), Wise and Olson (1998) and by the National Reading Panel (2000), which were generally positive. However, Torgesen and Horen (1992) pointed out that much work needed to be done on the integration of the computer with the existing curriculum that was highly teacher-driven. The qualitative studies conducted by Van der Leij (1994) and Wise and Olson (1998) both concerned the use of computers with reading-disabled children. Van der Leij (1994) found that studies that concentrated on a specific subskill were generally more effective than multi-component programmes. Wise and Olson (1998) concluded that talking computers combined with phonological awareness training had a positive effect on learning outcomes, especially in children with relatively stronger phonological skills. The National Institute for Literacy report (2008) also concluded that talking computers show promise.

Although most of the recent studies seem to be positive about effects of CAI, the two studies that analysed effect sizes within a meta-analytic approach do however not give much reason for optimism, as mean effects of about .20 with a standard error of around 0.07 are reported. In the terminology of Cohen's (1988) these are small effects. However, it is likely that due to improvements in computer hardware and software and the integration of the computer in classroom learning activities, CAI has become more effective. Therefore, Blok et al. (2002) analysed studies undertaken in the 1990–2000 period. They categorised the studies, which all were concerned with beginning reading, along a variety of criteria in order to be able to find out what the elements are that make computer programmes work. In particular, they looked in 45 studies that reported on 75 experimental conditions at effect sizes and characteristics such as year of publication, language of instruction, experimental design (with or without control group, with or without pretest), subject assignment



(blocking, randomisation, matching, within-subjects), size of control and experimental group, population (normally developing, reading-disabled), age of participants at the beginning of the study, type of programme (phonological awareness, speech feedback, flash words, reading while listening, or mixed), duration of the programme in weeks and in hours, type of the dependent variable (phonological skills, letter identification, word accuracy, word speed, text accuracy, text reading speed, mixed), type of posttest score (observed score, gain score, score adjusted for covariates). The combined effect size was 0.254 with a standard error of 0.056. Experimental subjects thus were on the average 0.254 standard deviations better off than students in the control condition or compared with a baseline score. The variance of the effect sizes was 0.083, which means that there were considerable differences in effect sizes between the studies. Thirty-four per cent of the variance could be explained by entering the effect size at pretest into the equation. Language of instruction explained another 27% of the variance; studies conducted with English as medium of instruction obtained effect sizes that were 0.319 *SD* larger than non-English studies. No other variable was related to effect size at the posttest. The conclusions were very straightforward: computer-assisted instruction has little effect. As said, another 10 years of further developments in hardware and software has not produced any better results than in the decade before. The language effect comes, however, as a surprise. The authors explained it as an effect of the transparency of the language. If this explanation however were viable the same would be expected for the Danish studies (there were two Danish studies in the sample), because Danish is nearly as deep as English with respect to the orthography of the language (Seymour et al. 2003). The language effect may reflect that there is more room for improvement in deep orthographies, as reading development lags behind in deep orthographies compared with shallow orthographies.

## 5 Purpose of the Study

The aim of the current systematic review is to analyse the studies that were conducted after the Blok et al. (2002) review, that is, studies published between 2000 and 2010, an equally long period. The review was extended with e-books that became widely available during that period, together with TV/Video. Furthermore, defining characteristics of the studies associated with the effect sizes are examined. We expected that multimedia applications would be more effective than before, because of the following technological changes. Availability of the Internet in schools made it possible to have access to large databases of learning materials. Generally, also, video and audio animations improved, and, due to new programming methods, programming computers, tablets and smart phones became easier.

## 6 Method

### 6.1 Search Criteria

Specific key terms and phrases related to multimedia and early literacy were identified by reviewing the following reference books: *Handbook of Early Literacy Research* (Neuman and Dickinson 2001), *Handbook of Research on New Literacies* (Coiro et al. 2008), and *International Handbook of Literacy and Technology, Volume II* (McKenna et al. 2006). The first two authors independently devised key word search strings, and then cross-referenced these, resulting in the following list of primary search key words: *children, young children, children at risk, minority children, language minority children, cultural minority children, low SES children, disadvantaged children, children with reading disabilities, dyslexic children*. Secondary search key words were: *literacy, emergent and early literacy, reading, early and beginning reading, writing, early writing, beginning writing*. Finally, the following tertiary search key words were used: *media, multimedia, electronic media, digital media, technology, ICT, information technology, educational technology, interactive technology, digital books, on-line books, talking books, digital books, electronic books (e-books), CD-ROM, computers, computer-assisted learning, computer-based learning, CAI, internet, World Wide Web, television (educational television, children's television), Sesame Street, Between the Lions, DVD, mobile phones*.

### 6.2 Search Strategy

The Educational Resources Information Centre (ERIC) and PsychINFO were searched simultaneously using the aforementioned key word search strings. The broadest terms were input first and 'find all search terms', 'apply related words', and 'also search full text' were options selected in order to attain the highest number of hits. In PsychINFO, a selection was made to narrow the subject age range by selecting the age group 'childhood (birth – 12 years)'. These databases were then searched for peer-reviewed articles published in English between 2000 and 2010. In addition, the following key journals published in the same period were manually searched: *Journal of Early Childhood Literacy, Journal of Research in Reading, Journal of Early Childhood Literacy, Reading Research Quarterly, Early Childhood Research Quarterly, Journal of Literacy Research, Reading and Writing, Computers & Education, and Journal of Computer Assisted Learning*. Finally, the following special issues on technology and young children were searched: 'Technology in early childhood education' in *Early Education and Development* (Vol. 17, 1, 2006), 'Using technology as a teaching and learning tool' in *Young Children* (November 2003), 'Literacy and technology: Questions of relationship' in *Journal of Research*

*in Reading* (Vol. 32, 1, 2009), 'Technology special issue' in *Contemporary Issues in Early Childhood* (Vol. 3, 2, 2002), 'Technology and young children', downloaded from [www.technologyandyoungchildren.org](http://www.technologyandyoungchildren.org). References for several hundred potential studies were located. After reviewing the abstracts of each, 92 studies were acquired through the library or, if published in an E-journal, downloaded for further evaluation. The search and the review process were carried out by each of the first two authors independently and then cross-checked.

### **6.3 Selection of Relevant Studies**

Studies were included in the meta-analysis by meeting the following criteria, based on the content of the article abstract, if it provided the necessary information, and full-text, if the abstract was not sufficient. (1) Quantitative research on literacy interventions published in peer-reviewed journals between 1 January 2000 and 1 May 2010. (2) Studies in which participants were classified as 'early childhood', that is, subjects 0–8 years old. (3) Studies that included children at risk for literacy failure (e.g., dyslexia, low SES and/or language/cultural minority children). (4) Studies that included mainstream children. (5) Studies that measured *at least one* of the following literacy outcome variables: phonological awareness, reading comprehension, spelling, accuracy of reading words, accuracy of reading nonwords, fluency of reading, learning about print concepts, vocabulary learning, letter learning, rapid automatized naming and listening comprehension. (National Institute for Literacy 2008). (6) Studies that were published in English.

### **6.4 Coding**

The first two authors independently coded all studies as to the following study characteristics. (1) Age group of the participants. Categories included kindergarten, preschool and kindergarten, first graders, second graders, kindergarten through second graders, second and third graders. (2) Specificity of treatment. Studies were coded as either training one subskill or training more than one subskill. (3) Risk of reading failure: at-risk (low SES, second language learner, or reading failure) or not-at-risk students. (4) Language of instruction: English, Dutch, French or Hebrew. (5) Country in which the study was conducted: US, UK, Canada, Netherlands, France or Israel. (6) Media type: e-book, computer-assisted instruction, TV/Video. (7) Type of control group/treatment: traditional medium/curriculum, alternative reading treatment, alternative non-reading treatment (e.g., math), pretest used as baseline assessment or no-risk group used as control. (8) Grouping of participants for intervention/treatment: mixed groupings, individual, whole class, small groups.

(9) Type of test used to assess learning outcome: standardised test, experimental test. (10) Transfer of training: training to test, transfer of training/curriculum-based. (11) Duration of treatment in weeks. (12) Number of sessions over whole treatment period. (13) Average session duration in minutes. (14) Type of posttest score: raw observed, adjusted (e.g., for pretest score), transformed (e.g., standardised score). (15) Design – experimental: pretest-posttest untreated control group, posttest untreated control group (with gain-scores analysed), pretest-posttest control group with alternative reading treatment, posttest no control group, pretest-posttest no control group. (16) Design – statistical: between classes, within classes, between schools, within schools, counterbalancing within class. In addition, publication order was computed by using the year of publication (2000–2010) and the issue number (1–4 or 6) of the journals into a scale that ranged from 1 to 10. From 12 and 13 the total time-on-task in minutes was computed. For analysis purposes, this number was divided by 100 and centred around 10. See Table 1 for the coding of all studies.

**Coding of Literacy Outcomes** The selected studies were also coded for type of literacy outcome, according to generally accepted definitions (see Stanovich 2000). However, we have reported elsewhere about whether the various literacy outcomes are differentially affected by the use of multimedia applications (Van Daal and Sandvik 2013). The results are summarised in Appendix 2. In this paper the different literacy outcomes are amalgamated, see below.

*Phonological awareness (PA)* is defined as the ability to detect, manipulate, or analyse the auditory aspects of spoken language, including the ability to distinguish or segment words, syllables, or phonemes, independent of meaning. *Reading comprehension* is the ability to comprehend and recall a written story and to make inferences. Both conventional (‘write the word or the sentence’) and invented spelling tasks (for preschoolers) are used to tap *spelling ability*. The *accuracy of reading* is defined as the ability to correctly read real words, sentences or text. The *accuracy of reading nonwords* is defined as the ability to correctly read nonwords or low-frequency words. In some studies, lexical decision-making (decide whether a string of letters is a word or not) was used as a reading accuracy task. *Fluency of reading* is measured with timed reading of words, sentences or texts tasks. *Learning about print concepts* is defined as knowledge of print conventions (e.g., from left to right and from top to bottom of a page reading, and going through a book from front to back) and concepts such as book cover, author, and purpose of books. *Vocabulary learning* comprises of being able to use words actively and passively. *Letter learning* entails knowledge of the names and sounds associated with printed letters, including letter naming fluency, sound discrimination, and letter-sound relations. *Rapid automatized naming (RAN)* is defined as the ability to rapidly name a sequence of random letters, digits, colours, or objects. Finally, *listening comprehension* is the ability to comprehend and recall an oral story and to make inferences.

**Table 1** Study coding

Study number	Treatment/ condition	Author(s)	Publication order <sup>a</sup>	Age group	Specificity of treatment	Risk of reading failure	Language of instruction	Country	Media type	Type of control	Grouping
1	Sound – meaning	Brabham et al.	6.2	1	2	2	1	1	1	2	1
2	CD-ROM – tape	Brabham et al.	6.2	1	2	2	1	1	1	2	1
3		Cassady and Smith	3.8	1	2	1	1	1	2	1	1
4		Chambers et al.	8.2	3	2	3	1	1	3	2	3
5		Chera and Wood	3.2	2	1	1	1	2	1	3	2
6	Analytic phonics	Comaskey et al.	9.2	1	2	3	1	3	2	2	4
7	Synthetic phonics	Comaskey et al.	9.2	1	2	3	1	3	2	2	4
8		De Jong and Bus	4.8	2	1	1	2	4	2	4	2
9		Doty et al.	1.4	4	1	1	1	1	2	4	2
10		Ecalte et al.	9.6	3	1	4	3	5	2	2	2
11		Hecht et al.	2.4	1	2	2	1	1	2	4	2
12		Howell et al.	0.6	3	2	4	1	1	2	4	4
13		Karemaker et al.	10.2	3	2	1	1	2	2	1	3

(continued)

Table 1 (continued)

Study number	Treatment/condition	Author(s)	Publication order <sup>a</sup>	Age group	Specificity of treatment	Risk of reading failure	Language of instruction	Country	Media type	Type of control	Grouping
14		Karemaker et al.	10.2	3	2	4	1	2	2	1	4
15		Kegel et al.	9.8	1	1	4	2	4	1	1	2
16	Kindergarten	Korat	10.2	1	2	1	4	6	1	1	2
17	First graders	Korat	10.2	3	2	1	4	6	1	1	2
18		Korat and Shamir	7.4	1	2	2	4	6	1	1	2
19	At risk	Linebarger et al.	10.4	6	2	2	1	1	3	1	4
20	Not at risk	Linebarger et al.	10.4	6	2	1	1	1	3	1	4
21		Lonigan et al.	3.8	2	1	4	1	1	2	1	2
22		Macaruso and Walker	8.6	1	1	4	1	1	2	4	2
23		Macaruso et al.	6.4	3	1	4	1	1	2	4	2
24		Mathes et al.	1.4	3	1	4	1	1	2	4	4
25		McKenney and Voogt	9.2	1	1	1	2	4	2	1	2
26		Mioduser et al.	0.2	1	1	4	4	6	2	1	2
27		Mitchell and Fox	1.8	2	1	4	1	1	2	1	2

28		Regtvoort and Van der Leij	7.2	1	2	4	2	4	2	4	2
29	Synthetic phonics	Savage et al.	9.6	3	2	1	1	3	2	4	4
30	Analytic phonics	Savage et al.	9.6	3	2	1	1	3	2	4	4
31	Not at risk	Segers and Verhoeven	3.8	1	1	1	2	4	2	3	2
32	At risk	Segers and Verhoeven	3.8	1	1	3	2	4	2	3	2
33		Segers and Verhoeven	2.6	1	2	3	2	4	1	4	2
34		Segers et al.	4.4	1	2	3	2	4	1	1	2
35		Segers and Verhoeven	5.2	1	1	3	2	4	1	3	2
36		Shamir et al.	8.8	1	2	2	4	6	1	1	2
37		Silverman and Hines	9.4	5	1	3	1	1	3	1	3
38		Van Daal and Reitsma	0.4	1	2	1	2	4	2	1	2
39		Verhallen and Bus	10.2	1	1	3	2	4	1	3	2
40	Multimedia	Verhallen et al.	6.4	1	2	3	2	4	1	3	2
41	Static	Verhallen et al.	6.4	1	2	3	2	4	1	3	2
42		Wild	9.8	3	1	1	1	2	2	3	2

(continued)

Table 1 (continued)

Study number	Treatment/condition	Author(s)	Publication order <sup>a</sup>	Age group	Specificity of treatment	Risk of reading failure	Language of instruction	Country	Media type	Type of control	Grouping
1	Sound – meaning	Brabham et al.	1	1	4	20			1	3	1
2	CD-ROM – tape	Brabham et al.	1	1	4	20			1	3	1
3		Cassady and Smith	1	1	40	80	20	1600	1	1	3
4		Chambers et al.	1	1	28		90		2	2	4
5		Chera and Wood	2		4	10	10	100	1	1	4
6	Analytic phonics	Comaskey et al.	1	1	13	39	15	585	1	3	2
7	Synthetic phonics	Comaskey et al.	1	1	13	39	15	585	1	3	2
8		De Jong and Bus	2			12	15	180	3	4	2
9		Doxy et al.	2	1					1	1	1
10		Ecalle et al.	1	1	5	10	15	150	1	1	2
11		Hecht et al.	2	1	24		15		1	1	3
12		Howell et al.	1	1	16	74	30	2220	1	1	2



13		Karemaker et al.	1	1	1	5	60	300	1	4	2
14		Karemaker et al.	1	1	5	2	60	120	1	4	5
15		Kegel et al.	2	1	16	15	10	150	1	1	4
16	Kindergarten	Korat	2	2					1	1	2
17	First graders	Korat	2	2					1	1	2
18		Korat and Shamir	2	2	3	3	30	90	1	1	2
19	At risk	Linebarger et al.	2	2					1	4	2
20	Not at risk	Linebarger et al.	2	2					1	4	2
21		Lonigan et al.	1	1	8				1	1	2
22		Macaruso and Walker	1	1	24	48	20	960	1	4	4
23		Macaruso et al.	1	1	24	64	20	1280	1	1	4
24		Mathes et al.	1	1	8	32	25	800	1	2	2
25		McKenney and Voogt	1	1	5	4	20	80	1	1	2
26		Mioduser et al.	2	1					1	2	2
27		Mitchell and Fox	1	1	4	15	20	300	1	1	2

(continued)

Table 1 (continued)

Study number	Treatment/condition	Author(s)	Publication order <sup>a</sup>	Age group	Specificity of treatment	Risk of reading failure	Language of instruction	Country	Media type	Type of control	Grouping
28		Regtvoort and Van der Leij	1	1	14	70	10	700	1	1	2
29	Synthetic phonics	Savage et al.	1	1	12	48	20	960	1	3	2
30	Analytic phonics	Savage et al.	1	1	12	48	20	960	1	3	2
31	Not at risk	Segers and Verhoeven	2	2	15	30	15	450	1	1	3
32	At risk	Segers and Verhoeven	2	2	15	30	15	450	1	1	3
33		Segers and Verhoeven	2	2		3	25	75	1	5	
34		Segers et al.	2	2		1			1	1	3
35		Segers and Verhoeven	1	1	40	40	15	600	1	1	3
36		Shamir et al.	1	1					1	1	2

37		Silverman and Hines	2	2	12	36	45	1620	1	1	2
38		Van Daal and Reitsma	1	1					1	1	2
39		Verhallen and Bus	2	2					1	1	2
40	Video	Verhallen et al.	2	2		1	6	6	1	3	2
41	Static	Verhallen et al.	2	2		1	6	6	1	3	2
42		Wild	1	1	10	20	20	400	1	4	2

<sup>a</sup>Publication order: 0–11 (rescaled conversion from publication year and volume number of journal. Age group: (1) kindergarten, (2) kindergarten and first graders, (3) first graders, (4) second graders, (5) kindergarten through second graders, (6) second and third graders. Specificity of treatment: (1) one subskill, (2) more than one subskill. Risk of reading failure: (1) not at risk (2) low SES, (3) second language learners, (4) reading failure. Language of instruction: (1) English, (2) Dutch, (3) French, (4) Hebrew. Country: (1) US, (2) UK, (3) Canada, (4) Netherlands, (5) France, (6) Israel. Media type: (1) e-book, (2) computer-assisted instruction, (3) TV/Video. Type of control: (1) traditional medium/curriculum, (2) alternative reading treatment, (3) alternative non-reading treatment, (4) pretest used as baseline assessment or no-risk group used as control. Grouping: (1) mixed groupings, (2) individual, (3) whole class, (4) small groups. Type of test: (1) standardised test, (2) experimental test. Transfer of training: (1) training to test, (2) transfer of training/curriculum-based. Duration of treatment: weeks. Number of sessions. Average session duration: minutes. Time-on-task: minutes. Type of posttest score: (1) raw observed, (2) adjusted, (3) transformed. Design – experimental: (1) pretest-posttest control group, (2) pretest-posttest experimental group only. Design – statistical: (1) between classes, (2) within classes, (3) between schools, (4) within schools, (5) counterbalancing within class. Empty cells: missing or n/a. See also text.

## 7 Results

### 7.1 *Descriptive Statistics*

After reviewing the abstracts or full-text of each article collected, 51 studies met the selection criteria.<sup>2</sup> If studies included more than one treatment or more than one experimental group, we treated them as separate studies. Nine studies were later excluded for missing relevant statistics (number of participants, means, standard deviations, or non-aggregated statistics; only five corresponding authors replied positively to our request to supply us with missing statistics). Twenty-eight articles reported on single studies, whereas seven contained multiple treatments/experimental groups. Of the remaining 42 studies, 26 studies included children at risk of reading failure. Of the studies of children at risk, 11 studies reported on interventions with second language learners, most stemming from cultural or language minority groups, six studies included children of low socio-economic status, and nine studies dealt with underachieving readers. Twelve studies on the effects of multimedia interventions in mainstream children were found. The 35 studies that were submitted to the meta-analysis are marked with an asterisk (\*) in the References.

The majority of studies were conducted in English-speaking countries, USA (15), UK (4), and Canada (4). Thirteen studies were conducted in The Netherlands (Dutch), one in France and five in Israel (Hebrew). Two studies dealt with embedded multimedia (TV/Video) in teachers' reading lessons, two with subtitled video, 14 with e-books, and 24 with Computer-Assisted Instruction. Most of the studies were published in the last 16 months of the period we examined (16), five in 2003 and five in 2006, whilst other publications were evenly spread over the other years. Thirty-eight studies were carried out with participants from preschool and kindergarten. About half of the studies trained a single subskill (18). Seventeen studies used the traditional medium/curriculum, six an alternative reading treatment, eight an alternative non-reading treatment and 11 a pretest baseline or a no-risk group as a control condition. Twenty-seven studies provided an individual treatment. Twenty-three studies used a standardised test to assess the learning outcomes, whereas 19 studies used experimental tests. Twenty-seven studies trained to the test, whereas 13 aimed at transfer or tested targets from the existing curriculum, whilst two studies were unclear about what sort of test was used. Duration of the treatment varied from 3 to 40 weeks, whereas the number of sessions varied from 1 to 74 with average session duration varying from 6 to 90 min. The intensity of the training in terms of total time-on-task varied between 6 and 2220 min. All but three studies analysed raw observed scores, whereas eight studies did not include a control group at all. Finally, three studies compared treatments between classes, 26 within classes,

---

<sup>2</sup>Our searches produced only one reference to a study on the Fast ForWord intervention programme. As this programme has extensively been evaluated by others without finding any effects, we decided not to include this study (which didn't find any effects either). See What Works Clearinghouse (2006, 2007) and Strong et al. (2010).

6 between schools, and 5 within schools. In one study treatments were counterbalanced within classes. In total 2525 children participated across all studies, 1201 as experimental subjects (on the average 28.6 per study) and 1324 as control subjects (on the average 31.5 per study).

## 7.2 *Meta-analysis*

For each treatment/experimental group/literacy outcome Hedge's  $g$ , to account for small sample sizes, was computed, that is, the difference between the means (of either the experimental group and the control group, or the posttest and the pretest in case there was no control group) was divided by the pooled standard deviations, as different units of measurement were used across studies (see Cornell and Mulrow 1999). Within each study, the effect sizes of multiple literacy outcomes were averaged. In Table 2 multiple and averaged effect sizes of all studies are presented, together with the numbers of participants of the studies.

However, it was first checked whether literacy outcomes could be averaged without losing information by running a principal component analysis on the results of two studies in our sample that contained the widest range of literacy outcomes. In the Savage et al. (2009) study in total 11 outcome measures were taken, of which the raw data were made available to us. We combined the two measures for PA (elision and blending) and the two for RAN (objects and letters). All nine remaining outcomes loaded between .466 and .878 on one factor that explained 58.3% of the variance. Steve Hecht ran a similar analysis on the primary data set of the Hecht and Close (2002) study and kindly shared the SPSS output with us. Six assessments explained 58.5 of the total variance and loaded between .631 and .870 on one single factor. Although there were differences between the studies as to instruments used, the age of the participants, and the types of computer programmes, the results of the principal component analyses, which both examined effects of analytic and synthetic phonics in both studies, definitely converge. It was therefore concluded that it is appropriate to average literacy outcomes within studies. Because the factor contains both outcomes that are close to reading (e.g., PA, reading fluency) and literacy (e.g., listening comprehension, vocabulary), we prefer to keep using the term 'literacy outcomes'. The results of both principal component analyses are presented in Appendix 1.

## 7.3 *Multilevel Modelling*

Multilevel modelling (MlwiN) was used to assess the effect that study characteristics have on the effect sizes reported in the studies (Rasbash et al. 2005). For this analysis, studies were regarded as nested under publication year. An average effect size of .645 ( $SE = .112$ ) was found, whilst 19.53% (.126,  $SE = .085$ ) of the total

**Table 2** Aggregated effect sizes and effects sizes for different learning outcomes

Study number	Treatment/condition	Author(s) (n exp-control)	d <sup>a</sup>	PA	Comp	Spell	Word	Nword	Fluency	P-concepts	Voc	Letter	RAN	List comp
1	Sound – meaning	Brabham et al. (25–55)	-0.04	0.18			0.16				-0.09	-0.42		
2	CD-ROM – tape	Brabham et al. (42–30)	-0.26	-0.67			-0.14				-0.02	-0.22		
3		Cassady and Smith (26–62)	0.50	0.64						0.37				
4		Chambers et al. (43–56)	0.56		0.56		0.75	0.46				0.47		
5		Chera and Wood (15–15)	0.07	0.37			-0.23							
6	Analytic phonics	Comaskey et al. (27–27)	1.01	0.73			0.64	0.76				1.92		
7	Synthetic phonics	Comaskey et al. (26–26)	1.29	1.49			0.80	0.91				1.94		
8		De Jong and Bus (18–18)	2.25		2.25									
9		Doty et al. (20–19)	1.14		1.14									
10		Ecalie et al. (14–14)	1.26			1.22	1.29							
11		Hecht et al. (42–34)	0.76	1.14		1.20	1.11			0.23	0.24	0.14		
12		Howell et al. (35–35)	1.55	1.77		0.80	2.06							
13		Karemaker et al. (61–61)	0.21	0.18			0.23							
14		Karemaker et al. (17–17)	0.13	-0.03			0.43	-0.01						
15		Kegel et al. (45–45)	0.49	0.24		0.73								
16	Kindergarten	Korat (20–20)	1.13				0.62				1.64			
17	First graders	Korat (25–25)	1.16				0.97				1.35			
18		Korat and S. (25–14)	0.42	0.19			0.10				0.96			

19	At risk	Linebarger et al. (6–8)	1.15	1.92	1.65	1.01	0.02					
20	Not at risk	Linebarger et al. (16–14)	0.30	1.21	-1.01	1.29	-0.28					
21		Lonigan et al. (20–21)	0.42	0.42								
22		Macaruso and Walker (12–12)	1.23	1.21			0.66	1.80				0.49
23		Macaruso et al. (15–15)	0.43		0.31				0.55			
24		Mathes et al. (43–42)	-0.07	-0.28	-0.16	0.12	0.05					
25		McKenney and Voogt (19–18)	0.63		0.63							
26		Mioduser et al. (16–15)	0.67	1.65	0.05				0.30			
27		Mitchell and Fox (24–24)	0.70	0.70								
28		Regtvoort and v.d. Leij (31–16)	0.35	0.29					0.41	0.55		
29	Synthetic phonics	Savage et al. (43–57)	0.20	0.41		0.20	-0.06	0.18	0.20	-0.12	0.46	
30	Analytic phonics	Savage et al. (44–57)	0.14	0.24		0.10	-0.20	0.13	0.37	-0.22	0.39	
31	Not at risk	Segers and Verhoeven (67–97)	0.41					0.41				
32	At risk	Segers and Verhoeven (67–97)	0.04					0.04				
33		Segers and Verhoeven (25–25)	0.87					0.87				

(continued)

Table 2 (continued)

Study number	Treatment/condition	Author(s) (n exp-control)	$d^a$	PA	Comp	Spell	Word	Nword	Fluency	P-concepts	Voc	Letter	RAN	List comp
34		Segers et al. (30–30)	-0.05		-0.27						0.17			
35		Segers and Verhoeven (26–36)	0.26	0.26								0.26		
36		Shamir et al. (27–27)	0.30	0.17			0.34			0.38				
37		Silverman and Hines (42–42)	0.37								0.37			
38		Van Daal and Reitsma (9–9)	0.53				0.51	0.40		0.31		0.89		
39		Verhallen and Bus (29–29)	0.97								0.97			
40	Video	Verhallen et al. (10–10)	1.87		1.55									
41	Static	Verhallen et al. (10–10)	0.80		0.64						2.19			
42		Wild (44–40)	0.16	0.30		0.02								

<sup>a</sup> $d$ : aggregated effect size, PA phonological awareness, Comp reading comprehension, Spell spelling, Word word reading accuracy, Nword nonword reading accuracy, Fluency reading fluency, P-concepts print concepts, Voc vocabulary, Letter letter learning, RAN rapid automatized naming, List Comp listening comprehension. See also text



**Table 3** Results of multilevel modelling for the effect of study characteristics on effect sizes

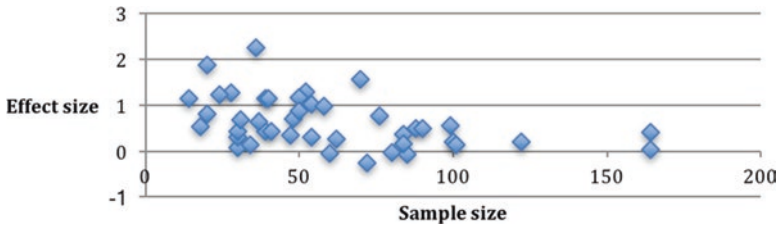
Variable	Parameter estimate	SE
Number of sessions (range: 1–74)	–.0354	.010
Time-on-task (divided by 100, centred at 10; max. = 22.2)	.153	.048
Age: Older than Kindergarten	–.740	.239
Control: Alternative reading treatment	1.139	.357
Control: Pretest as baseline or no-risk group	.758	.265
Design: Gain scores	–.881	.410

variation in effect sizes was explained by the year, in which the study was published and 31% (.200,  $SE = .064$ ) was due to differences between studies. In Table 3 the parameter estimates and the standard errors of the estimates of the final model with only significant effects are presented.

Factors that positively affect effect size include total time-on-task (an increase of .153 with every 100 min more,  $SE = .051$ ), which was slightly moderated by the number of sessions the participants engaged in (a decrease of –.035 with every additional session,  $SE = .010$ ). Effects are .854 larger for preschoolers and kindergartners in comparison to first graders and older children ( $SE = .248$ ). In comparison with studies in which the traditional medium or curriculum forms the control condition, effects sizes are larger if the control condition consists of an alternative reading intervention (1.139,  $SE = .357$ ) and if there is a pretest as base-line or if a no-risk group is used as the control group (.758,  $SE = .266$ ). A study design in which gain scores are analysed gave smaller effect sizes (–.882,  $SE = .410$ ).

Effect sizes are not influenced if the control condition consists of a non-reading task. Publication year, specificity of the training, type of risk factor, language of instruction, media type, grouping of the participants, type of test used, type of scores analysed, and design (statistical comparison) all did not affect the effect sizes of the studies. Nor did any interaction between significant parameters in the final model.

Finally, we checked whether publication bias affects the current meta-analysis. Publication bias refers to a tendency to publish studies with significant results, thus with sizable effect sizes. The presence of publication bias is assessed by examining the correlation of the effect size of studies with a measure of precision, such as sample size, standard error, or the inverse of the standard error (Cornell and Mulrow 1999). This can be done by visually inspecting the scatterplot of the correlation and by statistically testing the correlation under the assumption that studies are symmetrically distributed in a funnel shape with precise studies having less variable and less precise studies having more variable effect sizes, if publication bias is absent. In Fig. 1 effect sizes of the primary studies are plotted against the sample sizes. Visual inspection shows that there are relative few studies with many (over 80) participants. A funnel-like shape can be recognised in the studies with less than 80 participants. For these 30 studies the Kendall rank correlation is –.130 ( $p = .317$ ). Given the relative high  $p$ -value, it is unlikely, even given a relatively small number



**Fig. 1** Standardised effect sizes plotted against sample sizes

of primary studies, that publication bias forms a threat to the validity of this meta-analysis. However, the studies with large numbers of participants show clearly no funnel shape. This may well be due to the fact that very few large-scale studies can be conducted at all, due to financial constraints.

## 8 Discussion

This systematic review sought to assess how effective multimedia applications were in the 2000–2010 period, in which major developments in hard- and software took place. In addition, this study examined which characteristics of the primary studies are positively related to the effect sizes obtained. The hypothesis that multimedia applications would be more effective than before, is supported, as a medium overall effect of .65 was obtained, which is substantially larger than reported by Blok et al. (2002) in their review of CAI and by Zucker et al. (2009), who reviewed the efficiency of e-books. Moreover, this study shows that effects can and have been replicated in non-English speaking countries, though on a small scale. It also complements a previous report (Van Daal and Sandvik 2013), in which the effect of multimedia applications on specific literacy outcomes was evaluated.

Nineteen percent of variation in effect sizes in the current study can be ascribed to the year in which the study was published, whilst 31% reflected overall differences in effect size between the primary studies. Note that effect sizes vary between studies according to year of publication, but there is no significant association between effect size and year of publication. Time-on-task and being a preschooler or kindergartner were positively related to effect size obtained in the primary study. Three aspects of the design affected the effect size of studies. Larger effect sizes were obtained in studies that compared interventions/treatments with a traditional medium or curriculum. Also, effect sizes were larger for studies that used a pretest-posttest design without a control condition or took a no-risk group for comparison. Smaller effects were obtained, if gain scores were analysed.

The largest effect size, 2.25 on a comprehension measure, was obtained in a study by De Jong and Bus (2004), which compared effects of electronic books and being read aloud by parents with a counterbalanced design. In 10 other studies aggregated effect sizes were greater than 1. In most studies with multiple literacy

outcomes a considerable variation in effect sizes across literacy outcomes was found. This is probably due to different contents and different forms of practice. For example, in the Comaskey et al. (2009) study both the analytic and the synthetic phonics training was very effective for letter learning and PA but less so for word and nonword reading, whilst in the Hecht and Close (2002) study a *combination* of analytic and synthetic phonics training was more effective for word reading and PA but not for letter learning.

In contrast to these two studies, experimental groups were often compared with a control group that did ‘nothing’, which may produce inflated effect sizes. A more telling comparison would be to look at the so-called ‘added value’ of multimedia applications. Several of the studies included in this review offer such a possibility in addition to the aforementioned studies that compared multimedia interventions with regular classroom instruction. For example, in the study by Chambers et al. (2008) computer-assisted tutoring was compared with embedded multimedia. Effect sizes were larger for embedded multimedia than for computer-assisted tutoring with respect to comprehension (.56), word reading (.75), nonword reading (.46), and letter learning (.47). Another way to learn more about how multimedia may work is to include different kinds of experimental groups, as Verhallen et al. (2006) did. The effect size for the experimental group that was presented video pictures was larger than the effect size of the experimental group that was presented static pictures. In addition, the current study clearly showed that larger effect sizes were obtained in studies that compared the experimental group with an alternative reading treatment group, or using the pretest as a baseline, if there is no control group, or using a no-risk group as a control group. The latter should be positively interpreted: if at-risk children can catch up with their not-at-risk peers with help of multimedia, Matthew effects can be turned around (Stanovich 2000), which is also supported by the finding that effects were larger in preschoolers and kindergartners compared to older children. In other words, the earlier you intervene, the greater the chances of a positive response to intervention. In addition, population (at-risk, not at-risk) did not matter; it can thus be inferred that multimedia applications were equally effective in both populations: at-risk children did not get further behind their peers.

Over the years the methodological quality of the primary studies has definitely increased. Whereas Blok et al. (2002) observed that only 25 of 75 had a rigorous design, that is, included a control group and did not lack essential statistics, the current study includes 35 (out of 42 studies) with a control group. In most studies possible differences at pretest between experimental and control groups were accounted for. Also, the use of standardised tests has increased, and unreliable assessments, such as the use of gain or difference scores have become rare. Nevertheless, three studies that analysed gain scores were included in the current meta-analysis and yielded significantly smaller effect sizes. This is due to a relative large error variance of such compound scores and a reduction of the true variance, which leads to an underestimation of the effect size (see for a discussion of the use of compound scores Adèr et al. 2008, p. 261).

We were able to demonstrate that time-on-task between studies makes a difference. This has also been a topic of investigation within studies. For example, Hecht and Close (2002) found that time spent on using the Waterford Early Reading Programme uniquely contributed to effects four out of six literacy outcomes. A similar result was obtained by Segers and Verhoeven (2005), who found that the more time spent on playing a computer game that promoted phonological skills in native Dutch and immigrant children, the more was learnt. The moderating effect that was found for the number of sessions across primary studies in our meta-analysis could be due to regression to the mean, that is, with very many sessions an asymptote of the effectiveness is reached.

Consistent with Blok et al. (2002) no influence of study characteristics such as year of publication, design (statistical), population (being at risk or not at risk), and specificity of training (one or more subskills trained) was found. Cheung and Slavin (2012) found no effect of year of publication either. It is remarkable that year of publication did not affect the effect sizes, as one would expect that researchers gain insights from previous studies and build more effective multimedia. On the other hand, effect sizes are based on *mean* differences and variances. This means that some children profit more from interventions with multimedia than others. It could well be that multimedia interventions across different years of publication are beneficial for different subgroups of children without showing an overall increase of effect size. Design-statistical characteristics of the study (between classes, within classes, between schools, within schools, and counterbalancing within class) most probably did not affect the effect sizes, because when weaker designs such as the between comparisons were used, it was usually checked whether still valid conclusions could be drawn, for example, differences between groups at pretest could be excluded as a possible confounder.

Study characteristics, which Blok et al. (2002) did not examine, but made no difference in our review, include media type (e-book, CAI, TV/Video), country, language, grouping of students (mixed, individual, whole class, small groups), transfer of training (train to test, transfer of training/curriculum-based), and type of scores analysed (raw observed, adjusted, transformed). Unfortunately, if these effects existed at all, the design of the current study would not have had sufficient power to detect them.

### ***8.1 Comparisons with Other Meta-analyses: CAI***

It seems useful to compare our results with the results from other meta-analyses. As far as we know, two recent studies are relevant here, Hattie (2008), who conducted over 800 meta-analyses of existing meta-analyses, which encompassed 52,637 original studies and Cheung and Slavin (2012), who focussed on the impact of technology in literacy learning, synthesizing 84 studies. Hattie (2008) synthesized meta-analyses of CAI, of which only three original studies focussed on literacy

learning, including the aforementioned study by Blok et al. (2002) and two others with respective effect sizes of .19, .27 (published in 2000) and .31 (published in 1995). Cheung and Slavin (2012) using very stringent inclusion criteria, that is, only studies with print-related outcomes, – no phonological awareness or listening comprehension were included – found a 95% confidence interval for the effect size that ranged from .12 to .21. Cheung and Slavin (2012) also reported a relatively larger effect size for comprehensive models of instruction, that is, using CAI along with other non-computer activities supported by teachers.

Where does the difference between the current study's results and the results obtained by Blok et al. (2002) and Cheung and Slavin (2012) come from? We think that, whereas our inclusion criteria were similar to the ones used by Blok et al. (2002), there are disparities with the ones used by Cheung and Slavin (2012). Firstly, Cheung and Slavin (2012) selected 84 studies from the 1980s onwards, of which 47 were published in the 2000–2010 period, including not only journal articles as we did (they selected 15), but also unpublished doctoral theses (11), web publications (4), and reports (17). All studies they included were American, of which 4 were included in our meta-analysis. On the other hand, we included 11 more American studies and 21 studies conducted outside the US. In addition, the selection of studies by Cheung and Slavin (2012) was narrower with respect to literacy outcomes as they selected only print-related outcomes, but much wider with respect to the context in which the multimedia applications were used and the age of the participants. Thus, it could well be that selecting studies from the 1980s onwards, conducted in a wider educational context and with older participants in the original studies has led to finding relatively lower effect sizes of multimedia applications. Please note that we found that older participants profit relatively less from multimedia applications.

## ***8.2 Implementation Variables in CAI***

In intervention research, pilot and efficacy studies are first run in the lab and under controlled circumstances in schools. Then, a manualised intervention is implemented in real-world settings and it is evaluated whether intervention outcomes which are generalizable across various settings and participants (Kaderavek and Justice 2010). Pilot studies and efficacy research are carried out in a controlled setting to assess the causal relation between an intervention and an outcome, for example, whether a phonics programme influences phonological awareness. Maximum control is usually achieved by random allocation of participants to the experimental group, which receives the treatment, or to a comparison group, which receives an alternative treatment and/or a control group, which engages in 'business as usual', combined with pre- and post-testing. Efficacy research results in identification of the 'active ingredients' of an intervention; it answers the questions of why the intervention produces positive outcomes, of how and why an intervention is effective

and of why it works better than other interventions (Longabaugh et al. 2005). In addition, efficacy research informs the effectiveness of the intervention (characteristics) in terms of effect sizes.

Finally, through effectiveness research it is examined how effective an intervention is *as implemented* (Hulleman and Cordray 2009), that is, how treatment effectiveness reduction can be countered when moving from the lab to the field. The reduction in treatment effectiveness is examined by studying treatment fidelity. Treatment fidelity is defined as the degree to which field implementation of an intervention corresponds to the prototype implementation (Hulleman and Cordray 2009). There are two sources of treatment infidelity which decrease treatment strength: (1) in the experimental condition the treatment may not be implemented as prescribed (the teacher does not follow the manual or missed professional development training sessions), so that the intervention becomes less effective, and (2) in the control condition a teacher may add components from the experimental treatment or an alternative treatment, so that the control becomes more effective than it otherwise would have been. In sum, evidence-based practice (EBP) is based on results of both efficacy and effectiveness research.

Thus, in order for multimedia to be successful in the classroom situation or at home we already mentioned that the use of computers should be integrated with other teaching/learning activities (Cheung and Slavin 2012). Archer et al. (2014) examined therefore the moderating effects of (1) the quality of training and support teachers received for implementing a CAI intervention, and (2) the degree of implementation fidelity by combining three comparable meta-analyses. These meta-analyses comprised of original US studies conducted between 1990 and 2007. The overall effect size was .18, whereas there was an added effect size of .58 for training and support, a result that corroborates the finding of relatively larger effect sizes for comprehensive models of instruction (Cheung and Slavin 2012). However, no effect of treatment fidelity was found.

### 8.3 Comparisons with Other Meta-analyses: e-Books

For a very comprehensive systematic review of storybooks, see Bus et al. (2015). As far as we know, there is one meta-analysis specifically on the effectiveness of multimedia and interactive features in storybooks (Takacs et al. 2015). They analysed 57 effects on 5 outcomes from publications between 1980 and 2014 with 2147 participants, aged between 3 and 10 years of age. Effects were .17 ( $p = .04$ ), .20 ( $p = .04$ ),  $-.08$  (n.s.), .16 (n.s.) and .26 (n.s.) for story comprehension, expressive vocabulary, receptive vocabulary, code-related literacy skills and engagement and child-initiated communication during reading, respectively. In addition, Takacs et al. (2015) found that animated pictures, music and sound effects were beneficial, whereas hotspots, games, and dictionaries were distracting. It seems difficult to compare effect sizes from this study with ours, as Takacs et al. (2015) also included TV, video and more, whereas we seem to have included interventions based on the very first lab studies, which had been tested in the field by researchers.

## 8.4 *Limitations*

As with all research, this study has its limitations. We discuss one of them that pertain to meta-analyses in general: whether causal conclusions can be drawn.<sup>3</sup> For example, it would seem sensible to conclude from this study that especially young children should expand the time they spend on learning with multimedia applications, because it was found that effect sizes were relatively larger for younger children in comparison to older children and for studies, in which children spend more time on using the multimedia applications in comparison to studies, in which less time was spent. This is however not necessarily true, because we don't know *how exactly* the multimedia are used. We suggest an examination of how multimedia applications are actually used. Generally, smart phones and tablets offer opportunities for more interactivity through touch screens that can be used by even very young children. However, evaluating apps is a challenging task for the following reasons. (1) There are very many apps available,<sup>4</sup> which makes it difficult to choose from, not only by teachers and parents, but also by researchers. (2) It is unlikely that any commercially available app is fully adaptive with respect to instruction and testing a (literacy) learner, because a sizable item bank is usually lacking. It is therefore unlikely that many primary studies with adaptive interventions can be run, let alone conduct a systematic review. A future methodology that exploits advantages of smart phones and tablets (Dufau et al. 2011) could entail to design an app based on a proven adaptive learning system available to very many children and to tag the devices over the Internet in order to collect data.

## 9 **Conclusion**

Multimedia applications evaluated over the 2000–2010 period have proven to be effective, especially when delivered to preschoolers and kindergartners, and if they are used intensively.

We expect that CAI will continue to be used in schools and homes. However, it seems unlikely that tablets and smart phones equipped with touch screen technology will soon be replaced. However, as we indicated, it will be hard to conduct evaluation studies for these hand-held systems. Nevertheless, this should be done.

An interesting topic for future research is, in our opinion, to look at when children are ready to use educational apps and games on smart phones and tablets. Looking at school readiness, Duncan et al. (2007) found that school-entry maths, reading and attention predicted later achievement best. Others, for example, Diamond (2012) and Nicolson (2016) have suggested that children are ready for

---

<sup>3</sup>There exist many more limitations. Statistical issues are discussed by Bergeron and Rivard (2017).

<sup>4</sup>As of June 2015, over 80,000 educational apps were found on the Apple App Store.

learning maths and reading if their attentional skills are well developed. Moreover, Diamond (2013) found that attentional skills could be trained. Going back to educational apps and games, it would therefore be worthwhile to research how children can best be trained to use educational apps, thereby avoiding distraction by ‘bells and whistles’ (Bus et al. 2015; Takacs et al. 2015).

## Appendices

### *Appendix 1: Principal Component Analysis of Literacy Outcomes in Two Primary Studies*

Literacy outcome	Hecht and Close (2002)	Savage et al. (2009)
PA	.870	.803
Letter knowledge	.706	.578
Word reading	.791	.864
Spelling	.849	.880
Vocabulary	.631	.878
Print concepts	.715	–
Listening comprehension	–	.466
Reading comprehension	–	.837
Nonword reading	–	.863

### *Appendix 2: Effect Sizes for Separate Literacy Outcomes (Van Daal and Sandvik 2013)*

Literacy outcome <sup>a</sup>	ES	95% confidence interval	Number of studies
Comprehension	.52	.27–1.31	12
Letter learning	.89	.66–1.13	6
Nonword reading	.53	.39–.67	13
PA	.75	.68–.83	51
Print concepts	.86	.61–1.11	6
RAN	.21	.05–.38	3
Spelling	1.11	.90–1.32	5
Vocabulary	.68	.57–.80	28
Word reading	.60	.52–.68	44

<sup>a</sup>Literacy outcomes were slightly differently grouped, e.g., listening and reading comprehension were taken together



## References<sup>5</sup>

- Adèr, H. J., Mellenbergh, G. J., & Hand, D. J. (2008). *Advising on research methods: A consultant's companion*. Huizen: Johannes van Kessel Publishing.
- 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.
- Baddeley, A., & Gathercole, S. (1992). Learning to read: The role of the phonological loop. In J. Alegria, D. Holender, J. J. Morais, & M. Radeau (Eds.), *Analytic approaches to human cognition* (pp. 153–167). Amsterdam: Elsevier.
- Bast, J., & Reitsma, P. (1998). Analyzing the development of individual differences in terms of Matthew effects in reading: Results from a Dutch longitudinal study. *Developmental Psychology*, 34, 1373–1399.
- Bergeron, P.-J., & Rivard, L. (2017). How to engage in pseudoscience with real data: A criticism of John Hattie's arguments in VISIBLE LEARNING from the perspective of a statistician. *McGill Journal of Education*, 52, 237–246.
- Blok, H., Oostdam, R., Otter, M. E., & Overmaat, M. (2002). Computer-assisted instruction in support of beginning reading instruction: A review. *Review of Educational Research*, 72, 101–130.
- Bowey, J. A. (1995). Socioeconomic status differences in preschool phonological sensitivity and first-grade reading achievement. *Journal of Educational Psychology*, 87, 476–487.
- Bowey, J. A. (2005). Predicting individual differences in learning to read. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 155–172). Malden: Blackwell.
- \*Brabham, E. G., Murray, B. A., & Bowden, S. H. (2006). Reading alphabet books in kindergarten: Effects of instructional emphasis and media practice. *Journal of Research in Childhood Education*, 20, 219–234.
- Buckingham, D. (2000). *After the death of childhood: Growing up in the age of electronic media*. Oxford: Polity Press.
- Burnett, C. (2009). Research into literacy and technology in primary classrooms: An exploration of understandings generated by recent studies. *Journal of Research in Reading*, 32, 22–37.
- Bus, A. G., Takacs, Z. K., & Kegel, C. A. T. (2015). Affordances and limitations of electronic storybooks for young children's emergent literacy. *Developmental Review*, 35, 79–97.
- Bus, A. G., Sari, B., & Takacs, Z. K. (this volume). The promise of multimedia enhancement in children's storybooks. In J. Kim & B. Hassinger-Das (Eds.), *Reading in the digital age: Young children's experiences with e-books*. Cham: Springer.
- Byrne, B., Coventry, W. L., Olson, R. K., Samuelsson, S., Corley, R., Willcutt, E. G., Wadsworth, S., & DeFries, J. C. (2009). Genetic and environmental influences on aspects of literacy and language in early childhood: Continuity and change from preschool to Grade 2. *Journal of Neurolinguistics*, 22, 219–236.
- \*Cassady, J. C., & Smith, L. L. (2003). The impact of a reading-focused integrated learning system on phonological awareness in kindergarten. *Journal of Literacy Research*, 35, 947–964.
- \*Chambers, B., Slavin, R. E., Madden, N. A., Abrami, P. C., Tucker, B. J., Cheung, A., & Gifford, R. (2008). Technology infusion in success for all: Reading outcomes for first graders. *The Elementary School Journal*, 109, 1–15.
- \*Chera, P., & Wood, C. (2003). Animated multimedia 'talking books' can promote phonological awareness in children beginning to read. *Learning and Instruction*, 13, 33–52.
- Cheung, A. A. K., & Slavin, R. E. (2012). How features of educational technology applications affect student reading outcomes: A meta-analysis. *Educational Research Review*, 7, 198–215.
- Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, 3, 149–210.

<sup>5</sup>References marked with an asterisk indicate studies included in the meta-analysis.

- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. Hillsdale: Lawrence Erlbaum.
- Coiro, J., Knobel, M., Lankshear, C., & Leu, D. J. (2008). *Handbook of research on new literacies*. New York: Routledge.
- Coltheart, M. (1984). Writing systems and reading disorders. In L. Henderson (Ed.), *Orthographies and reading* (pp. 67–79). Hove: Lawrence Erlbaum.
- \*Comaskey, E. M., Savage, R. S., & Abrami, P. (2009). A randomized efficacy study of web-based synthetic and analytic programmes among disadvantaged urban Kindergarten children. *Journal of Research in Reading, 32*, 92–108.
- Cornell, J., & Mulrow, C. (1999). Meta-analysis. In H. J. Adèr & G. J. Mellenbergh (Eds.), *Research methodology in the life, behavioural and social sciences* (pp. 285–323). London: SAGE.
- Courage, M. L. (this volume). From print to digital: The medium is only part of the message. In J. Kim & B. Hassinger-Das (Eds.), *Reading in the digital age: Young children's experiences with e-books*. Cham: Springer.
- De Jong, M. T., & Bus, A. G. (2002). Quality of book-reading matters for emergent readers: An experiment with the same book in a regular or electronic format. *Journal of Educational Psychology, 94*, 145–155.
- \*De Jong, M. T., & Bus, A. G. (2004). The efficacy of electronic books in fostering kindergarten children's emergent story understanding. *Reading Research Quarterly, 39*, 378–393.
- De Jong, P. F., & Van der Leij, A. (1999). Specific contributions of phonological abilities to early reading acquisition: Results from a Dutch latent variable longitudinal study. *Journal of Educational Psychology, 91*, 450–476.
- Diamond, A. (2012). Activities and programs that improve children's executive functions. *Current Directions in Psychological Science, 21*, 335–341.
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135–168.
- \*Doty, D. E., Popplewell, S. R., & Byers, G. O. (2001). Interactive CD-ROM storybooks and young readers' reading comprehension. *Journal of Research on Computing in Education, 33*, 374–384.
- Dufau, S., Dunabeitia, J. A., Moret-Tatay, C., McGonigal, A., Peeters, D., et al. (2011). Smart phone, smart science: How the use of smartphones can revolutionize research in cognitive science. *PLoS One, 6*(9), e24974. <https://doi.org/10.1371/journal.pone.0024974>.
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., et al. (2007). School readiness and later achievement. *Developmental Psychology, 43*, 1428–1446.
- \*Ecalte, J., Magnan, A., & Calmus, C. (2009). Lasting effects on literacy skills with a computer-assisted learning using syllabic units in low-progress readers. *Computers & Education, 52*, 554–561.
- Elbro, C., Borstrøm, I., & Petersen, D. K. (1998). Predicting dyslexia from kindergarten: The importance of distinctness of phonological representations of lexical items. *Reading Research Quarterly, 33*, 36–60.
- Etta, R. A. (this volume). Parent preferences: E-books versus print books. In J. Kim & B. Hassinger-Das (Eds.), *Reading in the digital age: Young children's experiences with e-books*. Cham: Springer.
- Fletcher, J. D., & Atkinson, R. C. (1972). Evaluation of the Stanford CAI program in initial reading. *Journal of Educational Psychology, 63*, 597–602.
- Harm, M. W., McCandliss, B. D., & Seidenberg, M. S. (2003). Modeling the successes and failures of interventions for disabled readers. *Scientific Studies of Reading, 7*, 155–182.
- Hattie, J. (2008). *Visible learning: A Synthesis of over 800 meta-analyses relating to achievement*. London/New York: Routledge.
- \*Hecht, S. A., & Close, L. (2002). Emergent literacy skills and training time uniquely predict variability in responses to phonemic awareness training in disadvantaged children. *Journal of Experimental Child Psychology, 82*, 93–115.
- Hisrich, K., & Blanchard, J. (2009). Digital media and emergent literacy. *Computers in the Schools, 26*, 240–255.

- \*Howell, R. D., Erickson, K., Stanger, C., & Wheaton, J. E. (2000). Evaluation of a computer-based program on the reading performance of first grade students with potential for reading failure. *Journal of Special Education Technology, 15*, 5–14.
- Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness, 2*, 88–110.
- Kaderavek, J. N., & Justice, L. M. (2010). Fidelity: An essential component of evidence-based practice in speech-language pathology. *American Journal of Speech-Language Pathology, 19*, 369–379.
- Kamil, M., Intractor, S., & Kim, H. (2000). The effects of other technologies on literacy and literacy learning. In M. Kamil, P. Mesenthal, D. Reason, & R. Bar (Eds.), *Handbook of Reading Research: Volume 3* (pp. 771–788). Mahwah: Lawrence Erlbaum.
- \*Karemaker, A. M., Pitchford, N. J., & O'Malley, C. (2010a). Enhanced recognition of written words and enjoyment of reading in struggling beginner readers through whole-word multimedia software. *Computers & Education, 54*, 199–208.
- \*Karemaker, A. M., Pitchford, N. J., & O'Malley, C. (2010b). Does whole-word multimedia software support literacy acquisition? *Reading and Writing, 23*, 31–51.
- Keenan, J. M., Betjemann, R. S., Wadsworth, S. J., DeFries, J. C., & Olson, R. K. (2006). Genetic and environmental influences on reading and listening comprehension. *Journal of Research in Reading, 29*, 75–91.
- \*Kegel, C. A. T., Van der Kooy-Hofland, V. A. C., & Bus, A. G. (2009). Improving early phoneme skills with a computer program: Differential effects of regulatory skills. *Learning and Individual Differences, 19*, 549–554.
- Kintsch, W., & Rawson, K. A. (2005). Comprehension. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 209–226). Oxford: Blackwell.
- Kirschner, P. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning. *Learning and Instruction, 12*, 1–10.
- Klinkenberg, S., Straatemeier, M., & Van der Maas, H. L. J. (2011). Computer adaptive practice of math ability using a new item response model for on the fly and difficulty estimation. *Computers & Education, 57*, 1813–1824.
- \*Korat, O. (2010). Reading electronic books as a support for vocabulary, story comprehension and word reading in kindergarten and first grade. *Computers & Education, 55*, 24–31.
- \*Korat, O., & Shamir, A. (2007). Electronic books versus adult readers: Effects on children's emergent literacy as a function of social class. *Journal of Computer Assisted Learning, 23*, 248–259.
- \*Korat, O., & Shamir, A. (2008). The educational electronic book as a tool for supporting children's emergent literacy in low versus middle SES groups. *Computers & Education, 50*, 110–124.
- Krendl, K. A., & Williams, R. B. (1990). The importance of being rigorous: Research on writing to read. *Journal of Computer-based Instruction, 17*, 81–86.
- Kulik, C. C., & Kulik, J. A. (1991). Effectiveness of computer-based instruction: An updated analysis. *Computers in Human Behavior, 7*, 75–94.
- Lankshear, C., & Knobel, M. (2003). New technologies in early childhood literacy research: A review of research. *Journal of Early Childhood Literacy, 3*, 59–82.
- \*Linebarger, D., Piotrowski, J. T., & Greenwood, C. R. (2010). On-screen print: The role of captions as a supplementary literacy tool. *Journal of Research in Reading, 33*, 148–167.
- Longabaugh, R., Donovan, D. M., Karno, M. P., McCrady, B. S., Morgenstern, J., & Tonigan, J. S. (2005). Active ingredients: How and why evidence-based alcohol behavioural treatment interventions work. *Alcoholism: Clinical and Experimental Research, 29*, 235–247.
- \*Lonigan, C. J., Driscoll, K., Phillips, B. M., Cantor, B. G., Anthony, J. L., & Goldstein, H. (2003). A computer-assisted instruction phonological sensitivity program for preschool children at risk for reading problems. *Journal of Early Intervention, 25*, 248–262.
- \*Macaruso, P., & Walker, A. (2008). The efficacy of computer-assisted instruction for advancing literacy skills in kindergarten children. *Reading Psychology, 29*, 266–287.

- \*Macaruso, P., Hook, P. E., & McCabe, R. (2006). The efficacy of computer-based supplementary phonics programs for advancing reading skills in at-risk elementary students. *Journal of Research in Reading, 29*, 162–172.
- Mann, V. A., & Liberman, I. Y. (1984). Phonological awareness and verbal short-term memory. *Journal of Learning Disabilities, 17*, 592–599.
- Marsh, J. (2005). The techno-literacy practices of young children. *Journal of Early Childhood Research, 2*, 51–66.
- \*Mathes, P. G., Torgesen, J. K., & Allor, J. H. (2001). The effects of peer-assisted literacy strategies for first-grade readers with and without additional computer-assisted instruction in phonological awareness. *American Educational Research Journal, 38*, 371–410.
- McCandliss, B. D., Beck, I., Sandak, R., & Perfetti, C. (2003). Focusing attention on decoding for children with poor reading skills: Design and preliminary tests of the word building intervention. *Scientific Studies of Reading, 7*, 75–104.
- \*McKenna, M. C., Labbo, L. D., Kieffer, R. D., & Reinking, D. (2006). *International handbook of literacy and technology, volume II*. Mahwah: Lawrence Erlbaum.
- McKenney, S., & Voogt, J. (2009). Designing technology for emergent literacy: The PictoPal initiative. *Computers & Education, 52*, 719–729.
- \*Mioduser, D., Tur-Kaspa, H., & Leitner, I. (2000). The learning value of computer-based instruction of early reading skills. *Journal of Computer Assisted Learning, 16*, 54–63.
- \*Mitchell, M. J., & Fox, B. J. (2001). The effects of computer software for developing phonological awareness in low-progress readers. *Reading Research and Instruction, 40*, 315–332.
- NAEYC. (1996). *Technology and young children – Ages 3 through 8*. Washington, DC: National Association for the Education of Young Children.
- Naslund, J. C., & Schneider, W. (1996). Kindergarten letter knowledge, phonological skills, and memory processes: Relative effects on early literacy. *Journal of Experimental Child Psychology, 62*, 30–59.
- National Reading Panel. (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*. Rockville: NICHD Clearinghouse.
- National Institute for Literacy. (2008). *Developing early literacy: Report of the National Early Literacy Panel*. Jessup: National Institute for Literacy.
- Neuman, S. B., & Dickinson, D. K. (2001). *Handbook of early literacy research, volume I*. New York: Guilford Press.
- Nicolson, R. I. (2016). Developmental dyslexia: The bigger picture. In A. Davis (Ed.), *Dyslexia: Developing the debate*. London: Bloomsbury Academic.
- Oliver, B. R., Dale, P. S., & Plomin, R. (2005). Predicting literacy at age 7 from preliteracy at age 4: A longitudinal genetic analysis. *Psychological Science, 16*, 861–865.
- Ouyang, R. (1993). *A meta-analysis: Effectiveness of computer-assisted instruction at the level of elementary education (K-6)*. Philadelphia: University of Pennsylvania.
- Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.
- Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227–247). Oxford: Blackwell.
- Plowman, L., & Stephen, C. (2003). A ‘benign addition’? Research on ICT and pre-school children. *Journal of Computer Assisted Learning, 19*, 149–164.
- Rasbash, J., Steele, F., Browne, W., & Prosser, B. (2005). *A user’s guide to MlwiN, Version 2.0*. Bristol: Centre for Multilevel Modelling, University of Bristol.
- \*Regtvoort, A. G. F. M., & Van der Leij, A. (2007). Early intervention with children of dyslexic parents: Effects of computer-based reading instruction at home on literacy acquisition. *Learning and Individual Differences, 17*, 35–53.
- Rideout, V. (2014). *Learning at home: Families educational media use in America*. Retrieved from <https://eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED555586>

- Rideout, V., & Hamel, E. (2006). *The media family: Electronic media in the lives of infants, toddlers, preschoolers, and their parents*. Menlo Park: Kaiser Family Foundation.
- Sadoski, M., & Paivio, A. (2007). Toward a unified theory of reading. *Scientific Studies of Reading, 11*, 337–365.
- Samuelsson, S., Olson, R. K., Wadsworth, S., Corley, R., DeFries, J. C., Willcutt, E., Hulslander, J., & Byrne, B. (2007). Genetic and environmental influences on prereading skills and early reading and spelling development: A comparison among United States, Australia, and Scandinavia. *Reading and Writing, 20*, 51–75.
- \*Savage, R. S., Abrami, P., Hippius, G., & Deault, L. (2009). A randomized controlled trial study of the ABRACADABRA reading intervention program in Grade 1. *Journal of Educational Psychology, 101*, 590–604.
- Scarborough, H. S. (1990). Very early language deficits in dyslexic children. *Child Development, 61*, 1728–1743.
- \*Segers, E., & Verhoeven, L. (2002). Multimedia support of early literacy learning. *Computers & Education, 39*, 207–221.
- \*Segers, E., & Verhoeven, L. (2003). Effects of vocabulary training by computer in kindergarten. *Journal of Computer Assisted Learning, 19*, 557–566.
- \*Segers, E., & Verhoeven, L. (2005). Long-term effects of computer training of phonological awareness in kindergarten. *Journal of Computer Assisted Learning, 21*, 17–27.
- \*Segers, E., Takke, L., & Verhoeven, L. (2004). Teacher-mediated versus computer-mediated storybook reading to children in native and multicultural kindergarten classrooms. *School Effectiveness and School Improvement, 15*, 215–226.
- Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94*, 143–174.
- \*Shamir, A., Korat, O., & Barbi, N. (2008). The effects of CD-ROM storybook reading on low SES kindergartners' emergent literacy as a function of learning context. *Computers & Education, 51*, 354–367.
- Share, D. L. (1995). Phonological recoding and self-teaching: *Sine qua non* of reading acquisition. *Cognition, 55*, 151–218.
- \*Silverman, R., & Hines, S. (2009). The effects of multimedia-enhanced instruction on the vocabulary of English-language learners and non-English-language learners in pre-kindergarten through second grade. *Journal of Educational Psychology, 101*, 305–314.
- Slavin, R. E. (1991). Reading effects of IBM's 'Writing to Read' program: A review of evaluations. *Educational Evaluation and Policy Analysis, 13*, 1–11.
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360–407.
- Stanovich, K. E. (2000). *Progress in understanding reading: Scientific Foundations and new frontiers*. New York: Guilford Press.
- Stephen, C., & Plowman, L. (2003). Information and communication technologies in per-school settings: A review of the literature. *International Journal of Early Years Education, 11*, 224–234.
- Strong, G. K., Torgerson, C. J., Torgerson, D., & Hulme, C. (2010). A systematic meta-analytic review of evidence for the effectiveness of the "Fast ForWord" language intervention program. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 52*, 224–235.
- Takacs, Z. K., Swart, E. K., & Bus, A. G. (2015). Benefits and pitfalls of multimedia and interactive features in technology-enhanced storybooks: A meta-analysis. *Review of Educational Research, 85*, 698–739.
- Torgesen, J. K., & Horen, N. M. (1992). Using computers to assist in reading instruction for children with learning disabilities. In S. A. Vogel (Ed.), *Educational alternatives for students with learning disabilities* (pp. 159–181). New York: Springer.
- Turbill, J. (2001). A researcher goes to school: Using technology in kindergarten literacy curriculum. *Journal of Early Childhood Literacy, 1*, 255–279.

- \*Van Daal, V. H. P., & Reitsma, P. (2000). Computer-assisted learning to read and spell: Results from two pilot studies. *Journal of Research in Reading*, 23, 181–193.
- Van Daal, V. H. P., & Sandvik, J. M. (2013). The effects of multimedia on early literacy development of children at risk: A meta-analysis. In A. Shamir & O. Korat (Eds.), *Technology as a support for literacy achievements for children at risk* (pp. 73–122). Dordrecht: Springer.
- Van Daal, V. H. P., Reitsma, P., & Van der Leij, A. (1994). Processing units in word reading by disabled readers. *Journal of Experimental Child Psychology*, 57, 180–210.
- Van der Leij, A. (1994). Effects of computer-assisted instruction on word and pseudoword reading of reading-disabled students. In K. P. van den Bos, L. S. Siegel, & D. J. Bakker (Eds.), *Current directions in dyslexia research* (pp. 251–267). Lisse: Swets & Zeitlinger.
- \*Verhallen, M. J. A. J., & Bus, A. G. (2010). Low-income immigrant pupils learning vocabulary through digital picture storybooks. *Journal of Educational Psychology*, 102, 54–61.
- \*Verhallen, M. J. A. J., Bus, A. G., & de Jong, M. T. (2006). The promise of multimedia stories for kindergarten children at risk. *Journal of Educational Psychology*, 98, 410–419.
- Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., & Burgess, S. R. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A 5-year longitudinal study. *Developmental Psychology*, 33, 468–479.
- What Works Clearinghouse. (2006). *Fast ForWord language: English language learners*. Washington, DC: US Department of Education, Institute of Education Sciences.
- What Works Clearinghouse. (2007). *Fast ForWord: Beginning reading*. Washington, DC: US Department of Education, Institute of Education Sciences.
- \*Wild, M. (2009). Using computer-aided instruction to support the systematic practice of phonological skills in beginning readers. *Journal of Research in Reading*, 32, 413–432.
- Wimmer, H., Mayringer, H., & Landerl, K. (2000). The double-deficit hypothesis and difficulties in learning to read a regular orthography. *Journal of Educational Psychology*, 92, 668–680.
- Wise, B. W., & Olson, R. K. (1998). Studies of computer-aided remediation for reading disabilities. In C. Hulme & R. M. Joshi (Eds.), *Reading and spelling: Development and disorders* (pp. 473–487). Mahwah: Lawrence Erlbaum.
- Yelland, N. (2005). The future is now: A review of the literature on the use of computers in early childhood education (1994–2004). *AACE Journal*, 13, 201–232.
- Zucker, T. A., Moody, A. K., & McKenna, M. C. (2009). The effects of electronic books on pre-kindergarten-to-grade 5 students' literacy and language outcomes: A research synthesis. *Journal of Educational Computing Research*, 40, 47–87.