



HHS Public Access

Author manuscript

Prev Sci. Author manuscript; available in PMC 2019 May 23.

Published in final edited form as:

Prev Sci. 2018 February ; 19(2): 159–173. doi:10.1007/s11121-017-0820-2.

Promoting Afterschool Quality and Positive Youth Development: Cluster Randomized Trial of the Pax Good Behavior Game

Emilie Phillips Smith,

The University of Georgia

D. Wayne Osgood,

The Pennsylvania State University

Yoonkyung Oh, and

The Pennsylvania State University

Linda C. Caldwell

The Pennsylvania State University

Abstract

This randomized trial tested a strategy originally developed for school settings, the Pax Good Behavior Game (PAX GBG), in the new context of afterschool programs. We examined this approach in afterschool since seventy percent (70%) of all juvenile crime occurs between the hours of 3–6pm, making afterschool an important setting for prevention and promotion. Dual-career and working families need monitoring and supervision for their children in quality settings that are safe and appropriately structured. While substantial work has identified features of quality afterschool programs, increasing attention is being given to how to foster quality. PAX GBG, with its focus on shared norms, cooperative teams, contingent activity rewards, and liberal praise, could potentially enhance not only appropriate structure and supportive relationships, but also youth self-regulation, co-regulation, and socio-emotional development. This study examined the PAX GBG among 76 afterschool programs, serving 811 youth ages 5–12, who were diverse in race-ethnicity, socio-economic status, and geographic locale. Demographically matched pairs of afterschool programs were randomized to PAX GBG or treatment-as-usual. Independent observers conducted ratings of implementation fidelity and program quality across time; along with surveys of children's problem and prosocial behavior. Interaction effects were found using hierarchical linear models such that experimental programs evidencing higher implementation fidelity demonstrated better program quality than controls, (i.e., less harshness, increased appropriate structure, support, and engagement), as well as reduced child-reported hyperactivity and intent-to-treat effects on prosocial behavior. This study demonstrates that best practices fostered by PAX GBG and

The corresponding author, Emilie Smith is the Janette McGarity Barber Distinguished Professor and Department Head of Human Development and Family Science; she can be reached at emilie.smith@uga.edu.

Compliance with Ethical Standards

The authors declare that they have no conflict of interest. Research involving Human Participants was approved and monitored by The Pennsylvania State University Institutional Review Board (IRB # 23990).

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. A process for obtaining informed consent from all individual participants was included in the study.

implemented with fidelity in afterschool result in higher quality contexts for positive youth development.

Keywords

afterschool quality; child socio-emotional outcomes; co-regulation; implementation fidelity; PAX GBG; positive youth development; PAX GBG; randomized trial; self-regulation; setting-level effects

Introduction

Over the past 2 and a half decades, afterschool settings have emerged as important contexts for the prevention of youth problem behaviors and, the promotion of positive youth development (Catalano et al., 2002; Heath & McLaughlin, 1994; Pittman, 1991). Owing to the fact that afterschool is a part of the lives of over 10.2 million children and families in the United States (America After 3pm, 2014), this cluster randomized trial examines the impact of an approach originally developed for school settings, the PAXIS Institute's version of the Good Behavior Game (PAX GBG), on afterschool program quality and youth socio-emotional outcomes (Domitrovich et al., 2010; Embry, Richardson, Schaffer et al. 2010 Kellam et al., 2008). Of particular interest is the degree to which implementation fidelity is an important moderator of effects on program quality and, youth problem and prosocial behavior. In the following sections, we present a rationale for prevention research in afterschool, and the promise of PAX GBG in fostering both program quality and youth socio-emotional development.

Out-of-school time is important in that 70 percent of all juvenile crime in the U.S. occurs between the hours of 3 to 6.00 pm (Snyder & Sickmund, 2006). According to opportunity theory, when youth are unmonitored, lacking adult supervision, this provides an opportunity for unstructured socializing with peers that is associated with increased levels of juvenile delinquency (Osgood & Anderson, 2004). Adult monitoring and supervision for youth during the out-of-school time is an important work-family issue, given that nearly 70% of married couples and up to 85% of single parents work outside of the home and have children between the ages of 6 and 17 years old (Gottfredson, Gerstenblith, Soulé, Womer, & Lu, 2004). One in four families have a child enrolled in afterschool but 11.3 million children, (1 in 5) still return home alone and unsupervised (America after 3pm, 2014). Racial-ethnic minority children are even more likely to have working parent(s) and be in need of quality, affordable afterschool care (Hynes & Sanders, 2011). Providing monitoring and supervision afterschool is critical to supporting working families, as well as to reducing juvenile problem behavior, delinquency and substance use.

The growth of afterschool programs was fueled by the federal 21st Century Community Learning Centers Program (21st CCLC), legislation designed to provide a safe and supervised place coupled with academic and social enrichment for young people (Mahoney & Zigler, 2006; Smith, Boutte, Zigler, & Finn-Stevenson, 2004). In research on afterschool settings, Mahoney and colleagues have found that participants in urban afterschool programs demonstrated higher reading grades and performance on literacy assessments (Mahoney,

Lord, & Carryl, 2005). For urban middle school students, afterschool participation was related to reduced delinquency in programs that emphasized evidence-based practices and social skill development (Gottfredson, et al., 2004). In studies with ethnic minority samples, it has been found that culturally-oriented afterschool programs demonstrated not only increased socio-emotional skills, but also an enhanced sense of ethnic identity, self-worth, and reduced aggression and drug use (Belgrave et al., 2004; Riggs et al., 2010; Tebes et al., 2007). However, the research on afterschool does not all find positive results; some studies have reported finding little or no benefit of afterschool programming (James-Burdumy, Dynarski-Moore, Deke, Mansfield & Pistorino, 2005; Gottfredson, Cross, Wilson, Rorie, & Connell, 2010; Mahoney Stattin & Lord, 2004). In the national evaluation of 21C Centers, wide variations in levels of participation, attendance, and quality of programming likely resulted in these less positive findings (James-Burdumy et al., 2005). The quality of programming matters in that international research on recreational centers has shown that insufficient monitoring and supervision has iatrogenic effects on youth by attracting deviant peers who engage in problem behaviors (Mahoney, Stattin & Lord, 2004). On the other hand, in meta-analytic studies that have accounted for variations in program design and content, out-of-school time programming has been found to have modest benefits to youth reading and math skills (Lauer et al., 2006). Further, afterschool programs that were identified as S.A.F.E., that is sequenced (appropriately structured), active, focused (on skill development) and explicit (goal-oriented) fostered significantly improved academic achievement, socio-emotional development, and reduced problem behavior (Durlak, Weissberg, & Pachan, 2010). Thus, the quality and content of afterschool programming factored prominently in the degree to which benefits were found for youth.

Aspects of Quality Afterschool Program Settings

Quality in afterschool contexts can be understood using the lens of setting theory that posits that social processes, such as the nature of adult and peer relationships, shared power, and engagement in decision making, are important in youth social and behavioral adjustment (Barker, 1968; Fairweather, 1972; Larson, 2000; Tseng and Seidman, 2007). Using a setting-level perspective, changing these social processes at the program-level has the potential to benefit not only the current participants, but if sustained, future participants as well. Thus improving afterschool program quality holds promise for longer-term impacts.

Over the past decade, the Wallace Foundation has been one of the key organizations, along with the Mott and William T. Grant Foundations, seeking to find ways to build systematic approaches to funding, managing, and fostering quality afterschool programs in cities across the United States. In the Wallace Foundation Report, *Growing Together, Learning Together*, (, 2015). the Wallace Foundation has identified key elements to access including strong city leadership, coordination among multiple youth-serving organizations, the effective use of data, and a comprehensive approach to quality (Wallace, 2015).

Scholarship on afterschool has sought to identify characteristics of quality programming (Eccles & Gootman, 2002). Appropriate structure has been found to be an important aspect of afterschool quality because when monitoring and supervision is lacking in afterschool, youth become involved with negative peers and problem behavior (Little, Weimer, & Weiss,

2008; Mahoney, Stattin, & Lord, 2004). Appropriate management of behavior in afterschool, coupled with adult support, has been found to be related to higher quality experiences in which youth exercise opportunities to self-regulate, and to lead (Cross, Gottfredson, Wilson, Rorie, & Connell, 2010; Eccles & Gootman, 2002; Durlak, Pachan, & Weissberg, 2010). Youth need both direction and warmth from caring adults in their lives.

Bonding with caring adults is another important social process characterizing quality programs in that youth who feel connected to adults in their lives, adopt the behavioral goals and means of the adults they respect (Hirschi, 1969). Empirical research has demonstrated that youth who feel more connectedness evidence fewer emotional symptoms, less problem behavior, and exercise more collective efficacy, that is a sense of connectedness and willingness among groups to positively influence each other's behavior for good (Odgers, Moffett, Tach, 2009; Sampson, Raudenbush, & Earls, 1997; Smith, Osgood, Caldwell, Hynes, & Perkins, 2013). There are multiple studies of afterschool quality that have revealed that children who feel supported by the adults in their programs, participate more and perform better. Miller (2005) in her study of afterschool programs found that having engaging staff was associated with engaged youth, who were willing to tackle challenging social and academic activities. Pierce, Hamm and Vandell (2010) in their study of youth in Midwestern afterschool programs found that supportive interactions with staff in afterschool (measured by observed staff enthusiasm, warmth, and smiling) was a key variable related to higher reading and math scores in 2nd and 3rd grades for youth. Clearly, feeling supported by the adults in their lives is meaningful to children's social and academic development.

However, though interactions with adults are important, there are some lesser-studied youth-oriented dimensions, such as agency and engagement that are found to be key in youth development (Larson, 2000). Larson found that youth are more likely to report being actively engaged and focused in activities with their peers (Larson, 2000). In afterschool settings with more stable staffing, effective management and climate, youth report more positive experiences (Cross, Gottfredson, Wilson, Rorie, & Connell, 2010). Yet, while there are numerous studies of adult support, youth experience and engagement has received relatively less empirical attention (Fredricks, Bohnert, & Burdette, 2014).

Previous research has identified and examined several key aspects of quality afterschool programs, including appropriate structure, supportive relationships with adults and peers, and opportunities for engagement. These are all proximal social processes and interactions that are potentially salient to positive youth development (Eccles & Gootman, 2002; Kuperminc, Smith, & Henrich, 2013; Larson, 2000; Lerner et al., 2005; Yohalem & Wilson-Ahlstrom, 2010). In light of substantial scholarship focusing upon conceptualizing and assessing quality in afterschool, increasingly, more empirical research is turning to the question of how to *foster* quality in afterschool (Eccles & Gootman, 2002; Granger, 2010).

To date, some of the more effective initiatives in afterschool have used a continuous quality improvement (CQI) process of assessment, training, coaching, and data feedback, approaches that have evidenced effects upon local cities, programs *and* upon children, their attendance and participation (Wallace, 2015). Sheldon and colleagues found that they could increase the quality of afterschool reading instruction using CQI-type approaches and,

quality of instruction was in turn related to student gains in reading, with youth in higher quality programs making larger gains than those in lower quality programs (Sheldon, Arbretton, Hopkins, & Grossman, 2010). Another initiative in afterschool sought to build the capacity of youth program leaders and staff to engage in data-informed CQI processes and resulted in increased staff program monitoring and improved practices (C. Smith et al., 2012). The value of these approaches is that they go beyond single workshops to investigate how to implement and sustain the quality of implementation of best practices in afterschool settings. The current study contributes to this growing body of work on strengthening afterschool by providing technical assistance to afterschool programs using an evidence-based program, the Pax Good Behavior Game (PAX GBG). We extend the work on PAX GBG in afterschool by examining the impact on the overall quality of the afterschool setting and, on youth outcomes.

Approaches to Improving Afterschool Settings: A Science Migration Study

Providing appropriate structure for afterschool participants is paramount to engaging them in academic and developmental activities; little can be accomplished in a disorderly afterschool program (Cross et al., 2010). In this science-migration study, we examine a creative approach to behavior management, a cooperative, team-based game developed for schools, the Good Behavior Game (GBG, Barrish, Saunders, & Wolf, 1969), in afterschool. GBG is based upon life-course, social field theory that suggests that structuring the environment of youth at a critical developmental juncture fosters self-regulation, and socio-emotional skill enhancement that poises youth for long-term adaptive outcomes. These premises were tested in a randomized trial of first and second grade classrooms in Baltimore in which short-term positive effects of GBG were found on reduced hyperactivity, conduct problems, and improved reading and math achievement, particularly for the most aggressive boys (Ialongo, et al., 1999). Later work revealed that sixth graders, exposed to GBG in 1st and 2nd grade, were reported to have less problem behavior, were less likely to be suspended from school or, need mental health services (Ialongo, Poduska, Werthamer, & Kellam, 2001). Longitudinal effects of GBG have been found at ages 19–21 with participating males evidencing reduced rates of drug and alcohol use, delinquency, and incarceration (Kellam et al., 2008). These findings support the premise that intervention at a critical development point affects longer term outcomes in early adulthood.

Yet, less is known about the actual processes by which the game might attain effects upon children's behavior. One study examined the degree to which the game might affect teacher praise and student behavior (Lannie & McCurdy, 2007). However, this study did not find the effects upon teacher behavior as anticipated. At this juncture, more research is needed on the degree to which this cooperative game affects various hypothesized social processes in afterschool. This investigation explores the degree to which GBG might lead to warmer, more supportive and engaging interactions among adults and youth in afterschool that also include appropriate amounts of adult monitoring and supervision; we surmise that these social processes could be impacted by the intervention, GBG.

In recent years, the PAXIS Institute has developed a commercialized, manualized, disseminable, packaged version called PAX (which means *peace* in Latin) GBG (Embry, et

al., 2010). PAX GBG begins by involving the adult staff and youth in creating a shared vision of their afterschool program that includes a set of verbal and visual cues reminding the youth to behave their best. Praise is used liberally in both vocal and written forms throughout the settings. Youth are assigned to teams of children that encourage each other to behave during the timed game in order to earn contingent team-based activity rewards that are allowed for short periods of times (e.g. pencil-tapping, or active dancing); adults are encouraged to join in the activity rewards in ways that are enjoyable for all. With this package of features, PAX GBG potentially affects several social processes: 1) appropriate structure and adult support in that staff are involved in providing clear instructions, without harsh criticism, coupled with ample praise and involvement in the activity rewards and; 2) youth agency, belonging, and connectedness in that youth are engaged in envisioning their desired afterschool program and, in monitoring and encouraging their peers. Other pilot studies have found that PAX GBG demonstrated some promise for impacting key characteristics of afterschool settings and youth outcomes (Frazier, Capella, & Atkins, 2007; Frazier et al., 2013). In the current study, PAX GBG was adapted to account for the distinctive features of afterschool including multiple ages of children (versus classrooms of similarly-aged children), multiple staff (versus 1 classroom teacher) with various educational backgrounds in fluid afterschool locations such as gymnasiums and cafeterias (Hynes, Smith, & Perkins, 2009).

Of particular interest to the research team was the degree to which we could foster high levels of implementation fidelity across various afterschool programs in diverse socio-geographic locales. Implementation essentially examines the degree to which program practices are conducted with fidelity, true to the original program design (Fixsen, Naoom, Blase, & Friedman, 2005; Wanless & Domitrovich, 2015). Substantial research has demonstrated that programs with greater implementation fidelity evidence better effects (see for example Cross, et al., 2010; Moncher & Prinz, 1991). The current study examines implementation fidelity as a potential moderator of effects upon individual and program-level ecologies.

Summary and Research Questions

In summary, afterschool programs are important contexts for prevention and promotion given the large amount of youth risky behavior that occurs in out-of-school time, provided they are quality programs. This science-migration study tests an evidence-based approach originally developed for school classrooms in afterschool. The degree to which staff in these settings implement these practices with fidelity likely matters in terms of effects upon social processes such as appropriate structure, support, youth engagement, belonging, and ultimately reduced problem behavior. Two main research questions guided this study. First we asked, to what degree staff in the experimental settings implemented PaxGBG with fidelity. Second we examined the intent-to-treat (ITT) impact of PAX GBG, along with the potentially moderating impact of implementation fidelity on afterschool program quality and child outcomes.

Methodology

Sample and Design

A randomized trial was conducted among 76 afterschool sites in a northeastern state within a 240 mile radius including urban, suburban, and rural locales. Afterschool programs serving public elementary-school children in grades kindergarten to fifth were identified via multiple approaches: 1) contacting and searching the websites of local school districts for their afterschool care providers; and 2) systematically searching for local community-based agencies such as the YM/YWCA, Boys and Girls Club (BGC), and local Parks and Recreation Commissions (PR) that provided afterschool programming. Agencies offering programs for this age group most days of the week throughout the academic school year were included, and no programs meeting this criteria were excluded. The programs typically served youth from the end of the school day until 5:30 pm or 6:00 pm; all operated five days per school week. Across three years (i.e., 2009–10, 2010–11, and 2011–12) providers were approached in spring and began participation the following fall. The team successfully recruited 12 of 14 program providers who operated a range of 2–12 program sites, an 86 percent recruitment rate. At the site level, we retained 76 of 83 program sites recruited (92 percent). Two sites whose match dropped were excluded from the analyses. Further, one site that failed to collect sufficient child data across time was not included in the analysis, and its matching site was excluded as well resulting in 72 total sites.

In this study, we were interested in effects upon both the programs and the participating youth. Consent forms were sent to the parents of participating children in grades 2–5. Parents and children could refuse at any time and any previous data from their children would be deleted. Participation in the child survey varied across the sites ranging from 72–90 percent (kindergarten and first grade children were not included due to lower literacy levels at these stages of development). Children completed the surveys during the afterschool program using PDA's that they read on their own or with assistance from research staff. Children received an incentive (a string bag or water bottle) for their participation. The survey ranged in time from 45–60 minutes with short cartoon and joke breaks programmed between 15-minute sections on the PDA's. The survey included measures of youth problem and prosocial behavior (the focus of this study).

Program sites (regardless of provider) were matched on geographic locale (urban, suburban, rural), racial-ethnic composition (similar proportions of minority/non-minority students), and socio-economic status (measured by free/reduced lunch status of the elementary school served) and randomly assigned to condition. In an effort to engender trust and offer transparency for staff who might be wary of research, a project kick-off was held following recruitment to remind and inform program staff and directors of the nature of the project, the research timeframe, and tasks; and importantly, to hear and any address concerns (Smith, Wise, Rosen, Rosen, Childs, & McManus, 2014). Randomization was determined by calling up staff representing the matched program pairs, and having one person in each pair to flip a coin to determine the experimental condition. The staff and directors of the experimental programs discretely remained 1.5 hours following the kickoff to learn the dates of the upcoming trainings and to be introduced to their coach. The experimental conditions

consisted of PAX GBG versus the business-as-usual control condition, (data collection but no intervention). The demographic characteristics from director report and census data are provided in Table 1.

Intervention Procedures

This project tested the impact of PAX GBG upon both afterschool quality and youth behavior. PAX GBG encompassed a set of strategies that have at its core a cooperative game played among teams of children who earn contingent group-based rewards by minimizing off-task behavior during the timed-game (Embry, et al., 2010). PAX GBG was introduced by having afterschool staff and youth join in creating a shared vision for their “Wonderful Afterschool Program.” PAX GBG included strategies for managing transitions to new activities, settings, and the appropriate voice or activity levels. At the crux of Pax was the use of advance instructions, coupled with liberal contingent praise for being on-task and gentle redirection of disruptive behavior. Youth and staff periodically wrote notes of praise and gratitude to each other. In this study, teams were comprised of a mix of 4–5 children who varied in age, gender, and other behavioral characteristics. The teams were comprised in collaboration with staff and children and, were varied throughout the year to reduce boredom with the game. The time-period of the game ranged from 1–30 minutes and increased as youth became experienced in the game. Teams displaying 3 or less misbehaviors could “win” the game earning an activity reward involving both adults and children to enhance bonding (e.g. active dancing, pencil taps, jumping). Additional features were added after mastering the basic elements, including team jobs in which the youth assisted in leading and monitoring the game.

Our processes for recruiting and engaging afterschool staff involved strategies that were collaborative and attuned to garnering the support of multiple levels of management in the programs (Smith et al., 2014). Afterschool staff and directors in the experimental condition received four trainings sessions in PAX GBG lasting 3–4 hours, each comprised of didactic instruction and interactive activities for staff that facilitated opportunities to apply their learning and plan for implementation in their own sites. The last training allowed the site staff to review and plan for the upcoming summer or academic year thus encompassing a gradual sequencing of training designed to facilitate implementation fidelity and sustainability. Sites in the experimental condition received a coach who first visited to observe the site, returning weekly to provide technical assistance across 20–24 weeks of intervention. (The detailed PAX GBG Afterschool Manual is available upon request to the corresponding author or the PAXIS Institute).

The Observational Measurement Protocol

We utilized observational approaches to characterize the quality of the social processes within the afterschool settings focusing on the levels of adult and peer support, appropriate structure, and youth belonging and engagement in the program activities (Shinn & Rapkin, 2000; Tseng & Seidman, 2007). For each cohort, trained observers, blind to condition, visited programs on varying days for a 90–120 minute visit, rating them 5 times over the course of an academic year; the 2 pre ratings in fall and the 2 post ratings in spring were utilized in this study. To assess inter-rater reliability, 50 percent involved 2 live,

simultaneous and independent ratings. Observers received a two-day, 16-hour training and 4-hour booster trainings in fall and spring before the data collection waves (for details see Oh, Osgood, Smith, 2015). A group of scientific experts in education and developmental science established “gold standard” scoring via a consensus process in which the experts reached rates of 80 to 90 percent agreement on afterschool videos. Before being deployed, the data collectors matched the GSV scores at 80 percent or higher to prevent drift and promote reliability and accuracy (Oh, Osgood, & Smith, 2015; Stuhlman, Hamre, Downer, & Pianta, 2010). The observational protocol included measures of implementation fidelity and afterschool program quality.

Implementation Quality: The Afterschool Climate Assessment (ACA).—In trials in which groups are randomly assigned to experimental and treatment groups, intent-to-treat analyses examine comparisons between the randomized treatment and control group even though some of the treatment group may have not fully participated. Low implementing experimental sites might not demonstrate effects as strongly as ones that implement with more fidelity. Further, some savvy staff in control sites might be integrating empirically-based strategies into their own program practices, like clear guidelines and ample praise that are at the foundation of PAX GBG.

To examine these possibilities, we assessed implementation fidelity using an index of evidence-based practices promoted by PAX GBG and fostered by training and coaching. Implementation fidelity was rated by independent observers blind to experimental condition. This binary index sum of 10 items (yes/no) was developed specifically for this project to assess the use of evidence-based practices in both the experimental and control conditions. Sample items included “positive verbal reinforcement,” “clear and concise directions for activities,” “standard discipline programs used” and “clear rules/expectations posted.” We computed this measure, and all others, as means across items and (when relevant) across raters. Variation across sites in terms of implementation might have contributed to the items of the implementation measure being inconsistently intercorrelated across sites. In sites implementing all the strategies, these items would be closely correlated, whereas in sites implementing only a few strategies, the items would be less correlated, affecting the internal consistency of the measure. Because afterschool sites might be demonstrating a range of these strategies, high internal consistency reliability was not expected. This was evident in the lower Cronbach’s α of .62 indicating the level of internal consistency reliability. We assessed interrater reliability as an intraclass correlation coefficient, computed for the entire dataset of five waves of observations for all three cohorts. The interrater reliability of the ACA was .77, which is acceptable (Fleiss, 1981; Raudenbush, Martinez, Bloom, Zhu, & Lin, 2012). The internal consistency (ranging from .55 - .92) and inter-rater reliability (ranging from .34 - .77) of all of the observational measures are presented in Table 2. According to criteria proposed by Fleiss (1981; i.e., <.40, poor; .40-.59, fair; .60-.74, good; >.74, excellent), inter-rater reliability values for a majority of our scales and subscales were fair to good. The specific ratings are described below with their accompanying measures.

Setting Quality Measures.—The observational protocol gathered data on afterschool program quality. Because researchers have emphasized the importance of multiple measures

to capture the nature of theorized interactions in educational settings (Pianta & Hamre, 2009; Yohalem & Wilson-Ahlstrom, 2010), we used three reliable and valid tools popular in research on both afterschool and early childhood settings: the Caregiver Interaction Scale (CIS, Arnett, 1989); the Promising Practices Rating Scale (PPRS, Vandell et al., 2004); and the Youth Program Quality Assessment (YPQA, C. Smith & Hohmann, 2005). The CIS was particularly helpful in providing descriptions of the care-giving styles of individual staff (e.g. harsh, sensitive, permissive, uninvolved) aggregated across the program. In contrast, the PPRS, focused upon characterizing the nature of social processes such as the interaction between adults and children, or among the children themselves (e.g. adult support, peer support). The YPQA, also focused upon program-level characteristics, assessed the more unique attributes of youth belonging and responsibility in the afterschool programs.

Arnett's Caretaker Interaction Scale (CIS).—Developed by Arnett (1989), the CIS examined the interactions of caregiving staff with children in the following 4 areas: 1) harshness and criticism; 2) sensitivity - warmth and communication; 3) detachment - disinterest, and involvement in adult-oriented activities excluding children; and 4) permissiveness - staff failure to appropriately provide guidance and redirection when necessary. Observers rated up to 3 permanent, non-volunteer staff in each afterschool program on a 4-point response scale indicating the extent to which they engaged in a particular behavior or practice, where 1 represented *never* (0%); 2, *few instances* (1–30%); 3, *many instances* (31–60%); and 4, *consistently* (>61%) over the course of the entire program session for the day. The interrater reliabilities on this measure ranged from .58 - .77 across time and the internal consistency assessed using Cronbach's alpha ranged from .75 – .92. (Table 2).

Promising Practices Rating Scale (PPRS).—The PPRS is an observational tool developed by Vandell and Colleagues (2004) to study afterschool program quality and practices. In this study we measured 5 of the PPRS concepts salient to our conceptual model: 1. Supportive relations with adults (SRA), 2. Supportive relations with peers (SRP), 3. Appropriate structure (AS), 4. Level of engagement (LE), and 5. Chaos. These concepts were rated on a 4-point scale indicating the extent to which a given construct was characteristic of the program: where 1=*highly uncharacteristic*; 2=*somewhat uncharacteristic*; 3=*somewhat characteristic*; and 4=*highly characteristic*. In that single-item measures may have limited reliability (Nunnally & Berstein, 1994), we adapted the single-item indices by using the provided descriptive exemplars as items to produce an average score on the subscale (Oh, Osgood, & Smith, 2015; Table 2). The interrater reliabilities on this measure ranged from .50 - .67 and the internal consistency assessed using Cronbach's alpha ranged from .67 - .89.

The Youth Program Quality Assessment (YPQA).—The YPQA was administered following the PPRS; it focused upon several overarching dimensions of program quality thought to be salient to our conceptual model, youth engagement, belonging, conflict management, responsibility, and choice, including some potential concepts that we expected would be impacted by the use of a cooperative game among staff and youth that gave youth opportunities for team involvement and leadership (C. Smith & Hohmann, 2005). The

YPQA was used in this study to rate the overarching program. The YPQA was rated using scores of 1, 3, and 5 where 1 indicated that no children had access to this experience, 3, some children had access to this experience, and 5, most children had access to this experience (Table 2). The interrater reliabilities on this measure ranged from .34 - .58 and the internal consistency assessed using Cronbach's alpha ranged from .55 - .89.

Child Outcomes

Strengths and Difficulties Questionnaire (SDQ).: Children's behavioral outcomes were assessed using child reports of the Strengths and Difficulties Questionnaire (SDQ; Goodman, Meltzer, & Bailey, 2003; Mellor, 2004). The SDQ in this study was comprised of 22 items to which participants responded on a 3-point scale indicating the degree to which each item was "not true, sometimes true, or very true." The SDQ items were used to calculate an average total score and subscale scores on hyperactivity (inability to sit still or concentrate, internal consistency $\alpha = .79$), emotional symptoms (headaches, worries, unhappy, nervous, $\alpha = .76$), conduct problems (loses temper, lies, cheats, $\alpha = .65$), and prosocial behavior (considerate, shares, helpful, kind, $\alpha = .65$). (Means and standard deviations available in Table 1).

Problem behaviors and substance use (PBSU): were assessed by a developmentally-appropriate self-report measure for children obtained from Loeber and colleagues' Pittsburgh longitudinal study of delinquency (Russo et al., 1993). These items began by asking children if they knew how and where to obtain fairly mundane items like apples or money, progressing to riskier items like cigarettes or alcohol. The five items assessed involvement in experimenting with substances and problem behaviors to which youth could respond yes or no. A count variable was created measuring the total number of problem behaviors for which children reported an affirmative response with scores that ranged from 0 to 5; $M = .11$, $SD = .23$. Items included theft (taking things from others that don't belong to you), vandalism (destroying or damaging something that doesn't belong to you), smoking cigarettes, drinking alcohol, and experimenting with marijuana. The low mean indicated that at this age, few children were involved in these riskier, pre-delinquent behaviors. This scale exhibits a moderately high internal consistency reliability with a Cronbach's alpha of .71.

Analyses and Findings

A priori power calculations indicated that 72 programs would be sufficient to achieve power of .9 to detect a small program effect (i.e., Cohen's d of .2). These calculations assumed 25 children per program and used reliabilities and variances from pilot tests for the observational measures. Post-hoc power calculations based on values from the study as implemented (e.g., final sample sizes and observed variances) revealed power of .92 to .98 for a slightly larger effect of Cohen's d equals .3, and power of over .99 for a medium effect of d equals .5. The analytical approach was designed to examine the impact of PAX GBG upon setting-level processes and youth outcomes by addressing 3 major issues: 1) the first of which was necessary to evaluate demographic comparability in the experimental and control conditions assessing to what extent was random assignment successful; 2) the degree to which staff in experimental settings implemented with fidelity; and, 3) the impact of PAX

GBG upon program quality and youth outcomes. The last aim was examined using intent-to-treat analyses comparing experimental and control sites followed by more nuanced examinations of the degree to which implementing PAX GBG as designed (implementation fidelity) moderated these outcomes.

Comparability of Treatment and Control Programs

The first step in the analyses assessed the initial demographic equivalence of the experimental conditions before intervention to assure that random assignment of matched pairs resulted in relatively equivalent experimental groups at the onset of the study. There were no significant differences in baseline characteristics between children in the experimental and control sites except the proportion of the 3rd graders was slightly higher in the experimental condition, $t = -2.02$, $p < .05$ (Table 1).

Effects on Implementation Fidelity and Program Quality

Preliminary analyses revealed significant variation among observers in their mean ratings on most measures of fidelity and quality. Because all observers rated comparable numbers of both treatment and control program sites, these differences were not systematically confounded with program effects. Even so, these rater effects represented a form of error variance, that we removed by controlling for a set of thirteen dummy variables that captured differences among raters (D_{ijk}). The model also controls for the pretest assessment of the outcome measure ($Y_{Pre,ki}$), as well as for any mean difference between the two post-test rounds of observations.

The MLwiN multi-level analysis software (Rasbash, Steele, Browne, & Goldstein, 2012) was used to conduct intent-to-treat analyses of the impact of the treatment program on observed implementation fidelity (ACA) and afterschool setting quality. Testing group differences on fidelity is a necessary manipulation check that helps to verify that higher scores on the implementation of evidence-based practices were fostered by PAX GBG. These analyses used a four level MLwiN model (Rasbash et al., 2012) in which the rating by a single observer served as the Level 1 unit (one or two observers each visit), that was nested within a visit to the program site (Level 2, two observational visits during the post-test period), nested within the program site (Level 3). The pairings of program sites for random assignment served as Level 4 for the analysis, and the model allowed for dependence through residual intercept variance at all levels (Table 2 contains descriptive information on these measures).

The manipulation check reported in Table 3 revealed that programs assigned to the experimental group (versus the control sites) demonstrated increased levels of evidence-based practices fostered by PAX GBG and captured by the implementation fidelity index, the ACA, as compared to control sites. The program impact estimate of .16 for this measure corresponded to a standardized effect size of .77 (i.e., difference between treatment and control in standard deviation units), as rated by independent observers. Further, intent-to-treat analyses (comparison of experimental versus control sites) in Table 3 demonstrated that observers rated treatment program sites as having a significantly higher level of belonging for students ($p < .05$). There is some reason for concern that this finding might reflect a

chance contribution of significance testing for multiple correlated outcomes, given the significance level and number of observational measures. That concern does not apply to the ACA, which is the single fidelity index and the probability of the observed difference is $p < .001$.

Next, we found that programs in the treatment group implementing the strategy with higher fidelity (assessed by the implementation fidelity index, the ACA) also evidenced other positive social processes in the afterschool program sites. It should be noted that there was a tendency for several dimensions of afterschool quality to decrease across the academic year for afterschool programs in both conditions (Table 2). However, the interaction of implementation fidelity and experimental group status helped to sustain positive practices in some cases. The statistically significant interactions in this direction between treatment assignment and higher fidelity experimental group programs resulted in several enhanced aspects of program quality: less staff harshness/criticism (CIS-H), greater supportive relations with adults (PPRS-SRA), appropriate structure (PPRS-AS), and youth level of engagement (PPRS-LE) (Table 4). Even with the highly conservative Bonferroni correction, the effects for appropriate structure and youth level of engagement would be significant at $p < .05$ and for supportive relations with adults at $p < .10$.

Effects on Child Outcomes

The MLwiN multi-level analysis software (Rasbash, Steele, Browne, & Goldstein, 2012) was used to conduct intent-to-treat (experimental versus control) analyses of the effects of PAX GBG upon the child problem and prosocial behavioral outcomes. The outcome variables were post-test measures assessed for each child (Level 1). Our random intercept multi-level model took into account the nesting of children within the afterschool program sites (Level 2), and the nesting of those sites within the pairs from which random assignment occurred (Level 3). These analyses compared the youth in the experimental and control conditions while controlling for gender, race-ethnicity (dummy variables for African American, Latino, and youth of other racial-ethnic groups, compared to Euro-White children), grade (three dummy variables), the three annual cohorts of the study (two dummy variables), and pre-intervention scores on the outcome measure.

In our intent-to-treat analyses (Table 5), one statistically significant program effect emerged; youth in the PAX GBG experimental group had higher levels of youth-reported prosocial behavior (i.e., caring, sharing, and listening) measured by the SDQ. Subsequently, we tested models examining interactions between treatment status, implementation fidelity, and child outcomes. These analyses added the interaction between observed implementation fidelity (measured by the ACA) and treatment assignment to the analysis of child outcomes. These results, which appear in Table 6, indicated that implementation fidelity was associated with greater reductions in child-reported hyperactivity at experimental program sites. Statistically significant effects were not detected upon emotional symptoms, conduct problems, or problem behaviors (a sum of involvement in experimentation with illicit substances, theft, or vandalism) among this sample of elementary-age children. We note that for both the ITT effect on prosocial behavior and the interactive effect on hyperactivity, tests for only one five child outcomes reached significance at the $p < .05$ level. These results should be interpreted

cautiously in that light. However, previous research has noted that GBG most impacts children's hyperactivity and self-regulation.

Summary and Discussion

The purpose of this study was to examine the effectiveness of PAX GBG in afterschool programs using a training and technical assistance model thought to enhance implementation fidelity. Randomizing programs to matched pairs of experimental and control conditions was successful in achieving comparability between the groups on race-ethnicity and gender, boosting our confidence in the findings. We posed two main research questions namely: (1) did the experimental PAX GBG programs implement with fidelity and (2) were the practices associated with PAX GBG and, implemented with fidelity, effective in improving program quality and child outcomes?

Afterschool staff who received training and coaching in the experimental sites did implement PAX GBG with higher fidelity. This finding mirrors research findings that suggest a training and coaching program with afterschool staff may be useful in helping them hone their skills when offering a structured program (han, Bradshaw, Domitrovich, & Ialoongo, 2013; Durlak & DuPre, 2008; Sheldon et al., 2010; C. Smith et al., 2012). These professional development processes might be particularly important in afterschool settings where staff may not be trained professionals, with lower pay and higher turnover than teachers (Baldwin & Wilder, 2014).

High levels of implementation fidelity was critical to achieving most of the desired program-level outcomes. In the case of this study, higher fidelity sites were characterized by staff using less harsh language, more supportive relationships among adults and children, and higher levels of youth engagement. The PAX GBG program alone affected observed belonging. These outcomes are all clearly targeted in PAX GBG. PAX GBG also had an influence on children's prosocial behavior (caring, sharing, and listening) and, when PAX GBG was implemented with fidelity, children reported less hyperactivity (inability to stay seated, pay attention, etc.). As noted above, we cannot rule out that these may be chance differences and replication is needed to justify confidence in these results. Yet, the focus of PAX GBG, a cooperative game that offers group-based contingent awards shows promise that it may be targeting hypothesized salient dimensions of children's self-regulation and positive youth development.

Strengths and Limitations.

Though we detected effects on a number of program level and child outcomes (i.e. hyperactivity and prosocial skills) we did not detect effects upon youth emotional symptoms (e.g. worry and anxiety), conduct problems (lying, aggression), or problem behaviors (theft, vandalism, and experimentation with substances). This may or may not be a limitation but simply a finding. This study used specific measures of problem behavior as opposed to the total score on the SDQ that summarized across these behaviors, to avoid potentially masking important distinctions. It is possible that this game that required teams of youth to self-regulate, reduced hyperactivity, and in the longer term, might lead to effects on other areas of problem behavior. It is also possible that the timed game was better tooled to impact

hyperactivity than to impact emotional symptoms such as worry or anxiety. In this study, we were interested in both prevention and promotion, and wanted to examine the degree to which we might begin to see impact on early delinquency among children (Reid, Patterson, & Snyder, 2002). However, as anticipated, these were low frequency behaviors among this sample of elementary children; 11 percent of the youth reported such behaviors (e.g. theft and taking things that do not belong to them, vandalism/destroying things that do not belong to them, and tasting/experimentation with substances), potentially restricting variance and the ability to detect effects on this outcome. Some of our initial findings suggested that the number of minutes playing PAX GBG might be the factor most influencing problem behaviors of this sort (Smith et al., 2013), thus, the development of “go/no go” skills might support developmental social-field theory that early skill development impacts longer term outcomes. In future research, we could examine whether decreasing hyperactivity and increasing prosocial skills interrupts a potentially negative developmental trajectory.

The current study focused upon a sample of programs that were quite diverse both in terms of the racial-ethnic and social backgrounds of youth served, but also in terms of geographic locales. Thus, the sample is representative of a broad spectrum of programs and participating youth for both African-American and White youth, across urban, suburban, and rural locales and a variety of socio-economic backgrounds. However despite our best efforts, the sample was more limited in terms of representing Latino youth, a growing population in the U.S.

Afterschool is a burgeoning research setting and reliable and valid measurement tools are still being developed for this context. In general, our measures exhibited acceptable internal consistency, but lower inter-rater reliability even with our substantial initiatives to prevent drift and promote agreement. The intraclass coefficient (ICC) is a more stringent measure affected by the number of observers, 1.7 average in our study. According to criteria proposed by Fleiss (1981; i.e., <.40, poor; .40-.59, fair; .60-.74, good; >.74, excellent), inter-rater reliability values for a majority of our scales and subscales were fair to good. Reliability was likely affected by the substantial variability across raters and time conducting observations of these live, real-time social processes. This variability is not totally surprising in that live observations in afterschool usually encompass multiple adults, and several groups of youth, sometimes engaged in a variety of activities, across a range of physical spaces. This is a lot for trained observers to capture at the same time even with substantial training. Yet, with these levels of reliability, we were successful in detecting important setting-level effects upon program quality and youth outcomes.

Implications for Future Research and Practice.

Future research should consider the critical question of how to foster both implementation fidelity and setting-level quality in afterschool programs. A number of scholars have emphasized the need to get into “the black box” to understand the process that produces effects upon program quality and youth outcomes (Granger, 2010; Yohalem, & Wilson-Ahlstrom, 2010). This study contributes data across time helping to understand how a strategy implemented with fidelity might impact setting-level social processes and youth outcomes.

Implementation fidelity is best understood from a multilevel framework that recognizes the dialectical associations among context, program content, and individual actors (e.g., Domitrovich, 2008; Durlak & Dupre, 2008; Dusenbury, Brannigan, Falco, & Hansen, 2003; Fixsen et al., 2005; Han & Weiss, 2005; Pas, Waasdrop & Bradshaw, 2015). Taking into account the afterschool context, our focus was on training and supporting individual staff, critical and among the more common forms of improving implementation fidelity (Durlak & Dupre, 2008; Fixsen et al., 2005; Han & Weiss, 2005). Literature suggests that although training is important, it is not sufficient to enhance lasting implementation fidelity. Regular support and supervision are also required in real life situations (e.g., Fixsen et al., 2005). Support and supervision in response to naturally occurring situations is linked to better implementation fidelity and student outcomes (e.g., Domitrovich, Gest, Gill, Bierman, Welsh, & Jones, 2009; Rohrbach, Gunning, Sun, & Sussman, 2010). In particular, the two-phased coaching model may be useful for tailoring coaching support to staff working in culturally and socially diverse contexts to achieve implementation fidelity while at the same time making necessary but non-contraindicated program adaptations. From a multilevel framework perspective, we reasoned that change at the contextual level would contribute to change at the individual level.

More longitudinal work would allow us to parse out whether changes in these social processes produced changes in youth outcomes. Further, it is possible that different typologies of children, staff, programs, as well as variations in neighborhood context, differentially benefit from improvements made to afterschool programs. Another important consideration, is the process for developing organizational capacity to continue and sustain the improvement process using collaborative data feedback and action plans. Oftentimes, those most in need of intervention are less-well capacitated to accept change efforts. Capacity-building initiatives are needed to help a broad cross-section of programs benefit from intervention.

Another course of future research could examine varying formats, timing, and lengths of training programs. So often programs deliver a large dose when less training and technical assistance might be sufficient. Identifying the optimal amount of assistance could also be more cost-effective in the long run (Collins, Murphy, & Strecher, 2007). In conclusion, PAX GBG is a cooperative game that was adapted in this study for afterschool staff and youth, helping them to clarify shared behavioral expectations, engaging staff in using more supportive approaches, and encouraging teams of youth to self-regulate and support their peers in order to “win” the game. This large-scale, cluster randomized trial used a sample of staff and youth, diverse in their racial-ethnic, socio-economic, and geographic backgrounds, demonstrating that we can improve afterschool setting-level quality in ways that benefit positive youth development.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We acknowledge funding support from William T. Grant Foundation [Grant # 8529]; the Wallace Foundation [Grant #20080489]; and the National Institute for Drug Abuse [Grant # R01 DA025187]. We acknowledge former W. T. Grant Executives, Robert Granger, Edward Seidman and Vivian Tseng, whose feedback on this study was invaluable. We are also grateful for the many staff, parents, and children whose participation made this study possible.

References

- Afterschool Alliance (2014). *America After 3pm; Afterschool Programs in Demand*. Washington, D.C.
- Arnett J (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology*, 10(4), 541–552.
- Baldwin CK, & Wilder Q (2014). Inside Quality: Examination of Quality Improvement Processes in Afterschool Youth Programs. *Child & Youth Services*, 35(2), 152–168. doi:10.1080/0145935X.2014.924346
- Barker RG (1968). *Ecological psychology: Concepts and methods for studying the environment of human behavior*. Stanford University Press.
- Barrish H, Saunders M & Wolf MM(1969). Good Behavior Game: Effects of individual contingencies for group consequences on disruptive behavior in a classroom. *Journal of Applied Behavioral Analysis*, 2(2), 119–124.
- Becker KD, Bradshaw CP, Domitrovich C & Ialongo NS (2013). Coaching teachers to improve implementation of the Good Behavior Game. *Administration and Policy in Mental Health*, 40, 482–493. [PubMed: 23504295]
- Belgrave FZ, Reed MC, Plybon LE, Butler DS, Allison KW, & Davis T (2004). An evaluation of Sisters of Nia: A cultural program for African American girls. *Journal of Black Psychology*, 30(3), 329–343.
- Bronfenbrenner U (1986). Ecology of the Family as a Context for Human Development: Research Perspectives. *Developmental Psychology*, 22(6), 723–742.
- Catalano RF, Berglund ML, Ryan JA, Lonczak HS, & Hawkins JD (2002). Positive Youth Development in the United States: Research Findings on Evaluations of Positive Youth Development Programs. *Prevention & Treatment*, 5(1), 15a.
- Collins LM, Murphy SA, & Strecher V (2007). The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): New methods for more potent eHealth interventions. *American Journal of Preventive Medicine*, 32(5), S112–S118. [PubMed: 17466815]
- Cross AB, Gottfredson DC, Wilson DM, Rorie M, & Connell N (2010). Implementation quality and positive experiences in after-school programs. *American Journal of Community Psychology*, 45, 370–380. [PubMed: 20229226]
- Domitrovich CE, Bradshaw CP, Greenberg MT, Embry D, Poduska JM, & Ialongo NS (2010). Integrated models of school-based prevention: Logic and theory. *Psychology in the Schools*, 47(1), 71–88. [PubMed: 27182089]
- Durlak JA, Weissberg RP, Pachan M (2010). A meta-analysis of after-school programs that seek to promote personal and social skills in children and adolescents. *American Journal of Community Psychology*, 5, 294–309. DOI 10.1007/s10464-010
- Durlak JA, & DuPre E (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41, 327–350. [PubMed: 18322790]
- Eccles J & Gootman JA (Eds.). (2002). *Community programs to promote youth Development*. National Academies Press.
- Embry DD, Richardson C, Schaffer K, et al. 2010 PAX Good Behavior Game, 3rd ed. Tucson, AZ: PAXIS Institute.
- Embry DD (2002). “The Good Behavior Game: a best practice candidate as a universal behavioral vaccine.” *Clinical Child & Family Psychology Review*, 5(4): 273–297. [PubMed: 12495270]

- Fairweather GW (1972). *Social change: The challenge to survival*. Morristown, NJ: General Learning Press.
- Fixsen DL, Naoom SF, Blase KA, & Friedman RM (2005). *Implementation research: A synthesis of the literature*. <http://www.popline.org/node/266329>
- Fleiss JL (1981). The measurement of interrater agreement In Fleiss JL (Ed.), *Statistical methods for rates and proportions* (pp. 212–236). New York, NY: John Wiley.
- Frazier SL, Cappella E, & Atkins MS (2007). Linking mental health and after school systems for children in urban poverty: Preventing problems, promoting possibilities. *Administration and Policy in Mental Health and Mental Health Services Research*, 34(4), 389–399. [PubMed: 17340183]
- Frazier SL, Mehta TG, Atkins MS, Hur K, & Rusch D (2013). Not just a walk in the park: Efficacy to effectiveness for after school programs in communities of concentrated urban poverty. *Administration and Policy in Mental Health and Mental Health Services Research*, 40(5), 406–418. [PubMed: 22843303]
- Fredricks JA, Bohnert AM, & Burdette K (2014). Moving beyond Attendance: Lessons Learned from Assessing Engagement in Afterschool Contexts. *New Directions For Youth Development*, (144), 45–58.
- Goodman R, Meltzer H, & Bailey V (2003). The Strengths and Difficulties Questionnaire: A pilot study on the validity of the self-report version. *International Review of Psychiatry*, 15, 173–177. [PubMed: 12745329]
- Gorman-Smith D, Tolan PH, & Henry DB (2000). A developmental-ecological model of the relation of family functioning to patterns of delinquency. *Journal of Quantitative Criminology*, 16(2), 169–198.
- Gottfredson D, Cross AB, Wilson D, Rorie M, & Connell N (2010). Effects of Participation in After-School Programs for Middle School Students: A Randomized Trial. *Journal Of Research On Educational Effectiveness*, 3(3), 282–313. doi:10.1080/19345741003686659
- Gottfredson DC, Gerstenblith SA, Soulé DA, Womer SC, & Lu S (2004). Do after school programs reduce delinquency? *Prevention Science*, 5(4), 253–266. [PubMed: 15566051]
- Granger RC (2010). Understanding and Improving the Effectiveness of After-School Practice. *American Journal of Community Psychology*, 45(3–4), 441–446. [PubMed: 20238158]
- Heath SB, & McLaughlin MW (1994). The best of both worlds: Connecting schools and community youth organizations for all-day, all-year learning. *Educational Administration Quarterly*, 30(3), 278–300.
- Hirschi T (1969). *Causes of Delinquency*. Berkeley, CA: University of California Press.
- Hynes K and Sanders F (2011). Diverging experiences during out-of-school time: The race gap in exposure to after-school programs. *The Journal of Negro Education*, 80(4), 464–476.
- Hynes K, Smith E, & Perkins D (2009). Piloting a Classroom-Based Intervention in After-School Programmes: A Case Study in Science Migration. *Journal of Children's Services*, 4(3), 4–20.
- Ialongo N, Poduska J, Werthamer L, & Kellam S (2001). The distal impact of two first grade preventive interventions on conduct problems and disorder and mental health service need and utilization in early adolescence. *Journal of Emotional and Behavioral Disorders*, 9, 146–160.
- Ialongo NS, Werthamer L, Kellam SG, Brown CH, Wang S, & Lin Y (1999). Proximal impact of two first-grade preventive interventions on the early risk behaviors for later substance abuse, depression, and antisocial behavior. *American Journal of Community Psychology*, 27(5), 599–641. [PubMed: 10676542]
- James-Burdumy S, Dynarski M, Moore M, Deke J, Mansfield W and Pistorino C (2005). *When Schools Stay Open Late: The National Evaluation of the 21st Century Community Learning Centers Program: Final Report*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, 2005 Available at <http://www.ed.gov/ies/ncee>
- Kellam SG, Brown CH; Poduska JM; Ialongo NS; Wang W; Toyinbo P; Petras H; Ford C; Windham A; Wilcox CH (2008). Effects of a universal classroom behavior management program in first and second grades on young adult behavioral, psychiatric, and social outcomes. *Drug and Alcohol Dependence*, 95S5–S28. Doi:10.1016/j.drugalcdep.2008.01.004

- Kuperminc GP, Smith EP, & Henrich CC (2013). Introduction to the special issue on “Social and motivational processes in after-school settings: Bridging gaps between theory, research, and practice.” *The Journal of Early Adolescence*, 33(1), 5–16.
- Lannie AL, & McCurdy BL (2007). Preventing Disruptive Behavior in the Urban Classroom: Effects of the Good Behavior Game on Student and Teacher Behavior. *Education & Treatment of Children*, 30(1), 85–98.
- Larson R (2000). “Toward a psychology of positive youth development.” *American Psychologist*, 55, 170–183. [PubMed: 11392861]
- Lauer PA, Akiba M, Wilkerson SB, Apthorp HS, Snow D, & Martin-Glenn M (2006). Out-of-school-time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research*, 76(2), 275–313.
- Lerner RM, Lerner JV, Almerigi JB, Theokas C, Phelps E, Gestsdottir S, ... & Von Eye A (2005). Positive Youth Development, Participation in community youth development programs, and community contributions of fifth-grade adolescents findings from the first wave of the 4-H study of Positive Youth Development. *The Journal of Early Adolescence*, 25(1), 17–71.
- Lillehoj CJ, Griffin KW, & Spoth R (2004). Program provider and observer ratings of school-based preventive intervention implementation: Agreement and relation to youth outcomes. *Health Education and Behavior*, 31, 242–257. [PubMed: 15090124]
- Little P, Wimer C, & Weiss HB (2008). After school programs in the 21st century: Their potential and what it takes to achieve it. *Issues and opportunities in out-of-school time evaluation*, 10, 1–12.
- Mahoney JL, Lord H, & Carryl E (2005). An ecological analysis of after-school program participation and the development of academic performance and motivational attributes for disadvantaged children. *Child Development*, 76(4), 811–825. [PubMed: 16026498]
- Mahoney JL, Stattin H & Lord H (2004). Unstructured youth recreation centre participation and antisocial behavior development: Selection influences and the moderating role of antisocial peers. *International Journal of Behavioral Development*, 28(6), 553–560.
- Mahoney JL, & Zigler EF (2006). Translating science to policy under the No Child Left Behind Act of 2001: Lessons from the national evaluation of the 21st-Century Community Learning Centers. *Journal of Applied Developmental Psychology*, 27(4), 282–294.
- Mellor D (2004). Furthering the use of the Strengths and Difficulties Questionnaire: Reliability with younger child respondents. *Psychological Assessment*, 16, 12 2004, 396–401. [PubMed: 15584800]
- Miller BM (2005). Pathways to success for youth: What counts in after-school. Wellesley, MA, National Institute on Out-of-School-Time. <http://www.uwmb.org/news/05-mars-study.html>
- Moncher FJ, & Prinz RJ (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review*, 11(3), 247–266.
- Nunnally J, & Berstein I (1994). *Psychometric theory*. New York: McGraw-Hill.
- Ogders CL.; Moffett TE; Tach ML, Sampson RJ; Taylor A; and Matthews CL (2009). The protective effects of neighborhood collective efficacy on British children growing up in deprivation: A developmental analysis. *Developmental Psychology*, 45(4)942–957. [PubMed: 19586172]
- Oh Y, Osgood DW, & Smith EP (2015). Measuring afterschool program quality using setting-level observational approaches. *The Journal of Early Adolescence*, 35(5–6), 681–713. [PubMed: 26819487]
- Osgood DW, & Anderson AL (2004). Unstructured Socializing and Rates of Delinquency. *Criminology*, (3), 519.
- Pettigrew J, Graham JW, Miller-Day M, Hecht ML, Krieger JL, & Shin YJ Adherence and delivery: Implementation quality and program outcomes for the seventh-grade keepin’ it REAL program. *Prevention Science*, 16, 90–99.
- Pianta RC, & Hamre BK (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, 38(2), 109–119.
- Pierce KM; Bolt DM; & Vandell DL (2010). Specific features of after-school program quality: Associations with children’s functioning in middle childhood. *American Journal of Community Psychology*, 45(3–4), 381–393. DOI 10.1007/s10464-010-9304-2 [PubMed: 20336364]

- Pierce KM, Hamm JV, & Vandell DL (1999). Experiences in after-school programs and children's adjustment in first-grade classrooms. *Child Development*, 70(3), 756–767. [PubMed: 10368920]
- Pittman KJ (1991). *Promoting Youth Development: Strengthening the Role of Youth Serving and Community Organizations*. Washington, D.C: Center for Youth Development and Family Research.
- Rasbash J, Steele F, Browne WJ, Goldstein H, & Charlton C (2012). A user's guide to MLwiN.
- Raudenbush SW, Martinez A, Bloom H, Zhu P, & Lin F (2011). Studying the reliability of group-level measures with implications for statistical power: A six-step paradigm. Retrieved from <http://wtgrantfoundation.org/FocusAreas#youthsocial-settings>.
- Reid JB, Patterson GR, & Snyder JE (2002). *Antisocial behavior in children and adolescents: A developmental analysis and model for intervention*. Washington, D.C: American Psychological Association.
- Riggs NR, Bohnert AM, Guzman MD & Davidson D (2010) Examining the potential of community-based after-school programs for Latino youth. *American Journal of Community Psychology* 45(3–4), 417–429. [PubMed: 20300821]
- Russo MF, Stokes GS, Lahey BB, Christ MAG, McBurnett K, Loeber R, ... & Green SM (1993). A sensation seeking scale for children: Further refinement and psychometric development. *Journal of Psychopathology and Behavioral Assessment*, 15(2), 69–86.
- Sampson RJ, Raudenbush SW, & Earls F (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science*, 277(5328), 918–924. [PubMed: 9252316]
- Shinn M, & Rapkin BD (2000). Cross-level research without cross-ups in community psychology. In *Handbook of Community Psychology* (pp. 669–695). Springer US.
- Sheldon J, Arbreton A, Hopkins L, & Grossman JB (2010). Investing in success: Key strategies for building quality in after-school programs. *American Journal of Community Psychology*, 45(3–4), 394–404. [PubMed: 20352361]
- Smith C, & Hohmann C (2005). Full findings from the Youth PQA validation study. Ypsilanti, MI: High/Scope Educational Research Foundation.
- Smith C, Akiva T, Sugar SA, Devaney T, Lo Y-J, Frank K, Peck S, & Cortina K (2012). Continuous quality improvement in afterschool settings: Impact findings from the Youth Program Quality Intervention study Ypsilanti MI: David P Weikart Center for Youth Program Quality. Washington, DC: Forum for Youth Investment.
- Smith EP, Boutte GS, Zigler E, and Finn-Stevenson Matia. (2004). Opportunities for Schools to Promote Resilience in Children and Youth In Maton KI, Schellenbach CJ, Leadbetter BJ and Solarz AL (Eds.) *Investing in children, youth, families, and communities: Strengths-based research and policy*. Washington, D. C.: American Psychological Association.
- Smith EP, Osgood DW, Caldwell L, Hynes K, & Perkins DF (2013). Measuring collective efficacy among children in community-based afterschool programs: Exploring pathways toward prevention and positive youth development. *American Journal of Community Psychology*, 1–14. doi: 10.1007/s10464-013-9574-6.
- Smith EP, Wise E, Rosen H, Rosen A, Childs S, & McManus M (2014). Top-down, Bottom-up, and Around the Jungle Gym. A Social Processes and Networks Approaches to Building Learning Communities in Afterschool. *American Journal of Community Psychology*, 53, 491–502, DOI 10.1007/s10464-014-9656-0 [PubMed: 24781678]
- Snyder HN and Sickmund M (2006) *Juvenile Offenders and Victims: 2006 National Report*. U.S. Department of Justice pp 1–261.
- Stuhlman MW, Hamre BK, Downer JT, & Pianta RC (2010). A practitioner's guide to conducting classroom observations: What the research tells us about choosing and using observational systems.
- Tebes JK, Feinn R, Vanderploeg JJ, Chinman MJ, Shepard J, Brabham T, ... & Connell C (2007). Impact of a positive youth development program in urban after-school settings on the prevention of adolescent substance use. *Journal of Adolescent Health*, 41(3), 239–247. [PubMed: 17707293]
- Tseng V and Seidman E (2007). A systems framework for understanding social settings. *American Journal of Community Psychology*. 39:217–228. DOI 10.1007/s10464-007-9101-8 [PubMed: 17436080]

- Vandell DL, Reisner ER, Brown BB, Pierce KM, Dadisman K, & Pechman EM (2004). The study of promising after-school programs: Descriptive report of the promising programs. Retrieved from <http://www.wcer.wisc.edu/childcare/statements.html>.
- Wallace Foundation (2015). Growing Together, Learning Together: What Cities Have Discovered About Building Afterschool Systems. New York, NY.
- Wanless SB, & Domitrovich CE (2015). Readiness to Implement School-Based Social-Emotional Learning Interventions: Using Research on Factors Related to Implementation to Maximize Quality. *Prevention Science*, 1–7. [PubMed: 23943148]
- Weiss HB, Little P, & Bouffard SM (2005). More than just being there: Balancing the participation equation. *New Directions for Youth Development*, 2005(105), 15–31.
- Yohalem N and Wilson-Ahlstrom A (2010). Inside the black box: Assessing and improving quality in youth programs. *American Journal of Community Psychology* 45:350–357, DOI 10.1007/s10464-010-9311 [PubMed: 20300822]

Table 1

Demographic Description of the Programs and Participants: Initial Group Equivalence

Variable	Total		Control		GBG	
	N	% or M (SD)	N	% or M (SD)	N	% or M (SD)
<u>Child Characteristics</u>	811		430		381	
<i>Gender</i>						
Girl	406	50.10%	212	49.30%	194	50.90%
Boy	405	49.90%	218	50.70%	187	49.10%
<i>Race</i>						
White	391	48.20%	194	45.10%	197	51.70%
Black/African American	233	28.70%	131	30.50%	102	26.80%
Hispanic/Latino/a	54	6.70%	29	6.70%	25	6.60%
Other	133	16.40%	76	17.70%	57	15.00%
<i>Grade</i>						
2 nd	240	29.60%	130	30.20%	110	28.90%
3 rd	232	28.60%	110	25.60%	122	32.00%
4 th	202	24.90%	108	25.10%	94	24.70%
5 th	137	16.90%	82	19.10%	55	14.40%
<i>Behavioral Outcomes at Pre</i>						
SDQ-Hyperactivity	639	1.49 (0.52)	349	1.50 (0.51)	290	1.47 (0.53)
SDQ-Emotional Symptoms	648	1.60 (0.49)	352	1.63 (0.50)	296	1.57 (0.49)
SDQ-Prosocial Behaviors	650	2.55 (0.43)	354	2.56 (0.42)	296	2.55 (0.44)
SDQ-Conduct Problems	635	1.42 (0.44)	341	1.42 (0.44)	294	1.42 (0.44)
Prob Beh/Substance Use	622	.11 (.23)	340	0.12 (0.23)	282	0.11 (0.22)
<u>Program Characteristics</u>	73		36		37	
<i>Locale</i>						
Urban	22	30.10%	10	27.80%	12	32.40%
Suburban	45	61.60%	24	66.70%	21	56.80%
Rural	6	8.20%	2	5.60%	4	10.80%
<i>% Minority</i>						
Less than 25%	23	31.50%	12	33.30%	11	29.70%
25% ~ 50%	15	20.50%	8	22.20%	7	18.90%
51% ~ 75%	14	19.20%	7	19.40%	7	18.90%
more than 75%	21	28.80%	9	25.00%	12	32.40%
<i>% Free/reduced lunch eligible</i>						
Less than 25%	22	30.10%	11	30.60%	11	29.70%
25% ~ 50%	18	24.70%	10	27.80%	8	21.60%
51% ~ 75%	15	20.50%	7	19.40%	8	21.60%
more than 75%	18	24.70%	8	22.20%	10	27.00%

Note: There was no significant difference in baseline characteristics between participants in control sites and those in GBG sites, except the proportion of the 3rd graders, $t=-2.02$, $p<.05$

Table 2 Psychometric and Descriptive Data for Implementation Fidelity and Afterschool Program Quality Scales

Afterschool Program Quality Scales	Item N	α	IRR	Mean (SD)					
				Fall (Baseline)			Spring		
				GBG (n=37)	Control (n=36)	Total (N=73)	GBG (n=37)	Control (n=36)	Total (N=73)
Afterschool Climate Assessment (ACA, index of implementation fidelity)	10	0.62	0.77	0.50 (0.16)	0.49(0.16)	0.49(0.16)	0.69(0.17)	0.52(0.11)	0.60(0.16)
Caregiver Interaction Scale (CIS)									
Sensitivity/Detachment	11	0.92	0.77	2.94 (0.40)	2.97 (0.50)	2.95 (0.45)	2.74 (0.52)	2.73 (0.42)	2.74 (0.47)
Harshness	6	0.75	0.56	1.22 (0.15)	1.34 (0.31)	1.28 (0.25)	1.34 (0.30)	1.30 (0.31)	1.32 (0.30)
Permissiveness	3	0.84	0.58	1.89 (0.46)	1.90 (0.52)	1.90 (0.49)	2.15 (0.46)	2.22 (0.50)	2.18 (0.48)
Promising Practices Rating Scale (PPRS)									
Supportive relations with adults	5	0.88	0.59	3.09 (0.35)	3.06 (0.45)	3.08 (0.40)	2.91 (0.47)	2.81 (0.46)	2.86 (0.47)
Supportive relations with peers	3	0.89	0.50	3.20 (0.27)	3.23 (0.37)	3.22 (0.32)	3.13 (0.46)	3.02 (0.35)	3.08 (0.41)
Appropriate structure	4	0.67	0.67	3.27 (0.36)	3.22 (0.33)	3.25 (0.34)	3.21 (0.36)	3.10 (0.38)	3.15 (0.37)
Level of engagement	3	0.84	0.56	3.21 (0.27)	3.20 (0.34)	3.20 (0.30)	3.07 (0.42)	2.92 (0.41)	3.00 (0.42)
Chaos	2	0.81	0.63	1.40 (0.30)	1.53 (0.39)	1.47 (0.35)	1.44 (0.45)	1.55 (0.40)	1.49 (0.43)
Youth Program Quality Assessment (YPQA)									
Active engagement	3	0.75	0.45	2.74 (0.67)	2.87 (0.68)	2.81 (0.68)	2.58 (0.88)	2.41 (0.79)	2.49 (0.84)
Sense of belonging	4	0.55	0.34	3.37 (0.43)	3.41 (0.45)	3.39 (0.43)	3.44 (0.49)	3.20 (0.43)	3.32 (0.47)
Conflict resolution	4	0.89	0.59	2.53 (0.92)	2.66 (0.98)	2.95 (0.95)	2.70 (1.05)	2.39 (1.06)	2.54 (1.06)
Staff engagement	4	0.80	0.61	3.86 (0.74)	3.61 (0.78)	3.74 (0.77)	3.56 (0.72)	3.52 (0.65)	3.54 (0.68)
Responsibility	2	0.80	0.42	3.49 (0.84)	3.32 (0.74)	3.40 (0.79)	3.44 (0.85)	3.27 (0.71)	3.36 (0.78)
Choice	1	--	0.58	3.77 (0.80)	3.70 (0.95)	3.74 (0.87)	3.72 (0.91)	3.63 (0.89)	3.67 (0.90)

Note: Inter-rater reliability (IRR) was assessed with the inter-class correlation coefficient (ICC)

Table 3.

Descriptive Statistics, and Settings-level (experimental versus control group) effects

	N	Min.	Max.	Mean	S. D.	γ	S.E.
ACA-Afterschool Climate Assessment	219	0.1	1.0	0.59	0.21	0.16 ***	0.03
CIS-Sensitivity detachment	223	1.2	4.0	2.72	0.58	0.02	0.09
CIS-Harshness	223	1.0	3.1	1.31	0.37	0.07	0.06
CIS-Permissiveness	223	1.0	3.7	2.19	0.67	-0.12	0.08
PPRS-Supportive relations with adults	223	1.1	4.0	2.87	0.57	0.09	0.09
PPRS-Supportive relations with peers	223	1.3	4.0	3.10	0.54	0.12 ⁺	0.07
PPRS-Appropriate structure	223	1.8	4.0	3.17	0.47	0.08	0.07
PPRS-Level of Engagement	223	1.3	4.0	3.01	0.56	0.14 ⁺	0.08
PPRS-Chaos	223	1.0	3.7	1.50	0.53	-0.06	0.10
YPQA-Active engagement	222	1.0	5.0	2.47	1.13	0.10	0.15
YPQA-Belonging	222	1.5	5.0	3.31	0.64	0.23 *	0.10
YPQA-Conflicts	142	1.0	5.0	2.44	1.23	0.19	0.17
YPQA-Adult engagement	222	1.0	5.0	3.56	0.88	0.02	0.15
YPQA-Responsibility	219	1.0	5.0	3.32	1.07	0.09	0.14
YPQA-Choices	222	1.0	5.0	3.71	1.22	0.14	0.18

⁺
p < .10*
p < .05**
p < .01

Note: Estimates from four level MLwiN models that also control for mean differences among raters and between

Note: The unit of analysis is a single observer's visit (1 or 2 raters) to a program site.

Table 4.

Interactive effects on observational measures of fidelity (ACA) and treatment assignment

Outcome	γ	S.E.
CIS-Sensitivity/detachment	1.16 ⁺	0.67
CIS-Harshness/criticism	-1.11 [*]	0.50
CIS-Permissiveness	-0.19	0.62
PPRS-Supportive relations with adults	1.83 ^{**}	0.67
PPRS-Supportive relations with peers	0.88	0.60
PPRS-Appropriate structure	1.54 ^{**}	0.52
PPRS-Level of Engagement	1.82 ^{**}	0.61
PPRS-Chaos	-1.28	0.79
YPQA-Active engagement	1.95 ⁺	1.09
YPQA-Belonging	1.01	0.74
YPQA-Conflicts	0.66	1.50
YPQA-Adult engagement	1.71	1.17
YPQA-Responsibility	2.15 ⁺	1.16
YPQA-Choices	-1.86	1.46

⁺
p < .10^{*}
p < .05^{**}
p < .01

Note: Estimates from four level MLwiN models that also control for mean differences among raters and between the two post-test assessment periods.

Table 5.

Intent-to-treat and ACA Moderated effects on child outcomes, full MLwiN results.

Explanatory Variable	SDQ-Hyperactivity		SDQ-Emotional Symptoms		SDQ-Prosocial Behaviors		SDQ-Conduct Problems		Problem Behaviors and Substance Use (PBSU)	
	<i>g</i>	S.E.	<i>g</i>	S.E.	<i>g</i>	S.E.	<i>g</i>	S.E.	<i>g</i>	S.E.
Constant	1.36 ***	0.07	1.55 ***	0.06	2.57 ***	0.05	1.30 ***	0.06	0.08 **	0.03
Gender (1 = male)	0.09 †	0.05	-0.07 †	0.04	-0.17 ***	0.04	0.13 **	0.04	0.03 †	0.02
Black	0.19 **	0.06	0.06	0.05	0.03	0.05	0.08	0.05	0.08 **	0.02
Hispanic	0.26 **	0.10	0.05	0.08	-0.01	0.07	0.18 *	0.07	0.05	0.04
Other	0.09	0.06	0.10 †	0.06	0.02	0.05	-0.01	0.05	0.02	0.03
3rd grade	0.02	0.06	0.01	0.06	-0.07	0.05	0.09 †	0.05	-0.01	0.02
4th grade	-0.04	0.06	-0.05	0.06	-0.03	0.05	0.04	0.05	-0.05 *	0.03
5th grade	-0.18 *	0.07	-0.10	0.07	-0.01	0.06	-0.03	0.06	-0.06 †	0.03
Program Cohort 1	0.09	0.06	0.01	0.05	0.07	0.04	-0.05	0.06	0.02	0.03
Program Cohort 2	0.13 †	0.07	0.04	0.05	0.05	0.05	0.06 †	0.06	0.02	0.03
Pretest outcome measure	0.39 ***	0.05	0.44 ***	0.05	0.39 ***	0.04	0.42 ***	0.04	0.42 ***	0.04
Treatment Versus Control	-0.03	0.05	-0.01	0.04	0.08 *	0.04	-0.06	0.05	-0.01	0.02
<i>ACA moderated</i>	-0.76 *	0.37	.10	.32	.31	.27	.34	.35	-0.02	.15
<i>Residual Variance Components</i>	<i>t</i>	S.E.	<i>t</i>	S.E.	<i>t</i>	S.E.	<i>t</i>	S.E.	<i>t</i>	S.E.
Random Assignment Pair	0.003	0.007	0.000	0.000	0.000	0.000	0.004	0.006	0.000	0.000
Afterschool program site	0.003	0.009	0.000	0.000	0.000	0.003	0.009	0.008	0.001	0.001
Individual child	0.219	0.016	0.183	0.012	0.138	0.010	0.130	0.010	0.033	0.002
N Site assignment pairs	35		35		35		35		35	
N Program sites	71		71		71		70		71	

† *p* < .10
 * *p* < .05
 ** *p* < .01

100% < d

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 6.

Interactions of treatment assignment with ACA in effects on child outcomes.

Outcome	ACA	
	<i>g</i>	S.E.
SDQ-Hyperactivity	-0.76 *	0.37
SDQ-Emotional Symptoms	0.10	0.32
SDQ-Prosocial Behaviors	0.31	0.27
SDQ-Conduct Problems	0.34	0.35
Problem Behaviors and Substance Use (PBSU)	-0.02	0.15

* $p < .05$

Note. Estimates from three level MLwiN models that also control for all variables in Table 5: gender, race/ethnicity, grade, program cohort, and the pretest measure of the outcome variable.