Routledge
Taylor & Francis Group

Journal of Research on Educational Effectiveness

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uree20

Evaluation of a Social Skills Program for Early Elementary Students: *We Have Skills*

Keith Smolkowski, Hill Walker, Brion Marquez, Derek Kosty, Claudia Vincent, Carey Black, Gulcan Cil & Lisa A. Strycker

To cite this article: Keith Smolkowski, Hill Walker, Brion Marquez, Derek Kosty, Claudia Vincent, Carey Black, Gulcan Cil & Lisa A. Strycker (2022): Evaluation of a Social Skills Program for Early Elementary Students: *We Have Skills*, Journal of Research on Educational Effectiveness, DOI: 10.1080/19345747.2022.2037798

To link to this article: https://doi.org/10.1080/19345747.2022.2037798

Published online: 14 Mar 2022.

JOURNAL OF RESEARCH ON EDUCATIONAL EFFECTIVENESS
https://doi.org/10.1080/19345747.2022.2037798

Routledge
Taylor & Francis Group

INTERVENTION, EVALUATION, AND POLICY STUDIES

Evaluation of a Social Skills Program for Early Elementary Students: *We Have Skills*

Keith Smolkowski[a], Hill Walker[b], Brion Marquez[c], Derek Kosty[a], Claudia Vincent[b], Carey Black[a], Gulcan Cil[a] and Lisa A. Strycker[a]

[a]Oregon Research Institute, Eugene, OR, USA; [b]University of Oregon, Eugene, OR, USA; [c]IRIS Educational Media, Eugene, OR, USA

ABSTRACT

This paper reports the results of a randomized controlled trial evaluating *We Have Skills,* a brief curriculum designed to teach early elementary students academic and social skills as well as improve teacher efficacy in classroom management. Intervention efficacy was tested with 127 teachers, randomly assigned to condition, and 2,817 of their students. On key outcomes of student academically related behavioral skills and classroom adjustment, intervention teachers reported greater gains for their students than comparison teachers (Hedges's $g \geq 0.19$). Baseline scores moderated effects, demonstrating greater differences between conditions for initially struggling students. Intervention teachers also reported greater improvements on their sense of self-efficacy for classroom management and concerns about student behavior ($|g| \geq 0.30$) relative to comparison teachers. Effect sizes were similar to or greater than those reported for similar programs, and an economic analysis suggested that *We Have Skills* was less costly than many. The findings support cost-effective solutions that teach social–behavioral skills in early elementary grades—and suggest that such programs may be especially beneficial for students who struggle with academically related behaviors.

ARTICLE HISTORY
Received 17 January 2021
Revised 10 August 2021
Accepted 10 December 2021

KEYWORDS
Academic enablers; cluster-randomized trial; cost-effectiveness; noncognitive skills; social skills; student behavior

Over the past two decades, challenging behaviors displayed by very young children have increased in early education settings. It is not uncommon for even preschoolers to be suspended or expelled (Zeng et al., 2019). Room clears—removing all students from classrooms to reestablish order and routines after episodes of extreme acting-out behavior by a student—have emerged as a standard routine in many elementary classrooms (Navarrete, 2019). Disruptive behavior patterns exhibited by young children often persist throughout their school years, placing them at serious long-term risk for destructive outcomes, such as delinquency, school failure and dropout, violence, and drug and alcohol abuse (Hawkins et al., 1999; Patterson et al., 1992; Walker et al., 1995). Early intervention approaches that target both social–emotional and educational dimensions have shown promise in diverting students away from a negative developmental path (Dunlap

CONTACT Keith Smolkowski keiths@ori.org Oregon Research Institute, 1776 Millrace Drive, Eugene, OR 97403, USA.

& Fox, 2014; Webster-Stratton et al., 2004). The sooner systematic intervention occurs, the greater the likelihood that positive outcomes can be achieved.

On a broad scale, educators have turned to social skills training to address this challenge. This strategy allows school personnel to address, positively, the many behavioral challenges that students may bring with them to school. But existing social skills interventions have not worked well for students with well-established, disruptive behavior patterns (Gresham, 1998, 2002).

Efficacy of Social Skills Training

Social skills have been defined as actions for engaging in positive behaviors such as joining a group, having conversations, and extending invitations (Gresham, 1986; Gresham & Elliott, 2014), with training typically focused on relationships with peers and adults. Ideally, social skills programs recognize and remediate both skill acquisition deficits (skill not mastered) and skill performance deficits (mastered skill not displayed). Over the past several decades, educators have embraced social–emotional learning (SEL) interventions, but with little effect at times due to poor implementation fidelity, poor teaching methods, instructional content unrelated to key outcomes, and addressing disruptive behavior problems which are highly resistant to change (Gresham, 1998, 2002). Due possibly to implementation challenges (Low et al., 2016), a recent cluster-randomized trial of the widely disseminated SEL program *Second Step* in early elementary (Committee for Children, www.cfchildren.org) yielded small effect sizes, Hedges's $|g| = 0.02$ to 0.13, for social-emotional skills (Low et al., 2015, 2019), yet moderation analyses indicated stronger benefits for students with initially lower scores. Meta-analyses of universal SEL programs have generally documented benefits on behavioral outcomes, with effect sizes from 0.13 to 0.24 (Durlak et al., 2011; Taylor et al., 2017; excludes academic performance and skills tests not measured in the present study). Hence, school professionals and researchers continue to turn to social skills training as a preferred response to behavioral challenges.

Promising Approaches in Social Skills Training

Gaps in the social skills knowledge base provide a roadmap for innovation. Below, we summarize needs for improved methods and outcomes, as well as newer developments with potential to meet those needs.

There is a critical need for efficacious interventions targeting the conduct problems and academic deficits that hinder school success for many students. Curricula to help students negotiate teacher- and peer-related relationships are required to address this need in school settings (Walker et al., 1995). With this approach, students gain skill sets to meet teacher expectations in behavioral and academic areas, and to make friends and build supportive peer networks. For students, forging positive relationships with teachers and peers at school is a key social task. Existing school-based social skills training efforts, however, have focused largely on social–emotional dimensions, ignoring nonacademic behaviors or noncognitive factors that can greatly facilitate academic achievement (Farrington et al., 2012).

A major conceptual advance has recently broadened the relevance of social skills to school achievement and success (DiPerna, 2006). The social skills rubric now covers not only relationships but also social competence, which is fundamental to academic learning (Suldo et al., 2014). Academically related social skills can be framed as *academic enablers*—nonacademic behaviors that enable academic achievement (DiPerna & Elliott, 2002)—which include listening to teacher instructions, following directions, doing assignments, following rules, and asking clarification questions. Academic enablers emerged from research linking students' social skills and motivation to their academic achievement (Wentzel, 2009). This more comprehensive definition implies that systematic social skills training can enhance school success.

Along with refining the content of social skills programs and broadening the definition of social skills, developers have made progress in creating new models for delivery. Nelson et al. (2009) applied a multitiered intervention to 407 K-3 students: the *Behavior and Academic Support and Enhancement (BASE) Program* at the universal level (Nelson, 1996), the *First Step to Success* early intervention as a selected intervention (Walker et al., 1997), and the *Multisystemic Therapy* treatment at the indicated level (Henggeler et al., 1998). Longitudinal results showed that the universal intervention prevented problems from emerging, and the selected intervention yielded sustained social skills improvements and problem behavior reductions; the indicated intervention did not lead to significant gains. Evans and colleagues (2014) compared the implications of two behavioral support approaches for delivering mental health services in school settings: the *Schoolwide Positive Behavior Interventions and Supports* (SWPBIS; www.pbis.org) and the *Life Course Model* of children's behavioral health care. The authors concluded that both models were suitable for delivering complex interventions for the 80% of students needing mental health care and services; their work provides important guideposts for the efficient delivery of social–emotional skills content in schools.

Video technology and computer graphics offer promising avenues for enhancing social skills training programs, but have had a relatively limited role to date. When media resources are combined with effective instruction to define, model positive, and critique negative social skills, the result can be quite powerful for teachers and students. Video-based behavior modeling can improve skill development and generalization across a wide range of individuals, from social behavior modeled by peers for students with autism (e.g., Jones et al., 2014) to workplace professional development (Taylor et al., 2005). Well-produced videos help students easily discriminate negative from positive examples (Carnine, 1980; Taylor et al., 2005) and allow for consistency in instruction across classrooms. Critical features of the target skill can be highlighted and discussed as part of the instruction. Short, memorable songs can also improve memory (Batchelor & Bintz, 2012; Rainey & Larsen, 2002). Adapting technology in this manner, while embracing evidence-based principles of multimedia instruction (Mayer, 2008), can make teaching and learning social skills more interesting and reduce teacher burden.

Areas for Further Research and Development

Logistical problems and obstacles remain in the development and implementation of the innovations described above. Perhaps the most serious is lack of acceptance and use by

end users. Hoagwood et al. (2007, 2012) argued that the main reason schools do not implement effective interventions is that the programs do not fit their normal routines. Low et al. (2016), for example, found that nearly a quarter of teachers struggled to implement *Second Step*, which requires a 25- to 45-min block of intervention time to deliver scripted lessons. To improve program efficacy, social skills programs must be adaptable to a variety of school regimens, operations, and routines. Teachers must also adopt, implement, and maintain social skills instruction. Enhancing implementation requires approaches to increase teachers' willingness to invest their time and effort or programs that teachers can more easily implement.

Motivational interviewing (MI) strategies can enhance the acceptance and treatment integrity of school-based social skills training. Based on the work of Miller and Rollnick (2012), MI relies on the precept that a person's social interactions have significant and positive effects on intrinsic motivation, leading to better behavioral outcomes. With MI, individuals (including parents and teachers) are counseled to accept and commit to change. The use of MI in social skills instruction offers a promising new avenue for boosting treatment outcomes and social validity but has not been empirically tested. An MI session with a behavioral coach at the start of social skills training could have a positive impact, as studies show that people exposed to MI adhere more closely to treatment protocols and have better outcomes (Saunders et al., 1993).

Alternatively, the programs that reduce teacher training and preparation time, ease instruction, end increase acceptance by teachers and students—programs that reduce obstacles—may also enhance fidelity of program implementation. Oakes and colleagues (2018) showed that an integrated, flexible intervention delivery approach can improve end user acceptance and outcomes. Brief universal interventions may reduce the social and academic behavioral problems that are easiest to change before delivering more elaborate strategies to deal with persistent challenges. Models are also needed that coordinate intervention and measurement components so schools can easily adopt and adapt instruction—especially in elementary grades. Programs that offer practical, efficient implementation are more likely to be sustained in the classroom. The present study focuses on a program called *We Have Skills* (*WHS*) that took this second approach.

The We Have Skills Program

WHS aims to maximize school success, incorporating best practices informed by research on general education teachers' expectations for student classroom behavior (Marquez et al., 2014; Walker, 1986; Walker & Rankin, 1983). In the context of multitiered support systems, *WHS* provides universal instruction designed to improve student academic and social skills as well as increase teacher self-efficacy in managing the classroom. The intervention targets early elementary teachers and their students and comprises three interrelated elements: (a) a curriculum with eight lessons and classroom materials, (b) multimedia classroom instruction assets, and (c) an assessment instrument to screen student performance on the skills taught by *WHS*. The curriculum lays out each day's lesson with specific activities and recommendations for the teacher (e.g., "Give your students **immediate** feedback that is **descriptive** to let them know what they are doing well and what they need to do to improve," p. 21, emphasis in original). The

overarching goal is to provide effective classroom social–behavioral lessons and curriculum-based assessment to improve student outcomes and teacher self-efficacy.

Curriculum and Classroom Materials

Over eight lessons, the curriculum provides an overview for students and then teaches seven core skills: (a) listen, (b) ask for help, (c) follow directions, (d) do the best you can, (e) follow the rules, (f) work out strong feelings, and (g) get along. Each skill is taught in small daily doses across a week, minimizing time taken away from core curricula. Instruction in each skill includes several short activities, which represent the core components of *WHS*: skill introduction, video presentation (4–6 min.), brief discussion, a song associated with the skill (60 90 s), student practice initiated by teachers, and prompts to use their skills throughout each day. *WHS* overlays the instructional content onto established routines of general education classrooms. Teachers encourage students to practice their skills and praise or correct students' use of skills throughout the school day. Teachers may also use additional materials, such as skill practice booklets, skill posters, coloring pages, skill cards that define skills, certificates of mastery, and notices for parents. *WHS* was designed to require about 1 h per week.

The lesson plan overview provides brief suggestions for students who have difficulty mastering certain skills after completing the eight-lesson program. Labeled "Tier 2 Instruction," the activities would not suffice as a Tier 2 intervention in typical multitiered support systems. Rather, the suggestions represent the types of supports that teachers may offer students with mild behavior challenges, such as extra practice opportunities, increased descriptive praise of efforts to use the skills in the program, and precorrections, delivered individually or in small groups. Students may also rewatch the videos. Although the nature of these supports is limited and likely interpreted by teachers as optional, they may help students who struggle to learn the skills.

Multimedia Classroom Instruction

The first lesson for each skill includes an instructional video. In the first week, the video defines classroom behavioral skills and explains how students can learn through practice. Each subsequent week's video shows students the correct applications of a skill in varying classroom contexts and demonstrates to teachers how to reinforce the skills. Within each video, Mr. Lopez and his students discuss the new skill and interact with a "magic" chalkboard that presents graphics and realistic vignettes of students modeling examples and nonexamples of the skill. Mr. Lopez also introduces an animated animal friend, one for each skill, who sings a catchy song about the skill. Video lessons are carefully sequenced and feature racially and ethnically diverse child and adult actors, and each skill's song uses a different musical genre to engage all students. Marquez and colleagues (2014) reported that "child actors, realistic vignettes, animal characters, and memorable songs successfully appealed to elementary students" (p. 151) and that students frequently asked teachers to start each day with one or more of the 30-s songs.

Screening Tool

WHS includes access to the Elementary Social Behavior Assessment (ESBA; Marquez et al., 2014; Pennefather & Smolkowski, 2015) in an online tool to help teachers assess and track the skill levels of individual students. The instrument allows teachers to rate and screen whole classrooms in about 20–30 min, using a color-coded rating system (green = skill mastered, yellow = needs improvement, and red = cause for concern) to assess each student's status on 12 items. The assessment can help teachers evaluate the entire classroom, identify individual students who struggle on one or more skills, or determine if multiple students have yet to master the one of the skills. Screening is not stipulated by the curriculum (lesson plans), but trainers encourage teachers to administer the ESBA three times a year to gauge classroom progress.

Professional Development

During field tests and pilot evaluation, teachers implemented the universal elements of *WHS* with little difficulty and did not require extensive training (Marquez et al., 2014). The *WHS* professional development (PD) presents a description of *WHS* and its lesson plans. Like *WHS* itself, the PD emphasizes that social–behavioral skills (a) are not inherent features of character, (b) can be taught with evidence-based approaches, (c) improve through practice, and (d) should be regularly assessed to identify students who need to improve. The presentation previews the skills videos and other *WHS* materials, offers a brief summary of research support for *WHS* activities and materials, and suggests strategies to offer regular skill rehearsal with descriptive performance feedback. The PD also introduces teachers to universal screening with the ESBA and explains how to use the data for decision-making. The PD takes less than 3 h.

Development of WHS and Initial Evaluation Results

WHS was developed from evidence-based principles of explicit or direct instruction (e.g., Boxer, 2019; Engelmann & Carnine, 1991), including (a) effective sequencing, (b) appropriate pacing, (c) modeling with positive and negative examples, (d) opportunities for students to respond or practice, and (e) descriptive praise for correct behavior, and (f) descriptive corrections when students make mistakes. The multimedia assets, including the videos, songs, posters, and student print materials, were designed to incorporate evidence-based principles of multimedia instruction (Mayer, 2008). For instance, the videos signal key principles, segment instruction, provide demonstrations with narration, and use a simple, conversational style while minimizing extraneous material and reducing unnecessary redundancy.

In an initial study of *WHS*, Marquez et al. (2014) randomly assigned 70 classrooms to intervention (37) or control (33) conditions. Compared to control teachers, intervention teachers rated students higher on the ESBA (Hedges's $g = 0.27$) and reported higher self-efficacy for classroom management on the Teacher Sense of Efficacy Scale (described below; Cohen's $d = 0.79$) at posttest. The analyses included baseline measures as covariates, and the student-level analyses relied on a multilevel model to account for the assignment of intact classrooms to condition. Teachers also perceived *WHS* as easy to implement and highly recommended its use.

In the initial evaluation of *WHS*, teachers received only 8 weeks to implement the program during the fall of the school year. Given typical interruptions, such as holidays and in-service days, it is likely that not all teachers completed all eight lessons. Most teachers reported that they spent less than 8 h on *WHS* delivery during the 8 weeks.

The Present Study

This paper reports the results of an evaluation of *WHS*, which aimed to promote the academically related social–behavioral success of early elementary students. This *efficacy in real-world settings* "focuses on implementation by indigenous providers in school settings" (Smolkowski et al., 2019, p. 197) and falls near the center of the continuum from basic-science to scale up research. The study was designed to compare classrooms where teachers taught *WHS* to classrooms in which teachers used their usual practices to improve academically related behavioral skills. It extends the practical and efficient implementation strategies in the promising areas of research and practice. It also builds on the initial evaluation of *WHS* (Marquez et al., 2014) by recruiting a larger sample, adding measures of relevant student and teacher behavior, and allowing teachers the time and flexibility to teach *WHS* at their own pace and support students may not have initially mastered all skills. The investigation addresses three hypotheses:

1. Relative to students in control classrooms, students in classrooms that use *WHS* will demonstrate improved academically related behavior and classroom adjustment.
2. As some students arrive at school versed in behavioral skills, baseline levels of student behavior will moderate treatment-group differences in student outcomes.
3. Teachers implementing *WHS* will increase their descriptive praise and corrections of skills taught by *WHS*, report fewer concerns about student behavior, and report greater self-efficacy for classroom management than control teachers.

This study also evaluates implementation costs and estimates cost-effectiveness ratios.

Method

This study took place from 2015 to 2019 with early elementary school teachers and their students in 60 schools from multiple states (AZ, CA, CO, NV, OH, OR, NM, UT, WA). Investigators randomly assigned teachers to either an immediate-intervention condition or a wait-list, business-as-usual control condition in which no additional intervention or services were offered. Random assignment occurred within each school, when possible, district, state, and wave (cohort) and to experimentally control for extraneous influences, such as funding levels, proportion of students who receive free or reduced-price lunch, and similar variables.

Sample

The study included 127 teachers and 2,817 students within 60 schools: 66 teachers and 1,515 students in the *WHS* condition (23.0 students/teacher), plus 61 teachers and 1,302 students in

Table 1. Teacher and student characteristics by intervention condition.

	Intervention		Control	
	M or %	*SD*	*M* or %	*SD*
Teachers				
Age	40.3	8.7	42.8	10.5
% Female	100%		97%	
% Hispanic	9%		7%	
Race				
% White	86%		87%	
% African–American	2%		0%	
% Asian	6%		5%	
% Pacific Islander or Native Hawaiian	0%		2%	
% More than one race	3%		3%	
% Other	2%		2%	
% Prefer not to answer	2%		2%	
Experience (years)	13.0	9.0	13.0	10.0
Number of students in classroom	23.6	3.5	22.4	4.0
% Taught first grade only[1]	61%		62%	
Educational background				
% Bachelor's degree	41%		38%	
% Master's degree	52%		52%	
% Doctoral or professional degree	2%		0%	
% Other	6%		10%	
Students				
% Female	49%		49%	
% Hispanic	24%		25%	
Race				
% White	59%		56%	
% African–American	6%		5%	
% Asian	13%		14%	
% American Indian or Alaskan Native	2%		1%	
% Pacific Islander or Native Hawaiian	2%		1%	
% More than one race	3%		3%	
% Other	17%		19%	
% Receiving ESL services	17%		19%	
% With IEP	8%		9%	

Note. The sample consisted of 127 teachers (66 intervention, 61 control) and 2,817 students (1,515 intervention, 1,302 control).
[1]Although most teachers taught first grade, some taught kindergarten or mixed-grade classrooms with students in both kindergarten and first grade.

the control condition (21.3 students/teacher). Twenty-seven schools included only one teacher, and two or more teachers participated in the remaining 33 schools. We invited first-grade teachers to participate but about 39% taught kindergarten or kindergarten and Grade 1. Table 1 reports teacher and student characteristics by treatment condition. Teachers were similar across conditions in age, experience, education, and other characteristics. Students were similar across conditions in demographics and receipt of services (e.g., special education).

Procedures

The study used identical recruitment, implementation, and assessment procedures across four waves. All procedures were consistent across time and with school and district policies.

Recruitment

Project staff members recruited first-grade teachers via email, flyers, word of mouth, a product website, and industry conferences, approaches not dissimilar from those used by developers to market similar programs. Interested teachers received a description of study requirements and signed a letter of intent to participate. For teachers who enrolled, the study included all their students whose parents did not decline their child's participation—a passive consent procedure. After teachers completed assessments, those in the *WHS* condition began to implement the program. Control teachers were offered *WHS* for the following school year.

Comparison Classrooms

Teachers assigned to the control condition were asked to teach students as they normally would—business as usual. Comparison teachers reported using a range of programs that targeted student behavior in their classrooms: 29 reported positive behavior interventions and supports (PBIS; www.pbis.org); 9 reported *Second Step*; and 3 or fewer reported *CHAMPS* (Sprick, 2009), the Check-in/Check-out program (Crone et al., 2010), *Love & Logic* (Funk & Fay, 1995), *MindUP* (mindup.org), *Steps to Respect* (Brown et al., 2011), the *Good Behavior Game* (Dolan et al., 1989), or *Tools for Getting Along* (Smith et al., 2016). All teachers in the control condition except one received access to *WHS* in their second year of participation as described next. (The project lost contact with one teacher due to adversity within her school unrelated to this study.)

Intervention Classrooms

Teachers in the intervention condition received the WHS curriculum, access to an online screening system, a brief initial training, and coaching at no cost.

WHS Professional Development. The project provided a short, initial training for teachers in the *WHS* condition. Trainers previewed *WHS*, summarized its research support, introduced principles of effective delivery, described program details, and explained how to make data-driven decisions based on the ESBA. Training required about 2 h, including teacher introductions and follow-up questions. Project staff initially trained teacher in person. By the second year, teachers requested online training to simplify scheduling and to reduce the time commitment, disruption during the school day, and travel. After initial training, project staff members offered in-person or online coaching to address questions or concerns and to support maintenance of the program. After 2 years, only one teacher requested coaching, which consisted of a brief conversation, so we discontinued coaching thereafter. Teachers frequently reported that *WHS* was easy to implement in their classrooms and that additional supports were unnecessary.

WHS Classroom Implementation. Staff members asked teachers to complete the introduction and seven skills lessons at a rate of about one lesson per week. The pace varied due to school- or district-initiated breaks, teacher in-service days, or other interruptions. Teachers generally completed the eight lessons within 12 weeks. Teachers were asked to provide students with opportunities for independent practice

in the classroom and to assign home activities to support class instruction. Teachers received the ESBA via an online system to screen students before and after implementation to (a) assess student progress and (b) identify the students or skills that required additional support. For students who struggled to learn the skills, teachers could provide more instruction and practice or additional review of videos and other materials.

Other Programs. *WHS* teachers reported using other behavior or social skills programs at a frequency similar to control teachers: 29 reported PBIS; 11 reported *Second Step*; and 3 or fewer reported *CHAMPS*, the Check-in/Check-out program, *Love & Logic*, *MindUP*, *Steps to Respect*, the *Good Behavior Game*, or *Tools for Getting Along*. Teachers in the *WHS* condition also reported using Safe & Civil Schools' *Foundations* (Sprick et al., 2014), PATHS (Greenberg et al., 1995), and *Tribes Learning Communities* (tribes.com); one teacher used each program.

Measures

Teachers completed surveys in the fall, between October and December, and again in the spring, in May and June. They reported their own demographic data, self-efficacy for classroom management, and concerns about student behavior management. They also reported on students' background characteristics, skills taught by *WHS*, and classroom adjustment. During the course of *WHS* implementation, teachers completed *WHS* implementation fidelity checklists. Twice per year, after teachers began teaching *WHS*, staff members observed teachers' use of praise and corrections and reported impressions of student behavior in the classroom.

Elementary Social Behavior Assessment

Teachers used the ESBA for two purposes. All teachers completed the measure for their students for this evaluation. Teachers in the intervention condition were also encouraged but not required to use the ESBA to screen students as part of the *WHS* intervention. Intervention and control teachers were exposed to the ESBA at the same time.

The ESBA was developed for teachers to assess the student skills targeted by the *WHS* intervention (Walker et al., 2015) and was derived from the research of Walker and colleagues (Hersh & Walker, 1983; Walker, 1986). The ESBA has been validated in English (Pennefather & Smolkowski, 2015) and with a version adapted for Norwegian (Arnesen et al., 2018). Teachers rate students on an intuitive 3-point scale corresponding to the current degree of student mastery consistent with PBIS or other multitiered systems: *skill mastered* (3), *needs improvement* (2), and *cause for concern* (1). Example items include "Listens to and respects the teacher" and "Follows the teacher's directions." Short definitions operationalize each behavior along its critical dimensions and features. The definition of the "Listens" item states that the student "faces you while you are talking," "keeps eyes and ears on you," and "is attentive and waits turn to talk."

In the current study, the average of the 12 items comprised a reliable summary score ($\alpha = .94$ at baseline) and demonstrated stability over time within the comparison-group sample ($r = .74$). In previous research, the ESBA conformed to a single factor (Pennefather &

Smolkowski, 2015), demonstrated strong reliability ($\alpha = .94$–$.95$), correlated highly with multiple criterion measures ($r > .77$; Arnesen et al., 2018), and was sensitive to change (Marquez et al., 2014).

Walker–McConnell Scale of Social Competence

As the ESBA focused on skills specifically targeted by *WHS*, we collected a similar but independently developed measure of classroom adjustment. The Walker–McConnell Scale of Social Competence and School Adjustment–Elementary Version (WM; Walker & McConnell, 1995) asks teachers to report on 43 positively worded student behaviors that support successful school social competence. To reduce assessment burden, we collected the 19-item Classroom Adjustment Behaviors subscale as our primary measure of *WHS* impact and cost-effectiveness analysis. The subscale included key components of academic competence (i.e., listening skills, participation, responsiveness, and quality of work) with items such as "Other children seek child out to involve her/him in activities" and "Has good work habits." Teachers rated items on a scale from *never* (1) to *sometimes* (3) to *frequently* (5). The WM has demonstrated adequate score reliability ($\alpha \geq .89$) for subscales, 5-week test-retest reliability ($r > .76$), and concurrent validity with the Social Skills Rating System (Gresham & Elliott, 1990). The present sample produced excellent score reliability ($\alpha = .96$ at pretest) and stability over time ($r = .73$ in control sample).

Teacher Sense of Efficacy Scale

Teachers reported their self-efficacy for classroom management and instruction with the short form of the Teacher Sense of Efficacy Scale (TSES; Tschannen-Moran & Hoy, 2001). On a scale from *nothing* (1) to *a great deal* (9), teachers responded to 12 items, such as "How much can you do to control disruptive behavior in the classroom?" and "How much can you do to motivate students who show low interest in school work?" This instrument has shown strong reliability and construct validity (Tschannen-Moran & Hoy, 2001), and has been related to a variety of outcomes, including student achievement (Moore & Esselman, 1992; Ross, 1992), teacher planning and organization (Allinder, 1995), inclination to refer students to special education (Soodak & Podell, 1993), and commitment to teaching (Trentham et al., 1985). A summary score was computed for this study as the mean of the 12 items ($\alpha = .90$ at pretest).

Teacher Concerns Inventory

Teachers reported concerns about student behavior with the Discipline and Motivation subscale of the modified Teacher Concerns Inventory (TCI; Fimian, 1984). The subscale consists of six items rated on a scale ranging from *no strength; not noticeable* (1) to *major strength; very noticeable* (5) with items such as "I feel frustrated because of discipline problems in my classroom" and "I feel frustrated having to monitor pupil behavior." In previous research, the TCI demonstrated strong reliability ($\alpha \geq .75$) and good construct and content validity (Fimian, 1984). In this study, ratings were averaged to create a summary score ($\alpha = .80$ at pretest).

Implementation Fidelity

The *WHS* Implementation Fidelity Checklist asked teachers to document their completion of *WHS* lessons at the end of each lesson. The checklist for Lesson 1 asked four questions that did not align with later lessons, so we focused on checklists for Lessons 2–8, which asked 16 questions about whether teachers completed instruction, practice, feedback, and problem solving activities. Three instruction items asked about use of classroom discussion, the video, and skill cards. A practice item asked if teachers practiced the skill 3–5 times per day. Six feedback items tallied use of descriptive praise, descriptive correction, skills tickets, tally sheets, happy notices, and certificates of mastery. Teachers also reported use of problem-solving discussion, role-play, songs, games, coloring pages, and skills booklets. We summarized the fidelity data as the percent of teachers who reported completing each item, averaged across Lessons 2–8.

Direct Observations

Independent, in-class observations offer a valuable perspective on the behavior of study participants (Snyder et al., 2006). Independent observers assessed each teacher's behavior after they began using *WHS*. The lessons reminded teachers to use descriptive praise to help students know what they did well and descriptive corrections to let students know which skills they needed to improve. We framed these two key styles of feedback with nondescriptive praise and corrections as well as directions but focused on descriptive praise and corrections. Descriptive praise included statements such as "Thank you for raising your hand" or "Great job listening when others are talking." Nondescriptive praise might consist of, simply, "Great job" or "Way to go." Corrections were similar but reminded students about the behavior they needed to improve, such as "Please remember to keep your hands and feet to yourselves" (descriptive correction).

Observers, trained in stages (i.e., Smolkowski & Gunn, 2012), received an overview of the system, an explanation of the codes, and procedures for using the observation codebook as a reference. The trainer and the observers practiced coding video clips and debriefing observations as a group and then practiced in nonstudy classrooms with the trainer. The trainer and observers established an agreement rate of 80% or higher prior to observing project classrooms. The trainer periodically retested them to maintain interobserver agreement. Whenever possible, observers coded classrooms for 60 min in five 10-min intervals with 2-min breaks. Because some observation periods were cut short by interruptions, we summarized praise or corrections as rates per minute of observation time. Observations were collected for 114 of the 127 classrooms: 58 in the *WHS* condition and 56 from the control condition. Some classrooms could not be observed due to scheduling, travel, or other limitations.

We assessed interrater reliability with intraclass correlations: the proportion of variance between versus within paired observations for each measure. The 42 paired observations produced intraclass correlations of .88 for descriptive praise and .55 for descriptive corrections, representing nearly perfect and moderate reliability, respectively (Landis & Koch, 1977). Nondescriptive praise and feedback produced intraclass correlations of .84 and .43, respectively.

After each observation session, observers rated their impressions of student behavior from *rarely* (1) to *always* (6) on seven behaviors: (a) listens to and respects the teacher, (b) follows teacher's directions, (c) works with effort, (d) does seat assignments as

directed, (e) makes assistance needs known appropriately, (f) follows rules, and (g) gets along with peers. An intraclass correlations of .77 indicated substantial interobserver reliability (Landis & Koch, 1977). Observer impressions of student behavior were available from 107 classrooms.

Statistical Analysis

We tested condition differences on change in student outcomes with a multilevel Time × Condition model (Murray, 1998) represented by following composite equation:

$$Y_{tjk} = (\gamma_{000} + \gamma_{001}C_k + \gamma_{100}T_{tjk} + \gamma_{101}T_{tjk}C_k) + (\boldsymbol{u_{00k}} + \boldsymbol{u_{10k}}T_{tjk} + \boldsymbol{r_{0jk}} + \boldsymbol{e_{tjk}}).$$

Y_{tjk} represented a score for assessment occasion t on student j in classroom k. The model included three predictors: time, T_{tjk} (coded 0 at pretest, 1 at posttest); condition, C_k (coded 0 for control, 1 for intervention); and their interaction. The model produced estimates of the pretest intercept for the control condition, γ_{000}, the difference between conditions at pretest, γ_{001}, the change over time for the control condition, γ_{100}, and the difference in change between conditions, γ_{101}. The latter parameter estimates intervention efficacy. The model included residual terms for the classroom-level intercept, $\boldsymbol{u_{00k}}$; classroom-level change, $\boldsymbol{u_{10k}}T_{tjk}$; student-level intercept, $\boldsymbol{r_{0jk}}$; student-level change, $\boldsymbol{r_{1jk}}T_{tjk}$; and individual observations, $\boldsymbol{e_{tjk}}$. In this model, with just two time points, $\boldsymbol{e_{tjk}} = 0$. The student-level intercept, $\boldsymbol{r_{0jk}}$, was also equivalent to the within-student covariation between pretest and posttest assessments (Murray, 1998).

We examined whether baseline student behavior moderated group differences in change in teacher-reported student outcomes. This analysis expanded the statistical model to include the moderator as well as its interaction with Condition, Time, and Time × Condition terms. The three-way interaction estimated whether condition effects varied by pretest level. We relied on Jaccard and Turrisi (2003) and Preacher et al. (2006) for interpretation.

We analyzed teacher self-reports with Time × Condition models that excluded classroom-level variances. Observations were averaged for each classroom, and condition differences were tested with an analysis of covariance that included pretest WM as a covariate.

Model Estimation

Models specified with SAS PROC MIXED version 9.2 (SAS Institute, 2016) used full-information maximum likelihood (ML) methods. ML estimation uses all available data, reducing potential bias—even in the face of substantial attrition—provided data are missing at random (Graham, 2009). Compared to complete-case analyses, ML relies on relatively benign assumptions and does not introduce bias (Allison, 2009; Collins et al., 2001). Teacher and student dropout were very low, <4% and 4.4%, respectively, due primarily to school transfers.

Interpretation of Results

To interpret results, we focus on Hedges's g effect sizes, their 95% confidence intervals (CI), and model probabilities for hypothesis tests. As recommended by the American Statistical Association (Wasserstein & Lazar, 2016), we abstained from using bright-line rules such as

claims of "statistical significance" when $p < .05$. p Values measure the incompatibility between the observed data and all assumptions of the statistical model, including the null hypothesis, H_0 (Greenland et al., 2016). This awkward definition determines neither which assumptions are incorrect nor the importance of the association. To complement p values, we report effect sizes, g, and *model probabilities*, w. The model probabilities indicate the strength of evidence for one model when compared with others, given the data at hand. Based on the Akaike Information Criterion, Burnham et al. (2011) describe w as the probability of selecting the same model with a "replicate data set from the same system" (p. 30) and "allow statements such as "the probability of [H_A] is 0.78" (p. 26). Model probabilities better characterize the chance of a replicated result than p values. In this study, we compared models for two hypotheses: a model with the intervention effect (H_A) and one without (H_0). We reported the model probability for the model with the condition effect (H_A), and with only two models, the model probability for H_0 is $1 - w$.

Cost Analysis

Estimates of *WHS* costs relied on the ingredients method (Levin et al., 2018). Cost estimates considered all resources required to implement *WHS*, including new expenditures and opportunity costs associated with reallocation of existing resources, and used a societal

Table 2. Ingredients and costs for *We Have Skills* by stage of implementation.

Stage	Ingredients	Quantity	Unit Costs	Costs†
Curriculum	Curriculum* (videos, skill tickets, 166-page teacher's guide, skill cards, picture cards, coloring pages, certificates, feelings cards, etc.)	1 per teacher	$399.00 + 5% shipping	$418.95
	Print materials	32 pages per student for 22 students	$0.10 per page	$70.40
Training	Trainer fee*	2 h per training session	$400.00	$100.00
	Teacher time*	2 person-hours	$60.81 per hour	$121.62
	Equipment* (computer with video conferencing capability)	2 h per teacher	$0.16 per hour	$0.32
Screening	Teacher time	2 person-hours	$60.81 per hour	$121.62
	Materials	1 per each student	$9.99	$9.99
	Equipment (computer)	2 h per teacher	$0.16 per hour	$0.32
Delivery	Teacher preparation time	2 person-hours per class	$60.81 per hour	$121.62
	Teacher instructional time	8 person-hours per class	$60.81 per hour	$486.48
	Equipment (computer with USB or DVD player capability or DVD player)	8 h per class	$0.16 per hour	$1.28
	Equipment (projector)	8 h per class	$0.08 per hour	$0.64
	Facility (900 square feet of classroom space in school building)	8 h per class	$8.95 per hour	$71.60
	First Year or 1-Year Program Cost			
Per teacher				$1,525
Per student				$69
	Subsequent Years Cost			
Per teacher				$884
Per student				$40
	5-Year Intervention Life Cost per Year			
Per teacher				$962
Per student				$44

†Costs per teacher unless otherwise specified in row headings. *Startup cost, not relevant for subsequent years.

perspective, including resources regardless of who pays for them. We calculated the incremental cost of *WHS* implementation relative to a no-intervention control condition, assuming that *WHS* is an add-on that does not replace another program with similar target outcomes.

Table 2 lists all ingredients by stage of implementation. Teachers represented the unit of measurement in cost calculations. We assumed 22 students per teacher, the average in this study and the nation (National Center for Education Statistics, 2015), to convert per-teacher costs into per-student costs. We used national prices to improve the generalizability of estimates. The study included multiple waves of teachers who may have implemented the program across multiple years. We assumed the same resource use across cohorts and presented costs in 2018 dollars.

Ingredients and Prices

Curriculum, Other Materials, and Training. Teachers receive the *WHS* curriculum as a package, which includes videos, skill tickets, and a detailed 166-page teacher's guide. The guide provides skill cards, picture cards, student booklets, coloring pages, certificates, and feelings cards. Although teachers received *WHS* and screening materials free of charge in this study, we included the cost of these items in our calculations at the price the developers charged at the time of publication (https://www.irised.com/products/we-have-skills). Print materials, provided to teachers electronically, represented the printing costs for teachers. We based training time and fees on the reports from developers and the fees charged by *WHS* trainers. In this study, about four teachers on average participated in each training session. Accordingly, we spread trainer fees across four teachers when calculating per teacher costs.

Teacher Time. Teacher time dedicated to training, delivery, and screening was priced as total hourly wages plus benefits, assuming all teachers screened students. We used a national median salary of $58,230 for elementary school teachers (U.S. Bureau of Labor Statistics, BLS, 2019) converted to an hourly wage, assuming 1,440 work hours per year for K-12 teachers (Institute of Education Sciences, 2020). The benefits represented 33.5% of total compensation (BLS, 2018). Teachers' time was priced at $60.81 per hour ($40.44 wages + $20.37 benefits).

Equipment. Teachers and trainers connected via video conference, teachers screened students online using computers or other devices, and *WHS* delivery required a computer and a projector or TV in each classroom. The U.S. Department of Education indicated that at least 98% of U.S. public schools had computers with internet access in 2008 and over 97% had computers with LCD or DLP projectors in classrooms (Gray et al., 2010). We nonetheless included costs associated with the use of this equipment as they represented reallocation of school resources to *WHS* activities. We used the purchase price of a desktop computer and a projector suggested by CostOut, a cost tool kit (Hollands et al., 2015), spread over a standard 3-year lifetime following straight-line amortization where annual use was assumed to be 1,440 h.

Facilities. The delivery of *WHS* took place in regular elementary classrooms. To calculate costs of classroom use for *WHS*, we used the median construction cost of an

elementary school building from CostOut and uprated construction costs by 21% to include site preparation, furniture, furnishings, and fees (Levin et al., 2018). We then annualized these total building costs over 30 years using the conventional 3% interest rate to obtain the cost per square foot of school space per year. We assumed a standard classroom size of 900 square feet and defined annual use as the duration of the academic year (1,440 h per year).

Startup vs. Maintenance Costs

Costs of the *WHS* curriculum package and training were considered startup costs incurred only in the first year. The *WHS* license fee grants teachers lifetime access, so teacher can use the purchased *WHS* program for many years, up to the duration of their tenure in the position. Maintenance beyond the first year includes yearly costs for print materials and screening for delivery to new cohorts of students. We calculated per-year cost of *WHS* implementation and, as teachers or their schools adopt new programs over time, a 5-year life by adding the first-year costs to discounted maintenance costs over the next 4 years with a conservative 3% discount rate.

Cost Effectiveness Analysis

The cost-effectiveness ratio, a measure of cost effectiveness of a program or intervention, represents the cost associated with per-unit change in the outcome of interest. It is calculated by dividing the cost estimate by the effect size measured for the same unit (e.g., per student). To calculate the cost-effectiveness ratio for *WHS*, we divided the per-student costs by the effect sizes for the WM, the primary outcome measure. The cost calculations considered all resource use beyond the resources used in a no-treatment control condition. The effect sizes reflected the improvement in outcomes relative to a business-as-usual comparison group. Accordingly, we reported incremental cost-effectiveness ratios relative to a no-treatment control condition.

Results

Table 1 presents sample characteristics for teachers and students. About two-thirds of teachers taught first grade only; others taught kindergarten or mixed-age groups of kindergarten and first-grade students. Table 3 provides descriptive statistics for the dependent variables. We found no differences between conditions at baseline on primary measures.

Joiners and Attrition

Teacher attrition was less than 4%: only two intervention and three control teachers did not complete the posttest assessment. Teachers missing posttest assessments represent cluster-level attrition. As recommended by the What Works Clearinghouse (2020), the reference sample in the analysis of student-level nonresponse included students from clusters with complete teacher data. This reference sample consisted of 2,708 students, of which teachers reported data at both time points for 2,570 (94.9%), at posttest but

Table 3. Descriptive statistics for teacher and student outcomes by condition and assessment time.

	Intervention		Control	
Measure	Pretest	Posttest	Pretest	Posttest
Teacher-report ESBA				
M	2.57	2.72	2.60	2.66
(SD)	(0.49)	(0.42)	(0.46)	(0.44)
n	1,504	1,363	1,294	1,226
Teacher-report WM				
M	3.85	4.26	3.96	4.17
(SD)	(0.85)	(0.76)	(0.79)	(0.76)
n	1,504	1,363	1,293	1,226
Teacher self-report TSES				
M	7.15	7.59	7.21	7.34
(SD)	(0.87)	(0.80)	(0.82)	(0.71)
n	66	64	61	58
Teacher self-report TCI				
M	2.74	2.62	2.59	2.76
(SD)	(0.80)	(0.86)	(0.68)	(0.73)
n	66	64	61	58
Observed rate of teachers' descriptive praise per minute				
M		0.26		0.19
(SD)		0.19		0.12
n		58		56
Observed rate of teachers' descriptive corrections per minute				
M		0.29		0.29
(SD)		0.20		0.20
n		58		56
Observer impressions of student behavior				
M		5.15		4.94
(SD)		0.87		0.85
n		54		53

Note. ESBA: Elementary Social Behavior Assessment; WM: Walker-McConnell Classroom Adjustment Behaviors subscale; TSES: Teacher Sense of Efficacy Scale; TCI: Teacher Concerns Inventory.

not pretest (joiners) for 19 (0.7%), and at pretest but not posttest (attrition) for 119 (4.4%); no students were missing data at both times.

Among students with pretest data, the overall attrition rate was 4.4%, with a differential rate of 2.5%: 3.1% for control and 5.6% for intervention. On the pretest ESBA score, the 4.4% of students without posttest data differed between condition (Hedges's $g = -0.12$) more than students with posttest data ($g = -0.03$), but the interaction between condition and missingness implied minimal bias (interaction = 0.07, 95% CI [−0.11, 0.25], $t_{117} = 0.75$, $p = .4535$, $w = .33$). In this model, w represented the probability of the hypothesis that included the missingness-by-condition interaction compared to a hypothesis without the interaction. The interaction between condition and posttest missingness on pretest WM scores similarly indicated little influence of attrition (interaction = 0.09 [−0.20, 0.38], $t_{117} = 0.60$, $p = .5478$, $w = .31$). The effect sizes for condition differences were similar for students without posttest data ($g = -0.16$) and for students who completed the study ($g = -0.12$). We did not examine differential effects for joiners, who comprised only 0.7% of sample.

The interpretation of attrition results is not straightforward. Differential rates of attrition offer little information (Foster & Bickman, 1996) and baseline condition differences for students missing posttest data only rely on small samples. Attrition bias may be best conveyed by the missingness-by-condition interaction, although the approach has limits (Graham & Donaldson, 1993). We therefore chose ML estimation with all available data to balance effects of nonresponse and minimize bias (Collins et al., 2001; Graham, 2009).

Table 4. efficacy results from time × condition analysis of change in student and teacher outcomes.

Effect or statistic	Teacher-report ESBA	Teacher-report WM	Teacher self-report TSES	Teacher self-report TCI
Model probability (*w*)	.99	.99	.82	.75
Fixed effects				
Intercept	2.60	3.94	7.21	2.59
	(0.02)	(0.05)	(0.10)	(0.10)
Time	0.06	0.20	0.16	0.13
	(0.02)	(0.04)	(0.09)	(0.08)
Condition	−0.03	−0.11	−0.06	0.15
	(0.03)	(0.08)	(0.14)	(0.14)
Time × Condition	0.08	0.19	0.28	−0.24
	(0.02)	(0.06)	(0.12)	(0.11)
Variances				
Classroom-level intercept	0.02	0.12		
	(0.00)	(0.02)		
Classroom-level gain	0.01	0.04		
	(0.00)	(0.01)		
Student-level intercept	0.13	0.35	0.42	0.39
	(0.00)	(0.01)	(0.07)	(0.06)
Student-level gain	0.06	0.13	0.22	0.20
	(0.00)	(0.00)	(0.03)	(0.03)
Intraclass correlation	.10	.23	.00	.00
Time × Condition				
Hedges's *g*	0.19	0.25	0.36	−0.30
95% CI	[0.08, 0.31]	[0.11, 0.40]	[0.05, 0.68]	[−0.58, −0.02]
p value	.0008	.0008	.0237	.0385
Degrees of freedom	123	123	125	125

Note. Table entries show parameter estimates with standard errors in parentheses except for model probabilities, intraclass correlations, Hedges's *g* values, *p* values, and degrees of freedom. ESBA: Elementary Social Behavior Assessment; WM: Walker-McConnell Classroom Adjustment Behaviors subscale; TSES: Teacher Sense of Efficacy Scale; TCI: Teacher Concerns Inventory; Two teachers did not report student behavior, leaving 125, while all 127 completed the TSES and TCI.

Intervention Effects on Teacher Reports of Student Behavior

WHS teachers reported greater gains than control teachers in their students' academically related behavioral skills measured by the ESBA ($g = 0.19$ [0.08, 0.31], $t_{123} = 3.46$, $p = .0008$, $w = .99$) and Classroom Adjustment Behaviors subscale of the WM ($g = 0.25$ [0.11, 0.40], $t_{123} = 3.46$, $p = .0008$, $w = .99$). Model results (Table 4) suggested that the hypothesis of a difference between conditions, measured by the Time × Condition effect, fit the data; that is, models for both ESBA and WM scores which included the Time × Condition interaction had considerably higher probabilities ($w = .99$) than models without the condition difference ($w = .01$).

Baseline Moderators of Intervention Effects on Student Behavior

Baseline levels of student behavior moderated treatment-group differences for both student outcomes: ESBA ($t_{117} = -5.51$, $p < .0001$, $w > .99$) and WM ($t_{117} = -1.96$, $p = .0523$, $w = .95$; see Table 5). In these models, *w* describes the probability for the model with the test of moderation compared to an equivalent model without the Pretest × Time × Condition interaction. Models with the moderation effect were much more likely than models without, given the data, and students who struggled at pretest generally gained more from the intervention.

Table 5. Moderation results from mixed time × condition analysis of change in student outcomes.

Effect or statistic	Teacher-report ESBA	Teacher-report WM
Model probability (*w*)	>.99	.71
Fixed effects		
Intercept	2.59	3.90
	(0.01)	(0.02)
Time	0.06	0.21
	(0.02)	(0.03)
Condition	0.00	0.00
	(0.01)	(0.03)
Time × Condition	0.08	0.16
	(0.02)	(0.05)
Pretest	1.00	1.00
	(0.01)	(0.01)
Pretest × Condition	0.00	0.00
	(0.02)	(0.02)
Pretest × Time	−0.30	−0.29
	(0.02)	(0.02)
Pretest × Time × Condition	−0.13	−0.05
	(0.02)	(0.02)
Variances		
Classroom-level intercept	0.00	0.00
	(0.00)	(0.00)
Classroom-level gain	0.01	0.03
	(0.00)	(0.00)
Student-level intercept	0.00	0.00
	(0.00)	(0.00)
Student-level gain	0.04	0.10
	(0.00)	(0.00)
Pretest × Time × Condition effects		
p Value	<.0001	.0523
Degrees of freedom	117	117

Note. Table entries show parameter estimates with standard errors in parentheses except for model probabilities, *p* values, and degrees of freedom. Pretest covariates for each respective measure were centered at the mean. ESBA: Elementary Social Behavior Assessment; WM: Walker-McConnell Classroom Adjustment Behaviors subscale; TSES: Teacher Sense of Efficacy Scale; TCI: Teacher Concerns Inventory.

Figure 1 shows the estimated difference between conditions on the vertical access and the range of pretest scores on the horizontal axis. The dark, decreasing line across baseline scores illustrates the moderation effect, surrounded by confidence bounds, and a zero on the vertical axis represents no difference between conditions. Within the confidence bounds, vertical lines represent sample percentiles, similar to a boxplot. For instance, in the graph of the ESBA, the vertical lines show that about 50% of students scored below about 2.7, 25% below 2.3, and 5.0% below 1.6. The confidence bounds exclude zero at values of 2.8 or below, for the lower-scoring 59% of students. This shows that students below the 59th sample percentile scored better on the ESBA after exposure to the *WHS* intervention than those in classrooms that conducted business as usual. To assist with interpretation, we estimated effect sizes of 0.67 at the minimum baseline ESBA score, 0.49 at the 5th sample percentile, 0.26 at the 25th, 0.13 at the median, and 0.05 for 30% students who scored at the maximum ESBA value at baseline.

The lower graph shows that *WHS* positively affected all WM pretest scores. The confidence bounds exclude zero for the entire range of scores. The intervention effect was nonetheless larger for students with lower baseline scores. Effect sizes ranged from 0.40 at the lowest baseline WM score, 0.31 at the 5th percentile, 0.25 at the 25th,

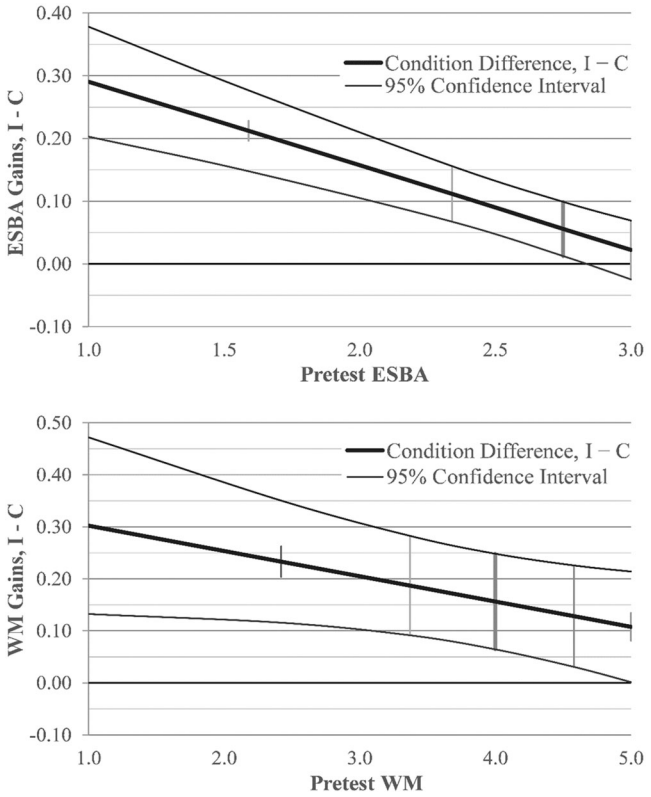

Figure 1. Differential effects of *We Have Skills* on the Elementary Social Behavior Assessment and Walker–McConnell Scale of Social Competence Classroom Adjustment Subscale. *Note.* The vertical axis shows the difference between conditions—zero on the vertical axis represents no difference between conditions—and the horizontal axis represents the range of pretest scores. The heavy decreasing line depicts the mean difference between conditions at each pretest value. The two thinner, outer lines show the 95% confidence bounds around the mean estimate. To show the location of the sample on the graphs, the vertical lines within the confidence bounds depict the median (heavier vertical line), 25th and 75th percentiles (thinner long lines), and the 5th and 95th percentiles (short outer lines). For example, on Graph A, a score of about 2.3 represents the lower 25th sample percentile at pretest.

0.21 at the median, 0.17 at the 75th, to 0.14 for the 10% of students with the highest score.

Intervention Effects for Observations of Student Behavior and Teacher Praise

Observers recorded higher ratings of student behavior in *WHS* classrooms than control classrooms ($g = 0.28$ [−0.10, 0.67], $t_{101} = 1.47$, $p = .1435$, $w = .50$). Although observer reports were of a magnitude consistent with teacher reports on the ESBA and WM, they were less reliable as indicated by the low model probability and wide confidence interval.

Observers coded greater rates of descriptive praise among *WHS* teachers than control teachers ($g = 0.56$ [0.19, 0.93], $t_{108} = 2.98$, $p = .0035$, $w = .96$), but the rate of descriptive corrections differed only minimally ($g = 0.08$ [−0.29, 0.45], $t_{108} = 0.43$, $p = .6680$, $w = .37$). Although not targets of the intervention, we found little change in the rate of

Table 6. Percent of teachers reporting each implementation activity of *We Have Skills* ($N = 54$).

	Mean rate (%)	Minimum rate (%)	Maximum rate (%)
Instruction:			
Discussion	99	67	100
Video	99	67	100
Skill card	88	0	100
Practice:			
3–5 times per day	95	43	100
Feedback:			
Descriptive praise	95	20	100
Descriptive correction	86	0	100
Skills tickets	75	0	100
Tally sheets	28	0	100
Happy notices	41	0	100
Certificates of mastery	37	0	100
Activities:			
Problem-solving discussion	83	0	100
Role-play	64	0	100
Songs	91	0	100
Games	31	0	100
Coloring pages	61	0	100
Skill booklets	52	0	100

nondescriptive praise, $g = 0.07$ [−0.31, 0.44], and nondescriptive corrections, $g = -0.07$ [−0.44, 0.31].

Intervention Effects on Teacher Self-Efficacy and Teacher Concerns

Compared to controls, *WHS* teachers reported greater gains on teachers' sense of self-efficacy (TSES scores, $g = 0.36$ [0.05, 0.68], $t_{125} = 2.29$, $p = .0237$, $w = .82$) and reduced concerns about discipline and motivation (TCI scores, $g = -0.30$ [−0.58, −0.02], $t_{125} = -2.09$, $p = .0385$, $w = .75$). The model probabilities imply that the hypothesis of a condition difference was 3 (TCI) to 4 (TSES) times as likely as those without.

Implementation Fidelity

Among intervention teachers, 54 of 66 (82%) completed at least one fidelity checklist during weeks 2–8. Table 6 summarizes fidelity data. High rates of implementation were reported for the discussion, video, and skill card instructional elements (88–99% completion rate) and practice (95%). Implementation of feedback-specific elements of *WHS* ranged from 28% for tally sheets to 95% for descriptive praise. Implementation of classroom activities were lowest for games at 31% and among other activities ranged from 52% for skill booklets to 91% for songs. Reports of implementation did not vary appreciably by the number of checklists completed.

Cost and Cost-Effectiveness

Table 2 presents the ingredients used to determine costs and summarizes total startup and maintenance costs of *WHS* implementation, including the curriculum, equipment, materials, supplies, fees, student instructional materials, and technical support. *WHS* cost $1,525 per teacher for the first year, $884 for each subsequent year, and $962 per year for the 5-

year life cycle. For 22 students per classroom, *WHS* cost $69 per student for the first year, $40 for subsequent years, and $44 per year for the 5-year life cycle. These costs represent the incremental costs over a no-treatment control. Replacing an existing program would require adjustments to the cost estimates (see Blonigen et al., 2008).

The cost-effectiveness ratio represents the cost associated with per-student change in the WM. We estimated a first-year cost-effectiveness ratio of $276 per student for *WHS* from the first-year costs of $69 per student and the effect size of 0.25 for the WM (i.e., $69/0.25). The cost-effectiveness ratio represents the incremental cost-effectiveness ratio relative to the no-treatment control condition. Assuming similar effects in subsequent cohorts of students, the cost-effectiveness ratio for the 5-year intervention life cycle could be $176 per student. The cost-effectiveness ratio may also vary by the initial skill level as implied by Figure 1.

Discussion and Conclusions

Marquez and colleagues (2014) developed *WHS* to include lessons grounded in evidence-based practices that appeal to early elementary school children and take minimal teacher time. This study aimed to assess whether the highly efficient classroom intervention, which requires minimal PD to implement, could improve students' classroom behavior and increase teachers' perceptions of their ability to manage classroom behavior and motivate their students. The study compared 66 teachers who delivered *WHS* with 1,515 of their students to 61 control-group teachers and their 1,302 students in a cluster-randomized trial.

The evidence suggests benefits from the *WHS* multimedia program on early elementary students' academically related behavioral skills (per DiPerna, 2006). This study finds stronger effects for the distal measure of classroom adjustment, the WM, $g = 0.25$, 95% CI [0.11, 0.40], than for the intervention-aligned ESBA, $g = 0.19$ [0.08, 0.31], suggesting the successful transfer of skills to a somewhat broader set of behaviors. This replicates the initial *WHS* evaluation, which reported similar effects for the ESBA, $g = 0.27$ [0.11, 0.43] (Marquez et al., 2014). Observer impressions of student behavior corroborated teacher reported measures in this study, $g = 0.28$ [−0.10, 0.67], but unfortunately these results were less reliable. Moderation tests indicated that students who struggled initially gained more from *WHS* when rated by the ESBA and WM.

Observers coded a higher rate of descriptive praise among intervention teachers than controls, $g = 0.56$ [0.19, 0.93], but not descriptive corrections. Intervention teachers also reported improved self-efficacy for behavior management, $g = 0.36$ [0.05, 0.68] and fewer concerns about student behavior, $g = -0.30$ [−0.58, −0.02].

Comparison to Related Programs

Effects from *WHS* are similar to those from other school-based SEL programs (e.g., Durlak et al., 2011; Taylor et al., 2017), with effect sizes from 0.13 to 0.24 on behavioral measures. For example, the *PATHS* program led to reported effects of $d = 0.00$–0.18 on social competence and externalizing measures (Kam et al., 2004). Belfield et al. (2015)

reported effects of $g = 0.14$ on social competence for the *4Rs* (*Reading, Writing, Respect, & Resolution*) program. A recent study of the widely disseminated *Second Step* program has demonstrated effect sizes of $|g| = 0.02–0.13$ across measures of social-emotional skills (Low et al., 2015, 2019), but as in the current study, moderation tests suggested stronger benefits for students with initially lower scores. Differential response based on initial skills may be more common than reported.

Many programs, however, require more time and effort to implement, as described below. The *Social Skills Improvement System–Classwide Intervention Program* (*SSIS-CIP*) is perhaps most similar to *WHS* in scope, teacher effort to implement, and materials (e.g., videos, lesson plans, student materials). *SSIS-CIP* and *WHS* both teach students how to ask for help and get along but diverge on other skills. *WHS* behaviors intended to enable academic achievement (DiPerna & Elliott, 2002) and teaches students how to listen to teachers, follow directions, complete their work, follow classroom rules, ask clarification questions, and work out strong feelings. The *SSIS-CIP* focuses on more traditional social skills: listening to others, cooperation, self-control, assertion, responsibility, and empathy. For the *SSIS-CIP*, DiPerna et al. (2018) reported effects of $g = 0.18$ [0.03, 0.33] on a teacher-rated social skills composite.

Cost and Cost-Effectiveness

The cost of *WHS* per student, $69 for the first year and $44 per year over 5 years, compares favorably to programs that teach social and classroom behavioral skills. Belfield et al. (2015) estimated per-student costs of the *4Rs* program at $733 and *Second Step* at $474 (prices adjusted to 2018). Hunter et al. (2018) reported a much lower cost per student of $20 for *SSIS-CIP* (price adjusted to 2018), but their estimate excluded time for screening, teacher preparation and instruction, and facilities. Hunter and colleagues averaged the first-year cost with the cost of one subsequent year. This illustrates the need to carefully review the ingredients and timeframes used in cost analyses to ensure comparability. Excluding those ingredients omitted by Hunter et al., for example, reduced the costs of *WHS* to $713 per teacher and $72 for subsequent years, or $32 and $3, respectively, per student. To directly compare *WHS* to *SSIS-CIP*, we averaged the *WHS* cost of the first year with one subsequent year to arrive at $18 per student, which illustrates that *WHS* produces similar or larger effect sizes at slightly lower cost. These lower cost estimates more closely estimate the costs of replacing an existing program (Blonigen et al., 2008).

The low cost of *WHS* stems from the limited requirements for training and instruction. *WHS* takes 2–3 h for training; other programs (e.g., *PATHS*) require up to 2 full days. *WHS* requires only about 60 min per lesson, or 8 h total for a classroom. *SSIS-CIP* requires about 50% more time (York, 2013), *PATHS* takes 15–32 h of instruction (Kam et al., 2004), and *Second Step* requires about 28 h to implement the program (Low et al., 2015).

The cost-effectiveness of *WHS* depends on students' differential response to the program based on initial student behavior. As shown in Figure 1, the effects of *WHS* vary by baseline level. Effect sizes may increase and incremental cost-effectiveness ratios decrease in settings where many students struggle with academically related behavioral

skills, making *WHS* more valuable. Conversely, classrooms with behaviorally skilled and well-adjusted students may benefit less from *WHS*. Although often unreported, differential response to initial skills may similarly affect the results from other SEL programs.

Implications

The success of *WHS* likely derives from its foundation in evidence-based practices, practical and efficient implementation, appeal to early elementary school children, and focus on students' academically related behaviors. To interest children who are continually exposed to the internet, TV, and other entertainment sources, *WHS* offers multimedia lessons featuring child actors, realistic vignettes, animal characters, and songs. *WHS* takes minimal preparation time from teachers, who can implement it with ease. Developers hypothesized that *in vivo* activities such as using descriptive praise to reinforce the use of skills and that corrections could help students avoid future mistakes or lapses yet take little time away from instruction.

Combined, the two trials of *WHS* suggest a functional relationship between *WHS* and student behavior. The larger effect size from the initial trial (Marquez et al., 2014) is also consistent with differential response to *WHS* by baseline performance, as the earlier trial included students with slightly lower average initial ESBA scores. Although the moderation results might suggest using *WHS* as a supplemental program, developers hypothesized that students would benefit most from whole-class implementation. As a universal intervention, *WHS* offers teachers and their students shared language and expectations for appropriate behavior. Teachers understand what their students have learned and can reinforce the use of skills or correct misbehavior throughout the day. Generalization is more difficult in pullout programs.

The goal of universal programs in multitiered support systems is to prevent challenging behaviors and improve behaviors that are easiest to change. Students who initially misunderstand expectations may respond quickly and positively to classwide instruction, allowing teachers to focus more intensive efforts on struggling students. As *WHS* serves this function with less effort from teachers than other universal interventions, it allows teachers to conserve precious time for instruction and for students who require additional support. In a tiered model that includes *WHS* at Tier 1, Tier 2 would include more intensive programs, such as *Skillstreaming the Elementary Child* (McGinnis & Goldstein, 1997) or *Second Step*, delivered to small groups of students who require additional demonstrations and practice of key behavioral or social skills. A smaller set of students may require highly intensive, Tier 3 interventions, such as *First Step to Success* (Walker et al., 2005), the *Good Behavior Game* (Dolan et al., 1989), or function-based interventions (Bambara & Kern, 2005; Matson, 2012) to mitigate the most challenging behaviors. *WHS* fits nicely in a tiered system that emphasizes efficient instruction of basic skills with minimal teacher effort, leaving time to address challenging behavior with more-intensive approaches.

Limitations

Several limitations temper the interpretation of the results. The most reliable evidence of intervention effects relied on teacher self-reports and reports of their students. We could

not blind teachers to condition. Nonetheless, a large number of SEL program evaluations rely on teacher reports (e.g., Durlak et al., 2011; Taylor et al., 2017), and we find no evidence in the literature that teacher reports compromise internal validity. Concerns about bias from self-reports often stem from misconceptions (Chan, 2009). Walker et al. (2015) reported that

> there is a long-standing body of research showing that teacher ratings predict important student outcomes, such as academic achievement and peer social status .... Even against the so-called gold standard of direct classroom observations, teacher ratings predict longitudinal outcomes about as well or better (pp. 365–366)

Our observations of student behavior also suggest similar differences in behavior between treatment conditions, although these results were less reliable.

Students were assessed in only one school year. Low et al. (2019) showed that for some skills similar to those measured in this study, students tended to return to preintervention levels by the fall of the next school year. The same phenomenon may have occurred in this study.

Observation of implementation would have strengthened conclusions about *WHS* fidelity. Doing so, however, would have required additional, cost-prohibitive observations. Investigators chose to maximize generalizability with a varied set of classrooms from several states at the expense of more-intensive data on *WHS* implementation in the classroom.

Conclusions

WHS developers set out to demonstrate that highly efficient instruction in specific skills associated with academic success (Marquez et al., 2014) could produce behavioral change in students commensurate with more-intensive and time-consuming programs. This cluster-randomized trial achieves that goal with established measures of student behavior, an experimental design to strengthen the internal validity, and a large sample to support generalizability. *WHS* offers a complete student curriculum for teachers who may lack the time, training, or expertise to provide social skills instruction, and the *WHS* instructional materials are flexible enough to support varying school settings, teachers' approaches to instruction, and students' needs. Evaluations of school-based social–emotional or behavioral programs have not always demonstrated lasting effects (Low et al., 2019; Sklad et al., 2012), and follow-up effects have not been evaluated for *WHS*. But given its ease of implementation, efficient instruction, and size of effects ($g = 0.19$ & 0.25) compared to similar programs, *WHS* may offer favorable results at a reasonably low cost.

Acknowledgments

We graciously acknowledge the many children, teachers, administrators, parents, schools, and research staff who made this study possible.

Disclosure Statement

Brion Marquez receives royalties from the sale of *WHS*, but in this study he did not participate in the randomization of teachers, collection or analysis of data, or presentation of results. All other authors declare that they have no conflicts of interest.

Funding

The U.S. Department of Education, Institute of Education Sciences, funded this research through Grant R305A150046 to IRIS Educational Media (dba Trifoia). The contents in this publication are those of the authors and do not represent views or policies of the Institute of Education Sciences.

ORCID

Keith Smolkowski http://orcid.org/0000-0003-2565-3297
Gulcan Cil http://orcid.org/0000-0002-8589-1468
Lisa A. Strycker http://orcid.org/0000-0001-8264-400X

References

Allinder, R. (1995). An examination of the relationship between teacher efficacy and curriculum-based measurement and student achievement. *Remedial and Special Education*, *16*(4), 247–254. https://doi.org/10.1177/074193259501600408

Allison, P. D. (2009). Missing data. In R. E. Millsap, A. Maydeu-Olivares, R. E. Millsap, & A. Maydeu-Olivares (Eds.), *The Sage handbook of quantitative methods in psychology* (pp. 72–89). Sage. https://doi.org/10.4135/9780857020994.n4

Arnesen, A., Smolkowski, K., Ogden, T., & Melby-Lervåg, M. (2018). Validation of the elementary social behavior assessment: Teacher ratings of students' social skills adapted to Norwegian, Grades 1 to 6. *Emotional and Behavioural Difficulties*, *23*(1), 39–54. https://doi.org/10.1080/13632752.2017.1316473

Bambara, L. M., & Kern, L. (2005). *Individualized supports for students with problem behaviors: Designing positive behavior plans*. Guilford.

Batchelor, K. E., & Bintz, W. P. (2012). Hand-clap songs across the curriculum. *The Reading Teacher*, *65*(5), 341–345. https://doi.org/10.1002/TRTR.01052

Belfield, C., Bowden, B., Klapp, A., Levin, H., Shand, R., & Zander, S. (2015). *The economic value of social and emotional learning* (Rev. ed.). Teachers College, Columbia University, Center for Benefit-Cost Studies in Education. www.cbcse.org

Blonigen, B., Harbaugh, W., Singell, L., Horner, R. H., Irvin, L. K., & Smolkowski, K. (2008). Application of economic analysis to school-wide positive behavior support (SW-PBS) programs. *Journal of Positive Behavior Interventions*, *10*(1), 5–19. https://doi.org/10.1177/1098300707311366

Boxer, A. (Ed.) (2019). *The researchED guide to explicit and direct instruction: An evidence-informed guide for teachers*. John Catt.

Brown, E. C., Low, S., Smith, B. H., & Haggerty, K. P. (2011). Outcomes from a school-randomized controlled trial of *Steps to Respect*: A bullying prevention program. *School Psychology Review*, *40*(3), 423–443. https://doi.org/10.1080/02796015.2011.12087707

Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, *65*(1), 23–35.10.1007/s00265-010-1029-6 https://doi.org/10.1007/s00265-010-1029-6

Carnine, D. (1980). Relationships between stimulus variation and the formation of misconceptions. *The Journal of Educational Research*, *74*(2), 106–110. https://doi.org/10.1080/00220671.1980.10885292

Chan, D. (2009). So why ask me? Are self-report data really that bad?. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends* (pp. 309–336). Routledge.

Collins, L. M., Schafer, J. L., & Kam, C.-M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, *6*(4), 330–351.

Crone, D. A., Hawken, L. S., & Horner, R. H. (2010). *Responding to problem behavior in schools: The Behavior Education Program*. Guilford.

DiPerna, J. C. (2006). Academic enablers and student achievement: Implications for assessment and intervention services in the schools. *Psychology in the Schools*, *43*(1), 7–17. https://doi.org/10.1002/pits.20125

DiPerna, J. C., & Elliott, S. (2002). Promoting academic enablers to improve student achievement. *School Psychology Review*, *31*(3), 293–298. https://doi.org/10.1080/02796015.2002.12086156

DiPerna, J. C., Lei, P., Cheng, W., Hart, S., & Bellinger, J. (2018). A cluster randomized trial of the *Social Skills Improvement System–Classwide Intervention Program* (*SSIS-CIP*) in first grade. *Journal of Educational Psychology*, *110*(1), 1–16. http://dx.doi.org/10.1037/edu0000191 https://doi.org/10.1037/edu0000191

Dolan, L., Turkkan, J., Werthamer–Larrson, L., & Kellam, S. G. (1989). *The Good Behavior Game training manual*. The Johns Hopkins Prevention Research Center.

Dunlap, G., & Fox, L. (2014). Supportive interventions for young children with social, emotional and behavioral delays and disorders. In H. Walker & F. Gresham (Eds.), *Handbook of evidence-based practices for emotional and behavioral disorders* (pp. 503–517). Guilford.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, *82*(1), 405–432.

Engelmann, S., & Carnine, D. (1991). *Theory of instruction: Principles and applications* (Rev. Ed.). ADI Press.

Evans, S. W., Rybak, T., Strickland, H., & Owens, J. S. (2014). The role of school mental health models in preventing and addressing children's emotional and behavioral problems. In H. Walker & F. Gresham (Eds.), *Handbook of evidence-based practices for emotional and behavioral disorders* (pp. 394–409). Guilford.

Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D. W., & Beechum, N. O. (2012). *Teaching adolescents to become learners. The role of noncognitive factors in shaping school performance: A critical literature review*. University of Chicago Consortium on Chicago School Research.

Fimian, M. J. (1984). The development of an instrument to measure occupational stress in teachers: The Teacher Stress Inventory. *Journal of Occupational Psychology*, *57*(4), 277–293. https://doi.org/10.1111/j.2044-8325.1984.tb00169.x

Foster, E. M., & Bickman, L. (1996). An evaluator's guide to detecting attrition problems. *Evaluation Review*, *20*(6), 695–723. https://doi.org/10.1177/0193841X9602000603

Funk, D., & Fay, J. (1995). *Teaching with Love & Logic: Taking control of the classroom*. The Love and Logic Press, Inc.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*, 549–576. https://doi.org/10.1146/annurev.psych.58.110405.085530

Graham, J. W., & Donaldson, S. I. (1993). Evaluating interventions with differential attrition: The importance of nonresponse mechanisms and use of follow-up data. *Journal of Applied Psychology, 78*(1), 119–128.

Gray, L., Thomas, N., & Lewis, L. (2010). *Educational technology in U.S. public schools: Fall 2008 (NCES 2010–034)*. U.S. Department of Education, National Center for Education Statistics.

Greenberg, M. T., Kusche, C. A., Cook, E. T., & Quamma, J. P. (1995). Promoting emotional competence in school-aged children: The effects of the *PATHS* curriculum. *Development and Psychopathology*, *7*(1), 117–136. [Database] https://doi.org/10.1017/S0954579400006374

Greenland, S., Senn, S., Rothman, K., Carlin, J., Poole, C., Goodman, S., & Altman, D. (2016). Statistical tests, p-values, confidence intervals, and power: A guide to misinterpretations. *The American Statistician*, *70*(2, supplement to Wasserstein & Lazar), 1–12.

Gresham, F. M. (1986). Conceptual issues in the assessment of social competence in children. In P. Strain, M. Guralnick, & H. Walker (Eds.), *Children's social behavior: Development, assessment and modification* (pp. 143–180). Academic Press.

Gresham, F. M. (1998). Social skills training: Should we raze, remodel, or rebuild? *Behavioral Disorders*, *24*(1), 19–25. https://doi.org/10.1177/019874299802400103

Gresham, F. M. (2002). Teaching social skills to high-risk children and youth: Preventive and remedial approaches. In M. Shinn, H. Walker, & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Preventive and remedial approaches* (pp. 403–432). National Association of School Psychologists.

Gresham, F. M., & Elliott, S. (1990). *Social skills rating system*. American Guidance Service.

Gresham, F. M., & Elliott, S. (2014). Social skills assessment and training in emotional and behavioral disorders. In H. M. Walker & F. M. Gresham (Eds.), *Handbook of evidence-based practices for emotional and behavioral disorders* (pp. 152–172). Guilford.

Hawkins, J. D., Catalano, R. F., Kosterman, R., Abbott, R. D., & Hill, K. G. (1999). Preventing adolescent health-risk behaviors by strengthening protection during childhood. *Archives of Pediatrics & Adolescent Medicine*, *153*(3), 226–234.

Henggeler, S. W., Schoenwald, S. K., Borduin, C. M., Rowland, M. D., & Cunningham, P. B. (1998). *Multisystemic treatment of antisocial behavior in children and adolescents*. Guilford.

Hersh, R., & Walker, H. M. (1983). Great expectations: Making schools effective for all students. *Policy Studies Review*, *2*, 147–188.

Hoagwood, K. E., Olin, S. S., Kerker, B. D., Kratochwill, T. R., Crowe, M., & Saka, N. (2007). Empirically based school interventions targeted at academic and mental health functioning. *Journal of Emotional and Behavioral Disorders*, *15*(2), 66–92. https://doi.org/10.1177/10634266070150020301

Hoagwood, K. E., Jensen, P. S., Acri, M. C., Olin, S., Lewandowski, R., & Herman, R. (2012). Outcome domains in child mental health research since 1996: Have they changed and why does it matter? *Journal of the American Academy of Child and Adolescent Psychiatry*, *51*(12), 1241–1260. https://doi.org/10.1016/j.jaac.2012.09.004

Hollands, F. M., Hanisch-Cerda, B., Levin, H. M., Belfield, C., Menon, A., Shand, R., Pan, Y., Bakir, I., & Cheng, H. (2015). *CostOut – the CBCSE cost tool kit*. Center for Benefit-Cost Studies of Education, Teachers College, Columbia University. www.cbcsecosttoolkit.org

Hunter, L. J., DiPerna, J. C., Hart, S. C., & Crowley, M. (2018). At what cost? Examining the cost effectiveness of a universal social–emotional learning program. *School Psychology Quarterly : The Official Journal of the Division of School Psychology, American Psychological Association*, *33*(1), 147–154.

Institute of Education Sciences (2020). *Cost analysis: A starter kit* (Version 1.0, IES 2020–001rev). Author.

Jaccard, J., & Turrisi, R. (2003). *Interaction effects in multiple regression* (2nd ed.). Sage.

Jones, J., Lerman, D. C., & Lechago, S. (2014). Assessing stimulus control and promoting generalization via video modeling when teaching social responses to children with autism. *Journal of Applied Behavior Analysis*, *47*(1), 37–50. https://doi.org/10.1002/jaba.81

Kam, C., Greenberg, M., & Kusche, C. (2004). Sustained effects of the *PATHS* curriculum on the social and psychological adjustment of children in special education. *Journal of Emotional and Behavioral Disorders*, *12*(2), 66–78. https://doi.org/10.1177/10634266040120020101

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.

Levin, H. M., McEwan, P. J., Belfield, C., Bowden, A. B., & Shand, R. (2018). *Economic evaluation in education: Cost-effectiveness and benefit-cost analysis*. (3rd ed.). Sage.

Low, S., Cook, C. R., Smolkowski, K., & Buntain-Ricklefs, J. (2015). Promoting social–emotional competence: An evaluation of the elementary version of *Second Step®*. *Journal of School Psychology*, *53*(6), 463–477.

Low, S., Smolkowski, K., & Cook, C. (2016). What constitutes high-quality implementation of SEL programs? A latent class analysis of *Second Step®* implementation. *Prevention Science : The Official Journal of the Society for Prevention Research*, *17*(8), 981–991.

Low, S., Smolkowski, K., Cook, C., & Desfosses, D. (2019). Two-year impact of a universal social-emotional learning curriculum: Group differences from developmentally sensitive trends over time. *Developmental Psychology*, *55*(2), 415–433. https://doi.org/10.1037/dev0000621

Marquez, B., Marquez, J., Vincent, C. G., Pennefather, J., Sprague, J. R., Smolkowski, K., & Yeaton, P. (2014). The iterative development and initial evaluation of We Have Skills!, an innovative approach to teaching social skills to elementary students. *Education and Treatment of Children, 37*(1), 137–161.

Matson, J. L. (2012). *Functional assessment for challenging behaviors*. Springer.

Mayer, R. E. (2008). Applying the science of learning: Evidence-based principles for the design of multimedia instruction. *The American Psychologist*, *63*(8), 760–769.10.1037/0003-066X.63.8.760 https://doi.org/10.1037/0003-066X.63.8.760

McGinnis, E., & Goldstein, A. P. (1997). *Skillstreaming the elementary school child: New strategies and perspectives for teaching prosocial skills*. Research Press.

Miller, W. R., & Rollnick, S. (2012). *Motivational interviewing: Helping people change* (3rd ed.). Guilford.

Moore, W., & Esselman, M. (1992). Teacher efficacy, power, school climate and achievement: A desegregating district's experience [Paper presentation]. Annual Meeting of the American Educational Research Association, San Francisco, CA.

Murray, D. M. (1998). *Design and analysis of group-randomized trials*. Oxford University Press.

National Center for Education Statistics. (2015, July 31). *Digest of education statistics* (Table 209.30). https://nces.ed.gov/programs/digest/d14/tables/dt14_209.30.asp

Navarrete, K. (2019, October 7). 'Room Clear' becoming the norm in classrooms across Arizona. *ABC15 News Arizona*. https://www.abc15.com/news/region-phoenix-metro/central-phoenix/room-clear-becoming-the-norm-in-classrooms-across-arizona

Nelson, J. R. (1996). Designing schools to meet the needs of students who exhibit disruptive behavior. *Journal of Emotional and Behavioral Disorders*, *4*(3), 147–161. https://doi.org/10.1177/106342669600400302

Nelson, J. R., Hurley, K. D., Synhorst, L., Epstein, M. H., Stage, S., & Buckley, J. (2009). The child outcomes of a behavior model. *Exceptional Children*, *76*(1), 7–30. https://doi.org/10.1177/001440290907600101

Oakes, W. P., Lane, K., Common, E., & Buckman, M. (2018). Systematic screening for behavior in early childhood settings: Early identification and intervention within a tiered prevention framework. *Perspectives on Early Childhood Psychology and Education*, *3*, 10–38.

Patterson, G. R., Reid, J. B., & Dishion, T. J. (1992). *Antisocial boys*. Castalia Press.

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interaction effects in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*, *31*(4), 437–448. https://doi.org/10.3102/10769986031004437

Pennefather, J., & Smolkowski, K. (2015). Validation of the Elementary Social Behavior Assessment: A measure of student prosocial school behaviors. *Assessment for Effective Intervention, 40*(3), 143–154. https://doi.org/10.1177/1534508414557562

Rainey, W. D., & Larsen, J. D. (2002). The effects of familiar melodies on initial learning and long-term memory for unconnected text. *Music Perception*, *20*(2), 173–186. https://doi.org/10.1525/mp.2002.20.2.173

Ross, J. A. (1992). Teacher efficacy and the effects of coaching on student achievement. *Canadian Journal of Education, 17*(1), 51–65.

SAS Institute (2016). *SAS/STAT® 14.2 user's guide*. SAS Institute, Inc. SAS Institute, Product Documentation. http://support.sas.com/documentation/index.html

Saunders, J. B., Aasland, O. G., Babor, T. F., de la Fuente, J. R., & Grant, M. (1993). Development of the alcohol use disorders identification test (AUDIT): Who collaborative project on early detection of persons with harmful alcohol consumption-II. *Addiction (Abingdon, England)*, *88*(6), 791–804. https://doi.org/10.1111/j.1360-0443.1993.tb02093.x

Sklad, M., Diekstra, R., Ritter, M. D., Ben, J., & Gravesteijn, C. (2012). Effectiveness of school-based universal social, emotional, and behavioral programs: Do they enhance students' development in the area of skill, behavior, and adjustment? *Psychology in the Schools*, *49*(9), 892–909. https://doi.org/10.1002/pits.21641

Smith, S. W., Daunic, A. P., Aydin, B., Van Loan, C. L., Barber, B. R., & Taylor, G. G. (2016). Effect of *Tools for Getting Along* on student risk for emotional and behavioral problems in upper elementary classrooms: A replication study. *School Psychology Review*, *45*(1), 73–92. https://doi.org/10.17105/SPR45-1.73-92

Smolkowski, K., Crawford, L., Seeley, J. R., & Rochelle, J. (2019). Introduction to implementation science for research on learning disabilities. *Learning Disability Quarterly, 42*(4), 192–203. https://doi.org/10.1177/0731948719851512

Smolkowski, K., & Gunn, B. (2012). Reliability and validity of the Classroom Observations of Student-Teacher Interactions (COSTI) for kindergarten reading instruction. *Early Childhood Research Quarterly, 27*(2), 316–328. https://doi.org/10.1016/j.ecresq.2011.09.004

Soodak, L. C., & Podell, D. M. (1993). Teacher efficacy and student problem as factors in special education referral. *The Journal of Special Education*, *27*(1), 66–81. https://doi.org/10.1177/002246699302700105

Sprick, R. S. (2009). *CHAMPS: A proactive and positive approach to classroom management*. Pacific Northwest Publishing.

Sprick, R. S., Isaacs, S., Booher, M., Sprick, J., & Rich, P. (2014). *Foundations: A proactive and positive behavior support system* (3rd ed.). Northwest Publishing.

Suldo, S. M., Gormley, M. J., DuPaul, G. J., & Anderson-Butcher, D. (2014). The impact of school mental health on student and school-level academic outcomes: Current status of the research and future directions. *School Mental Health*, *6*(2), 84–98. https://doi.org/10.1007/s12310-013-9116-2

Taylor, P. J., Russ-Eft, D. F., & Chan, D. W. L. (2005). A meta-analytic review of behavior modeling training. *The Journal of Applied Psychology*, *90*(4), 692–709. https://doi.org/10.1037/0021-9010.90.4.692

Taylor, R. D., Oberle, E., Durlak, J. A., & Weissberg, R. P. (2017). Promoting positive youth development through school-based social and emotional learning interventions: A meta-analysis of follow-up effects. *Child Development*, *88*(4), 1156–1171. https://doi.org/10.1111/cdev.12864

Trentham, L. L., Silvern, S., & Brogdon, R. (1985). Teacher efficacy and teacher competency ratings. *Psychology in the Schools*, *22*(3), 343–352. https://doi.org/10.1002/1520-6807(198507)22:3<343::AID-PITS2310220317>3.0.CO;2-0

Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, *17*(7), 783–805. https://doi.org/10.1016/S0742-051X(01)00036-1

U.S. Bureau of Labor Statistics (2018). *Employer costs for employee compensation – March 2018*. U.S. Department of Commerce, Bureau of Labor Statistics. https://www.bls.gov/news.release/pdf/ecec.pdf

U.S. Bureau of Labor Statistics (2019). Occupational employment statistics, occupational employment and wages, May 2018: 25-2021 elementary school teachers except special education. U.S. Department of Commerce, Bureau of Labor Statistics. https://www.bls.gov/oes/2018/may/oes252021.htm

Walker, H. M. (1986). The AIMS (Assessments for Integration Into Mainstream Settings) assessment system: Rationale, instruments, procedures and outcomes. *Journal of Clinical Child Psychology*, *15*(1), 55–63. https://doi.org/10.1207/s15374424jccp1501_7

Walker, H. M., Colvin, G., & Ramsey, E. (1995). *Antisocial behavior in school: Strategies and best practices*. Brooks/Cole.

Walker, H. M., Golly, A., McLane, J., & Kimmich, M. (2005). The Oregon First Step to Success Replication Initiative: Statewide results of an evaluation of the program's impact. *Journal of Emotional and Behavioral Disorders*, *13*(3), 163–172. https://doi.org/10.1177/10634266050130030401

Walker, H. M., & McConnell, S. (1995). *Walker–McConnell scale of social competence and school adjustment: Elementary version*. Singular Publishing Group.

Walker, H. M., & Rankin, R. (1983). Assessing the behavioral expectations and demands of less restrictive settings. *School Psychology Review*, *12*(3), 274–284. https://doi.org/10.1080/02796015.1983.12085043

Walker, H. M., Stiller, B., Golly, A., Kavanagh, K., Severson, H., & Feil, E. (1997). *First Step to Success: Helping young children overcome antisocial behavior (an early intervention program for grades K–3)*. SoprisWest.

Walker, H. M., Marquez, B., Yeaton, P., Pennefather, J., Forness, S., & Vincent, C. (2015). Teacher judgment in assessing students' social behavior using a response-to-intervention framework: Using what teachers know. *Education and Treatment of Children*, 38, 363–382.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose. *The American Statistician*, *70*(2), 129–133. https://doi.org/10.1080/00031305.2016.1154108

Webster-Stratton, C., Reid, M. J., & Hammond, M. (2004). Treating children with early-onset conduct problems: Intervention outcomes for parent, child, and teacher training. *Journal of Clinical Child and Adolescent Psychology: The Official Journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53*, *33*(1), 105–124. https://doi.org/10.1207/S15374424JCCP3301_11

Wentzel, K. (2009). Peers and academic functioning at school. In K. Rubin, W. Bukowski, & B. Laursen (Eds.), *Handbook of peer interactions, relationships and groups* (pp. 531–547). Guilford.

What Works Clearinghouse (2020). *What Works Clearinghouse standards handbook* (Version 4.1). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation, WWC. http://ies.ed.gov/ncee/wwc/

York, H. (2013). *Evaluating the effectiveness of a classwide social skills intervention with preschoolers and kindergarteners* [Master's thesis]. Louisiana State University. https://digitalcommons.lsu.edu/gradschool_theses/1958/

Zeng, S., Corr, C. P., O'Grady, C., & Guan, Y. (2019). Adverse childhood experiences and preschool suspension expulsion: A population study. *Child Abuse & Neglect*, *97*, 104149. https://doi.org/10.1016/j.chiabu.2019.104149