MAX PLANCK INSTITUTE
FOR HUMAN DEVELOPMENT

This paper was originally published by Sage as:
Van Steensel, R., McElvany, N., Kurvers, J., & Herppich, S. (2011). **How effective are family literacy programs? Results of a meta-analysis**. *Review of Educational Research*, *81*(1), 69–96. https://doi.org/10.3102/0034654310388819

This publication is with permission of the rights owner freely accessible due to an Alliance licence and a national licence (funded by the DFG, German Research Foundation) respectively.

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, nontransferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public. By using this particular document, you accept the above-stated conditions of use.

Provided by:

Max Planck Institute for Human Development
Library and Research Information
library@mpib-berlin.mpg.de

Check for updates

Review of Educational Research
March 2011, Vol. 81, No. 1, pp. 69–96
DOI: 10.3102/0034654310388819
© 2011 AERA. http://rer.aera.net

How Effective Are Family Literacy Programs? Results of a Meta-Analysis

Roel van Steensel
University of Amsterdam

Nele McElvany
Technische Universität Dortmund

Jeanne Kurvers
Tilburg University

Stephanie Herppich
University of Kiel

*This meta-analysis examines the effects of family literacy programs on children's literacy development. It analyzes the results of 30 recent effect studies (1990–2010), covering 47 samples, and distinguishes between effects in two domains: comprehension-related skills and code-related skills. A small but significant mean effect emerged (*d *= 0.18). There was only a minor difference between comprehension- and code-related effect measures (*d *= 0.22 vs.* d *= 0.17). Moderator analyses revealed no statistically significant effects of the program, sample, and study characteristics inferred from the reviewed publications. The results highlight the need for further research into how programs are carried out by parents and children, how program activities are incorporated into existing family literacy practices, and how program contents are transferred to parents.*

Keywords: meta-analysis, literacy, parents, families.

Literacy development is a major goal of education and one of the fundamental prerequisites for academic success and participation in modern society (Snow, Burns, & Griffin, 1998). As recent international studies have shown, however, a considerable number of students systematically lag behind their peers in literacy skills (Mullis, Martin, Gonzales, & Kennedy, 2003; Organisation for Economic Co-operation and Development, 2001). One way of working to prevent delays in children's literacy development is by means of extending and improving their literacy experiences outside school through so-called family literacy programs.

Hannon (2003) defined family literacy programs as "programmes to teach literacy that acknowledge and make use of learner's family relationships and engagement in family literacy practices" (p. 100). This definition implies that family literacy programs can include a broad array of activities. Various researchers have attempted to classify these. One of the first was Nickse (1989, 1991), who suggested a taxonomy based on two dimensions: type of intervention (direct or indirect) and type of participant (adult or child). She distinguished among programs that provide direct instruction for both parents and children (i.e., adult education combined with center- or classroom-based activities for children), programs in which no direct instruction is provided but where services have an informal nature, targeting parents' and children's skill development only indirectly (e.g., library activities such as read-alouds), programs that provide direct instruction for parents and target children indirectly (e.g., adult education programs that also offer parenting support), and programs that provide direct instruction for children and target parents indirectly (e.g., classroom-based literacy interventions with take-home activities). Morrow and Paratore (1993) proposed a more limited typology and distinguished between two categories of programs: home–school partnerships and intergenerational programs. The former concern programs designed to involve parents in literacy activities and events that support school-based goals, whereas the latter have the broader goal of improving the literacy development of both parents and children by means of systematic instruction (either to parents and children separately or combined). Cairney (2002) added so-called partnership programs to this classification, that not only intend to reinforce the ties between parents and schools but also involve the broader communities of which they are part. More recently, Sénéchal and Young (2008) distinguished among school-based involvement, which concerns parental activities in the school environment, home–school conferencing, which involves, for example, parent–teacher communication about children's literacy development, and home-based involvement, where parents provide literacy-learning activities at home.

Given the plethora of possible activities illustrated by these taxonomies, we decided to focus the current study on one specific category of programs, namely, those that target children's literacy development directly by providing stimulating parent–child activities to be carried out at home and by training parents to transfer the contents embedded in these activities; this resembles Sénéchal and Young's (2008) category of home-based involvement. The rationale for such programs stems from various strands of research. The research base first of all comprises studies that highlight the strong relationship between school literacy development and family variables such as socioeconomic status (SES) and the stimulating effects of parental involvement in education, particularly for disadvantaged groups (for overviews, see Hannon, 2003; Nickse, 1989; Purcell-Gates, 2000). Equally important for the development of family literacy programs has been the research into children's emergent literacy development (Lancy, 1994; Teale & Sulzby, 1986; Whitehurst & Lonigan, 1998), which has shown that many children develop a profound knowledge of the functions and forms of written language before school, as the result of observing and participating in stimulating reading and writing activities at home. These emergent literacy skills provide an important basis for school-related literacy learning.

The main conclusion from this body of literature—that families are important contexts for literacy learning—formed the basis for the development of interventions that have as their primary goal to teach parents, mostly those of children who are likely to experience literacy delays because of risk factors such as low SES or low parental literacy levels, to incorporate stimulating literacy practices in their homes (Purcell-Gates, 2000). Family literacy programs are thought to have several advantages over, for example, literacy interventions in the educational context (Hannon, 1995; Kağıtçıbaşı, 1999; McElvany & Artelt, 2009; McElvany & van Steensel, 2009; Van Tuijl, Leseman, & Rispens, 2001). First, family-based interventions in principle provide ample opportunity for one-to-one teaching and learning interactions between parent and child and thus for intensive practice and individual feedback. Second, as these programs aim to make permanent, positive changes in the routines of family life, they can be expected to promote literacy skills for the long term. Third, capitalizing on the family as the primary context of intervention increases these programs' sensitivity to the social and cultural conditions of child development, which may be particularly relevant when family and school cultures differ.

Although the arguments for the effectiveness of family-based programs are compelling, their potential is met by some serious challenges. Based on an in-depth investigation of two family interventions, McElvany and van Steensel (2009) identified and elaborated three dimensions of implementation quality issues of family literacy interventions: (a) intensity and quality of parent–child interactions, (b) intensity and quality of the support and training provided for parents, and (c) selectivity of participation. These areas provide challenges to the way programs are conducted, and thus to their effectiveness, in several ways (also see McElvany, 2008). First of all, although these programs aim to foster parents' didactic abilities, they may presuppose skills and knowledge that are not necessarily present (particularly in disadvantaged parents) and that cannot be developed fully simply by participating in the intervention. Second, the relationship between parents and child is sensitive and emotional and may be disrupted by the pressures arising from a teaching–learning situation (also see Grolnick, 2003). Third, the busy schedules of family life may interfere with both participation in training and intervention time. Finally, the conditions of parent training may hinder the transfer of program contents, for example, because training is conducted by paraprofessional volunteers. These may have difficulty in grasping the theoretical basis of the program or in showing flexibility in matching the theory to practical situations (Gray & Wandersman, 1980).

There is as yet no generalizable answer to the question of whether family literacy programs are effective. One reason is that many of the reviews published to date are very broad in scope and do not permit specific conclusions to be drawn about the impact of family literacy programs (Blok, Fukkink, Gebhardt, & Leseman, 2005; Erion, 2006; Fishel & Ramirez, 2005; Mattingly, Prislin, McKenzie, Rodriguez, & Kazyar, 2002; White, Taylor, & Moss, 1992). Some of these reviews have reported encouraging results regarding the effects of specific forms of parent involvement on academic achievement (Erion, 2006; Fishel & Ramirez, 2005). Others are less optimistic, especially when family-based activities are compared to center-based activities (Blok et al., 2005; White et al., 1992). All

of these syntheses, however, examined a relatively large variety of parent- or family-centered interventions—beyond family literacy programs they covered, for example, parental volunteering at school (Fishel & Ramirez, 2005) and coaching of (general) parenting skills (Blok et al., 2005). They also included a variety of outcome measures beside those relating directly to literacy, usually combined into one or two general categories (e.g., cognitive skills; Blok et al., 2005). This issue is often referred to as the *apples-and-oranges problem* of meta-analysis (Kulik & Kulik, 1989; Lipsey & Wilson, 2001).

Few meta-analyses have focused solely on family literacy programs. Recently, Sénéchal and Young (2008) synthesized the results of studies investigating the effects of family literacy interventions on reading acquisition in kindergarten and primary school. They found a substantial overall effect of interventions: the weighted mean effect size was $d = 0.65$. In the same year, Mol, Bus, De Jong, and Smeets (2008) summarized the outcomes of studies examining the effects of a specific shared reading program (Dialogic Reading) on the vocabulary skills of children in preschool and kindergarten. They found a moderate mean effect size of $d = 0.42$. Although these meta-analyses provide valuable and specific insights into the effects of family literacy interventions, some aspects warrant critical appraisal. First, both meta-analyses used a database of 16 primary studies—a rather small sample size, particularly when the studies are split into subsamples in the case of moderator analyses (a claim that is supported by the relatively large confidence intervals). Second, some features of the two reviews make it difficult to draw conclusions about the general effects of the family literacy approach. Mol et al. (2008) focused on the effects of a very specific form of family literacy interventions—Dialogic Reading—which is usually considered appropriate for only (very) young children. Moreover, they focused exclusively on the intervention's influence on vocabulary, disregarding any potential (lack of) effects on other outcomes. Sénéchal and Young (2008), on the other hand, excluded vocabulary from their effect measures. Although this decision was based on a sound theoretical model of (emergent) literacy (Sénéchal, LeFevre, Smith-Chant, & Colton, 2001), it may have increased the risk of missing important results. Finally, neither of the studies accounted for the possibility that the moderator variables under investigation were correlated. In other words, it is possible that the effects found for certain variables were confounded with the effects of other variables.

These observations prompted us to conduct a new meta-analysis, differing from previous reviews in several respects. First, the meta-analysis presented in this article is specific in the sense that it focuses on family literacy interventions as defined previously but comprehensive in the sense that it does not concentrate on particular program types or exclude certain effect measures. Second, it aims to paint a more detailed picture of program effects—and to minimize the apples-and-oranges problem—by distinguishing two categories of effect measures. Specifically, we distinguished between effects in the domains of *comprehension-related skills* and *code-related skills*. This distinction was informed not only by theoretical models of emergent literacy development (Storch & Whitehurst, 2002; Whitehurst & Lonigan, 1998) that distinguish between oral language skills (i.e., semantic, syntactic, and conceptual abilities) and code-related skills (abilities necessary for deciphering the written language code) but also by process models of reading that conceptualize the reading process as a constant interaction between

lower-order decoding skills and higher-order comprehension skills (Hoover & Gough, 1990; Kintsch, 1998; LaBerge & Samuels, 1974; Stanovich, 1980). Third, our analysis includes a relatively large set of studies (30 studies covering 47 samples), making it possible to draw more robust conclusions.

Our first aim was to analyze the overall effect of family literacy programs and to examine whether outcomes differed for comprehension- and code-related effect measures. As variability in the effectiveness of the programs was to be expected, our second aim was to identify the variables (moderators) that determined these differences. We distinguished among three types of variables: program characteristics, sample characteristics, and study characteristics.

With respect to program characteristics, interventions may first of all differ in the types of activities offered (for an overview, see Hannon, 2003). For example, Sénéchal and Young (2008) distinguished between shared reading programs—comprising programs in which parents read to their child as well as programs in which parents listen to their child read—and programs in which parents tutor specific skills. They found the latter programs to yield the largest effect sizes. Second, programs can differ in the training and support provided for parents (Hannon, 2003). Training can be provided at home or in centers or schools, it can be delivered by professional educators, semi- or paraprofessionals, or volunteers, and it can be complemented by the provision of resources (such as storybooks). Some researchers have found indications for effects of these variables. In a review of 13 more general parent involvement programs, for example, Olds and Kitzman (1993) found that most programs using professional parent trainers had positive effects on children's cognitive and language development, whereas most programs using semiprofessionals were not effective. Third, programs effects may be influenced by duration, although there is as yet no clear evidence that longer interventions are more effective than shorter ones. Both Blok et al. (2005) and Sénéchal and Young (2008) conducted moderator analyses including program duration but found no evidence for effects.

Family (literacy) interventions may be directed toward different target groups (Gomby, Culross, & Behrman, 1999; Hannon, 2003). Two sample characteristics have been of particular interest in previous meta-analyses: the at-risk status of the targeted children and their age. The former is generally indicated by parental education, job level, or income (Blok et al., 2005; Mol et al., 2008; Sénéchal & Young, 2008) or by preintervention reading level (Sénéchal & Young, 2008). Findings on the effects of these two variables are inconclusive, however. Sénéchal and Young (2008) found no difference in effects between children from lower SES families and children from higher SES families, between children who read at a normal level and children who read below the normal level, or between kindergarteners and primary schoolers. In contrast, Mol et al. (2008) reported larger program effects for non-at-risk children than for at-risk children, and they found larger effects for preschoolers than for kindergarteners.

One of the most important study characteristics to be considered is sample selection procedure. Intervention studies can be either experimental or quasi-experimental: in the former participants are randomly assigned to program and control groups, in the latter they are not (Raudenbush, 2005). Lipsey (2003) demonstrated that nonrandomization leads to inflated effect sizes, which may, in the case of family programs, be from volunteer effects in self-selected program groups. Some of the

reviewers mentioned above were able to examine the influence of sample selection on program effects. Sénéchal and Young (2008) compared the effect sizes of 12 randomized studies to those of 4 nonrandomized studies but found no differences. In their review of (general) parent involvement programs, in contrast, Fishel and Ramirez (2005) concluded that studies with effective methodology (including randomization) "failed to demonstrate significant change in child outcomes, and studies with large effect sizes had flawed methodology" (p. 393). Other relevant study characteristics include pretesting and time of measurement (Blok et al., 2005; Sénéchal & Young, 2008). Blok et al. (2005) found a negative effect for the latter: Effect sizes were smaller in the case of follow-up tests than in the case of immediate posttests. The authors attributed this finding to a fading out effect. Sénéchal and Young (2008), in contrast, found no effect of measurement time.

Research Questions

Given the importance attributed to promoting literacy via family literacy programs and the inconclusive scientific evidence in this field, the present study aimed at summarizing recent research on the effectiveness of these programs. The main research questions guiding our analyses were as follows:

1. Do family literacy programs positively affect children's literacy skills?
2. Do the effects of these programs vary with the type of effect measure, that is, is there a difference between comprehension- and code-related effect measures?
3. To what extent are program effects moderated by (a) program characteristics, (b) sample characteristics, and (c) study characteristics?

To address the first research question, we combined the results of a set of studies on family literacy programs by aggregating the data reported on program effects in the form of Cohen's *d* effect sizes to produce a single overall mean. Effect sizes represent the magnitude and direction—positive or negative—of a program effect. To answer the second question, we further categorized the effect sizes on the basis of the theoretical distinction between comprehension- and code-related effect measures. With respect to the third question, we drew on the literature described above to identify and examine the effects of the following moderators:

a. Program characteristics: types of parent–child activities offered, delivery of parent training, staff quality, duration. In addition, we considered two further variables. First, given our distinction between comprehension- and code-related effect measures, we also distinguished between programs focusing on the former versus the latter skills. Second, we examined the effects of family literacy activities complemented by a similar program conducted at school or in a center because we found such combined activities to occur.
b. Sample characteristics: at-risk status, age group.
c. Study characteristics: sample selection, pretesting, time of measurement.

Method

Search Strategy and Study Selection Criteria

We conducted a computer-assisted search of the PsycINFO and ERIC databases using the following groups of keywords: (a) *program*, *intervention*, *training*; (b) *home*, *family*, *parents*; and (c) *literacy*, *reading*. Each term in the first group was combined with each term in the second and third groups, resulting in 18 combinations (e.g., *program*, *home*, *literacy*; *program*, *home*, *reading*; etc.). The searches were restricted to studies published in peer-reviewed journals between January 1990 and April 2010. The peer-review criterion was chosen to maximize the possibility of finding methodologically sound studies (consistent with, for example, the meta-analytic methodology followed by the National Reading Panel and the National Early Literacy Panel; see Schatschneider, Westberg, & Shanahan, 2008). The searches were done in two stages. In the first stage of our study we searched for studies published between 1990 and 2007; this search yielded 1,884 PsycINFO hits and 1,315 ERIC hits. In the second stage we updated our analysis with studies published in the period until April 2010. This yielded an additional 1,093 hits from PsycINFO and 777 hits from ERIC. The titles and abstracts were screened by three researchers, who also did the coding. Studies were selected for closer examination if the abstract indicated relevance to the topic of the meta-analysis. A study was added to the sample if it met the following criteria:

1. The study measured the effects of a family literacy program on children's literacy skills.
2. The participants were preschoolers, kindergarteners, and/or primary school children.
3. The study compared an experimental group that participated in the intervention to a control group that did not.
4. The study provided effect sizes (Cohen's *d*) or information (means, standard deviations, results of statistical testing, program and control *n*s) allowing effect sizes to be calculated.
5. The effect measures reported were indicators of literacy skills that could be categorized as either comprehension-related or code-related. In two studies (Miller & Kratochwill, 1996; Saracho, 1997), a combined measure was used. The outcome of this measure was included, but only in the analysis of the general literacy ability effect sizes (see below).
6. The total sample size was at least 10.

There was one exception to Criterion 3. In the case of combined family and center or school interventions, we only included those studies in which a comparison was made between the combined condition and a condition in which children had participated only in the center or school intervention. This was necessary to get a precise estimate of the effect of the family component by partialling out the effect of the center or school component. This also implies that studies in which a combined condition was compared to a no-treatment control group (we came across five such studies) were not included in the sample.

Studies focusing on specific target groups (e.g., children with learning disabilities, children with speech or language impairments, children with emotional or behavioral disorders, and children with mental retardation) were excluded from the meta-analysis.

The final sample consisted of 30 studies covering a total of 4,326 children or families (1,866 who had received treatment and 2,460 who had not).

Coding Procedure

All studies were coded according to a standardized coding scheme with the following sections: article information, program characteristics, sample characteristics, study characteristics, and program effects. These data were then entered in an SPSS data file.

The article information section covered the title of the article, the name(s) of the author(s), the name of the journal, and the year of publication.

The program characteristics section recorded the name of the program as well as data on the following program variables: type of parent–child activities, program focus, delivery of parent training, staff quality, location, and duration. The type of activities variable concerns the type(s) of parent–child activities that are offered in the program: We distinguished among programs that offered only shared reading activities, programs that offered shared reading plus other types of activities, and programs that offered what can be termed "literacy exercises," that is, training of specific skills. The program focus variable reflects whether the emphasis in the program was on comprehension-related skills, code-related skills, or both. Delivery of parent training related to whether parents were trained and supported by means of home visits and/or group meetings and whether parental support was complemented by the provision of resources (more specifically, books). Staff quality refers to the question of whether parents were trained by professionals, semiprofessionals, or both. Location reflects whether the program was conducted solely in the children's homes or also had a center- or school-based component, that is, whether the parent–child activities were paralleled by activities in preschool or day care centers or at school. In terms of duration, we distinguished between programs lasting up to half a school year (<5 months) and programs lasting half a school year or more. This distinction was based on the distribution of durations in the sample. Appendix A provides more detailed information on the way the categorization of program characteristics was operationalized.

The sample characteristics section recorded the at-risk status of the participating children, that is, whether or not they were at risk of delays in educational and/or literacy development. In all but one case at-risk status was based on socioeconomic indicators (such as parental education and family income), sometimes complemented by cultural or linguistic indicators (Did families belong to cultural and/or linguistic minority communities?) or developmental indicators (such as the score on a standardized test). In one case the children's at-risk status was based on their families' history of reading failure. In addition, we categorized the children's age group, distinguishing between preformal education (preschool, kindergarten) and formal education (Grade 1 and beyond). This distinction was based on the fact that formal reading instruction generally does not start before Grade 1.

The study characteristics section coded whether or not the program and control groups were formed via randomization. We coded studies as having applied

randomization only if this was done at the level of individual participants. Studies in which groups (e.g., classes) were randomly assigned to program and control conditions were placed in the no randomization category. Studies in which individuals or groups were not randomly assigned but were matched were also placed in the latter category. In addition, we recorded whether or not there was a pretest and when the posttest was administered (short term or follow-up).

Finally, the program effects section recorded the effect sizes, either taken directly from the articles or calculated following the guidelines provided by Lipsey and Wilson (2001; also see below). The effect measures were coded as reflecting either comprehension-related skills or code-related skills. The former category covered knowledge and skills such as (active or receptive) vocabulary, narrative comprehension, reading comprehension, story (re)telling, and story (re)writing, whereas the latter covered emergent literacy skills such as letter identification, concepts about print, visual matching or discrimination, phonological awareness, initial or final consonant recognition, rhyme recognition, and more formal skills such as alphabet knowledge, word reading, reading rate, accuracy, fluency, and spelling.

As indicated above, there were three coders, each of whom coded one third of the studies. Before coding, each coder received the coding scheme and a detailed description and explanation of the information to be extracted and interpreted. Each coder also coded half of the studies initially coded by one of the other coders (these studies were selected at random). The level of interrater agreement ranged from 73% to 91%. Discrepancies were discussed until a consensus was reached. Based on the conclusions of these discussions, the coders checked all of their original codings, discussed cases of doubt, and made any necessary changes.

Calculating Effect Sizes

The 30 studies reported 152 measures for which effect sizes could be established. If effect sizes were given in the article, these were used in the analyses. This was the case for 18 measures. If effect sizes were not given, they were calculated using one of the formulas presented in Appendix B (taken from Lipsey & Wilson, 2001). All effect size calculations were checked using Wilson's Effect Size Determination Program (Wilson, 2001). Note that two or more formulas could be applicable within one study.

Aggregating and Weighting Effect Sizes

To avoid statistical dependency of observations, a meta-analysis should not include more than one effect size per construct per sample (Lipsey & Wilson, 2001). We therefore aggregated the effect sizes by means of averaging. On the basis of our research questions, we distinguished two levels of aggregation. The central dependent variable in Research Question 1 is literacy. All effect sizes in each sample were therefore averaged to produce a single general literacy ability effect size per sample. Effect sizes for comprehension- and code-related measures were then averaged separately to address Research Question 2. In many cases, this procedure resulted in one aggregated comprehension-related effect size and one aggregated code-related effect size per sample.

Aggregations were made at the sample level rather than at the study level. These levels are not necessarily the same. Some studies included more than one sample,

for example, because two groups in different locations or times or of different ages participated in the program. Other studies examined the effects of different versions of the same program, using different samples. In cases where the same group of children was assessed more than once, in a short-term and a follow-up assessment, the study was coded as having two (or more) samples to be able to determine the effect of time of measurement. A similar procedure was used by Blok et al. (2005).

The effect sizes of some studies were disproportionately low or high. Following Lipsey and Wilson (2001), effect sizes more than two standard deviations (2 × 0.50) from the mean of all 152 initial effect sizes (0.27) were reduced to either the mean plus two *SD*s, that is, 1.27, or the mean minus two *SD*s, that is, –0.73.

Study samples naturally differ in size. In this meta-analysis, sample sizes ranged from 15 to 781. To prevent studies with very small sample sizes having a disproportionate influence on our findings, we weighted the effect sizes by the inverse of their sampling error variance.

Checking for Publication Bias

Although there are valid reasons for doing so, limiting the selection to studies published in journals may lead to unwanted effects. It has been shown that published studies have larger mean effect sizes than unpublished studies (Lipsey & Wilson, 1993), implying that researchers tend to leave studies showing no or limited effects "in the file drawer."

One way of checking for this file-drawer effect is by computing the *fail-safe N*, that is, the number of unpublished studies reporting no results needed to reduce the weighted mean effect size to the point of nonsignificance (Lipsey & Wilson, 2001; Orwin, 1983). With a criterion effect size of 0.10 (for no effect), the fail-safe *N* for this meta-analysis is 38, indicating that 38 unpublished studies (or, more precisely, samples) showing no effects would be needed to reduce the weighted mean to 0.10.

Another way of detecting publication bias is by generating a *funnel plot* (Sterne, Egger, & Davey Smith, 2001; Thornton & Lee, 2000), in which the treatment effects estimated in individual studies are plotted against some measure of precision. Visual inspection of the funnel graph (see Appendix C) shows a rather symmetrical, funnel-shaped distribution around the (weighted) mean effect, which implies there is no real indication of publication bias. This conclusion was corroborated by the result of Egger's linear regression test for asymmetry (Egger, Davey Smith, Schneider, & Minder, 1997): The value of the intercept did not deviate significantly from zero ($p = .516$).

Data Analysis

Data analysis was conducted in SPSS and Microsoft Excel using the procedures provided by Lipsey and Wilson (2001). The analysis procedure comprised three steps. First, the weighted mean effect size was calculated. Second, a homogeneity analysis was conducted to examine whether all effect sizes in the sample estimated the same population effect or whether there was excess variability that needed to be explained. Third, moderator analyses were conducted to examine the relations between program effects and program, sample, and study characteristics.

Results

We begin this section by describing the characteristics of the programs included in the meta-analysis and then present the meta-analysis itself.

Analysis of Program Characteristics

Information on each separate study is provided in Table 1. The 30 studies covered 21 different programs, of which 5 were evaluated more than once. Dialogic Reading was the subject of 5 studies (Studies 3, 4, 9, 13, and 30). Paired Reading was evaluated in 3 studies (Studies 15, 18, and 19). Three programs were each evaluated twice: HIPPY (Studies 2 and 28), Project PRIMER (Studies 5 and 6), and Project EASE (Studies 11 and 23). The remaining 16 programs were the subject of one study each. In 3 studies different versions of the same program were evaluated. In two cases (Studies 4 and 13), a home-based only program was compared to a home- and classroom-based program; in one case (Study 20) the difference in program versions was based on the materials used: One version was based on children's literature books and one on basal reading materials from participating children's classrooms.

In this section we further discuss the distribution of program characteristics as well as the relationships among the program characteristics and between the program and sample characteristics; the latter are presented only if a χ^2 test proved a relationship to be statistically significant. Since the same program can be implemented in different ways, we decided to take as the basis for our comparison the number of programs or program versions as they were offered in each specific study. This decision resulted in 32 (instead of 30) cases.

In most cases ($n = 18$) the programs involved shared reading plus other types of literacy activities. In 9 cases the interventions provided only shared reading, whereas in 5 cases they involved only literacy exercises. Programs targeted comprehension-related skills and both comprehension- and code-related skills in equal numbers of cases (both $n = 13$), whereas in only 6 cases programs focused solely on code-related skills. In the vast majority of cases ($n = 22$) parents were trained and supported by professionals, whereas training provided by semiprofessionals only or by both professionals and semiprofessionals appeared in relatively few cases ($n = 4$ and $n = 2$, respectively; 4 missing cases). In most cases program delivery occurred through group meetings ($n = 26$); in relatively few programs trainers visited families in their homes ($n = 9$). Interestingly, in 5 cases both group meetings and home visits were provided. Additional resources, mostly in the form of books, were provided in 24 of the cases. In nearly all cases the programs consisted of home-based activities only ($n = 28$); in only 4 cases did programs provide both home- and classroom-based activities. Programs with a shorter and a longer duration were about equally distributed across the cases ($n = 18$ and $n = 14$, respectively). In 17 cases the programs targeted at-risk families, whereas in 14 cases they targeted non-at-risk families (1 missing case). Finally, in about half of the cases the programs were intended for children in the preformal education phase ($n = 15$). In 16 cases they targeted children in the formal education phase, and in one case they targeted both categories.

Activity type and program focus were related, $\chi^2(4) = 30.93$, $p < .001$: The shared-reading-only programs focused either on comprehension skills only ($n = 7$)

TABLE 1
Description of the studies included in the meta-analysis

No.	Author 1	Year	Name	Act.	Focus	Staff	Delivery	Loc.	Dur.	At risk	Age	Sample	Pretest	Time	n_p; n_c	Mean ES (all/Cm/Cd)
1	Allen	2007	PCHP	SR+	Cm	Semi	V/B	Home	≥5	Y	Pre	NRA	N	F/U	63/77; 42/56	–0.35†; –0.38†; –0.33†
2a	BarHava-Monteith	1999	HIPPY	SR+	Both	Semi	V/M/B	Home	≥5	Y	Both	NRA	N	Short	77; 704	0.24†; NA; 0.24†
2b															29; 29	0.33†; 0.43†; 0.22†
3	Chow	2003	DR	SR	Cm	Prof	B	Home	<5	N	Pre	RA	Y	Short	27/29; 27/28	0.15#; –0.18†; 0.47*
4a	Chow	2008	DR	SR	Cm	Prof	M/B	Home	<5	N	Form	RA	Y	Short	37; 36	0.19†; 0.40†; 0.12†
4b				SR+	Both										38; 36	0.22†; 0.39†; 0.16†
5	Cronan	1999	PRIMER	SR+	Cm	Semi	V/B	Home	≥5	Y	Pre	RA	Y	F/U	50; 75	–0.09†; –0.09†; NA
6a	Cronan	1996	PRIMER	SR+	Cm	Semi	V/B	Home	≥5	Y	Pre	RA	Y	Short	14; 14	0.74†; 0.74†; NA
6b															14; 14	0.25†; 0.25†; NA
6c															60; 60	0.42†; 0.42†; NA
6d															60; 60	0.03†; 0.03†; NA
7a	Drouin	2009	—	SR+	Both	Prof	M	Both	<5	N	Pre	NRA	Y/N	Short	30; 18	0.20†; NA; 0.20†
7b														F/U1	30; 18	0.31†; NA; 0.31†
7c														F/U2	30; 18	0.22†; –0.06†; 0.32†
8	Evangelou	2007	PEEP	SR+	Both	—	V/M/B	Home	≥5	Y	Pre	NRA	N	Short	174; 123	0.31*; 0.31*; NA
9a	Fielding-Barnsley	2003	DR	SR+	Both	Semi	V/B	Home	<5	Y	Pre	NRA	N	Short	26; 23	0.50†; 0.58†; 0.48†
9b														F/U	26; 23	0.81†; 0.47†; 0.87†
10	France	1993	CS	LE	Cd	Prof	M	Home	<5	N	Form	NRA	Y	Short	22; 10	–0.73†; NA; –0.73†
11	Jordan	2000	EASE	SR+	Cm	Prof	M/B	Home	≥5	N	Pre	NRA	Y	Short	177; 71	0.34*; 0.64*; 0.20*
12	Kim	2010	—	SR	Cm	—	M/B	Home	<5	Y	Form	RA	Y	Short	110; 108	–0.12†; –0.12†; NA

(continued)

TABLE 1 (continued)

No.	Author 1	Year	Name	Act.	Focus	Staff	Delivery	Loc.	Dur.	At risk	Age	Sample	Pretest	Time	n_p; n_c	Mean ES (all/Cm/Cd)
13a	Lonigan	1998	DR	SR	Cm	Prof	M/B	Home	<5	Y	Pre	RA	Y	Short	7; 16	0.05†; 0.05†; NA
13b															9; 11	0.26†; 0.26†; NA
13c								Both							7; 15	0.03†; 0.03†; NA
13d															10; 16	0.49†; 0.49†; NA
14	McElvany	2009	BPCRP	SR+	Both	—	B	Home	<5	N	Form	NRA	Y	Short	104; 393	0.13*; 0.19*; –0.04*
15	Miller	1996	PR	LE	Cd	Prof	V	Home	<5	Y	Form	NRA	Y	Short	23; 23	–0.19† (combined); NA; NA
16	Morrow	1992	—	SR+	Cm	Prof	M/B	Both	≥5	N	Form	NRA	Y	Short	56; 46	0.22†; 0.22†; NA
17	Morrow	1997	—	SR+	Both	Both	V/M/B	Home	≥5	Y	Form	NRA	Y	Short	28; 28	0.81†; 0.81†; NA
18	Murad	2000	PR	LE	Cd	Prof	M/B	Home	<5	N	Form	RA	Y	Short	22; 24	0.62†; 0.61†; 0.63†
19	Overett	1998	PR	SR	Cm	Prof	M/B	Home	<5	Y	Form	NRA	Y	Short	29; 32	0.49†; 0.48†; 0.50†
20a	Powell-Smith	2000	—	SR	Both	Prof	M/B-1	Home	<5	N	Form	RA	Y	Short	12; 12	–0.30†; NA; –0.30†
20b														F/U	12; 12	–0.30†; NA; –0.30†
20c							M/B-2							Short	12; 12	–0.32†; NA; –0.32†
20d														F/U	12; 12	–0.38†; NA; –0.38†
21	Rasinski	2005	FSR	SR+	Cd	Prof	M	Home	<5	N	Form	RA	Y	Short	15; 15	0.07†; NA; 0.07†
22	Reutzel	2006	WTG	LE	Cd	Prof	M	Home	≥5	Y	Form	NRA	Y/N	Short	66; 77	0.20†; 0.37†; 0.12†
23	Rolla San Francisco	2006	EASE	SR+	Cm	Prof	M/B	Home	≥5	Y	Pre	RA	Y	Short	30; 55	–0.13†; –0.08†; –0.15†
24	Saint-Laurent	2005	—	SR+	Both	Prof	M/B	Home	≥5	N	Form	NRA	Y	Short	53; 55	0.47†; 0.44†; 0.57†
25	Saracho	1997	—	SR+	Cm	Prof	M	Home	≥5	—	Pre	NRA	N	Short	48; 54	0.33†; NA; 0.22†; 0.44† (combined)

(continued)

TABLE 1 (continued)

No.	Author 1	Year	Name	Act.	Focus	Staff	Delivery	Loc.	Dur.	At risk	Age	Sample	Pretest	Time	n_p; n_c	Mean ES (all/Cm/Cd)
26	St. Clair	2006	MEES	SR+	Both	Prof	M/B	Home	≥5	Y	Pre	NRA	Y	F/U	6–14; 11/14	0.89†; 0.74†; 1.03†
27	Sylva	2008	SPOKES	SR+	Both	Prof	V/M/-	Home	<5	Y	Form	RA	Y/N	Short	58; 44–51	0.23†; 0.15#; 0.26#
28a	Van Tuijl	2001	HIPPY/ OO	SR+	Both	Both	V/M/B	Home	≥5	Y	Pre	NRA	Y	Short	122; 59	0.14*; 0.14*; NA
28b														Short	83; 55	–0.18*; –0.18*; NA
29a	Watson	2008	Funnix	LE	Cd	—	M	Home	≥5	N	Pre	NRA	Y	Short	7; 9	1.08†; NA; 1.08†
29b											Form			Short	8; 7	0.38†; NA; 0.38†
30a	Whitehurst	1994	DR	SR	Cm	Prof	M/B	Both	<5	Y	Pre	RA	Y	Short	19; 26	0.22†; 0.22†; NA
30b														F/U	15/16; 23	0.23†; 0.23†; NA

Note. Act. = activity type; Loc. = location; Dur. = duration (in months); n_p = sample size program group; n_c = sample size control group; ES = effect size; Cm = comprehension; Cd = code; SR = shared reading; SR+ = shared reading plus; LE = literacy exercises; Semi = semiprofessional; Prof = professional; V = (home) visits; M = (parent) meetings; B = book provision; Y = yes; N = no; Pre = preformal education; Form = formal education; RA = random assignment; NRA = no random assignment; F/U = follow-up; * = given by the authors; † = calculated by the analysts; # = combination of * and †; NA = not applicable; Abbreviations of the programs: PCHP = Parent-Child Home Program; HIPPY = Home Instruction Programme for Preschool Youngsters; DR = Dialogic Reading; PRIMER = PRoducing Infant/Mother Ethnic Readers; PEEP = Peers Early Education Partnership; CS = Cued Spelling; EASE = Early Access to Success in Education; BPCRP = Berlin Parent–Child Reading Program; PR = Paired Reading; FSR = Fast Start Reading; WTG = Words to Go; MEES = Migrant Education Even Start; SPOKES = Supporting Parents on Kids Education in School; OO = Opstap Opnieuw.

or on both comprehension and code skills (n = 2), the shared-reading-plus programs mostly focused on both comprehension and code skills (n = 11) or on comprehension skills only (n = 6), and the programs involving literacy exercises targeted only code skills (n = 5). In addition, activity type was associated with home visits, $\chi^2(2) = 6.06$, $p = .048$: In all but one of the cases where programs provided home visits, these programs offered a broad range of activities. Activity type was also related to book provision, $\chi^2(2) = 12.29$, $p = .002$: In all but one case where programs provided books, these were shared-reading or shared-reading-plus programs. Finally, there was a relationship between activity type and program duration: The longer programs were mostly those that offered shared reading plus other types of activities, $\chi^2(2) = 10.87$, $p = .004$.

Apart from activity type, program focus was related only to book provision, $\chi^2(2) = 15.71$, $p < .001$: In all but one case where books were provided, the programs targeted comprehension skills or both comprehension and code skills.

Staff quality was first of all associated with home visiting, $\chi^2(2) = 19.09$, $p < .001$: In all cases where programs made use of semiprofessionals in addition to or instead of professionals, home visits were offered; in the majority of cases where professionals were used (20 out of 22) no home visiting occurred. Staff quality was also related to the provision of group meetings, $\chi^2(2) = 10.49$, $p = .005$: In most cases where training was provided by professionals in addition to or instead of semiprofessionals (20 out of 24) group meetings were offered, whereas in most cases where only semiprofessionals were used (3 out of 4) no group meetings were offered. Finally, staff quality was associated with program children's at-risk status, $\chi^2(2) = 6.17$, $p = .046$. Remarkably, in all cases where semiprofessionals were used (n = 6) the programs targeted at-risk children; this was the case for only 9 out of 21 cases where professionals were used.

Apart from activity type and staff quality (see previously), home visiting was related to the provision of group meetings, $\chi^2(1) = 5.43$, $p = .020$, and children's at-risk status, $\chi^2(1) = 10.44$, $p = .001$. Regarding the former, where home visits were provided, more than half of the programs (5 out of 9) also provided group meetings; where group meetings were provided, only about a fifth of the programs also provided home visits (5 out of 26). Regarding the latter, remarkably, in all cases where home visiting was offered (n = 9) at-risk children were the target group; where no home visits were provided this was the case for only 8 out of 22 programs.

Meta-Analysis

Weighted mean effect size. Our first aim was to establish the overall effectiveness of the programs under investigation and to examine whether effects differed when the analysis was based on a more refined operationalization of literacy ability. The first step in the meta-analysis was therefore to compute the weighted mean of the effect sizes aggregated to the highest level (general literacy ability) as well as the weighted means of the comprehension- and code-related effect sizes. This resulted in weighted mean effect sizes of $d = 0.18$, $d = 0.22$, and $d = 0.17$, respectively. Following Cohen (1988), effect sizes around 0.20 can be categorized as small. The results of z tests showed that all three effect sizes deviated significantly

from zero: for the effect sizes aggregated to the highest level $z = 5.51$ (95% confidence interval [CI] = [0.11, 0.24]), for the comprehension-related effect sizes $z = 6.04$ (95% CI = [0.15, 0.29]), and for the code-related effect sizes $z = 3.96$ (95% CI = [0.08, 0.25]), in all three cases $p < .001$.

The programs included in the meta-analysis thus seemed to have a small but statistically significant effect on literacy abilities. The minor difference between the weighted means of the comprehension- and code-related effect sizes implies that there is no real indication that, overall, the programs had a differential impact on the two types of skills. In the subsequent analyses we thus drew solely on the effect sizes aggregated to the highest level.

Homogeneity analysis. We computed the homogeneity statistic Q (Hedges & Olkin, 1985; Lipsey & Wilson, 2001) to examine whether the effect sizes included in the analysis all estimated the same population effect. If Q is statistically nonsignificant, it can be assumed that the distribution of effect sizes is homogeneous and that individual effect sizes differ from the population mean by sampling error only. If Q is significant, the distribution of effect sizes is heterogeneous, which means that there are differences among the effect sizes that have a source other than participant-level sampling error (Lipsey & Wilson, 2001). The Q test proved to be significant ($Q = 66.67$, $df = 46$, $p < .05$). Therefore, the assumption of homogeneity was rejected; there is evidently excess variability in the effect sizes that needs to be explained.

Accounting for excess variability: Moderator analyses. Variability in effect sizes that is not the result of participant-level sampling error can be explained in more than one way. In a random effects model, excess variability results from random differences at the study level. Application of a random effects model requires the recalculation of the weighted mean effect size to include an additional random variance component (Lipsey & Wilson, 2001). New analyses including this random component did not produce different results, however; in fact, the weighted mean effect size was exactly the same as in the initial (fixed effects) model ($d = 0.18$). In a fixed effects model, it is assumed that excess variability in effect sizes originates from systematic, identifiable differences between studies (Lipsey & Wilson, 2001). To test this assumption, we conducted moderator analyses following the *analog to the ANOVA* procedure (Hedges, 1982; Lipsey & Wilson, 2001), using the program, sample, and study characteristics discussed above as moderators. In the analog to the ANOVA procedure, effect sizes are grouped into categories (i.e., values of a moderator variable). The homogeneity among the effect sizes within the categories and the differences between the categories are tested by means of a Q test. If the result of the between-group comparison ($Q_{between}$) is significant, it can be assumed that the variable concerned is related to effect size variability. Table 2 presents the results of the moderator analyses.

In none of the cases the moderator analyses revealed statistically significant effects, but in the case of sample selection there was a trend ($p < .10$). The mean effect sizes in both the randomization and no randomization categories still differed significantly from zero, however (randomized studies: $z = 2.11$, $p < .05$;

TABLE 2
Moderator analyses of program, sample, and study characteristics

Variable	Groups	*N* comparisons	Weighted ES	95% CI	Q_{within}	$Q_{between}$
Program characteristics						
1. Activity type	Shared reading only	14	0.05	[–0.11, 0.20]	10.28	3.59
	Shared reading + other activities	27	0.21***	[0.14, 0.28]	40.79*	
	Literacy exercises	6	0.17	[–0.06, 0.40]	12.01*	
2. Program focus	Comprehension	20	0.13*	[0.03, 0.23]	25.03	2.00
	Code	7	0.16	[–0.06, 0.38]	12.08	
	Both	20	0.22***	[0.13, 0.31]	27.56	
3. Staff quality	Professionals	29	0.21***	[0.11, 0.30]	26.21	0.59
	Semiprofessionals	10	0.18**	[0.05, 0.31]	21.45*	
	Both	3	0.12	[–0.09, 0.33]	9.14*	
4. Home visits	Yes	16	0.18***	[0.09, 0.28]	33.75**	0.01
	No	31	0.18***	[0.09, 0.26]	32.91	
5. Group meetings	Yes	36	0.20***	[0.13, 0.27]	43.74	1.17
	No	11	0.12*	[0.00, 0.24]	21.76*	
6. Book provision	Yes	36	0.18***	[0.11, 0.25]	55.83*	0.00
	No	10	0.18*	[0.00, 0.35]	10.76	
7. Location	Home-based only	39	0.17***	[0.11, 0.24]	65.63**	0.36
	Home- and center-based	8	0.24*	[0.03, 0.44]	0.68	
8. Duration	< 5 months	26	0.13**	[0.03, 0.23]	27.81	1.21
	≥ 5 months	21	0.21***	[0.13, 0.29]	37.65**	
Sample characteristics						
9. Educational status	At risk	27	0.16***	[0.08, 0.24]	45.32*	0.24
	Not at risk	19	0.20***	[0.09, 0.30]	20.51	
10. Age group	Preformal education	26	0.19***	[0.10, 0.28]	36.65	1.00
	Formal education	19	0.14**	[0.04, 0.25]	28.93*	
	Both	2	0.26*	[0.04, 0.47]	0.10	
Study characteristics						
11. Sample selection	Randomization	26	0.11*	[0.01, 0.21]	21.00	2.83
	No randomization	21	0.22***	[0.14, 0.30]	42.84**	
12. Pretesting	Yes	35	0.15***	[0.07, 0.23]	49.18*	1.41
	No	8	0.24***	[0.12, 0.35]	15.78*	
13. Time of measurement	Short term	38	0.20***	[0.13, 0.27]	45.00	2.69
	Follow-up	9	0.04	[–0.14, 0.22]	18.98*	

Note. ES = effect size; CI = confidence interval.
*$p \leq .05$. **$p \leq .01$. ***$p \leq .001$.

nonrandomized studies: $z = 5.37$, $p < .001$), which implies that although the mean effect size for the nonrandomized studies was larger, even for the randomized studies the effect was nontrivial.

The results for program focus need some further clarification. Table 2 presents the findings of a moderator analysis using only the effect sizes aggregated to the level of general literacy ability. The information value of this outcome is limited, however: The more interesting question is whether skills of a particular type were more strongly affected by programs focusing on that type of skills than on other skills. Therefore, we conducted analyses with program focus as the moderator variable separately for comprehension- and code-related skills. The results were not as expected: For neither of the two skill categories did the between-group statistic reveal significant effects (for comprehension skills $Q_{between} = 3.99$, $df = 2$, $p > .05$, for code skills $Q_{between} = 0.99$, $df = 2$, $p > .05$). For comprehension-related skills the largest mean effect size was—counterintuitively—found for code-focused programs ($d = 0.43$), but this result was based on only two observations. For programs focused on comprehension skills and programs focused on both skills, the mean effect sizes were $d = 0.15$ and 0.25, respectively. For code-related skills the mean effect sizes for programs focused on comprehension skills, code skills, and both types of skills were $d = 0.11$, 0.18, and 0.20, respectively.

Finally, a sensitivity analysis was performed in which the results of the fixed effects model assumed in the analog to the ANOVA procedure were compared to the results of a mixed effects model, which assumes that the effects of moderator variables are systematic but that there is a remaining (unmeasured) random effect in the effect size distribution in addition to sampling error (Lipsey & Wilson, 2001). The latter model revealed—similar to the fixed effects model—that none of the moderators had a significant effect; the trend for sample selection disappeared ($Q_{\text{between}} = 1.75$, $p > .05$).

Conclusions and Discussion

The first aim of this meta-analysis was to describe the general effectiveness of recently evaluated family literacy programs. The second aim was to examine whether effects differed when the analysis was based on a more refined operationalization of literacy ability. To this end, we made a theoretically driven distinction between comprehension- and code-related skills. The third aim was to further investigate the relationships between program effectiveness and program, sample, and study characteristics.

To provide an overview of the types of programs that have been developed, implemented, and systematically evaluated over the past two decades, we started the description of our results with an analysis of the characteristics of the programs in our database as well as the relationships between these characteristics. A first interesting observation was the frequent occurrence of programs offering a broad range of activities and the infrequency of programs focusing on (primarily code-related) literacy exercises. This is likely a reflection of the move away from the focus on reading readiness in beginning literacy education—advocating the training of specific preliteracy skills—to the holistic approach advocated in the past two to three decades, particularly by representatives of the emergent literacy paradigm (Hall, 1987; Teale & Sulzby, 1986). Some of the observed relationships among the program characteristics and between the program and sample characteristics are as could be expected. It is likely that programs focusing on comprehension-related skills transfer these skills by means of activities such as shared reading, in which parents and

children (assumedly) enter in a process where they construct meaning together. Such programs would also provide resources in the form of storybooks. It can also be expected that programs that offer a broad range of activities last longer than programs that focus on a very specific kind of activity (either shared reading only or literacy exercises). The observed relationships between staff quality on one hand and delivery mode and children's at-risk status on the other call for further discussion. It seems surprising that semiprofessionals were so often responsible for training the parents of children at risk of educational or reading delays. In some cases, however, the use of semiprofessionals for at-risk families seemed to be a deliberate strategy. In the HIPPY program (evaluated by BarHava-Monteith, Harré, & Field, 1999, and Van Tuijl et al., 2001), for example, parent training was provided by mothers from the target communities who were slightly better educated than the mothers they supported. Using mothers from the same communities was seen as a way of approaching parents who are sometimes hard to reach. This also partly explains why programs using semiprofessionals more often provided home visits: It is probably easier for such trainers to be invited to families' homes. Although one could wonder whether semiprofessionals are able to transfer program contents as intended, particularly in the case of disadvantaged parents (see, e.g., Gray & Wandersman, 1980), it is noteworthy that programs using semiprofessionals for training at-risk families did not yield a significantly lower mean effect size than programs using professionals (in both cases $d = 0.16$, $Q_{\text{between}} = 0.00$, $df = 1$, $p > .05$).

Our analysis revealed small but significant weighted mean effect sizes for general literacy ability and for both comprehension-related abilities and code-related abilities. Family literacy interventions thus seem to make a modest contribution to children's literacy skills. This seems in line with the conclusions of Blok et al. (2005) and White et al. (1992) that the added value of parent involvement programs is generally small, but it does not entirely correspond with the results of the earlier meta-analyses conducted by Sénéchal and Young (2008) and Mol et al. (2008), who established medium to large mean effect sizes ($d = 0.65$ and 0.42, respectively). How can this difference be explained? First of all, it could be a result of the decision in the latter two meta-analyses to include or exclude certain types of effect measures—Sénéchal and Young left out the results of vocabulary measures, whereas Mol et al. focused solely on vocabulary—which might have increased the overall mean effect size. Second, we were more conservative in adjusting notably large effect sizes—following Lipsey and Wilson (2001) we drew the line for outlier effect sizes at two standard deviations from the mean, whereas both Mol et al. and Sénéchal and Young drew the line at three. This decision probably affected the mean effect sizes to some extent, particularly since the samples in these studies were relatively small. Finally, the results of all three meta-analyses must be understood in light of the precision with which they estimated the weighted mean effect sizes. The CIs in the meta-analyses by Sénéchal and Young (2008) and Mol et al. (2008) were 0.53 to 0.76 and 0.16 to 0.54, respectively, whereas in the analysis presented in this article the CI was 0.11 to 0.24, which implies that in the former two analyses the estimate was less precise. The same was true for the CIs in the moderator analyses.

The small effect resulting from our meta-analysis can be interpreted in two ways. On one hand, one could say that a small effect is not what one might hope for given the investment in time and manpower some of these programs take. The

overall mean effect size of 0.18 would not correspond to more than a three-point gain for program children compared to control children on a standardized test such as the well-known Peabody Picture Vocabulary Test (e.g., Dunn & Dunn, 1997). On the other hand, one could argue that it is important to value an effect size in light of the context in which it was obtained (McCartney & Rosenthal, 2000). The activities in family literacy programs are not conducted by professional educators but by parents, who have to fit the program activities in the busy schedules of everyday life and who, given the fact that many of them are low educated, do not necessarily have the didactic skills (implicitly) required by the program. The fact that, in spite of these possible obstacles to an optimal implementation, the programs analyzed here on average show a nontrivial, positive effect may therefore still be considered meaningful.

The small difference between the comprehension- and code-related effect sizes indicates that the interventions do not have a clear differential impact on these two types of skills. This interpretation is further supported by the observation that neither for comprehension-related skills nor for code-related skills were there differences in effects between program types: Code-focused programs, for example, did not yield higher effects on code-related skills than did comprehension-focused programs. A possible reason for these findings is that, although programs may aim to focus on a specific kind of skill, parents cannot be forced to act accordingly. For example, parents who are instructed to focus on comprehension during book reading interactions might nevertheless give feedback on reading errors or, conversely, elaborate on the content of a story when they are supposed to be supporting the child's reading aloud. This explanation warrants further research into how program activities are actually carried out by parents and children (see Implications for Research and for Practice).

Moderator analyses revealed no significant effects of program, sample, or study characteristics. What are the possible reasons for the lack of moderator effects? First of all, it might be that there is not very much heterogeneity to be explained (the homogeneity statistic was not highly significant: $Q = 66.67$, $df = 46$, $p < .05$). It might also be that we failed to include more influential moderator variables. It can be assumed that the most important moderator is implementation quality or treatment fidelity, which refers to the question of whether parents and children actually carried out the activities in the way intended by the program. If treatment fidelity is low, the transfer of program contents is likely to be inadequate and program effects will be tempered. However, only 12 of the 30 studies in our database included some measure of treatment fidelity, and in most cases these measures were quite superficial (records or ratings of frequency, duration, and completion of parent–child activities; records or ratings of parental attendance at training sessions). In only one case were systematic observations made of parent–child activities. Interestingly, the results of these observations point in the direction of the implementation problems mentioned before: McElvany and Artelt (2009) found that "neither the structure of the conversations between parent and child, the quantity of parental feedback, nor the extent of parental guidance through the sessions proved satisfactory" (p. 428).

Limitations

This study has certain limitations. First, the amount of statistical information provided by the primary studies differed. Only 12% of the studies provided

Cohen's *d* effect sizes. For the other studies, these had to be computed on the basis of the available statistical information. However, the level of detail of this information varied. Some researchers provided both (unadjusted) pre- and posttest means, standard deviations, and the results of statistical testing. Others reported only adjusted posttest means, gain scores, or *t* or *F* scores. These differences may have affected our results (McGaw & Glass, 1980). Second, one of the main drawbacks of meta-analyses is the apples-and-oranges problem (see the introduction): Meta-analysts combine data from different studies, although these data might not always be entirely comparable (Kulik & Kulik, 1989; Lipsey & Wilson, 2001). We took this problem into account by distinguishing between comprehension- and code-related effect measures. The lack of differential effects found for these two categories may imply that this distinction was still too crude. It is certainly the case that both categories cover a range of abilities and that further categorizations might have yielded different results. In this light, one could also speculate that the proximity of measures to program contents plays a role: Among the variety of measures used some will be more closely aligned to the skills supported in a program and some will be more distal; the latter will probably yield smaller effects than the former. In our case, however, finer categorizations would have made the number of effect sizes too small for meaningful analyses. Third, although the way programs are conducted in practice is likely a substantial factor in their success, in-depth information about program implementation or treatment fidelity was provided in hardly any of the studies examined. Whether overall effectiveness was decreased by the way parents and children carried out program activities therefore remains a matter of speculation and should be the subject of future research (see below).

Implications for Research and for Practice

Our findings have implications for the developers and users of family literacy programs, for those who make decisions about their implementation, and for the researchers who examine their effectiveness. The conclusion that the overall effects of the programs are small should give program developers, policymakers, and educators pause for thought as the high expectations they might have of these programs are not necessarily justified. This does not mean, however, that the programs should be abandoned. First of all, even small effects can be meaningful when viewed in light of the context in which they were obtained. Moreover, it remains to be determined how program activities are actually implemented by parents and children and how these activities interact with existing family (literacy) practices. Particularly in sociocultural minority groups the question is to what extent these programs agree with the informal pedagogical models of participating parents. Probably the most important recommendation following from this meta-analysis, then, is to conduct more thorough research into how programs are actually carried out by families. Such studies are likely to provide insights into how program contents are transferred from parents to children and thus into what might temper the effectiveness of some programs. This will probably help developers to make programs (more) effective. Equally important is research into the way program contents are transferred from trainers to parents: If training is not optimal, implementation in the homes likely suffers. In summary, our findings call

for studies that combine rigorously designed, product-oriented effect analyses with careful, process-oriented research examining whether and how family literacy programs bring about change in children's home environments.

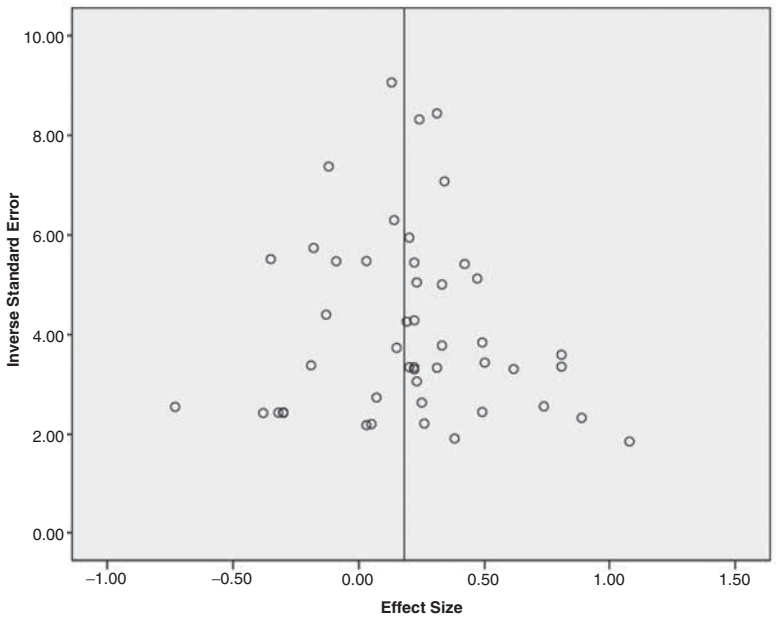
APPENDIX A
Additional information about the categorization of program characteristics

Moderator variable	Operationalization
Activity type	This variable refers to the type(s) of parent–child activities that are offered in the program: (a) shared reading, that is, joint parent–child activities around a storybook or picture book, with the focus on the interactive transaction of meaning; (b) literacy exercises, which include phonics training and storybook-based activities that do not focus on the transaction of meaning, but on practicing correct reading; (c) shared reading plus, that is, programs that complemented shared reading activities by, for example, library visits, songs, storytelling, vocabulary teaching, joint writing, or literacy exercises.
Program focus	This variable reflects whether the emphasis in the program was on (a) comprehension-related skills, (b) code-related skills, or (c) both. Programs were categorized as comprehension focused if they aimed to enhance abilities such as lexical knowledge, narrative and exposition skills, reading comprehension, and knowledge about story structure and/or to promote the application of interaction strategies such as relating story contents to the child's own experience, expanding the child's utterances, decontextualizing from story contents, posing open-ended or wh-questions, predicting story lines, and discussing story contents. Programs were categorized as code focused if they sought to enhance abilities such as word decoding, fluent reading, spelling, phonological skills, concepts about print, alphabet knowledge, and rhyme awareness and/or to promote parent–child activities such as word reading, word writing, simultaneous reading aloud of text, and manipulating letter cards.
Staff quality	This variable refers to the professional qualifications of those responsible for the training and support of parents. We distinguished between (a) programs that only used professionals, that is, trainers who were educated and experienced in the field of literacy training, education, and/or parent training (including teachers and researchers); (b) programs that only used semiprofessionals, that is, trainers who were not (yet) fully educated/experienced in these fields, such as (teacher) students and trained parents; and (c) programs that used both.

APPENDIX B
Formulas used to calculate effect sizes

Formula	Explanation
(1) $(M_{p(rogram)} - M_{c(ontrol)})/\sqrt{[(SD_p^2 + SD_c^2)/2]}$	Basic formula
(2) $(M_p - M_c)/\sqrt{\{[SD_p^2 * (n_p - 1)] + [SD_c^2 * (n_c - 1)]/(n_p + n_c - 2)\}}$	Applicable when there are considerable differences between program and control *n*s
(3) $t * \sqrt{[(n_p + n_c)/(n_p * n_c)]}$	Applicable when only a *T*-score is available
(4) $2 * \sqrt{(F/N)}$	Applicable when only an *F* score is available

APPENDIX C
*Funnel graph of the mean effect sizes per study (*x*-axis) plotted against the inverse of their standard errors (*y*-axis)*



Note

We would like to thank the editor and three anonymous reviewers for their valuable comments on an earlier draft of the paper. We would also like to thank Susannah Goss for the English language corrections.

References

**References marked with an asterisk indicate studies included in the meta-analysis.*

*Allen, L. R., Sethi, A., & Astuto, J. (2007). An evaluation of graduates of a toddlerhood home visiting program at kindergarten age. *NHSA Dialog*, *10*, 36–57.

*BarHava-Monteith, G., Harré, N., & Field, J. (1999). A promising start: An evaluation of the HIPPY program in New Zealand. *Early Child Development and Care*, *159*, 145–157.

Blok, H., Fukkink, R. G., Gebhardt, E. C., & Leseman, P. P. M. (2005). The relevance of delivery mode and other programme characteristics for the effectiveness of early childhood intervention. *International Journal of Behavioral Development*, *29*, 35–47.

Cairney, T. H. (2002). Bridging home and school literacy: In search of transformative approaches to curriculum. *Early Child Development and Care*, *172*, 153–172.

*Chow, B. W.-Y., & McBride-Chang, C. (2003). Promoting language and literacy development through parent–child reading in Hong Kong preschoolers. *Early Education and Development*, *14*, 233–248.

*Chow, B. W.-Y., McBride-Chang, C., Cheung, H., & Chow, C. S. L. (2008). Dialogic reading and morphology training in Chinese children: Effects on language and literacy. *Developmental Psychology*, *44*, 233–244.

Cohen, J. (1988). *Statistical power analysis for the social sciences*. Hillsdale, NJ: Erlbaum.

*Cronan, T. A., Brooks, L. B., Kilpatrick, K., Bigatti, S. M., & Tally, S. (1999). The effects of a community-based literacy program: One-year follow-up findings. *Journal of Community Psychology*, *27*, 431–442.

*Cronan, T. A., Cruz, S. G., Arriaga, R. I., & Sarkin, A. J. (1996). The effects of a community-based literacy program on young children's language and conceptual development. *American Journal of Community Psychology*, *24*, 252–272.

*Drouin, M. A. (2009). Parent involvement in literacy intervention: A longitudinal study of effects on preschoolers' emergent literacy skills and home literacy environment. *Early Childhood Services*, *3*, 1–18.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test* (3rd ed.). Circle Pines, MN: American Guidance Service.

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*, 629–634.

Erion, J. (2006). Parent tutoring: A meta-analysis. *Education and Treatment of Children*, *29*, 79–106.

*Evangelou, M., Brooks, G., & Smith, S. (2007). The Birth to School Study: Evidence on the effectiveness of PEEP, an early intervention for children at risk of educational under-achievement. *Oxford Review of Education*, *33*, 581–609.

*Fielding-Barnsley, R., & Purdie, N. (2003). Early intervention in the home for children at risk of reading failure. *Support for Learning*, *18*, 77–82.

Fishel, M., & Ramirez, L. (2005). Evidence-based parent involvement interventions with school-aged children. *School Psychology Quarterly*, *20*, 371–402.

*France, L., Topping, K., & Revell, K. (1993). Parent-tutored cued spelling. *Support for Learning*, *8*, 11–15.

Gomby, D. S., Culross, P. L., & Behrman, R. E. (1999). Home visiting: Recent program evaluations. Analysis and recommendations. *Future of Children*, *9*, 4–26.

Gray, S. W., & Wandersman, L. P. (1980). The methodology of home-based intervention studies: Problems and promising strategies. *Child Development, 51*, 993–1009.

Grolnick, W. S. (2003). *The psychology of parental control: How well-meant parenting backfires*. Mahwah, NJ: Erlbaum.

Hall, N. (1987). *The emergence of literacy*. Dunton Green, England: Hooder & Stoughton.

Hannon, P. (1995). *Literacy, home and school: Research and practice in teaching literacy with parents*. London, England: Falmer.

Hannon, P. (2003). Family literacy programmes. In N. Hall, J. Larson, & J. Marsh (Eds.), *Handbook of early childhood literacy* (pp. 99–111). London, England: Sage.

Hedges, L. V. (1982). Fitting categorical models to effect sizes from a series of experiments. *Journal of Educational and Behavioral Statistics*, *7*, 119–137.

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

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal*, *2*, 127–160.

*Jordan, G. E., Snow, C. E., & Porche, M. V. (2000). Project EASE: The effect of a family literacy project on kindergarten students' early literacy skills. *Reading Research Quarterly*, *35*, 524–546.

Kağıtçıbaşı, Ç. (1999). Empowering parents and children: The case of the Turkish Early Enrichment Project. In L. Eldering & P. P. M. Leseman (Eds.), *Effective early education: Cross-cultural perspectives* (pp. 235–258). New York, NY: Falmer.

*Kim, J. S., & Guryan, J. (2010). The efficacy of a voluntary summer book reading intervention for low-income Latino children from language minority families. *Journal of Educational Psychology*, *102*, 20–31.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, England: Cambridge University Press.

Kulik, J. A., & Kulik, C. L. C. (1989). Meta-analysis in education. *International Journal of Educational Research*, *13*, 221–340.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*, 293–323.

Lancy, D. F. (1994). The conditions that support emergent literacy. In D. F. Lancy (Ed.), *Children's emergent literacy: From research to practice* (pp. 1–20). Westport, CT: Praeger.

Lipsey, M. W. (2003). Those confounded moderators in meta-analysis: Good, bad, and ugly. *Annals of the American Academy of Political and Social Science*, *587*, 69–81.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, *48*, 1181–1209.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

*Lonigan, C. J., & Whitehurst, G. J. (1998). Relative efficacy of parent and teacher involvement in a shared-reading intervention for preschool children from low-income backgrounds. *Early Childhood Research Quarterly*, *13*, 263–290.

Mattingly, D. J., Prislin, R., McKenzie, T. L., Rodriguez, J. L., & Kazyar, B. (2002). Evaluating evaluations. The case of parent involvement programs. *Review of Educational Research*, *72*, 549–576.

McCartney, K., & Rosenthal, R. (2000). Effect size, practical importance, and social policy for children. *Child Development*, *71*, 173–180.

McElvany, N. (2008). *Förderung von Lesekompetenz im Kontext der Familie* [Promoting reading literacy in the family context]. Münster, Germany: Waxmann.

*McElvany, N., & Artelt, C. (2009). Systematic reading training in the family: Development, implementation, and initial evaluation of the Berlin Parent–Child Reading Program. *Learning and Instruction*, *19*, 79–95.

McElvany, N., & van Steensel, R. (2009). Potentials and challenges of family literacy interventions: The question of implementation quality. *European Educational Research Journal*, *8*, 418–433.

McGaw, B., & Glass, G. V. (1980). Choice of the metric for effect size in meta-analysis. *American Educational Research Journal*, *17*, 325–337.

*Miller, B. V., & Kratochwill, T. R. (1996). An evaluation of the Paired Reading program using Competency-Based Training. *School Psychology International*, *17*, 269–291.

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.

*Morrow, L. M. (1992). The impact of a literature-based program on literacy achievement, use of literature, and attitudes of children from minority backgrounds. *Reading Research Quarterly*, *27*, 252–275.

Morrow, L. M., & Paratore, J. (1993). Family literacy: Perspective and practices. *Reading Teacher*, *47*, 194–200.

*Morrow, L. M., & Young, J. (1997). A family literacy program connecting school and home: Effects on attitude, motivation, and literacy achievement. *Journal of Education Psychology*, *89*, 736–742.

Mullis, I. V. S., Martin, M. O., Gonzales, E. J., & Kennedy, A. M. (2003). *PIRLS 2001 international report: IEA's study of reading literacy achievement in primary schools in 35 countries*. Chestnut Hill, MA: Boston College.

*Murad, C. R., & Topping, K. J. (2000). Parents as reading tutors for first graders in Brazil. *School Psychology International*, *21*, 152–171.

Nickse, R. S. (1989). The noises of literacy: An overview of intergenerational and family literacy programs. Washington, DC: Office of Educational Research and Development.

Nickse, R. S. (1991, April). *A typology of family and intergenerational literacy programs: Implications for evaluation*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Olds, D. L., & Kitzman, H. (1993). Review of research on home visiting for pregnant women and parents of young children. *Future of Children*, *3*, 53–92.

Organisation for Economic Co-operation and Development. (2001). *Knowledge and skills for life: First results from PISA 2000*. Paris, France: Author.

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, *8*, 157–159.

*Overett, J., & Donald, D. (1998). Paired reading: Effects of a parent involvement programme in a disadvantaged community in South Africa. *British Journal of Educational Psychology*, *68*, 347–356.

*Powell-Smith, K. A., Shinn, M. R., Stoner, G., & Good, R. H. (2000). Parent tutoring in reading using literature and curriculum materials: Impact on student reading achievement. *School Psychology Review*, *29*, 5–27.

Purcell-Gates, V. (2000). Family literacy. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 853–870). Mahwah, NJ: Erlbaum.

*Rasinski, T., & Stevenson, B. (2005). The effects of Fast Start Reading: A fluency-based home involvement reading program, on the reading achievement of beginning readers. *Reading Psychology*, *26*, 109–125.

Raudenbush, S. W. (2005). Learning from attempts to improve schooling: The contribution of methodological diversity. *Educational Researcher, 34*(5), 25–31.

*Reutzel, D. R., Fawson, P. C., & Smith, J. A. (2006). Words to Go! Evaluating a first-grade parent involvement program for "making" words at home. *Reading Research and Instruction*, *45*, 119–159.

*Rolla San Francisco, A., Arias, M., Villers, R., & Snow, C. (2006). Evaluating the impact of different early literacy interventions on low-income Costa Rican kindergartners. *International Journal of Educational Research*, *45*, 188–201.

*Saint-Laurent, L., & Giasson, J. (2005). Effects of a family literacy program adapting parental intervention to first graders' evolution of reading and writing abilities. *Journal of Early Childhood Literacy*, *5*, 253–278.

*Saracho, O. N. (1997). The influence of a home literacy program and children's development of literacy. *Perceptual and Motor Skills*, *85*, 185–186.

Schatschneider, C., Westberg, L., & Shanahan, T. (2008). Methodology of the National Early Literacy Panel. In National Early Literacy Panel (Ed.), *Developing early literacy: Report of the National Early Literacy Panel* (pp. 1–54). Washington, DC: National Institute for Literacy.

Sénéchal, M., LeFevre, J. A., Smith-Chant, B. L., & Colton, K. V. (2001). On refining theoretical models of emergent literacy: The role of empirical evidence. *Journal of School Psychology*, *39*, 439–460.

Sénéchal, M., & Young, L. (2008). The effect of family literacy interventions on children's acquisition of reading from kindergarten to grade 3: A meta-analytic review. *Review of Educational Research*, *78*, 880–907.

Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Stanovich, K. E. (1980). Toward an interactive compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, *16*, 32–71.

*St. Clair, L., & Jackson, B. (2006). Effect of family involvement training on the language skills of young elementary children from migrant families. *School Community Journal*, *16*, 31–41.

Sterne, J. A. C., Egger, M., & Davey Smith, G. (2001). Systematic reviews in health care: Investigating and dealing with publication bias and other biases in meta-analysis. *BMJ*, *323*, 101–105.

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading. Evidence from a longitudinal structural model. *Developmental Psychology*, *38*, 934–947.

*Sylva, K., Scott, S., Totsika, V., Ereky-Stevens, K., & Crook, C. (2008). Training parents to help their children read: A randomized control trial. *British Journal of Educational Psychology*, *78*, 435–455.

Teale, W. H., & Sulzby, E. (1986). Introduction: Emergent literacy. In W. H. Teale & E. Sulzby (Eds.), *Emergent literacy: Writing and reading* (pp. vii–xxv). Norwood, NJ: Ablex.

Thornton, A., & Lee, P. (2000). Publication bias in meta-analysis: Its causes and consequences. *Journal of Clinical Epidemiology*, *53*, 207–216.

*Van Tuijl, C., Leseman, P. P. M., & Rispens, J. (2001). Efficacy of an intensive home-based educational intervention programme for 4- to 6-year-old ethnic minority children in the Netherlands. *International Journal of Behavioral Development*, *25*, 148–159.

*Watson, T., & Hempenstall, K. (2008). Effects of a computer based beginning reading program on young children. *Australasian Journal of Educational Technology*, *24*, 258–274.

White, K. R., Taylor, M. J., & Moss, V. D. (1992). Does research support claims about the benefits of involving parents in early intervention programs? *Review of Educational Research*, *62*, 91–125.

*Whitehurst, G. J., Arnold, D. S., Epstein, J. N., Angell, A. L., Smith, M., & Fischel, J. E. (1994). A picture book reading intervention in day care and home for children from low-income families. *Developmental Psychology*, *30*, 679–689.

Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development*, *69*, 848–872.

Wilson, D. B. (2001). *Effect Size Determination Program [Computer software]*. College Park: University of Maryland.

Authors

ROEL VAN STEENSEL is a linguist who works as an educational researcher at the Research Institute of Child Development and Education at the University of Amsterdam, P.O. Box 94208, Amsterdam, 1090 GE, Netherlands; e-mail: *R.C.M.vanSteensel@uva.nl*. His research focuses on the literacy development of at-risk students, in the preschool phase as well as in primary and secondary education.

NELE MCELVANY is an educational psychologist who is a professor at Technische Universität Dortmund; e-mail: *McElvany@ifs.tu-dortmund.de*. This research project was conducted during her time at the Max Planck Institute for Human Development in Berlin. Her research focuses on individual, social, and institutional conditions for literacy development, students with immigrant backgrounds, and teacher competencies in regard to teaching reading literacy.

JEANNE KURVERS works as a psycholinguist at the department of Language and Culture Studies of Tilburg University; e-mail: *J.J.Kurvers@uvt.nl*. The main focus of her research is the emergent and beginning literacy of children and adults in relation to multilingualism in the context of migration in the Netherlands and in developing countries.

STEPHANIE HERPPICH is a psychologist, Ph.D. student, and research assistant at the Leibniz Institute for Science and Mathematics Education at the University of Kiel; e-mail: *herppich@ipn.uni-kiel.de*. Her research interests are tutoring and tutors' assessment accuracy.