Intervention fidelity of Getting Ready for School: Associations with classroom and teacher characteristics and prescamer's school readiness skills

Maria Marti a,⁎, Samantha Melvin b,⁎, Kimberly G. Noble c, Helena Duch a

a Heilbrunn Department of Population & Family Health, Mailman School of Public Health, Columbia University, 60 Haven Avenue, B-2, New York, NY 10032, United States
b Teachers College, Columbia University, 525 W 120th St, New York, NY 10027, United States
c Teachers College, Columbia University, Thorndike Hall, 525 W 120th St., Box 199, New York, NY 10027, United States

A R T I C L E   I N F O
Article history:
Received 29 December 2016
Received in revised form 27 December 2017
Accepted 5 February 2018

A B S T R A C T
Getting Ready for School (GRS) is a new school readiness intervention for teachers and parents, designed to help children develop early literacy, math, and self-regulation skills. GRS was implemented in 19 Head Start classrooms. In the present study we examined variability in different aspects of intervention fidelity including dosage, adherence, and child engagement. In addition, we studied the association among classroom, teacher and student characteristics and fidelity, and whether measures of fidelity were associated with children’s growth in math, early literacy, and self-regulation skills across the preschool year. Findings indicate that on average teachers reported completing almost 80% of the activities assigned, and that they were observed to adhere fairly well to the lessons. Child engagement was observed to be moderate to high across classrooms. Classroom quality, as measured by the CLASS, and age of children were positively associated with adherence. Teachers that had participated in GRS for two years were more likely to complete more activities. Different components of fidelity were associated with child outcomes. Percentage completion of math and literacy activities were positively associated with growth in math and literacy skills. Children in classrooms in which teachers adhered more faithfully to the curriculum made significantly greater gains in literacy, math, and self-regulation skills. Child engagement was positively associated with a measure of executive function. Results highlight the importance of examining implementation fidelity. Implications for preschool teachers are discussed.

© 2018 Elsevier Inc. All rights reserved.

1. Introduction

School readiness—the combination of skills that includes pre-academic knowledge, such as basic literacy and math foundations, as well as self-regulatory skills, such as social–emotional regulation, patience, and focus—has become a critical component of Early Childhood Education (ECE) programs in the United States. Decades of research has shown that socioeconomically disadvantaged children tend to enter school behind their peers in terms of these skills (Bradley & Corwyn, 2002; Noble, Norman, & Farah, 2005; Noble et al., 2012), creating the foundation for an achievement gap that only grows over time (Reardon, 2011; West, Denton, & Germino-Hausken, 2001). Notably, a growing body of research suggests that the quality of ECE programs is a critical element in boosting school readiness (Duncan, 2003), yet “high-quality” is difficult to define and even more difficult to achieve in the resource-limited programs that serve the children who need the boost the most (Kagan, 2009). As an increasing number of states and cities expand their early childhood services, emphasis is placed not only on the number of programs and seats available to families, but also on their quality and success in helping children be prepared for school.

Recent studies have shown that returns on investment in ECE may only be significant for high-quality programs (Duncan & Magnuson, 2013), certain components of which (including class size reduction and teacher credentials) are difficult and expensive to infuse into existing programs. Thus, numerous interventions in recent years have focused on targeting specific school readiness skills, in the hopes that building teacher practice in terms of expanding these skills will be associated with increasing quality more generally (Clements & Sarama, 2013; Clements, Sarama, Wolfe, & Spiteri, 2013; Wilson & Farran, 2012). This method...
seems to be particularly effective for relatively low-cost, easy-to-implement interventions.

Even so, it is only more recently becoming common for interventions that take this approach to discuss intervention fidelity when reporting outcomes in early childhood interventions (Clements, Sarama, Wolfe, & Spitzler, 2015; Mendive, Weiland, Yoshikawa, & Snow, 2016). Intervention fidelity (IF) refers to the degree to which intervention practices and core components are implemented as intended; herein, the degree to which teachers implement activities assigned by coaches and adhere to the lesson plans is discussed. Without an understanding of the feasibility and fidelity of a new intervention, it is impossible for educators and policymakers to know whether or not an intervention could be brought to scale. In this paper, we will present IF findings from a novel school readiness intervention, Getting Ready for School, as well as associations among IF, classroom- and teacher-level variables, and child outcomes.

1.1. Getting Ready for School

Getting Ready for School (GRS) is an intervention that aims to simultaneously target three main school readiness areas: literacy, math, and self-regulation. Based on the principle that children learn best when receiving high-quality, reinforced learning opportunities both at home and in school settings (Bronfenbrenner, 2005), GRS is a flexible, supplemental curriculum for teachers and parents, designed to help children develop the early literacy, math, and self-regulation skills they need to start primary school ready and eager to learn.

1.1.1. Intervention history

Initially designed as a math and literacy curriculum for parents in Central and Eastern Europe who lacked access to traditional preschools, the original GRS intervention was developed and implemented in partnership with the Open Society Foundation in several countries in Central and Eastern Europe, and aimed to prepare children for primary school by helping parents create more effective home learning environments. Parents attended monthly workshops and received printed materials with instructions for play-based learning activities that could be done with everyday materials. In 2010, the intervention was piloted in the USA and demonstrated potential to improve literacy and math skills in participating children relative to those receiving Head Start–as-usual (Noble et al., 2012). Therefore, the program sought to expand based on promising results and positive feedback from parents. The expansion involved three components: (1) design of a complementary classroom component, coupled with similar training for teachers, in order to reinforce skill building in both contexts and to foster parent–teacher communication, (2) the integration of self-regulation skills throughout both curricula (classroom and home), and (3) design of a video library of parent-based activities, to allow parents to observe demonstrations of home-based activities. This expansion was undertaken as a four-year process, with the first year dedicated to curriculum and intervention development, the second year dedicated to piloting the intervention with an iterative revision process and assessing its feasibility, and the third and fourth years aimed to pilot the preliminary efficacy of the GRS intervention collecting fidelity, child outcome data, and classroom quality measures. The three years of piloting the GRS intervention are the subject of the present paper.

1.1.2. Intervention rationale

School readiness skills such as early literacy, language development, math, and self-regulation do not develop in isolation from one another. Recent research has identified relevant correlations and predictions across domains, and the directionality of these relationships remains unclear. For instance, oral language has been shown to predict trajectories of self-regulation (Vallotton & Ayoub, 2011), and self-regulation at kindergarten has been associated with reading and math skills at 6th grade (McClelland, Acoc, & Morrison, 2006; McClelland et al., 2014). There is also strong evidence showing correlations between math and literacy skills in preschool children (Purpura, Hum, Sims, & Lonigan, 2011). Yet, there is no clear evidence that one school readiness skill predicts academic development above and beyond the others (Bailey, Duncan, Watts, Clements, & Sarama, 2018). Therefore, it is crucial to offer teachers the tools to foster the development of all school readiness skills with equal emphasis. For these reasons, the primary aim of GRS is to comprehensively address school readiness by promoting literacy, math, and self-regulation skills with equal emphasis. GRS was designed as a supplemental curriculum that addresses multiple school readiness skills simultaneously, so that teachers can experience the benefits of content-focused goal-directed curricula that explicitly support skill development, without the additional burden of implementing three completely separate content curricula in addition to other programmatic requirements.

Indeed, results from the Preschool Curriculum Evaluation Research Initiative revealed that all but two comprehensive preschool curricula (out of 14 evaluated) failed to increase children’s academic readiness (Preschool Curriculum Evaluation Research Consortium, 2008). Recent work has also suggested that content-specific curricula (targeted at a single area such as math or literacy) may be more effective at generating gains in corresponding skills compared to global curricula such as High/Scope (Hohmann, Weikart, & Epstein, 2008) and Creative Curriculum (Dodge, Colker, Heroman, & Bickart, 2002) that focus more on holistic development through exploratorium than on explicit skill growth through a focus on content areas (Duncan et al., 2015). While the most effective approach to boosting school readiness skills might involve implementing multiple content-focused curricula at once, it is understandably quite difficult for teachers to incorporate multiple supplemental curricula into their daily routines (Duncan et al., 2015). Numerous interventions explicitly target one of these components of school readiness (Clements & Sarama, 2007; Jones, Bailey, & Jacob, 2014; Lonigan, Farver, Phillips, & Clancy-Menchetti, 2011), and several target two (Barnett et al., 2008; Bierman, Dimitrovich et al., 2008; Bierman, Nix, Greenberg, Blair, & Domitrovich, 2008); however, GRS is one of the first interventions to take a fully integrated perspective, incorporating three specific components of skill-based content into a single package. Moreover, while some studies have found evidence of spillover effects into readiness domains outside of the targeted content area, e.g. improvement in literacy or executive function skills observed with a mathematics intervention; (Sarama, Lange, Clements, & Wolfe, 2012; Weiland & Yoshikawa, 2013), the sizes and sustainability of these effects are not yet clear. We posit that by allowing teachers to explicitly target multiple domains of school readiness simultaneously, it is more likely that gains will be larger across each of the addressed domains compared to the small spillover effects that may result from focusing on a single content area.

1.1.3. Intervention design

GRS is made up of classroom, home, and teacher–parent communication components. Broadly speaking, the classroom component consists of a series of supplemental classroom activities designed to promote literacy, math, and self-regulation development. Literacy and math activities include learning games (such as “Match the Letters” or “Racing Shapes”), movement activities (such as “Clap the Syllables” or “Jump to the Number”), or poster activities (involving tasks such as analyzing poems, talking about images, or sorting and counting objects).
Table 1
Scope and sequence of GRS parent component.

<table>
<thead>
<tr>
<th>Unit</th>
<th>Math</th>
<th>Literacy</th>
<th>Self-regulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Counting to determine quantity, counting sets of objects up to 5, comparing groups with “more” and “less”</td>
<td>Letter recognition, letter–sound correspondence, print conventions</td>
<td>Paying attention, self-control skills</td>
</tr>
<tr>
<td>2</td>
<td>Sorting objects into groups, counting sets of objects up to 5, recognizing numerals 1–5</td>
<td>Print conventions, sound and letter recognition name, identifying rhymes, letter–sound correspondence</td>
<td>Recognizing, managing, and talking about feelings</td>
</tr>
<tr>
<td>3</td>
<td>Recognizing numerals 6–10, counting sets of objects up to 10, sorting objects, comparing groups of objects</td>
<td>Expressive language, creating stories, phonological awareness, identifying rhymes</td>
<td>Talking about feelings of characters in stories</td>
</tr>
<tr>
<td>4</td>
<td>Recognizing numerals 6–10, counting sets of objects up to 10</td>
<td>Identifying favorite words, expressive language, identifying rhymes, creating stories</td>
<td>Working memory, identifying and talking about feelings</td>
</tr>
<tr>
<td>5</td>
<td>Measuring length and weight, comparing objects, making predictions</td>
<td>Letter recognition, print conventions, expressive language, writing, letter–sound correspondence</td>
<td>Paying attention, working memory, self-control, managing emotions</td>
</tr>
<tr>
<td>6</td>
<td>Identifying, recognizing, and counting shapes; sorting objects by shape, color, and size</td>
<td>Expressive language, rhyming, story sequencing</td>
<td>Thinking about feelings, making lists to help children follow directions, remember, and complete tasks</td>
</tr>
<tr>
<td>7</td>
<td>Identifying and making patterns</td>
<td>Identifying words, expressive language, creating stories, story sequencing</td>
<td>Working memory, paying attention using the “if-then” rule</td>
</tr>
<tr>
<td>8</td>
<td>Practicing simple addition</td>
<td>Reading and writing, expressive language, recognizing words</td>
<td>Taking turns, cooperation and working together</td>
</tr>
<tr>
<td>9</td>
<td>Practicing simple subtraction, learning about zero</td>
<td>Reading and writing, story sequencing, expressive language</td>
<td>Paying attention, discussing feelings about kindergarten</td>
</tr>
</tbody>
</table>

Following the model originally constructed by Whitehurst and Lonigan (1998), GRS targets emergent literacy through the development of both “outside-in” and “inside-out” skills. “Outside-in” skills represent children’s understanding of the linguistic context and include the semantic and conceptual knowledge that comprises oral language; knowledge of narrative structure and narrative-building skills; and print conventions (e.g. left-to-right and front-to-back). “Inside-out” skills comprise children’s knowledge of the rules for translating the particular writing they are trying to read into sounds, and includes knowledge of graphemes, phonological awareness, phoneme–grapheme correspondence, and emergent writing. Experience with “outside-in” skills fosters development of oral language and vocabulary, whereas experience with “inside-out” skills fosters development of early reading skills (Bowyer-Crane et al., 2008; Connor, Morrison, & Slominski, 2006).

GRS literacy activities include a “Morning Message” activity that grows in complexity in each unit to target conventions of print, as well as stand-alone literacy activities that build various important concepts.

To support the acquisition of math skills, GRS includes activities involving numbers and operations; geometry and spatial sense; measurement; algebra; and data analysis (Clements, Sarama, & DiBiase, 2004; National Association for the Education of Young Children & National Council of Teachers and Mathematics, 2002). These skills are taught through five main processes: problem solving, reasoning, communicating, connecting, and representing (National Association for the Education of Young Children & National Council of Teachers and Mathematics, 2002). For example, when children are prompted to compare the size of different objects in their environment, they learn about measurement. Similarly, when children count sides of shapes, they learn about numbers and geometry and make the connection between these two mathematical concepts (Butera et al., 2014). In GRS, each unit tackles a different math skill, including one-to-one correspondence, counting, sorting and comparison, pattern identification, and measurement.

To support the development of executive function and social emotional development, self-regulation activities were developed in partnership with the team of investigators who designed the Social Emotional Cognitive Understanding and Regulation (SECURE) curriculum (Jones et al., 2014). SECURE acts as a bridge between the self-regulation skills children need and the characteristics of high-quality teachers and classrooms (Bailey & Jones, 2013). Key themes and activities from SECURE’s comprehensive cognitive, emotional, and behavioral self-regulation intervention were carefully integrated into the GRS curriculum. GRS incorporates self-regulation skills into every unit, using both standalone activities and additions to literacy and math activities. The skills developed in these lessons include cognitive skills (working memory, inhibitory control, attention control, and cognitive flexibility) and social–emotional skills (emotional knowledge, expression, and management, empathy, prosocial behavior, coping skills, and conflict resolution). Classroom activities include daily “Brain Games,” which are fun, engaging, often-familiar games (such as a modified “Freeze Dance”), designed to build executive function skills by increasing in complexity over time. Standalone self-regulation lessons, which may be implemented up to twice a week, teach specific skills and strategies through signals, classroom tools, and teaching practices. These self-regulation tools can be integrated into other activities as well as normal classroom interactions and transitions.

These literacy, math, and self-regulation activities are organized into a single classroom manual composed of nine units, following a developmentally appropriate trajectory (see Table 1). It is important to emphasize that the GRS manual and the intervention more broadly were designed to provide high-quality, content-focused activities that can be easily integrated with the requirements of other widely used, more global curricula. All activities last under 15 min; can be done with the whole group, in small groups, or one-on-one; have suggestions for implementation and scaffolding; are designed to be repeated; and are flexible to maximize integration with any primary curriculum or schedule (see Appendices for a sample lesson). Moreover, the manual includes additional tools to help teachers choose activities that will easily integrate with their chosen themes or target skills, such as corresponding standards (from the Head Start Quality Framework) for each activity, and an Appendix of additional tools such as skill and content indices, to assist in customization and integration.

Teachers are asked to implement two to three GRS activities each day, for a total of about 30–45 min. The full classroom manual includes 79 literacy activities, 47 math activities, and 50 self-regulation activities. In a given week, “Morning Message” and “Brain Games” were designed to be implemented daily, along with two to three stand-alone activities that could be repeated over the course of the week. To the extent possible, each activity was
designed to synthesize multiple readiness skills, with the goal of being truly integrative across targeted content areas.

Throughout program implementation, teachers received a full-day introductory training and individualized support from a classroom coach, who was also a member of the research team. Coaches met with teachers once a week (alternating between weeks of coaching and weeks of observation) for approximately eight months. During coaching meetings, which typically lasted between 30 and 60 min, teachers planned the implementation of GRS activities and reflected on previous activities implemented. On alternate weeks, coaches observed classroom activity implementation, with modeling and live coaching support sometimes accompanying these sessions. Coaching procedures were modeled after the principles of Practice-Based Coaching (PBC; National Center on Quality Teaching and Learning, 2015), whereby a cyclical model of teacher–coach planning, focused observation, and reflection is used to guide classroom practices. All classroom coaches either held or were pursuing a Masters or PhD degree in educational or psychological fields. Coaches received rigorous training at the beginning of the year, including on principles of PBC, and participated in biweekly group meetings.

The home component of the curriculum consists of literacy, math, and self-regulation activities that are very simple and can be done with everyday household materials. Activities are meant to integrate targeted developmentally appropriate learning into everyday family life. As with the classroom component, these activities are organized into nine units following a developmentally appropriate trajectory. Parents have the flexibility to work through this book on their own or to follow along with parallel activities suggested by the child’s teacher in the weekly letters (the primary instrument of teacher–parent communication in the program). We also provide one tablet per classroom that has a collection of videos demonstrating how to implement the program’s activities at home by showing other parents and their children doing the activities. Tablets are available in every classroom and can be checked out for a week, like a library book. Parents also have access to a program website where they can see all the videos and download all the activities. All materials are available in both Spanish and English and were designed with special attention to culturally diverse and low-literacy parents.

Parent knowledge and practice is supplemented with twelve events, led by GRS staff, over the course of the school year. Parents are invited to attend workshops where they can learn about learning and skill development, as well as in-class events where teachers and GRS staff work with parent–child dyads in the classroom on activities from the parent book.

1.2. Measures of intervention fidelity (IF) and links with child outcomes

A wealth of research suggests that the level of intervention fidelity, or the degree to which intervention practices and core components are implemented as intended, affects the program’s outcomes (for a review: Durlak & DuPre, 2008). Recent research has differentiated intervention fidelity from fidelity of implementation, which refers to the degree to which any procedures—such as coaching, in-service training, or any other kind of evidence-based professional development practice—are used as intended to support the adoption and use of core intervention practices (Dunst, Trivette, & Raab, 2013; Hulleman, Rimm-Kaufman, & Abry, 2013). The current study focuses on intervention fidelity, which is most commonly measured by two constructs: dosage and adherence (Darrow, 2010). Dosage refers to the amount of time that the intervention is actually implemented, while adherence refers to the degree to which implementation by the teacher matches the design of the intervention (Durlak, 2010). Evidence is mixed about the degree to which dosage and/or adherence is linked to student outcomes. For example, Hamre et al. (2010) found significant associations with dosage but not adherence, and others have found similar mixed effects of one measure of IF or the other (see Mendive et al., 2016).

The lack of consistent findings is due in part to the lack of consensus measuring IF. Researchers measure these constructs in a variety of different ways. Dosage is most commonly measured by counting the number of intervention activities completed by the teacher in the classroom or number of minutes spent on these activities in a given time period (Hamre et al., 2010). Adherence is typically measured by observing the actual implementation of activities to determine fidelity to the intervention’s supported practices (Stein et al., 2008). Measures of adherence vary widely because they are informed by the intervention. Some researchers advocate for directly observing implementation rather than relying on teacher-reported measures of fidelity (Schulte, Easton, & Parker, 2009), though this is more challenging methodologically.

In addition to dosage and fidelity, research has pointed to other dimensions of implementation, such as participant responsiveness, which has been conceptualized as participants’ reaction to the intervention, including sustained interest and attention (Dmitrovich et al., 2008; Dmitrovich, Gest, Jones, Gill, & DeRousie, 2010; Dusenbury, Brannigan, Hansen, Walsh, & Falco, 2005). This has been operationalized by measuring children’s affective and behavioral responses to the intervention (e.g. to what extent is the child engaged in and enjoying the activity). Although participant responsiveness has been often omitted in IF research, some studies have linked IF to child outcomes. For example, in the implementation study of the Head Start REJI intervention, child engagement was positively associated with a range of social–emotional skills, including teachers’ ratings of social competence and aggression, and quality of children’s problem-solving (Dmitrovich et al., 2010). In the present study, we use multiple dimensions of IF (dosage, adherence, and child engagement) to operationalize and disentangle the various aspects of IF.

1.3. Impact of classroom, teacher, and curriculum characteristics on intervention fidelity

Studies show that numerous ecological factors affect IF (Durlak & DuPre, 2008), and as such it is critical to examine both IF and program outcomes through a contextual lens. However, evidence is mixed across studies about the degree to which ecological factors matter. For example, it is unclear whether or not teacher experience and education levels influence IF (Baker, Kupersmidt, Voegler-Lee, Arnold, & Willoughby, 2010; Knoche, Sheridan, Edwards, & Osborn, 2010), or whether even more proximal measures such as perceived need for a program are more predictive (Durlak, 2010). Classroom characteristics such as larger classroom size and teacher–child ratio have a negative impact on classroom quality (Barnett, Schulman, & Shore, 2004; Pianta, La Paro, Payne, Cox, & Bradley, 2002), and recent research suggests that larger classroom size is associated with lower adherence fidelity (Zvolch, 2009). In the present study, we examine both classroom-level factors such as classroom size, teacher–child ratios, child age distribution, and classroom quality as measured by the CLASS (Pianta, Karen, Paro, & Hamre, 2008) and teacher–level factors (including years of education and experience, as well as years in the center) in relation to fidelity to GRS.

1.4. The present study

Although GRS incorporates both classroom and home components, in the present paper, only teacher fidelity to the classroom component of the curriculum will be analyzed. The results regarding the home component are presented in a separate study.
Table 2

Teacher- and coach-reported measures of fidelity (averages by classroom).

<table>
<thead>
<tr>
<th>Question</th>
<th>Response type</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coach meeting teacher feedback worksheet</td>
<td>Approximately how many activities were completed each week?</td>
<td>Number (converted from activity names above)</td>
<td>19</td>
<td>2.43</td>
<td>.92</td>
</tr>
<tr>
<td></td>
<td>How many minutes did you spend on GRS activities each day?</td>
<td>Likert scale [1 = less than 10 min, 2 = 10–30 min, 3 = 30–60 min, 4 = more than 60 min]</td>
<td>14</td>
<td>1.99</td>
<td>.41</td>
</tr>
<tr>
<td>Coach meeting coach feedback worksheet</td>
<td>How effective was the meeting?</td>
<td>Likert [1 (extremely ineffective) to 5 (extremely effective)]</td>
<td>19</td>
<td>3.89</td>
<td>.52</td>
</tr>
<tr>
<td></td>
<td>To what extent do you think GRS is being done with correct dosage?</td>
<td>Likert [1 (never) to 5 (always)]</td>
<td>14</td>
<td>2.95</td>
<td>.97</td>
</tr>
<tr>
<td></td>
<td>To what extent do you think GRS is being done with fidelity?</td>
<td>Likert [1 (never) to 5 (always)]</td>
<td>14</td>
<td>3.68</td>
<td>.77</td>
</tr>
<tr>
<td>GRS activity observation form-coach report</td>
<td>Adherence</td>
<td>Likert [1 (strongly disagree) to 5 (strongly agree)]</td>
<td>19</td>
<td>4.27</td>
<td>.37</td>
</tr>
<tr>
<td></td>
<td>Used required materials or adequate substitutes</td>
<td></td>
<td>19</td>
<td>4.76</td>
<td>.24</td>
</tr>
<tr>
<td></td>
<td>Emphasized the listed vocabulary</td>
<td></td>
<td>19</td>
<td>4.31</td>
<td>.41</td>
</tr>
<tr>
<td></td>
<td>Targeted core concepts from the GRS lesson plan</td>
<td></td>
<td>19</td>
<td>4.23</td>
<td>.51</td>
</tr>
<tr>
<td></td>
<td>Executed the lesson in its entirety (as written in lesson plan)</td>
<td></td>
<td>19</td>
<td>3.80</td>
<td>.66</td>
</tr>
<tr>
<td></td>
<td>Child engagement</td>
<td></td>
<td>19</td>
<td>4.32</td>
<td>.32</td>
</tr>
<tr>
<td></td>
<td>Seemed enthusiastic about starting the activity</td>
<td></td>
<td>19</td>
<td>4.34</td>
<td>.30</td>
</tr>
<tr>
<td></td>
<td>Paid attention during the activity</td>
<td></td>
<td>19</td>
<td>4.27</td>
<td>.41</td>
</tr>
<tr>
<td></td>
<td>Seemed to enjoy the activity</td>
<td></td>
<td>19</td>
<td>4.45</td>
<td>.29</td>
</tr>
<tr>
<td></td>
<td>Appeared to understand the activity</td>
<td></td>
<td>19</td>
<td>4.19</td>
<td>.44</td>
</tr>
</tbody>
</table>

Note: Some questions were only used in Cohorts 2 and 3 indicated by an N= 15.

In order to draw conclusions about the potential promise of this integrated intervention, we examined the feasibility and fidelity of the classroom component of the intervention. Our research questions in this study were as follows:

1. How feasible is a supplemental, integrated school readiness intervention that taps skills across the domains of literacy, math, and self-regulation?
2. How do classroom and teacher characteristics inform intervention fidelity?
3. Are different measures of IF (dosage, adherence, and child responsiveness) associated with growth on child's self-regulatory and preacademic skills during the academic year?

2. Method

2.1. Participants

Twenty-two Head Start classrooms across 3 Cohorts participated in the Getting Ready for School (GRS) intervention. Each Cohort had different classrooms, children, and teachers. In Cohort 1, the GRS curriculum and implementation process was still in the development phase, so it primarily focused on the feasibility and fidelity of the intervention, along with the iterative development and finalization of curricular materials. In contrast, both the curriculum and the intervention design were complete and final in Cohorts 2 and 3, and so a more formal pilot—including updated IF measures, teacher surveys, and child outcomes—was conducted. The pilot study took place over three consecutive years—Cohort 1 in 2013–2014, Cohort 2 in 2014–2015, and Cohort 3 in 2015–2016—with each lasting approximately eight months, beginning in November and ending in June.

The 22 classrooms that participated were distributed across Head Start programs in a large urban setting. Seven classrooms in three centers were enrolled in Cohort 1 of implementation. The lead teachers in four classrooms had previous experience with GRS content, because they had been involved in the first year of the intervention development process. In Cohort 2 (full pilot), seven new classrooms in four new centers were enrolled. All teachers were new to GRS. In Cohort 3 (full pilot), eight new classrooms in four centers were enrolled, all of which completed the program. Across the three Cohorts, ten classrooms served exclusively 4-year-old children who were preparing for kindergarten, nine classrooms comprised rising kindergarteners and 3-year-old children who were not transitioning to kindergarten (mixed age classrooms), and three classrooms in Cohort 1 comprised all 3-year-old children. Across Cohorts, the average percentage of 3-year-old children per classroom not transitioning to kindergarten was 21.85% (SD = 30.83%; see Table 2 for more detailed information). On average, the student–teacher ratio was 8.28 students to each teacher (SD = 1.32, range = 6–10 students to each teacher). It is important to note that the lead teachers in three classrooms had participated in GRS before (two in Cohort 1 and one in Cohort 2). Therefore, in subsequent analysis, we exclude the first round of data for these three teachers given that data collected from the same teacher in two Cohorts could not be treated as independent. The final sample of classrooms included in this study is nineteen.

Nineteen lead teachers participated in this study, all of whom were female. Most teachers were Latina (77.3%), three were Asian (13.6%), and two were African American (9.1%). Additional demographic and experiential information was collected for teachers in Cohort 2 and 3, as shown in Table 2. Most teachers held a Master's degree (85.7%), and on average, they had been teaching in an early childhood center for 11 years. While assistant teachers occasionally assisted in the implementation of activities, lead teachers were the
primary participants of interest in this study and therefore we only include data on lead teachers.

In addition, in Cohorts 2 and 3, we examined the effects of GRS on children’s school readiness skills. A total of 274 children enrolled in participating classrooms during these years (128 in Cohort 2 and 146 in Cohort 3). Overall, 85% (total n = 233) of families in selected classrooms consented to participate in the evaluation, 4.7% (n = 13) dropped school before consent, 2.9% (n = 8) enrolled at the end of the academic year, 6.2% (n = 17) did not return the consent form, and three families refused to participate. Of the 233 families that consented to participate, 15 left the site shortly after consenting to the study, and 4 signed the consent past the baseline testing period; and therefore lacked baseline data. The majority of children were Hispanic (78%), 47% were male, and mean age at pretest ranged from 40 to 59 months (M = 50.34 months; SD = 4.18).

All procedures presented in this study were approved by the Columbia Medical Center’s Institutional Review Board.

2.2. Implementation design

The coaching structure was the same across all Cohorts of implementation, with teachers receiving a full-day introductory training at the beginning of the school year, along with an introductory meeting with their classroom coach, wherein they could ask follow-up questions about the program. Teachers also received weekly coaching support (alternating between planning meetings and observation, as described above) by a trained GRS staff member, who observed implementation, assisted in planning integration to maximize child learning, and facilitated reflective discussion on implementation.

In Cohort 1, the implementation schedule was prescriptive. At the beginning of each unit, teachers were given a calendar of pre-selected activities to implement each week. In a typical week, the target dosage was 4–5 different activities per week, including a daily “Morning Message” activity, a daily self-regulation “Brain Game;” and 2–3 activities on math, literacy, or self-regulation, each to be repeated twice that week. The only teacher-level individualization with this schedule was in terms of pace, because three Cohort 1 classrooms comprised 3-year-old children, and as such they moved more slowly through the units, repeating activities more often. The weekly implementation dosage (number of activities completed) was typically the same for these teachers, but they did not get as far through the curriculum.

In Cohorts 2 and 3, the implementation schedule was flexible to meet the varying needs of different classrooms. This approach was designed with end-of-year feedback from teachers in Cohort 1. During each planning meeting in Cohort 2 and 3, coaches would walk through teachers’ goals and plans for the coming two weeks and choose activities that complemented those plans. This shift in implementation schedules was more consistent with the intended goals of GRS, and more aligned with the intentions of Practice-Based Coaching (National Center on Quality Teaching and Learning, 2015). In many cases, if teachers did not complete an activity in a given week, they were assigned it again for the following week. Research highlights the need to allow individualized variation to maximize student learning, particularly across multiple learning domains (Lieber et al., 2009). In the case of GRS, coaches support teachers to choose activities based on children’s existing ability levels and specific areas targeted for growth.

2.3. Measures

2.3.1. Intervention fidelity measures

Intervention fidelity measures were collected in all three Cohorts. GRS coaches met with teachers weekly and collected data at each contact point with teachers (meeting or observation). During biweekly planning meetings, coaches and teachers selected 4–5 activities for each of the 2 following weeks which included Morning Message, one Brain Game (self-regulation), one math activity, one literacy activity, and one additional activity. During the subsequent two weeks, teachers kept track of the activities they completed and how often they repeated them. In the following planning meeting, coaches recorded which of the assigned activities they completed and how many times they implemented each activity. After each planning meeting, coaches answered Likert-style questions about their perceptions of meeting efficacy and the general effectiveness of implementation. During biweekly classroom observations, coaches completed a detailed rating form for each activity observed, including Likert-scale ratings of adherence to lesson plans and child engagement. A subset of these data was used in subsequent analyses (see Table 2 for complete questions, answer types, and descriptive statistics). Based on teachers’ self-report of activities completed and coaches’ observations, we defined the following measures of dosage fidelity and adherence fidelity.

2.3.1.1. Dosage fidelity. Teachers reported the activities they had completed at least once during a period of two weeks in their biweekly coaching meeting (1 = completed activity and 0 = did not complete activity). The biweekly dosage scores were summed across the entire year to create a measure of dosage fidelity that describes the total number of activities teachers have completed at least once. We then summed dosage score by domain across the year to obtain a total dosage score for literacy, math, and self-regulation activities. This measure is similar to the one used by Dimitrovich et al. (2010) in the evaluation of Head Start REDI, where teacher reported weekly the activities implemented. Teachers were also strongly encouraged to repeat activities as described above. However, there was some inconsistency in measurement of activity repetition as, for example, some teachers responded with how many times they repeated the activity with each child in small groups, while others responded with how many times they repeated the activity in total, across children. Thus, the measure of dosage captured the number of different activities completed at least once but does not account for repetition.

2.3.2. Adherence to dosage fidelity

The number of activities teachers completed at least once was divided by the number of activities assigned to yield a completion percentage score. This was done separately for literacy, math, and self-regulation activities, as well as an aggregate of all activities. This variable thus measures not only dosage, but also teachers’ adherence to the recommended dosage, and as such may conflate dosage and adherence. However, this proportional approach reflects the fact that there was variation in the number of assigned activities across teachers, as some teachers repeated past lessons more frequently until their children were ready to move on. Because some teachers had better overall compliance than others, this percentage measure provides a more nuanced approach to capture this variability.

2.3.3. Adherence fidelity

Each classroom coach observed at least one GRS lesson every other week during the eight months of coaching. Each GRS activity has a description of materials required, vocabulary to use, core concept being target (i.e., rhyming, one to one correspondence, and working memory; see sample activity), and step-by-step lesson plan to facilitate implementation and assessment. For each activity observed, coaches filled out 4-item adherence form rating the degree to which (1) teachers’ used required materials or adequate substitutes, (2) emphasized the listed vocabulary on the lesson, (3) targeted core concepts from the GRS lesson plan, and (4) executed the lesson in its entirety as written in lesson plan on a Likert scale
from 1 (strongly disagree) to 5 (strongly agree). On average, 14 (SD = 6.64, range = 4–31) lessons were observed and rated per classroom across the year. We averaged the scores of all the observations to create a classroom mean score for adherence fidelity (Table 2). This procedure produced internally consistent composite scores (α = .77) with normal distributions (skewness < .30).

2.3.4. Child engagement

Although there is agreement that measures of child engagement should capture the extent to which children or students are attentive and respond positively to the program implemented, this construct has been measured very differently across studies. For example, Domitrovich et al. (2010) created a single item for coaches to rate during monthly classroom observations: “How many children were positively engaged and interested in the activity,” on a scale from 1 (very few) to 4 (nearly all). Others have created multiple-item scales, where observers rated the degree to which students were engaged, attentive, and expressed their opinions on a seven-point scale (7 being very engaged). Because of the lack of consistency between studies, we created a measure of child engagement for the present study to capture how students responded and focused on the activities implemented. During classroom observations every other week, coaches rated different aspects of child engagement, while teachers were implementing GRS activities. For each GRS activity observed, coaches filled out a four-item form rating the following items: children (1) seemed enthusiastic about starting the activity, (2) paid attention during the activity, (3) seemed to enjoy the activity, and (4) appeared to understand the activity. Each item was rated on a Likert scale from 1 (strongly disagree) to 5 (strongly agree). We averaged the scores of all the observations to create a classroom mean score for child engagement that yielded internally consistent composite scores (α = .88) with normal distributions (skewness < .30; Table 2).

2.3.5. Child outcomes

In Cohorts 2 and 3, child data were collected pre- and postintervention. Prior to the start of the intervention, participating children were pulled out of their regular classroom twice on different days by a bilingual research assistant for 20–30 min to participate in a brief assessment of early literacy, oral language, early mathematics, and self-regulation skills. This procedure was repeated at the end of the intervention period, approximately seven months after pretest (M = 220.36 days between testing sessions, SD = 24.17 days).

2.3.6. Language dominance

Children who spoke Spanish as reported by their parents were administered three subtests from the Preschool Language Assessment Scale (Duncan & De Avila, 1998; PreLAS, 2000) to determine language dominance: Simon Says/following directions, Art Show/answering questions about pictures, and Say What You Hear/repeat sentences. Items were administered to children in both English and Spanish, and the experimenter recorded the language in which the child’s score was higher. Once language dominance was determined in this manner, the dominant language was used for all subsequent assessments. Children who spoke Chinese as reported by their parents were administered the Expressive/Receptive One Word Picture Vocabulary Test (EOWPVT; Gardner, 1990). We administered the first 15 items of the test in English and Chinese, and the higher scoring language determined dominance. Once language dominance was determined in this manner, the dominant language was used for all subsequent assessments. Chinese-dominant children only received Head-toes–Knees-Shoulders and Toy Wrap (see below) tasks, as other assessments were standardized and have not been normed in Chinese.

Based on language dominance tests, at pretest, 75.9% children were dominant in English, 21.7% in Spanish, and 2.4% in Chinese. At posttest, 83.7% were English dominant, 15.3% Spanish dominant, and 1% Chinese dominant.

2.4. Emergent literacy skills

Woodcock-Johnson Test of Academic Achievement (WJ III)/Bateria III Woodcock Munoz: Letter–Word Identification and Picture Vocabulary Subscales (WJ III; Woodcock, McGrew, & Mather, 2001). WJIII is a widely used, standardized battery of academic skills, with strong validity and reliability in both English and Spanish. Letter–Word Identification involves identifying printed letters, and later reading printed words aloud. Picture Vocabulary assesses oral language and lexical knowledge. Internal consistency (Cronbach’s alpha) for these subtests ranges from .81 to .98. Test–retest reliability ranges from .89 to .92 (Bradley-Johnson, Morgan, & Nutkins, 2004). The subscale Picture Vocabulary was used in Cohort 2 to assess expressive language. Because of low variability, Expressive One-Word Picture Vocabulary Test, 4th Edition (EOWPVT-4; Brownell 2000) and the Spanish-Bilingual Edition Test Kit (EOWPVT-SBE; Brownell, 2001) were used instead in Cohort 3 to assess expressive language. Internal consistency for EOWPVT-4 ranges from .93 to .97 (Berry, Bridges, & Zaslow, 2004). WJ-III Picture Vocabulary and EOWPVT-4/EOWPVT-SBE produce standardized scores (M = 100 and SD = 15) based on national samples. In order to create a measures of expressive language that could be used across Cohorts 2 and 3, we created a z-score using WJ-III Picture Vocabulary standard scores for Cohort 2 and EOWPVT-4/EOWPVT-SBE standard scores for Cohort 3.

Clinical Evaluation of Language Fundamentals, Phonological Awareness Subtest (CELF-Preschool-2 and CELF-Preschool-2 Spanish; Semel, Wiig, & Secord, 2003). This measure has been normed in both English and Spanish, and tests rhyming, blending, segmenting, and identifying sounds and syllables in words and sentences. The English version of the CELF-2 Phonological Awareness subtest has a test–retest reliability of .82–.86 and a Cronbach’s alpha of .88. The Spanish version has a test–retest reliability of .81–.93 and internal consistency of .82–.88. We created a composite by summing scores on the syllabic blending, sentence segmentation, and syllabic segmentation tests that are used both in the English and Spanish versions.

2.5. Early math skills

WJIII/Bateria Woodcock-Munoz, Applied Problems (Woodcock et al., 2001). As above, this battery is available in both English and Spanish. Applied Problems measure the child’s ability to analyze and solve math problems. Initial items involve counting and identifying the number of objects in a picture, progressing to more complex calculations. Cronbach’s alpha ranges from .90 to .92 (Bradley-Johnson et al., 2004).

Test of Early Mathematics Ability (TEMA-3; Ginsburg & Baroody, 2003). We administered Form A of the TEMA-3 to assess mathematics performance. The test measures informal and formal concepts and skills in the following: counting skills, number-comparison facility, number literacy, number facts, calculation skills, and understanding of number concepts. The TEMA-3 has been normed for children between the ages of 3 and 8:11 and provides standard scores (M = 100, SD = 15). The Spanish version was provided to us by the developer of the TEMA-3. Internal consistency reliabilities have been reported all above .92. Performance on TEMA-3 is highly correlated with the Woodcock-Johnson III Tests of Achievement (Woodcock et al., 2001), TEMA-3 was only used in Cohort 2.

Research-based Elementary Math Assessment Short Form (REMA-SF; Weiland et al., 2012) was used in Cohort 3 to assess math ability instead of TEMA. The decision to change was driven by
the length of the TEMA. REMA-SF is the short form of the validated REMA (Clements, Sarama, & Liu, 2008). It consists of 19 items that measure comparing and ordering, verbal counting, arithmetic, number recognition and, composing number, shape identification, shape composition, and patterning. Internal consistency reliability ranges from .75 to .94 and is highly correlated with the Woodcock-Johnson III Tests of Achievement. The REMA-SF has been normed for preschool children between and provides standard scores ($M = 50$, $SD = 15$). The Spanish version was provided to us by the developer of the REMA-SF.

2.6. Self-regulation skills

**Head-Toes-Knees-Shoulders (HTKS; McClelland et al., 2014).** HTKS measures behavioral self-regulation integrating aspects of executive function such as inhibitory control, attention, working memory, and cognitive flexibility (McClelland & Cameron, 2012; Ponitz, McClelland, Matthews, & Morrison, 2009). HTKS has been used for children aged 4–8 years. Children first follow one of two commands naturally, and then are instructed to respond with a conflicting, nonautomatic action. For example, if the administrator says, “Touch your head;” the correct response would be to touch one’s toes. HTKS increases task complexity by adding two additional commands, including “touch your shoulders” and “touch your knees.” Commands get progressively more complex. On each item, children score 2 points for responding correctly, 1 point for self-correcting (initial movement to the incorrect response but ending with the correct response), and 0 points for responding incorrectly. In previous research, children who perform better on the task receive higher teacher and parent ratings of behavior, and fall prekindergarten scores predict achievement level and gains in the spring (McClelland et al., 2007; Ponitz et al., 2009). HTKS demonstrate construct and predictive validity in preschool through first grade, in both English and Spanish-speaking children (McClelland et al., 2014; Ponitz et al., 2008). It has an inter-rater reliability of .98 (Connor et al., 2010).

**Toy-wrap task.** This measure is part of the Preschool Self-Regulation Assessment (PSRA), a brief, structured battery of tasks designed to assess self-regulation in attentional, emotional, and behavioral domains among preschool children (Smith-Donald, Raver, Hayes, & Richardson, 2007). It is available in English and in Spanish and has been used in both the Chicago School Readiness Project (Smith-Donald et al., 2007) and in the Head Start REDI Program (Bierman, Domitrović et al., 2008; Bierman, Nix et al., 2008). In the Toy Wrap task, the child is instructed not to peek for a minute while the researcher noisily rotates a “surprise.” Scores indicate how many times the child peeks (zero, one, two, or more) and how much time happened before the child peeks.

**Preschool Self-Regulation Assessment (PSRA)—Assessor report.** After each testing session (2 in pretest and 2 in post-test), the 28-item PSRA Assessor Report was completed (Smith-Donald et al., 2007). The PSRA Assessor Report provides a global picture of children’s emotions, attention, and impulsivity throughout the assessor–child interaction. Items are coded using a Likert scale ranging from 0 to 3, with some items reverse coded to minimize automatic responding. Previous research yielded robust evidence for two factors: Attention/Impulse Control and Positive Emotion, demonstrating good internal consistency for both factors ($\alpha = .89$ and .87, respectively). These subscales have also showed moderate correlations with teacher-reported behavior problems and social competence in the BPI and SCBE30 (Smith-Donald et al., 2007). For the current study, we used the Attention/Impulse control factor as an additional measure of self-regulation. This factor includes 10 items measuring attention, concentration, impulse control patience, accuracy, distractibility, focus, planning ability, compliance, and ability to be seated during the evaluation. Because assessors filled out the PSRA at each testing session, we averaged PSRA scores for the two testing sessions to create an average Attention/Impulse control factor for pretest and one for post-test. The Attention/Impulse control factor showed excellent internal consistency ($\alpha = .92$).

2.6.1. Parent attendance in GRS events

Parents were offered 12 events (including an orientation, eight workshops, and three in-classroom events) throughout the year, where parents learned about GRS and practiced activities. On average, parents attended 2 events out of the 12 offered. We tracked attendance to these events to include as a covariate. Attendance to GRS events is used as a proxy of parent involvement in the intervention.

2.6.2. Classroom and teacher characteristics measures

In all three Cohorts, we collected structural classroom characteristics. Classroom characteristics include percentage of 3-year-old children who are not transitioning to kindergarten, teacher–student ratios, and whether or not the lead teacher in that classroom had prior experience with GRS. In Cohorts 2 and 3, instructional quality was also measured using the Classroom Assessment Scoring System (CLASS; Pianta et al., 2008) at the beginning of the year. Certified CLASS coders observed each classroom for four cycles of 20 min each. In each cycle, the quality of teacher–child interactions was assessed in 10 dimensions: Positive Climate, Negative Climate, Teacher Sensitivity, and Regard for Student Perspectives, Behavior Management, Instructional Learning Formats, and Productivity, Concept Development, Quality of Feedback, and Language Modeling. Each dimension is scored on a seven-point scale ranging from 1 (low) to 7 (high). A score for each dimension is obtained by averaging the 4 cycles. For the present paper, we were interested in obtaining an overall score of teaching quality. Thus, for each classroom, an aggregate score was created by averaging the ten average dimension scores. A similar procedure has been used in previous research (Hamre et al., 2010). The mean for total CLASS score was 4.78 ($SD = .78$, range = 3.46–6.10) and showed excellent internal consistency with $\alpha = .91$.

Teachers reported on their race and gender. In Cohorts 2 and 3, additional teacher characteristics were collected via surveys, completed by the head teachers and assistant teachers in the 14 classrooms at the end of the year. Teacher characteristics include level of training/education completed, length of time in the early childhood education field, and length of time at the current center (see Table 3). In all classrooms, both head and assistant teachers completed the survey. Because head teachers took the lead in implementing GRS, we only report head teacher characteristics.

2.7. Analysis plan

To examine the first research question, we used descriptive statistics for each IF variable to understand the degree to which activities are being implemented across classrooms and to what extent GRS is being implemented with fidelity. To examine the second research question, we used correlations and t-tests to understand which classroom and teacher characteristics, if any, were associated with IF. When association between teacher or classroom characteristics and IF were significant, we further examined the strength of these relationships while controlling for the effect of other teacher and classroom characteristics using partial correlations. Research question three, examining the association between fidelity measures and child outcomes, involved child and classroom-level data, using multilevel modeling with children nested by classroom. We assessed the extent to which intervention fidelity (level 2) was uniquely associated with post-test child math, literacy, and self-regulation outcomes, controlling for the
Table 3  
Classroom and head teacher characteristics.

<table>
<thead>
<tr>
<th>Question</th>
<th>N</th>
<th>%</th>
<th>M</th>
<th>SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classroom characteristics</td>
<td>19</td>
<td>-</td>
<td>-</td>
<td>2.29</td>
<td>12–20</td>
</tr>
<tr>
<td>Class size (num. students)</td>
<td>19</td>
<td>-</td>
<td>-</td>
<td>1.41</td>
<td>6–10</td>
</tr>
<tr>
<td>Student–teacher ratio</td>
<td>19</td>
<td>-</td>
<td>8.32</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Type of classroom</td>
<td>19</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>All 3 year olds</td>
<td>1</td>
<td>6%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>All 4 year olds</td>
<td>1</td>
<td>47%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Mixed ages (3 and 4 year olds)</td>
<td>9</td>
<td>47%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Percentage of 3 year old children by classroom</td>
<td>19</td>
<td>-</td>
<td>18.37</td>
<td>26.93</td>
<td>0–100</td>
</tr>
<tr>
<td>Total CLASS score +</td>
<td>14</td>
<td>4.78</td>
<td>.78</td>
<td>3.46–6.10</td>
<td></td>
</tr>
<tr>
<td>Head teacher characteristics</td>
<td>19</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Gender (female)</td>
<td>19</td>
<td>100%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Previous experience with GRS</td>
<td>19</td>
<td>7</td>
<td>36.80%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Yes</td>
<td>7</td>
<td>63.20%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>No</td>
<td>12</td>
<td>14.30%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Level of education completed +</td>
<td>14</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Bachelor’s</td>
<td>2</td>
<td>85.70%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Master’s</td>
<td>12</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Length of time in early childhood field (years) +</td>
<td>14</td>
<td>-</td>
<td>11.74</td>
<td>7.27</td>
<td>3–30</td>
</tr>
<tr>
<td>Length of time at the current center (years) +</td>
<td>14</td>
<td>-</td>
<td>6.42</td>
<td>8.07</td>
<td>.42–28</td>
</tr>
</tbody>
</table>

Note: + indicates that these questions were only administered to Cohort 2 and 3 teachers.

same child outcomes assessed at pretest (level 1); child characteristics including age, gender, English language learner status, and time between baseline and post-test assessment (level 1); number of GRS events attended by the parents (level 1); and CLASS score (level 2). Given the different types of fidelity measures and the small sample size, we ran separate models for each measure of fidelity. Analyses that included child-level data were conducted using SAS software, version 9.4 of the SAS System (SAS Institute Inc., 2014) using Proc MIXED.

3. Results

3.1. Descriptive statistics of dosage and adherence measures

Intervention fidelity measures were collected in Cohorts 1, 2, and 3. Collapsing across Cohorts, teachers completed an average of 94 activities (SD = 30.95) with high variation between classrooms, ranging from 27 to 136 activities completed. On average, coaches assigned 115 different activities (SD = 30.84). Because coaches assigned activities when teachers did not complete them as described previously, there was also high variability in the number of assigned activities, ranging from 47 to 160. Completion percentages, based on number of activities assigned, were relatively high across the board. On average, teachers completed 80% of assigned activities at least one time (SD = 12%, range = 57–97%). Additionally, teachers across Cohorts adhered very closely to the written curriculum, as measured by direct observation (M = 4.27 out of 5, SD = .37, range = 3.58–4.95). Child engagement (M = 4.34 out of 5, SD = .37, range = 3.82–5) during GRS activities was, on average, quite high.

Considering in particular whether the integrated nature of the intervention is feasible, on average, there were slight but not significant differences in completion across literacy (M = 85%, SD = 12%, range = 54–100%), math (M = 77%, SD = 16%, range = 33–92%) and self-regulation activities (M = 76%, SD = 19%, range = 43–100%), indicating that teachers implemented activities across domains in relatively equal measure. Despite that, on average, there were no significant differences in completion percentage across domains; we further examined the difference in percentage completion between the domain with highest completion percentage and the domain with the lowest completion percentage for each classroom. Overall, 36% of teachers had a difference higher than 20% (M = 37.62%, SD = 9.82%) between the domain with the highest and the lowest percent completion. In contrast, 64% of teachers had a percent difference of less than 20% between the high and low percent completion domain, with an average difference of 12.71% (SD = 4.76%). Therefore, despite the fact that GRS was implemented with equal emphasis across domains in the majority of classrooms, cross-domain differences were much larger in about one-third of the participating classrooms.

Similarly, there were slight but no significant differences in adherence to lesson plans across literacy (M = 3.40, SD = .62, range = 2.08–5), math (M = 4.25, SD = .34, range = 3.50–4.80), and self-regulation activities (M = 4.17, SD = .68, range = 2.33–5), indicating high adherence across the board. Because classrooms received different numbers of observations across the three domains, we report overall adherence in subsequent analyses.

3.2. Associations among measures of intervention fidelity

We examined association among measures of IF across Cohorts. As seen in Table 4, correlations between domains revealed a positive association between total number of math and literacy activities completed (r = .78, p < .01) and between math and literacy completion percentage (r = .64, p < .01), such that teachers who did more math activities were doing more literacy activities, and teachers that had a higher completion percentage on math activities also had a higher completion percentage on literacy. Total number of completed self-regulation activities was positively correlated with total number of completed math (r = .74, p < .01) and literacy activities (r = .78, p < .01). However, percentage of completion for self-regulation was uncorrelated with percentage completion for math and literacy. Finally, adherence fidelity showed a positive trend toward significance with child engagement (r = .41, p = .06) but was not associated with any measure of dosage or adherence to dosage.

3.3. Associations among classroom, teacher, and curriculum characteristics and intervention fidelity measures

We used Pearson correlations and t-test to examine the relationships between classroom and teacher characteristics and IF measures. Data from the three cohorts were used when available. As shown in Table 5, collapsing across cohorts, larger classroom sizes were loosely associated with fewer total number of activities completed (r = −.39, p = .09) and lower total percentage of completion (r = −.39, p = .10). Student–teacher ratio was also loosely
Table 4
Correlations among intervention fidelity measures (n = 19 classrooms).

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. General quantity</td>
<td></td>
<td></td>
<td></td>
<td>.74</td>
<td></td>
<td>.87</td>
<td>.66</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. % All completed</td>
<td>.95</td>
<td></td>
<td>.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Literacy quantity</td>
<td>.95</td>
<td></td>
<td>.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. % Literacy completed</td>
<td>.95</td>
<td></td>
<td>.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Math quantity</td>
<td>.95</td>
<td></td>
<td>.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. % Math completed</td>
<td>.95</td>
<td></td>
<td>.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Self-regulation quantity</td>
<td>.95</td>
<td></td>
<td>.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. % Self-regulation completed</td>
<td>.95</td>
<td></td>
<td>.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Adherence fidelity</td>
<td>.95</td>
<td></td>
<td>.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Child engagement</td>
<td>.95</td>
<td></td>
<td>.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5
Bivariate associations between classroom characteristics, teacher characteristics, and measures of intervention fidelity.

<table>
<thead>
<tr>
<th>N</th>
<th>Total number of activities</th>
<th>Total completion percentage</th>
<th>Adherence fidelity</th>
<th>Child engagement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Class size</td>
<td>19</td>
<td>-.39</td>
<td>-.39</td>
<td>-.01</td>
</tr>
<tr>
<td>2. Student–teacher ratio</td>
<td>19</td>
<td>-.39</td>
<td>-.39</td>
<td>-.01</td>
</tr>
<tr>
<td>3. Percentage of 3-year-old children</td>
<td>19</td>
<td>.04</td>
<td>.04</td>
<td>-.01</td>
</tr>
<tr>
<td>4. Total CLASS score</td>
<td>14</td>
<td>.47</td>
<td>.47</td>
<td>.50</td>
</tr>
<tr>
<td>5. GRS previous experience</td>
<td>19</td>
<td>t = 3.07</td>
<td>t = 3.07</td>
<td>.01</td>
</tr>
<tr>
<td>Yes</td>
<td>6</td>
<td>119 (13.72)</td>
<td>82.13 (13.11)</td>
<td>4.23 (1.39)</td>
</tr>
<tr>
<td>No</td>
<td>14</td>
<td>82.85 (28.03)</td>
<td>78.66 (10.42)</td>
<td>4.27 (1.37)</td>
</tr>
<tr>
<td>7. Years of classroom experience</td>
<td>14</td>
<td>.48</td>
<td>.48</td>
<td>.50</td>
</tr>
<tr>
<td>8. Years in center</td>
<td>14</td>
<td>.51</td>
<td>.51</td>
<td>.50</td>
</tr>
</tbody>
</table>

Note: Coefficients reported are person correlations and t-test (indicated with a t).

<table>
<thead>
<tr>
<th></th>
<th>p &lt; .10.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p &lt; .05.</td>
</tr>
<tr>
<td></td>
<td>p &lt; .01.</td>
</tr>
</tbody>
</table>

associated with lower percentage of math activities completed (r = -.41, p = .08). No other significant associations were found between classroom size or student–teacher ratio and other measures of IF. In addition, having a higher percentage of 3-year-old children in the classroom was associated with lower levels of adherence fidelity (r = -.53, p < .05). Correlations also revealed that classroom quality, as measured by CLASS observation at the beginning of implementation, was positively associated with observed adherence to GRS lesson plans (r = .59, p < .05) and loosely associated with higher total completion percentage (r = .47, p = .09), total number of activities (r = .47, p = .09), and number of self-regulation activities (r = .51, p = .06).

Teacher education level was not associated with any measure of dosage or adherence fidelity, perhaps because of the lack of variability (80% of head teachers held a Master’s degree). Years teaching in the center was positively associated with total completion percentage (r = .58, p < .05), and loosely associated with higher number of activities implemented (r = .51, p = .06), number of math activities completed (r = .50, p = .07), and literacy completion percentage (r = .52, p = .06). Years of experience in early childhood classrooms was not associated with adherence or child engagement, but showed a positive trend toward significance with total number of activities completed (r = .48, p = .09), as well as number of literacy (r = .48, p = .08) and self-regulation activities completed (r = .51, p = .06). In addition, teachers with prior GRS experience completed a significantly greater number of activities (M = 119) than teachers with no prior experience (M = 82.85; t = 3.03, p < .01). This is because teachers with prior GRS experience progressed more quickly through the curriculum and were therefore assigned significantly more activities. There were no associations between prior GRS experience and any measure of adherence or child engagement.

Finally, we explored what classroom and teacher characteristics were associated with a more balanced domain implementation. Teachers with prior GRS experience showed a smaller difference in percentage completion between the domain with highest completion percentage and the domain with the lowest completion (M = 13.67%) than teachers with no prior experience (M = 26.46%; t = 2.61, p < .05). Years teaching in the center showed a negative trend so that teachers that had been teaching in the center for more years showed smaller difference in percentage completion between domains (r = -.50, p = .07). No other significant associations were found.

Last, we used partial correlations when more than one teacher or classroom characteristics were significantly associated with any measure of IF. The association between percent of 3-year-old children in the classroom and adherence became nonsignificant when controlling for CLASS scores. The association between CLASS scores and adherence became a trend (r = .47, p = .10). These results may be due to the small sample size (n = 11).

3.4. Fidelity effects on child outcomes at post-test

All child outcome analyses included participating classrooms and children from Cohorts 2 and 3 (as child outcome data were not collected in Cohort 1). Descriptive data for child outcomes at pre and post-test are presented in Table 6. To examine associations between fidelity measures and child outcomes at post-test, all variables were standardized to have a mean of zero and standard deviation of 1 to ease interpretability. Table 7 presents standard-
ized coefficients ($b$) and standardized errors (SE) that identify the direction and magnitudes of the association between each separate fidelity measure (level 2) and child literacy, math, and self-regulation outcomes controlling for pretest child outcomes (matching post-test child outcome), time spent between baseline and post-test assessment, child characteristics, and parent’s attendance to GRS events (level 1), as well as CLASS scores (level 2). Pretest child outcomes (level-1 predictor), in the case of all measures, were strongly and significantly associated with post-test child outcomes (b’s ranging from .66 to .35).

When we examined effects of IF measures in separate models, dosage (number of activities completed), adherence to dosage (percentage of activities completed), adherence, and child engagement showed significant associations with child outcomes (see Table 7). We first examined the total number of activities completed and completion percentage of each component of the curriculum (math, literacy, and self-regulation) on all child outcomes. There was a positive main effect of the total number of literacy activities completed on expressive vocabulary ($b = .10$), as well as a positive trend toward significance on assessor ratings of attention and impulse control ($b = .13, p = .07$). Literacy completion percentage was loosely associated with increased gains in basic math ($b = .10, p = .09$). For the math activities, total number of math activities was not associated with any child outcome, but math completion percentage showed a positive main effect on basic math ($b = .18$), and a positive trend toward significance on expressive vocabulary ($b = .08, p = .06$). Total number of self-regulation activities and self-regulation completion percentage were positively associated with attention and impulse control ($b = .18$ and .15 respectively). Contrary to our hypothesis, total number of self-regulation activities and self-regulation completion percentage were associated with smaller gains on phonological awareness ($b = −20$ and $−19$, respectively), even when controlling for percentage of literacy activities. Self-regulation completion percentage was associated with smaller gains on math ability in Cohort 2 ($b = −.42$). When controlling for the effects of math completion percentage on math ability in Cohort 2, the effects of self-regulation completion percentage became nonsignificant. Finally, total number of activities completed was positively associated with growth on expressive vocabulary ($b = .10$), and showed a positive trend toward significance on attention and impulse control ($b = .13, p = .07$). Total percentage completion was not associated with any child outcome.

Next, we examined the effects of adherence fidelity on change in child outcomes. Controlling for level-1 and level-2 variables, there was a significant positive main effect of adherence fidelity on post-test phonological awareness ($b = .20$), basic math ($b = .15$), and cognitive flexibility ($b = .23$), and a positive trend toward significance on math ability as measured by the TEMA-3 in Cohort 2 ($b = .16, p = .08$). Finally, child engagement showed a positive significant association with growth in cognitive flexibility ($b = .15$).

### 4. Discussion

The purpose of the current study was threefold. First, we used intervention fidelity data to examine the feasibility of a skill-specific, integrated school readiness intervention implemented in Head Start classrooms. Second, we examined the associations between intervention fidelity and teacher/classroom characteristics in order to better understand what contributes to high IF. Third, we examined the association between intervention fidelity and change in child outcomes across the literacy, math, and self-regulation domains.

This study suggests that on average, Getting Ready for School was implemented with moderate-to-high levels of dosage and adherence fidelity, yet variability was observed between classrooms. Intervention fidelity—both in terms of dosage and adherence—was associated with gains in children’s literacy, math, and self-regulation skills. Children’s engagement was also positively associated with growth in self-regulation skills. These results are promising as they suggest that an integrated curriculum that specifically targets literacy, math, and self-regulation can be feasible for Head Start teachers to implement and that high fidelity can lead to growth in school readiness skills.

#### 4.1. Variability in teachers’ implementation of GRS: feasibility of an integrated supplemental curriculum

Recent evidence suggests that implementing multiple content-focus curricula could potentially generate more gains in child school readiness (Duncan et al., 2015). However, it is unclear how feasible this approach would be given the burden that doing so could represent for teachers. The GRS curriculum was designed as a single supplemental curriculum that supports skills across three key domains for school readiness—math, literacy, and self-regulation—through instructional activities that can be implemented daily as a part of normal classroom routines with the support of a coach.

The results show that the implementation of a supplemental school readiness curriculum that tackles multiple educational domains is feasible and can be implemented with fidelity. In the present study, teachers implemented the curriculum with moderately high fidelity across the board, yet there was high variability in adherence to dosage. Teachers did not quite reach the target dosage of implementation of different activities, demonstrated by an average completion percentage of 80%. These results are consistent with previous research examining intervention fidelity in early childhood programs, where typically dosage varies between 55% and 80% (Hamre et al., 2010; Odom et al., 2010). It is also possible that an alternate measure of dosage that captures repetition of activities, as well as the total time spent on GRS activities daily, may have provided a better estimate of the amount of time teachers...
Table 7
Multilevel models implementation variables and child outcomes.

<table>
<thead>
<tr>
<th></th>
<th>Literacy</th>
<th>Math</th>
<th>Self-regulation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-literacy (WJ-LWID) (n = 182)</td>
<td>Expressive vocabulary (WJ-PV/EO WPVT) (n = 159)</td>
<td>Phonological awareness (CELF) (n = 181)</td>
</tr>
<tr>
<td>Total quantity</td>
<td>.01 (.05)</td>
<td>.10 (.05)</td>
<td>−.16 (.09)</td>
</tr>
<tr>
<td>% All completed</td>
<td>.02 (.05)</td>
<td>.08 (.05)</td>
<td>−.05 (.09)</td>
</tr>
<tr>
<td>Literacy quantity</td>
<td>.01 (.05)</td>
<td>.10 (.05)</td>
<td>−.13 (.09)</td>
</tr>
<tr>
<td>% Literacy completed</td>
<td>.01 (.06)</td>
<td>.07 (.05)</td>
<td>−.05 (.09)</td>
</tr>
<tr>
<td>Math quantity</td>
<td>.04 (.05)</td>
<td>.08 (.05)</td>
<td>−.15 (.09)</td>
</tr>
<tr>
<td>% Math completed</td>
<td>.04 (.05)</td>
<td>.08 (.05)</td>
<td>−.06 (.09)</td>
</tr>
<tr>
<td>Self-regulation</td>
<td>.02 (.06)</td>
<td>.09 (.05)</td>
<td>−.20 (.09)</td>
</tr>
<tr>
<td>% Self-regulation</td>
<td>.01 (.05)</td>
<td>.04 (.06)</td>
<td>−.19 (.08)</td>
</tr>
<tr>
<td>Adherence fidelity</td>
<td>.01 (.06)</td>
<td>.09 (.05)</td>
<td>−.20 (.10)</td>
</tr>
<tr>
<td>Child engagement</td>
<td>−.01 (.05)</td>
<td>−.01 (.05)</td>
<td>.06 (.08)</td>
</tr>
</tbody>
</table>

Note: Entries show parameter estimates with standard errors in parentheses. All models control for baseline child outcome, time between baseline and posttest assessment, parents’ attendance to GRS events, CLASS scores, and nesting effects. To preserve power, models control for child age, gender, and language of testing (English vs. Spanish) when correlations were found significant in bivariate analysis. Estimation method = ML; Kenward–Roger degrees of freedom.

* p < .10
** p < .05
*** p < .01
**** p < .001.
were actually spending on GRS in a given week. The GRS activities are meant to be repeated twice per week, but repetition is not reflected in the dosage measure presented here. Future work should more precisely measure both the total number of times an activity was implemented, as well as the amount of time spent on GRS on a weekly basis.

Additionally, most teachers implemented activities across all three domains relatively consistently, yet there was some variability across classrooms in regard to which domain had higher percentage completion. These results suggest that despite the feasibility of an integrated supplemental curriculum such as GRS, some teachers may have preference for one or two domains over the others. In general, teachers who had more experience and who had previous experience with GRS showed a more balanced implementation, demonstrated by smaller differences in percentage completion between domains. However, other variables not examined may explain these differences. Teachers may customize their practice based on the needs of their classrooms. For example, in classrooms where more children show poor social–emotional skills, teachers may be inclined to implement more self-regulation activities in detriment of literacy and math. Other teacher characteristics, such as confidence in their ability to teach certain concepts, may explain these differences. For example, teachers with math anxiety may tend to do less math activities compared with literacy and self-regulation (Chen, McCray, Adams, & Leow, 2014; Geist, 2015). Future research should explore differences between classrooms where teachers implement the three domains with equal emphases versus classroom where teachers emphasize one domain over others and its implications for child growth.

Finally, teachers implemented GRS activities with high adherence to the written curriculum. Based on observational ratings, teachers used required materials, used the listed vocabulary, targeted core concepts from the GRS lesson plan, and executed the lesson in its entirety as written in lesson plan most of the time. This suggests that curriculum implementation is straightforward. However, because classroom coaches conducted these observations, it is possible that these ratings were upwardly biased. Future work should ensure that other individuals periodically perform observations to get even more accurate estimates of adherence fidelity.

### 4.2. Associations among classroom characteristics, teacher characteristics, curriculum characteristics, and intervention fidelity

A range of classroom and teacher characteristics were associated with IF measures. Interestingly, classroom size, student–teacher ratio, years working in the center, and previous experience with GRS were associated with dosage, while classroom quality and child’s age distribution were associated with adherence fidelity. These different patterns of association have important implications for scalability and can inform future research studying GRS and other school readiness programs.

Perhaps unsurprisingly, classroom size showed trends toward significance, such that classrooms with more children implemented fewer activities and had lower completion percentages. This is in line with other work that shows that classroom quality is generally associated with smaller class sizes and indicates the importance of complying with recommendations for structural quality (Barnett et al., 2004). Moreover, classrooms with larger student–teacher ratios showed lower percentages of math completion. It may be that math activities in particular lend themselves to work in smaller groups and/or require more individualized support, and therefore teachers who have more students have difficulty finding the time. These results have important implications for future rounds of implementation. GRS allows a flexible implementation, as many activities can be adapted to small or large groups. However, in classrooms with larger student–teacher ratio, teachers will need more support to adapt activity formats to their classrooms’ needs, either by implementing activities in larger groups or making sure that one teacher can monitor and oversee the classroom while the other implements small group activities.

In addition, years of experience in the center but not years of teaching experience in early childhood was positively associated with dosage: teachers that had been teaching in the center for more years implemented more activities overall and showed higher levels of percentage completion. It may be that teachers with more years of experience in the center are more familiarized with the school program and lesson plans, schedule, and classroom organization, and therefore it is easier for them to integrate more supplemental activities in their daily routines and lesson plans. Finally, teachers who were exposed to the GRS curriculum prior to starting the intervention completed more activities than teachers who were new to GRS. These results suggest that although the activities are relatively easy for teachers to implement regardless of their previous experience in GRS, it may take more time and experience to reach the desired dosage. Taken together, these findings have important implications for school readiness interventions as they suggest that teachers may need time to familiarize themselves with new interventions to optimize implementation and subsequent outcomes for children. Future research with larger sample sizes must explore the impact of teachers’ previous experience in GRS on growth in child outcomes and examine the role of dosage mediating this relationship.

It is worth noting that teacher characteristics such as years of education, years of classroom experience, years in the center, and classroom size and student–teacher ratio were not associated with adherence fidelity to GRS activities. This is promising, as it suggests that teachers from a wide variety of backgrounds and level of expertise can implement supplemental curricula such as GRS with similar adherence fidelity. Instead, other factors were associated with adherence fidelity. Teachers from classrooms with better quality as measured by the CLASS at the beginning of the year were more likely to adhere faithfully to GRS lesson plans, while classroom quality was not associated with dosage. These results suggest that GRS is feasible regardless of quality, yet classrooms with lower quality may require more intense professional development and extra support from coaches to implement the curriculum with high fidelity. Additionally, as the proportion of 3-year-old children in the classroom increased, adherence fidelity to GRS decreased. Because three and four year olds can be at different developmental stages, teachers in classrooms with more 3-year-old children may have to adapt the activities to meet the child’s developmental level, resulting in lower rates of adherence. Recent research suggests that in preschools serving low-income children, increased proportion of 3-year-old children in mixed-age classrooms may have a negative effect on the school readiness of 4-year-old children (Ansari, Purtell, & Gershoff, 2015; Bell, Greenfield, & Bultotsky-Shearer, 2013). Our results call for further research examining whether differences in instruction could be the mechanism by which rising kindergarteners in classrooms with a high proportion of 3-year-old children develop school readiness skills at a slower pace. In other words, we need to better understand if teachers adapt their instruction and adhere less faithfully to a curriculum to meet the needs of younger children, while not providing enough challenging opportunities for older children.

Other factors that were not measured here may also be crucial in understanding variability in implementation. For example, teacher stress may interfere with their capacity to implement interventions to the desired dosage and adherence (Wehby, Maggin, Partin, & Robertson, 2012). In addition, there is growing agreement that the use of coaching models with individualized support is more effective in improving teaching quality than one-time trainings (Powell,
Diamond, Burchinal, & Koehler, 2010; Wayne, Yoon, Zhu, Cronen, & Garet, 2008). Less is known about the impact of coaching quality on different dimensions of program implementation. The GRS intervention offers an intense coaching model with weekly meetings to support implementation. While the coach–teacher interaction could have affected intervention fidelity, it is beyond the scope of this study to examine these possible associations. Future research should explore how to define quality of coaching and what the implications may be for fidelity to school readiness interventions such as GRS.

4.3. Associations between intervention fidelity and child outcomes

There has been a growing interest in connecting intervention fidelity and child outcomes in early childhood research (Clements et al., 2013; Domitrovich et al., 2010; Guo et al., 2016; Knoche et al., 2010). Despite research that shows a generally positive relationship between intervention fidelity and child outcomes, these relationships varied in the present study depending on the fidelity measure used (Clements & Sarama, 2008; Hamre et al., 2010; Domitrovich et al., 2008; Justice, Mashburn, Hamre, & Pianta, 2008). The present study provides evidence that both dosage of math and literacy activities, and adherence to the GRS curriculum are associated with child gains on pre-literacy and math skills, while dosage of self-regulation activities showed both positive and negative effects on gains in child outcomes. While most fidelity research has defined dosage as number of activities and/or time spent doing activities, in the present study, we took a different approach according to the GRS coaching model. We defined two measures of dosage: the number of activities completed and the percentage of activities completed based on activities assigned by the coach. While the measure of number of activities completed is a pure measure of dosage, percentage of completed activities captures compliance and adherence to the coaching plan and accounts for variation across teachers. Overall, domain-specific percentages of activities completed were related to growth in children’s pre-academic skills, but differences emerged based on the outcome measured. Associations were less robust for number of activities completed. Adherence was associated with child outcomes across domains.

For literacy outcomes, children in classrooms where teachers provided higher number of literacy activities showed greater growth in expressive vocabulary. Interestingly, higher percentages of math activities were loosely associated with more growth in expressive vocabulary. These results are consistent with previous research suggesting that math activities may also have positive effects on children’s language and pre-literacy skills (Clements et al., 2004; Clements & Sarama, 2008). It could also suggest that GRS activities, specifically math activities, are integrative in nature and promote literacy development while focusing on math concepts.

Adherence fidelity was associated with positive gains in phonological awareness suggesting that adherence to the lesson plans is critical to support skills such as rhyming or identifying sounds and syllables in words. Surprisingly, higher self-regulation activity completion was associated with less growth in phonological awareness even when controlling for literacy dosage and adherence to dosage. The scope of this paper does not allow us to fully understand these unexpected results. However, while expressive vocabulary could be more easily supported through both literacy and math activities as our results suggest, phonological awareness skills may require more targeted activities that specifically target skills such as rhyming, identifying sounds and syllables in words, or blending and segmentation. One could argue that these targeted activities may require more effort and time, and thus, teachers that prioritize self-regulation skills may dismiss key literacy activities that support phonological awareness, which may account for the negative effects of self-regulation dosage and adherence to dosage on phonological awareness. Alternatively, it may be that in classrooms where more children have challenging behaviors, teachers completed more of the assigned self-regulation activities. Therefore, it could be that behavior problems such inattention hinder gains in other academic skills (Lonigan et al., 1999). Future research should examine the interaction among children’s behavior problems, dosage of self-regulation activities and its effects on academic gains.

For math outcomes, children in classrooms where teachers completed a higher percentage of assigned math activities and exhibited more adherence fidelity showed greater growth in applied problems over the year. Interestingly, the total number of math activities completed was not associated with growth. These results could suggest that adherence to the dosage of activities may be more important than the number of math activities that they complete. Therefore, it may be important to assign teachers an amount of activities they can handle based on their level of experience and children’s needs. Although in order to examine the efficacy of early childhood educational interventions we must define a recommended dosage, we must consider some level of individualization to meet teachers’ and children’s needs and assure successful implementation. Alternatively, the lack of association between total number of math activities and child skill growth may relate to the fact that we were not able to successfully account for repetition of activities in our measure of dosage. Higher percentages of literacy activities were loosely associated with more growth on applied problems highlighting again the integrative nature of GRS activities and the possible spillover effects. Finally, higher self-regulation completion was associated with smaller gains in math ability in Cohort 2. Nonetheless, this association became nonsignificant when controlling for math completion percentage. In addition, the negative association between self-regulation completion and math was neither replicated in Cohort 3 with a different measure of math ability nor with a measure of basic math across Cohorts.

For self-regulation outcomes, children in classrooms where teachers provided higher number of self-regulation activities and completed a higher percentage of assigned activities showed greater growth in attention and impulse control as measured by the PSRA assessor report. However, these positive associations were not observed with the Head-Toes-Shoulders-Knees task, a measure of behavioral self-regulation that integrates inhibitory control, working memory, and cognitive flexibility. Conversely, adherence fidelity and child engagement were positively associated with growth in HTKS. These different patterns of association among dosage, adherence fidelity, and child engagement and gains in self-regulation require further discussion. The lack of a positive association between dosage and HTKS could be due to our dosage measure which does not capture repetition. This may be crucial for self-regulation activities, because a substantial part of the GRS self-regulation curriculum is on games and activities that need to be practiced on a daily basis.

Finally, total number of activities was only associated to gains in expressive vocabulary suggesting that the impact of dosage and adherence to dosage on most skills may be domain specific. Therefore, it may be crucial that coaches support a balanced implementation of activities across the three domains.

Overall, it is important to note that adherence fidelity to GRS was associated with growth in children’s pre-academic skills and self-regulation over and above teaching quality, as measured by the CLASS. This finding is similar to other works showing positive effects of skill-focused curricula when accounting for classroom quality (Domitrovich et al., 2010) and contradicts the notion that instruction-based, content-specific curricula are unnecessary when high-quality teaching is in place (Elkind, 2001). This is critical, as it suggests that fidelity to GRS has the potential to support child’s...
school readiness in literacy, math, and self-regulation over and above of the quality of teaching per se, which has important implications for scalability and generalizability of the intervention.

Finally, the present study is one of the few to examine the association between child engagement and growth in child outcomes (Domitrovich et al., 2010). Child engagement with GRS activities was positively associated with growth in self-regulation as measured by HTKS. These results add to previous research showing a positive relation between child engagement and social–emotional outcomes, and highlight the importance of taking into consideration students’ responses. Despite child engagement being high across observations, future research should explore which specific activities (by domains) and formats (whole group vs. small group) are most engaging.

4.4. Limitations

The findings of this study must be considered in light of several limitations. First, although we focused on the most commonly used constructs of IF, dosage and adherence, these measures were specific to the intervention presented in this study and had no prior validation. While this is a limitation of most IF studies, it makes comparing results across interventions difficult (Mendive et al., 2016; Halle, Metz, & Martinez-Beck, 2013). Second, measures of dosage were self-reported, as teachers were asked to report the activities they completed every two weeks. Although this is a common limitation of IF studies (Domitrovich et al., 2008; Mendive et al., 2016), future research should explore methods to assess dosage more accurately, such as using daily logs to report activities completed. Additionally, studies have shown that associations between dosage and child outcomes are stronger when measures are more specific (e.g., number of minutes of implementation rather than number or proportion of activities implemented (Hamre et al., 2010; Strasser & Lissi, 2009)). Our measures of dosage only took into account number and percentage of activities completed once and did not capture time spent doing the activity or repetition. Because GRS activities could be implemented in whole groups or small groups, it was challenging for teachers to keep track of how many times each child completed an activity and for how long. Data on these components of dosage would have provided a more comprehensive picture of IF and potentially brought more variability within our sample.

Third, our measure of adherence fidelity was not domain specific. Although adherence fidelity was associated with growth in basic math and language skills, we cannot explain the mechanisms through which adherence supports child growth in each specific domain. Additionally, we did not consider variation in adherence between observations nor differences in dosage over the implementation period. Previous research examining other school readiness programs has suggested changes over time in intervention fidelity for adherence but not dosage (Domitrovich et al., 2010). Future research should explore patterns of implementation to understand variation over time and its impact on child outcomes. In addition, adherence fidelity was based on observations made by the classroom coach, who is not an independent observer. One of the benefits of having coaches assess adherence is that we were able to collect multiple observations across the year, more than what is usually reported in IF studies (Domitrovich et al., 2010). However, we were not able to have multiple independent observers assessing adherence for each classroom preventing us from examining inter-rater reliability. In addition, we did not assess the reliability of GRS coaches’ adherence ratings. To address these pitfalls, coaches received training on program implementation and evaluation, and through the year, coaches attended biweekly supervision, where they discussed any issue in the implementation and evaluation of GRS. Furthermore, the curriculum describes activity implementation step-by-step to facilitate adherence fidelity evaluation; however, some variability between coaches can be assumed.

Fourth, despite controlling for a range of classroom- and individual-level variables in our models, it is likely that other variables not included here are relevant to the analyses presented. For example, the relationships that coaches establish with teachers may also have an effect on dosage and adherence. This is a limitation of any intervention that provides a coaching component. Future studies with larger samples will allow us to add another nesting level in our analyses and control for coach effects. Moreover, while we did control for parent attendance to GRS workshops, the current study did not examine how parents use GRS at home to support their child’s development, and that may explain some variability in child growth. Because GRS is an integrated program that targets both teachers and parents as key actors in supporting child development, future research will include additional data on parental involvement at home. Additionally, although we examined the association between a wide range of teacher and classroom characteristics and IF, we did not examine some factors that have been linked to IF in other work. In a review of the literature, Durlak (2010) highlighted that organizational capacity, program resources and funding, and education policies are associated with program implementation and thus could potentially hinder or support IF (Durlak & DuPre, 2008).

Finally, the small sample size in the current study may have prevented us from detecting unique associations among dosage, adherence, child engagement and child growth. Therefore, the associations found in the present study between IF and child outcomes are exploratory in nature and cannot be generalized to other samples. In addition, due to sample size limitations, we could not examine the associations of dosage, adherence, and child engagement with child outcomes in a unique model. Future research should explore whether different implementation measures are equally important for child growth. Further, it is important to note that these findings do not represent a randomized experiment, but rather examine differences within teachers receiving the intervention. The lack of statistical power also prevented us from examining whether or not child baseline characteristics and performance interacted with IF predicting child outcomes (Hamre et al., 2010). Larger scale studies are warranted in order to establish a robust relationship between IF and child growth and explore possible interactions. This investigation only explored the link between intervention fidelity and child growth without a comparison group. A research study is underway examining the efficacy of GRS in comparison with a control group and will help expand current findings and examine the moderator role of intervention fidelity.

5. Conclusion

In conclusion, this study contributes to the field of early childhood curriculum research by demonstrating that the implementation of an integrated supplemental curriculum that includes content-rich math, literacy and self-regulation activities is feasible in Head Start classrooms and that varying levels of intervention fidelity are associated with child gains across domains. Study findings support current efforts to measure multiple indicators of intervention fidelity as they relate differently with child growth. This study also expands previous research by highlighting how teacher and classroom characteristics may be differentially related to various measures of IF, thus highlighting some areas where classrooms may need more support to implement school readiness interventions as intended. The study findings suggest that teachers may need time to familiarize themselves with new interventions to optimize implementation and reach the desired dosage. Therefore, before examining the effectiveness of any school readiness inter-
vention or program we may need to consider giving teachers at least one year to get familiarized with new interventions so they feel confident in their ability to implement them. In addition, classrooms with lower instructional quality may need additional support, such as more intense professional development, to implement supplemental programs with high adherence fidelity. Therefore, policies that require the evaluation and monitoring of classroom quality can help us identify what teachers may need additional support to adhere to classroom curricula.

Nonetheless, programs such as GRS that offer activities with clear, targeted instruction have the potential, when implemented faithfully, to improve children’s school readiness skills across domains, above and beyond the general quality of classroom instruction. This type of research-based, integrated curriculum has the potential to be a successful policy lever for promoting school readiness at-risk populations, by emphasizing the inclusion of explicit, content-focused activities within the broader frameworks of widely used curricula. As pre-kindergarten programs grow more popular with cities and states adopting universal policies, it is critical to keep in mind that teachers’ implementation of developmentally appropriate, targeted activities has a greater chance of improving the literacy, math, and self-regulation skills that children need to succeed.

Funding

This work was supported by the Institute of Education Sciences under the FY 2012 Early Learning Programs and Policies research program [award#R305A120783, 2012].

Acknowledgments

The authors gratefully acknowledge the help of Saskia Op den Bosch, Maria Gabriele, Kelsey Repka, and administrators, teachers, families, and children who participated in the study. The authors thank the Open Society Foundation and of the Farber Foundation for their support in the initial development and implementation of the Getting Ready for School initiative.

References


