



Contents lists available at SciVerse ScienceDirect

Journal of Applied Developmental Psychology

Effects of *Sesame Street*: A meta-analysis of children's learning in 15 countries^{☆,☆☆}Marie-Louise Mares^{*}, Zhongdang Pan¹

Department of Communication Arts, University of Wisconsin-Madison, 821 University Ave., Madison, WI 53706, USA

ARTICLE INFO

Article history:

Received 21 March 2012
 Received in revised form 18 December 2012
 Accepted 4 January 2013
 Available online xxxx

Keywords:

Early education
 Educational television
Sesame Street
 Meta-analysis
 International

ABSTRACT

Sesame Street is broadcast to millions of children globally, including in some of the world's poorest regions. This meta-analysis examines the effects of children's exposure to international co-productions of *Sesame Street*, synthesizing the results of 24 studies, conducted with over 10,000 children in 15 countries. The results indicated significant positive effects of exposure to the program, aggregated across learning outcomes, and within each of the three outcome categories: cognitive outcomes, including literacy and numeracy; learning about the world, including health and safety knowledge; social reasoning and attitudes toward out-groups. The effects were significant across different methods, and they were observed in both low- and middle-income countries and also in high-income countries. The results are contextualized by considering the effects and reach of the program, relative to other early childhood interventions.

© 2013 Elsevier Inc. All rights reserved.

As the most enduring exemplar of children's educational television programming, *Sesame Street* has been broadcast around the world. It is currently aired in over 130 countries, in many instances as an international co-production based on the original US template, tailored to local conditions. As such, the program in all its versions represents a major early education intervention. This meta-analysis is the first systematic evaluation of the effectiveness of this global intervention, based on summative and other studies of the program's effects, conducted in 15 countries. It examines the extent to which children outside the US may learn from viewing *Sesame Street* in diverse social, political, and economic circumstances – including in some of the world's poorest regions.

Benefits and challenges of early education

A recent report from UNESCO (2011) indicated that the United Nations' goal of universal primary school enrollment by 2015 is almost certain to fall well short. Moreover, the report reiterates findings that among children who attend school in low-income regions of the world, sizable proportions do not finish even primary school and/

or fail to reach minimal standards of literacy and numeracy (see Glewwe & Kremer, 2006 for an extensive review; UNESCO, 2011).

There is a growing consensus among researchers that failures of formal primary education in low-income areas are partly attributable to conditions that exist in infancy and early childhood (Baker-Henningham & Boo, 2010; Walker, 2011). Naudeau, Martinez, Premand, and Filmer (2011) reported on longitudinal findings in Cambodia, Ecuador, Nicaragua, Madagascar, and Mozambique. They concluded that cognitive delays began in the first few years of life and were strongly linked to the child's socioeconomic background, in part through differential access to cognitive stimulation and nutrition. Other studies show that children who start primary school at a developmental disadvantage are the most likely to repeat grades, to drop out of school altogether, and to engage in risky behaviors (see Grantham-McGregor et al., 2007, for a review).

These findings are consistent with Heckman and Masterov's (2007) widely shared view that investing in early education for children from disadvantaged backgrounds should be more cost-effective than later interventions because of the cumulative nature of learning and cognitive development (see also Cunha, Heckman, & Schennach, 2010). Various international teams of scholars from across disciplines have concluded that initial cognitive delays translate into life-long inequities of achievement and opportunity and foster the intergenerational transmission of poverty, thereby reducing the efficient use of public spending on primary or secondary education (Engle et al., 2007; Grantham-McGregor et al., 2007; Naudeau et al., 2011; Walker et al., 2007, 2011).

Despite these conclusions, recent reports indicate that education for children under age 5 continues to be of low priority to many governments and most private donors (UNESCO, 2006, 2011). Widespread

[☆] Sesame Workshop commissioned this study and provided full access to the proprietary reports.

^{☆☆} Preparation of this manuscript was supported by a fellowship for the first author at the Netherlands Institute for Advanced Study in the Humanities and Social Sciences (NIAS).

^{*} Corresponding author. Tel.: +1 608 263 2350; fax: +1 608 262 9953.

E-mail addresses: mares@wisc.edu (M.-L. Mares), zhongdangpan@wisc.edu (Z. Pan).

¹ Tel.: +1 608 263 3920; fax: +1 608 262 9953.

lack of funding means that early child-care and education in many low- and middle-income countries are provided by caregivers with little or no training, and low levels of education (UNESCO, 2006; Walker et al., 2011). In fact, Walker et al. (2007, 2011) argue that inadequate cognitive stimulation is one of the most urgent modifiable risk factors encountered by young children in low-income regions. In a recent interview, the Chief of UNICEF's Early Childhood Development Unit, Dr. Ulkuer, argued that there is clear need for effective, scalable, and affordable interventions to provide enrichment to children in low-income regions, before developmental delays accumulate (UNICEF Innocenti, 2011).

Television as a possible source of informal early education

It is against this backdrop that television programming can be considered as a source of informal education and enrichment. In some populations, programming supplements many other sources of early learning; in some regions (especially in the poorest areas of the world) it may be one of the few sources of cognitive stimulation available to provide systematic curricula. For example, in Bangladesh, preschool enrollment is estimated at roughly 10% of the relevant age group (UNESCO Institute of Statistics, 2011a, 2011b, 2011c, 2011d, 2011e), yet almost half of a national sample of children watched television daily in 2007 (Khan, Chakaborty, Rahman, & Nasrin, 2007). Of those, 83% of children in urban locations and 58% of children in rural locations watched *Sisimpur*, the local version of *Sesame Street*. The key question is whether and what children learn from viewing.

To help address this question, we first briefly review prior research on educational television, outlining key findings and controversies. Next, we outline the scope and focus of the meta-analysis and then consider other, non-media early childhood interventions as a context for evaluating the findings.

Research on educational TV in the US

A comprehensive review of the educational media literature is beyond the scope of this paper. Readers are referred to a number of detailed discussions, both of educational content in general and of *Sesame Street* in particular (Anderson, Lavigne, & Hanson, 2013; Fisch & Truglio, 2001; Mares, 2009). There are, however, three points from the extant literature that are relevant to the current project.

The first point is that a growing body of research suggests negligible benefits or even some negative effects of viewing television programs or DVDs for children less than 2 years old (Krcmar, Grela, & Lin, 2007; Linebarger & Walker, 2005; Robb, Richert, & Wartella, 2009). Of these, the most relevant is Linebarger and Walker's (2005) finding that early exposure to *Sesame Street* was associated with slightly lower expressive vocabulary scores in a longitudinal study of 51 children followed from 6- to 30-months of age. It remains unclear whether these relationships indicate effects of exposure, and whether they vary by parental education and the availability of other, non-media sources of stimulation.

The second point is that *Sesame Street* faced early criticisms that exposure to the program potentially contributed to socioeconomic achievement gaps in early learning. In their re-analysis of the initial *Sesame Street* evaluation data, Cook et al. (1975) concluded that children from less affluent research sites watched the program less often (even when encouraged to view) and showed marginally smaller improvements in literacy and numeracy over six months than those from more affluent sites. They noted that the results of another study (Minton, 1975) suggested small effects of exposure to *Sesame Street* among socioeconomically advantaged children and negligible effects for disadvantaged children. Cook et al. (1975) ended by questioning whether the money and effort expended on *Sesame Street* might more appropriately be focused on preschool or other interventions, targeted exclusively at children from low socioeconomic backgrounds who would be at most risk of remaining illiterate.

Since then, no longitudinal studies have systematically examined the varying effects of *Sesame Street* or other educational programs on early education achievement across different socioeconomic strata. However, subsequent studies indicated that exposure to educational TV did not vary substantially by socio-economic status or by other family stressors, and that there were positive relationships between educational TV viewing and children's educational outcomes, including in low-income populations (Vandewater & Bickham, 2004; Zill, Davies, & Daly, 1994).

The final point is that longitudinal research in the US has found significant positive effects of exposure to educational programming, including *Sesame Street*. Wright et al. (2001) focused their study on children in low- and middle-income households in Kansas. They found that viewing of educational programs (primarily *Sesame Street*) at ages 2 and 3 was positively associated with scores on reading, math, receptive vocabulary and school readiness, both concurrently and later when children reached 5 years of age. In contrast, viewing educational programs was not associated with achievement among children aged 4–7 at the onset of the study. Similar age patterns were observed for exposure to *Sesame Street* and children's subsequent vocabulary scores in another longitudinal study that did not specifically target lower-income households (Rice, Huston, Truglio, & Wright, 1990).

Further, early educational viewing may initiate trajectories that endure long past the initial effects of learning letters and numbers. Anderson, Huston, Schmitt, Linebarger, and Wright (2001) recontacted 570 adolescents whose childhood viewing patterns had been measured approximately ten years earlier (some of them from Rice et al., 1990). They found positive relationships between amount of viewing of *Sesame Street* at age 5 and subsequent high school science grades, time spent reading books for leisure, and attitudes toward achievement. For boys (but not girls), there were additional positive associations for English grades, overall GPA, and level of difficulty of math classes taken in high school. Viewing of other educational programs at age 5 did not predict high school grades, but early viewing of *Mister Rogers' Neighborhood* (a program emphasizing imagination) predicted adolescent creativity scores.

Taken together, these studies suggest that early exposure to educational programming can have positive effects. What the research thus far does not indicate is whether there are similar effects outside the US.

The current project

The present meta-analysis is intended to address this gap in the literature. We focus specifically on *Sesame Street* due to its unique longevity, global reach, adaptation to different national contexts, and reliance on evaluative research. This project is an analysis of the proprietary summative research conducted on *Sesame* international co-productions, together with the few other independent non-US studies that had been conducted over the years.

International co-productions

Although many countries simply air *Sesame Street* dubbed into the local language, there are numerous versions of the program that are created specifically for a particular country, co-produced with companies in that country. In 2011, there were 39 different international co-productions of *Sesame Street*. As described by Cole, Richman, and McCann Brown (2001), all co-productions share the program's immediately identifiable style and target age group, and all overlap in the core learning goals for that age group. However, the specific educational goals of each country's co-production are developed by early childhood specialists from that particular country in conjunction with the local production team. Sets and characters are developed specifically for that production; live-action videos and animations

are used to present educational content in contexts that are familiar and relevant to the local audience. At least half the material is produced locally; the remainder comes from Sesame Workshop's library of material.

The primary studies

In addition to the above characteristics, international co-productions are shaped by the Children's Television Workshop (CTW) model (Mielke, 1990; Truglio & Fisch, 2001), which emphasizes the integration of formative research into the production process and the use of summative research to examine the educational impact of the program once it is broadcast. That summative, proprietary research forms the core of the current study, together with a few other studies located through on-line databases. The proprietary studies were carried out by research companies within each country or by academics, under contract to Sesame Workshop and in consultation with Sesame Workshop personnel.

Obviously, it would clearly be preferable if more of the studies had been conducted by wholly independent researchers. However, the fact that most of the studies are internal reports means that we have a chance to add publicly unavailable findings to an area where systematic empirical evidence is much needed. One example illustrates this point. In their recent review, Engle et al. (2011) noted that educational television programming "might be a viable option for improving early child development" in low-income regions (p. 1343), but only two studies could be cited for this claim. Calling attention to the general dearth of empirical evidence in this area, Walker et al. (2011) noted the continued need for research in low-income regions on the effects of interventions on children's socio-emotional development, as well as cognitive development.

In this context, the current meta-analysis assesses a novel and relevant data set, gathered with thousands of children, many from low- and middle-income countries. Sesame Workshop asked us to conduct the meta-analysis. The agreement was that we would receive all the reports that were available, regardless of their quality or findings, and that there would be no restrictions on our analyses or publication of our findings.

Key questions

The chief question that we hoped to answer with the data was whether there were significant positive effects of exposure to *Sesame Street* in non-US contexts.

The second question was whether the program would be more effective at teaching some types of content than others. The international co-productions covered a wide array of topics, and the studies assessing those productions included a wide range of measures. We organized those outcomes into three broad categories (cognitive outcomes, learning about the world, and social reasoning and attitudes) to allow for comparison with prior literatures and meta-analytic findings.

Learning outcome categories. The first conceptual category encompassed measures of traditional preschool content such as knowledge of letters, numbers, colors, shapes, and relationships of size and distance. Such outcomes have previously been grouped together as "cognitive outcomes" (Nores & Barnett, 2010).

The second category, labeled "learning about the world," encompassed measures of learning about the physical and social environment, including knowledge of natural features, health/hygiene, and national/local culture. With regard to health, the international co-productions have focused not only on providing general health knowledge (e.g., it is good to brush teeth, wash hands, see a doctor when you're hurt) but also on teaching basic information related to local health challenges, such as HIV/AIDS in South Africa or malaria and bike safety in Indonesia.

The third category, labeled "social reasoning and attitudes" encompassed measures related to moral reasoning and attitudes toward social out-groups. *Sesame Street* characters are intended to model prosocial interactions and scripts are often written with the intent of fostering positive attitudes toward social or political out-groups. In some instances, the program touches upon deep-seated conflicts. For example, in Kosovo, one goal of *Rruga Sesam* and *Ulica Sezam* was to foster positive feelings between Serbs and Albanians; in Northern Ireland, *Sesame Tree* emphasized positive images of Catholic and Protestant traditions; *Rechov Sumsum* and *Shara'a Simsim* focused in part on messages of mutual respect between Israelis and Palestinians. In other productions, the emphasis has been on inclusiveness and respect toward those with disabilities (Bangladesh) or HIV/AIDS (Nigeria).

Other moderators. The relatively small number of studies limited the capacity to examine other moderators of effect size besides learning outcome. One key issue was whether the results were robust across studies with different research designs and other methodological characteristics. We also examined whether effects were significant and positive in low- and middle-income countries as well as in high-income countries, and whether they were significant and positive even for studies with targeted sampling of children from low-income populations. We also conducted supplementary analyses on the few available effect size estimates from demographic sub-samples defined by the child's sex, age, and region of residence (urban vs. rural).

Context: What could be expected as a result of exposure?

Hill, Bloom, Black, and Lipsey (2008) note that effect sizes generally tend to be small in education research and advise against the use of Cohen's (1988) broad guidelines. Instead, they recommend comparing effect sizes for a particular educational intervention against normative expectations for developmental change or observed effect sizes for other interventions. Given the challenges of establishing developmental norms for different countries and contexts (Naudeau et al., 2011), the second approach is more relevant here. Below, we describe the prior research on cognitive outcomes, and note the dearth of comparisons for the two other outcomes considered in the current project.

In a recent meta-analysis on early childhood interventions in non-US contexts, Nores and Barnett (2010) examined 56 experimental or quasi-experimental studies of 30 interventions in 23 countries in Europe, Africa, Asia, Central and South America. Those labeled educational interventions contained at least some component designed to improve the quality and amount of cognitive stimulation available to children, whether through training of childcare providers or enrollment in preschool. Roughly a third of the educational interventions were also combined with nutritional supplements. The authors reported that children's cognitive outcomes (measures of IQ, vocabulary, literacy, and math) were most affected by educational/combined interventions (Cohen's $d = 0.35$) as opposed to nutritional supplements only ($d = 0.25$) or cash transfers to parents ($d = 0.17$).

Nores and Barnett (2010) noted that the overall effect size across intervention types was significantly smaller in low- and middle-income countries than in high-income countries, though the difference occurred primarily for health rather than for cognitive outcomes. Because the interventions varied from country to country, it is hard to untangle from Nores and Barnett's (2010) analyses whether the country's income level or the type of intervention (or both) resulted in differential effectiveness. The current project, with its examination of *Sesame Street*, offers the opportunity to explore whether the effects of a consistent (though locally targeted) intervention vary substantially by the country's income level.

For the other two outcomes considered in the current project, there is little information about the effectiveness of interventions among preschool-aged children. With regard to health, safety, or cultural learning, there are few meta-analyses that present effects for such young ages. [Rispen, Aleman, and Goudena \(1997\)](#) reported that children aged five or younger showed substantial learning from preschool or kindergarten instruction how to avoid sexual abuse ($d = .97$ based on 17 effect size estimates from 10 studies).

With regard to social reasoning and attitudes toward out-groups, [Durlak, Weissberg, Dymnicki, Taylor, and Schellinger \(2011\)](#) examined the effectiveness of 213 school-based programs intended to promote social and emotional learning. The authors reported significant, positive effects on socio-emotional outcomes, but the study did not indicate the effect size for children aged five or younger, even though kindergarten interventions were in the scope of the study. Other studies suggest that it is often difficult to change young children's beliefs and attitudes about social groups, particularly for those groupings that are visually salient, such as those based on disability or racial differences ([Bigler, 1999](#); [Elliott & Byrd, 1983](#); [Persson & Musher-Eizenman, 2003](#)). Given this, we expected the effects of exposure on children's social attitudes to be heterogeneous and, on average, relatively small.

There is yet another way of contextualizing the effects of *Sesame Street*, beyond making comparisons with the effect sizes observed for other interventions. That is, it is worth considering how many children are reached and what other options are available to children in that context. Thus in the final section of the results, we present data on the reach of the program within the various countries included in the meta-analysis and consider how it compares with national enrollment in pre-primary education.

Methods

Sample

Inclusion/exclusion criteria

To be included in the meta-analysis, a study had to assess the effects of exposure to *Sesame Street* (or a co-production in a particular country, *Sesame Street* henceforth) on at least one of the key learning outcomes and the report had to provide the statistical information needed to estimate at least one such effect. The data had to have been gathered on children (rather than adults) from countries other than the US. Reports had to have been written in English and to have appeared in online databases or been made available to us by Sesame Workshop by September 20th, 2010.

Our study was on the effects of exposure to *Sesame Street* as aired, thus we did not include experiments in which the researchers manipulated and examined the effects of different features of the content (e.g., intelligibility, presence or absence of material about out-group members). Nor did we include formative research if it was clear that the programming being tested was in the early stages of development and would probably be modified before being aired.

Report identification and selection

We began with a pool of 52 internal reports and published articles presented to us by Sesame Workshop, of which 20 provided quantitative data on the effects of children's exposure to *Sesame Street* on at least one of the learning outcomes. Of these 20, 18 contained sufficient statistical information for us to calculate at least one effect size estimate. Of the 32 other reports, 12 contained qualitative data only, 13 assessed viewership rather than effects, 4 examined the effects of outreach materials rather than exposure to the program, and 3 evaluated different television programs.

In addition, we carried out a systematic search of seven databases: Communication and Mass Media Complete, PsycINFO, ProQuest Research Library, ProQuest Theses/Dissertations, Social Science Full Text, Web of Knowledge, ERIC, and WorldCat Theses/Dissertations

using the term “Sesame Street” in the title, abstract, subject or key words. We also examined the reference lists of chapters in *G is for Growing*, a book written by core personnel from Sesame Workshop in 2001, reviewing the prior thirty years of research. This search yielded only seven additional, non-redundant reports that met the inclusion criteria. This low number reflects the fact that we already had the proprietary research and most of the non-proprietary research on the effects of *Sesame Street* was conducted in the US (or, in a few instances, was on outcomes not examined in the current project, such as learning of English as a second language). Of the seven relevant reports, three contained adequate statistical information to calculate at least one effect size estimate.

This left us with a final sample of 21 reports with the necessary statistical information for computing effect sizes. [Table 1](#) summarizes the scope of this sample and [Appendix A](#) lists the reports.

Limitations of the sample

As noted earlier, two of the quantitative reports from the Sesame archives and four nonproprietary reports found via on-line searches were not included because of inadequate statistical information. These deficiencies could not be rectified – we did not have access to the raw data of the *Sesame Street* proprietary studies and given the age of the four nonproprietary studies (all of which were from the 1970s) we did not attempt to contact the authors. Also, the reports varied substantially in the degree of detail given to findings, particularly non-significant differences between groups. Such inconsistency among the reports rendered it meaningless to conduct any vote-counting analysis on the implied effect assessments (see [Borenstein, Hedges, Higgins, & Rothstein, 2009](#)). Instead, [Web Appendix B](#) outlines the main findings of the studies we could not include, and we briefly review those findings in the [Results](#) section as additional context for the meta-analytic findings.

In the few instances where we had access to both a published article and the original research report, or to two published articles describing the same study, we used the version with the most complete information. Had there been contradictory or inconsistent information, we would have used the published version (or in the case of two publications, the most recent version), and had there been unique information in both, we would have used both versions.

Implications for analytic approach. Given that we only had data from 21 reports, we could not assess the effects of report-level or study-level moderators in a multivariate context. This places severe constraints

Table 1

A summary of the scope of the meta-analysis.

	Reports		Studies		Effect sizes	
	N	%	N	%	N	%
Total	21	100.0	24	100.0	242	100.0
Sub-sample effect size	6	28.6	6	25.0	108	44.6
Type of report						
Internal reports	17	80.9	18	75.0	195	80.6
Published articles	3	14.3	5	20.8	36	19.0
Conference presentation	1	4.8	1	4.2	1	0.4
Year conducted						
Before 2000	4	19.0	4	16.7	55	22.7
2000 or later	17	81.0	20	83.3	187	77.3
Commissioned by						
Sesame Workshop/CTW	17	81.0	20	83.3	185	76.4
Method						
Experiment	4	19.0	4	16.7	134	55.4
Quasi-experiment	7	33.3	8	33.3	56	23.1
Survey	10	47.6	12	50.0	52	21.5
Research design						
Longitudinal/pre-post	18	85.7	21	87.5	221	91.3
Cross-sectional	3	14.3	3	12.5	21	8.7

Note. Two of these reports have subsequently been published in peer-reviewed journals ([Borzekowski & Henry, 2011](#); [Borzekowski & Macha, 2010](#)).

on the interpretation of any individual moderator effect – the studies reported in these reports varied not only in location and context but also in the year, the type of population sampled, and the outcomes assessed. We are, of necessity, cautious in presenting moderator analyses.

We present these limitations here, rather than in the **Conclusion and discussion** section as is more normative, both to explain our analytic focus and to provide the appropriate cautious framework for interpreting the results.

Measures

Exposure to Sesame Street

For studies that relied on survey measures of viewing, we gave precedence to children's self-reports of exposure, and used caregivers' reports only when children's data were not given. In some studies (e.g., Cole et al., 2003) there were no direct measures of children's exposure – in such cases we relied on information about the number of weeks in which *Sesame Street* was broadcast between pretest and posttest. For studies that differentiated levels of exposure by groups (experimentally induced exposure vs. control, or high vs. low exposure groups, or statistically created groups based on “naturally” occurring levels of exposure), we coded such groupings into corresponding levels of exposure. We did not code effect sizes for analyses where the predictor was memory for *Sesame Street* characters (often labeled “receptivity”), given Cook and Connor's (1976) point that measures of character recognition share some of the skills required for measures of letter and number recognition, thus potentially leading to over-estimation of the relationship between viewing and learning.

Outcomes

Our reasoning about which learning outcomes to include was based on examination of the core educational curriculum of Sesame Workshop and the frequency with which outcomes appeared across the studies. As summarized in Table 2, we grouped the outcomes into three conceptual categories (cognitive outcomes, learning about the world, social attitudes and reasoning) in order to facilitate comparison with prior research. For each outcome, whenever a composite

index was used by the primary study, we coded the results based on that index measure rather than on individual items.

Moderator variables

We began by coding over 20 possible moderators, but we focus here on the subset that was most meaningful given the number of cases. Table 3 summarizes the key moderators and gives Krippendorff's alpha for inter-coder reliability, assessed across three coders on a random sample of 17 studies from 9 of the reports (Krippendorff, 2004). Country income codes were entered from World Bank codes for each country at the time the data were gathered (World Bank, 2011), and collapsed into a standard dichotomy of low- and middle-income (LAMI) and high-income countries. For each effect estimate, we also recorded whether the dependent variable was assessed with a multi-versus single-item measure, and the number of covariates in the analysis from which the estimate would be calculated.

Subsample

We coded relevant effect estimates for any of nine different types of subsamples. These were differentiated by sex (boys or girls), location (urban or rural), family's socioeconomic status (low SES), and age (3-year-olds, 4-year-olds, 5-year-olds, and 6-year-olds). These subsamples could overlap within a study, so aggregating the effect sizes was only meaningful within each of the subsample categories.

Computing effect size estimates

The version of *d* reported in this analysis (denoted as *g* in Hedges & Olkin, 1985) is multiplied by a correction factor (Borenstein et al., 2009), to compensate for the known upward bias (i.e., overestimation of effect size) in Cohen's *d*.

To compute effect sizes that involved comparisons between children's test scores taken before and after exposure to *Sesame Street*, we needed to correct for the pretest and posttest correlation (see Borenstein et al., 2009; Morris & DeShon, 2002). However, this information was frequently missing from the reports. Smith, Glass, and Miller (1980) suggest that the correlations (for adults) should be estimated at .50 for an established psychological inventory assessed with a two to six months interval, or at .70 for cognitive measures assessed four months apart or more. We used a fairly conservative estimate of .50, based in

Table 2

Learning outcomes: Categories and operationalization.

Outcome category	Examples of operationalizations
Cognitive outcomes	
Literacy	Letter recognition, naming, writing, vocabulary
Numeracy	Number recognition, counting
Relationships & classifications	Knowledge of colors, shapes, sizes, ordering (e.g., size, distance), spatial relations
Learning about the world	
Environment & science knowledge	Knowledge of physical features such as coral reefs, plants, animals; recycling, water conservation
Health & safety knowledge & behavior	How diseases (e.g., malaria, HIV/AIDS) are transmitted, hygiene, nutrition, fire & bike safety, crossing road safely. How often child washes hands, brushes teeth, uses bed net.
Cultural knowledge	Knowledge of own country name, musical instruments, cultural artifacts
Social attitudes/reasoning	
Prosocial reasoning	Perspective-taking, rules for fair play, cooperation, reasoning about conflict resolution & inclusiveness
Attitudes: Disabilities or HIV/AIDS	Empathy or friendship interest toward child with disability; okay to be friends or hug someone with HIV/AIDS
Attitudes: Religion, nationality, ethnicity	Stereotyping, perceived differences & similarities, interest in interacting with child based on religion, nationality or ethnicity, e.g., Protestants & Catholics, Arabs & Jews, Serbians & Albanians, Roma, African descent
Attitudes: Other sources of difference	Stereotyping, perceived differences & similarities, interest in interacting with child based on sex, body size, familiarity

Table 3

Key moderators: Operationalization and inter-coder reliability.

Variables	Operationalization	α	<i>N</i>
Study level			17
Method	Cross-sectional survey, panel surveys, quasi-experiment, experiment	.76	
Design ^a	Cross-sectional, longitudinal/pre-post	.76	
Quality index	Presence or absence of six quality indicators:	.78	
	o Random sampling or assignment at individual level		
	o Multiple indicators for key variables		
	o Reliability assessment for key indices		
	o Quality control in field operations		
	o Experimental or statistical controls		
	o Strong basis for causal inferences (panel design, between-group or pre-post experimental design)		
Low-SES samples	Study only sampled from low-income/low-SES populations; sample not exclusively low SES	.80	
Country income	Low- or middle-income vs. high-income World Bank ratings for country in year of study		From World Bank (2011)

Note. Reliability (calculated as Krippendorff's alpha) was assessed across three coders.

^a For “design” one coder consistently deviated from the other two, so reliability was assessed only for those two coders.

part on observations of our data set. For example, in Borzekowski and Wahyuningrum (2009), the authors presented correlations between baseline and post-intervention measures at $r = .55$ for early cognitive skills, $r = .70$ for early literacy scores, and $r = .67$ for arithmetic skills. These numbers provided an empirically plausible range. However, we expect test-retest correlations (where reported) to be smaller for attitudinal outcomes and for younger children. Our choice of .50 thus involved balancing these two sets of considerations. Analyses using an alternate estimate of $r = .70$ yielded the same conclusions.

Structure of the data set

As shown in Table 1, the data came from 21 reports that contained 24 studies (a study being defined as data collection on an independent group of individuals). These studies reported 242 effect sizes, of which 108 are from overlapping sub-samples. The data thus have a nested structure with each effect size nested in its corresponding study (For simplicity, study is treated as the higher-level unit, given that only two reports contained more than one study.). The studies in our analysis generated 1 to 14 effect sizes based on whole samples, with an average of 5.58 effect sizes per study. Six of the studies reported additional statistical information for estimating effect sizes in at least one subsample, with one study reporting 70 effect sizes from eight of the nine differentiated subsamples.

Analytic strategy

Model specification

The basic model of a meta-analysis may be specified as follows (see Hox & de Leeuw, 2003):

$$d_j = \delta + \mu_j + \varepsilon_j \tag{1}$$

where, $j = 1, 2, 3, \dots, J$ mutually independent effect sizes (when there is only one effect size from each study, J is thought of as number of studies), d_j = reported j th effect size, δ = the population effect size (or the “true” effect size), μ_j = deviation of the j th effect size from

the population parameter ($N \sim [0, \sigma_{\mu_j}^2]$), and ε_j = sampling error for j th study ($N \sim [0, \sigma_{\varepsilon_j}^2]$). Normally, $\sigma_{\varepsilon_j}^2$ is reported in the study and is entered as data.

In our case, almost all studies had multiple effect sizes, or effect sizes on multiple outcomes. Given that effect sizes within a study are all measured on the same children (or a subset of those same children), they can reasonably be expected to be more similar than effect sizes across studies. That is, the bulk of heterogeneity ($\sigma_{\mu_j}^2$) is expected to be between studies. We can specify the heterogeneity as a random-effect variance at the study K level, leading to a three-level model as follows:

$$d_j = \delta + \mu_j + \mu_k + \varepsilon_j \tag{2}$$

where μ_j = unique effect of the effect size j ($N \sim [0, \sigma_{\mu_j}^2]$), while μ_k = unique effect of the study k ($N \sim [0, \sigma_{\mu_k}^2]$).

Model estimation

We began by testing the more parsimonious model specified in Eq. (1), then estimated the model specified in Eq. (2) and conducted a statistical test of the between-study random-effect variance ($\sigma_{\mu_k}^2$). The two-level model yielded a $-2 \times \log$ likelihood at -94.13 , the corresponding figure for the three-level model is -50.16 . The difference was 43.97 ($\chi^2, df = 1, p < .001$). Thus, the assumption of the more parsimonious model that effect sizes within a study would be independent was not justified, and the three-level model was clearly preferable. Indeed, as can be seen from the final two columns in Table 4, virtually all heterogeneity among effect sizes is a reflection of between-study variation.

Results

Whole sample effect sizes

Table 4 summarizes the findings based on the 134 whole sample effect size estimates, calculated on 10,596 children from 15 countries.

Table 4
Whole sample effect size estimates.

Based on ...	<i>J</i>	\bar{d}	$SE_{\bar{d}}$	C.I.	<i>z</i>	$\sum N_k$	$\sigma_{\mu_j}^2$	$\sigma_{\mu_k}^2$
Whole sample (<i>K</i> = 24)								
Two-level model	134	.260	.023	.215–.304	11.41	10,596	.133***	–
Three-level model		.292	.049	.197–.387	6.01		.000	.037***
Studies with								
Cross-sectional design (<i>K</i> = 3)	21	.460	.136	.194–.727	3.39	615	.000	.041
Longitudinal/pre-post design (<i>K</i> = 21)	113	.266	.050	.168–.364	5.30	9981	.000	.033***
Studies that were								
Experiments (<i>K</i> = 4)	31	.377	.107	.168–.586	3.53	1403	.000	.036*
Quasi-experiments (<i>K</i> = 8)	52	.281	.082	.120–.443	3.41	2294	.000	.038**
Surveys (<i>K</i> = 12)	51	.262	.071	.124–.400	3.71	7336	.000	.034**
Tests of								
Cognitive outcomes (<i>K</i> = 15)	68	.284	.061	.165–.403	4.67	7604	.000	.038***
Learning about the world (<i>K</i> = 13)	29	.339	.078	.188–.491	4.38	7797	.000	.049***
Social reasoning/attitudes (<i>K</i> = 17)	37	.189	.038	.114–.265	4.93	5837	.000	.000
Studies conducted in countries with								
Low/medium income (<i>K</i> = 16)	110	.293	.055	.185–.400	5.34	8319	.000	.036***
High-income (<i>K</i> = 8)	24	.285	.105	.080–.490	2.73	2277	.000	.043*
Studies sampling from								
Low SES groups only (<i>K</i> = 9)	48	.413	.080	.255–.570	5.14	1725	.000	.037**
Not only low SES groups (<i>K</i> = 15)	86	.221	.051	.121–.321	4.32	8871	.000	.022***

Note. *J* = number of tests (effect sizes), *K* = number of studies, \bar{d} = unbiased estimate of the average effect size, $SE_{\bar{d}}$ = standard error of the effect size estimate, C.I. = confidence interval of the effect size estimate, $\sum N_k$ = sum of all individuals across the *K* studies, $\sigma_{\mu_j}^2$ = random variance across *J* effect sizes, $\sigma_{\mu_k}^2$ = random variance across *K* studies. Significance of the average effect size estimate at $p < .05$ is indicated by a *z* greater than 1.96.
* $p \leq .05$. ** $p \leq .01$. *** $p \leq .001$.

Overall effect

Based on the three-level model, the overall estimate of effects of exposure to *Sesame Street* was significant and positive, $d = .292$. The 95% confidence interval suggests that the effect size parameter is likely to be between .197 and .387. On average, those who watched more (either in comparison to other children, or in comparison to themselves earlier) scored between one- and two-fifths of a standard deviation higher, taken across the different types of outcomes. The fact that 113 (84%) of the estimates were from longitudinal or pre-post designs, controlling for the child's initial level of ability, enhances the credibility of the overall effect size estimate.

Moderation by methodological features

Five methodological features were considered as potential moderators of overall effect size (see Table 3). Due to the small number of studies, each of them was tested in a separate regression analysis. In fact, none of the features moderated the effect size (all p values $> .15$).

Effects by outcome category

As can be seen in Table 4, there were significant positive effects for each of the three outcome categories, ranging from $d = .189$ for social attitudes to $d = .339$ for learning about the world. However, the regression analysis found no significant moderation of the average effect size by outcome category. Neither the comparison between average effect sizes for cognitive outcomes versus learning about the world ($b = .089$, $p = .127$), nor the comparison between cognitive outcomes and social reasoning/attitudes ($b = .011$, $p = .846$) was significant. It is worth noting that cognitive outcomes were assessed with over 7000 children from 10 countries, learning about the world with over 7000 children from 9 countries, and social attitudes and reasoning with over 5000 children from 10 countries.

Effects by country income

The substantial majority of whole-sample effect size estimates (82%) came from studies conducted in low- and middle-income countries. As shown in Table 4, the average effect size from low- and middle-income countries was significant and positive ($d = .293$).

Most effect size estimates from low- and middle-income countries came from experimental or quasi-experimental studies (74%), strengthening the plausibility of causal claims about the effectiveness of the program for those who participated. However, given concerns that children in low- and middle-income countries are relatively unlikely to be in caregiving or preschool environments where they would be encouraged to watch educational programs, the predominance of effect size estimates from studies in which children were required to watch the program may potentially undercut the generalizability of the findings.

To examine this issue, we estimated the average effect size for the sub-set of studies conducted in low- and middle-income countries that were surveys. In these studies, viewing was more likely to be self-selected by the child rather than required by caregivers or teachers as part of the research protocol. There were six such studies, yielding 29 individual effect size estimates, assessed on a total of 5967 children. The average effect estimate for surveys conducted in low- and middle-income countries was significant and positive ($d = .213$, $SE = .070$, 95% confidence interval = .076–.351, $z = 3.04$).

In addition to these significant effects observed in low- and middle-income countries, there was a virtually identical effect observed in studies conducted in high-income countries ($d = .285$). The regression analysis found no moderation of effect size by country income level ($b = .004$, $p = .973$). We regard the equivalence of these effect size estimates with considerable caution, given that there are only 24 whole-sample estimates from high-income countries and that various factors are confounded with country income.

One such factor is the year in which the research was conducted: 91% of the effect sizes from low/middle-income countries were from studies conducted after 2000, compared to 25% of effect sizes from

high-income countries. Other factors related to methodology: 74% of effect sizes from low/middle-income countries were from experimental or quasi-experimental studies, compared to 8% from high-income countries. More than 97% of the effect sizes from low/middle-income countries were based on longitudinal assessments, compared to only 25% of those from high-income countries. Studies conducted in low/middle-income countries also had more desirable methodological features – on average, 3.91 out of a total of 6 – than studies conducted in high-income countries.

It is useful to establish that the effects were significant and positive for both lower- and higher-income countries. However, given all the differences outlined above, it is inappropriate to compare the strength of the effects or to conclude that the effects really are identical.

Effects by sample SES

There were nine studies in which researchers explicitly reported sampling children exclusively from low-SES populations. These nine studies yielded 48 whole-sample effect size estimates. Overall, the effect of exposure to *Sesame Street* in low-SES samples was positive and significant ($d = .413$). The credibility of this relatively strong effect is enhanced by the fact that these studies had, on average, four of six indicators of methodological quality ($M = 4.25$, $SD = 1.04$).

There was also a significant positive effect in samples that were not exclusively from low-SES populations ($d = .221$). We strongly caution against comparing the two effect sizes, despite the fact that they are significantly different ($b = .197$, $p = .029$) because other factors are confounded with this variable, notably the distribution of effect estimates across learning outcomes, methods used, and study quality. With regard to the distribution of outcomes, comparing low-SES-only studies versus other studies, 60% versus 45% of estimates were for cognitive outcomes, 29% versus 17% were for learning about the world, and 10% versus 37% were for social attitudes. With regard to method, 81% of the effect size estimates from low-SES-focused studies were obtained via experiments or quasi-experiments, as opposed to 51% of the estimates from other studies. The studies that sampled exclusively from low-SES populations also had more desired methodological features than the other studies ($M = 4.25$ vs. $M = 3.17$).

Subsample effects

Table 5 shows the effect size estimates from subsamples. As can be seen, there were very few studies that contained adequate data for calculation of subsample estimates. We place little value on the findings as a diagnosis about the conditions under which exposure may be more or less effective, given that the sex and location comparisons were each based almost entirely on one study. We did not include information on low SES subsamples, because there were only two estimates that could be calculated, both from the same study. The findings serve primarily to highlight the need for future systematic examination of possible subgroup differences.

Sex

There were 11 estimates that could be calculated for boys alone, of which 8 were from a single study conducted in Bangladesh (Research & Computing Services, 2006). Based on those 11 estimates, 5 of which were for social attitudes/reasoning, the average effect for boys was not statistically significant. There were 10 estimates that could be calculated for girls alone, of which 8 were from the same study in Bangladesh. Based on those 10 estimates (4 of which were for social attitudes/reasoning) there was a significant, positive effect for girls ($d = .250$).

Location

There were 13 effect size estimates each for urban and rural subsamples, 12 of which came from one study conducted in Egypt (Rimal

Table 5
Effect size estimates from subsamples.

Based on...	<i>J</i>	\bar{d}	$SE_{\bar{d}}$	C.I.	<i>z</i>	$\sum N_k$	$\sigma_{\mu_j}^2$	$\sigma_{\mu_k}^2$
Sex subsamples								
Boys (<i>K</i> = 4)	11	.148	.079	–.006–.302	1.88	1194	.005	.000
Girls (<i>K</i> = 3)	10	.250	.081	.090–.409	3.09	707	.000	.000
Location subsamples								
Urban residents (<i>K</i> = 2)	13	.085	.087	–.086–.255	0.97	510	.033	.000
Rural residents (<i>K</i> = 2)	13	.244	.071	.105–.384	3.44	506	.000	.000
Age subsamples								
3-year-olds (<i>K</i> = 1)	9	.579	.126	.332–.825	4.60	52	.000	.000
4-year-olds (<i>K</i> = 2)	19	.560	.279	.013–1.11	2.01	216	.000	.144
5-year-olds (<i>K</i> = 3)	20	.487	.116	.259–.714	4.19	498	.017	.019
6-year-olds (<i>K</i> = 2)	11	.066	.122	–.173–.305	0.54	423	.085*	.000

Note. *J* = number of tests (effect sizes), *K* = number of studies, \bar{d} = unbiased estimate of the average effect size, $SE_{\bar{d}}$ = standard error of the effect size estimate, C.I. = confidence interval of the effect size estimate, $\sum N_k$ = sum of all individuals across the *K* studies, $\sigma_{\mu_j}^2$ = random variance across *J* effect sizes, $\sigma_{\mu_k}^2$ = random variance across *K* studies. Significance of the average effect size estimate at $p < .05$ is indicated by a *z* greater than 1.96.

* $p \leq .05$.

et al., 2006). Based almost wholly on that one study, there was no significant effect of exposure on urban residents, but there was a significant effect for rural residents ($d = .244$).

Age

As can be seen in Table 5, there were few studies that presented data on age subsamples – and only one that gave data for 3-year-olds ($n = 52$). Nonetheless, the limited data suggest an age pattern whereby effects were stronger for 3- to 5-year-olds than for 6-year-olds.

Context: Studies not included

Web Appendix B describes the primary findings of the six quantitative reports that could not be included in the meta-analysis. The studies varied substantially in the number of outcomes assessed, and most of them reported significant effects of exposure only for a minority of those outcomes, or for a subset of children (e.g., field experiment in Mexico by Diaz-Guerrero, Reyes-Lagunes, Witzke, & Holtzman, 1976). To the extent that significant effects were reported, they mostly occurred for numeracy and literacy outcomes (e.g., letter recognition) that were explicitly taught by the program (Hammer, 1972; Taylor & Skanes, 1977; Salomon, 1974; Ulitsa Sezam Team, 1999) rather than more general gains in IQ or vocabulary scores (Taylor & Skaynes, 1977). In addition, there was little evidence of significant positive effects on social attitudes (Fox, Killen, & Leavitt, 2005; Ulitsa Sezam Team, 1999). Overall, there is no indication that this set of studies would have altered our meta-analytic findings.

In addition, there was some indication of positive effects of exposure on low SES samples, but only within the target age range of the program (i.e., 2- to 5-year-olds). Both Hammer (1972) and Salomon (1974) reported stronger effects for kindergarten children from lower SES backgrounds than those from higher SES backgrounds. However, for children older than the target age range of the program, there was less evidence of positive effects for low SES groups. In one study, 5–7 year old children from low-income, isolated communities in Canada showed very little evidence of positive effects of two years of in-school exposure to the program (Taylor and Skanes, 1977). Similarly Salomon (1974) found that “lower class” 2nd and 3rd graders benefited less from exposure to *Sesame Street* than those from middle class backgrounds, despite watching and enjoying the program more – no such pattern was observed with kindergarteners. These findings are consistent with the age and SES analyses of the meta-analysis.

After the analyses were completed, we found an additional non-redundant publication (Borzekowski & Macha, 2010) that fell within our inclusion parameters, including the cut-off date. Although

the authors had listed *Sesame Street* as one of their key terms, the title of the article referenced the name of the co-production (*Kilimani Sesame*), and the article did not emerge in any of our online searches. This raised concern about other articles we may have missed because of this issue. We re-ran the searches using the names of the co-productions, but did not find any additional articles published before September 2010. Since our cut-off date, another of the proprietary reports has been published (Borzekowski & Henry, 2011), but in that case, there was no additional data contained in the published article.

With regard to the Borzekowski and Macha's (2010) study, we had used the proprietary report written by those authors, which described the effects of exposure to *Kilimani Sesame* on five measures of health knowledge, behaviors, and attitudes toward those with HIV/AIDS. Those data are included in the meta-analysis. However, the published article gave data for seven additional outcomes observed with the same children, and those effect estimates are not included in the meta-analysis. Three relate to literacy: pre-literacy ($d = .225$), letter recognition ($d = -.020$), and early literacy ($d = -.131$). Three relate to numeracy: number recognition ($d = .051$), counting ($d = .614$) and arithmetic ($d = .188$). One measure assessed prosocial reasoning: ($d = .856$). The estimates for literacy and numeracy are within the range of other effect size estimates computed for these outcomes, and do not alter the conclusions about the significance, direction, or magnitude of the effect size for cognitive outcomes. The single estimate for prosocial reasoning is somewhat higher than any of the other 37 estimates of social reasoning/attitudes (see Web Appendix A), but the conclusion of a generally small, significant effect for these outcomes remains unaltered.

Context: Exposure to *Sesame Street* and availability of other resources

Table 6 summarizes survey data about children's exposure to *Sesame Street* in those low- and middle-income countries for which such information was available. In addition, it lists UNESCO data for national enrollment in pre-primary education, and data on adult literacy rates as a rough indicator of the likelihood that children would have caregivers who could help them learn to read and write. As can be seen, rates of reported exposure to *Sesame Street* were generally comparable to or higher than enrollment in pre-primary education, including in countries such as Bangladesh, Egypt, or India where adult literacy rates were relatively low.

Conclusion and discussion

Overall, this meta-analysis indicates that there was a significant positive effect of exposure to the various versions of *Sesame Street*.

Table 6

Context in low and middle-income countries: Pre-primary enrollment, adult literacy, and exposure to *Sesame Street*.

Country	Year	Author	Sample	Caregiver reports			Child reports			UNESCO Institute of Statistics	
				N	Child watches <i>Sesame</i> (%)	Viewing measure	N	Watches <i>Sesame</i> (%)	Viewing measure	Pre-primary enrollment GER % ^a	Adult literacy %
Bangladesh	2007	ACPR (2007)	Nationally representative	6742	70	Watches	6710	77	Usually	11	55.9
Egypt	2006	Rimal et al. (2006)	Nationally representative	426	70	Once per month or more	487	57	What programs do you watch?	24	66.4 (2006)
India	2009	GyanVriksh (2010)	Stratified cluster sample, 8 states	–	–	–	4301	48	Watch regularly	28 (2002)	62.8 (2006)
Indonesia	2009	Synovate (2009)	Nationally representative	1767	47	Ever watched	–	–	–	48	92.2 (2008)
Kosovo	2005	Prism Research (2005)	Random sample ethnic Serbs	497	31	Watches	–	–	–	2.8 (2000–2001)	91.9 (2007)
South Africa	2002	Nielsen (2003)	Urban & rural random samples	716	72	Watches	–	–	–	65	88.7 (2007)

^aGross Enrollment Ratios (GER) refer to the number of pupils enrolled in a given education level regardless of age, expressed as a percentage of the population in the theoretical age group for that level of education. Unless noted otherwise, UNESCO Institute of Statistics data are from 2009. Data for Northern Ireland are only available aggregated with the U.K. Data for Kosovo for pre-primary enrollment are from Ilazi (2003); and for adult literacy, from the US Department of State (2012).

On average, across 24 studies conducted with more than 10,000 children from 15 countries, those who watched more performed better than those who watched less. This effect was observed across a variety of methodological approaches, in low- and middle-income countries as well as in high-income countries, in samples from low-SES populations as well as in samples that were not specifically targeted to be low-SES.

Watching *Sesame Street* was associated with learning about letters, numbers, shapes, and sizes – the elements of basic literacy and numeracy that remain fraught for millions of children globally. It was also associated with learning about science, the environment, one's culture, and health and safety-related practices such as washing one's hands or wearing a bike helmet. Finally, it was also associated with more prosocial reasoning about social interactions and more positive attitudes toward various out-groups, including those that were associated with long-standing hostilities or stereotyping. The fact that over 90% of estimates contained some control for the child's initial performance on that outcome considerably enhances the plausibility that these were causal effects on those children who were selected to participate or who chose to watch.

Part of the value of this project lies in the fact that the research was conducted outside the US, allowing us to compare the findings with effects observed within the US. Studies by Wright et al. (2001) and Rice et al. (1990) showed that exposure to *Sesame Street* in the US predicted concurrent and subsequent educational achievement for children who watched at ages 2 and 3, but not at ages 4 or 5, suggesting a particular developmental window of effectiveness. This window may be somewhat wider in other countries, as our results showed significant effects for 3-, 4-, and 5-year old subsamples. Continued investigation of these developmental patterns, and assessment of effects with 2-year-olds, would be valuable. We also found some indication of sex and rural/urban subsample differences not observed in the US. Such findings, although not robust enough because of the very small number of effects involved, suggest directions for further investigation.

An additional issue that arose in the US concerned Cook et al.'s (1975) critique that *Sesame Street* appeared to have little benefit for children of low SES backgrounds and could ultimately widen socio-economic achievement gaps. As noted earlier, we cannot meaningfully compare the effect sizes across different levels of SES. However, significant, positive effects were observed in low-SES-only samples, as well as in mixed-SES samples; in World-Bank-designated low- and middle-income countries, as well as in high-income countries. Thus, the data

suggest that there are educational benefits of exposure to *Sesame Street* for children from economically disadvantaged backgrounds as well as those in more affluent circumstances, and that the positive effects observed in the US were also observed in a variety of other countries.

How important or valuable are the effects? This question has two intertwined but conceptually distinct parts. The first involves examining the numerical value of the overall effect size and the second involves examining the effect size in a substantive context, often relative to other options.

For the first part, the straightforward answer is that there was an average overall effect of more than a quarter of standard deviation ($d = .292$), and that the effect size estimates were somewhat above that average for “learning about the world” and somewhat below that for social reasoning. These numbers are undoubtedly crude estimates of the effects. Our best guess is that the “real world” effects lie somewhere within the confidence intervals – as low as .20 or as high as .39.

These estimates are, of course, based only on the data we could use. They would probably be somewhat different had we been able to include all relevant data from all the reports. As noted in the Methods section, we had reports containing effects that we could not estimate because of inadequate information, and for some reports we could only calculate effect sizes for some of the outcomes. Despite these limitations, the evidence of positive overall effects on learning outcomes was robust across a variety of research conditions.

The second part of the question concerns how to interpret the numbers in terms of real world differences that they may represent. The US Department of Education provides guidelines for reporting on educational interventions, and those guidelines recommend translating effect sizes into “improvement indexes” (What Works Clearinghouse, 2011). Such a strategy, based on Cohen's (1988) U_3 index, involves looking up the z score associated with the effect size and observing the area under the normal curve. Thus, an effect size of .29 corresponds to a z score of .29, representing a percentile gain or “improvement” of 11.6 (either within an individual or between those who watch less vs. more). The counter-factual causal reasoning implies that the hypothetical average child who watched *Sesame Street* would be at roughly the 62nd percentile, whereas if that hypothetical average child had not watched, he or she would (by definition) be at the 50th percentile.

Another way of assessing the social significance of the findings is to consider the effects and availability of other early childhood interventions. As noted earlier, Nores and Barnett (2010) analyzed experiments

and quasi-experiments in which children were assigned to receive early childhood educational/child-care interventions, often combined with nutritional supplements. They reported an average effect of .35 on cognitive outcomes. This is not much bigger in magnitude than the effect of .28 on cognitive outcomes for exposure to *Sesame Street*, observed in the current meta-analysis. Indeed, the effect estimates of .38 for experiments and .28 for quasi-experiments in the current project (taken across outcome categories) suggest that the average effect of being assigned to watch the program is comparable to the average effect on cognitive outcomes of other early interventions in which children are made to participate. However, a relatively unique aspect of the current intervention is that watching *Sesame Street* is something that children can potentially choose to do of their own volition, rather than being assigned to do it by adults. The significant, positive effect size from surveys suggests that self-initiated viewing may also contribute to children's learning.

This is not to undercut the vital importance of investing in high quality preschools or providing caregiver training to improve the quality of early childcare. However, UNICEF and various international teams of researchers have pointed to the urgent need for effective, scalable, and affordable interventions to provide enrichment to children in low-income regions before developmental delays accumulate (Engle et al., 2007, 2011; UNICEF Innocenti, 2011; Walker et al., 2007, 2011).

We have no data for the cost-efficiency of creating and broadcasting *Sesame Street* co-productions relative to other early education interventions. However, it is clear that *Sesame Street* broadcasts offer consistency of quality at the content-delivery end: all children (with viewing access) within a country can receive the same educational curricula, presented in the same way. Across countries, the co-productions all share the same visual style, core learning goals, and teaching strategies.

Moreover, with regard to scalability and access, the percentages listed in Table 6 translate into millions of children who watched the program (*Sesame Workshop*, 2011). For example, in Bangladesh, nearly 7 million children aged 4–7 had watched *Sisimpur* (Khan et al., 2007). In Indonesia, 7.5 million children aged 3–6 had watched *Jalan Sesama* (Synovate, 2009). In South Africa, 2 million children aged 3–9 had watched *Takalani Sesame* (Nielsen, 2003). In Egypt, nearly 12 million children aged 2–8 had watched *Alam Sim Sim* (Synovate, 2007). Even in India, where only 58% of children are estimated to have access to television, 21% of children aged 2–8 had watched *Galli Galli Sim Sim* in the past week — over 20 million children were watching on a regular basis (GyanVriksh Technologies, 2008).

The findings of the current meta-analysis, taken together with data on the reach of the program, indicate that *Sesame Street* is an enduring example of a scalable and effective early childhood educational intervention. The significant, positive effects on cognitive, learning, and socio-emotional outcomes observed in the current meta-analysis represent real educational benefits for the millions of preschool-age children around the world who visit *Sesame Street* via their televisions.

References

- Anderson, D. R., Huston, A. C., Schmitt, K. L., Linebarger, D. L., & Wright, J. (2001). Early childhood television viewing and adolescent behavior: The recontact study. *Monographs of the Society for Research in Child Development*, 68(1), 1–143. <http://dx.doi.org/10.1111/1540-5834.00120>.
- Anderson, D. R., Lavigne, H. J., & Hanson, K. G. (2013). The educational impact of television: Understanding television's potential and limitations. In A. Valdivia, & E. Scharrer (Eds.), *The international encyclopedia of media studies: Media effects/media psychology* (pp. 635–656). Malden, MA: Wiley-Blackwell.
- Baker-Henningham, H., & Boo, F. L. (2010). Early childhood stimulation interventions in developing countries: A comprehensive literature review. *IZA Discussion Paper No. 5282* (Retrieved from www.iadb.org/document.cfm?id=35349131)
- Bigler, R. S. (1999). The use of multicultural curricula and materials to counter racism in children. *Journal of Social Issues*, 55, 687–705. <http://dx.doi.org/10.1111/0022-4537.00142>.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex, UK: Wiley & Sons. <http://dx.doi.org/10.1002/9780470743386>.
- Borzekowski, D., & Wahyuningrum, Y. (2009). *The educational impact of Jalan Sesama on Indonesian preschool children: An experimental study*. John Hopkins Bloomberg School of Public Health. Unpublished research report prepared for Sesame Workshop.
- Borzekowski, D. L. G., & Henry, H. K. (2011). The impact of *Jalan Sesama* on the educational and healthy development of Indonesian preschool children: An experimental study. *International Journal of Behavioral Development*, 35, 169–179. <http://dx.doi.org/10.1177/0165025410380983>.
- Borzekowski, D. L. G., & Macha, J. E. (2010). The role of *Kilimani Sesame* in the healthy development of Tanzanian preschool children. *Journal of Applied Developmental Psychology*, 31, 298–305. <http://dx.doi.org/10.1016/j.appdev.2010.05.002>.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: LEA.
- Cole, C. F., Arafat, C., Tidhar, C., Tafesh, W. Z., Fox, N. A., Killen, M., et al. (2003). The educational impact of Rechov Sumsum/Shara'a Sim sim: A Sesame Street television series to promote respect and understanding among children living in Israel, the West Bank and Gaza. *International Journal of Behavioral Development*, 27, 409–422.
- Cole, C. F., Richman, B. A., & McCann Brown, S. A. (2001). The world of Sesame Street research. In S. M. Fisch, & R. T. Truglio (Eds.), *"G" is for growing: Thirty years of research on children and Sesame Street* (pp. 147–179). Mahwah, NJ: Lawrence Erlbaum.
- Cook, T. D., Appleton, H., Conner, R. F., Shaffer, A., Tamkin, G. A., & Weber, S. J. (1975). *Sesame Street revisited*. New York, NY: Russell Sage Foundation.
- Cook, T. D., & Connor, R. F. (1976). *Sesame Street around the world. The educational impact. Journal of Communication*, 26, 155–164. <http://dx.doi.org/10.1111/j.1460-2466.1976.tb01395.x>.
- Cunha, F., Heckman, J., & Schennach, S. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78, 883–931. <http://dx.doi.org/10.3982/ECTA6551>.
- Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. B., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, 82, 405–432. <http://dx.doi.org/10.1111/j.1467-8624.2010.01564.x>.
- Elliott, T. R., & Byrd, E. K. (1983). Attitude change toward disability through television portrayal. *Journal of Applied Rehabilitation Counseling*, 14, 35–37 (<http://nrca-net.org/jarc.html>)
- Engle, P. L., Black, M. M., Behrman, J. R., de Mello, M. C., Gertler, P. J., Kapiriri, L., et al. (2007). Child development in developing countries 3: Strategies to avoid the loss of developmental potential in more than 200 million children in the developing world. *The Lancet*, 369, 229–242. [http://dx.doi.org/10.1016/S0140-6736\(07\)60112-3](http://dx.doi.org/10.1016/S0140-6736(07)60112-3).
- Engle, P. L., Fernald, L. C. H., Alderman, H., Behrman, J., O'Gara, C., Yousafzai, A., et al. (2011). Strategies for reducing inequalities and improving developmental outcomes for young children in low-income and middle-income countries. *The Lancet*, 378, 1339–1353. [http://dx.doi.org/10.1016/S0140-6736\(11\)60889-1](http://dx.doi.org/10.1016/S0140-6736(11)60889-1).
- Fisch, S. M., & Truglio, R. T. (2001). *"G" is for growing: Thirty years of research on children and Sesame Street*. Mahwah, NJ: Lawrence Erlbaum.
- Glewwe, P., & Kremer, M. (2006). Schools, teachers, and education outcomes in developing countries. In E. A. Hanushek, & F. Welch (Eds.), *Handbook of the economics of education*, Vol. 2. (pp. 945–1017) New York, NY: Elsevier. [http://dx.doi.org/10.1016/S1574-0692\(06\)02016-2](http://dx.doi.org/10.1016/S1574-0692(06)02016-2).
- Grantham-McGregor, S., Cheung, Y. B., Cueto, S., Glewwe, P., Richter, L., Strupp, B., et al. (2007). Developmental potential in the first 5 years for children in developing countries. *The Lancet*, 369, 60–70. [http://dx.doi.org/10.1016/S0140-6736\(07\)60032-4](http://dx.doi.org/10.1016/S0140-6736(07)60032-4).
- Heckman, J. J., & Masterov, D. V. (2007). The productivity argument for investing in young children. *Review of Agricultural Economics*, 29, 443–493. <http://dx.doi.org/10.1111/j.1467-9353.2007.00359.x>.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York, NY: Academic Press.
- 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. <http://dx.doi.org/10.1111/j.1750-8606.2008.00061.x>.
- Hox, J. J., & de Leeuw, E. E. (2003). Multilevel models for meta-analysis. In S. P. Reise, & N. Duan (Eds.), *Multilevel modeling: Methodological advances, issues, and applications* (pp. 90–111). Mahwah, NJ: Lawrence Erlbaum.
- Ilazi, H. (2003). *Gender review in education in Kosovo*. Kosovo: UNICEF (Retrieved from http://www.unicef.org/kosovo/kosovo_media_pub_educ_005.04.pdf)
- Khan, M. S. H., Chakaborty, N., Rahman, A. P. M., & Nasrin, T. (2007). *2007 follow-up (wave II) evaluation of the reach and impact of Sisimpur: A technical report*. Bangladesh: Associates for Community and Population Research (ACPR).
- Krcmar, M., Grella, B., & Lin, K. (2007). Can toddlers learn vocabulary from television? An experimental approach. *Media Psychology*, 10, 41–63. <http://dx.doi.org/10.1080/15213260701300931>.
- Krippendorff, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. *Human Communication Research*, 30, 411–433. <http://dx.doi.org/10.1111/j.1468-2958.2004.tb00738.x>.
- Linebarger, D. L., & Walker, D. (2005). Infants' and toddlers' television viewing and language outcomes. *American Behavioral Scientist*, 48, 624–645. <http://dx.doi.org/10.1177/0002764204271505>.
- Mares, M. L. (2009). Educational programming and learning. In R. Nabi, & M. B. Oliver (Eds.), *The SAGE handbook of media processes and effects* (pp. 331–344). Thousand Oaks, CA: Sage.
- Mielke, K. W. (1990). Research and development at the Children's Television Workshop. *Educational Technology Research and Development*, 38, 7–16. <http://dx.doi.org/10.1007/BF02314640>.

- Minton, J. H. (1975). The impact of *Sesame Street* on readiness. *Sociology of Education*, 48, 141–151. <http://dx.doi.org/10.2307/2112472>.
- 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. <http://dx.doi.org/10.1037/1082-989X.7.1.105>.
- Naudeau, S., Martinez, S., Premand, P., & Filmer, D. (2011). Cognitive development among young children in low-income countries. In H. Alderman (Ed.), *No small matter: The impact of poverty, shocks and human capital investment in early childhood development* (pp. 9–50). Washington, DC: The World Bank.
- Nores, M., & Barnett, W. S. (2010). Benefits of early childhood interventions across the world: (Under) investing in the very young. *Economics of Education Review*, 29, 271–282. <http://dx.doi.org/10.1016/j.econedurev.2009.09.001>.
- Persson, A., & Musher-Eizenman, D. R. (2003). The impact of a prejudice-prevention television program on young children's ideas about race. *Early Childhood Research Quarterly*, 18, 530–546. <http://dx.doi.org/10.1016/j.ecresq.2003.09.010>.
- Rice, M. L., Huston, A. C., Truglio, R., & Wright, J. (1990). Words from *Sesame Street*: Learning vocabulary while viewing. *Developmental Psychology*, 26, 421–428. <http://dx.doi.org/10.1037/0012-1649.26.3.421>.
- Rispens, J., Aleman, A., & Goudena, P. P. (1997). Prevention of child sexual abuse victimization: A meta-analysis of school programs. *Child Abuse & Neglect*, 21, 975–987. [http://dx.doi.org/10.1016/S0145-2134\(97\)00058-6](http://dx.doi.org/10.1016/S0145-2134(97)00058-6).
- Robb, M. B., Richert, R. A., & Wartella, E. A. (2009). Just a talking book? Word learning from watching baby videos. *British Journal of Developmental Psychology*, 27, 27–45. <http://dx.doi.org/10.1348/026151008X320156>.
- Sesame Workshop (2011). *Country reach extrapolation*. Unpublished report prepared by Sesame Workshop.
- Smith, M. L., Glass, G. V., & Miller, T. I. (1980). *The benefits of psychotherapy*. Baltimore, MD: The Johns Hopkins University Press.
- Synovate (2007). *Alam Simsim survey*. Unpublished research report prepared for Sesame Workshop.
- Synovate (2009). *Survey of reach and perception of Jalan Sesama (wave II, 2009)*. Unpublished research report prepared for Sesame Workshop.
- Truglio, R. T., & Fisch, S. M. (2001). Introduction. In S. M. Fisch, & R. T. Truglio (Eds.), *G is for growing – Thirty years of research on children and Sesame Street* (pp. xv–xxi). Mahwah, NJ: Lawrence Erlbaum.
- UNESCO (2006). Strong foundations – Early childhood care and education. *Education for All Global Monitoring Report 2007* (Retrieved from <http://www.unesco.org/new/en/education/themes/leading-the-international-agenda/efareport/reports/2007-early-childhood/> September 10, 2011)
- UNESCO (2011). The hidden crisis: Armed conflict and education. *EFA Global Monitoring Report*. Paris, France: UNESCO Publishing (Retrieved from <http://unesdoc.unesco.org/images/0019/001907/190743e.pdf>)
- UNESCO Institute of Statistics (2011a). Education (all levels) profile – Bangladesh. Retrieved from http://stats.uis.unesco.org/unesco/TableViewer/document.aspx?ReportId=121&IF_Language=en&BR_Country=500
- UNESCO Institute of Statistics (2011b). Education (all levels) profile – Egypt. Retrieved from http://stats.uis.unesco.org/unesco/TableViewer/document.aspx?ReportId=121&IF_Language=eng&BR_Country=2200
- UNESCO Institute of Statistics (2011c). Education (all levels) profile – India. Retrieved from http://stats.uis.unesco.org/unesco/TableViewer/document.aspx?ReportId=289&IF_Language=eng&BR_Country=3560
- UNESCO Institute of Statistics (2011d). Education (all levels) profile – Indonesia. Retrieved from http://stats.uis.unesco.org/unesco/TableViewer/document.aspx?ReportId=121&IF_Language=en&BR_Country=3600
- UNESCO Institute of Statistics (2011e). Education (all levels) profile – South Africa. Retrieved from http://stats.uis.unesco.org/unesco/TableViewer/document.aspx?ReportId=289&IF_Language=eng&BR_Country=7100
- UNICEF Innocenti (2011). The debate – Early childhood development. A debate with Nobel Laureate Prof. J. Heckman, the Editor-in-Chief of The Lancet, Dr. R. Horton, and the Chief of UNICEF's ECD Unit, Dr. N. Ulker. Uploaded by UNICEF Innocenti. Retrieved from http://www.youtube.com/watch?v=vhxDGC6_g64
- US Department of State (2012). Background note: Kosovo. Retrieved from <http://www.state.gov/r/pa/ei/bgn/100931.htm>
- Vandewater, E. A., & Bickham, D. S. (2004). The impact of educational television on young children's reading in the context of family stress. *Applied Developmental Psychology*, 25, 717–728. <http://dx.doi.org/10.1016/j.appdev.2004.09.006>.
- Walker, S. P. (2011). Promoting equity through early childhood development interventions for children from birth through three years of age. In H. Alderman (Ed.), *No small matter: The impact of poverty, shocks and human capital investment in early childhood development* (pp. 115–154). Washington, DC: The World Bank.
- Walker, S. P., Wachs, T. D., Gardner, J. M., Lozoff, B., Wasserman, G. A., Pollitt, E., et al. (2007). Child development: Risk factors for adverse outcomes in developing countries. *The Lancet*, 369, 145–157. [http://dx.doi.org/10.1016/S0140-6736\(07\)60076-2](http://dx.doi.org/10.1016/S0140-6736(07)60076-2).
- Walker, S. P., Wachs, T. D., Grantham-McGregor, S., Black, M. M., Nelson, C. A., Huffman, S. L., et al. (2011). Inequality in early childhood: Risk and protective factors for early child development. *The Lancet*, 378, 1325–1338. [http://dx.doi.org/10.1016/S0140-6736\(11\)60555-2](http://dx.doi.org/10.1016/S0140-6736(11)60555-2).
- What Works Clearinghouse (2011). *Procedures and standards handbook (version 2.1)*. US Department of Education Institute of Education Sciences (Retrieved from <http://ies.ed.gov/ncee/wwc/documentsum.aspx?sid=19>)
- World Bank (2011). *Country analytical history operational category change*. Retrieved from Siteresources.worldbank.org/DATASTATISTICS/resources/OGHIST.XLS.
- Wright, J. C., Huston, A. C., Murphy, K. C., St. Peters, M., Pinon, M., Scantlin, R., et al. (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, 1347–1367. <http://dx.doi.org/10.1111/1467-8624.t01-1-00352>.
- Zill, N., Davies, E., & Daly, M. (1994). *Viewing of Sesame Street by preschool children in the United States and its relation to school readiness*. Rockville, MD: Westat.

Appendix A. List of reports providing effects and reach data by country

Australia

Lemercier, K.I., & Teasdale, G.R. (1973). "Sesame Street": Some effects of a television programme on the cognitive skills of young children from lower SES backgrounds. *Australian Psychologist*, 8, 47–51.

Bangladesh

Khan, M.S.H., Chakaborty, N., Rahman, A.P.M., & Nasrin, T. (2007). *2007 follow-up (wave II) evaluation of the reach and impact of Sisimpur: A technical report*. Bangladesh: Associates for Community and Population Research (ACPR).

Research and Computing Services (2006). *The educational impact of Sisimpur: Results of an experimental study of children's learning*. Bangladesh: Research and Computing Services.

Canada

Ellis, E.N., Reid, M.J., & Hoen, R. (1972). The impact of "Sesame Street" on primary pupils in Vancouver. *Research report 72–16*. Vancouver: Dept of Planning and Evaluation. Retrieved from <http://www.eric.ed.gov/>

Egypt

Rimal, R.N., Figuero, M.E. & Federowicz, M. (2006). *Impact of Alam SimSim among 4- to 6-year olds in Egypt: Effects on math ability, literacy skills, and gender attitudes*. John Hopkins University and El-Zanaty & Associates. Unpublished research report prepared for Sesame Workshop.

India

GyanVriksh Technologies (2008). *Experimental design undertaken to understand the impact of Galli Galli Sim Sim*. Unpublished research report prepared for Sesame Workshop.

GyanVriksh Technologies (2010). *The reach and impact of the Galli Galli Sim Sim television show in India: End-line report of a naturalistic longitudinal study*. Unpublished research report prepared for Sesame Workshop.

Indonesia

Borzekowski, D., & Wahyuningrum, Y. (2009). *The educational impact of Jalan Sesama on Indonesian preschool children: An experimental study*. John Hopkins Bloomberg School of Public Health. Unpublished research report prepared for Sesame Workshop.

Borzekowski, D. (2010). *The educational impact of Jalan Sesama: Evidence of children's learning from a longitudinal study*. John Hopkins Bloomberg School of Public Health. Unpublished research report prepared for Sesame Workshop.

* Synovate (2009). *Survey of reach and perception of Jalan Sesama (wave II, 2009)*. Unpublished research report prepared for Sesame Workshop.

Israel and Palestine

Cole, C.F., Arafat, C., Tidhar, C., Tafesh, W.Z., Fox, N.A., Killen, M., Ardila-Rey, A., Leavitt, L.A., Lesser, G., Richman, B.A., & Yung, F. (2003). The educational impact of *Rechov Sumsum/Shara'a Sim sim: A Sesame Street television series to promote respect and understanding among children living in Israel, the West Bank and Gaza*. *International Journal of Behavioral Development*, 27, 409–422.

Walden, T., Arafat, S.H., Yahia, G.H., Sandbank, A., Wertheim, C., & Zimmerman, H. (2009). *The contribution of initiated viewing of Rechov Sumsum to the social perceptions of kindergarten-aged children*. Beit Berl Academic College. Unpublished research report prepared for Sesame Workshop.

Kosovo

Fluent (2008). *Assessment of educational impact of Rruga Sesam and Ulica Sezam in Kosovo: Report of findings prepared for Sesame Workshop.*

* Prism (2005). *Reception and reach of Rruga Sesam and Ulica Sezam in Kosovo.* Unpublished research report prepared for Sesame Workshop.

Mexico

Diaz-Guerrero, R., & Holtzman, W.H. (1974). Learning by televised "Plaza Sesamo" in Mexico. *Journal of Educational Psychology*, 66, 632–643.

Nigeria

Sesame Street Nigeria (2009). *Sesame Street Nigeria pilot child study: Findings from informal assessment of children.* Unpublished research report prepared for Sesame Workshop.

Northern Ireland

Connolly, P., Kehoe, S., Larkin, E. and Galanouli, D. (2008) *A cluster randomised controlled trial evaluation of the effects of watching Sesame Tree on young children's attitudes and awareness*, Belfast: Centre for Effective Education, Queen's University Belfast.

Larkin, E., Connolly, P., and Kehoe, S. (2009) *A Longitudinal study of the effects of young children's natural exposure to Sesame Tree on their attitudes and awareness (report 2)*, Belfast: Centre for Effective Education, Queen's University Belfast.

Larkin, E., Connolly, P., and Kehoe, S. (2009) *A cluster randomised controlled trial evaluation of the effects of the Sesame Tree schools outreach pack on young children's attitudes and awareness (report 3)*, Belfast: Centre for Effective Education, Queen's University Belfast.

South Africa

* AC Nielsen Summary report of reach and awareness of Takalani Sesame and HIV/AIDS knowledge. Unpublished research report prepared for SABC Education.

Human Sciences Research Council (2003). *Impact assessment of Takalani Sesame Season One.* Unpublished research report prepared for SABC Education.

Khulisa (2005). *Impact assessment of Takalani Sesame Season II programme.* Unpublished research report prepared for Sesame Workshop.

Tanzania

Borzekowski, D., & Macha, J. (2009). *An evaluation of the impact of Kilimani Sesame's health messaging in Tanzania.* John Hopkins Bloomberg School of Public Health. Unpublished research report prepared for Sesame Workshop.

Turkey

Sahin, N. (July, 1992). *Preschoolers' learning from educational television.* Paper presented at the XXVth International Congress of Psychology, Brussels, Belgium.

* indicates a report that was only used for the reach statistics presented in Table 6.

Appendix B. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.appdev.2013.01.001>.