NORTHWESTERN UNIVERSITY

Were They Ready To Learn? Short- and Long-Term Effects of Ready To Learn Media on Young Children's Literacy

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Media, Technology & Society

By

Lisa B. Hurwitz

EVANSTON, ILLINOIS

June 2017

Abstract

The U.S. Department of Education's Ready To Learn (RTL) initiative funds (a) mass media and related community outreach intended to promote school readiness among all children—especially at-risk populations—, and (b) formative and summative research by third party evaluators (Michael Cohen Group, 2012). Today, the majority of American preschoolers have consumed RTL media, and RTL has become one of the largest funders of educational media research (Corporation for Public Broadcasting [CPB] & Public Broadcasting Service [PBS], 2011). This dissertation primarily aimed to assess RTL's short- and long-term effectiveness in promoting school readiness, specifically foundational literacy skills. As a secondary goal, this dissertation also sought to shed insight into methodological and theoretical issues concerning children's learning from media more broadly.

Article 1 provides a meta-analytic review of the accumulated research on the short-term effectiveness of RTL's literacy-themed media. Results indicate that RTL media resulted in small but positive impacts on children's early literacy skills, comparable to comprehensive early childcare programs such as Head Start (Kay & Pennucci, 2014), but considerably smaller than targeted nonmediated literacy interventions, such as in-person phonics instruction (Langenberg et al., 2000).

Article 2 outlines how a developmental researcher might re-recruit families who participated in research studies, such as early childhood RTL evaluations, for later research. It highlights newer technological tools that facilitate locating participants who moved between waves of data collection and scheduling them for follow-up sessions (e.g., people-centric search engines built on algorithms that scrape the Internet for publicly available contact information).

Researchers conducting non-prospective longitudinal, prospective longitudinal, or even cross-sectional research could benefit from many of these strategies.

Article 3 builds on Articles 1 and 2. Preschool and kindergarten students who participated in one of the evaluations included in the Article 1 meta-analysis were re-recruited for a follow-up study six years later, using recruitment strategies detailed in Article 2. Children were in preschool and kindergarten during the original intervention, and were in 5th and 6th grade at follow-up testing. Children completed age-appropriate literacy assessments, and their parents provided complementary survey data. Findings suggested the effects of early exposure to RTL media sustained into middle childhood, but only for children who had below and above average literacy skills in early childhood prior to the original intervention. These findings are interpreted in light of the Early Learning Hypothesis, and the Traveling Lens, Capacity, and Differential Susceptibility to Media Effects Models.

Altogether, these results provide relatively positive support in favor of the continued funding of RTL, and speak to larger theoretical debates concerning children's learning from media. Continued monitoring is warranted, as the specific foci and execution of the RTL initiative has evolved across grant cycles.

Acknowledgements

I want to take a moment to thank the very many wonderful people in my life, starting with my brilliant dissertation committee. I give my deepest thanks to my mentor and doctoral advisor Ellen Wartella, who has taught me so much about this field specifically and life in general, always encouraging me to think of big ideas. Further, I thank Alexis Lauricella for providing so much professional guidance and emotional support over the past five years and for always making the time to help me think through both minor and major decisions. Finally, I thank Kelly Schmitt, without whose support this project would not have been possible. From thoughtfully considering how my findings relate to broader educational policy debates to discussing finer details such as contradictory race-ethnicity reporting across surveys, every interaction with Kelly has helped me to elevate the quality of my work and thinking. And of course I am grateful to Kelly and her team for their careful efforts evaluating PBS KIDS Island in 2010, facilitating a relatively straightforward follow-up effort for Articles 2 and 3.

I also would like to thank the many other Northwestern faculty who have shaped my thinking these past five years. I owe particular and immeasurable thanks to Daniel J. O'Keefe, who provided me hours of private tutoring and feedback in support of Article 1. I also thank Thomas D. Cook; I never would have thought to do this project had I not taken his excellent program evaluation class.

In addition, I want to acknowledge the gifted undergraduate students, graduate students, and postdoctoral fellows who contributed so much to this project. I am very thankful to Meredith Ford, Dashia Kwok, Zachary Lochmueller, Francesca Pietrantonio, and Mo Ran, who, perhaps unknowingly, risked their lives driving with me around Chicagoland and collecting data for

Finally, I would like to thank my family and friends from my life beyond Northwestern for encouraging me in and supporting my decision to pursue higher education, and for volunteering your children to be guinea pigs in my studies over the year. You genuinely were interested in my work in moments where I wanted to discuss it, and allowed me to live vicariously through your lives when I didn't. Thank you; thank you.

Table of Contents

List of Tables and Figures

Tables

Figures

Introduction

In 1994, the U.S. Department of Education, Congress, and the Corporation for Public Broadcasting launched the Ready To Learn (RTL) initiative in response to two public concerns: (a) America's children, particularly certain disadvantaged populations, were entering elementary school ill-prepared to succeed in formal educational environments; and (b) these same children were spending hours watching television in a media environment largely characterized by program-length commercials devoid of learning content (Bryant, Bryant, Mullikin, McCollum, & Love, 2001). RTL provides funding for educational mass media intended to promote school readiness; outreach initiatives complementing and extending this learning; and supportive implementation and effectiveness research (Bryant et al., 2001; Singer & Singer, 1998). Most U.S. preschool children have consumed RTL media, and RTL has become one of the largest funders of research into children's learning from media (CPB & PBS, 2011). Originally, RTL funding was used primarily for television, providing seed money for new properties (e.g., *Super WHY!),* as well as added support for legacy properties that existed prior to the launch of the grant (e.g., *Sesame Street*; CPB & PBS, 2011). RTL producers also have moved to the forefront of experimenting with delivering lessons via newer media platforms, such as computer games and mobile apps (Michael Cohen Group, 2012).

A handful of narrative literature reviews written or commissioned by grant recipients have examined RTL's media's overall effectiveness in promoting early literacy (and to a much lesser extent numeracy), highlighting cases for which evaluators noted positive effects for children who consumed RTL media (e.g., Cohen, Hadley, & Marcial, 2016; CPB & PBS, 2011; Michael Cohen Group, 2012). Typically, evaluators in the studies included in these reviews

assessed effects immediately after children complete one- to three-month interventions in which they were exposed to RTL media. However, in a small number of cases, researchers examined if effects persisted up to a year after initial exposure, generally finding supportive evidence in favor of RTL's longer-term effects, at least for select literacy outcomes (Linebarger, 2010; Neuman, Newman, & Dwyer, 2011).

Despite this generally supportive evidence in favor of RTL, accountability evidence for the initiative is incomplete at present. Past reviews typically focused on single grant cycles, rather than looking at RTL's more holistic impact, and did not assess the *consistency* of positive effects. It is also unclear the extent to which the authors of those reviews searched for and included research conducted by parties not funded by the grant. Taking these steps to consider and fairly weight all existing evidence is necessary for a rigorous, unbiased review (Borenstein, Hedges, Higgins, & Rothstein, 2009). It is additionally uncertain whether boosting early learning skills through RTL media would have continued downstream positive impacts on children's academic performance once they enter formal schooling. Exploring the longer-term impacts of early intervention allows for a more compelling policy argument for or against continued funding, particularly for initiatives explicitly attempting to make children better equipped for formal education (e.g., Barnett, 2013). This dissertation address these gaps in accountability evidence by examining the short- and long-term effects of RTL media on young children's literacy across two research studies described in three articles.

The general theoretical assumption across all articles is that children **can** learn from media, provided such media is developmentally appropriate and is intended to convey academic lessons (Hirsh-Pasek et al., 2015). Because the effects of educational media are often relatively

small, others may question this assumption (e.g., Torgerson, 2007). However, RTL was launched under the belief that educational media could be leveraged to promote school readiness (Bryant et al., 2001), even though the specific theoretical focus adopted by content producers and evaluators has varied over the course of the grant's history. Thus, the assumption about educational media's potential is consistent with the theory of the RTL program.

Coloring the overall positive outlook on the potential of educational media espoused in this dissertation, it is likely that certain content, context, and child variables have the potential to positively or negatively moderate educational media's effects (Guernsey, 2012). The potential impact of many moderators is considered across the articles in this this dissertation.

The word "content" is sometimes used to refer different form of media such as curriculum-based vs. aggressive fare (Guernsey, 2012). For this project, a looser definition of content is used, referring to media product itself, and all affordances, storylines, and characters encompassed therein. Some space in this dissertation is dedicated to assessing whether certain media properties are more effective than others. Additionally, this dissertation explores whether new media (e.g., computer games, mobile apps) are effective at all or as effective as television, a more well-studied platform (Fisch & Truglio, 2000). On the one hand, new media might be able to promote strong learning by automatically "leveling up" to present children with increasingly challenging but developmentally appropriate content (Grant et al., 2012; McManis & Gunnewig, 2012; Roberts, Chung, & Parks, 2016), fostering in children a sense of self-efficacy as they advance through gaming environments (Educational Development Center & SRI International, 2012; Ronimus & Lyytinen, 2015), and using highlighting and other features to draw children's attention to educational lessons (Guernsey & Levine, 2015). However, the cognitive and

perceptual resources needed to operate new media may prevent children from processing intended lessons in a deep enough fashion to facilitate transfer to more distal learning measures (Aladé, Lauricella, Beaudoin-Ryan, & Wartella, 2016). And new media products are often designed in a manner devoid of plot, such that engaging with new media would not result in transportation – a sense of losing oneself in a storyline in a way that facilitates deeper processing and learning (Lu, Baranowski, Thompson, & Buday, 2012). Likewise, new media often feature distracting hotspots that draw children's attention away from intended educational lessons (Guernsey & Levine, 2015).

Regarding potential contextual variables that may moderate learning, this dissertation also examines learning in homes and in schools. This dissertation further explores how various methodological decisions (e.g., employment of a between- or within-subjects design) alter both the context of learning and ability to detect significant learning effects.

In terms of child characteristics, this dissertation tests whether media's effects vary as a result of measurable socio-demographic characteristics (e.g., socio-economic status) in the short-term, and children's cognitive skills, independent of family background, in the long-term. Thanks in part to Valkenburg and Peter (2013), who propose that various child-level factors can predict and moderate media's effectiveness in their Differential Susceptibility Media Effects Model, children's media scholars beginning to shift from primarily asking *whether* media has an impact on children to asking *who* experiences the greatest media effects. Extending this theoretical work, Piotrowski and Valkenburg (2015) distinguish between "dandelion children" (p. 4), who would flourish in any environment and are largely insusceptible to media, and "orchid children" (p. 4), who flourish under positive media effects but are particularly vulnerable

to negative media, such as content modeling aggressive behaviors. Linebarger and Barr (2017) assert that low-income children may be one population of "orchid children".

In addition to probing into potential content-, context-, and child-level moderators, this dissertation also explore how long-lasting media's effects may be. As will be explained in more detail in Article 3, Anderson and colleagues (2001) propose that early educational television exposure can spark long-term academic achievement, an idea they refer to as the "Early Learning Hypothesis" (p. 3). They substantiate this notion with supportive correlational evidence linking preschool television consumption to high school academic achievement, and other scholars have begun to collect causal evidence into whether the effects of media, including new media, are still measurable up to a year after initial exposure (e.g., Reitsma & Wesseling, 1998; Segers & Verhoeven, 2005). Although the RTL producers and evaluators never explicitly refer to the Early Learning Hypothesis, this notion is highly resonant with the RTL theory of program. Nonetheless, the Early Learning Hypothesis may be more polemic than the idea that children can consume educational media and then repeat back what they learned shortly after consumption, especially given that short-term media effects tend to be relatively small (i.e., it is debatable whether small effects could sustain over time).

Across this dissertation, I broadly assess the impact of the RTL initiative as a means of providing accountability evidence speaking to the value of this federal program (Rossi, Lipsey, & Freeman, 2003). Specifically, I examine whether exposure to RTL literacy-themed media results in robust, positive effects in the short-term, and whether any positive effects sustain into late elementary/early middle school. In the process, I consider how my results shed insight—or

fail to shed insight—into some of the potential moderators identified in the preceding paragraphs.

Article 1 systematically assesses the average short-term impact of RTL media exposure on young children's early literacy skills via a meta-analytic review of the accumulated short-term RTL evaluations. This article also examines if effects differ as function of content, context, and child variation. This meta-analysis includes both studies funded and not funded by RTL.

In Article 3, I re-contact children who participated in one of the evaluations in my meta-analysis. In this study, I ask whether the effects of playing RTL literacy-themed computer games in early childhood sustain into middle childhood. I also explore whether effects vary as a function of children's preliteracy skills prior to the RTL intervention. The latter analyses allow me to address whether the media helped to close achievement gaps by supporting the children most in need of intervention.

Article 2 does not RTL initiative specifically but provides a broader reflection on the process of re-recruiting families who participate in research when children are in early childhood (e.g., in studies such as RTL evaluations) for follow-up waves of research. Re-recruiting families for longitudinal research requires (a) searching across a variety of sources to locate participants and leveraging new online tools designed to yield contact information; (b) writing multifaceted scheduling scripts highlighting a study's value; and (c) confirming appointments in a way that conveys a casual tone to make parents feel comfortable rescheduling if need be. Rather than burying these insights in the Method section of a traditional dissertation, Article 2 will be published as a standalone piece to help guide other researchers working with families. After all, there have been calls for more longitudinal research specifically in the area of children's media

(Wartella et al., 2016), as well as more broadly in other areas of child development (e.g.,

Karmiloff-Smith, 1998). Demonstrating how to address the formidable challenge of re-recruiting

families for follow-up waves of data collection may help to facilitate more longitudinal research

with children, and may even inform cross-sectional recruitment protocols.

This dissertation concludes by considering the larger implications these findings have for

both policy and scholarship. Overall, the findings described herein positively reflect on the RTL

initiative, at least for its impact on children's early literacy skill as assessed in this dissertation.

However, these findings raise as many new questions as they answer in terms of theory, calling

for more nuanced thinking about the factors that moderate learning from media.

Article 1: Getting a Read on Ready To Learn Media: A Meta-Analytic Review of Effects on Literacy

Abstract

Most U.S. preschoolers have consumed media created with funding from the U.S. Department of Education's Ready To Learn (RTL) initiative, which was established to promote school readiness. Synthesizing data from 45 evaluations, this meta-analysis examined the short-term effects of RTL media exposure on young U.S. children's literacy skills. Results indicate positive effects of RTL exposure on children's literacy outcomes, especially vocabulary and phonological processing. Findings are robust across a variety of research designs and for exposure to both television and new media. Effects are comparable in size to comprehensive early childcare programs such as Head Start. These results are discussed in terms of accountability evidence for RTL and larger debates in scholarly understanding of educational media effects.

Keywords: Ready To Learn, literacy, meta-analysis, educational media, early childhood

Responding to public concerns about growing numbers of U.S. children entering kindergarten ill-prepared for formal schooling, Congress, the U.S. Department of Education (DoEd), and the Corporation for Public Broadcasting (CPB) launched the Ready To Learn (RTL) initiative in 1994 (Bryant et al., 2001). Through this initiative, DoEd has awarded millions of dollars annually - in recent years over $25 million per year - to facilitate the creation, dissemination, and research into the effectiveness of media products promoting young children's (ages 2-8) school readiness, with an emphasis on reaching disadvantaged youth (Michael Cohen Group, 2012). Funding has been used for the continued production of legacy media properties such as *Sesame Street,* as well as the creation of newer properties such as *Between the Lions* and *Super WHY!* (CPB & Public Broadcasting Service [a U.S. national public television broadcaster and distributor; hereafter PBS], 2011). Today, the majority of U.S. preschoolers have been exposed to RTL media in some capacity, with millions of children accessing these resources regularly (CPB & PBS, 2011). Historically, early literacy has been one of RTL's primary focus areas. This article provides a meta-analytic review of the accumulated research on the effectiveness of RTL's literacy-themed media.

Early Literacy

As young children grow to become literate, they begin to master a variety of related early literacy skills. Young readers must learn to recognize letters of the alphabet (alphabet knowledge), develop increasing mastery over the sounds used in language (phonological processing), recall definitions of words (vocabulary), understand how texts work (e.g., in English, words are read from left to right and books often begin with a cover and title page; print concepts), and be able to follow story structure (narrative comprehension; see Table 1 for more

on each of these literacy skills and Langenberg et al., 2001 for more on literacy fundamentals).

Children typically learn these skills in early childhood, beginning to develop print concepts at

around the age of 2.5 years, alphabetic knowledge and phonological processing at around the age

of 3 years, narrative comprehension of written texts at around the age of 7 years, and vocabulary

across childhood (Grant et al., 2012). To a great extent, young children's skills in these areas are

mutually reinforcing and predictive of adult literacy (Langenberg et al., 2000). For example,

early phonological processing (Calfee, Lindamood, & Lindamood, 1973) and vocabulary

(Cunningham & Stanovich, 1997) predict reading at the end of formal schooling. Likewise, early

narrative abilities (Griffin, Hemphill, Camp, & Wolf, 2004) and letter identification (Adlof,

Catts, & Lee, 2010) also predict later reading. Of all these skills, phonological processing is

arguably the strongest predictor of later reading (Grant et al., 2012).

Table 1. Literacy Categories

Literacy Category	Examples of Relevant Skills	Sample Assessments
Alphabet knowledge	Letter identification; letter sequencing	Alphabet Knowledge test from the Phonological Awareness Literacy Screening, Pre-K (PALS-PreK; Invernizzi, Sullivan, Meier, & Swank, 2004)
Print concepts	A variety of related skills concerning knowledge of parts of a book and knowledge of print conventions	Conventions subtest from the Test of Early Reading Ability, Third Edition (TERA-3; Reid, Hresko, & Hammill, 2001)
Phonological processing	Phonemic awareness (knowledge of language sounds, demonstrated through skills such as blending phonemes or joining together	Beginning Sound Awareness and Letter Sounds tests from PALS-

	language sounds to form words or pseudowords); Phonics (knowledge of the relation between letters and sounds)	PreK (Invernizzi et al., 2004)
Vocabulary	Expressive vocabulary (accurately labeling objects/states/phenomena); Receptive vocabulary (knowing the meaning of words others say aloud)	Peabody Picture Vocabulary Test, Third Edition (PPVT-III; Dunn & Dunn, 1997)
Narrative comprehension	Following and interpreting story plotlines	Meanings subtest from the TERA-3 (Reid et al., 2001)

Note. More information on literacy categories.

Early literacy is often a target for intervention, and promoting literacy has been a major focus of U.S. educational policy, especially over the past two decades (e.g., National Conference of State Legislatures, 2016). Past meta-analyses point to the effectiveness of programs promoting early phonological processing, vocabulary, and reading, particularly for preschool-age children (e.g., Bus & van Ijzendoorn, 1999; Mol, Bus, de Jong, & Smeets, 2008). Because of such evidence, a National Reading Panel of experts convened by the U.S. Congress in the late 1990's recommended literacy instruction promote learning in these areas (Langenberg et al., 2000).

Educational Children's Media

A considerable number of U.S. teachers and parents expose young children to screen media intended to promote one or more of the literacy skills outlined above. Among the 80% of early childcare teachers with access to television, 84% with access to computers, and 29% with access to tablets, most use these devices to promote children's literacy (62%, 76%, and 79% of teachers respectively per device; Wartella, Blackwell, Lauricella, & Robb, 2013). Many parents complement children's in-school media use by exposing or allowing children to consume more

literacy-themed media at home. Among the families with access to television, computers, and

tablets (96%, 76%, and 40% respectively), 61% use television, 46% use computers, and 51% use

tablets to expose children to educational content such as literacy-themed media (Rideout &

Saphir, 2013). Television use dominates children's time with screen media, especially at home,

but, as access becomes more widespread, both teachers and parents increasingly are providing

children access to newer devices such as tablets (Rideout & Saphir, 2013; Wartella et al., 2013).

A variety of theoretical frameworks suggest children can learn skills such as vocabulary

from such media, provided that media products are intended to be educational and are designed

in a manner that supports learning (Guernsey, 2012). For the sake of brevity, some of the more

recent and widely accepted thinking in this area is reviewed. The Capacity Model posits that

children should learn from educational media when the intended educational lessons, plotline,

and (when relevant) game mechanics are all mutually reinforcing (Fisch, 2004, 2016). For

example, a computer game focused on the /p/ sound might require children to hover their mouses

over images of foods like "pepperoni" and "peppers" to hear the words said aloud while they

create a menu for a p-themed pizza parlor. This model also suggests well designed media

products will (a) base narratives in simple, familiar settings (e.g., a playground) so that children

can assimilate lessons into existing schemas, and (b) repeat lessons across multiple contexts

(e.g., include a second game in which children create a shopping list of "pears", "peaches" and

other "p-" foods; Fisch, 2004). Extending the notion of the importance of repetition, others have

suggested caregivers should complement media's lessons with real-world activities promoting

the same learning goals - a concept termed Experiential Mediation (Piotrowski, Jennings, &

Linebarger, 2012). More broadly, the idea that other people can engage children in discussions to

scaffold their learning from media is known as Joint Media Engagement (Takeuchi & Stevens, 2011). To a lesser extent, the media itself may be able to mimic strong scaffolding behavior (e.g., with characters asking children questions to enhance their comprehension and critical thinking; Strouse, O'Doherty, & Troseth, 2013).

Scholars currently are debating whether new media such as computer games or mobile applications (apps), and television can embody these production and design principles equally well (Hirsh-Pasek et al., 2015). That is, researchers are asking if new media is more or less effective than television. On the one hand, new media has the potential to facilitate strong learning by automatically becoming more challenging and "leveling up" to enhance engagement (Hirsh-Pasek et al., 2015) and by incorporating learning aids such as clickable dictionaries or word narration (Guernsey & Levine, 2015). But on the other hand, the interactivity required to engage with these newer platforms may require a fair amount of troubleshooting and technical assistance from caregivers to allow for successful usage, which consequently may hamper learning (Guernsey, 2012).

For most of the 20th century, the bulk of research substantiating the premise that children could learn literacy skills from well-designed media focused on *Sesame Street,* but, in the past twenty years, there has been an explosion of new literacy-themed media properties and products available across platforms (i.e., television, computer, mobile), as well as related research (Guernsey & Levine, 2015). A few meta-analyses and narrative literature reviews have explored *Sesame Street* and these newer media properties and products consistently enhance literacy (e.g., Moses, Linebarger, Wainwright, & Brod, 2010; Pasnik, Penuel, Llorente, Strother, & Schindel, 2007). The literature generally suggests educational media has small but positive effects on early

literacy skills, typically as measured by custom assessments designed to reflect the learning goals of the media stimuli (Anderson & Collins, 1988). However, not every media product is equally effective in this regard, and many open questions remain. For example, it is unclear if media benefits *all* literacy skill areas. Prior reviews provide strong evidence that media can promote vocabulary (e.g., Moses et al., 2010) and some evidence it can promote other skill areas, such as phonological processing (Ehri et al., 2001; Langenberg et al., 2000), narrative comprehension, and print concepts (Pasnik et al., 2007). However, Pasnik and team (2007) failed to find significant gains for alphabet knowledge. Moreover, Gola and colleagues (2012) question if media can promote relatively complex, multifaceted skills such as phonological processing (also see Guernsey, 2012). It also is unclear if effects are equal for all children. Concerningly, some research suggests low-income children learn less from educational media than their more affluent counterparts, which in turn may exacerbate achievement gaps (Cook et al., 1975). Furthermore, these reviews have yet to address whether television or new media typically yield larger effects.

The RTL Initiative

RTL has placed a strong emphasis on promoting early literacy among both general and at-risk populations across every five-year grant cycle, aligned with early literacy's emphasis in larger U.S. educational policy discussions (e.g., National Conference of State Legislatures, 2016). During the first two cycles (i.e., 1994-2000 and 2000-2005), RTL funding recipients CPB and PBS prioritized early literacy by encouraging parents to read books complementing RTL media (Horowitz et al., 2005), and in subsequent grant cycles, producers increasingly began developing turnkey curricula for teachers and parents to help them supplement RTL's media content (Llorente, Pasnik, Penuel, & Martin, 2010). Starting in 2005, shortly after the passage of

the No Child Left Behind Act (which focused in part on promoting literacy skills such as

alphabet knowledge, phonological processing, vocabulary, and comprehension in a manner

aligned with National Reading Panel recommendations; National Conference of State

Legislatures, 2016), DoEd explicitly tasked content producers with promoting foundational

literacy skills (Michael Cohen Group, 2012). While RTL producers have espoused a variety of

different theories of learning across media properties and grant cycles, the initiative's consistent

general underlying assumption is that well-designed, developmentally appropriate media should

enhance early literacy and other school readiness skills, especially when complemented with

Joint Media Engagement .

Over the course of the grant's history, DoEd also has called for increasingly rigorous

evaluations into the effectiveness of RTL media. For example, starting in the 2005-2010 grant

cycle, DoEd encouraged RTL evaluators to use randomized controlled trials (when feasible) and

required content producers to partner with third-party researchers, who, although receiving

funding from RTL, presumably would produce more objective results than content producers

might find themselves (Michael Cohen Group, 2012). Consequently, RTL became one of the

largest funders of research into children's learning from media (CBP & PBS, 2011).

The Current Study

While extant literature suggests educational media can promote early literacy, systematic

exploration of the effectiveness of products created with RTL funding still is needed to fully

address whether curriculum-based media in general consistently achieves such effects and

whether the RTL products specifically are successful in this regard. That is, it is unclear whether

RTL media embody the strong production and design principles described above. Given the

majority of U.S. children are consuming RTL-funded media (CPB & PBS, 2011) and hundreds

of millions of taxpayer funds have been dedicated to this initiative since the grant's inception

(DoEd, 2015), arriving at such an understanding provides accountability evidence for the RTL

initiative. Some research has attempted to address this gap in the literature through meta-analytic

work that included a mix of RTL and non-RTL media (Moses et al., 2010; Pasnik et al., 2007);

however, those studies did not specifically address the effectiveness of the RTL initiative.

Moreover, since the time those reports were prepared, researchers have completed dozens of new

evaluations of RTL literacy-themed media, with increasing attention to methodological rigor

(Michael Cohen Group, 2012).

The present exploratory study aimed to update this literature. Specifically, this study asked

(RQ1) whether RTL media exposure positively impacts young children's early literacy skill

development. The study also explored (RQ2) whether these effects varied as a function of

participant characteristics (e.g., socio-economic status) or research design (e.g., within- vs.

between-subjects design). Finally, this study assessed the relation between grant-specific factors

and observed effects, examining if (RQ3a) effects changed over time as content producers might

have gained experience or as evaluations became more rigorous, and if (RQ3b) effects varied

between researchers who were supported by RTL funding and those who were not. The findings

from this study provide accountability evidence for RTL spending and address broader concerns

regarding young children's learning from media.

Method

Identification of Relevant Investigations

Literature search. To locate evaluations funded by RTL, studies referenced in lists of

evaluations prepared by RTL grant recipients (e.g., CPB & PBS, 2011) and in compendia of

research on educational children's media (e.g., Fisch, 2004) were identified. Many evaluations

were available online as technical reports from RTL research and production websites. To find

additional evaluations referenced in RTL reviews but not publicly available, RTL research

Principal Investigators and the Director of Research at CPB were contacted directly. These

individuals shared all available reports, including some that failed to support the value of a focal

property's effectiveness. Lastly, on August 8 and 9, 2016, a systematic search was carried out

across six academic databases: Communication Source, Eric, ProQuest Dissertations & Theses

Global, PsycINFO, Web of Science, and WorldCat Dissertations & Theses. Search efforts

focused on titles, abstracts, and keywords or topics for the titles of every media property

identified in outside literature as being funded by RTL, names of major grant recipients, and

marketing taglines used in public promotions for RTL (e.g., Bryant et al., 2001; Cohen et al.,

2016; CPB & PBS, 2011; Llorente et al., 2010; Michael Cohen Group, 2012; see Appendix A for

the full list of search terms).

      Inclusion criteria. For the purposes of this analysis, evaluations had to meet five criteria:

1. *Use of final versions of RTL media products without any added content.* Studies were

   included that used full versions of media content (e.g., Piotrowski et al., 2012), clips from

   RTL shows (e.g., Neuman et al., 2011) or versions of episodes that paused videos

   occasionally for discussion, so long as no new mediated content was added (e.g., Penuel

   et al., 2012).

         In contrast, studies were excluded in the case of formative evaluations of media

   that would change before commercial distribution (e.g., McCarthy et al., 2012); studies

testing versions of episodes with added scenes that did not air nationally (e.g.,

Linebarger, 2006) or using still images of RTL media characters as stimuli (e.g.,

Golinkoff, Jacquet, Hirsh-Pasek, & Nandakumar, 1996); or evaluations of spinoffs of

properties that received RTL funding (e.g., *Seseame Beginnings*, a *Sesame Street* spinoff;

Roseberry, Hirsh-Pasek, Parish-Morris, & Golinkoff, 2009), unless there was evidence

producers received RTL funding specifically for the spinoff (e.g., *Duck's Alphabet*, an

RTL-funded *WordWorld* spinoff; Michael Cohen Group, 2012).

     If evaluations were conducted outside the U.S., authors needed to indicate they

tested versions of media created exclusively with RTL funds. Some content producers

who received RTL funding also partnered with international groups to create versions of

media tailored for local audiences (e.g., Mares & Pan, 2013). Thus, a property's content

in a foreign market might feature a mix of content created with and without RTL funds.

Searches did not yield any international studies in which authors specified they used the

RTL-funded versions of media; thus, all studies in this meta-analysis sampled in the U.S.

2. *Data collected between 1994 and the present.* Studies of legacy properties conducted

    before the launch of RTL were excluded (e.g., Wright et al., 2001).

3. *Presence of at least one direct assessment of children's literacy in areas such as alphabet*

    knowledge, print concepts, phonological processing, vocabulary, and narrative

    comprehension (see Table 1 for examples). Studies focused on learning domains other

    than literacy (e.g., mathematics; Cohen et al., 2016) and studies where children's

    comprehension of the television show's plot was the sole cognitive outcome measure

    (e.g., Sanchez & Lorch, 1999) were excluded.

4. *Clear demarcation between the effects of RTL vs. non-RTL media.* Studies in which researchers exposed children to both RTL *and* other children's media in a way that confounded RTL and non-RTL media effects were excluded (e.g., Nathanson & Rasmussen, 2011).

5. *Sufficient statistical information to estimate effect sizes.* In cases in which such detail was not provided in available reports, this information was requested from authors. About two out of every three contacted research teams provided useable data, with recent RTL grant recipients tending to provide more data than RTL-funded researchers who conducted research in the first grant cycle (many of whom had retired) and non-funded researchers.

Final sample. Data for this meta-analysis came from 45 RTL evaluations (33 of which were funded by RTL). In many instances, a combination of an unpublished research report, conference presentation, or published journal article all described the same evaluation. Under these circumstances, all texts were treated as being part of the same evaluation, and available information was consolidated across texts as appropriate to provide a comprehensive portrayal of the evaluation. These 45 evaluations resulted in 783 effect sizes for further analyses. See Appendix B for the full list of evaluations.

Coding of Study Descriptors

Each evaluation was coded for literacy learning outcomes, participant characteristics, design and protocol decisions, and grant-related factors. Basic descriptive information about each code can be found in Tables 2 and 3 and Appendix B.

Table 2. Descriptive Statistics for Nominal and Ordinal Moderators

Moderator	n	m	k	Cohen's κ
Literacy categories				.84

Alphabet knowledge	18	31	56	
Print concepts	8	14	21	
Phonological processing	20	35	97	
Vocabulary	30	53	536	
Narrative comprehension	10	16	20	
Multiple skills	16	29	31	
Other	12	16	22	
Design and procedure-related				
Study design				1.00
Between subjects	35	65	404	
Within subjects	11	14	379	
Scope of study				.59
One geographic region	30	62	719	
Multiple geographic regions	14	16	62	
Setting				.87
School	36	61	686	
Home	11	18	97	
Media platform				.79
TV only	35	66	563	
Incorporates new media	12	13	220	
Media property				.75
Between the Lions	9	23	129	
Martha Speaks	6	9	356	
Sesame Street	6	10	45	
Super WHY!	5	5	44	
Other property	17	23	101	
Combination of properties	8	9	108	
Extension Activities/Materials				.89
Activities/Materials	27	44	553	
No Activities/Materials	22	29	220	
Control				.50
Business as usual	17	33	162	
Non-RTL media	12	13	83	
Alternate curriculum	7	13	151	
Assessment				.72
Custom narrow	27	41	528	
Custom broad	13	16	30	
Standardized and validated	30	54	225	
Grant-related				
Researchers funded by RTL				1.00
Yes	33	55	711	
No	12	24	72	
Grant Cycle				.93
1994-2000	5	8	26	

2000-2005	5	15	63
2005-2010	26	45	666
2010-2015	9	11	28

Note. Number of studies (*n*), comparisons (*m*), and effect sizes (*k*) associated with each nominal and ordinal moderator, and intercoder reliability (Cohen's κ) for each set of moderators based on double coding 30% of evaluations.

Literacy learning outcomes. Each evaluation assessed at least one literacy skill. An outcome was treated as a literacy outcome if it concerned any of the following: alphabet knowledge, print concepts, phonological processing, vocabulary, narrative comprehension, or other literacy skill areas such as interest in reading, English conversation skills, or syntax. Each literacy category is described in greater detail in Table 1, with descriptive information in Table 2.

Participant characteristics. Given RTL's focus on reaching all children, especially ones from low-income households or those who otherwise might be at-risk for poor school outcomes (CPB & PBS, 2011), data were recorded on each sample's composition with respect to family income (% low-income with annual household incomes below $30,000 or eligible for free or reduced school lunch), race-ethnicity (% Caucasian, African American, Hispanic or Latino, Asian, and Native American), gender (% Female), school grade (% school age = Kindergarten and above), and English language learner status (% ELL). See Table 3.

Table 3. Descriptive Statistics for Continuous Moderators

Moderator	*n*	*m*	*k*	*M*	*SD*	Median	Range	ICC
Sample-related								
% Low-income	24	35	514	59	28	62	9 - 100	.99
% Caucasian	33	56	656	40	32	34	0 - 92	.94
% African American	31	50	652	29	25	26	0 - 95	.95
% Hispanic/Latino	26	40	614	29	31	20	0 - 100	.99
% Asian	21	34	572	9	9	6	0 - 29	.77
% Native American	9	14	56	8	27	0	0 - 100	1.00
% Female	38	64	712	51	7	50	27 - 88	.65
% School-age	43	77	771	54	47	69	0 - 100	.97

% ELL	17	33	267	35	38	14	0 - 100	1.00
Duration of media consumption								
Weeks of exposure	43	75	773	14.07	26.34	6	1 - 120	.87

Note. Number of studies (*n*), comparisons (*m*), and effect sizes (*k*) providing information on each moderator, along with means, standard deviations, medians, and ranges (calculated at the comparison level), and $ICC_{(3,1)}$ based on double coding 30% of evaluations.

Design and protocol. The dataset varied along the following dimensions (see Table 2):

1. *Study design:* Studies varied in whether they used a between- or within-subjects design.

   In a between-subjects design, a group exposed to RTL media was compared to a group

   not exposed to RTL media. Conversely, in a within-subjects design, children's

   performance was assessed before and after RTL exposure, and there was no comparison

   group.

2. *Scope:* Studies varied in whether data were collected in one geographic region or across

   multiple regions. Early into the history of RTL, evaluators concluded that it was

   challenging to ensure fidelity of implementation when fielding evaluations across

   multiple geographic regions (Horowitz et al., 2005), but this assumption has not been

   formally tested.

3. *Setting:* Some studies examined media consumption in school or another childcare

   setting, but others had children consume all or some media at home.

4. *Duration of media consumption:* Evaluations varied in the number of weeks in which

   researchers tasked children with consuming media.

5. *Nature of RTL treatment.* Stimulus materials presented to children were coded based on

   (a) *media platform* (TV only vs. incorporating some new media), (b) *media property*

   (properties examined in five or more evaluations—*Between the Lions, Martha Speaks,*

Sesame Street, Super WHY!--; other properties examined in fewer than five evaluations;

and combinations of multiple of these properties), and (c) use of supplementary *extension*

activities or materials.

6. *Nature of control.* In studies with between-subjects designs, comparison children were

   either given an alternate, non-mediated curriculum; exposed to media outside the purview

   of RTL; or not given any specific directions (business as usual).

7. *Assessment.* Generally, evaluations of both literacy interventions (e.g., Langenberg et al.,

   2000) and educational media programs (Anderson & Collins, 1988) often measure

   learning with custom measures closely reflecting the curricula to which children are

   exposed, although occasionally will employ validated measures. Thus, in the present

   meta-analysis, outcomes were scored as being assessed by custom measures designed to

   either narrowly assess specific items or skills shown in media (*custom narrow)* or to

   broadly assess literacy skills not explicitly modeled in media (coded as *custom broad*)*,* or

   by *standardized and nationally validated assessments* (see Table 1 for examples).

   Grant-related factors. Each evaluation was coded for the grant cycle in which data were

collected (1994-2000, 2000-2005, 2005-2010, 2010-2015) and for whether researchers were

supported by RTL funding.

Meta-Analytic Procedures

   Effect size measure. Cohen's *d* was used to calculate effect sizes. For most between-

subjects design studies, effect sizes were calculated using procedures outlined by Morris (2007),

comparing gains of the treatment groups to gains of control groups based on provided means and

standard deviations. When insufficient information was available to use this procedure, effects

were calculated by drawing on other study data such as post-test scores (Borenstein et al., 2009) and t-statistics (Rosnow & Rosenthal, 2009). For within-subjects designs, pre- and post-exposure performance was compared, following procedures described by Morris and DeShon (2002).

Structure of dataset. The data reported herein had a nested structure. Evaluations often contained more than one comparison (e.g., comparing a treatment group for whom parents received supplemental extension activities to a control group, and comparing another treatment group for whom parents did not receive supplements to that same control group; Piotrowski et al., 2012). Additionally, evaluators often administered multiple literacy measures to children across these different conditions and comparisons. Thus, measures were nested within comparisons, and comparisons were nested within evaluations. In total, the dataset contained $m = 79$ comparisons (per evaluation $M = 1.76$ comparisons, $SD = 1.05$, *range*: 1-6) and $k = 783$ effect sizes (per comparison $M = 9.73$ effects, $SD = 22.68$, *range*: 1-144).

Analytic strategy. Analyses were conducted in R-3.13 (R Core Team, 2016) via the rma.mv function in the metafor package (Viechtbauer, 2010). A precision-weighted, random effects, multi-level model was specified using procedures outlined by van Houwelingen and colleagues (2002). Measures were nested within comparisons, which were nested within evaluations, and it was assumed that effects might vary across comparisons and across evaluations due to coded methodological variation. To avoid making assumptions about the inter-relations of the random effects, a compound symmetric variance structure was used. Others conducting similar meta-analyses on the effects of educational media have specified comparable models (e.g., Mares & Pan, 2013). In the present data, examination of the variance components revealed that 1% of the variance was accounted for at the comparison level, and 40% was

accounted for at the evaluation level. To compare the nominal moderators described in the section on Coding of Study Descriptors, meta-analytic procedures by Viechtbauer (2015) were used (see Table 4, Model 1). To explore the effects of the ordinal and continuous moderators, meta-regression was used (Viechtbauer, 2007) (see Table 5, Model 1). To test the robustness of these findings, all analyses were repeated using two more conservative models (Models 2 and 3), described in more detail in a Robustness Check section below. Finally, publication bias (the notion researchers may have refrained from publishing studies with null or negative results; Borenstein et al., 2009) also was assessed, although the likelihood of this bias was low because so many unpublished RTL evaluations were available due to grant reporting requirements.

Results

Main Effects

Overall, exposure to RTL media significantly predicted an increase in children's literacy ($d = .21$, $SE = .04$, $n = 45$, $m = 79$, $k = 783$, $p < .001$). Children who consumed RTL media scored approximately one-fifth of a standard deviation higher on literacy assessments. This result included all effect sizes across studies of varied designs and across all literacy outcomes, although there was evidence of heterogeneity across the dataset ($Q(782) = 3751.42$, $p < .001$).

Table 4. Nominal Moderator Effects

Moderator	Model 1		Model 2		Model 3	
	d	$SE\ (d)$	d	$SE\ (d)$	d_t	$SE\ (d_t)$
Literacy categories						
Alphabet knowledge	$.09_a^\dagger$	.05	$.12_{ab}$	.08	$.10_a^\dagger$	.05
Print concepts	$.23_{ab}^\dagger$	.12	$.10_{ab}$	.07	$.23_{ab}^\dagger$	.13
Phonological processing	$.23_b^{***}$	.06	$.21_b^{**}$	.06	$.25_b^{***}$	.05
Vocabulary	$.27_b^{***}$	.06	$.27_b^{**}$	.08	$.26_b^{***}$	.06
Narrative comprehension	$-.01_a$	.06	$-.07_a$	.11	$-.05_c$	.05
Multiple skills	$.15_{ab}^\dagger$	.08	$.06_{ab}$	.08	$.16_{ab}^*$	.07
Other	$.14_{ab}$	.09			$.14_{ab}^\dagger$	.08

Design and procedure-related						
Study design						
Between subjects	$.17_a^{***}$	.04			$.17_a^{***}$	.03
Within subjects	$.37_b^{***}$	.09			$.36_b^{***}$	.08
Scope of study						
One geographic region	$.21_a^{***}$	.05	$.24_a^{**}$	.08	$.21_a^{***}$	.05
Multiple geographic regions	$.23_a^{***}$	.04	$.18_a^{***}$	.05	$.20_a^{***}$	.05
Setting						
School	$.22_a^{***}$	.05	$.25_a^{**}$	.08	$.22_a^{***}$	.04
Home	$.20_a^{***}$	.07	$.17_a^{***}$	.05	$.19_a^{***}$	.05
Media platform						
TV only	$.21_a^{***}$	.05	$.24_a^{**}$	.08	$.21_a^{***}$	.04
Incorporates new media	$.20_a^{***}$	.03	$.20_a^{***}$	.04	$.20_a^{***}$	.03
Media property						
Between the Lions	$.07_a$	.05	$.11_a$	.09	$.08_a^{\dagger}$	.05
Martha Speaks	$.12_{ab}^{\dagger}$	.06	$.05_a$	.15	$.11_a^{*}$	.05
Sesame Street	$.52_c^{**}$	.18	$.60_b^{***}$	.08	$.50_b^{**}$	.17
Super WHY!	$.28_{bc}^{***}$	.06	$.27_a^{***}$	.07	$.30_b^{***}$	.05
Other property	$.16_{abc}^{*}$	.06	$.21_a^{\dagger}$	.12	$.16_{ab}^{**}$	.05
Combination of properties	$.20_{abc}^{**}$	.06	$.11_a$	.08	$.20_{ab}^{**}$	.06
Extension Activities						
Activities	$.25_a^{***}$	.06	$.28_a^{**}$	.10	$.24_a^{***}$	.05
No Activities	$.15_a^{***}$	.03	$.12_a^{**}$	.04	$.15_a^{***}$	.03
Control						
Business as usual	$.15_a^{**}$	.05	$.25_a^{**}$	.09	$.16_a^{**}$	.05
Non-RTL media	$.18_a^{**}$	.07	$.25_a^{*}$	.12	$.16_a^{**}$	.05
Alternate curriculum	$.19_a^{*}$	.09	$.26_a$	.19	$.17_a^{\dagger}$	.10
Assessment						
Custom narrow	$.33_b^{***}$	.06	$.36_b^{***}$	.10	$.33_b^{***}$	.05
Custom broad	$.16_{ab}^{**}$	.07	$.19_{ab}^{\dagger}$	.10	$.16_a^{*}$	.07
Standardized and validated	$.12_a^{***}$	.03	$.13_a^{**}$	.04	$.12_a^{***}$	.03
Grant-related						
Researchers funded by RTL						
Yes	$.21_a^{***}$	.04	$.25_a^{***}$	.07	$.21_a^{***}$	.03
No	$.21_a^{\dagger}$	.12	$.11_a$	.17	$.20_a^{\dagger}$	.10

Note. Model 1: Effect sizes and corresponding standard errors calculated for the full dataset. Model 2: Examining the 17 highest quality evaluations as per Slavin (1986). Model 3: Data corrected for clustering, as per procedures outlined in Hedges (2007) ($^{\dagger} p < .10$, $*p < .05$, $** p < .01$, $*** p < .001$). Different letter subscripts denote significant differences between moderators ($p < .05$).

As shown in Table 4, Model 1, effects varied somewhat across literacy outcomes. There were significant positive effects for phonological processing and vocabulary, and nonsignificant but positive effects for alphabet knowledge ($p = .07$), print concepts ($p = .051$), and tests of multiple literacy skills (i.e., tests simultaneously assessing more than one literacy skill; $p = .052$). Effects for vocabulary and phonological processing were significantly larger than alphabet knowledge and narrative comprehension.

Methodological Moderator Analyses: Model 1

Table 5. Continuous and Ordinal Moderator Effects

Moderator	Model 1		Model 2		Model 3	
	b	$SE(b)$	b	$SE(b)$	b_t	$SE(b_t)$
Sample-related						
% Low-income	.0030[†]	.0017	.0026	.0020	.0027[†]	.0014
% Caucasian	.0017	.0012	.0010	.0022	.0012	.0011
% African American	-.0010	.0011	-.0006	.0020	-.0006	.0010
% Hispanic/Latino	-.0015	.0012	-.0011	.0017	-.0015	.0012
% Asian	.0024	.0040	.0050	.0075	.0036	.0039
% Native American	.0016	.0015	.0538	.0473	.0016	.0015
% Female	.0009	.0047	.0045	.0083	.0004	.0056
% School-age	-.0017[*]	.0008	-.0005	.0013	-.0014[*]	.0007
% ELL	.0014	.0013	.0017	.0022	.0013	.0013
Duration of media consumption						
Weeks of exposure	.0002	.0022	.0292[†]	.0165	.0000	.0019
Grant-related						
Grant period	.10034[*]	.0483	.0368	.1083	.0868[†]	.0452

Note. Model 1: Meta-regression beta estimates and corresponding standard errors calculated for the full dataset. Model 2: Examining the 17 highest quality evaluations as per Slavin (1986). Model 3: Data corrected for clustering, as per procedures outlined in Hedges (2007) ([†] $p < .10$, *$p < .05$, ** $p < .01$, *** $p < .001$).

Sample-related moderators. Effects were relatively stable across sub-populations (see Table 5, Model 1). Results did not vary based on the racial composition or the proportions of girls or ELLs in a sample. However, evidence suggested media products were more effective for

preschool-aged children: A sample of all elementary school students would be expected to score about one-fifth a standard deviation lower than a sample of preschoolers. Interestingly, effects were nonsignificantly larger for samples with a higher proportion of low-income children ($p = .08$). A sample with all low-income children would be expected to score about one-third of a standard deviation higher than a sample of all affluent children.

Design and procedure-related moderators.

Study design, scope, setting, and length. Results were fairly homogenous across a variety of different study designs. As shown in Table 4 (Model 1), effects were positive and significant for between- and within-subjects designs; studies conducted in one and multiple geographic regions; and studies conducted in homes and schools. For the most part, results did not vary between these different designs. Effects likewise did not vary as a function of exposure duration, as can be seen in Table 5, Model 1. Nevertheless, effects were larger for within-subjects designs than between-subjects designs.

Nature of treatment.

Media platform. Both studies focused solely on the impact of television and those that included new media components yielded significant and positive effects (see Table 4, Model 1).

Property. Across platforms, all properties, excluding *Between the Lions* and *Martha Speaks,* yielded effects greater than zero. (see Table 4, Model 1). Nonetheless, the effects for *Martha Speaks* were still estimated to be positive, with $p = .051$. Some media properties produced larger impacts than others. Effects for *Sesame Street* were significantly larger than *Between the Lions* and *Martha Speaks* and nonsignificantly larger than "other" properties examined in fewer than five evaluations (which included properties such as *Barney and Friends,*

The Electric Company, and *WordWorld*; $z = 1.95$, $p = .051$) and combinations of multiple

properties ($z = 1.77$, $p = .08$). Likewise, the effects for *Super WHY!* were significantly larger than

Between the Lions and nonsignificantly larger than *Martha Speaks* ($z = 1.90$, $p = .06$).

Supplemental Activities or Materials. Across platforms and properties, effects were

significant and positive for both studies with and without supplementary materials or activities,

as illustrated in Table 4, Model 1.

Nature of Control. As shown in Table 4, Model 1, when examining the subset of studies

with between-subjects designs, effects were significant for evaluations that allowed the control

group to go about business as usual, asked control children to consume media created without

RTL funding, and provided control children with an alternate (non-mediated) curriculum.

Nature of assessment. Effects were positive and significant for custom narrow, custom

broad, and standardized measures (see Table 4, Model 1). However, effects were larger for

custom narrow measures than standardized measures.

Grant-related moderators.

Time trend. As demonstrated in Table 5, Model 1, effects grew over the course of RTL's

history. A child who participated in an intervention during the most recent past grant period

(2010-2015) would be expected to score about two-fifths of a standard deviation higher than a

child who participated in the first grant cycle (1994-2000).

Results by funding status. As mentioned previously, nearly three-quarters of the

evaluations in the sample were funded by RTL. Only these RTL-funded evaluations produced

statistically significant mean effects. Nonetheless, effects for non-funded evaluations were

identical in magnitude (i.e., $d = .21$ for both funded and non-funded evauations); the standard

error was larger for non-funded evaluations such that these evaluations did not achieve conventional levels of statistical significance ($p = .07$).

Robustness Checks

  Alternate Models. The robustness of findings was assessed by examining two additional, conservative analytical models (Models 2 and 3). Findings consistent across all three models (main Model 1 + robustness Models 2 and 3) provide strong evidence supporting patterns in the data, while findings only present in one model should be interpreted with more caution.

  For Model 2, evaluations were screened for rigor using criteria suggested by Slavin (1986). Specifically, evaluations that were in the smallest quartile in terms of sample size (fewer than 51 participants) and duration (shorter than 3.5 weeks), evaluations that failed to use random assignment, and evaluations that failed to specify these three criteria were removed. This left 17 high quality evaluations for subsequent analyses ($m = 29$, $k = 252$), which included data for all moderators except measures of multiple literacy skills.

  In Model 3, all analyses were rerun on the full dataset, attenuating estimates to account for clustering when relevant. As Hedges (2007) explains, many studies assessing learning outcomes sample by large aggregated units like schools, where children are clustered within classrooms and thus might all achieve similar scores regardless of intervention. Indeed, most studies in this dataset used clustered sampling ($n = 30$, $m = 57$, $k = 601$). To provide more conservative estimates of effects in such cases, Hedges (2007) outlined procedures that attenuate estimates, which were used to estimate Model 3. In doing so, a compendium of K-12 intraclass correlation (ICC) values Hedges and Hedberg (2007) estimated for such purposes was consulted, using kindergarten ICCs for the samples in the present dataset composed of preschoolers.

Across both robustness models (see Tables 4 and 5, Models 2 and 3), the general pattern of results was largely unchanged. The overall estimated effect size was significant, positive, and roughly comparable in magnitude for both Models 2 ($d = .23$, $SE = .06$, $n = 17$, $m = 29$, $k = 252$, $p < .001$) and 3 ($d = .21$, $SE = .03$, $n = 45$, $m = 79$, $k = 783$, $p < .001$). However, some particulars varied across models. For literacy skills, in Model 2, there was no longer a significant difference between alphabet knowledge vs. vocabulary and phonological processing, although there was now a nonsignificant difference between vocabulary and tests of multiple skills, with the effects on vocabulary greater than multiple skills ($z = 1.86$, $p = .06$). In Model 3, the estimated effect for narrative comprehension was significantly smaller than all other literacy skills.

Regarding sample characteristics, the significant effect for school-age children from Model 1 disappeared in Model 2. While still nonsignificant, given RTL's focus on low-income samples, it is worth noting that the beta estimate for low-income samples was similar in magnitude across models ($p = .056$ in Model 3).

Concerning study design, in Model 2, there was a nonsignificant relation between length of an intervention in weeks and estimated effect size ($p = .08$), such that estimated effects increased for each week of intervention. Regarding the treatment itself, in Model 2, the effects for *Sesame Street* were estimated to be greater than any other property. However, in Model 3, the difference between *Sesame Street* and *Martha Speaks* observed in the prior two models disappeared. For the control activity, effects for studies in which control children used alternate curricula were nonsignificant in both robustness models (although $p = .08$ in Model 3). In terms of assessments, the estimated effect for custom broad measures was no longer significant in Model 2 ($p = .054$), and in Model 3, custom broad measures yielded significantly smaller effects

than custom narrow measures.

Finally, concerning grant-related factors, the estimate for grant period was no longer significant in the robustness models (however, $p = .055$ in Model 3).

Publication bias. While many evaluations in this meta-analysis were unpublished reports, it is possible the present dataset still was biased by missing data (see Borenstein et al., 2009 for a discussion of bias in meta-analysis), especially if researchers not accountable to DoEd or content producers opted not to share null findings. However, the average effect size of published ($d = .27$, $SE = .08$, $n = 18$, $m = 33$, $k = 217$, $p = .001$) and unpublished studies ($d = .24$, $SE = .06$, $n = 30$, $m = 46$, $k = 566$, $p < .001$) were comparable, $z = .28$, $p = .78$, suggesting publication bias was not an issue for the present dataset. Rosenthal's (1979) Fail-safe N, which determines how many additional null effects would need to be added to a meta-analytic dataset to "nullify" the estimated effect, suggested it would take 173,218 additional null effects before results would fail to achieve statistical significance. This high number also indicates publication bias or any other bias caused by missing data may not be problematic in this dataset. As shown in Figure 1, a funnel plot of the results provided a visual measure of bias (Light & Pillemer, 1984). The placement of effects in this plot was generally symmetrical and Egger's test was nonsignificant ($p = .15$), also suggesting the dataset did not suffer from bias related to missing data (Egger, Smith, Schneider, & Minder, 1997).
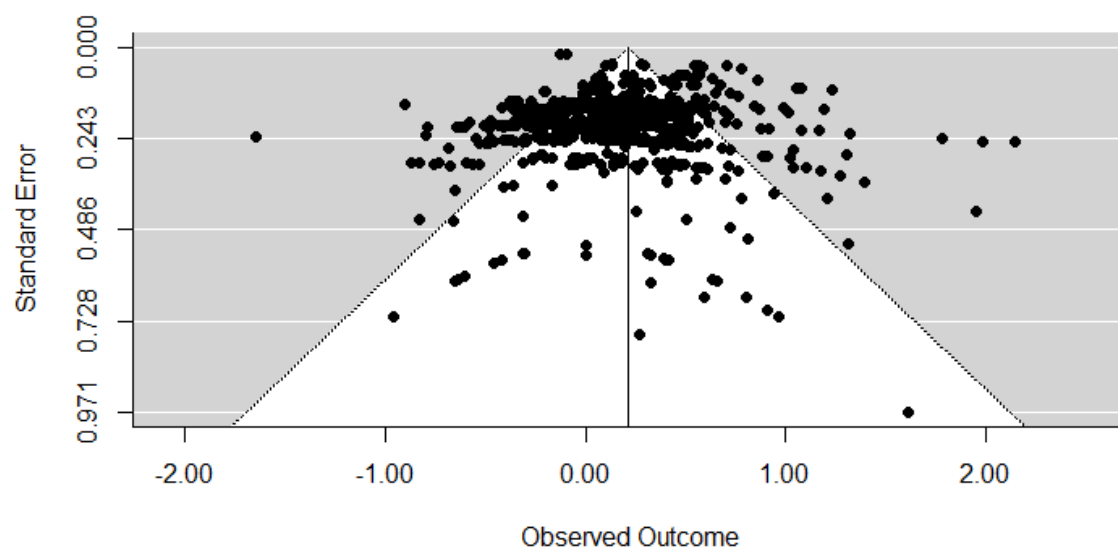
Figure 1. Funnel Plot of Each Study's Mean Effect Size by Study-Average Standard Errors

Discussion

Overall, this meta-analysis suggests a positive effect of exposure to RTL media on children's early literacy (especially their vocabulary and the harder-to-promote skill area of phonological processing). Findings were fairly robust across research designs, although effects were somewhat more pronounced in within-subjects designs and in cases where researchers used custom narrow assessments closely mirroring focal media content. Several different media properties yielded effects on par with *Sesame Street.*

Interpretation of Overarching Findings

There are multiple ways to interpret the real-world meaning of these findings (see Cooper, 1981). For education interventions, the What Works Clearinghouse (2014) recommends translating effect sizes into an "improvement index", or an expected change in percentile rank resulting from intervention. This entails converting an effect size into a z-score, interpreting that

z-score as a U3 index (the area under the standard normal curve corresponding to said z-score), and subtracting off 50% (the area under the curve corresponding to "average" performance not impacted by treatment). For the present-analysis, the improvement index would be 8.32. In other words, being exposed to RTL media corresponded to gains in literacy achievement of about 8 percentiles. Judging against normative growth rates set forth by Hill and colleagues (2008), the estimated effects obtained herein are equivalent to about one-and-a-half months of literacy learning above and beyond typical growth.

The results in this study were on par with those obtained in meta-analyses examining computerized reading interventions (Langenberg et al., 2000) and versions of *Sesame Street* produced in 15 countries outside the U.S. (Mares & Pan, 2013). The present effects also were comparable to meta-analyses probing into the impact of comprehensive U.S. early childcare (e.g., state preK, Perry Preschool, and Head Start; Kay & Pennucci, 2014). However, the current effects were considerably smaller than those reported in other meta-analyses of sometimes longer, largely nonmediated targeted literacy interventions focusing on programs where teachers or researchers taught children reading strategies or where researchers taught parents coaching strategies like dialogic reading (e.g., Bus & van Ijzendoorn, 1999; Mol et al., 2008).

Implications for RTL Stakeholders

RTL media had small but positive impacts on children's early literacy, providing fairly positive accountability evidence in favor of RTL. The effects of consuming content from some newer properties (e.g., *Super WHY!)* for the most part seemed to be about equivalent to the effects of RTL legacy property *Sesame Street*, which previously had been considered the gold standard for educational children's media (Fisch, 2004; Guernsey & Levine, 2015). RTL also

successfully promoted vocabulary acquisition, replicating prior literature (Moses et al., 2010), and impacted the multifaceted skills of phonological processing, an area previous researchers supposed might not be sensitive to media intervention (Gola et al., 2012). Given the particularly strong link between phonological processing and later reading competence (Grant et al., 2012), this finding is especially noteworthy. Moreover, RTL media appeared to be just as, if not more, effective across grant cycles.

Most of the demographic findings suggested in this meta-analysis also are consistent with RTL's goals. RTL media were equally effective for both boys and girls, ELLs and native speakers, and children of a variety of different racial-ethnic backgrounds. There was also evidence that RTL media were as, or more, effective for low-income audiences (an RTL target population; Michael Cohen Group, 2012). In light of past work, which suggested educational children's media might be less effective for low-income learners (e.g., Cook et al., 1975), this finding is quite encouraging. Although inconsistent with RTL's aim to reach children age 8 and younger, it is not surprising that media were less impactful for school-age audiences, especially given prior meta-analytic work reporting similar age-based findings for a variety of literacy interventions (Langenberg et al., 2000; Mol et al., 2008). RTL evaluators noted the school-age children in many samples were already scoring at ceiling at pre-test, especially for simpler measures such as assessments of alphabet knowledge (e.g., Penuel et a., 2012). In a similar vein, literacy research more generally suggests that children typically master many skills related to print concepts and phonological processing by kindergarten or first grade (Grant et al., 2012). Additionally, school-age children might be more likely to receive formal literacy instruction in school and have less to gain from educational media than their younger counterparts.

When considering RTL's overall effectiveness, however, one must consider findings indicating that (a) effects were not consistently significant when the control group engaged with non-mediated alternate curricula; (b) only RTL-funded researchers found significant outcomes on average; and (c) effects were largest for custom narrow assessments mirroring the content presented in media. When considering findings about differences based on control task, it is important to keep in mind that in all three models, there was no significant difference in estimated effects between studies where control groups used alternate curricula or media, or where control groups engaged in business as usual. Given that only 7 evaluations used alternate curricula, there may be an issue of statistical power for this moderator. Regarding the link between funding status and outcomes, it similarly is important to note that the effects for RTL and non-RTL researchers were generally comparable in magnitude, and that the difference between these two groups of researchers never achieved statistical significance. Nonetheless, it is possible that RTL researchers might have conducted studies in qualitatively different ways more likely to yield positive results (e.g., invested more resources creating custom narrow measures) or simply may have been more skilled than some of their non-RTL-funded counterparts (qualifying them to be awarded competitive RTL research subcontracts). Alternatively, given that only 12 evaluations were authored by teams without RTL funding, it again is possible that the effects for non-RTL researchers might have failed to achieve significance simply due to a lack of power. It would be helpful for more fully independent researchers (i.e., researchers not accountable to RTL grant recipients or in any way benefiting from RTL funding) to elect to use media from some of the properties identified in this review in future studies to help clarify this.

Concerning effects being larger for custom narrow assessments, this is unfortunately a

common weakness of intervention research across sub-fields. In evaluations of both mediated and nonmediated literacy interventions, researchers tend to rely heavily on such measures, which in turn yield larger effects than validated measures (Anderson & Collins, 1988; Langenberg et al., 2000). These measures may reflect partial or otherwise biased manifestations of the skillsets they are intended to measure. In this particular dataset, many of the standardized assessments researchers administered or emulated in custom broad measures were criterion-referenced and thus not necessarily sensitive to growth over short interventions (e.g., Invernizzi et al., 2004). It therefore may possible that the effects for standardized and custom broad measures would have been larger if researchers solely relied on norm-referenced tests. Then again, it might be unrealistic and perhaps somewhat frightening if 6-14 weeks of media exposure resulted in large and rather generalized gains (Cohen et al., 2016).

Moving forward, RTL funders and producers may wish to consider continuing to emphasize skills not as well supported by the current body of RTL media, such as narrative comprehension. Although it may be more challenging for an early learning program to effect change in this area, given that it is a more advanced skill than many of the other literacy skills targeted by RTL (Grant et al., 2012), previous research suggests narrative comprehension can be promoted via a variety of narrative storytelling formats (Linebarger & Piotrowski, 2009).

Implications for Developmental Scientists

These findings also speak to broader questions in scholarship on children and media (e.g., the value of Experiential Mediation). Although the findings of this meta-analysis were fairly robust across a variety of research designs, results nonetheless underscore conventional wisdom and theory about designing studies of this nature on some fronts. As might be intuitive, there is

weak evidence that longer interventions and strong evidence that interventions without comparison groups result in larger effects. Additionally, consistent with prior scholarship on learning from media (Fisch, 2004) and on learning more generally (Hill et al., 2008), the present results suggest children's learning might appear greatest when they are given assessments closely mirroring the educational content they have consumed, as it is challenging for them to generalize lessons learned in one context to assessments that do not closely resemble that context.

It may be frustrating to some that these data do not resolve debates into whether television or new media are more effective in promoting literacy. Because RTL producers have been on the forefront of creating content for new platforms, it may be that children and caregivers struggled to fully utilize media products running on newer platforms or that families lacked needed resources such as reliable Internet access (Llorente et al., 2010). Relatedly, RTL producers may have faced a steep learning curve in designing effective new media products, given that designers and scholars alike are only just beginning to explore the new media design principles that best support learning without boring or distracting children (Hirsh-Pasek et al., 2015). Consequently, device and service penetration may "catch-up" with the set of RTL new media products and the products themselves may continue to improve, eventually providing more of an advantage to new media.

Somewhat surprisingly (see Horowitz et al., 2005), effects were comparable for studies conducted in both one and multiple geographic regions. This suggests that with diligence, evaluators can indeed execute studies in more than one site.

The lack of support for the benefits of extension activities or materials may seem quite unexpected, failing to support literature on Experiential Mediation (Piotrowski et al., 2012) and

Joint Media Engagement (Takeuchi & Stevens, 2011). However, some RTL researchers

suspected adults facilitating interventions sometimes failed to implement extension activities as

intended by media producers (e.g., Horowitz et al., 2005). Also, it is possible the extension

materials researchers provided did not always complement mediated lessons. RTL producers

have continuously refined their development of extension resources to try to make them more

caregiver-friendly (Llorente et al., 2010), but these improvements may not be reflected in the

present dataset. Anecdotally, in studies where researchers explicitly laid out curricula with

detailed instructions on extending mediated lessons and when facilitators followed these

instructions, effects were quite large (e.g., Neuman et al., 2011).

Limitations

These results should be interpreted in light of three limitations. First, because this study

focused on a U.S.-based grant, all the studies in this sample were conducted in the U.S. Even

though these results are aligned with recent international meta-analyses of similar media

products (Mares & Pan, 2013), results obtained here may not generalize to all languages or

cultures. Second, there was a limit to questions this meta-analysis could answer. After all, only

so much nuance can be captured via nominal moderators. For example, some extension activities

or materials might have been stronger than others, but such qualitative variation is not reflected

in the present analyses. Likewise, some media properties might be more successful at enacting

the production and design principles described in the introduction, but this study did not engage

in a content analysis of the media itself to be able to directly address this possibility. Third, the

results reported above do not account for dependence between measures (i.e., that two effect

sizes from the same comparison are likely correlated). Because 38 evaluations did not provide

information on the correlation between measures and because methodological work suggests the impact of ignoring dependence between effects is minor for large datasets like this one (Kim & Becker, 2010), this decision seemed justified. Indeed, in yet another alternate model (not reported herein), in which covariances between effects within comparisons were imputed from information in the 7 evaluations that did provide some correlational information, the results were nearly identical to those of Model 1. However, it is possible that results may have varied more substantially had fuller correlational information been available.

Conclusion

The results from the present study provide evidence that consuming RTL media is a valuable use of children's time, especially for preschoolers, and that spending on initiatives such as RTL is a valuable use of taxpayer dollars. These findings demonstrate RTL media can successfully promote early literacy skills (e.g., vocabulary; phonological processing), and that these findings are robust across a variety of different research designs and under a variety of different assumptions.

Article 2: "Are You Sure My Child Participated in that Study?" Recruiting and Re-Recruiting for Developmental Research

Abstract

Recruiting children and families for research studies can be challenging, and re-recruiting former participants for longitudinal research can be even more difficult, especially when a study was not prospectively designed to encompass continuous data collection. This article discusses how researchers can set up initial studies to potentially facilitate later waves of data collection; locate former study participants using newer, often digital, tools; schedule families using recruitment scripts that highlight the many benefits to continued study participation; and confirm appointments using other newer digital tools. This article reflects on my experience conducting a non-prospective longitudinal study with urban parents and their early adolescent children, as well as prior methodological pieces and longitudinal articles. The primary aim is to provide suggestions to others wishing to re-recruit families for longitudinal studies; however, many of the strategies shared in this article also could enhance cross-section recruitment.

Keywords: recruitment, methods, developmental research, longitudinal

Across developmental science subfields, researchers have called for more longitudinal research (e.g., Karmiloff-Smith, 1998; Wartella et al., 2016). Such research provides compelling information about developmental trajectories, patterns, sequences, and pathways (Nicholson, Sanson, Rempel, Smart, & Patton, 2002).

However, conducting longitudinal research presents formidable challenges, chief among which is re-recruiting sizeable samples across waves of research (Cotter, Burke, Loeber, & Navratil, 2002; Ribisl et al., 1996). Recruiting for cross-sectional studies with children is already quite demanding, requiring researchers to devise thoughtful sampling schemes (Bornstein, Jager, & Putnick, 2013), set up databases tracking contact with families, convince families to follow through and schedule study appointments, and confirm these appointments in a sensitive manner, all while working around families' busy schedules (Striano, 2016). In addition to these challenges, re-recruiting for longitudinal work presents the added difficulty of working with a finite pool of families who might become fatigued from participating in multiple waves of research or who might have moved or changed contact information over time (Agrawal, Kellam, Klein, & Turner, 1978; Barakat-Haddad, Elliott, Eyles, & Pengelly, 2009; Cotter et al., 2002).

In the present paper, I first provide general background information on developmental scientists' interest and success in re-recruiting for longitudinal work. I then describe my own recent foray re-recruiting families who participated in a brief educational intervention in 2010 for a follow-up study in 2016. I next outline strategies for arranging studies to facilitate later waves of data collection, locating original families, and scheduling them for new testing sessions. I contextualize these strategies in light of previously shared wisdom in this area. My primary aim

is to inform the work of researchers conducting longitudinal research; however, many of the suggestions I share also could enhance cross-sectional participant management.

Longitudinal Research in Developmental Science

Numerous developmental scientists are attempting longitudinal research, recruiting participants for a study at one time, and then re-recruiting them for subsequent waves of research at later time points. Large-scale multi-disciplinary cohort and panel studies such as the Dunedin Multidisciplinary Health and Development Study (e.g., Silva, 1990; Stanton & Silva, 1992) and studies falling under the umbrella of the Early Childhood Longitudinal Study (ECLS) (e.g., Heatly, Bachman, & Votruba-Drzal, 2015) are the most renowned sort of longitudinal work in child development. Such studies require extensive and continuous planning and resources (Nicholson et al., 2002).

In addition to these large-scale efforts, many other scholars have conducted smaller scale follow-up studies, sometimes deciding to re-recruit families who participated in one (often lab-based) study after extensive periods without any contact. Researchers have conducted such longitudinal work across a host of domains, including spatial cognition (e.g., Lauer & Lourenco, 2016), language (e.g., Can, Ginsburg-Block, Golinkoff, & Hirsh-Pasek, 2013), temperament (e.g., Schwartz, Wright, Shin, Kagan, & Rauch, 2003), personality (e.g., Harris, Brett, Johnson, & Deary, 2016), self-regulation (e.g., Ayduk et al., 2000), health and physical development (e.g., Fein, Li, Chen, Scanlon, & Grummer-Strawn, 2014), mental health (e.g., Agrawal et al., 1978), and media use (e.g., Hanson, 2017). Similarly, researchers who evaluate interventions also have engaged in comparable, sometimes non-prospective, longitudinal research years after the conclusion of interventions. Such studies help assess the long-term impact of programs

attempting to directly improve children's outcomes in areas such as education (e.g., Campbell et al., 2012) and health and fitness (e.g., Lazorick et al., 2014), and indirectly influence child outcomes by providing parents supports such as cash supplements (e.g., Huston et al., 2005) and drug counseling (e.g., Haggerty et al., 2008). Thus, a variety of developmental scholars are responding to calls for longitudinal research, even though in practicality there are often sizeable gaps in time between waves of research in these studies.

Without the resources of large-scale cohort and panel studies, researchers have met the challenges of re-recruiting original study participants with varied success. Some report it to be particularly challenging re-recruiting urban individuals with racial-ethnic minority and low socio-economic status (SES) backgrounds (e.g., Barakat-Haddad et al., 2009; Fein et al., 2014; Poehlmann-Tynan et al., 2015; Ribisl et al., 1996), older children (e.g., Cotter et al., 2002; Cotter, Burke, Stouthamer-Loeber, & Loeber, 2005), and individuals with common names (e.g., Barakat-Haddad et al., 2009; Masson, Balfe, Hackett, & Phillips, 2013). Conversely, others suggest very high-SES families are especially challenging to re-recruit (Silva, 1990). In the scholarship reviewed in the preceding paragraph, researchers' re-recruitment rates, when reported or inferable across publications, ranged from less than 15% (Barakat-Haddad et al., 2009; Schwartz et al., 2003) to nearly 100% (Campbell et al., 2012; Silva, 1990). However, researchers often fail to report retention rates or information that can be used to infer these rates, or to even use consistent definitions of what they consider to be successful retention (e.g., locating a parent AND child versus only locating the parent; Ribisl et al., 1999).

Along these lines, scant guidance exists for re-recruiting families for longitudinal research. A small number of developmental researchers have shared strategies to facilitate

multiple rounds of data collection, foremost among which include (a) using contact information

participants provided at the beginning of the original study (e.g., Ayduk et al., 2000), (b)

employing—often fee-based—databases to find updated contact information (e.g., Haggerty et

al., 2008), and (c) providing increasing compensation across waves of data collection (e.g.,

Cotter et al., 2002), among other techniques described in more detail in the following pages.

Nonetheless, typically, published studies in child development focus more on the measures

administered across waves of data collection than the recruitment process. Much of the existing

re-recruitment guidance exists in articles targeting general populations, with advice not always

relevant to children and families (e.g., Ribisl et al., 1996), or in clinical- or practitioner-oriented

journals, which often include strategies very specific to small, special populations of children

(e.g., Masson et al., 2013). The primary goal in the present manuscript is to extend this body of

literature by outlining strategies for re-recruiting a wide variety of children and families for a

wide variety of potential follow-up studies. To do so, I reflect on my own recent experiences re-

recruiting urban families for a longitudinal study.

Current Context

I recently re-contacted families (original $N = 136$) who had participated in an 8 week

educational computer game intervention when children were in preschool and kindergarten (M

age at Time 1 = 5.24 years, $SD = .71$). I collected follow-up data from former study participants

six years after the original intervention when children were in late elementary school (M age at

Time 2 = 11.28 years, $SD = 1.30$). This was not a prospective study that was planned from the

outset of the original intervention; consequently, the research team had no intermediate contact

with families between the intervention and follow-up study. The original sample was American,

urban, and racially and ethnically diverse (28% Caucasian, 28% Hispanic, 20% African American, 24% Other or Mixed race), with approximately 60% of families with incomes below $40,000 and 45% receiving or eligible for government aid. Given the literature suggesting older children (Cotter et al., 2002), individuals from urban communities and low-income or racial-ethnic minority backgrounds are particularly difficult to re-recruit (Ribisl et al., 1996), I recognized the challenge ahead of me in re-recruiting for this project.

I had modest financial but considerable human resources at my disposal. My budget was a little over US$5,000 to recruit, travel to, and compensate participants. I also had part-time re-recruitment manpower from a faculty member, lab coordinator, doctoral student (who was collecting this data for her dissertation), and five undergraduate research assistants. The faculty member had conducted a similar longitudinal study nearly 20 years ago, before dramatic increases in the popularity of newer interpersonal communication tools, such as text messaging and social networking sites (Duggan, 2013; Purcell, 2011).

Preparing Initial Studies to Facilitate Later Waves of Data Collection

Researchers may not always know at the outset of an initial study if they will be able to conduct subsequent rounds of data collection, for a variety of reasons including tenuous funding (Nicholson et al., 2002). Likewise, the research literature may prompt new questions that were not under consideration as part of the original research, but which could be addressed by re-recruiting original participants. Indeed, this was the case for us. Even in these situations, there are several steps researchers can take to facilitate potential future rounds of data collection. First, researchers may partner with local organizations, including schools (see Bornstein et al., 2013; Ribisl et al., 1996). These relationships can assist with both initial recruitment (Striano, 2016)

and re-recruitment, depending on the initial language used in establishing said partnership and the nature of the study itself (see the discussion on locating participants below for more information).

Second, researchers can create a database with detailed participant contact information, including information such as participants' and their family members' full names and aliases (e.g., nicknames, maiden names), phone numbers, emails, mailing addresses, educational and employment information (if relevant), contact information for friends, neighbors and/or relatives, birthdates, plans to move or change names, physical descriptions, and favorite hangouts (Barakat-Haddad et al., 2009; Cotter et al., 2002; Haggerty et al., 2008; Masson et al., 2013; Ribisl et al., 1996). During original data collection, the research team collected names, phone numbers, email addresses, and employment information for up to two parents; full names, birthdates, and preschool/kindergarten names for children; and mailing addresses for the entire family. This helped ensure high participation over the course of the original intervention, and, as described in more detail below, was invaluable to subsequent re-recruitment efforts.

Third, researchers can adopt a blanket policy of including an optional element on all consent forms seeking permission for future contact about later research opportunities (e.g., Masson et al., 2013). This easy addition to consent forms provides an opportunity for future outreach should the need arise. In the absence of this initial consent, follow-up contact unexpectedly may be perceived as a privacy violation, and prevent former participants from choosing to dedicate more time to a research project. Further, university Institutional Review Boards (IRBs) may have an ethical obligation to prevent non-consensual follow-up contact. Indeed, this seems to be a common standard across universities in the U.S. (e.g., J. Hecht,

personal communication, July 1, 2016). It is important researchers take these initial safeguards,

because children are special vulnerable populations (Hartmann, 1992).

Fourth, researchers should attempt to build rapport with both parents and children across

research activities (Cotter et al., 2002) and establish clear branding via the use of university or

other logos (Haggerty et al., 2008; Ribisl et al., 1996; Striano, 2016). Depending on the level of

anonymity of the study, researchers could even establish a Facebook or similar social media

group for participants. Having a blog or newsletter to update families about study findings also

can be a way to maintain connections with families who have participated in prior studies. In

prospective studies, researcher sometimes send participants regular newsletters and

birthday/holiday greetings (Ribisl et al., 1996). Such techniques can help ensure the research

experience is positive for participants and help them identify with the research study (and

perhaps sign-up for other cross-sectional studies with researchers even if a team does not attempt

to follow up with a specific study).

Finally, researchers with sufficient foresight in certain sub-domains may wish to gain

consent from parents to contact their family members, friends, neighbors or other professionals

in their lives (e.g., clergy; case workers) for help locating them (i.e., participating parents and

children) at a later point, and to have parents prepare notes for these individuals consenting for

them to provide current family contact information (Cotter et al., 2002; Passetti, Godley, Scott,

& Siekmann, 2000; Ribisl et al., 1996). These permissions could later be leveraged to assist with

locating study participants for later waves of data collection. As a caveat, this approach may not

be appropriate for all topic areas. For example, seeking this information might be reasonable as

part of a lengthy intervention but could be intrusive in a one-time lab session.

Re-Recruiting for Longitudinal Research

In the discussion of re-recruitment for later waves of longitudinal research, I distinguish between *locating,* and actually *scheduling* families and engaging them in study *participation*. I consider a family to be *located* if a caregiver responds to re-recruitment attempts verbally or in writing, and acknowledges that I have identified the correct family, regardless of their level of interest in participating in the follow-up study (Haggerty et al., 2008). In contrast, I consider a family *to be scheduled and to have participated* in the study once they have arranged a time to meet with researchers and followed through with these appointments. Families who are difficult to locate are not necessarily difficult to schedule: In this study, the time I spent searching for a given family was negatively but nonsignificantly correlated with the time it took to schedule said family once located ($r = -.20$, $p = .06$).

Although I located, scheduled, and engaged families in study participation simultaneously, others have recommended dedicating some amount of time to locating participants prior to moving to the scheduling phase (Ribisl et al., 1996). That is, researchers may be well served by first locating as many families as they can and then scheduling them for follow-up appointments. Researchers may find certain families more time consuming to recruit than others (Masson et al., 2013), which in turn can prevent researchers from being able to speedily collect all data in a finite time period. When separating the locating and scheduling phases, families might receive a small monetary reward simply for providing updated contact information (followed eventually by a larger reward for participating and contributing data to subsequent waves of data collection; Ribisl et al., 1996). Additionally, more resources can be dedicated to families that researchers suspect might be challenging to locate based on prior

interactions (Cotter et al., 2002), especially in light of research suggesting some families are consistently more challenging to re-recruit across multiple waves of research than others (Fein et al., 2014). In hindsight, I think this approach would have served me well and would be valuable for other developmental researchers, in that doing so might help researchers conduct all testing sessions succinctly once they have located most target participants. Testing easier-to-locate children first and testing harder-to-locate children months later might allow significant, potentially confounding developmental differences to arise.

The remainder of this section discusses strategies for locating and scheduling families. The discussion of the former topic is particularly germane to those conducting follow-up studies, but the discussion of the latter topic includes strategies that also can be applied during the first round of data collection in a longitudinal study and during cross-sectional studies.

Locating Study Participants

The research team and I located 122 (90%) of the original participants in the present study. We waited at least two weeks between attempts to reach the same family, and made efforts to initiate contact during a variety of times of day. Since participating children were in late elementary school at Time 2, many lived in households where all caregivers had fulltime jobs, and accordingly, caregivers sometimes were more receptive to calls made outside of business hours (see Striano, 2016). However, other parents in the sample worked nontraditional jobs or hours, and thus were easier to reach midmorning. On average, it took 22.9 days ($SD =$ 35.04) or about 3 days of outreach attempts from the time the team first tried to locate families via any means to when we successfully located them. As is evident by the large standard deviation, there was considerable variability in the ease of locating families: We located 67

families (55% of located families) in less than one week (i.e., following one contact attempt day), 25 families (20% of located families) within one month (i.e., following about two contact attempt days), and the remaining 30 families (25% of located families) after one month. We may have been able to locate some families somewhat more rapidly had I not also been collecting data simultaneously. Located families had higher incomes, were more likely to be Caucasian, and were less likely to be of Other/Mixed race-ethnicity than non-located families ($ps < .05$). Otherwise, there were no differences between located and non-located families.

As will be described in more detail below and in Table 1, families were located using the contact information they provided at Time 1, and searching for updated contact information via free and paid tools. Below, I also point to strategies I did not believe would be successful for this study but that other researchers may find useful.

Originally provided contact information. The team used originally provided contact information to call, text, email, and/or mail parents (e.g., Anderson et al., 2001; Ayduk et al., 2000).. In contrast to other studies, where researchers only saw moderate success leveraging originally provided contact information (e.g., Haggerty et al., 2008), I found that reaching out to caregivers using the information they provided in the original study was by far the most successful location strategy employed. Traditional phone calls were particularly fruitful. Although many of the participants in the original study had moved from the specific homes they lived in at Time 1, most were still living in the greater metropolitan area and had retained their cell phone number or email address. Had the team exclusively relied on original contact information, we still would have located over 70% of the original participants, which some would argue is an acceptable retention rate (Ribisl et al., 1996). This thus could have saved many

Table 6. Techniques for Locating and Scheduling Participants (N = 136 original study participants)

Contact Method	Located			Participated			Days Taken to Locate via Successful Method		Days Between Location and Participation	
	n	% of original participants	% of located participants	*n*	% of original participants	% of participating participants	*M*	*SD*	*M*	*SD*
Original contact info.	101	74%	83%	86	63%	87%	11.26	21.54	21.98	28.14
Phone	76	56%	62%	62	46%	63%	11.21	23.62	22.92	29.58
Call	72	53%	59%	59	43%	60%	11.78	24.01	23.45	30.18
Text	4	3%	3%	3	2%	3%	1.00	0.00	12.33	8.50
Email	24	18%	20%	24	18%	24%	11.58	14.71	19.39	24.10
Mail	1	1%	1%	0	0%	0%	7.00			
Contact info. via free sources	18	13%	15%	13	10%	13%	14.72	13.81	19.38	28.69
Phone (calls)	8	6%	7%	7	5%	7%	13.63	14.60	13.29	10.75
Email	6	4%	5%	4	3%	4%	14.00	11.10	38.00	48.39
New email	4	3%	3%	2	2%	2%	22.33	5.86	17.50	6.36
New Gmail	2	2%	2%	2	2%	2%	1.00	0.00	58.50	72.83
Social media (Facebook)	4	3%	3%	2	2%	2%	18.00	18.96	3.50	2.12
Contact info. via paid sources	3	2%	3%	2	2%	2%	5.33	3.79	1.50	0.71
Phone (calls)	1	1%	1%	0	0%	0%	1.00			
Mail	2	2%	2%	2	2%	2%	7.00	9.00	1.50	

Note. Number of families who first responded to recruitment attempts via each contact method, percentage located of the original sample (N =136) and percentage of located sample (n = 122) responsive to each method, number of families who participated in the follow-up after responding to each contact method, percentage of the original sample (N = 136) and percentage of participating sample (n = 101) who participated after responding to each contact method, days taken to locate via successful method (the average number of days between when researchers first attempted to contact families and when they finally located them, categorized by the method that ultimately provided to be successful in locating them (where the day initially contacted = day 1)), the standard deviation on that number, days between location and participation (the average number of days between when researchers first located a family and when the family participated in the study per method), and the standard deviation on that number.

manpower hours attempting the strategies outlined below. Growing reliance on cell phones means that individuals, especially in urban areas such as where I collected my data and especially for higher-SES populations, are increasingly maintaining the same phone number even after they relocate (Dost & McGeeney, 2016). This may in part explain some of my team's success using several-years-old phone numbers to re-recruit relative to prior research conducted before increased cell phone penetration.

Searching for participants via free online tools. Multiple free strategies were employed to ascertain updated contact information in cases where originally provided information was outdated. First, a variety of free search engines and databases were leveraged (e.g., Cotter et al., 2002; Passetti et al., 2000). Again somewhat contrasting with prior work, where free resources only yielded a small amount of accurate contact information (e.g., Haggerty et al., 2008; Masson et al., 2013), this was the most successful search avenue for in the present study after exhausting original contact information. Mirroring recent prior similar studies (Barakat-Haddad et al., 2009; Cotter et al., 2002; Fein et al., 2014; Haggerty et al., 2008), general search engines such as Google, online databases known for contact information (e.g., the National Change of Address database) and people-specific search engines (i.e., search engines designed to scrape the Internet

just for publicly available contact information) were all leveraged. Whitepages.com, pipl.com, zabasearch.com, , and RefereneceUSA.com yieldied the most fruitful results, although be cautioned that there is high turnover across such websites (for a similar warning, see Masson et al., 2013). Several websites recommended in prior, relatively recent methods pieces were no longer live during the present data data collection. I suspect my team's success with free resources may in part be attributable to the large volumes of contact information some people now post online through personal websites, public social network profiles, and the like (Rainie, Kiesler, Kang, & Madden, 2013), and to newer blogs and articles showcasing high quality free people-centric search engines (e.g., Boswell, 2007). It also may be that the algorithms these services use have improved over time.

Second, my team tested plausible alternate email addresses. We located a small number of families by pairing the local part of email addresses parents had provided at Time 1 with Gmail domains (e.g., If a mother indicated her email address was something like sparklequeen78@sbcglobal.net at Time 1, we attempted to contact her at Time 2 using sparklequeen78@gmail.com). I noticed early into recruitment, before we began implementing this strategy, that several parents had changed their emails along these lines. Additionally, I know from outside reports that Gmail grew in popularity in between the time we originally collected data and when I conducted the second wave of data collection, while Yahoo, Hotmail and AOL declined in use (Creager, 2011; Dupre, 2014; Khan, 2015). More broadly, because people frequently use the same usernames repeatedly across accounts (Jacobsson Purewal, 2015), researchers may succeed in locating participants by pairing any usernames on file with the latest or most popular email or social networking providers.

Third, in cases in which the team were unable to find active contact information for families by the means described above, we searched for and attempted to contact parents through Facebook, Twitter, and LinkedIn. As per prior research (Masson et al., 2013), team members occasionally searched through publicly available lists of parents' contacts on these networks to ensure they had identified the correct family (e.g., searching if a potential parent was connected to his/her spouse as identified on the Time 1 contact forms).

The team debated creating clean study accounts vs. using personal accounts to message participants, but ultimately decided to use personal accounts for enhanced credibility (i.e., to come across like real people and not spammers; T. Rousse, personal communication, June 28, 2016). Accordingly, I suspected recruiters who were similar to the sample parents (i.e., fellow parents of school-age children) might be more successful recruiting in this manner (see Ribisl et al., 1996 for a similar discussion about recruiter-participant similarities in the context of more traditional location methods). Nonetheless, the data do not provide clear evidence for or against this supposition: The youngest undergraduate researcher on the team received a response from 0 out of the 3 families she contacted via social media, the faculty member (who had children) received a response from 1 out of 8 families she contacted, but I received a response and scheduled 1 out of 3 families she contacted.

Despite searching across multiple social networking sites, the team only successfully recruited (and scheduled) through Facebook, paralleling other recent non-prospective longitudinal studies that relied on social media for re-recruitment (Masson et al., 2013). These findings may reflect Facebook's popularity among middle-aged adults relative to other social networking sites at the time of data collection (Duggan, Ellison, Lampe, Lenhart, & Madden,

2015). Interestingly, in other work, participants reported preferring to be contacted through Facebook over traditional mail or telephone because they felt Facebook was private and could conveniently be accessed by both smartphone or computer (Masson et al., 2013).

Searching for participants via paid online databases. Like others (e.g., Cotter et al., 2002), after exhausting the aforementioned free options, I paid to access fee-based contact information databases. There are mixed opinions on the usefulness of these resources. Some prior studies have utilized such search engines early into their re-recruitment efforts and found them to be quite useful (e.g., Haggerty et al., 2008). Others, however, are more skeptical about their credibility (Cotter et al., 2002). Some paid search engines use marketing language that might give users unrealistic expectations about the quality of the results they can yield; these paid search engines may not provide much, if any, contact information that users could not find for free (Boswell, 2007). I did not begin using paid search engines until late into the recruitment process, after the team had already drawn heavily from available free resources. I located one participant via a $40 paid database (PhoneDetective.com), who ultimately requested not to participate in the Time 2 study, and two participants via a $30 paid database (Intelius.com), who did opt to participate. Altogether, this suggests that researchers with limited resources may be successful re-recruiting with free resources and should wait to resort to paid databases until they are confident that have exhausted the data available via free searches.

Additional location strategies. Articles also recommended additional search strategies, which I did not believe would be feasible or sensible for my team. However, other developmental researchers may find these approaches useful.

- Leveraging relationships with partner organizations (e.g., schools, after school

programs) to facilitate continued data collection (e.g., Agrawal et al., 1978; Lazorick et al., 2014). In some cases, organizations such as school districts may help re-locate original study participants (Tourangeau, Le, & Nord, 2005) or provide researchers useful contact information (Agrawal et al., 1978). These sorts of researcher-organization partnerships are common in prospective longitudinal studies with children (e.g., Tourangeau et al., 2005). Although policies differ from one school district to the next, the general consensus seems to be that for a school or similar organization to connect researchers with former participants, they (i.e., families) must have initially consented to future contact, as well as for the school district sharing their contact information (Ribisl et al., 1996). Nonetheless, even with these conditions in place, schools still would likely require submission to the school district research review board, which is a time-consuming process with no guarantee of additional participant recruitment information. Researchers also may need to budget to facilitate this form of collaboration. In some cases, it may be possible to obtain contact data from schools free-of-charge if a study benefits the district, for example, by providing data about a school-based program's effectiveness (P. Godard, personal communication, August 10, 2016). However, in other cases, school districts may charge researchers a fee regardless of educational relevance to cover the staff time invested in pulling the data (S. Dickson, personal communication, August 12, 2016). Relatedly, some school districts may be more helpful if they have received a grant related to the study.

- Visiting the neighborhoods participants lived in at the time of the original study (e.g.,

Haggerty et al., 2008; Ribisl et al., 1996). Because families with young children can develop attachments to their neighbors and first home, cumulative inertia (i.e., resistance to moving the longer a family stays in one location) sometimes sets in until additional life events impact the probability of moving (such as growing family size, change in jobs, desire for more space; Huff & Clark, 1978). As such, prior researchers have had modest success locating original study participants by visiting their former homes and neighborhoods, and reaching out to their friends and neighbors (e.g., Anderson, Huston, Schmitt, Linebarger, & Wright, 2001; Haggerty et al., 2008). Studies with a primarily middle-SES sample may note particularly high levels of cumulative inertia and residential stability, as prior research indicates that those with average income or education levels are less likely to move than persons at the extremes (Abu-Lughod & Foley, 1960). Accordingly, visiting the neighborhoods participants lived in at Time 1 could be fruitful for teams working with certain populations, but should only be attempted in cases where researchers gained appropriate consent as described above.

- Posting local advertisements. Barakat-Haddad and colleagues (2009) located a small number (less than 1%) of children in their longitudinal study through advertisements in local newspapers. In rural, small, or tight-knit communities, local and grassroots outreach could provide another avenue for locating former participants.

Concluding notes on location. I believe all the location approaches described above complement one another, and indeed, others have suggested multiple recruitment outreach methods often work in tandem (Ribisl et al., 1996). Several parents who previously ignored

telephone calls responded positively to calls after receiving a written letter describing the study.

One father emailed after receiving a re-recruitment letter and explained that it piqued his

daughter's interest, at which point I began addressing letters to both parents and children.

Likewise, all the parents who responded to recruitment attempts via text message had received

calls and emails from the research team first. Consequently, Table 6 might under-express the

value of certain recruitment channels (i.e., parents may have been primed to be receptive to calls

from researchers after receiving re-recruitment messages in other forms).

Experienced or committed research staff may be able to brainstorm additional ways to

search for contact information specific to local communities or more efficiently leverage the

latest search tools. As prior research suggests, more experienced researchers are often stronger

recruiters in general (Sugden & Moulson, 2015). I found undergraduate volunteers often needed

quite a bit of direction when searching to yield usable location data.

Throughout the location process, the research team maintained a detailed log of

communication attempts across platforms, somewhat aligned with recommendations from Cotter

and colleagues (2002) and Ribisl and team (1996). For efficiency's sake, we categorized phone

and email contact information according to whether we had (a) confirmed its connection to a

participant, (b) denied its connection to the participant (i.e., wrong or inactive phone number or

email), (c) not yet tested it for connection, or (d) tested it but not received a definitive

confirmation or denial. Though retaining this level of detail did not result in the most streamlined

database, this was necessary to avoid wasting time retesting communication avenues already

deemed unhelpful. In additional columns in the database, researchers recorded the date and time

of previous communication attempts, successful or not, keeping the most recent attempt on top

for ease of determining when someone had last been contacted. This allowed for the

diversification of timing of attempts.

Scheduling Families for Study Participation

After (or in my case during) the location process, researchers schedule participants for

study appointments. In this follow-up study, I completed appointments with 101 families (83%

of the located sample and 74% of the full original sample). The remaining located families either

refused to participate in the Time 2 study (6 families; 5% of located families), missed or

cancelled appointments and were unresponsive to attempts to reschedule (6 families; 5% of

located families), or failed to ever schedule appointments before the conclusion of data collection

(8 families; 7% of located families). Of the families I did test, it took on average 21.20 days (SD

$= 28.07$) from first locating them to completing research sessions with them. Again, I

encountered a great deal of variation in ease of scheduling, holding sessions with 40 families

(40% of participating families) within one week of location, 41 (41%) within one month, and 20

(20%) after one month had past. This data mirrors prior work; for example, Cotter and team

(2002) reported that about a third of their pediatric mental health clinic sample was easy to

schedule, a sixth required multiple contact attempts before scheduling, and 7-8% refused to

participate in some follow-up sessions. Table 6 provides a more detailed breakdown of how long

it took to schedule participants after locating them via each method described in the previous

section. Among the 122 located families, those who scheduled and attended Time 2

appointments were nonsignificantly less affluent than those located but not scheduled ($p = .09$).

However, no other fully or marginally significant differences emerged between these groups.

When reaching out to schedule appointments, I included language I thought might

motivate busy families to make time for this study (Striano, 2016). For credibility, I referenced

my university towards the beginning of most recruitment communications (Haggerty et al., 2008;

Silva, 1990; Sugden & Moulson, 2015). As a potential appeal, I also explained the overarching

study goal (Ribisl et al., 1996; Silva, 1990), which was to evaluate the long-term effectiveness of

educational computer games created with funding from the U.S. government. I assumed this

approach would speak to the parents in the sample, who had enrolled their children in an optional

educational computer game intervention when children were in preschool or kindergarten. Later

in the recruitment process, I began describing to parents roughly how many families had already

participated in the Time 2 study. I intended for this to both legitimize the study and to make

parents feel as if they were part of something large and important. I also mentioned that

participation would help me conclude my dissertation project and graduate, to associate

participating with an additional positive outcome. This personal connection may explain why I

had slightly more success recruiting via social media than the other team members.

Furthermore, parents and children were provided $20 for participating, and this was

highlighted in recruitment communications. This compensation seemed highly motivating to

children but less so to parents. Minimum wage was about $10 per hour in the city where I

collected my data, and the cost of living was higher than the national average. The relatively low

incentive for this area may explain why some of the affluent located families ultimately did not

schedule and follow through with appointments. Other longitudinal work providing larger

compensation yielded somewhat higher scheduling rates (e.g., Haggerty et al., 2008), although

over-compensating families, especially lower-income families like many in this sample could be

considered unethical or coercive (American Psychological Association, 2010; Hartmann, 1992).

Initially, I described to parents what I believed were salient aspects of the original study, attempting to trigger positive memories or loyalty to the intervention (see Cotter et al., 2002; Ribisl et al., 1996 for a discussion on the value of study loyalty and affiliation). However, it became clear over the course of re-recruitment that many parents struggled to recall the Time 1 intervention: Thirteen parents wrote in questionnaires that they did not recall the intervention or wrote comments clearly confusing this study with ones conducted by other groups on other topics, and additional parents verbally mentioned having forgotten the original study. Similar difficulties also have been suggested in other longitudinal studies (Barakat-Haddad et al., 2009). This may explain in part why the faculty member, who was the only Time 2 study team member involved in the original data collection, was not more successful recruiting via social media – weak personal connection. To further jog memory and enhance credibility, Ibegan taking care to name the preschool or kindergarten where families were initially recruited.

Across all recruitment messaging, some parents seemed to extrapolate a sense that participating in the study would somehow provide children an academic enrichment opportunity, aid them as parents in better facilitating children's continued education, or otherwise abstractly improve children's education. Similarly, some parents said that they ultimately decided to participate in the study hoping it would help inspire their children to attend the university where the research team was based. Since different appeals may speak to different families, it may be best to position a study as accruing a variety of tangible benefits (Striano, 2016; Sugden & Moulson, 2015).

Adhering to recommendations in prior methodological pieces, participants were given a great deal of flexibility in terms of scheduling (Cotter et al., 2002; Ribisl et al., 1996). For

convenience and comfort, families were given the choice of participating in their homes (an option chosen by 45 families, 45% of participants), university lab space in a suburb just outside of the urban city families lived in at Time 1 (22 families, 22% of participants), or local libraries with private study rooms (23 families, 23% of participants). It is likely that this flexibility increased participation rates, as logistical barriers such as inflexible work schedules, lack of transportation, or need for child care for younger children are known to impact the ability to retain low-income families in research (Duch, 2005). Additionally, 11 families (11%) who had moved out of the metropolitan area completed online surveys instead of participating in person (another strategy also recommended in other recent methodological reviews; e.g., Kalkhoff, Youngreen, Nath, & Lovaglia, 2014). This helped to address physical barriers found in other studies, such as where some participants were unable to be interviewed by phone due to physical constraints (e.g., SEAL training in Alaska, being in prison, and having re-located to Mexico; Anderson et al., 2001). Most appointment time requests were honred, excluding cases, for example, when families wanted to meet late at night or at local libraries during hours they were not open. As alluded to above, I tried to schedule appointments as soon after reaching participants as possible, ideally within one week of initially locating families (Kalkhoff et al., 2014; Ribisl et al., 1996). If I did not have much availability over the course of the following week, I held off attempting to contact parents until my schedule was more open rather than seeking appointments weeks in advance. This practice sought to minimize the window in which families could forget about appointments or experience other schedule changes.

Online and digital tools can further help with appointment management. I manually entered study appointments into a digital calendar shared by the research team, emailed parents

event appointments that could be added to any personal digital calendars they maintained, and

programmed these event appointments to send parents reminders the night before or morning of

their scheduled study sessions. Others have reported using digital interfaces such as

YouCanBook.Me to allow participants to privately and relatively independently sign-up for and,

if need be, reschedule study appointments, choosing among several session times researchers

make available (Kalkhoff et al., 2014). These interfaces also can send automated appointment

reminders in advance of sessions (Kalkhoff et al., 2014).

Confirming Appointments

As with all developmental research, the last step in the re-recruitment process was to

confirm appointments with parents, making it clear I would be happy to reschedule if need be

(Striano, 2016). Initially, I confirmed appointments exclusively by phone call and email, aligned

with methodological recommendations elsewhere (Kalkhoff et al., 2014), but later into

recruitment, began calling, emailing, or texting parents, depending on their preferences.

Even though only a small number of parents responded to early locating/scheduling

efforts by text message (see Table 1), this communication channel worked well for confirmation,

perhaps because families perceived text messaging to be a less formal means of communicating.

Four dyads (11% of in-person appointments scheduled at the time time) missed their

appointments before I started texting parents, but this only occurred once (2% of in-person

appointments) after I introduced texting (and this one appointment was missed due to an

unfortunate family emergency). Parents who needed to cancel or reschedule appointments

seemed more comfortable doing so via text rather than over the phone or through email.

Moreover, if researchers suspected a participant might forget about the appointment after official

confirmation, they would send casual "on my way" text messages to prompt parents' memories. I used my personal phone for such purposes, although it might have been wiser to buy a study-specific cell phone or set up a Google Voice account that researchers could share without compromising privacy (M. Smith, personal communication, April 27, 2016). Some of the digital study appointment management tools described in the previous section can automatically send confirmation emails or text messages, and may allow parents to reschedule appointments without needing to interface with a researcher (Kalkhoff et al., 2014).

Conclusion

Across developmental sub-fields – from basic cognition (e.g., Lauer & Lourenco, 2016) and language development (e.g., Can et al., 2013) to mental (e.g., Agrawal et al., 1978) and physical health (e.g., Fein et al., 2014) to applied interventions targeting children (e.g., Campbell et al., 2012) and their families (e.g., Huston et al., 2005) – researchers are engaging in longitudinal work, re-recruiting families who participated in one study to gain a better understanding of children's developmental trajectories. Frequently, research teams do not decide to begin embarking on this work until after an original study has concluded. My work demonstrates the feasibility of re-recruiting sizeable numbers of urban families after an extended gap in communication with limited financial resources. Future developmental research teams should be able to achieve high follow-up rates by (a) setting up initial studies in which parents provide detailed contact information, including contact information for multiple caregivers, and consent for later waves of research, (b) search across a variety of sources to locate participants, e.g., social media, people-centric search engines, etc., (c) write multifaceted scheduling scripts highlighting the study's value, and (d) confirm appointments in a way that conveys a casual tone

that makes parents feel comfortable, even if they need to reschedule. Many of these strategies may likewise enhance the recruitment process even for cross-sectional research. Developmental researchers also may wish to consult Table 1 to help inform timelines as they plan.

Relative to prior research, I had more success re-recruiting with free tools such as people-centric search engines, and less success using paid and other tools. Differences between my study and others may be attributable in part to advances in modern technology and my team's efforts to leverage popular technological services. Because parents are increasingly maintaining the same cell phone numbers, especially among high- and middle-income populations, even after they move (Dost & McGeeney, 2016) and because of the existence of a plethora of free people-centric search engines, I had more success using originally provided telephone numbers and free search engines than researchers reported previously. I similarly found text messaging, which is currently very popular in the U.S. (Duggan, 2013), to be helpful in ensuring participants maintained their appointments or felt comfortable rescheduling if necessary. Indeed, even researchers conducting cross-sectional work may wish to consider incorporating text messages into their appointment confirmation protocols. I only saw limited success locating participants via social media or paid databases, somewhat contrasting prior research where paid databases were more effective (e.g., Haggerty et al., 2008). However, I suspect my team's location findings would have varied had we engaged the various search methods in a different order.

Technology-related advances aside, this study also reinforces the value of strategies others have suggested to schedule participants and calls into question assumptions about the ease of scheduling particular groups of participants. My experiences underscore the importance of collecting detailed contact information during an initial study (Ribisl et al., 1996) and using

recruitment (or re-recruitment) scripts that make the university affiliation and study goals clear (Sugden & Moulson, 2015). Such re-recruitment scripts may be especially valuable in cases where researchers are focused on academic, prosocial, or other potentially positive outcomes. Moreover, emphasizing the university and study aims may even be more worthwhile than reminding families of the particulars of the original study. Like others (see Ribisl et al., 1996), I also believe my re-recruitment success is in part attributable to the fact that I planned Time 2 study activities in a way that allowed me to conduct research in a variety of settings (i.e., lab, library, home). However, I recognize such flexibility may not be feasible for all development sub-domains, such as when researchers are interested in collecting neurological data (e.g., Schwartz et al., 2003).

Like some studies, I had somewhat more success locating families with higher-incomes, as well as Caucasian families (e.g., Fein et al., 2014). At least for this study, I expect these two seemingly different findings may both trace back to stability issues related to family income, rather than anything cultural. Caucasian families in this study were more affluent than the rest of the sample ($p < .001$), a trend that bears out nationally in the U.S. (Wilson, 2015). In general and across racial-ethnic groups, low-income families are more likely to experience disruptions in phone service, changes to cell phone numbers (Ahlers-Schmidt et al., 2012), and physical relocation over time (Abu-Lughod & Foley, 1960), factors which in turn might make them particularly challenging to locate. However, mirroring findings that contradict my location results (e.g., Silva, 1990), I was arguably *less* successful *scheduling* higher-income participants, who may have been unmotivated to find time to participate. To increase scheduling rates, those with greater resources may wish to provide larger monetary compensation or even more fervent

appeals to the study's mission to help attract more affluent potential participants.

As is typical in longitudinal research (Ribisl et al., 1996), I attempted to re-recruit all of the original families and used all of the allotted appointment time during each research session to gain information about children's present, study-specific functioning. Consequently, I did not formally survey parents about their perceptions of the re-recruitment experience or experimentally compare the effectiveness of different recruitment strategies (which might have resulted in losing participants assigned to less successful recruitment strategy conditions). Moreover, I refined my recruitment approach over the course of data collection, as is common in research of this nature (see Ribisl et al., 1996). Future researchers with larger initial samples or more time and financial resources should consider formally testing some of the assumptions in this article; such work would address important gaps in the methodological literature.

Despite the aforementioned limitations, I hope my efforts can guide other developmental scientists interested in conducting longitudinal research. Moreover, some of these strategies may even positively impact recruitment for cross-sectional studies. Given the power longitudinal studies have to clarify developmental trajectories (Nicholson et al., 2002) and provide compelling accountability evidence for interventions (e.g., Barnett, 2013), and given the growing interest in work of this nature (e.g., Wartella et al., 2016) it is important developmental scientists feel capable of re-recruiting sufficiently large samples, even with limited resources and even when they decide to begin such work long after the conclusion of a particular study.

Article 3: Did We Succeed in "Raising Readers"? Effects of Ready To Learn Early Childhood Literacy Computer Games in Middle Childhood

Abstract

The majority of young children in the U.S. have consumed media created with funding from the Department of Education's Ready To Learn (RTL) initiative. RTL aims to promote foundational early learning skills (e.g., basic literacy) so that children are prepared to succeed in formal schooling. The Early Learning Hypothesis predicts early exposure to such educational media can catalyze long-term academic success. This study assessed whether the effects of RTL media in early childhood sustain into middle childhood. One-hundred-and-one youth who had participated in an evaluation of an RTL literacy-themed computer game in early childhood were re-contacted in middle childhood. Their present-day literacy skills were assessed, and parents provided complementary data. A curvilinear relationship between children's early childhood pretest scores and middle childhood outcomes was detected. The positive effects of the games sustained into middle childhood, but only for children with below and above average literacy skill prior to the original intervention. These findings are interpreted in light of the Traveling Lens and Capacity Models, two theoretical frameworks outlining the relationship between children's skillsets and their receptivity to educational media.

Keywords: Ready To Learn, educational media, longitudinal, early intervention

The U.S. Department of Education's (DoEd) Ready To Learn (RTL) initiative provides $25 million annually for the development of public mass media intended to promote school readiness (Bryant et al., 2001; DoEd, 2015; Singer & Singer, 1998). RTL has provided an added stream of funding for some longstanding educational media properties that premiered prior to the launch of the grant, such as *Sesame Street*, along with seed money for new properties that did not exist before RTL, including *Between the Lions, Super WHY!*, and *WordWorld* (Corporation for Public Broadcasting & Public Broadcasting Service [hereafter CPB and PBS, respectively], 2011). Today, the majority of young children in the U.S. have been exposed to media paid for in part by RTL (CPB & PBS, 2011). Research suggests this RTL-funded media successfully has promoted basic early academic skills (e.g., alphabet knowledge) in the short-term (see Article 1), but whether this expenditure has resulted in more children truly being ready to succeed in school is an open question. Theoreticians believe early educational media exposure can catalyze long-term learning, an idea known as the Early Learning Hypothesis (Anderson et al., 2001). To date, this hypothesis largely has been substantiated by correlational or indirect measures of the long-term effects of educational media (e.g., Kearney & Levine, 2015; Rosser et al., 2007).

In this study, children who participated in an evaluation of RTL-funded literacy-themed computer games in early childhood were re-contacted to determine if the positive effects of exposure to these games sustained into middle childhood. These findings provide accountability evidence for the RTL initiative, validate the Early Learning Hypothesis with causal data, and speak to other scholarly debates on children's learning from media.

Ready To Learn Initiative

RTL launched in 1994, a time when the American public was concerned about the dearth of high quality children's television programming and reports of large numbers of young children entering elementary school ill-prepared for formal schooling (Bryant et al., 2001). The intention of the initiative was to fund mass media that would help America's school children, especially at-risk populations, gain foundational early learning skills necessary for success in elementary school and beyond (Singer & Singer, 1998). At around the same time, commercial media providers in the U.S. concurrently began providing an increasing volume of educational fare (Bryant et al., 2001), and numerous international content producers have created similar media (e.g., Mares & Pan, 2013).

Unlike many of these other content providers, RTL grant recipients uniquely have made most of their content publicly available for free or at a low cost (CPB & PBS, 2011), have promoted positive parenting practices and to a lesser extent the RTL media via social marketing campaigns (with taglines like "Raising Readers" or "Anytime is learning time"; Hurtado, Galdo, Agin, & Heil, 2010), and have created outreach programming (e.g., media-themed summer camps) intended to complement and extend the lessons present in the media (Llorente et al., 2010). Together, these factors have helped to make RTL media wide-reaching (CPB & PBS, 2011). RTL media producers also have been at the forefront of experimenting with newer forms of media; starting in 2005, the grant explicitly tasked content producers with creating new media tools, such as educational computer games (Michael Cohen Group, 2012). The present study aims in part to evaluate the long-term success of one such tool: an educational computer gaming suite that promoted foundational reading skills (Schmitt, Sheridan Duel, & Linebarger, 2017).

Educational Media as a Catalyst for Long-Term Learning and Growth

With their Early Learning Hypothesis, Anderson and colleagues (2001) propose that early exposure to educational television, such as the programs funded by RTL, can spark growth and learning that persists throughout children's time in school. They suggest children who acquire key school readiness skills from educational television in early childhood may be better prepared for elementary school, initially placed in higher ability groups in school, and thus set on a trajectory of continued success. The present study explores whether exposure to interactive media such as RTL computer games also leads to similar long-term growth.

Some empirical evidence substantiates the Early Learning Hypothesis. A host of studies following up with educational media evaluations provide causal evidence that positive effects from educational media are detectable several months after initial exposure. For instance, in an evaluation of an RTL-funded vocabulary intervention, program participants retained vocabulary gains six months after the intervention had concluded (Neuman et al., 2011). Likewise, in another RTL intervention, young children who watched episodes of literacy-themed *Super WHY!* towards the beginning of the school year outperformed their peers at the end of the year on a measure of letter sound knowledge, and children who watched the show and played complementary online games outperformed peers on measures of lower case letter knowledge and rhyme awareness (Linebarger, 2010). However, contrary to the Early Learning Hypothesis, children in the study's control group who did not consume any media had the strongest long-term performance on a measure of beginning word sound awareness (i.e., understanding of the sounds at the beginning of words; Linebarger, 2010).

Similar studies looking at other media also found results consistent with the Early Learning Hypothesis. For example, the positive effects of a noncommercial, literacy-themed

computer game that promoted letter-sound knowledge sustained several months after an initial effectiveness evaluation (Segers & Verhoeven, 2005). Likewise, kindergarten students who played computer games that promoted blending (the ability to combine letter sounds into words) demonstrated stronger reading skills mid-way through first grade (Reitsma & Wesseling, 1998). And kindergartners who played a mix of literacy-themed computer games and apps evinced stronger literacy on composite measures assessing competence across several early literacy skills at multiple time points across first grade (Ponciano & Thai, 2016).

A small number of scholars have examined even longer-term impacts from educational media (mostly television) prior to the launch of the RTL initiative using correlational or indirect methods. The evidence these studies provide is mixed. To illustrate, in a correlational study, preschoolers who watched child-targeted television as in the early 1980's (including educational programs such as *Sesame Street*, which at the time had not yet received RTL funding) read more books and achieved better grades in high school English, math, and science (Anderson et al., 2001). In contrast, in another quasi-experimental study, there were no differences in high school reading, vocabulary, or math performance between preschoolers who lived in and outside communities with access to *Sesame Street* in the late 1960's (again many years before Sesame Street received RTL funding; Kearney & Levine, 2015). However, in the same study, those living in communities with *Sesame Street* demonstrated stronger labor force outcomes despite the lack of evidence of differential school performance (Kearney & Levine, 2015). Differences between these two studies may be driven in part by changes in *Sesame Street*'s curriculum over the past half-decade (Fisch & Truglio, 2000) or as a result of methodological differences.

Differential Susceptibility to Media Effects

Per other media theories, it may be overly simplistic to assume all children would realize comparable long-term benefits from early educational media exposure, and indeed, a focus on individual differences between children may help explain why different studies yield seemingly conflicting findings (Piotrowski & Valkenburg, 2015; Valkenburg & Peter, 2013). Turning to theories originally conceived to explain short-term learning from media may shed insight into which children are most likely to benefit from such exposure in the long-term.

The Traveling Lens Model posits children gain the most from media when they perceive its content as being moderately difficult – not too easy or too challenging (Huston & Wright, 1989; Rice, Huston, & Wright, 1982). Specifically, the model predicts children's interest in and attention to media will be strongest for media content they find moderately difficult, and this interest and attention will facilitate greater learning. The model goes on to suggest that as children's thinking becomes more advanced or their familiarity with the mediated content increases, their interest, attention, and learning from a given media stimulus may decline. In such situations, children might begin attending to more challenging aspects of favored media products or seek out more challenging media content entirely. In today's new media environment, replete with active and interactive games, the media itself might "level up" to continuously present children with content aligned to their evolving skillsets (Guernsey, 2012; Hirsh-Pasek et al., 2015; Walker, 2011). Extending this line of thinking, perhaps educational media may have more pronounced long-term effects for children who found content moderately difficult at initial exposure and in cases where media present children increasingly challenging content over time.

The evidence is mixed as to whether, in the short-term, RTL-funded media were most effective for children with below or above average abilities. That is, it is unclear if RTL media

content aligned best with the abilities of weaker or stronger early readers. For instance, in an evaluation of the RTL television show *WordWorld*, children with higher initial literacy skills prior to *WordWorld* exposure scored higher on posttest measures of word recognition, while children with lower initial literacy demonstrated greater gains in phonemic awareness (i.e., mastery of language sounds; Michael Cohen Group, 2009). In other studies focused on literacy summer camps using curricula and media from *Super WHY!* and on *Between the Lions* television episodes, children with average and above average literacy at pretest benefited most from media exposure (Jennings, 2013; Linebarger, Kosanic, Greenwood, & Doku, 2004). Conversely, in another evaluation, children under the age of 4.5 only benefited from RTL-funded *Pocoyo* apps if their English language skills were poor pre-exposure (Michael Cohen Group, 2013). This substantial variation makes it difficult to say whether the body of RTL short-term evaluations substantiate the Traveling Lens Model. It may be that media's alignment to children's skillsets varies across different RTL media products and/or across different samples of children.

The Capacity Model proposes somewhat different mechanisms undergirding learning from media (Fisch, 2000, 2004, 2016). This model is most well-known for its nuanced discussion of the characteristics of media that lead to learning, but, especially pertinent to the current discussion, it also outlines individual differences that may make certain children especially receptive to educational media (Aladé & Nathanson, 2016). The model posits children with stronger media-related subject matter knowledge, interest, and, for some metrics, verbal ability should gain more from educational media because they should be able to assimilate the narrative and/or educational messages presented in the media more efficiently, easily, and readily (Fisch, 2004). That is, the model implies a linear relationship between subject matter knowledge,

interest, and verbal ability, and learning (however, Fisch never uses the word "linear" to describe these relationships, and in a footnote focused on the characteristics of media that facilitate learning, acknowledges the possibility of nonlinear relationships in certain cases; Fisch, 2004).

In a validation study, Aladé and Nathanson (2016) examined these assumptions by asking preschool-age children to view a science-themed television episode and by administering a battery of cognitive and interest assessments pre- and post-exposure. Consistent with the Capacity Model, they found children with stronger verbal skills and higher levels of science subject matter knowledge learned more from the episode. Although Aladé (2013) noted children's interest in science was correlated with their verbal ability, children's scientific interest did not predict learning of science content like verbal ability did (Aladé & Nathanson, 2016). That said, Aladé and Nathanson (2016) questioned whether their pattern of results would have looked the same had they operationalized "interest" differently, perhaps asking children how motivated they were to attend to and learn from the science program.

Despite these generally supportive data from Aladé and Nathanson (2016), the RTL literature reviewed above calls into question the assumption that prior skill, verbal ability, and subject matter interest linearly predict learning (e.g., Michael Cohen Group, 2009, 2013). Also, similar to the Traveling Lens Model, to my knowledge, the Capacity Model's tenets about child characteristics have only been examined in the short-term, and it therefore is difficult to say if and how these individual differences would manifest in the long-term.

The Current Study

This investigation followed up with children who participated in an RTL literacy evaluation in early childhood, re-assessing their abilities in middle childhood, six years after the

original intervention. The primary aim was to better understand whether the RTL initiative has resulted in the creation of media capable of stimulating meaningful and long lasting impacts on children's school performance. The study intended to provide data extending scholarly understanding of the Early Learning Hypothesis, testing this hypothesis with causal data focused on the long-term effects of educational computer games intended to support literacy learning and exploring whether learning varies as a function of children's skillsets prior to media exposure.

Original early childhood RTL intervention. In 2010, Schmitt and colleagues (2017) evaluated an RTL literacy computer game with a diverse sample of preschool and kindergarten children living in a city in the Midwestern U.S. Researchers randomly assigned about one third of the sample to play a control suite of computer games created by a commercial company, which featured non-educational puzzles and arts activities. They assigned the remaining two-thirds of the sample to play a leveled RTL literacy computer game suite (PBS KIDS Island) featuring the media properties *Between the Lions, Martha Speaks, Sesame Street, Super WHY!,* and *WordWorld.*

The RTL literacy gaming suite presented eight sequenced levels that each contained four mini-games focused on different literacy skills. Once children successfully completed all four mini-games in one level, they advanced to a new, more challenging level. Early levels promoted relatively basic literacy skills (see Grant et al., 2012 for more on literacy learning sequences). For example, in levels one and two, children played mini-games in which they put together jigsaw puzzles featuring various letters of the alphabet as a means of enhancing letter recognition. Later levels promoted more advanced skills. For instance, mini-games in levels four - six presented children with challenges they solved by correctly choosing words that rhymed

with a target. All children completed the first three levels before the conclusion of the study, and 65% advanced all the way to the final level (i.e., level 8).

Relative to the control group, children who played the RTL literacy games for 6-8 weeks scored higher on three assessments measuring mastery of English language sounds (i.e., beginning sound awareness, letter-sound knowledge, letter-sound fluency, and phonological awareness), as well as on measures of letter sequencing and vocabulary. Parents also indicated the games promoted children's letter recognition, word recognition, spelling, and self-esteem.

Middle childhood follow-up. In the present evaluation, families who participated in the above early childhood RTL intervention were re-contacted. The aim was to assess whether the positive effects of early exposure to the RTL literacy games sustained into middle childhood, six years after the intervention. More specifically, this study asked (RQ1a) whether children who played the RTL literacy games still evinced stronger literacy skills in middle childhood, and (RQ1b) whether these effects varied as a function of children's literacy skill prior to the intervention.

Additionally, the study asked (RQ2a) whether long-term effects generalized to other learning domains beyond literacy. Even though RTL media often focus on specific early learning domains such as literacy (see Article 1), the intention of the initiative is to broadly prepare children for success in school (Singer & Singer, 1998). Additionally, other research suggests literacy undergirds success in a variety of school subjects, because so many school courses require children to read texts and process written and verbal directions (Schiefele, Schaffner, Möller, & Wigfield, 2012). Moreover, prior research also indicates that children might be able to apply or transfer the skills they learn in media to other contexts in cases in which they encode

lessons in a manner abstracted from the original media (e.g., consuming literacy-themed media focused on the /L/ sound and later sounding out the word "log" during story time in school or "lung" during a science lesson on the body; for a longer discussion, see Fisch, 2004; Fisch, Kirkorian, & Anderson, 2005). The study further questioned (RQ2ab) whether these long-term transfer effects also varied as a function of children's literacy prior to intervention.

Method

Sample

Families in the original study were recruited from 15 schools in a major city in the Midwestern U.S. in Winter and Spring 2010. One-hundred-thirty-six children (48% male) completed the full intervention. Of these, 93 were assigned to the RTL literacy group, and 43 were assigned to the control group. Across the full sample, 94 children were in preschool during the original study, and 42 were in kindergarten. The sample was racially and ethnically diverse (28% Caucasian, 28% Hispanic, 20% African American, 7% Asian, and 17% other/mixed background). An effort was made to recruit families across the socio-economic spectrum, with an average sample income-to-needs ratio[1] of 1.96 (SD = 1.42).

---

[1] Income-to-needs ratios were calculated via a three step process. First, researchers asked parents in early childhood to report their household income using a 9-point scale ranging from under $20,000 (1) to over $125,000 (9). Second, researchers identified the midpoint for each of these nine income ranges, and divided them by a correction factor to adjust for the cost of living in the city where the research originally took place, which was above the U.S. national average (American Chamber of Commerce Research, 2007). Finally, for each family, researchers divided these adjusted estimates of income by the federal poverty threshold for a family of that size, based on poverty data provided by the U.S. Census (Institute of Research on Poverty, 2010).

After receiving approval for the follow-up study from my university Institutional Review Board (IRB), families from the original intervention were re-contacted in Spring and Summer 2016. Of the 136 original participating families, 101 participated in the follow-up study (74% of original participants). This follow-up rate is on par with rates achieved in comparable longitudinal intervention evaluations (e.g., Huston et al., 2005). Of the 35 families who did not participate in the follow-up, six refused to participate, 15 were located but did not attend a study session before the conclusion of data collection, and 14 were unlocatable. Relative to the families who did participate in the follow-up study, children in the missing group were more likely to be of other/mixed race ($p < .05$). Besides race-ethnicity, there were no other significant differences between the families who participated in this follow-up and those who did not.

Table 7 provides full demographic information for the sample who participated in the middle childhood follow-up. Among the full follow-up sample, those who were in the RTL literacy group during the original intervention were slightly more affluent than their control group counterparts (a difference that was nonsignificant in the original evaluation). But otherwise, the makeup of families across both conditions was comparable.

Table 7. Demographics for Children Who Participated in the Middle Childhood Follow-up

	All ($N = 101$)	Treatment ($n = 65$)	Control ($n = 36$)
Age in Years in Middle Childhood	11.28 (1.30)	11.35 (0.77)	11 .14 (1.93)
Grade in Middle Childhood			
4th or below	17%	17%	18%
5th	52%	46%	62%
6th	27%	32%	18%
7th or above	4%	5%	3%

Female	55%	54%	56%
Child Race-Ethnicity			
Caucasian	31%	37%	19%
African American/Black	21%	15%	31%
Hispanic	29%	28%	31%
Other or Mixed	20%	20%	19%
Out-of-Area Residence in Middle Childhood	11%	12%	8%
Maternal Level of Education at Middle Childhood Follow-up			
Less than a Bachelor's	47%	42%	56%
Bachelor's or Higher	53%	58%	44%
Income-to-Needs Ratio in Early Childhood[†]	1.94 (1.37)	2.12 (1.39)	1.60 (1.29)
Income-to-Needs Ratio in Middle Childhood[*]	2.30 (1.44)	2.55 (1.41)	1.80 (1.40)
Receipt of Public Assistance	28%	24%	37%
Pretest Composite Scores in Early Childhood	-0.03 (1.00)	0.01 (1.04)	-0.11 (0.95)

Note. Means with standard deviations in parentheses for continuous variables and frequencies for categorical variables. Listwise deletion was used for item-level missing data. [†] $p < .1$, * $p < .05$.

Procedure

The protocol for the middle childhood data collection varied depending on families' residence at follow-up. Families still living in the area where the original study was conducted ($n = 90$) were given the option of completing an in-person research session in their home, at my university lab facility, or at a neighborhood library. During these sessions, parents completed a 25-minute questionnaire, while children completed a 40-60-minute researcher-led battery of assessments. Parents of families who had moved ($n = 11$) were emailed an online version of the

parent questionnaire. These parents also were given the option to complete the questionnaire verbally over the phone, although no parents opted to do so. As part of both in-person and online data collection, parents consented to allow researchers to access the data they and their children provided in early childhood and link it to the data provided at follow-up. All children who participated in the study received $20 cash. Parents who completed the questionnaire during an in-person session also received $20 cash, and parents who completed the online version of the survey received $20 Amazon e-gift cards.

Measures

      Early childhood direct measures of child literacy. In the original early childhood intervention, children completed a series of standardized and custom literacy assessments at pretest and posttest. At both time points, children completed the Phonological Awareness Literacy Screening for Preschool (PALS-PreK; Invernizzi et al., 2004) Alphabet Knowledge, Letter Sounds, Beginning Sound Awareness, and Rhyme Awareness subtests, as well as the Get Ready to Read! Screener, a composite measure of early literacy covering a variety of literacy sub-skills (Whitehurst & Lonigan, 2001). Children also completed researcher-developed measures of letter sequencing, phonemic awareness, and vocabulary. See Schmitt and team (2017) for a fuller description of these measures. Using the psych package in R (Revelle, 2016), an item cluster analysis suggested all of these measures clustered together at pretest (Cluster fit = .94, Pattern fit = .99, RMSR = .08, Cronbach's α = .86). For the current study, a pretest composite score for each child was created using regression weights based on the cluster analysis. The average early childhood pretest composite score for the full original sample was 0 (*SD* = .98, Range: -2.8 to 1.12), and for the subset of children who participated in the middle

childhood follow-up was -.03 (*SD* = 1.00, Range: -2.8 to 1.12).

Middle childhood direct measures of child literacy. Schmitt and colleagues' (2017)

original data provided fairly convincing evidence that the RTL literacy games promoted

children's phonological awareness (i.e., the ability to identify and manipulate language sounds),

with statistically significant findings at the .05 level for three measures of phonological

awareness, and an average effect size of Cohen's *d* = .2 across all measures of phonological

awareness. The present study focused on children's present day rhyming (a skill falling under the

broader umbrella of phonological awareness), deletion (the ability to remove sounds from the

beginning, end, or middle of words; another phonological awareness subskill), segmentation (the

ability to break up words into smaller linguistic units such as syllables or letter sounds; a third

phonological awareness subskill), decoding (applying letter-sound knowledge to decipher words

and pseudowords), spelling, and reading. Previous research suggests all the aforementioned

literacy skills are predicted by early phonological awareness (e.g., Juel, 1988; Scarborough, Ehri,

Olson, & Fowler, 1998; Wagner et al., 1997). To assess these skills, several sub-tests from the

Kaufman Test of Educational Achievement, II (KTEA-II; Kaufman & Kaufman, 2004a) were

administered. The KTEA-II was normed with 200 school-age children, and split-half reliability

coefficients for most subtests were .90 or above (Bonner & Carpenter, 2005).

Rhyming. Children's rhyming was assessed with the Rhyming section of the KTEA-II

Phonological Awareness sub-test. For this measure, children were asked to aurally discriminate

between rhyming and non-rhyming words. A researcher read children six sets of words with four

words each. Within each four-word set, three words rhymed, and one word did not. Children

earned one point for correctly identifying the non-rhyming word in each set, for a maximum

score of 6 points (*M* = 5.78, *SD* = .67, Range: 2 to 6).

  Deletion. Children's deletion was assessed with the Deletion section of the KTEA-II

Phonological Awareness sub-test, which required children to delete sounds from target words.

For example, a researcher asked children to "Say *make*, but without the /k/ sound", and children

earned a point if they correctly deleted the /k/ sound to verbally form the word "may". Children

completed six of these deletion items, again for a maximum potential score of 6 points (*M* =

5.19, *SD* = 1.13, Range: 1 to 6).

  Segmentation. Children's segmentation was assessed with the Segmentation section of

the KTEA-II Phonological sub-test, which required children to separate words into smaller

linguistic units. In the first part of this task, children broke four words into distinct syllables (e.g.,

spa…ghe...ti), earning one point for each word correctly segmented. Two of these four target

words contained two syllables, and the other two contained three. In the second part of the task,

children had to segment five words into distinct sounds or phonemes (e.g., say "crust" as

"/k/…/r/…/uh/…/s/…/t/"), again earning one point for each correctly segmented word. Thus,

altogether children could earn 9 points for Segmentation (*M* = 4.88, *SD* = 1.92, Range: 2 to 9).

  Decoding. Children's decoding was assessed via the KTEA-II Nonsense Word Decoding

sub-test. In this test, children read aloud an increasingly challenging set of pseudowords (e.g.,

"kimp", "sprewful"). Children could earn up to 50 points on this sub-test for correctly decoding

all pseuodowords; however, administration of the sub-test terminated once children incorrectly

decoded four words in a row (*M* = 26.01, *SD* = 9.04, Range: 0 to 42).

  Spelling. Children's spelling was assessed via the KTEA-II Spelling sub-test. In this test,

a researcher read a target word, contextualized it by reading it in a sentence, and read it one more

time. Children then had to spell (i.e., write out) each word, earning one point for each correct

spelling. Administration of the sub-test terminated after children spelled four items incorrectly in

a row. Children could earn up to 41 points on this test ($M = 25.51$, $SD = 7.73$, Range: 2 to 38).

Reading. Children's word-level reading was assessed with the KTEA-II Word

Recognition sub-test. Similar to the Nonsense Word Decoding sub-test, this battery required

children to read aloud a series of increasingly difficult words (real words this time), awarded one

point for each word children read correctly, and terminated after children read four in a row

incorrectly. Children could receive scores of up to 75 ($M = 40.78$, $SD = 10.04$, Range: 6 to 66).

Parent measures. Parents rated children's school performance across seven subject

areas: reading, writing, science, social studies, mathematics, speaking/listening, and following

directions. They rated how well children performed in each area relative to other children that

age, using a five-point scale ranging from Poor (1) to Excellent (5). There was excellent internal

consistency across parents' ratings in all seven subject areas (Cronbach's α = .92). A composite

measure of parents' ratings across all subject areas (averaging across their seven ratings[2]; $M =$

3.82, $SD = 0.90$, Range: 1.14 to 5.00) was examined, along with each indicator independently to

investigate long-term effects on children's literacy and downstream transfer effects on other

learning domains beyond literacy.

---

[2] a composite score averaging across all parent ratings (i.e., a mean) was used so that this score

would be on the same scale as each individual item, thereby facilitating ease of interpretation.

However, results were similar in alternate models that used a composite based on the same

procedure used for the early childhood pretest composite.

Parents also provided updated demographic information, including information on family size and income used to calculate family income-to-needs ratios[3] at follow-up.

Results

Data were analyzed in R-3.31 (R Core Team, 2016) using multiple regression to test whether assignment to the RTL literacy group predicted either children's middle childhood literacy outcomes or children's parental ratings of academic performance, and controlling for child sex, age at follow-up, income-to-needs ratio, change in income-to-needs ratio, residence, pretest score, and condition assignment (RTL literacy v. control). Additionally, to test RQ1b and RQ2b (whether effects varied based on children's skillsets prior to the intervention), models included a Condition x Pretest interaction term. Moreover, because it was unclear whether the media would be more effective for children with below average, average, or above average pretest scores in early childhood, models also included a squared pretest term and a Condition x

---

[3] Similar procedures to Schmitt and team (2017) were used to calculate updated income-to-needs ratios in the present study. First, parents indicated their family income using a 14-point scale ranging from under $5,000 (1) to over $200,000 (14). The midpoint for each range on this scale was then divided by a correction factor to account for the cost of living in the area where data collection activities occurred at the time of the follow-up (Council for Community and Economic Research, 2010). For each family, this adjusted income was divided by the latest estimates of the poverty threshold for a family of that size (United States Census Bureau, 2016). To avoid multicollinearity, models included families' income-to-needs ratios during the original testing, as well as a change in income-to-needs term, which subtracted families' income-to-needs ratios in early childhood from their income-to-needs ratios at follow-up. This term controlled for any relationship between increases or decreases in family wealth over the six years between data collection periods, and children's middle childhood outcomes.

Pretest^2 interaction term in the models. Cubic pretest and Condition x Pretest^3 interaction terms were initially added, but ultimately omitted these from the final models because neither were significant in any model (Stock & Watson, 2006).

In conducting these analyses, item-level missing data were imputed using predictive mean matching implemented via multiple imputation by chained equations with the mice package in R (van Buuren & Groothuis-Oudshoorn, 2011). One or more items included in the models were missing for 23% of the sample: Family income information was missing for 13 children, and one or more middle childhood outcome measures were missing for 12 children (with two children missing both family income and one or more outcomes). Gender, age, pretest, and intervention condition assignment information were available for all children. To impute the missing items, a subset of the data that included all variables in the statistical models, along with variables that might be related to those in the model (e.g., whether a family received welfare, which might predict family income) was isolated. For each variable, the software replaced missing cells with values provided by other families who seemed to be similar based on available data. The software repeated this process five times to create five imputed datasets. For a relatively small dataset like this one, five imputations were sufficient (Allison, 2015). When conducting regressions, the software pooled across these five plausible imputed datasets.

Long-Term Literacy Development

In the first set of analyses, children's middle childhood literacy assessment scores were set as dependent variables to test RQ1a and RQ1b. The beta estimate for the Condition term was nonsignificant across all models. However, there was consistent evidence across most outcomes for a positive Condition x Pretest^2 interaction. As shown in Table 8, this interaction term was

Table 8. Child Literacy Outcomes in Middle Childhood (N = 101)

	Rhyming		Deletion		Segmentation		Decoding		Spelling		Reading	
	b	SE	b	SE	b	SE	b	SE	b	SE	b	SE
Female	0.05	0.15	0.27	0.23	-0.25	0.42	1.48	1.74	1.86	1.45	-0.76	4.10
Child Age at Follow-up	-0.05	0.06	0.11	0.09	-0.12	0.17	0.13	0.73	0.75	0.62	-0.66	0.71
Income-to-Needs in Early Childhood	0.05	0.06	$0.17^{\dagger}$	0.09	0.20	0.18	1.23	0.72	0.94	0.63	1.74^{*}	0.66
Change in Income-to-Needs	0.04	0.10	-0.24	0.15	-0.44	0.31	1.01	1.24	-0.07	1.07	2.79^{*}	1.19
Out-of-Area Residence	-0.46	0.40	-0.78	0.54	-0.06	0.71	-2.48	3.15	-1.13	3.25	-0.76	4.10
Pretest in Early Childhood	0.18	0.14	0.22	0.19	0.73^{*}	0.36	3.11^{*}	1.51	3.60^{**}	1.28	5.80^{***}	1.47
Pretest in Early Childhood^2	-0.11	0.08	-0.14	0.11	0.06	0.21	-0.24	0.91	-0.19	0.75	0.50	0.84
Condition	-0.26	0.21	-0.35	0.29	0.29	0.54	-2.84	2.38	-2.43	2.00	0.68	2.24
Condition x Pretest	0.09	0.23	-0.18	0.33	$-1.18^{\dagger}$	0.63	2.76	2.82	1.09	2.31	1.50	2.52
Condition x Pretest^2	0.38^{**}	0.14	$0.38^{\dagger}$	0.21	-0.48	0.39	4.32^{*}	1.71	$2.72^{\dagger}$	1.42	3.18^{*}	1.60
R^2	0.30		0.28		0.18		0.26		0.35		0.46	

Note. Missing cases were imputed using predictive mean matching. $^{\dagger} p < .1$, $^{*} p < .05$, $^{**} p < .01$, $^{***} p < .001$

significant for Rhyming, Decoding, and Reading, and approached significance for Deletion and Spelling. This means that RTL literacy group children with low and high pretest scores in early childhood outperformed counterparts in the control group with comparably low and high pretest scores, but that children with average pretest scores tended to score similarly regardless of condition assignment.

To illustrate this more concretely, a child in the RTL literacy group whose early childhood pretest score was one standard deviation *below* the mean would be expected to score 2.51 points higher in Reading in middle childhood than a peer in the control group who also scored one standard deviation below the mean before the intervention. Put differently, an RTL literacy group participant whose pretest score was one standard deviation below the mean would be expected to be reading approximately one grade level higher than his/her peer in the control group with the same early childhood pretest score (i.e., 4th grade vs. 3rd grade reading level; Kaufman & Kaufman, 2004b). Likewise, an RTL literacy group participant whose early childhood pretest score was one standard deviation *above* the mean would be expected to out-perform a comparable control group peer by 5.13 points or two grade levels (i.e., 9th vs. 7th grade performance) in middle childhood (Kaufman & Kaufman, 2004b). Children, however, who scored "average" at early childhood pretest would be expected to score similarly in middle childhood regardless of condition (children in the RTL literacy group only demonstrated a .64 point advantage, with both groups scoring at a 6th grade level; Kaufman & Kaufman, 2004b). Thus, addressing RQ1a and RQ1b, the RTL literacy games did seem to have affected children's middle childhood literacy, but effects were strongest for children who scored on the lower and higher end at pretest prior to early childhood media exposure.

The only exception to the general pattern found for the literacy skills assessed (i.e., Rhyming, Decoding, Reading, as well as the trends for Deletion and Spelling) was for Segmentation. For this outcome, the Condition x Pretest term was negative and trended toward significance. This meant that for Segmentation, there is some evidence that the RTL literacy games had long-term benefits only for children with below average early childhood pretest scores, but led to lower scores for children with above average scores.

Parental Ratings of Children's Performance

Next, parents' ratings of children's school performance were examined as a means of determining if the effects of the intervention generalized/transferred beyond literacy (addressing RQ2a and RQ2b). As with the literacy results, the beta estimate for Condition was nonsignificant overall and for each individual skill area. However, again mirroring the literacy-specific findings, the key Condition x Pretest^2 interaction for parents' ratings of children's overall school performance was significant. Interestingly, the Condition x Pretest^2 interaction also was positive and significant for parents' ratings of children's performance specifically in science and social studies. Thus, based on parents' ratings, there is some evidence the RTL literacy games' long-term effects extended to school areas beyond English Language Arts, although only for children whose early childhood pretest scores were lower or higher than average. The Condition x Pretest^2 interaction term also trended towards significance for parents' ratings of children's reading and writing, consistent with the direct measure findings reported earlier.

Table 9. *Parent Ratings of Children's Academic Performance in Middle Childhood (N = 101)*

	Overall Performance		Reading		Writing		Science		Social Studies		Math		Speaking/ Listening		Following Directions	
	b	SE	b	SE	b	SE	b	SE	b	SE	b	SE	b	SE	b	SE
Female	0.53**	0.16	0.60**	0.20	0.94***	0.22	0.42*	0.19	0.41†	0.21	0.27	0.24	0.43*	0.17	0.65**	0.20
Child Age at Follow-up	-0.05	0.07	-0.11	0.09	-0.07	0.09	-0.06	0.08	0.02	0.11	-0.09	0.10	-0.04	0.07	-0.02	0.08
Income-to-Needs Early Childhood	0.02	0.06	0.05	0.08	0.02	0.09	0.08	0.07	0.06	0.08	0.02	0.10	-0.05	0.07	-0.01	0.08
Change in Income-to-Needs	0.14	0.12	0.39*	0.16	0.17	0.17	-0.02	0.14	0.10	0.16	0.13	0.17	0.15	0.13	0.07	0.15
Out-of-Area Residence	-0.33	0.26	-0.08	0.33	-0.29	0.35	-0.22	0.30	-0.55	0.34	-0.44	0.39	-0.36	0.28	-0.34	0.32
Pretest Early Childhood	0.36**	0.14	0.54**	0.17	0.31†	0.18	0.44**	0.16	0.36*	0.18	0.45*	0.20	0.24†	0.15	0.17	0.17
Pretest Early Childhood^2	0.00	0.08	0.06	0.10	-0.03	0.11	0.05	0.10	0.02	0.11	-0.08	0.12	0.00	0.09	0.00	0.10
Condition	-0.06	0.22	-0.19	0.28	-0.11	0.29	-0.24	0.26	-0.33	0.29	0.05	0.32	0.14	0.23	0.28	0.27
Condition x	0.03	0.25	0.05	0.32	0.13	0.34	0.07	0.29	0.00	0.33	-0.02	0.37	-0.16	0.27	0.14	

Pretest

Condition x Pretest^2	0.32*	0.16	0.35†	0.20	0.36†	0.21	0.47*	0.19	0.42*	0.21	0.31	0.24	0.15	0.17	0.20	0.20
R^2	0.33		0.34		0.30		0.28		0.22		0.23		0.22		0.21	

Note. Missing cases were imputed using predictive mean matching. Overall performance averages parents' rating across the individual academic categories listed in the remaining columns in this table (Cronbach's α = .92; † $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$).

Discussion

The RTL-funded early childhood literacy computer game had long-lasting impacts on children's literacy as assessed by standardized measures, as well as their performance in other school subjects as reported by parents. Even though the original intervention was fairly short (i.e., 6-8 weeks) and primarily focused on literacy (Schmitt et al., 2017), this early learning experience seems to have been powerful enough to have enhanced some children's learning trajectories in English Language Arts and in other subject areas throughout elementary school. To an extent, these findings substantiate the main principle of the Early Learning Hypothesis (i.e., that early educational media exposure can catalyze long-term scholastic success; Anderson et al., 2001) with strong, causal, longitudinal evidence, and indicate that the Early Learning Hypothesis extends beyond television to early educational computer play. As intended (Singer & Singer, 1998), the RTL media resulted in some children being more prepared to succeed in elementary school. However, effects were only detectable for children with below and above average literacy skills prior to exposure to the intervention in early childhood, with negligible effects for children with average pretest scores. Although there is some ambiguity as to the exact mechanisms and pathways that yielded these results due in large part to the large time gap between waves of data collection, the Traveling Lens (Huston & Wright, 1989; Rice et al., 1982) and Capacity Models (Fisch, 2000, 2004, 2016) can provide insights as to why this pattern of results have arisen.

It is possible that the learning content in the games may have initially aligned well with the skillsets of the weaker readers in the sample. Before the intervention, these children scored low on measures of very basic literacy such as alphabet letter knowledge. The RTL games,

which began by promoting this basic literacy content and leveled up to more advanced content (see Grant et al., 2012), might have engaged these children and scaffolded their learning in a manner consistent with the Traveling Lens Model (Huston & Wright, 1989; Rice et al., 1982). In other words, the RTL literacy games may have begun at just the right degree of difficulty for the weakest readers in the sample and slowly became more challenging in a way that continued to align with their growing literacy (Huston & Wright, 1989; Rice et al., 1982). Furthermore, the strong learning and engagement the games fostered for the children most in need of literacy remediation might have provided them a toolkit of literacy strategies they could draw from to excel in first grade reading classes or in other classes requiring reading and helped to improve their attitudes towards reading and literacy (Fisch, 2004; Fisch et al., 2005). This in turn may have helped them place into more advanced groups for reading or for other school subjects requiring skill in reading (Anderson et al., 2001). A combination of these circumstances may have initiated a chain of continued scholastic success as per the Early Learning Hypothesis (Anderson et al., 2001).

In contrast, children with above average literacy skills at early childhood pretest might have found the RTL gaming suite engaging and reinforcing despite less-than-perfect skill alignment at first. Initially, they may have been interested and motivated to play the beginning, easy games just because they enjoyed engaging with literacy-themed content, somewhat consistent with the Capacity Model (Fisch, 2004). That is, these children may have liked literacy and voluntarily allocated attentional resources to the literacy-themed games, even though they may have already known some of the lessons promoted in the early levels (Fisch, 2004; Hirsh-Pasek et al., 2015). Doing so may have helped them rehearse and strengthen foundational skills

(Fisch, 2004; Rice et al., 1982). They later may have continued to approach the gaming experience with comparable focus (Hirsh-Pasek et al., 2015), eventually reaching more advanced levels that introduced newer content. They also may have uniquely benefited from certain games with text elements that might have been incomprehensible to other children or from games with more linguistically complex instructions administered orally, again consistent with Capacity Model principles (Fisch, 2000; 2004). Altogether, the games may have strengthened, reinforced, and extended these children's literacy skills and/or their motivation to engage or interest in reading at a crucial time before they began formal schooling, setting them on an even more advanced academic trajectory than they were on prior to the intervention (Anderson et al., 2001).

Why might the findings for Segmentation not mirror the findings for the other literacy measures? Children's early childhood pretest scores significantly predicted their middle childhood Segmentation scores (see Table 8), and Segmentation scores were significantly correlated with all other middle childhood literacy measures except Rhyming and Deletion ($.11 \leq r$s $\leq .28$, ps $\geq .30$). However, children may not have fully understood the Segmentation task instructions. The training items for this task asked children to practice segmenting words with two syllables or phonemes (sounds), and many children consistently segmented words into two parts, regardless of the correct number of syllables or phonemes in a given test item (i.e., even if test items contained three or more syllables or phonemes). To address these sorts of issues, Pearson recently released a newer streamlined version of the KTEA (Scheller, 2014). The version of the Segmentation test used in the present study may not have provided as accurate a depiction of children's literacy as the other literacy measures.

Limitations

The findings from this study should be interpreted in light of four limitations. First and foremost, plausible pathways through which the games may or may not have influenced children's long-term outcomes were suggested. Although these proposed explanations align with the data collected and theory, additional data could have provided even stronger evidence for the arguments herein. For example, researchers did not directly measure engagement, motivation, or related constructs in a standardized, quantifiable fashion in the original study. Thus, statements about children's potential level of focus during gameplay cannot be verified. Future scholars conducting similar early learning evaluations could measure motivation via direct verbal measures or by monitoring biomarkers such as heart rate (Aladé & Nathanson, 2016). Likewise, data on children's performance in first grade or earlier in elementary school were not collected, and therefore one can only surmise that the pathways Anderson and colleagues (2001) proposed are correct. Future RTL evaluators or other scholars conducting similar research should consider prospectively planning follow-up studies at regular intervals to provide an even richer understanding of educational media's long-term effects.

Second, families were recruited to the original evaluation by convenience sampling, which presents a threat to the external validity of the games' impact in both the short- and long-term (Bornstein et al., 2013). The sample was diverse in terms of race and socio-economic status. However, these families fairly universally reported highly valuing education. All participating families chose to enroll children in an optional literacy intervention in preschool or kindergarten, and many families in both conditions anecdotally reported continuing to enroll children in optional scholastic enrichment programs. As such, the pattern of results may have been different for families who place less emphasis on education, especially in light of other research showing

parent beliefs about education predict children's educational attainment (e.g., Wu & Qi, 2006).

Third, the original intervention focused only on one suite of early learning computer games created with RTL funding and intended to promote literacy. Meta-analytic work suggests there is considerable variability in the effectiveness of RTL media properties in the short-term (Article 1). Although these particular games featured multiple media properties (*Between the Lions, Martha Speaks, Sesame Street, Super WHY!,* and *WordWorld*), it is possible different long-term results would have emerged had children played games focused on different media properties. Likewise, differing results may have arisen had the focal media content (a) initially presented more challenging content, (b) had the lessons been delivered via a television show or app or (c) drawn from other, non-RTL-funded educational media properties.

Finally, parent reports of children's school performance were used to address RQ2a and RQ2b. Measuring children's school performance directly would have allowed a more definitive determination of the degree to which results transferred beyond literacy.

Conclusion

The results from this longitudinal evaluation provide causal evidence that even relatively brief exposure to educational computer games in early childhood can have long-lasting effects detectable six years after initial media exposure for some children both in the learning domain directly promoted (in this case literacy), with additional downstream effects on other domains (e.g., science and social studies). These findings support expenditure on such educational tools through programs like the U.S. RTL initiative. Moreover, these findings, to a certain extent, substantiate and help add nuance to scholarly understanding of the Early Learning Hypothesis (Anderson et al., 2001): Early educational computer game exposure can have long-term effects

on children's school performance throughout elementary school. However, media may not be a silver bullet that equally affects all children (Piotrowski & Valkenburg, 2015; Valkenburg & Peter, 2013). Aligned with the Traveling Lens Model, early media's long-term effects may be particularly strong when mediated content closely reflects children current skillsets (Huston & Wright, 1989; Rice et al., 1982), or, somewhat in accordance with the Capacity Model, when children themselves are able to bring stronger skills and interest to the mediated experience (Fisch, 2000, 2004). To help ensure comparable long-term effects are realized by *all* children, it may be necessary to create games that can quickly adapt educational lessons to children's skillsets to make the gaming experience more universally engaging (Guernsey, 2012; Roberts et al., 2016).

Conclusion

While RTL media are not a magic bullet that ameliorate all achievement gaps at school entry, overall this dissertation suggests the RTL initiative is effective. RTL media have modest but positive effects on children's early literacy skills, and this early boost lasts for some children at least through elementary school and generalizes to other subject domains beyond literacy. Below, broader implications for the RTL initiative and for children's media theory are discussed, cutting across insights gleaned from all three articles.

Implications for the RTL Initiative

Even though school readiness, especially among low-income populations, is still a public policy concern in the U.S. (Bradbury, Corak, Waldfogel, & Washbrook, 2015), federal funding for public media and many educational programs/services like RTL is at risk as of the time of this writing. Yet, as stated above, this dissertation provides fairly positive accountability evidence in favor of the RTL initiative and its role in promoting early literacy and school readiness among both general and at-risk populations. RTL media yield effects on par with Head Start (Kay & Pennucci, 2014). Only $25 million in national taxpayer dollars are allocated for RTL each year (DoEd, 2015). While this may initially seem like a considerable sum, and indeed I note that this expenditure is nontrivial in Article 1, it is miniscule compared to other federal budget line items. For instance, Congress allocated upwards of $9 *billion* for Head Start last year (National Head Start Association, 2017). And other developed countries allocate considerably larger proportions of taxpayer dollars for public media (Lee, 2012). In this context, eliminating or significantly reducing such a small line item like RTL seems unfounded. Indeed, Kearney and

Levine (2015) previously have made similar arguments about the benefit of educational public media following an examination of the effects of *Sesame Street* prior to the launch of RTL.

Despite the overall positive findings of this dissertation, there is a need for continued monitoring of the RTL initiative (assuming it remains funded), especially as the program has continuously evolved with a sizable number of newer properties dedicated to promoting STEM learning and with increasing emphasis on various forms of new media (Pasnik, Llorente, Hupert, & Moorthy, 2016). It is difficult to say whether the effects found in this dissertation would generalize to the current RTL context. That is, it is hard to predict whether a comparable set of studies focused on STEM-themed properties would have found similar results. Article 1 implies that future reviews may note larger effects as RTL producers continue to refine their approach to community outreach (Llorente et al., 2010) and improve technology designed to automatically level play experiences to align with children's existing skillsets (Roberts et al., 2016). Additionally, the new media platforms producers currently are favoring might particularly lend themselves to promoting certain types of inquiry-based STEM learning (Fisch, Damashek, & Aladé, 2016). On the other hand, also in Article 1, RTL media's effects on literacy became larger across grant cycles, perhaps as producers became more skilled at promoting learning in that domain. It is possible that RTL's effects on STEM learning might similarly be small now but might grow over time. Relatedly, it may be harder for researchers to adequately measure STEM learning in preschool-age children due to a lack of relevant, validated measures (Pasnik, Llorente, Hupert, Dominguez, & Silander, 2015), such that evaluators may be less well-equipped to assess learning in STEM than literacy at this point.

It likewise is impossible to surmise how the effective older RTL literacy-themed media might be if lifted out of their original contexts and re-evaluated today, given that the specific focus of each RTL grant cycle has been very much a reflection of broader and concurrent educational policy discussions. The 2005-2010 grant cycle focused on literacy shortly after the dissemination of the conclusions by the National Reading Panel (Langenberg et al., 2000) and prompted media adhering to those recommendations (Michael Cohen Group, 2012). The 2010-2015 cycle focused on both literacy and mathematics aligned with the then new Common Core State Standards, which also centered on those topic areas (PBS, 2011). The current cycle concerns science and literacy, consistent with the new and popular Next Generation Science Standards (Educational Development Center & SRI International, 2016). Preschool and early elementary school teachers around the country likely have been drawing from curricula that emphasize and promote learning in a manner aligned with the same policy shifts. Thus, lessons children received in school may have complemented the media they consumed as part of RTL evaluations. Even though there was not support for the benefit of Joint Media Engagement as operationalized in Article 1 (i.e., parents/teachers doing any sort of extension activity), educators and parents still may be able to extend media's lessons through in-person lessons that mirror media's content but risk mitigating media's effects by delivering contradictory lessons (Savage et al., 2013).

Accordingly, it is possible some of the literacy media products examined in the present dissertation might be less effective today or in the future as learning science and related policy continue to evolve, and as teachers adopt new instruction methods that may reflect the latest pedagogical knowledge but may become increasingly dissimilar to the way older RTL media

delivered lessons. Therefore, it may be wise for content producers to take care to remove products created with dated pedagogy from circulation. It seems as if PBS may already be doing this through its PBS LearningMedia hub (https://www.pbslearningmedia.org/), but older products often remain accessible elsewhere online (Webster, 2014) for potential use by parents, teachers, or researchers. This is not an issue for evaluators testing or teachers using recently produced media, just a potential concern when considering whether a media product produced in a different educational policy context and deemed effective at that time would have the same effects today.

In sum, it seems unjustified that RTL and programs like it are at risk for elimination based on the present positive findings. However, that is not to say that RTL and related programs should be immune from continued monitoring. Moreover, teachers and caregivers should critically consider whether a given RTL media product is appropriate for a child based on his/her broader educational experiences, keeping in mind that an older product once judged effective may cease to mirror the most recent learning science.

Implications for Children's Media Theory

In addition to having practical policy ramifications, this dissertation also extends theoretical thinking concerning children's learning from media in some areas, and raises new questions in others. As discussed at length in Article 3, these findings provide causal evidence substantiating the Early Learning Hypothesis and indicate that this hypothesis applies to new media. Moreover, these results also suggest that it is necessary to consider individual differences when assessing media's long-term effects. Bridging both latter points, these findings further

signal that some children's media theories originally intended to describe short-term individual difference phenomena may continue to bear out in the longer-term.

Future research beyond this dissertation is needed, however, to further integrate these patterns of individual differences into a more coherent theoretical framework and to definitively differentiate between the potential benefits and detriments of new media. Each of these topics – the value of new media, and theory regarding individual differences – is discussed in greater detail below.

A content-based approach to considering new media. In the larger children's media research community, scholars recently and frequently have asked whether curriculum-based new media can promote learning at all (e.g., Blackwell, 2015) or whether it can do so as well as television (e.g., Aladé et al., 2016). In Article 1, the meta-analytic results suggest that on average, television and new media proved equally effective in promoting children's early literacy, but I speculate that new media may have not fully realized its potential as an educational tool and may become increasingly effective over time (Hirsh-Pasek et al., 2015).

It could be more sophisticated, however, to ask under *what circumstances* new media has the potential to be effective at all, or to be more effective than television. Just as children's media scholars are increasingly considering the content, context, and individual child when discussing television's potential effects on children (Guernsey, 2012), so too might it be necessary to consider such factors when determining whether an educational lesson would be delivered more effectively via television or new media. The overall null difference between television and new media observed in Article 1 may be attributable to content producers or evaluators either instinctively choosing to test properties via the platform that is best suited to a given content,

context, child combination, or to mistakes in choice of platform cancelling each other out (e.g., an evaluation of a highly effective app averaged with an evaluation of an ineffective app). In that vein, additional analyses of the Article 1 data not reported above reveal that researchers tended to use new media more often in home-based studies and television more often in school-based studies (χ^2(1, k = 783) = 41.65, p =.01). Evaluators or producers may have intuited that new media could work well in the home context, where children could easily play independently or with the support of other family members (Plowman, McPake, & Stephen, 2008), whereas television viewing might lend itself better to a whole class activity.

In additional exploratory analyses also not included in Article 1, results indicated that new media was more effective than television at promoting alphabet knowledge (z = 2.10, p = .04), although there were no other significant differences between television and new media for any other literacy outcomes. Alphabet knowledge is a more basic early literacy skill particularly germane to children on the younger end of RTL's target age range (Grant et al., 2012). Younger preschoolers tend to consume television in a less attentive fashion than their slightly older peers (Anderson & Lorch, 1983), but perhaps would be more focused playing a simple interactive game (Michael Cohen Group, 2011). In other words, child-level developmental state and the content of the media could explain this finding. Alphabet knowledge also hypothetically may receive less of a boost from the stronger narratives typically provided in television (unlike vocabulary, for example, where contextual information aids learning; Gola et al., 2012).

Although others similarly argued for the importance of considering a given platform's fit to an educational context (e.g., Fisch et al., 2016), this line of thinking has not permeated the way scholars currently are designing new media research or conceptualizing related theoretical

frameworks. When researchers find mixed results as to new media's effectiveness at all or effectiveness relative to television, frequently, they provide posthoc explanations for such results. The findings of this dissertation suggest a need for more theoretical work to help prospectively guide decisions to use or test new media or television, moving beyond *whether* a given platform is appropriate to *when* it might be.

Alternatively, the larger children's media research community may be overly fixated on the differences in affordances between television and new media (Wartella, 2015). Television and new media products may truly lead to roughly equal potential effects. Perhaps the potential boost well-designed interactivity could provide (Hirsh-Pasek et al., 2015) is equivalent to the benefit of a stronger narrative that television is more likely to offer (Lu et al., 2012).

Interchangeably referring to both computer games and mobile apps as new media was necessary for the present line of inquiry (there were not enough evaluations to examine computer games and apps separately in Article 1), but would be problematic going forward, especially in light of RTL's very young target audience. Children on the younger end of RTL's target age range have more difficulty using a mouse than children in early elementary school (Crook, 1992), and consequential navigational difficulties could limit children's learning from computers (Guernsey, 2012). Although similar age-related patterns are noted in studies of children's tablet engagement (Michael Cohen Group, 2011), difficulties may be less pronounced. Thus, younger preschool children might learn more from apps than similar computer games. In Article 3, the learning content of the 2010 RTL computer game began with a focus on relatively basic literacy skills that children typically begin learning around age 3 (Grant et al., 2012), even though children in the original sample ranged in age from 4-6 years. In app form, the game could have

been "aged down" to be developmentally appropriate for 3-year-old children, provided they had background familiarity with at least some letters of the alphabet to allow them to assimilate the in-game lessons into their existing schemas about literacy (Schmitt & Linebarger, 2009). This design choice could have led to stronger effects in both the short- and long-term.

Growing the garden. Elsewhere in Article 3, results provide evidence substantiating both the Traveling Lens Model's notion of media helping children whose skillsets align with its content (Huston & Wright, 1989; Rice et al., 1982) and the Capacity Model's notion of media benefiting the children who have the strongest verbal skills (Fisch, 2004). Reconsidering and reconceptualizing the Differential Susceptibility to Media Effects Model and its notions of orchid children who are susceptible to media vs. dandelion children who are impervious (Piotrowski & Valkenburg, 2015) in light of these findings may help to reconcile this apparent discrepancy.

Initially considering these results, one might conclude that the low-income children in Article 1 (aligned with Linebarger & Barr, 2017) and the below average early readers in Article 3 are orchid children, while children with average early literacy skills in Article 3 are dandelions. After all, the former groups of children particularly benefited from early RTL media exposure, while the latter group seemed more impervious. However, as argued above, I am not convinced that the children with average literacy skills were hopeless cases. According to the Traveling Lens Model (Huston & Wright, 1989; Rice et al., 1982), these children could have exhibited comparable effects had they played a game better aligned with their current abilities. That is, with a game that began with more challenging content or that automatically re-calibrated to better meet their needs, these children too may have grown like orchids. The right temperature, level of humidity, etc. may have been in place in 2010 for the below average children who

participated in the original evaluation, but the children with average preliteracy skills may have flourished in slightly warmer conditions. Future empirical work is needed with harder games or with games using item response theory to automatically level content to more definitively determine if there really are dandelion children, at least under circumstances mirroring the present study (i.e., when families opt-in to engage in an educational media intervention).

One could also make a case that the children with above average early literacy skills in Article 3 were more orchids. These children were already strong readers for their age prior to media exposure (a beautiful characteristic, reminiscent of a beautiful flower) and seemed to benefit from the intervention. However, the data does not support the notion that these children might have withered without the early boost provided by the game. Children in the control condition with comparable above average pretest scores were still scoring above average in middle childhood; their scores were just slightly lower than their RTL peers. It may be more apt to compare these gifted children to tulips. They would grow well in many conditions with minimal interference, but blossom even more with the right intervention.

Altogether, I am arguing that this intervention grew a garden with multiple genera of orchids (children who are not precocious but who can benefit from developmentally appropriate media, as per the Traveling Lens Model; Huston & Wright, 1989; Rice et al., 1992) and tulips (above average children who would still be gifted even without intervention but who benefit from extra exposure to engaging educational content, as per the Capacity Model; Fisch, 2004). I do not see any dandelions in this garden. But perhaps because the original intervention in Article 3 used convenience sampling, with families electing to engage in the optional intervention, only flowers with the potential to grow were planted (i.e., participated).

As a caveat, Piotrowski and Valkenburg (2015) specifically call for identifying child-level factors that simultaneously make children more susceptible to both positive *and negative* media effects – that these factors should help identify orchid children. Because the RTL evaluations explored herein did not expose children to any negative forms of media, I can only partially expand up their model and am unsure the extent to which these ideas would translate to violent media effects.

Closing Thoughts

Overall, these findings provide fairly supportive accountability evidence in favor of RTL's power to positively influence children's school readiness and school performance, and point to a need for more nuanced examination of undergirding causal mechanisms. RTL media have positive short-term effects on children's early literacy and lasting impacts on some children's performance in literacy and other school subjects throughout elementary school. The findings of this dissertation therefore support the continued funding of this initiative. Nonetheless, it is important not to overgeneralize these results. More research is needed to verify that the initiative continues to induce comparable change under evolving conditions and to help scholars better understand when television or new media are more appropriate tools to induce learning and when individual children may benefit the most from media exposure.

References

Abu-Lughod, J., & Foley, M. M. (1960). Consumer strategies. In N. N. Foote, J. Abu-Lughod,

    M. M. Foley & L. Winnick (Eds.), *Housing choices and housing constraints* (pp. 71-

    271). New York: McGraw Hill.

Adlof, S. M., Catts, H. W., & Lee, J. (2010). Kindergarten predictors of second versus eighth

    grade reading comprehension impairments. *Journal of Learning Disabilities, 43*, 332-

    345. doi:10.1177/0022219410369067

Agrawal, K. C., Kellam, S. G., Klein, Z. E., & Turner, J. (1978). The Woodlawn mental health

    studies: Tracking children and families for long-term follow-up. *American Journal of

    Public Health, 68*, 139-142. doi:10.2105/AJPH.68.2.139

Ahlers-Schmidt, C. R., Chesser, A. K., Nguyen, T., Brannon, J., Hart, T. A., Williams, K. S., &

    Wittler, R. R. (2012). Feasibility of a randomized controlled trial to evaluate Text

    Reminders for Immunization Compliance in Kids (TRICKs). *Vaccine, 30*, 5305-5309.

    doi:10.1016/j.vaccine.2012.06.058

Aladé, F. (2013). *What preschoolers bring to the show: The effects of cognitive abilities and

    viewer characteristics on children's learning from educational television* (Master's

    thesis). Retrieved from http://rave.ohiolink.edu/etdc

Aladé, F., Lauricella, A. R., Beaudoin-Ryan, L., & Wartella, E. (2016). Measuring with Murray:

    Touchscreen technology and preschoolers' STEM learning. *Computers in Human

    Behavior, 62*, 433-441. doi:10.1016/j.chb.2016.03.080

Aladé, F., & Nathanson, A. I. (2016). What preschoolers bring to the show: The relation between vewer characteristics and children's learning from educational television. *Media Psychology, 19*, 406-430. doi:10.1080/15213269.2015.1054945

Allison, P. (2015). Imputation by predictive mean matching: Promise and peril. Retrieved from http://statisticalhorizons.com/predictive-mean-matching

American Chamber of Commerce Research. (2007). Cost of living index for selected U.S. cities, 2005. Retrieved from http://www.infoplease.com/ipa/a0883960.html

American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.

Anderson, D. R., & Collins, P. A. (1988). *The impact on children's education: Television's influence on cognitive development (Working paper no. 2)*. Washington, D.C.: Office of Education Research and Improvement.

Anderson, D. R., Huston, A. C., Schmitt, K. L., Linebarger, D. L., & Wright, J. C. (2001). Early childhood television viewing and adolescent behavior: The recontact study. *Monographs of the Society for Research in Child Development, 66*, vii-147. doi:10.1111/1540-5834.00120

Anderson, D. R., & Lorch, E. P. (1983). Looking at television: Action or reaction? In J. Bryant & D. R. Anderson (Eds.), *Children's understanding of television* (pp. 1-33). New York, NY: Academic Press.

Ayduk, O., Mendoza-Denton, R., Mischel, W., Downey, G., Peake, P. K., & Rodriguez, M. (2000). Regulating the interpersonal self: Strategic self-regulation for coping with

rejection sensitivity. *Journal of Personality and Social Psychology, 79*, 776-792. doi:10.1037/0022-3514.79.5.776

Barakat-Haddad, C. P., Elliott, S. J., Eyles, J., & Pengelly, D. (2009). Predictors of locating children participants in epidemiological studies 20 years after last contact: Internet resources and longitudinal research. *European Journal of Epidemiology, 24*, 397-405. doi:10.1007/s10654-009-9364-5

Barnett, W. S. (2013). *Getting the facts right on pre-k and the President's pre-k proposal*. New Brunswick, NJ: National Institute for Early Education Research.

Blackwell, C. K. (2015). *Technology use in early childhood education: Investigating teacher access and attitudes toward technology and the effect of iPads on student achievement* (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global database. (3705215)

Bonner, M., & Carpenter, D. (2005). Kaufman Test of Educational Achievement, Second edition, Comprehensive form. In R. A. Spies & B. S. Plake (Eds.), *The sixteenth mental measurements yearbook*: University of Nebraska Press & the Buros Center for Testing. Retrieved from http://web.a.ebscohost.com.turing.library.northwestern.edu/.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Hoboken, NJ: Wiley.

Bornstein, M. H., Jager, J., & Putnick, D. L. (2013). Sampling in developmental science: Situations, shortcomings, solutions, and standards. *Developmental Review, 33*, 357-370. doi:10.1016/j.dr.2013.08.003

Boswell, W. (2007, July 23). Where to find public records online, *Lifehacker*. Retrieved from
http://lifehacker.com/

Bradbury, B., Corak, M., Waldfogel, J., & Washbrook, E. (2015). *Too many children left behind:*
The U.S. achievement gap in comparative perspective. New York, NY: Russell Sage
Foundation.

Bryant, J. A., Bryant, J., Mullikin, L., McCollum, J. F., & Love, C. C. (2001). Curriculum-based
preschool television programming and the American family: Historical development,
impact of public policy, and social and educational effects. In J. Bryant & A. Bryant, J.
(Eds.), *Television and the American family* (2nd ed., pp. 415-434). Mahwah, NJ:
Lawrence Erlbaum Associates.

Bus, A. G., & van Ijzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-
analysis of experimental training studies. *Journal of Educational Psychology, 91*, 403-
414. doi:10.1037/0022-0663.91.3.403

Calfee, R. C., Lindamood, P., & Lindamood, C. (1973). Acoustic-phonetic skills and reading:
Kindergarten through twelfth grade. *Journal of Educational Psychology, 64*, 293-298.
doi:10.1037/h0034586

Campbell, F. A., Pungello, E. P., Burchinal, M., Kainz, K., Pan, Y., Wasik, B. H., . . . Ramey, C.
T. (2012). Adult outcomes as a function of an early childhood educational program: An
Abecedarian Project follow-up. *Developmental Psychology, 48*, 1033-1043.
doi:10.1037/a0026644

Can, D. D., Ginsburg-Block, M., Golinkoff, R. M., & Hirsh-Pasek, K. (2013). A long-term
predictive validity study: Can the CDI Short Form be used to predict language and early

literacy skills four years later? *Journal of Child Language, 40*, 821-835.

doi:10.1017/S030500091200030X

Cohen, M., Hadley, M., & Marcial, C. (2016). The role of research and evaluation in Ready to Learn transmedia development. *Journal of Children and Media, 10*, 248-256. doi:10.1080/17482798.2016.1140488

Cook, T. D., Appleton, H., Conner, R. F., Shaffer, A., Tamkin, G., & Weber, S. J. (1975). *Sesame Street revisited*. New York, NY: Russell Sage Foundation.

Cooper, H. M. (1981). On the significance of effects and the effects of significance. *Journal of Personality and Social Psychology, 41*, 1013-1018. doi:10.1037/0022-3514.41.5.1013

Cotter, R. B., Burke, J. D., Loeber, R., & Navratil, J. L. (2002). Innovative retention methods in longitudinal research: A case study of the developmental trends study. *Journal of Child and Family Studies, 11*, 485-498. doi:10.1023/a:1020939626243

Cotter, R. B., Burke, J. D., Stouthamer-Loeber, M., & Loeber, R. (2005). Contacting participants for follow-up: How much effort is required to retain participants in longitudinal studies? *Evaluation and Program Planning, 28*, 15-21. doi:10.1016/j.evalprogplan.2004.10.002

Council for Community and Economic Research. (2010). Cost of living index for selected U.S. cities. Retrieved from http://www.infoplease.com/business/economy/cost-living-index-us-cities.html

CPB, & PBS. (2011). *Findings from Ready To Learn 2005-2010*. Arlington, VA: Corporation for Public Broadcasting and Public Broadcasting Service.

Creager, K. (2011, July 5). Is your email address preventing you from getting the job you want? [Blog post], *Brazen Blog*. Retrieved from http://www.brazen.com/blog/archive/job-search/is-your-email-address-preventing-you-from-getting-the-job-you-want/

Crook, C. (1992). Young children's skill in using a mouse to control a graphical computer interface. *Computers & Education, 19*, 199-207. doi:10.1016/0360-1315(92)90113-J

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology, 33*, 934-945. doi:10.1037/0012-1649.33.6.934

DoEd. (2015). Funding status. Retrieved from http://www2.ed.gov/programs/rtltv/funding.html

Dost, M., & McGeeney, K. (2016). *Moving without changing your cellphone number: A predicament for pollsters*. Washington, D.C.: Pew Research Center's Internet & American Life Project.

Duch, H. (2005). Redefining parent involvement in Head Start: A two-generation approach. *Early Child Development and Care, 175*, 23-35. doi:10.1080/0300443042000206237

Duggan, M. (2013). *Cell phone activities 2013: 50% of cell owners download apps to their phones; 48% listen to music services; video calling has tripled since 2011; texting remains a popular activity.* Washington, D.C.: Pew Research Center's Internet & American Life Project.

Duggan, M., Ellison, N. B., Lampe, C., Lenhart, A., & Madden, M. (2015). *Socail media update 2014: While Facebook remains the most popular site, other platforms see higher rates of growth*. Washington, D.C.: Pew Research Center's Internet & American Life Project.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test-Third Edition (PPVT-III)*. Circle Pines, MN: American Guidance Service.

Dupre, E. (2014, April 1). Growing up Gmail: 10 years of email marketing lessons, *DMN*. Retrieved from http://dmnews.com

Educational Development Center, & SRI International. (2012). *PBS KIDS transmedia suites gaming study*. Education Development Center, Inc. & SRI International. New York and Menlo Park, CA.

Educational Development Center, & SRI International. (2016). *Science + literacy: A formative research tool for CPB-PBS Ready To Learn producers*. Education Development Center, Inc. & SRI International. New York and Menlo Park, CA.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*, 629-634. doi:10.1136/bmj.315.7109.629

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly, 36*, 250-287. doi:10.1598/RRQ.36.3.2

Fein, S. B., Li, R., Chen, J., Scanlon, K. S., & Grummer-Strawn, L. M. (2014). Methods for the year 6 follow-up study of children in the Infant Feeding Practices Study II. *Pediatrics, 134*, S4.

Fisch, S. M. (2000). A Capacity Model of children's comprehension of educational content on television. *Media Psychology, 2*, 63-91. doi:10.1207/S1532785XMEP0201_4

Fisch, S. M. (2004). *Children's learning from educational television: Sesame Street and beyond*. Mahwah, NJ: Lawrence Erlbaum.

Fisch, S. M. (2016, October). The Capacity Model, 2.0: Cognitive processing in children's comprehension of educational games. In J. T. Piotrowski (Chair), *Digital media and early literacy.* Symposium conducted at the Society for Research on Child Development Special Topics Meeting: Technology and Media in Children's Development, Irvine, CA.

Fisch, S. M., Damashek, S., & Aladé, F. (2016). Designing media for cross-platform learning: Developing models for production and instructional design. *Journal of Children and Media, 10*, 238-247. doi:10.1080/17482798.2016.1140485

Fisch, S. M., Kirkorian, H. L., & Anderson, D. R. (2005). Transfer of learning in informal education. In J. P. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 371-393). Greenwich, CT: Information Age Publishing.

Fisch, S. M., & Truglio, R. (Eds.). (2000). *G is for "growing": Thirty years of research on children and Sesame Street*. Mahwah, NJ, USA: Lawrence Erlbaum

Gola, A. A. H., Mayeux, L., & Naigles, L. R. (2012). Electronic media as incidental language teachers. In D. G. Singer & J. L. Singer (Eds.), *Handbook of children and the media* (2nd ed., pp. 139-156). Thousand Oaks, CA: SAGE.

Golinkoff, R. M., Jacquet, R. C., Hirsh-Pasek, K., & Nandakumar, R. (1996). Lexical principles may underlie the learning of verbs. *Child Development, 67*, 3101-3119. doi:10.1111/j.1467-8624.1996.tb01905.x

Grant, A., Wood, E., Gottardo, A., Evans, M. A., Phillips, L., & Savage, R. (2012). Assessing the content and quality of commercially available reading software programs: Do they

have the fundamental structures to promote the development of early reading skills in children? *NHSA Dialog, 15*, 319-342. doi:10.1080/15240754.2012.725487

Griffin, T. M., Hemphill, L., Camp, L., & Wolf, D. P. (2004). Oral discourse in the preschool years and later literacy skills. *First Language, 24*, 123-147. doi:10.1177/0142723704042369

Guernsey, L. (2012). *Screen time: How electronic media - from baby videos to educational software - affects your young child*. Philadelphia, PA: Basic Books.

Guernsey, L., & Levine, M. H. (2015). *Tap, click, read: Growing readers in a world of screens*. San Francisco, CA: Jossey-Bass.

Haggerty, K. P., Fleming, C. B., Catalano, R. F., Petrie, R. S., Rubin, R. J., & Grassley, M. H. (2008). Ten years later: Locating and interviewing children of drug abusers. *Evaluation and Program Planning, 31*, 1-9. doi:10.1016/j.evalprogplan.2007.10.003

Hanson, K. G. (2017). *The influence of early media exposure on children's development and learning* (Unpublished doctoral dissertation). University of Massachusetts Amherst, Amherst, MA. Retrieved from

Harris, M. A., Brett, C. E., Johnson, W., & Deary, I. J. (2016). Personality stability from age 14 to age 77 years. *Psychology and Aging, 31*, 862-874. doi:10.1037/pag0000133

Hartmann, D. P. (1992). Design, measurement, and analysis: Technical issues in developmental research. In M. H. Bornstein & M. E. Lamb (Eds.), *Developmental psychology: An advanced textbook* (3rd ed., pp. 132-154). Hillsdale, NJ: Lawrence Erlbaum.

Heatly, M. C., Bachman, H. J., & Votruba-Drzal, E. (2015). Developmental patterns in the associations between instructional practices and children's math trajectories in elementary

school. *Journal of Applied Developmental Psychology, 41*, 46-59.

doi:10.1016/j.appdev.2015.06.002

Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics, 32*, 341-370. doi:10.3102/1076998606298043

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis, 29*, 60-87. doi:10.3102/0162373707299706

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives, 2*, 172-177. doi:10.1111/j.1750-8606.2008.00061.x

Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., & Kaufman, J. (2015). Putting education in "educational" apps: Lessons from the science of learning. *Psychological Science in the Public Interest, 16*, 3-34. doi:10.1177/1529100615569721

Horowitz, J. E., Juffer, K., Davis Sosenko, L., Stout Hoffman, J. L., Bojorquez, J. C., Dailey, K., & Wingren, G. (2005). *Data collection of the federal performance indicators for PBS Ready To Learn: Year five summary*. Los Alamitos, CA: WestEd.

Huff, J. O., & Clark, W. A. V. (1978). Cumulative stress and cumulative inertia: A behavioral model of the decision to move. *Environment and Planning A, 10*, 1101-1119. doi:10.1068/a101101

Hurtado, M., Galdo, J., Agin, L., & Heil, S. (2010). *A social marketing approach to promoting literacy-related behaviors among low-income families*. Paper presented at the 2010 Annual Meeting of the American Educational Research Association, Denver, Colorado.

Huston, A. C., Duncan, G. J., McLoyd, V. C., Crosby, D. A., Ripke, M. N., Weisner, T. S., & Eldred, C. A. (2005). Impacts on children of a policy to promote employment and reduce poverty for low-income parents: New Hope after 5 years. *Developmental Psychology, 41*, 902-918. doi:10.1037/0012-1649.41.6.902

Huston, A. C., & Wright, J. C. (1989). Forms of television and the child viewer. In G. A. Comstock (Ed.), *Public communication and behavior* (Vol. 2, pp. 103-159). Orlando, FL.: Academic Press.

Institute of Research on Poverty. (2010). What are the poverty thresholds and poverty guidelines? Retrieved from http://www.irp.wisc.edu/faqs/faq1.htm#year2000

Invernizzi, M., Sullivan, A., Meier, J., & Swank, L. (2004). *PreK teacher manual: PALS Phonological Awareness Literacy Screening*. Richmond, VA: University of Virginia.

Jacobsson Purewal, S. (2015, January 22). 5 tips for finding anything, about anyone, online, *CNET*. Retrieved from http://dmnews.com

Jennings, N. (2013). *Super readers at CET/ThinkTV: An evaluation of Super WHY! reading camps: Final report*. Children's Education and Entertainment Research (CHEER) Lab, University of Cincinnati. Cincinnati, OH.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80*, 437-447. doi:10.1037/0022-0663.80.4.437

Kalkhoff, W., Youngreen, R., Nath, L., & Lovaglia, M. J. (2014). Human participants in laboratory experiments in the social sciences. In M. J. Webster & J. Sell (Eds.),

Laboratory experiments in the social sciences (2nd ed., pp. 103-126). Waltham, MA: Elsevier.

Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. *Trends in Cognitive Sciences, 2*, 389-398. doi:10.1016/S1364-6613(98)01230-3

Kaufman, A. S., & Kaufman, N. L. (2004a). *Kaufman test of educational achievement* (2nd ed.). San Antonio, TX: Pearson.

Kaufman, A. S., & Kaufman, N. L. (2004b). *Kaufman test of educational achievement: Comprehensive form norm book* (2nd ed.). San Antonio, TX: Pearson.

Kay, N., & Pennucci, A. (2014). *Early childhood education for low-income students: A review of the evidence and benefit-cost analysis (Doc. No. 14-01-2201)*. Olympia, WA: Washington State Institute for Public Policy.

Kearney, M. S., & Levine, P. B. (2015). *Early childhood education by MOOC: Lessons from Sesame Street*. (Working Paper 21229), Retrieved from the National Bureau of Economic Research website: http://www.nber.org/papers/w21229

Khan, O. (2015, July 15). Major email provider trends in 2015: Gmail's lead increases [Blog post], *MailChimp*. Retrieved from http://blog.mailchimp.com/major-email-provider-trends-in-2015-gmail-takes-a-really-big-lead/

Kim, R.-S., & Becker, B. J. (2010). The degree of dependence between multiple-treatment effect sizes. *Multivariate Behavioral Research, 45*, 213-238. doi:10.1080/00273171003680104

Langenberg, D. N., Correro, G., Ehri, L., Ferguson, G., Garza, N., Kamil, M. L., . . . Yatvin, J. (2000). *Report of the National Reading Panel: Teaching children to read: An evidence-*

based assessment of the scientific research literature on reading and its implications for

reading instruction: Reports of the subgroups. Bethesda, MD: National Institute of Child

Health and Human Development, National Institutes of Health.

Lauer, J. E., & Lourenco, S. F. (2016). Spatial processing in infancy predicts both spatial and

mathematical aptitude in childhood. *Psychological Science, 27*, 1291-1298.

doi:10.1177/0956797616655977

Lazorick, S., Crawford, Y., Gilbird, A., Fang, X., Burr, V., Moore, V., & Hardison Jr, G. T.

(2014). Long-term obesity prevention and the Motivating Adolescents with Technology

to CHOOSE Health™ program. *Childhood Obesity, 10*, 25-33.

doi:10.1089/chi.2013.0049

Lee, S. (2012, October 11). Big Bird debate: How much does federal funding matter in public

broadcasting?, *ProPublica*. Retrieved from http://www.propublica.org

Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*.

Cambridge, MA: Harvard University Press.

Linebarger, D. L. (2006). *The Between the Lions American Indian literacy iniative research

component: Report prepared for the United States Department of Education*.

Philadelphia, PA: Annenberg School for Communication, University of Pennsylvania.

Linebarger, D. L. (2010). *Television's impact on children's reading skills: A longitudinal study*.

Philadelphia, PA: Annenberg School for Communication, University of Pennsylvania.

Linebarger, D. L., & Barr, R. (2017, April). *Differntial susceptibility to media effects: Who is

affected by what media content under which circumstances.* Poster presented at the

Biennial Meeting of the Society for Research on Child Development, Austin, TX.

Linebarger, D. L., Kosanic, A. Z., Greenwood, C. R., & Doku, N. S. (2004). Effects of viewing the television program *Between the Lions* on the emergent literacy skills of young children. *Journal of Educational Psychology, 96*, 297-308. doi:10.1037/0022-0663.96.2.297

Linebarger, D. L., & Piotrowski, J. T. (2009). TV as storyteller: How exposure to television narratives impacts at-risk preschoolers' story knowledge and narrative skills. *British Journal of Developmental Psychology, 27*, 47-69. doi:10.1348/026151008X400445

Llorente, C., Pasnik, S., Penuel, W. R., & Martin, W. (2010). *Organizing for literacy: How public media stations are raising readers in their communities: A case study report to the Ready to Learn Initiative*. Education Development Center, Inc. & SRI International. New York and Menlo Park, CA.

Lu, A. S., Baranowski, T., Thompson, D., & Buday, R. (2012). Story immersion of videogames for youth health promotion: A review of literature. *Games for Health Journal, 1*, 199-204. doi:10.1089/g4h.2011.0012

Mares, M.-L., & Pan, Z. (2013). Effects of Sesame Street: A meta-analysis of children's learning in 15 countries. *Journal of Applied Developmental Psychology, 34*, 140-151. doi:10.1016/j.appdev.2013.01.001

Masson, H., Balfe, M., Hackett, S., & Phillips, J. (2013). Lost without a trace? Social networking and social research with a hard-to-reach population. *British Journal of Social Work, 43*, 24-40. doi:10.1093/bjsw/bcr168

McCarthy, B., Atienza, S., Kao, Y., Tiu, M., Banes, L., & Hubbard, A. (2012). *Alpha testing of PBS KIDS transmedia suites: A report to the CPB-PBS Ready to Learn initiative*. Los Alamitos, CA: WestEd.

McManis, L. D., & Gunnewig, S. B. (2012). Finding the education in educational technology with early learners. *YC Young Children, 67*, 14-24.

Michael Cohen Group. (2009). *The effects of WordWorld viewing on preschool children's acquisition of pre-literacy skills and emergent literacy: A cluster-randomized controlled trial*. New York, NY: Michael Cohen Group.

Michael Cohen Group. (2011). *Your children, apps, & iPad*. New York, NY: Michael Cohen Group, LLC.

Michael Cohen Group. (2012). *Learning from Ready To Learn*. New York, NY: Michael Cohen Group, LLC.

Michael Cohen Group. (2013). *Effects of interaction with Pocoyo playsets on preschool (Spanish) ELL children's English Language Learning: A randomized controlled trial*. New York, NY: Michael Cohen Group, LLC.

Mol, S. E., Bus, A. G., de Jong, M. T., & Smeets, D. J. H. (2008). Added value of dialogic parent–child book readings: A meta-analysis. *Early Education and Development, 19*, 7-26. doi:10.1080/10409280701838603

Morris, S. B. (2007). Estimating effect sizes from the pretest-posttest-control group designs. *Organizational Research Methods*. doi:10.1177/1094428106291059

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7*, 105-125. doi:10.1037/1082-989X.7.1.105

Moses, A. M., Linebarger, D. L., Wainwright, D. K., & Brod, R. (2010). *Television's impact on children's vocabulary knowledge: A meta-analysis*. Philadelphia, PA: Annenberg School for Communication, University of Pennsylvania.

Nathanson, A. I., & Rasmussen, E. E. (2011). TV Viewing compared to book reading and toy playing reduces responsive maternal communication with toddlers and preschoolers. *Human Communication Research, 37*, 465-487. doi:10.1111/j.1468-2958.2011.01413.x

National Conference of State Legislatures. (2016). Literacy & No Child Left Behind (NCLB). Retrieved from http://www.ncsl.org/research/education/literacy-no-child-left-behind.aspx

National Head Start Association. (2017). *Top issues*. Retrieved from http://www.nhsa.org/knowledge-center/center-advocacy/top-issues

Neuman, S. B., Newman, E. H., & Dwyer, J. (2011). Educational effects of a vocabulary intervention on preschoolers' word knowledge and conceptual development: A cluster-randomized trial. *Reading Research Quarterly, 46*, 249-272. doi:10.1598/RRQ.46.3.3

Nicholson, J., Sanson, A., Rempel, L., Smart, D., & Patton, G. (2002). Longitudinal studies of children and youth: Implications for future studies. In A. Sanson (Ed.), *Children's health and development: New research directions for Australia* (pp. 38-62). Melbourne, Australia: Australian Institute of Family Studies.

Pasnik, S., Llorente, C., Hupert, N., Dominguez, X., & Silander, M. (2015). *Supporting parent-child experiences with Peg+Cat early math concepts: Report to the CPB-PBS Ready To*

Learn initiative. Education Development Center, Inc. & SRI International. New York and Menlo Park, CA.

Pasnik, S., Llorente, C., Hupert, N., & Moorthy, S. (2016). *Reflections on the Ready To Learn Initiative, 2010 to 2015: How a federal program in parternship with public media supported young children's equitable learning during a time of great change*. Education Development Center, Inc. & SRI International. New York and Menlo Park, CA.

Pasnik, S., Penuel, W. R., Llorente, C., Strother, S., & Schindel, J. (2007). *Review of research on media and young children's literacy: Report to the Ready to Learn Initiative*. Education Development Center, Inc. & SRI International. New York and Menlo Park, CA.

Passetti, L. L., Godley, S. H., Scott, C. K., & Siekmann, M. (2000). A low-cost follow-up resource: Using the World Wide Web to maximize client location efforts. *American Journal of Evaluation, 21*, 195-203. doi:10.1177/109821400002100206

PBS. (2011). Ready To Learn. Retrieved from http://pbskids.org/readytolearn/

Penuel, W. R., Bates, L., Gallagher, L. P., Pasnik, S., Llorente, C., Townsend, E., . . . VanderBorght, M. (2012). Supplementing literacy instruction with a media-rich intervention: Results of a randomized controlled trial. *Early Childhood Research Quarterly, 27*, 115-127. doi:10.1016/j.ecresq.2011.07.002

Piotrowski, J. T., Jennings, N. A., & Linebarger, D. L. (2012). Extending the lessons of educational television with young American children. *Journal of Children and Media, 7*, 216-234. doi:10.1080/17482798.2012.693053

Piotrowski, J. T., & Valkenburg, P. M. (2015). Finding orchids in a field of dandelions: Understanding children's differential susceptibility to media effects. *American Behavioral Scientist, 59*, 1776-1789. doi:10.1177/0002764215596552

Plowman, L., McPake, J., & Stephen, C. (2008). Just picking it up? Young children learning with technology at home. *Cambridge Journal of Education, 38*, 303-319. doi:10.1080/03057640802287564

Poehlmann-Tynan, J., Gerstein, E. D., Burnson, C., Weymouth, L., Bolt, D. M., Maleck, S., & Schwichtenberg, A. J. (2015). Risk and resilience in preterm children at age 6. *Development and Psychopathology, 27*, 843-858. doi:10.1017/S095457941400087X

Ponciano, L., & Thai, K. P. (2016, October). *Reducing the risk of failure in kindergarten with ABCMouse.* Poster presented at the Special Topics Meeting: Technology and Media in Children's Development, Irvine, CA.

Purcell, K. (2011). *Search and email still top the list of most popular online activities: Two activities nearly universal among adult internet users*. Washington, D.C.: Pew Research Center's Internet & American Life Project.

R Core Team. (2016). R: A language and environment for statistical computing [Computer software]. Retrieved from http://www.r-project.org

Reid, K., Hresko, W. P., & Hammill, D. D. (2001). *Test of Early Reading Ability (TERA-3)* (3rd ed.). Austin, TX: Pro-Ed.

Reitsma, P., & Wesseling, R. (1998). Effects of computer-assisted training of blending skills in kindergartners. *Scientific Studies of Reading, 2*, 301-320. doi:10.1207/s1532799xssr0204_1

Revelle, W. (2016). psych: Procedures for personality and psychological research. Evanston, IL: Northwestern University. Retrieved from http://CRAN.R-project.org/package=psych

Ribisl, K. M., Walton, M. A., Mowbray, C. T., Luke, D. A., Davidson, W. S., & Bootsmiller, B. J. (1996). Minimizing participant attrition in panel studies through the use of effective retention and tracking strategies: Review and recommendations. *Evaluation and Program Planning, 19*, 1-25. doi:10.1016/0149-7189(95)00037-2

Rice, M. L., Huston, A. C., & Wright, J. C. (1982). The forms and codes of television: Effects of children's attention, comprehension, and social behavior. In D. Pearl, L. Bouthilet & J. Lazar (Eds.), *Television and behavior: Ten years of scientific progress and implications for the eighties* (pp. 24-38). Washington, DC: Government Printing Office.

Rideout, V., & Saphir, M. (2013). *Zero to eight: Children's media use in America 2013*. San Francisco, CA: Common Sense Media.

Roberts, J. D., Chung, G. K. W. K., & Parks, C. B. (2016). Supporting children's progress through the PBS KIDS learning analytics platform. *Journal of Children and Media, 10*, 257-266. doi:10.1080/17482798.2016.1140489

Ronimus, M., & Lyytinen, H. (2015). Is school a better environment than home for digital game-based learning? The case of GraphoGame. *Human Technology: An Interdisciplinary Journal on Humans in ICT Environments, 11*, 123-147. doi:10.17011/ht/urn.201511113637

Roseberry, S., Hirsh-Pasek, K., Parish-Morris, J., & Golinkoff, R. M. (2009). Live action: Can young children learn verbs from video? *Child Development, 80*, 1360-1375. doi:10.1111/j.1467-8624.2009.01338.x

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*, 638-641. doi:10.1037/0033-2909.86.3.638

Rosnow, R. L., & Rosenthal, R. (2009). Effect sizes: Why, when, and how to use them. *Zeitschrift für Psychologie / Journal of Psychology, 217*, 6-14. doi:10.1027/0044-3409.217.1.6

Rosser, J. C. J., Lynch, P. J., Cuddihy, L., Gentile, D. A., Klonsky, J., & Merrell, R. (2007). The impact of video games on training surgeons in the 21st century. *Archives of Surgery, 142*, 181-186. doi:10.1001/archsurg.142.2.181

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2003). *Evaluation: A systematic approach* (7th ed.). Thousand Oaks, CA: Sage Publications.

Sanchez, R. P., & Lorch, E. P. (1999). Comprehension of televised stories by preschool children with ADHD. *Journal of Clinical Child Psychology, 28*, 376.

Savage, R. S., Abrami, P. C., Piquette, N., Wood, E., Deleveaux, G., Sanghera-Sidhu, S., & Burgos, G. (2013). A (Pan-Canadian) cluster randomized control effectiveness trial of the ABRACADABRA web-based literacy program. *Journal of Educational Psychology, 105*, 310-328. doi:10.1037/a0031025

Scarborough, H. S., Ehri, L. C., Olson, R. K., & Fowler, A. E. (1998). The fate of phonemic awareness beyond the elementary school years. *Scientific Studies of Reading, 2*, 115-142. doi:10.1207/s1532799xssr0202_2

Scheller, A. (2014). *KTEA-3: Overview: Part 1*. Retrieved from http://downloads.pearsonclinical.com/videos/KTEA3-Overview-021414/KTEA3OverviewPart1WebianrHandout-02142014.pdf

Schiefele, U., Schaffner, E., Möller, J., & Wigfield, A. (2012). Dimensions of reading motivation and their relation to reading behavior and competence. *Reading Research Quarterly, 47*, 427-463. doi:10.1002/RRQ.030

Schmitt, K. L., & Linebarger, D. L. (2009). *Summative evaluation of PBS KIDS Island: Home study*. KLMedia Research & Children's Media Lab, University of Pennsylvania. Chicago, IL.

Schmitt, K. L., Sheridan Duel, L., & Linebarger, D. L. (2017, April). *Can young children learn pre-literacy skills from online educational games?* Poster presented at the Biennial Meeting of the Society for Research on Child Development, Austin, TX.

Schwartz, C. E., Wright, C. I., Shin, L. M., Kagan, J., & Rauch, S. L. (2003). Inhibited and uninhibited infants "grown up": Adult amygdalar response to novelty. *Science, 300*, 1952.

Segers, E., & Verhoeven, L. (2005). Long-term effects of computer training of phonological awareness in kindergarten. *Journal of Computer Assisted Learning, 21*, 17-27. doi:10.1111/j.1365-2729.2005.00107.x

Silva, P. A. (1990). The Dunedin Multidisciplinary Health and Development Study: A 15 year longitudinal study. *Paediatric and Perinatal Epidemiology, 4*, 76-107. doi:10.1111/j.1365-3016.1990.tb00621.x

Singer, J. L., & Singer, D. G. (1998). *Barney & Friends* as education and entertainment: Evaluating the quality and effectiveness of a television series for preschool children. In J. K. Asamen & G. L. Berry (Eds.), *Research paradigms, television, and social behavior* (pp. 305-367). Thousand Oaks, CA: SAGE Publications.

Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher, 15*, 5-11. doi:10.3102/0013189X015009005

Stanton, W. R., & Silva, P. A. (1992). A longitudinal study of the influence of parents and friends on children's initiation of smoking. *Journal of Applied Developmental Psychology, 13*, 423-434. doi:10.1016/0193-3973(92)90010-F

Stock, J. J., & Watson, M. W. (2006). *Introduction to econometrics* (2nd ed.). Boston, MA: Addison Wesley.

Striano, T. (2016). Recruitment and access *Doing developmental research: A practical guide* (pp. 92-109). New York, NY: Guilford Press.

Strouse, G. A., O'Doherty, K., & Troseth, G. L. (2013). Effective coviewing: Preschoolers' learning from video after a dialogic questioning intervention. *Developmental Psychology, 49*, 2368-2382. doi:10.1037/a0032463

Sugden, N. A., & Moulson, M. C. (2015). Recruitment strategies should not be randomly selected: Empirically improving recruitment success and diversity in developmental psychology research. *Frontiers in Psychology, 6*. doi:10.3389/fpsyg.2015.00523

Takeuchi, L., & Stevens, R. (2011). *The new coviewing: Designing for learning through joint media engagement*. Retrieved from http://www.joanganzcooneycenter.org/wp-content/uploads/2011/12/jgc_coviewing_desktop.pdf

Torgerson, C. J. (2007). The quality of systematic reviews of effectiveness in literacy learning in English: a 'tertiary' review. *Journal of Research in Reading, 30*, 287-315. doi:10.1111/j.1467-9817.2006.00318.x

Tourangeau, K., Le, T., & Nord, C. (2005). *Early Childhood Longitudinal Study, Kindergarten class of 1998-99 (ECLS-K): Fifth-grade methodology report (NCES 2006-037).* Washington, DC: National Center for Education Statistics, U.S. Department of Education.

United States Census Bureau. (2016). Poverty thresholds by size of family and number of children. Retrieved from http://www.census.gov/data/tables/time-series/demo/income-poverty/historical-poverty-thresholds.html

Valkenburg, P. M., & Peter, J. (2013). The differential susceptibility to media effects model. *Journal of Communication, 63*, 221-243. doi:10.1111/jcom.12024

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software, 45*, 1-67. URL: http://www.jstatsoft.org/v45/i03/.

van Houwelingen, H. C., Arends, L. R., & Stijnen, T. (2002). Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine, 21*, 589-624. doi:10.1002/sim.1040

Viechtbauer, W. (2007). Accounting for heterogeneity via random-effects models and moderator analyses in meta-analysis. *Zeitschrift für Psychologie/Journal of Psychology, 215*, 104-121. doi:10.1027/0044-3409.215.2.104

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor Package. *Journal of Statistical Software, 36*, 1-48. URL: http://www.jstatsoft.org/v36/i03/.

Viechtbauer, W. (2015). Comparing estimates of independent meta-analyses or subgroups. Retrieved from http://www.metafor-project.org/doku.php/tips:comp_two_independent_estimates

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. R., . . . Garon, T. (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. doi:10.1037/0012-1649.33.3.468

Walker, H. (2011). Evaluating the effectiveness of apps for mobile devices. *Journal of Special Education Technology, 26*, 59-63.

Wartella, E. (2015). Educational apps: What we do and do not know. *Psychological Science in the Public Interest, 16*, 1-2. doi:10.1177/1529100615578662

Wartella, E., Beaudoin-Ryan, L., Blackwell, C. K., Cingel, D. P., Hurwitz, L. B., & Lauricella, A. R. (2016). What kind of adults will our children become? The impact of growing up in a media-saturated world. *Journal of Children and Media, 10*, 13-20. doi:10.1080/17482798.2015.1124796

Wartella, E., Blackwell, C. K., Lauricella, A. R., & Robb, M. B. (2013). *Technology in the lives of educators and early childhood programs: 2012 survey of early childhood educators*. Latrobe, PA: Fred Rogers Center at Saint Vincent's College.

Webster, J. G. (2014). *The marketplace of attention: How audiences take shape in a digital age*. Cambridge, MA: MIT Press.

What Works Clearinghouse. (2014). *Procedures and standards handbook* (3rd ed.). Washington, D.C.: What Works Clearinghouse, Institute of Education Sciences, DoEd.

Whitehurst, G. J., & Lonigan, C. J. (2001). *Get ready to read*. Columbus, OH: Pearson Early Learning.

Wilson, V. (2015, September 16). New Census data shows no progress in closing stubborn racial income gaps [Blog post], *Economics Policy Institute*. Retrieved from http://www.epi.org/blog/new-census-data-show-no-progress-in-closing-stubborn-racial-income-gaps/

Wright, J. C., Huston, A. C., Murphy, K. C., St. Peters, M., Piñon, M., Scantlin, R., & Kotler, J. (2001). The relations of early television viewing to school readiness and vocabulary of children from low-income families: The early window project. *Child Development, 72*, 1347-1366. doi:10.1111/1467-8624.t01-1-00352

Wu, F., & Qi, S. (2006). Longitudinal effects of parenting on children's academic achievement in African American families. *Journal of Negro Education, 75*, 415-429.

Appendix A:

Academic Search Terms

- Ready To Learn
- RTL
- Corporation for Public Broadcasting
- CPB
- Public Broadcasting Service
- PBS
- *Arthur*
- *Barney and Friends*
- *Between the Lions*
- *Cat in the Hat Knows a Lot About That*
- *Clifford the Big Red Dog*
- *Curious George*
- *Dinosaur Train*

- *Dragon Tales*
- *Duck's Alphabet*
- *Fizzy's Lunch Lab*
- *Great Word Quest*
- *Martha Speaks*
- *Miss Spider's Sunny Patch Friends*
- *Mission to Planet 429*
- *Nova the Robot*
- *Odd Squad*
- *Peg + Cat*
- *Pocoyo*
- *Postcards from Buster*
- *R U There*
- *RU There*
- *Reading Rainbow*
- *Ruff Ruffman*

- *Sesame Street*
- *Sid the Science Kid*
- *Super WHY*
- *SuperWHY*
- *The Electric Company*
- *UMIGO*
- *Word Girl*
- *WordGirl*
- *Word World*
- *WordWorld*
- World of Words
- Any time is learning time
- Anytime is learning time
- Learning Triangle
- Raising Readers

Appendix B:

Mean Effect Sizes and Key Codable Characteristics of Included Evaluations

Evaluation	*d*	Texts Available	Literacy Categories Tested	Sample Characteristics	Study Design	Scope	Setting	Weeks of Media	Nature of RTL Treatment	Nature of Control	Assessments	Grant Cycle	RTL Funding Status
Chiong and Shuler (2010); Rockman et al. (2010)	0.31	Unpublished reports	Alphabet knowledge; Phonological processing; Vocabulary; Narrative comprehension	64% Caucasian; 16% African American; 6% Hispanic; 3% Asian; 44% Female; 72% School-age; 39% Low-income	Within	Multiple regions	Home	2	*Martha Speaks* and *Super WHY!* apps without extension activities	N/A	Mix of custom broad and standardized	2005-2010	Funded
Dwyer (2010); Neuman and Dwyer (2011)	0.57	Journal article; Dissertation	Vocabulary	58% Caucasian; 27% African American; 0% School-age	Within and Between	One region	School	8	*Sesame Street* TV with extension activities	Business as usual	Mix of custom narrow, custom broad, and standardized	2005-2010	Funded
Educational Development Center and SRI International (2008); Penuel, Clements, Pasnik, and Llorente (2008)	0.01	Unpublished reports	Alphabet knowledge; Print concepts; Phonological processing	93% African American; 49% Female; 0% School-age	Between	One region	Home	8	*Between the Lions* and *Sesame Street* TV with extension activities	Non-RTL Media	Mix of custom broad and standardized	2005-2010	Funded
Ferrell (2002)	-0.2	Thesis	Alphabet knowledge; Other	56% Caucasian; 15% African American; 6% Asian; 2% Native American; 50% Female; 100% School-age;	Between	Multiple regions	Home		*Sesame Street* TV	Non-RTL Media and Alternate curriculum	All custom broad	1994-2000	Not Funded

Study	ES	Publication type	Skills	Demographics	Design	Region	Setting	N	Program	Comparison	Measures	Years	Funding
Garrity, Piotrowski, McMenamin, and Linebarger (2010)	0.14	Unpublished report	Phonological processing; Vocabulary	18% Caucasian; 71% African American; 51% Female; 100% School-age; 81% Low-income; 7% ELL	Between	One region	Home and school	10	*The Electric Company* TV, with and without computer games, with and without extension activities	Business as usual	Mix of custom narrow and standardized	2005-2010	Funded
Godfrey (2015)	0.55	Dissertation	Vocabulary; Narrative comprehension	65% Caucasian; 5% African American; 20% Asian; 57% Female; 0% School-age	Within	One region	School	6	*WordWorld* TV with and without extension activities	N/A	All standardized	2010-2015	Not Funded
Harper et al. (2006)	0.23	Unpublished report	Print concepts; Narrative comprehension; Multiple skills	0% School-age	Between	One region	School	20	*Between the Lions* TV with extension activities	Business as usual	All standardized	2005-2010	Funded
Jennings (2013)	-0.1	Unpublished report	Alphabet knowledge; Phonological processing; Other	39% Female	Between	One region	School	1	*Super WHY!* TV with extension activities	Business as usual	Mix of custom narrow, custom broad, and standardized	2010-2015	Funded
Kaefer and Neuman (2013)	0.46	Journal article	Vocabulary	36% Caucasian; 34% African American; 4% Hispanic; 11% Asian; 54% Female; 0% School-age; 77% Low-income; 100% ELL	Between		School	2	*Sesame Street* TV with extension activities	Business as usual	Mix of custom narrow and custom broad	2010-2015	Not Funded

Study	ES	Publication	Skills	Demographics	Design	Region	Setting	N	Program	Comparison	Measures	Years	Funding
Kelly (2003)	-0.3	Dissertation	Multiple skills	60% Caucasian; 4% African American; 27% Asian; 47% Female; 100% School-age; 9% ELL	Between	One region	Home and school	19	*Between the Lions* TV with extension activities	Business as usual	All standardized	2000-2005	Not Funded
Linebarger (2000); Linebarger, Kosanic, Greenwood, and Doku (2004)	0.09	Journal article; Unpublished report	Alphabet knowledge; Phonological processing; Vocabulary; Multiple skills	81% Caucasian; 6% African American; 7% Hispanic; 1% Native American; 49% Female; 100% School-age; 34% Low-income; 7% ELL	Between	One region	School	3.5	*Between the Lions* TV without extension activities	Business as usual	Mix of custom narrow, custom broad, and standardized	1994-2000	Funded
Linebarger (2009)	0.02	Unpublished report	Alphabet knowledge; Print concepts; Phonological processing; Vocabulary; Narrative comprehension; Multiple skills	46% Female; 0% School-age	Between	One region	School	30	*Between the Lions* TV with extension activities	Business as usual	All standardized	2005-2010	Funded
Linebarger and Piotrowski (2007, 2009)	-0.1	Journal article; Unpublished report	Narrative comprehension; Multiple skill	52% Female; 0% School-age	Between	One region	School	8	*Clifford the Big Red Dog* TV without extension activities	Non-RTL Media	Mix of custom broad and standardized	2005-2010	Funded
Linebarger and Walker (2005)	0.28	Journal article	Vocabulary	90% Caucasian; 55% Female; 0% School-age; 9% Low-income	Between	One region	Home	120	*Arthur, Clifford, Barney, Dragon Tales,* and *Sesame Street* TV	Mix of custom narrow and custom broad	All standardized	2000-2005	Not Funded

Linebarger, McMenamin, and Wainwright (2009)	0.27	Journal article; Unpublished report	Alphabet knowledge; Print concepts; Phonological processing; Vocabulary; Narrative comprehension; Multiple skills	68% Caucasian; 1% African American; 2% Hispanic; 44% Female; 0% School-age; 22% Low-income	Between	One region	Home	8	*Super WHY!* TV without extension activities	Non-RTL Media	Mix of custom narrow, custom broad, and standardized	2005-2010	Funded
Linebarger, Moses, Jennings, and McMenamin (2010); Linebarger, Moses, and McMenamin (2010a, 2010b); Linebarger, Moses, Garrity Liebeskind, and McMenamin (2013); Moses, Linebarger, McMenamin, and Liss-Mariño (2009)	0.2	Journal article; Conference presentation; Unpublished reports	Vocabulary	56% Female; 46% School-age; 95% Low-income	Between	Multiple regions	Home	4	*Martha Speaks* TV without extension activities	Business as usual	All custom narrow	2005-2010	Funded
Linebarger, Piotrowski, and Vaala (2007); Piotrowski, Vaala, and Linebarger (2009)	1.06	Conference presentation; Unpublished report	Vocabulary	100% School-age; 100% ELL	Between	One region	School	12	*Postcards from Buster* TV with extension activities	Business as usual and Non-RTL Media	All custom narrow	2005-2010	Funded
Marshall, Lapp, and Cavoto (2009)	0.39	Unpublished report	Alphabet knowledge; Phonological processing; Vocabulary; Other	25% Caucasian; 55% African American; 16% Hispanic; 48% Female; 59% School-age	Within	Multiple regions	School	1	*Super WHY!* TV with extension activities	N/A	All custom narrow	2005-2010	Funded

Study	ES	Publication	Skills	Demographics	Design	Region	Setting	Duration	Intervention	Comparison	Measures	Years	Funding
McCarthy et al. (2011)	0.23	Unpublished report	Phonological processing	100% School-age	Within	Multiple regions	School	5.5	*The Electric Company* TV and computer games with extension activities	N/A	All standardized	2010-2015	Funded
Meyer and Sroka (2010)	0.26	Unpublished report	Alphabet knowledge; Phonological processing; Vocabulary; Other; Multiple skills	53% Caucasian; 26% African American; 19% Hispanic; 2% Asian; 53% Female; 14% ELL	Between	Multiple regions	School	1	*Super WHY!* TV with extension activities	Alternate curriculum	Mix of custom narrow and standardized	2005-2010	Funded
Michael Cohen Group (2009)	0.15	Unpublished report	Alphabet knowledge; Print concepts; Phonological processing; Vocabulary	32% Caucasian; 30% African American; 21% Hispanic; 6% Asian; 1% Native American; 0% School-age; 33% Low-income	Between	Multiple regions	School	6	*WordWorld* TV without extension activities	Business as usual	Mix of custon narrow and standardized	2005-2010	Funded
Michael Cohen Group (2010a)	0.33	Unpublished report	Alphabet knowledge; Phonological processing; Multiple skills	6% Caucasian; 57% African American; 23% Hispanic; 47% Female; 0% School-age; 60% Low-income	Between	Multiple regions	Home	2	*Duck's Alphabet* computer games without extension activities	Non-RTL Media	Mix of custom broad and standardized	2005-2010	Funded
Michael Cohen Group (2010b)	0.19	Unpublished report	Vocabulary; Narrative comprehension; Other	5% Caucasian; 35% African American; 57% Hispanic; 47% Female; 78% Low-	Between	Multiple regions	School	2	*Mission to Planet 429* computer games without	Non-RTL Media	Mix of custom narrow and standardized	2005-2010	Funded

Study	ES	Publication type	Skills	Demographics	Design	Region	Setting	Age	Program	Comparison	Measures	Years	Funding
				income; 100% School-age						extension activities		
Michael Cohen Group (2010c)	-0.02	Unpublished report	Vocabulary; Narrative comprehension; Other	14% Caucasian; 51% African American; 21% Hispanic; 6% Asian; 56% Female; 100% School-age; 25% Low-income	Between	Multiple regions	School	2	*R U There?* TV and computer games without extension activities	Non-RTL Media	Mix of custom narrow, custom broad, and standardized	2005-2010	Funded
Michael Cohen Group (2013)	0.12	Unpublished report	Vocabulary; Other	100% Hispanic; 52% Female; 0% School-age; 56% Low-income; 94% ELL	Between	Multiple regions	School	3	*Pocoyo* apps without extension activities	Non-RTL Media	Mix of custom narrow and standardized	2010-2015	Funded
Michael Cohen Group (2015)	0.16	Unpublished report	Vocabulary; Multiple skills	6% Caucasian; 25% African American; 59% Hispanic; 13% Asian; 55% Female; 0% School-age; 59% Low-income; 69% ELL	Between	Multiple regions	School	4	*Pocoyo* apps with extension activities	Non-RTL Media	Mix of custom narrow and standardized	2010-2015	Funded
Naigles (2000)	0.24	Book chapter	Vocabulary	85% Caucasian; 59% Female; 0% School-age; 0% ELL	Between	One region	School	2	*Barney* TV without extension activities		All standardized	1994-2000	Not Funded
Naigles et al. (1997); Singer and Singer (1998)	0.15	Book chapter; Unpublished report	Other	88% Caucasian; 50% Female; 0% School-age; 0% ELL	Between	One region	School	2.5	*Barney* TV without extension activities	Business as usual	All custom narrow	1994-2000	Funded
Neuman and Kaefer (2013)	1.12	Journal article	Vocabulary	92% Caucasian; 3% African	Within	One region	School	8	*Sesame Street* TV with	N/A	All custom narrow	2010-2015	Not Fund...

Study	ES	Publication type	Skills	Demographics	Design	Region	Level	N	Program	Comparison	Measure	Years	Funding
				American; 4% Hispanic; 52% Female; 0% School-age; 100% Low-income					extension activities				
Neuman, Newman, and Dwyer (2010, 2011)	0.46	Journal article; Unpublished report	Vocabulary	37% Caucasian; 43% African American; 3% Hispanic; 12% Asian; 52% Female; 0% School-age; 100% Low-income; 4% ELL	Between	One region	School	8	*Sesame Street* TV with extension activities	Alternate curriculum	All custom narrow	2005-2010	Funded
Penuel et al. (2012); Penuel et al. (2009)	0.32	Journal article; Unpublished report	Alphabet knowledge; Phonological processing; Print concepts	6% Caucasian; 28% African American; 53% Hispanic; 10% Asian; 3% Native American; 51% Female; 0% School-age; 68% Low-income	Between	One region	School	10	*Between the Lions, Super WHY!* and *Sesame Street* TV and computer games	Non-RTL Media	Mix of custom broad and standardized	2005-2010	Funded
Phillips (2008)	0.41	Unpublished report	Alphabet knowledge; Phonological processing; Vocabulary; Other	26% Caucasian; 44% African American; 24% Hispanic; 46% Female; 68% School-age	Within	Multiple regions	School	1	*Super WHY!* TV with extension activities	N/A	All custom narrow	2005-2010	Funded
Piotrowski, Jennings, and Linebarger (2012); Piotrowski, Linebarger, and	-0.1	Journal article; Unpublished report	Alphabet knowledge; Phonological processing; Vocabulary; Multiple skills	49% Female; 65% School-age	Between	One region	School	4	*Between the Lions* TV with and without extension activities	Business as usual	All standardized	2005-2010	Funded

Study													
Jennings (2009)													
Prince, Grace, Linebarger, Atkinson, and Huffman (2002)	0.26	Unpublished report	Print concepts; Vocabulary; Narrative comprehension; Other; Multiple skills	20% Caucasian; 48% African American in one comparison; 100% Native American in another comparison; 71% School-age; 80% Low-income	Between	One region	School	26	*Between the Lions* TV with extension activities	Business as usual	All standardized	2000-2005	Funded
Register (2003, 2004)	0.05	Journal article; Dissertation	Alphabet knowledge; Print concepts; Phonological processing; Narrative comprehension; Multiple skills	100% School-age	Between	One region	School	3.4	*Between the Lions* TV with and without extension activities	Business as usual and Alternate curriculum (in separate comparisons)	All standardized	2000-2005	Not Funded
Rollins (2000)	-0.3	Thesis	Other	67% Female; 100% School-age	Within	One region	School	3	*Reading Rainbow* TV without extension activities	N/A	All custom narrow	1994-2000	Not Funded
Schmitt and Linebarger, (2009)	0.43	Unpublished report	Alphabet knowledge; Phonological processing; Other	43% Female; 0% School-age	Within	One region	Home	4	*Between the Lions, Martha Speaks, Super WHY!, Sesame Street,* and *WordWorld* computer games	N/A	All standardized	2005-2010	Funded

									without extension activities				
Schmitt, Sheridan Duel, and Linebarger (2016, 2017); Schmitt, Sheridan, McMenamin, and Linebarger (2010)	0.17	Conference presentation; Unpublished report	Alphabet knowledge; Phonological processing; Vocabulary; Multiple skills	31% Caucasian; 18% African American; 29% Hispanic; 8% Asian; 52% Female; 31% School-age; 62% Low-income; 15% ELL	Between	One region	Home	6.8	*Between the Lions, Martha Speaks, Super WHY!, Sesame Street,* and *WordWorld* computer games with and without extension activities	Non-RTL Media	Mix of custom narrow, custom broad, and standardized	2005-2010	Funded
Silverman (2009a, 2013)	0.18	Journal article; Unpublished report	Vocabulary	33% African American; 63% Hispanic; 1% Asian; 55% Female; 100% School-age; 80% Low-income; 67% ELL	Within	One region	School	4	*Martha Speaks* TV without extension activities	N/A	All custom narrow	2005-2010	Funded
Silverman (2009b)	0.14	Unpublished report	Vocabulary	6% Caucasian; 19% African American; 61% Hispanic; 8% Asian; 49% Female; 100% School-age; 94% Low-income	Within	One region	School	4	*Martha Speaks* TV and computer games (in separate comparisons) with extension activities	N/A	All custom narrow	2005-2010	Funded
Silverman (2009c, 2013)	-0.02	Journal article; Unpublished report	Vocabulary	10% Caucasian; 24% African American;	Within	One region	School	4	*Martha Speaks* TV with and	Alternate curriculum	All custom narrow	2005-2010	Funded

				53% Hispanic; 2% Asian; 49% Female; 100% School-age; 51% Low-income; 59% ELL					without extension activities				
Silverman (2009d); Silverman and Carlis (2010)	0.27	Conference presentation; Unpublished report	Vocabulary	12% Caucasian; 63% African American; 10% Hispanic; 8% Asian; 55% Female; 100% School-age; 15% ELL	Between	Multiple regions	School	8	*Martha Speaks* TV with extension activities	Alternate curriculum	All custom narrow	2005-2010	Funded
Silverman, Kim, Hartranft, Nunn, and McNeish (2016)	0.17	Journal article	Vocabulary; Narrative comprehension	62% Caucasian; 20% African American; 50% Female; 100% School-age; 61% Low-income; 8% ELL	Between	One region	School		*Martha Speaks* TV with extension activities	Business as usual	Mix of custom narrow and standardized	2010-2015	Funded
Tollefson (2013)	0.41	Thesis	Multiple skills	92% Caucasian; 5% African American; 3% Hispanic; 45% Female; 0% School-age; 31% Low-income	Between	One region	School	30	*Between the Lions* TV with extension activities	Alternate curriculum	All standardized	2010-2015	Not Funded
Uchikoshi (2004, 2005, 2006a, 2006b)	0.25	Journal articles; Dissertation	Phonological awareness; Vocabulary; Multiple skills	100% Hispanic; 47% Female; 100% School-age; 80% Low-income	Between	One region	School	18	*Arthur* or *Between the Lions* TV without extension activities	Business as usual	All standardized	2000-2005	Not Funded

Note. This table reflects the information that was coded from each evaluation, collapsing across comparisons. It only references measures for which there was sufficient statistical information to calculate effects. For the purposes of this meta-analysis, cases in which authors compared two different versions of RTL treatment to one another without a comparison group were coded as within-subjects designs in the present dataset, with each version of the RTL treatment considered a separate comparison.