Published in final edited form as:

Early Child Res Q. 2012 ; 27(1): 55–56. doi:10.1016/j.ecresq.2011.07.003.

Examining the Sustainability of an Evidence-Based Preschool Curriculum: The REDI Program

Rebecca M. Sanford DeRousie and Karen L. Bierman

Abstract

This study examined the extent to which an evidence-based preschool curriculum (Head Start REDI) was sustained by 20 teachers during the year following a randomized controlled efficacy trial, when teachers were no longer required by the research project to implement the curriculum. Two quantitative measures of sustainability (teacher ratings, REDI coach ratings) and a qualitative measure (teacher interview) were collected and compared. Sustainability varied by the specific curriculum component, with higher rates of sustainability for the social-emotional component (Preschool PATHS) than for the language and literacy components. Estimates of sustainability were affected by the method of measurement, with REDI coach ratings and qualitative teacher interviews more closely aligned than teacher ratings. Responses from qualitative interviews identified the main factors that teachers thought affected sustainability. Teacher responses suggest that efforts to promote sustainability are best targeted at reducing barriers, such as competing demands, rather than simply highlighting the benefits of the new curriculum.

Keywords

sustainability; Head Start teachers; evidence-based curriculum

Over the past decade, emphasis on the development and implementation of evidence-based programs has intensified from both research and policy perspectives (Rohrbach, Grana, Sussman, & Valente, 2006). Recent education initiatives, such as No Child Left Behind (2002) and Good Start Grow Smart (2002), now require that federally funded programs utilize “evidence-based” programs. Numerous lists of evidence-based programs have been compiled targeting mental, emotional, and behavioral problems, as well as educational outcomes (Collaborative for Academic, Social, and Emotional Learning, 2003; Gottfredson & Gottfredson, 2002).

As evidence-based programs are more broadly disseminated, growing concern is focused on the processes of implementation quality and sustainability (Greenberg, Domitrovich, Graczyk, & Zins, 2005; Griffin, 2010; Rohrbach et al., 2006). This new focus is critical, because even effective programs will produce few results if they are not implemented well, and the start-up costs of initiating evidence-based programs may not be worthwhile if the programs are not sustained over time. Despite the recent acknowledgements of the

© 2011 Elsevier Inc. All rights reserved.

Reprint requests can be sent to, Karen Bierman, The Pennsylvania State University, 110 Moore Building, University Park, PA 16802, (phone:814-865-3879; fax:814-863-7002).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

importance of building a knowledge base of sustainability, there is still limited research focused on understanding, measuring, and facilitating sustainability.

This lack of research on sustainability is of special concern for the field of early childhood interventions. Given the tremendous variation found in early childhood settings in areas such as financing, staff salaries, and turnover, it is crucial to investigate sustainability in such programs. As early childhood programs (e.g., preschools, daycares, and Head Start programs) have few discretionary funds to invest in new programs, once an evidence-based program is selected, it is important that it can be maintained with fidelity over time. The purpose of this study was to contribute to the understanding of sustainability in early childhood educational settings by examining the sustainability of a multi-component evidence-based program (REDI) in Head Start settings. Specifically, the study had three major aims: 1) to assess the degree to which sustained, high-quality implementation was achieved for three different intervention components, 2) to compare the utility of three different methods of assessing sustainability (e.g., REDI coach ratings, teacher ratings, and teacher qualitative interviews), and 3) to use qualitative interviews to explore the factors affecting sustainability from the teachers' perspective.

Previous Research on Sustainability

Generally speaking, sustainability refers to the continued implementation of programs over time (Pluye, Potvin, & Denis, 2004; Scheirer, 2005; Shediach-Rizkallah & Bone, 1998). The few studies that have examined the sustainability of school-based programs indicate that sustainability is not easy to attain. For example, Elias and Kamarinos (2003) examined the sustainability of social-emotional learning programs five years or more after implementation and found that only 6 of 14 sites (43%) had achieved program sustainability. A study of program dissemination of an alcohol prevention program found that only one in four teachers (25%) continued to teach the program in the second year, despite free training and materials and being encouraged to deliver the program (Rohrbach, Graham, & Hansen, 1993). To some extent, the level of sustainability attained depends upon the criteria used to define a sustained program. For example, Scheirer (2005) concluded that 14 of the 17 health service programs she reviewed (82%) achieved some level of sustainability, using the liberal criteria of continued delivery of at least one program component by at least 60% of the original sites.

However, this simple definition does not take into account the importance of ensuring that program implementation continues at high enough quality to assure the ongoing achievement of targeted outcomes (Durlak, 2010). Since it is not always feasible to continually document program outcomes, it is commonly assumed that if the program is sustained with fidelity, then the program impact on target child outcomes is also being achieved. Thus, sustainability should not be defined simply as program continuation, but as the "continued implementation of an intervention or prevention program, with ongoing implementation fidelity to core program principles" (Han & Weiss, 2005, p. 666).

Sustainability, Fidelity, and Adaptation

Part of the difficulty in establishing a standard threshold to indicate sustainability is that there is continued debate about what represents fidelity in sustainability studies. Fidelity is commonly defined as the "degree to which teachers and other program providers implement programs as intended by the program developers" (Dusenbury, Brannigan, Falco, & Hansen, 2003, p. 240). The literature suggests that fidelity consists of several distinct dimensions that influence program outcomes, including adherence (the degree to which the program is implemented as written), dosage (the amount of programming that is delivered), and quality of delivery (the skill of the teacher or implementer who is presenting the program)

(Domitrovich & Greenberg, 2000; Dusenbury et al., 2003; Odom et al., 2010). Depending upon the extent to which measures or assessment methods emphasize one of these aspects of fidelity over another, they might provide different estimates of sustainability.

In addition, the tension between maintaining strict fidelity and allowing local adaptations to evidence-based programs is a ubiquitous issue in the field of sustainability. Adaptations are considered to be “modifications to the manner or intensity with which various components or lessons are delivered” (Pankratz et al., 2006, p. 319). Some argue that allowing for adaptations is necessary and will help promote sustainability because adaptations can help a program meet local needs and promote local buy-in to a new program (Bumbarger & Perkins, 2009; Dusenbury et al., 2003; Johnson, Hays, Center, & Daley, 2004; Scheirer, 2005). This may be especially important in classroom-based programs when classroom makeup may differ from year to year (Han & Weiss, 2005). Others argue that maintaining strict fidelity is critical to sustainability (Elliott & Mihalic, 2004; Olds, Hill, O'Brien, Racine, & Moritz, 2003) in order to produce the positive benefits associated with the original program. Indeed, much of the literature suggests that a lack of fidelity may harm program outcomes (Durlak & DuPre, 2008; Fixsen, Naoom, Blase, Friedman, & Wallace, 2005). Specifically, Dusenbury, Brannigan, Hansen, Walsh, and Falco (2005) found that 63% of adaptations detracted from curriculum objectives.

An important factor to consider in this debate is whether the adaptation is positive or negative. One could imagine a classroom-based program where a creative teacher makes changes that actually enhance the original curriculum. This type of adaptation is qualitatively different from those which detract from the curriculum's purpose, such as omitting lessons or ignoring protocol. Unfortunately research on fidelity has usually been limited to identifying *if* changes were made and has not addressed the *type* of change. Most teachers and other providers do make some adaptations (Hill, Maucione, & Hood, 2007; Pankratz et al., 2006). For example, a study of the sustainability of a residential substance abuse treatment program found that 92% of projects were continued after grant support ended, but only 36% continued with no significant change (Porowski, Burgdorf, & Herrell, 2004). In an examination of the Life Skills Training program, Dusenbury and colleagues (2005) found that no teacher carried out the lessons as written. Hence, future research would benefit from a focus on the type of adaptation being made (Bumbarger & Perkins, 2009), and it also seems reasonable that assessments of program sustainability should allow for some adaptation.

Measuring Sustainability

To add to the difficulty of studying sustainability, there is no standard method of measuring fidelity or sustainability. It is generally agreed that independent observers are the best means of measuring fidelity (Dusenbury et al., 2003; Harachi, Abbott, Catalano, Haggerty, & Fleming, 1999), but such a means of monitoring fidelity is not always feasible due to limited staff and funds. With school-based interventions, it is often teachers who implement a program and thus have firsthand knowledge about the implementation process. Therefore, teachers are the usual source of information on implementation; however, findings on the validity of teacher reports are mixed. Lillehoj, Griffin, and Spoth (2004) reported significant agreement between provider and observer ratings of fidelity, but only the observer scores predicted student outcomes. Additionally, they found that the providers (i.e., teachers) tended to overly report how much they were implementing (Lillehoj et al., 2004). Melde and colleagues (Melde, Esbensen, & Tusinski, 2006) examined fidelity of a school-based victimization prevention program and found that providers and observers were in agreement 70% of the time. In contrast, a number of studies have found low rates of agreement

between teachers and observers (Hansen & McNeal, 1999; Noell et al., 2005; Wickstrom, Jones, LaFleur, & Witt, 1998).

To address some of these pressing issues in the study of the sustainability of evidence-based early childhood programs, the present study examined three program components, using multiple informants to assess sustainability (teachers and REDI coaches), and included measures that assessed quality as well as quantity of sustained program use. In addition, open-ended interviews with teachers were conducted to better understand teachers' perceptions regarding the challenges associated with program sustainability.

The Current Study

The current study examined the sustainability of the Head Start REDI (REsearch-Based, Developmentally Informed) preschool curriculum during the year following its introduction in the context of a randomized-controlled efficacy trial. REDI is an evidence-based, multi-component enrichment curriculum targeting social-emotional and language/emergent literacy skill domains. The social-emotional component is the Preschool PATHS (Promoting Alternative Thinking Strategies) curriculum, a universal evidence-based program taught by teachers that targets prosocial skills, emotion knowledge, self-control, and social problem solving skills (Domitrovich, Cortes, & Greenberg, 2007). To enhance vocabulary and oral language skills, REDI uses a dialogic reading program (DR), which includes books with content synchronized with the PATHS lessons, scripted to support interactive reading in small groups, and enriched with vocabulary-enhancing prop books (Wasik & Bond, 2001; Whitehurst et al., 1994). REDI also includes two sets of learning activities designed to promote specific emergent literacy skills (ELAs). A set of "Sound Games" based on the work of Lundberg and colleagues (Adams, Foorman, Lundberg, & Beeler, 1998) introduces phonological awareness skills, moving from easier to more difficult skills during the course of the year (e.g., listening, rhyming, alliteration, words and sentences, syllables, and phonemes). Alphabet Center activities, including letter stickers, a letter bucket, materials to create a Letter Wall, and craft materials for various letter-learning activities are included to promote letter knowledge (Ball & Blachman, 1991; Lonigan, 2006). To learn more about the REDI curriculum, please see Bierman et al. (2008).

The REDI efficacy trial

Forty-four Head Start classrooms participated in the randomized-controlled trial, which used multiple informants and measures to evaluate the impact of REDI on child outcomes. Twenty-two lead teachers received the intervention materials, along with extensive professional development support to implement the REDI curriculum and generalization strategies. In addition to an intensive three-day workshop before the school year began and a one-day booster training in January, teachers received ongoing support from REDI coaches, who met with them on a weekly basis to model appropriate techniques and to provide technical assistance and mentoring in curriculum implementation.

Implementation data gathered during the research trial indicated moderate to strong implementation (Domitrovich, Gest, Jones, Gill, & Sanford DeRousie, 2010; Domitrovich et al., 2009). Reported monthly, using a 7-point scale to score implementation quality, average REDI coach ratings placed teachers in the 4 ("adequate") to 5 ("strong") range for all program components, PATHS, $M = 4.61$, $SD = .74$; Emergent Literacy Activities [ELA], $M = 4.61$, $SD = .63$, and Dialogic Reading [DR], $M = 4.39$, $SD = .57$.

Child outcome measures were collected at the beginning and end of this research implementation year, and indicated that children who received REDI, compared to children assigned to "usual practice" Head Start classrooms, had significantly higher scores on a

variety of language, literacy, and social-emotional measures compared to children who received the standard Head Start program (Bierman et al., 2008). In addition, based upon research observations of teachers, participation in REDI promoted more positive teaching practices in the classroom, including more positive behavior management and enriched language use (Domitrovich et al., 2009).

The REDI sustainability study

The present study focuses on the year that followed the teachers' initial implementation of REDI, when teachers were no longer mandated by the research program to use REDI, although the Head Start programs encouraged teachers to do so. Three types of data were collected during the sustainability year to assess the quantity and quality of continued implementation of the three core components of REDI (e.g., PATHS, ELA, DR).

The first goal of this study was to compare the degree to which the three different REDI components were sustained with fidelity. The multi-component nature of REDI offered the unique opportunity to examine sustainability by the same teachers of curriculum components that varied considerably in terms of their content and burden. The PATHS curriculum required teachers to initiate instruction in a novel domain (none of the participating classrooms were using a social-emotional learning curriculum before the trial) and to implement a comprehensive curriculum that involved weekly lessons, follow-up extension activities, and a set of generalized strategies for supporting social-emotional development in the classroom. The DR program required teachers to change the way they implemented an existing facet of their teaching (e.g. read aloud), replacing their usual books and activities with a set of curriculum-based books and specific presentation style, with daily lessons (4 times per week). The ELA components were much less extensive than either PATHS or DR, and involved inserting brief, discrete, and explicit instruction activities into the classroom schedule -- alphabet center activities and sound games that involved listening, rhyming, and sound segmenting games (three times per week). Given the differences in the demands that these various curriculum components placed on teachers, it was anticipated that they might represent different kinds of sustainability challenges, and that comparing the degree and quality of their ongoing use after the research trial might reveal processes relevant for understanding sustainability more generally.

A second goal of this study was to compare the results obtained from three different sources of sustainability data – professional coach ratings, teacher ratings, and qualitative interviews with teachers. A key question involved the degree to which there might be agreement among these sources, as well as the degree to which different biases might appear in estimates of sustainability derived from each type of measure.

The third goal of the study was to explore the reasons why the three REDI curriculum components were sustained, modified, or abandoned, by examining teacher responses in the qualitative interviews. If sustainability were found to vary by component, then an examination of the teachers' own words could help to identify specific supports or barriers that contributed to the differential sustainability.

Method

Overview of Procedures and Timeline

The present study followed the lead intervention teachers, tracking their use of the REDI curriculum components during the year after their participation in the research trial, when Head Start programs encouraged them to continue using the program, but the research no longer required it. These intervention teachers were interviewed during the winter months (January and February) of the sustainability year. By this time, teachers had established their

daily routines and had finalized their decisions regarding curriculum implementation. Qualitative interviews with teachers and quantitative ratings by teachers were collected at this time. Teachers were compensated \$20.

During the sustainability year, REDI coaches checked in with teachers approximately once per month (stopping by classrooms or connecting via phone or email) to encourage ongoing program use, and to answer any questions teachers had. REDI coaches also made a formal classroom visit in the fall and again in the spring, observing a PATHS lesson and a DR lesson during each visit, and reviewing teacher lesson plans. In the spring of the year (April–May), they completed rating forms to summarize teachers' sustained use of the REDI components during the year, using information gleaned from the formal classroom observations, their reviews of teacher lesson plans, and their periodic discussions with teachers.

Participants

Twenty of the 22 lead teachers who had been randomly assigned to the intervention condition of the REDI efficacy trial continued to teach in Head Start the year after their research participation (the other two teachers left their Head Start positions). These teachers (95% female, 85% white) varied in their background and prior teaching experience when they began their involvement with REDI: ten had been teaching for some time (11+ years), four were moderately experienced (6–10 years), and six were relatively new teachers (0–5 years of prior teaching experience). In terms of formal education, 12 of these lead teachers (60%) had completed a four year degree (or more), and the rest were evenly split between an associate' degrees, ($n = 4$, 20%) or a high school degree ($n = 4$, 20%). In all cases, teachers had taught REDI for the first time the year prior to this study, when they participated in the research trial.

Measures

REDI coach ratings—REDI coaches rated the quality of REDI program sustainability for each curriculum component (PATHS, DR, ELA) using a 7-point scale (1= poor, 4=adequate, 7=exemplary). For each component, coaches rated this summary item: “Based on your classroom observations, review of lesson plans, and consultation meetings, what is your evaluation of the overall quality of [REDI component] implementation?” Due to scheduling constraints, ELA activities were not always observed directly during REDI coach classroom visits, resulting in four classrooms with missing data for REDI coach ratings of ELA sustainability.

Teacher ratings—Midway through the school year (January–February), teachers completed a written questionnaire. After rating several specific items about their use of the REDI program, teachers rated their sustainability of each REDI component using three items: 1) To what extent was REDI implemented as written? 2) To what extent have you made changes to REDI components by shortening the lessons? (reverse-coded), and 3) To what extent have you made changes by incorporating new materials/activities? (reverse-coded). Each question was rated on a 5-point Likert Scale (e.g. 1=not at all, 3= somewhat, 5= very much), and they were averaged to create a sustainability score for each component. Inter-item reliability was acceptable, $\alpha = .72$ for DR, $\alpha = .85$ for PATHS, $\alpha = .65$ for ELA.

Teacher interviews—During face-to-face interviews in January and February, teachers were asked open-ended questions regarding their continued use of the REDI components, factors that supported or hindered their continued use of REDI, and their understanding of how REDI benefits child development. Interviews lasted 20–60 minutes and were audio-taped, and later transcribed by the first author. Responses regarding the continued use of

REDI components were categorized into three groups: 1) not sustained - teachers said they were not using the component, or described so many changes that the component no longer represented the REDI version, 2) moderately sustained - teachers reported continued use of the REDI component, but described changes in implementation that varied from the REDI written version, and 3) well sustained - teachers reported using the REDI component as written, with no more than very minor changes.

The factors that teachers discussed as enhancing or hindering their sustained use of REDI were coded into fourteen categories - nine factors that facilitated sustainability and five barriers to continued use (described more fully in the results).

Results

Levels of Sustainability of the REDI Components

The first goal of this study was to explore the extent to which Head Start teachers sustained the three major components of the REDI curriculum, as assessed by three different sources of sustainability data. Means and standard deviations for each intervention component are presented in Table 1, as assessed by REDI coach ratings, teacher ratings and teacher interview responses. According to all three sources, PATHS was sustained at the highest level, Emergent Literacy Activities [ELA] fell in the middle, and Dialogic Reading [DR] was sustained at the lowest level (see means and ranges in Table 1).

In order to estimate the number of teachers who achieved a level of sustainability thought to be sufficient to produce an impact on child outcomes, a cut-off was established to indicate a “satisfactory” level of sustainability on each measure. This cut-off required teachers to demonstrate continued and regular use of a program component, but allowed for program modifications by the teacher. Using a rating of 4 (or better) for the coach ratings (representing “adequate” on the 7-point scale), coaches rated 90% of the teachers as sustaining PATHS, and 60% of the teachers as sustaining ELA and DR at a satisfactory level. Using a parallel cutoff of 3.5 (or better) on the 5-point teacher ratings (averaged across three items, indicating implementing the program “pretty much” to “somewhat” as written), 100% of the teachers rated themselves as sustaining PATHS, 60% rated themselves as sustaining ELA, and 60% rated themselves as sustaining DR at a satisfactory level. Using a similar criterion of “moderately sustained” (or better) according to the teacher interviews, 100% of the teachers indicated that they were sustaining PATHS, 85% reported that they were sustaining ELA, and 70% indicated that they were sustaining DR at a satisfactory level. Note that a much smaller proportion of teachers indicated that they were implementing the program exactly as written (without any modifications) during the interviews, with only 85% (for PATHS) and 10% (for ELA and DR) of the teachers indicating that they had not made any changes in the program. This finding replicates prior research documenting the tendency for teachers to adapt and embellish programs over time.

In summary, all three measures of sustainability (coach ratings, teacher ratings, and teacher interviews) revealed the same relative rankings regarding the sustained use of REDI components, and approximately the same level of “satisfactory” sustainability, defined as sustained use of core components allowing for teacher modifications. Estimates of a “satisfactory” level of sustainability ranged from 90% – 100% for PATHS, 60% – 85% for ELA, and 60% – 70% for DR across the three types of measurement.

Within-Rater and Inter-Rater Associations

The second goal of this study was to compare the results obtained from three different sources of sustainability data. A key question involved the degree to which there might be agreement among these sources, as well as the degree to which different biases might appear

in estimates of sustainability derived from each type of measure. First, we examine correlations for the different REDI components from the same source, and then the cross-rater correlations.

Table 2 shows within-rater correlations for the three REDI components. Coach ratings of sustainability revealed significant correlations among all three curriculum components, ranging from $r = .59$ to $r = .66$. Apparently, coaches tended to view teachers as high or low implementers no matter what the activity, leading to fairly uniform ratings across components. In contrast, teacher ratings of sustainability were more differentiated across program components, revealing only one significant correlation (for ELA and PATHS). Teacher interviews fell in the middle, as significant correlations emerged for two of the three component pairs (for ELA and PATHS, and for ELA and DR). These findings suggest that coach ratings, and to some extent the teacher interviews as well, were capturing some common factors that affected sustainability estimates across the different REDI components (such as teaching quality, enthusiasm for the curriculum). The teacher ratings, in contrast, seemed more affected by component-specific factors that influenced sustainability.

Table 2 also shows the inter-rater correlations comparing different sources estimating sustainability of the same REDI components. Coach and teacher ratings revealed significant agreement in rating the sustainability of DR, $r = .45$, but not in rating the sustainability of ELA or PATHS, which tended to be in the opposite direction from expected, $r = -.15$ and $r = -.30$, respectively (see Table 2). The negative correlation for PATHS was not significant, and is likely an anomaly associated with the constrained variability and ceiling effects that were apparent in the teacher ratings of PATHS, since all teachers reported sustaining PATHS at a relatively high level. Likely for the same reason, rates of sustainability reported by teachers on the ratings and in the interviews were not significantly correlated for PATHS ($r = -.03$). Correlations between teacher ratings and teacher interviews were in the right direction (although not significant) for ELA ($r = .30$), and were significant for DR ($r = .46$). Coach ratings and teacher interviews were significantly aligned for both DR, $r = .64$, and for PATHS, $r = .64$, although the ratings for ELA were not correlated, $r = .00$. These findings again suggest that coach ratings and the qualitative teacher interviews may have been more sensitive to some over-arching factors associated with the quality of sustained implementation that contributed to their shared variance, at least in their ratings of DR and PATHS.

To consolidate information across sources and create a cross-method assessment of sustainability, a summary measure was created that required consistently high scores across all three sources to be identified as sustaining a REDI component at a high level. Specifically, to be designated as sustaining a component at a high level, a teachers had to receive a score of 5 or higher on coach ratings (7-point scale) *and* a score of 4 or higher on teacher ratings (5-point scale) *and* describe practices in the teacher interview that were coded as “well-sustained” or “moderately sustained.” Conversely a teacher was designed to show a low level of sustainability if *any* of the three sources of information revealed low levels of use – e.g., if teachers scored a two or lower on the coach ratings *or* teacher ratings of sustainability, or they said that they were not continuing to use the component during the interview. Teachers who did not meet these criteria for the “low” or “high” sustaining groups were placed into the medium level of sustainability. These teachers scored between 2.33 and 4.89 on the teacher ratings of sustainability, 3 or 4 on coach ratings, and were in the high or middle group on the teacher interviews. Using these conservative criteria, 75% of teachers sustained PATHS at a high level, whereas relatively few teachers sustained ELA or DR at a high level (e.g., 30% and 25%, respectively). However, setting the criteria more modestly, at a “moderate” to “high” level of sustainability, a majority of teachers sustained all components, 95% PATHS, 80% ELA, and 70% DR. Corresponding proportions of

teachers showing low (or no) sustainability based upon this cross-method categorization were 5% PATHS, 20% ELA, and 30% DR.

Factors Contributing to Differential Sustainability

The third goal of the study was to examine the reasons why teachers chose to sustain or not sustain the different REDI curriculum components. To better understand the reasons why teachers chose to sustain or not sustain the REDI, we examined the frequency with which teachers listed different factors as supporting or impeding the sustainability of the REDI components in the open-ended interviews. The responses and frequency of mention are shown in Table 3. The most common reason cited to support the continued use of REDI was that 100% of the teachers perceived benefits for the children, including teachers who were not sustaining REDI. Across the teachers, all of the components were specifically identified as providing benefits – although PATHS and DR were most commonly mentioned, as in the following remarks.

"I think PATHS really builds their self confidence, their self-worth, it really makes them feel good about themselves and respect each other. I think the PATHS lessons teach them a lot about feelings and identifying, working with other kids."

"[REDI has helped] definitely their vocabularies, their appreciation of books and their understanding of authors and illustrators, kind of how a story is set up.... I think it's really helping them to be more expressive in their creative activities like telling stories, talking about their art and their other things that they do."

A majority of teachers (85%) also mentioned that students were interested and engaged in the program and 85% also indicated that the REDI curriculum enhanced their teaching skills. Teachers particularly expressed excitement about seeing children use newly taught PATHS skills, such as social problem solving, and some also mentioned the children's use of new vocabulary from DR and enhanced language skills. Teachers felt that REDI enhanced their teaching skills because it acted as a reminder, setting down in print strategies and activities that teachers knew were effective but did not engage in regularly. Teachers also credited REDI with introducing them to new concepts and approaches to learning, such as incorporating language recasting into their daily routine, or realizing the importance of social-emotional learning. One teacher responded:

"I think [REDI has] made us stop and think about the emotional view....I think it just really made me aware that it's a component of the child's developmental being. Before, probably when I first started, it was more....of an academic approach before. And now it's kind of just made me a little bit more aware that there's more to a child."

Eighty percent of the teachers indicated that sustaining REDI was made easier by the support they received from their co-teachers and teaching team, who helped maintain the continued use of REDI in the classroom. A majority of teachers (75%) also indicated that sustaining REDI was made easier by the way that it fit with their teaching style – continuing to use REDI integrated easily with things they had done in the past. This may have meant talking with children about their feelings, as is done in PATHS, or presenting children with books regularly, as done with DR. The following comments epitomize the responses in this category:

"And I like PATHs, it's very similar to my personal way of dealing with behaviors and all of that, like using their words and you know, telling their friends....that part was really easy for me becausethat's how I've always done things."

"And then the DR I really like because, it was a lot of what I did already, was ask them questions and stories. It just guides me a little bit more through it."

Over half of the teachers (65%) mentioned that they appreciated the way that REDI addressed specific concerns of the classroom, and 60% mentioned that REDI helped create a better classroom atmosphere. These comments applied particularly to PATHS, as teachers reported that REDI provided a social-emotional learning curriculum where there had never been one before, improving the classroom climate as a function of the explicitly teaching of emotions and social-problem solving skills. For example, teachers remarked:

"In the end, your classroom has just changed. The dynamic. Kinds of problems you had before, they do get resolved. I've noticed a big change. You don't have the aggression, the disruptions, it's run real smooth."

"You can feel the difference, at least to me, in a REDI classroom. There's community, the kids really care about each other. Because they're willing to tell each other their feelings. And it's just amazing to me the difference."

In contrast to these factors which supported the sustainability of REDI, five main factors emerged that were endorsed by at least eight teachers as contributing to why teachers were **not** continuing to sustain REDI. The largest complaint, mentioned by 70% of the teachers, was that the use of REDI conflicted with other program requirements, and there was insufficient time to do both. For example, many programs were required to use thematic curricular approaches (designing classroom activities around themes such as snow, beach, garden, etc) and often the REDI books did not match up with what were considered appropriate Head Start themes. In addition, one program wanted teachers to incorporate a structured alphabet-learning program in addition to the alphabet center activities used in REDI. Another program encouraged teachers to attend trainings in a different reading approach (shared reading) in order to use it in the classroom alongside REDI's dialogic reading. PATHS did not conflict with other requirements because there were no other organized social-emotional learning programs in these Head Start classrooms. Teachers expressed much frustration with the difficulties of finding time and energy to fit everything in (REDI, other Head Start requirements, personal favorite activities). Their frustration is evident in the following comment:

"It's just hard. We're trying to do everything, we're trying to do High/Scope, we're trying to do themes, and we're trying to do REDI and it's, you can't do everything."

While some teachers saw the multiple demands as competition with REDI, others simply felt that they lacked sufficient time to complete all the REDI activities and everything else. This feeling of a time crunch was intensified by the fact that during the research implementation year, the teachers were excused from certain Head Start requirements in order to focus on implementing REDI. Once the research project was over, the Head Start requirements were reinstated, placing constraints on the teachers, especially in terms of time demands, (and also in terms of competing activities as described above). Teachers who only had 3.5 hours (half-day classrooms) with their children felt particularly hindered in their ability to deliver the curriculum in a satisfactory manner:

"I think definitely, if we had the longer day, then we could fit in more of the stuff, especially with, we could fit in everything with REDI, the PATHS and DR, plus what Head Start curriculum is expecting from us too. I think we could fit it all in."

While teachers often talked about time issues in regards to the entire REDI curriculum, DR and the ELA were all specifically mentioned as being difficult to fit into the daily schedule. Ten teachers (50%) cited time shortage as reason why they were not sustaining REDI.

Fifty-percent of the teachers also mentioned that they were reluctant to sustain REDI because they did not feel that aspects of REDI were developmentally appropriate for their students. The sound games (part of the ELA) were most commonly addressed by this

complaint – eight of the ten teachers who voiced this concern were specifically concerned about the sound games. Others were simply concerned that in the mixed-age classrooms of Head Start, some parts of REDI were not appropriate, especially in a class with a majority of three year olds. The following comments illustrate this concern about developmental appropriateness:

“There are just some of the books that I think are way above this level... I mean I realize that you picked out books with vocabulary that you want to highlight but some of the books are just way, you know, you can see them glaze. And I'm like oh boy and then I quick try to just get through the end of the book sometimes.”

“The sound games, I'll put in my lesson plans but is the kids don't get them, then I don't do them. Because a lot of them are, for the three years olds, they just don't... get it.”

A smaller proportion of teachers (40%) indicated that limited or no administrative support for continuing to use REDI was a reason for reducing use of the curriculum. Eight teachers mentioned that their supervisors had little interest or knowledge about the REDI approach and related activities which resulted in little or no support for the continued use of the REDI components.

The last barrier, mentioned by eight teachers, was that aspects of REDI conflicted with the teachers' philosophy and style. Teachers commented on how REDI conflicted with their beliefs about teaching (such as the belief that children need plenty of time to play without the pressure of learning something specific). They also objected to REDI because they had to alter their preferred activities or because REDI simply required too much work to incorporate regularly. These concerns are exemplified in the following comments.

“So once this year, it wasn't a requirement, I got of a lot of it because it just wasn't a personal fit for me... I have a way of doing my circle. It's my circle, I'm very protective of it. And then [after training]... they tell you, hey you can't do your circle the way that you're doing it... It was very upsetting... it just wasn't me.”

“You know the biggest problem I have, I feel like we're pushing too much. Letters and sound games..., I feel like there was too much emphasis on having kids learn letters... Quite honestly I wouldn't do a whole lot with letters at all if I didn't have to. I would have fun with the kids. But I still would work towards, you know, teaching them, just backing off letters, you know, more until they're ready. If they are ready, take them and work with them, but if they're not, let them be...”

One final question was whether teachers who did or did not sustain REDI talked about the program components differently in these qualitative interviews. To address this question, we identified the six teachers with the highest and six teachers with the lowest qualitative sustainability based upon the teacher interviews, and compared their open-ended descriptions of REDI. Interestingly, to a considerable extent, high and low sustainers made very similar comments about the REDI program. No differences between these groups were apparent in terms of the benefits they perceived for children, student engagement, positive impact on teaching skills, or value of REDI for promoting a more positive classroom atmosphere. In addition, no differences emerged in the extent to which high and low sustainers perceived time shortage, conflicts between REDI and other program requirements, or aspects of REDI's developmental inappropriateness for 3-year-olds. There were just two categories of response that differentiated the high and low sustainers – perceptions of program support for sustaining REDI and responses to programmatic and time conflicts.

Specifically, teachers in the high sustainability group were more likely to report that using REDI was required by their program than were teachers in the low sustainability group (83% vs. 16%). In addition, whereas none of the teachers in the high sustainability group complained about a lack of program support for sustaining REDI, 67% of the teachers in the low sustainability group mentioned the lack of program support as hindering sustainability.

High and low sustaining teachers mentioned concerns about the time needed to do REDI and conflicts between REDI and other program requirements (such as theme-based modules, alternative emergent literacy practices) with the same frequency. However, whereas the high-sustaining group noted that they struggled to fit REDI in, and described strategies they used to cope with the conflicting program and time demands, teachers in the low-sustaining group comments were much more likely to indicate they had altered their behavior and sacrificed the sustainability of REDI due to program and time conflicts, as the following quotations reflect.

From a teacher in the high sustainability group: “Overall seems like the curriculum parts fit tighter. Part of the difficulty we face here is that we have to do themes. But then the REDI stories don't fit with the themes. The stories don't go with the theme so we just end up doing the theme around it. We do some out of order but that doesn't fit with the rest of the curriculum. It's the most difficult part of making it all fit.”

From a teacher in the low sustainability group: “If it had been up to us, we would have continued REDI because we were very pleased with how it fit into our schedule. And as far as this theme based, it has changed things.”

Discussion

Differential Sustainability of the REDI Components

The multi-component nature of the REDI program offered the unique opportunity to examine teacher sustainability of curriculum components that varied considerably in terms of their content and burden. The three sources of information (coach ratings, teacher ratings, teacher interviews) provided very similar estimates of overall sustainability; each revealed that Preschool PATHS was more effectively sustained than the language and emergent literacy components of REDI (ELA and DR). Most teachers sustained PATHS at a level judged to be adequate to affect child outcomes -- 90% of the teachers according to REDI coaches and 100% according to the teachers themselves. In contrast, coach and teacher ratings suggested that only 60% of the teachers sustained ELA and DR at a satisfactory level; comparable estimates based on teacher interviews were 85% sustainability for ELA and 70% sustainability for DR. These findings raise questions about the factors that led to the differential sustainability, given that these REDI components were all introduced and supported in a similar fashion during the research year. Specifically, during the research year, all three of these program components were implemented with high fidelity when teachers had the support and oversight provided by the research team, achieving coach ratings of implementation quality of 4.61 (PATHS), 4.61 (ELA), and 4.39 (DR), indicating implementation levels of “adequate” to “strong” on a 7-point scale. During the sustainability year, the corresponding coach ratings of implementation quality were 4.57 (PATHS), 4.11 (ELA) and 3.81 (DR). These findings indicate that high-fidelity implementation during an initial year of training and professional support is sometimes followed by high-fidelity sustainability, but not always. In addition, the type of curriculum may affect how implementation during a research trial affects high fidelity sustainability. An examination of the correlations between coach ratings of implementation quality during the research trial and coach ratings of sustainability revealed moderate correlations for all three intervention components, but only the correlation for PATHS was statistically significant ($r = .46$).

Although fidelity of implementation during the initial implementation year appears to play a role in promoting program sustainability, other factors are clearly important as well.

Several key differences between PATHS and the DR and ELA program components may have contributed to the differential rates of sustainability. First, PATHS required teachers to initiate instruction in a novel domain, which did not conflict with nor require changes in any of the programming that teachers used prior to the research trial. As evident in the qualitative interviews, PATHS addressed a domain (social-emotional development) that teachers believed was important, but was not well-addressed by their prior programming. In contrast, teachers were already doing book-reading and alphabet activities prior to the introduction of the REDI program. The DR and ELA programs required teachers to change the way they implemented these existing facets of their teaching, replacing (or in some cases adding to) their usual books and activities with a set of curriculum-based books and specific presentation style. At some Head Start centers, teachers were asked to introduce and implement additional emergent literacy interventions alongside the DR and ELA activities during the sustainability year, creating duplication of effort. This probably contributed to the high level of modification of the ELA and DR components by teachers during the sustainability period. According to the teacher interviews, only 10% of the teachers continued to implement DR and ELA as written (vs. 85% who did so with PATHS), whereas an additional 40%–65% sustained DR and ELA but modified the programs. In the qualitative interviews, teachers expressed a belief in the importance of emergent literacy skills for school readiness, but some felt that these skills were receiving too much emphasis.

In addition, the implementation burden was higher for the emergent literacy components (DR and ELA) than it was for PATHS. Specifically, the DR program required teachers to deliver scripted lessons (book-reading or prop book review) four times per week. ELA activities were short in duration, but were designed to occur six times per week (three sound games and three alphabet center visits each week). In contrast, PATHS was taught twice a week (one lesson and one extension activity each week).

All three of these factors may have contributed to the greater sustainability of PATHS compared with the other REDI components – a valued and unique emphasis on social-emotional development, the lack of conflict or duplication with other existing program requirements, and a lower level of burden in terms of frequency of implementation.

Measuring Sustainability: Variation by Source and Method of Assessment

A second goal of this study was to compare the estimates of sustainability obtained from three different methods of assessment – REDI coach ratings, teacher ratings, and qualitative teacher interviews. Interestingly, the within-rater correlations revealed different patterns depending upon the information source. Coach ratings of sustainability for PATHS, DR, and ELA were all significantly inter-correlated, suggesting that coaches tended to see teachers as sustaining REDI at similar levels across all three program components. Sustainability ratings based upon the teacher interviews also showed considerable within-rater consistency, with significant correlations emerging for two of the three pairs of components. In contrast, estimates of sustainability derived from the teacher ratings were more differentiated -- only one of the three correlations was significant, suggesting that teachers perceived considerable variability in the degree to which they were sustaining the various REDI components. To some extent, differences in the assessment items rated by coaches versus teachers may account for these findings. When rating sustainability, coaches were asked to consider the quality of REDI lesson delivery and generalization, as well as quantity of program delivery. Variation across teachers in their instructional skills (e.g., quality of REDI lesson delivery and generalization, enthusiasm for the program) likely contributed to the shared variance evident in coach ratings of sustainability across the three program components. In contrast,

the teacher rating items focused specifically on the degree to which the program was being implemented “as written.” Thus, the teacher ratings emphasized the degree to which lessons followed the manual scripts rather than the quality of lesson delivery and generalization, which may be why their ratings showed greater differentiation and less shared variance across program components.

Interestingly, the degree to which inter-rater agreement emerged across the assessment methods varied considerably depending upon the program component being rated. Assessments of the sustainability of Dialogic Reading were quite consistent across source, with REDI coach ratings, teacher ratings, and teacher interview ratings all significantly inter-correlated. These findings parallel similar levels of significant correlation linking teacher ratings and observer ratings of fidelity that have been reported in previous research (Lillehoj et al., 2004; Melde et al., 2006). Significant inter-rater correlations also emerged between coach ratings and teacher interview ratings assessing PATHS sustainability, although the teacher ratings of PATHS sustainability were not correlated with these other sources. The correlations were probably attenuated by the narrow range and ceiling effects on teacher ratings of PATHS sustainability. In contrast, the information sources did not agree on ELA sustainability; no significant inter-rater correlations emerged. The lack of agreement between the coaches and teachers in ratings of ELA sustainability is consistent with findings from other studies such as Noell and colleagues (Noell et al., 2005). In the current study, coaches were not able to complete observations of the ELA components for 4 of the 20 teachers, reducing the sample size for these comparisons, which may have attenuated the cross-source associations. In addition, given that ELA program activities are quite brief (10 minutes), the classroom visits may have provided coaches with an insufficient basis to evaluate ELA program delivery. In contrast, both PATHS and DR involve longer lessons that are highly interactive, and both also include generalized teaching strategies (e.g., emotion coaching, language expansions) that can be observed throughout the day. Hence, the amount of information that a coach can gather in a single visit about the quality of the PATHS and DR delivery is much greater than for ELA. These findings suggest that the assessment of sustainability is complex, and although observers are better than the teachers themselves at assessing the quality of sustained implementation, they may not be as able to evaluate sustained dose. When quality of lesson delivery and generalization are particularly important elements of a program (as in DR and PATHS, which involve interactive conversations with children), observers are likely to be the better assessors of sustainability than the teachers themselves. In contrast, when the program involves concrete lessons (as in ELA) with impact more dependent upon frequency of presentation rather than quality of teacher-student interaction, teacher documentation of dose may be a better measure of sustainability than an observer’s judgment. It should also be noted that, unlike previous studies (Lillehoj et al., 2004), the teachers in this study did not seem to overestimate their rates of sustainability.

It is notable that REDI coach ratings were more likely to show significant correlations with the teacher interviews than with the teacher ratings (significant for DR only). These findings suggest that teacher interviews may be a better means of gathering information about sustainability than teacher ratings, as previous researchers have suggested. For example, Resnicow and colleagues (Resnicow et al., 1998) found that a teacher self-report questionnaire was not a valid measure of implementation completeness (i.e. quantity or dosage) compared to a teacher interview procedure. In the present study, the REDI coach ratings of sustainability likely indexed quality of sustained implementation much more than dosage. Since teachers were asked to demonstrate the REDI components during the observers visit, observers could assess competence and fidelity of implementation, but they had to rely on indirect measures of teacher report and planning logs to assess dosage. In the detailed face-to-face interviews with the teachers about their sustainability, teachers gave

full descriptions of the ways in which they were sustaining the REDI components, providing information that allowed the interviewer to rate quality as well as quantity of sustained program use. The open-ended interviews also allowed teachers to more fully elaborate on what parts of the curriculum they were continuing to use and what changes they had made, providing contextualized accounts of their sustained program use. In contrast, the teacher ratings were limited, focusing on the extent to which they implemented REDI lessons as written. Possibly, a more elaborated rating scale might have elicited teacher evaluations that captured qualitative, as well as quantitative, features of program sustainability. However, it is also possible that such qualitative information is difficult to elicit without a more complete interview and/or observation. Although many evaluations may rely on basic teacher reports to assess program sustainability, this study suggests that a teacher rating scale alone is unlikely to capture the depth of information regarding fidelity in sustainability that can be attained by the combination of coach ratings or teacher qualitative interviews.

Issues in the Assessment of Fidelity in Sustained Program Delivery

An important issue that emerged in this study and one that requires greater study in the future involves the conceptualization and assessment of fidelity in sustained program delivery. For example, in this study, rates of sustainability described in the teacher interviews were considerably lower if one used a standard that allowed for little or no deviation from the program as written (10% for DR and ELA; 85% for PATHS) than if one included teachers who continued to use the program, but made modifications during the sustainability year (70% for DR, 85% for ELA, and 100% for PATHS). Some researchers suggest that some level of program adaptation should be allowed and even encouraged during the sustainability phase, in order to foster ownership of the intervention and promote long-term sustainability (Bumbarger & Perkins, 2009; Vaughn, Klingner, & Hughes, 2000). Yet, the danger is that program modifications might reduce program impact on child outcomes (Elliott & Mihalic, 2004; Scheirer, 2005; Greenberg et al., 2005). This study would have benefitted from a closer examination of the adaptations the teachers made. After all, characteristics of classrooms change from year to year depending on the children, and what works with a class one year may be a failure with the next class. Teachers need to have some flexibility built into a curriculum to address the needs of specific students. One of the benefits of the mixed-methods assessment approach was that the teachers were able to give details about changes they had made in their own words. While many changes involved omitting activities due to time constraints, there were other adaptations that undoubtedly strengthened the curriculum. Although this particular study does not capitalize on this wealth of data about adaptations, future research may use such mixed-method approaches in order to begin the complicated evaluation about what common changes may enhance or detract from a newly implemented curriculum. Such information may assist program developers to learn more about core components and identify when to be firm in training and how to support teachers to insert their own creative mark on the program.

While it is likely that a more lenient definition of sustainability would have resulted in higher rates of sustainability for the language and literacy components, at the same time it is important to avoid the pitfall of accepting adapted practice as satisfactory, without evidence that this adapted practice produces at least the same outcomes for children as the version of the intervention examined in the efficacy trial. As noted earlier, schools have a difficult time implementing evidence-based programs with fidelity. In a national study of substance use prevention programs, only 19% of district coordinators reported that their school was implementing a research-based curriculum with fidelity (Hallfors & Godette, 2002). Thus, more strict definitions of implementation and sustainability may help to strengthen the field of school-based intervention. Once knowledge about core components has accumulated,

along with knowledge about the best means of fostering high-quality implementation, then a more lenient definition which allows for adaptation may be appropriate.

Facilitating Sustainability

The third goal of this study was to explore the reasons why teachers were (or were not) sustaining the REDI curriculum components by examining the qualitative interviews. These factors shed some additional light on the question of why PATHS was sustained at higher rates than the language and literacy components. As noted earlier, PATHS, as a social-emotional curriculum, filled a need that many of the Head Start classrooms had, and this may have contributed to the high level of its continued use. In contrast, the REDI literacy components were just another approach to teaching literacy skills that were already underway to some extent in the classrooms, and this appeared to decrease teachers' commitment to sustaining the REDI language and literacy components with the same level of fidelity.

In general, competing activities and mixed messages regarding program commitment and support emerged clearly in the teacher interviews as barriers to the sustained use of the language and literacy REDI components, and these barriers seemed to have a greater direct impact on teaching behavior than did teacher perceptions of the quality or perceived benefits associated with the REDI. That is, teachers saw good things in all the REDI curriculum components (i.e. benefits to students, high levels of student engagement, etc), but that did not lead to equally high levels of sustainability for all components. Even teachers who had abandoned use of a particular component said positive things about it. On the other hand, there were fewer barriers mentioned, but when a teacher made a statement about a barrier, it usually was accompanied by statements reflecting reduced levels of sustainability. Interestingly, most of the barriers to sustainability mentioned in the qualitative interviews, including conflicts with other program requirements, insufficient time to implement, limited program support for implementation, and developmentally inappropriate for three-year-olds – referred to the language and literacy components of REDI (DR and ELA) and not to PATHS. PATHS filled a curriculum void, so there were no competing requirements and teachers made sure to allow time for PATHS. These findings suggest that future efforts to increase sustainability rates may be most productive if focused on reducing barriers such as competing requirements rather than simply trying to increase support for the program by identifying benefits.

Of particular note, perceptions that the Head Start program administrators required teachers to continue using REDI emerged as a particularly notable factor associated with sustained program use – mentioned by 83% of teachers who sustained all REDI components at a high level versus only 16% who sustained REDI at low levels. A strong commitment by the supervisors and program directors may be critical to promote sustainability.

Limitations

There were several limitations to this study. First, this study only examined the first year of sustainability. As monthly coach support was still available for the teachers, only the experience of the *transition* to sustainability was captured rather than the typical experiences in a year of sustainability when all outside funding and support has ended. Sustainability is often considered to be the extent to which a program is sustained for at least two years after the implementation year (Bumbarger & Perkins, 2009; Elias, Zins, Graczyk, & Weissberg, 2003; Glasgow, Vogt, & Boles, 1999). Hence, this study may have over-estimated sustainability by measuring it only during the first year after the research program.

Second, this study focused on the degree to which the REDI curriculum was sustained, and did not systematically address the many factors that may influence sustainability. Factors such as teacher experiences, classroom quality, administrative support, attributes of the intervention, teachers' perceptions of the curriculum materials, the quality/quantity of training, and the system of on-going technical assistance, as well as broader factors such as the political climate and the educational structure of the Head Start program, all may influence sustainability (Greenberg et al., 2005). Some of these factors were briefly mentioned in the interviews, some are explored elsewhere (Sanford DeRousie, 2008), but most were not addressed in this study.

Third, there were a number of design features that warrant mention as study limitations, including the small sample size and the use of a single assessment point to collect data on sustainability. In addition, teacher ratings and interviews were collected mid-winter (January–February), whereas coach ratings were collected at the end of the academic year (April–May), and the different time schedules for assessments may have reduced cross-rater correlations. The specific questions used to assess sustainability were different for teachers and coaches, and asked for global impressions. This precluded our ability to differentiate the degree to which quantity versus quality of sustainability was being assessed. Future research would benefit from the use of a larger set of items designed to tap these dimensions separately.

A final limitation is based on the fact that a critical piece of sustainability is documenting whether positive results are maintained over time (Pluye et al., 2004; Shediak-Rizkallah & Bone, 1998). Unfortunately, this study did not include measures to document whether the benefits of REDI (i.e. increased school readiness) were maintained with a new cohort of preschoolers who received the intervention during the sustainability period. Although the standard assumption is that continued high-quality implementation would translate into replication of positive effects, this is only an assumption. Future studies of sustainability would do well to include measures of the desired outcomes in the current cohort of participants as part of the examination of sustainability.

Conclusion

This study provides evidence that sustaining an evidence-based preschool curriculum in Head Start classrooms is feasible. That REDI was sustained in the year following the intervention year is encouraging given that sustained use was not a primary focus of the Head Start REDI efficacy trial. This study also suggests that future efforts to sustain preschool evidence-based programs may want to focus on the elimination of barriers, such as competing requirements, in order to maximize sustainability. Although this study serves as a starting point for studying the sustainability of early childhood evidence-based programming, the findings highlight the importance of including multiple methods for measuring sustainability, including both quantitative and qualitative means, in future research.

Research Highlights

We examined the sustainability of a preschool curriculum by Head Start teachers.
The social-emotional component was sustained more than the literacy component.
In-depth observations and interviews were best sources of fidelity in sustainability.
Teacher sustainability likely to be affected by barriers such as competing demands.

Acknowledgments

This project was supported by an ACF Head Start Graduate Student Research Grant, HHS-2006-ACF-OPRE-YD-0068, and by National Institute of Child Health and Human Development grants, HD046064 and HD43763. Appreciation is expressed to the teachers, students, parents, and program personnel who served as partners in this project in the Huntingdon, Blair, and York County Head Start Programs of Pennsylvania.

References

- Adams, MJ.; Foorman, BR.; Lundberg, L.; Beeler, T. Phonological sensitivity in young children: A classroom curriculum. Baltimore, MD: Brookes; 1998.
- Ball WW, Blachman BA. Does phoneme segmentation training in kindergarten make a difference in early word recognition and developmental spelling? *Reading Research Quarterly*. 1991; 26:49–66.
- Bierman KL, Domitrovich CE, Nix RL, Gest SD, Welsh JA, Greenberg MT, Gill S. Promoting academic and social-emotional school readiness: The Head Start REDI program. *Child Development*. 2008; 79:1802–1817. [PubMed: 19037951]
- Bumbarger BK, Perkins DF. After randomized trials: Issues related to dissemination of evidence-based interventions. *Journal of Children's Services*. 2009; 3:55–64.
- Collaborative for Academic, Social, and Emotional Learning. Safe and sound: An educational leader's guide to evidence-based social and emotional learning programs. Chicago, IL: Author; 2003.
- Domitrovich CE, Cortes RC, Greenberg MT. Improving young children's social and emotional competence: A randomized trial of the Preschool "PATHS" Curriculum. *The Journal of Primary Prevention*. 2007; 28:67–91. [PubMed: 17265130]
- Domitrovich CE, Gest SD, Gill S, Bierman KL, Welsh JA, Jones DJ. Fostering high-quality teaching in Head Start classrooms: Experimental evaluation of an integrated enrichment program. *American Educational Research Journal*. 2009; 46:567–597.
- Domitrovich C, Gest S, Jones D, Gill S, Sanford DeRousie R. Implementation quality: Lessons learned in the context of the Head Start REDI trial. *Early Childhood Research Quarterly*. 2010; 25:284–298.
- Domitrovich CE, Greenberg MT. The study of implementation: Current findings from effective programs that prevent mental disorders in school-aged children. *Journal of Educational and Psychological Consultation*. 2000; 11:193–221.
- Durlak JA. The importance of doing well in whatever you do: A commentary on the special section, "Implementation research in early childhood education." *Early Childhood Research Quarterly*. 2010; 25:348–357.
- Durlak JA, DuPre EP. Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*. 2008; 41:327–350. [PubMed: 18322790]
- Dusenbury L, Brannigan R, Falco M, Hansen WB. A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research*. 2003; 18:237–256. [PubMed: 12729182]
- Dusenbury L, Brannigan R, Hansen WB, Walsh J, Falco M. Quality of implementation: Developing measures crucial to understanding the diffusion of preventive interventions. *Health Education Research*. 2005; 20:308–313. [PubMed: 15522898]
- Elias, ME.; Kamarinos, P. Sustainability of school-based preventive, social-emotional programs: A model site study. Presentation at the annual meeting of the American Psychological Association; Toronto, Canada. 2003 August.
- Elias M, Zins JE, Graczyk PA, Weissberg RP. Implementation, sustainability and scaling-up of social-emotional and academic innovations in public schools. *School Psychology Review*. 2003; 32:303–319.
- Elliott DS, Mihalic S. Issues in disseminating and replicating effective prevention programs. *Prevention Science*. 2004; 5:47–53. [PubMed: 15058912]
- Fixsen, DL.; Naoom, SF.; Blase, KA.; Friedman, RM.; Wallace, F. Implementation research: A synthesis of the literature. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network; 2005.

- Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: The RE-AIM framework. *American Journal of Public Health*. 1999; 89:1322–1327. [PubMed: 10474547]
- Good Start Grow Start: The Bush Administration's Early Childhood Initiative. 2002. Retrieved from <http://www.whitehouse.gov/infocus/earlychildhood/toc.html>
- Gottfredson DC, Gottfredson GD. Quality of school-based prevention programs: Results from a national survey. *Journal of Research in Crime and Delinquency*. 2002; 39:3–35.
- Greenberg, MT.; Domitrovich, CE.; Graczyk, PA.; Zins, JE. The study of implementation in school-based preventive interventions: Theory, research, and practice. Rockville, MD: Center for Mental Health Services, Substance Abuse and Mental Health Services Administration; 2005.
- Griffin JA. Research on the implementation of preschool intervention programs: Learning by doing. *Early Childhood Research Quarterly*. 2010; 25:267–269. [PubMed: 20657800]
- Hallfors D, Goddette D. Will the “principles of effectiveness” improve prevention practice? Early findings from a diffusion study. *Health Education Research: Theory and Practice*. 2002; 17:461–470.
- Han S, Weiss B. Sustainability of teacher implementation of school-based mental health programs. *Journal of Abnormal Child Psychology*. 2005; 33:665–679. [PubMed: 16328743]
- Hansen WB, McNeal RB. Drug education practice: results of an observational study. *Health Education Research*. 1999; 14:85–97. [PubMed: 10537950]
- Harachi TW, Abbott RD, Catalano RF, Haggerty KP, Fleming CB. Opening the black box: Using process evaluation measures to assess implementation and theory building. *American Journal of Community Psychology*. 1999; 27:711–731. [PubMed: 10676545]
- Hill LG, Maucione K, Hood BK. A focused approach to assessing program fidelity. *Prevention Science*. 2007; 8:25–34. [PubMed: 16967341]
- Johnson K, Hays C, Center H, Daley C. Building capacity and sustainable prevention innovations: A sustainability planning model. *Evaluation and Program Planning*. 2004; 27:135–149.
- Lillehoj CJ, Griffin KW, Spoth R. Program provider and observer ratings of school-based preventive intervention implementation: Agreement and relation to youth outcomes. *Health Education and Behavior*. 2004; 31:242–257. [PubMed: 15090124]
- Lonigan CJ. Development, assessment, and promotion of pre-literacy skills. *Early Education and Development*. 2006; 17:91–114.
- Melde C, Esbensen FA, Tusinski K. Addressing program fidelity using onsite observations and program provider descriptions of program delivery. *Evaluation Review*. 2006; 30:714–740. [PubMed: 17093106]
- Noell GH, Witt JC, Slider NJ, Connell JE, Gatti SL, Williams KL, Resetar JL. Treatment implementation following behavioral consultation in schools: A comparison of three follow-up strategies. *School Psychology Review*. 2005; 34:87–106.
- No Child Left Behind Act. No Child Left Behind Act of 2001. Publ. L.; 2002. p. 107-110.
- Odom S, Fleming K, Diamond K, Lieber J, Hanson M, Butera G, Marquis J. Examining different forms of implementation in early childhood curriculum research. *Early Childhood Research Quarterly*. 2010; 25:314–328. [PubMed: 21874091]
- Olds D, Hill PL, O'Brien R, Racine D, Moritz P. Taking preventive intervention to scale: The nurse-family partnership. *Cognitive and Behavioral Practice*. 2003; 10(4):278–290.
- Pankratz M, Jackson-Newsom J, Giles S, Ringwalt C, Bliss K, Bell ML. Implementation fidelity in a teacher-led alcohol use prevention curriculum. *Journal of Drug Education*. 2006; 36:317–333. [PubMed: 17533804]
- Pluye P, Potvin L, Denis JL. Making public health programs last: Conceptualizing sustainability. *Evaluation and Program Planning*. 2004; 27:121–133.
- Porowski AW, Burgdorf K, Herrell JM. Residential substance abuse treatment for pregnant and parenting women. *Evaluation and Program Planning*. 2004; 27:191–198.
- Resnicow K, Davis M, Smith M, Lazarus-Yaroch A, Baranowski T, Baranowski J. How best to measure implementation of school health curricula: A comparison of three measures. *Health Education Research*. 1998; 13:239–250. [PubMed: 10181022]

- Rohrbach LA, Graham JG, Hansen WB. Diffusion of a school-based substance abuse program: Predictors of program implementation. *Preventive Medicine*. 1993; 22:237–260. [PubMed: 8483862]
- Rohrbach LA, Grana R, Sussman S, Valente TW. Type II translation: Transporting prevention interventions from research to real world settings. *Evaluation & the Health Professions*. 2006; 29:302–333.
- Sanford DeRousie, RM. Making changes last in the classroom: Factors contributing to the sustainability of an evidence-based preschool curriculum. San Francisco, CA: Poster presented at the annual meeting of the Society for Prevention Research; 2008 May.
- Scheirer MA. Is sustainability possible? A review and commentary on empirical studies of program sustainability. *The American Journal of Evaluation*. 2005; 26:320–347.
- Shediac-Rizkallah MC, Bone LR. Planning for the sustainability of community-based health programs: Conceptual frameworks and future directions for research, practice and policy. *Health Education Research*. 1998; 13:87–108. [PubMed: 10178339]
- Vaughn S, Klingner J, Hughes M. Sustainability of research-based practices. *Exceptional Children*. 2000; 66:163–171.
- Wasik BA, Bond MA. Beyond the pages of a book: Interactive book reading and language development in preschool classrooms. *Journal of Educational Psychology*. 2001; 93:243–250.
- Wickstrom KF, Jones KM, LaFleur LH, Witt JC. An analysis of treatment integrity in school-based behavioral consultation. *School Psychology Quarterly*. 1998; 13:141–154.
- Whitehurst GJ, Arnold D, Epstein JN, Angell AL, Smith M, Fischel JE. A picture book reading intervention in daycare and home for children from low-income families. *Developmental Psychology*. 1994; 30:679–689.

Table 1

Implementation Levels Based on Coach Ratings, Teacher Ratings, and Teacher Interviews

	REDI Program Components		
	DR	ELA	PATHS
<u>REDI Coach Ratings</u>			
Mean (<i>SD</i>)	3.81 (1.08)	4.11 (.93)	4.57 (1.03)
Range	1–6	2–6	2–6
% “Satisfactory” (4 or higher)	60%	60%	90%
<u>Teacher Ratings</u>			
Mean (<i>SD</i>)	3.55 (1.02)	3.98 (.65)	4.60 (.45)
Range	1.33–5	3–5	3.67–5
% “Satisfactory” (3.5 or higher)	60%	60%	100%
<u>Teacher Interviews</u>			
Mean (<i>SD</i>)	1.80 (.62)	1.95 (.51)	2.85 (.37)
Range	1–3	1–3	1–3
% “Satisfactory” (2 or 3)	70%	85%	100%

Note: DR = Dialogic Reading; ELA = Emergent Literacy Activities. Coach ratings used a 7-point scale (1–7), teacher ratings used a 5-point scale (1–5) averaged across three items, and interviews were rated using a 3-point scale (1–3). *N* = 20 teachers, except for coach ratings of ELA where *N* = 16 teachers.

Table 2

Correlations Among and Between Different Measures of Sustainability by Component

	Coach Ratings			Teacher Ratings			Teacher Interviews		
	DR	ELA	PATHS	DR	ELA	PATHS	ELA	PATHS	PATHS
Coach Ratings									
DR		.59*	.64*						
ELA			.66*						
PATHS									
Teacher Ratings									
DR	.45*	.17	.23		-.27	-.12			
ELA	-.20	-.15	-.34			.52*			
PATHS	-.18	-.25	-.30						
Teacher Interviews									
DR	.64*	.58*	.59*	.46*	-.22	-.38	.47*		.33
ELA	.07	.00	.25	.19	.30	-.12			.52*
PATHS	.30	.14	.64*	-.14	-.07	-.03			

* $p < .05$.

Table 3

Factors Affecting Sustainability from Qualitative Teacher Interviews

Factor	% of Teachers Who Made Comments
<i>Why Sustain</i>	
Teacher sees benefits of REDI in children	100%
Students engage with and use curriculum skills	85%
REDI helps teachers improve their teaching skills	85%
Positive support from teaching team	80%
REDI approach matches teaching style	75%
REDI addresses specific concerns of classroom	65%
REDI helps create a better classroom atmosphere	60%
REDI is easy to use	50%
REDI was required	45%
<i>Why Not Sustain</i>	
REDI conflicts with other program requirements	70%
Not enough time to do REDI	50%
REDI not developmentally appropriate	50%
Limited or no support for using REDI	40%
REDI conflicts with teacher's philosophy and style	40%