

What Works Clearinghouse



Children Classified as Having an Emotional Disturbance

March 2012

First Step to Success

Program Description¹

First Step to Success is an early intervention program designed to help children who are at risk for developing aggressive or antisocial behavioral patterns. The program uses a trained behavior coach who works with each student and his or her class peers, teacher, and parents for approximately 50 to 60 hours over a three-month period. *First Step to Success* includes three interconnected modules: screening, classroom intervention, and parent training. The screening module is used to identify candidates who meet eligibility criteria for program participation. Classroom intervention and parent training comprise the program intervention component of *First Step to Success*.

Research²

Two studies of *First Step to Success* that fall within the scope of the Children Classified as Having an Emotional Disturbance review protocol meet What Works Clearinghouse (WWC) evidence standards, and no studies meet WWC evidence standards with reservations.³ The two studies included 243 children in kindergarten through third grade who attended schools in New Mexico and Oregon. Based on these two studies, the WWC considers the extent of evidence for *First Step to Success* on children classified with an emotional disturbance (or children at risk for classification) to be small for all domains examined in this report (external behavior, emotional/internal behavior, social outcomes, reading achievement/literacy, and other academic performance domains).

Effectiveness

First Step to Success was found to have positive effects on external behavior, potentially positive effects on emotional/internal behavior, social outcomes, and other academic performance, and no discernible effects on reading achievement/literacy for children classified with an emotional disturbance.

| Report Contents | |
|---|-------|
| Overview | p. 1 |
| Program Information | p. 2 |
| Research Summary | p. 3 |
| Effectiveness Summary | p. 4 |
| References | p. 7 |
| Research Details for Each Study | p. 9 |
| Outcomes Measures for Each Domain | p. 12 |
| Findings Included in the Rating for Each Outcome Domain | p. 14 |
| Supplemental Single-Case Design Studies | p. 17 |
| Endnotes | p. 18 |
| Rating Criteria | p. 19 |
| Glossary of Terms | p. 20 |

Table 1. Summary of findings⁴

| Outcome domain | Rating of effectiveness | Improvement index (percentile points) | | Number of studies | Number of students ⁵ | Extent of evidence |
|------------------------------|------------------------------|---------------------------------------|------------|-------------------|---------------------------------|--------------------|
| | | Average | Range | | | |
| External behavior | Positive effects | +28 | +15 to +38 | 2 | 243 | Small |
| Emotional/internal behavior | Potentially positive effects | +10 | na | 1 | 46 | Small |
| Social outcomes | Potentially positive effects | +23 | +14 to +28 | 1 | 197 | Small |
| Reading achievement/literacy | No discernible effects | -2 | -5 to +2 | 1 | 193 | Small |
| Other academic performance | Potentially positive effects | +12 | na | 1 | 194 | Small |

na = not applicable

Program Information

Background

First Step to Success was developed by Hill M. Walker, Ph.D., a research scientist at the Oregon Research Institute and a professor at the University of Oregon and is distributed by Sopris West, 4185 Salazar Way, Frederick, CO 80504. Email: customerservice@sopriswest.com. Web: <http://www.sopriswest.com>. Telephone: (800) 547-6747. Fax: (888) 819-7767.

Program details

First Step to Success is used with students in kindergarten through third grade who are at risk for developing anti-social behavior patterns. No information on the scope of use or the demographic characteristics of program users is available. The program incorporates three interconnected modules: screening, school intervention, and parent training. Teachers use a screening tool to nominate students and rate their behavior using a standardized scale and definition of antisocial behavior. The school intervention module, Contingencies for Learning Academic and Social Skills (CLASS), focuses on reducing problem behavior and increasing adaptive, prosocial behaviors. A behavior coach works with the teacher while the teacher observes and learns the techniques necessary to implement the program. The student is taught to recognize and replace inappropriate behaviors with appropriate ones, which are subsequently reinforced by classroom peers who are taught positive strategies to support the student. The student accrues points toward his or her behavioral goal. If the student reaches a daily goal, he or she gets to choose an activity designed for the whole class to enjoy. The CLASS module requires 30 program days across three phases (coach, teacher, and maintenance) for completion. The parenting component (HomeBase) is implemented in concert with the CLASS program at school. The behavior coach meets with the student's parents/caregivers for approximately 45 minutes per week for six weeks. Parents are taught to focus on and encourage the following child competencies: communication, cooperation, limit setting, problem solving, friendship making, and confidence development.

Cost⁶

The cost of implementing the *First Step to Success* model is approximately \$500 per student and includes materials and the behavioral coach's time.

Research Summary

Twenty-one studies reviewed by the WWC investigated the effects of *First Step to Success* on children classified as having an emotional disturbance (or at risk for classification). Two studies (Walker et al., 1998; Walker et al., 2009) are randomized controlled trials (RCTs) that meet WWC evidence standards without reservations. Those two studies are summarized in this report. The remaining 19 studies do not meet either WWC eligibility screens or evidence standards. (See references beginning on p. 7 for citations for all 21 studies.)

Five additional studies were reviewed against the pilot Single-Case Design standards. Three studies met the pilot Single-Case Design standards and two did not meet pilot Single-Case Design standards. Studies reviewed against pilot Single-Case Design standards are listed in Appendix D and do not contribute to the intervention’s rating of effectiveness.

Table 2. Scope of reviewed research

| | |
|--|------------|
| Grade | K, 1, 2, 3 |
| Delivery method | Individual |
| Program type | Supplement |
| Studies reviewed | 21 |
| Meets WWC standards | 2 studies |
| Meets WWC standards with reservations | 0 studies |

Summary of studies meeting WWC evidence standards without reservations

Walker et al. (1998) randomly assigned 46 kindergarten children in the Eugene, Oregon school district to the *First Step to Success* intervention group or to a wait-list control condition. The study included two cohorts of students; Cohort 1 included 24 students who were in kindergarten during the 1993–94 academic year, and Cohort 2 included 22 students who were in kindergarten during the 1994–95 academic year. Students were described as exhibiting antisocial behaviors, including victimizing others, severe tantrums, and aggression. The study reported student outcomes after approximately three months of program implementation in kindergarten.⁷

Walker et al. (2009) examined the effects of *First Step to Success* on children in grades 1–3 attending schools in the Albuquerque, New Mexico school district. The analysis sample included approximately 198 students (analysis samples varied across outcomes). Teachers were randomly assigned to the *First Step to Success* intervention or to a usual care control condition across two cohorts (one for the 2005–06 academic year and the other for the 2006–07 academic year). The Systematic Screening for Behavior Disorders (SSBD) instrument was used to identify children in each classroom who were exhibiting the most severe behavioral concerns. Teachers then completed Stage 2 of the SSBD, which included ratings of student adaptive and maladaptive behavior. The student in each classroom with the highest average ranking across the SSBD Stage 2 measures was targeted for inclusion in the study. The study reported student outcomes after approximately three months of program implementation. The WWC based its effectiveness ratings on findings from comparisons of 101 students who received *First Step to Success* and 97 control students who received usual care.⁸

Summary of studies meeting WWC evidence standards with reservations

No studies of *First Step to Success* meet WWC evidence standards with reservations.

Effectiveness Summary

The WWC review of interventions for Children Classified as Having an Emotional Disturbance addresses student outcomes in seven domains: external behavior, emotional/internal behavior, social outcomes, reading achievement/literacy, math achievement, school attendance, and other academic performance. The two studies that contribute to the effectiveness rating in this report cover five domains: external behavior, emotional/internal behavior, social outcomes, reading achievement/literacy, and other academic performance. The findings below present the authors’ estimates and WWC-calculated estimates of the size and statistical significance of the effects of *First Step to Success* on children classified as having an emotional disturbance. For a more detailed description of the rating of effectiveness and extent of evidence criteria, see the WWC Rating Criteria later in this report.

Summary of effectiveness for the external behavior domain

Two studies reported findings in the external behavior domain.

Walker et al. (1998) found, and the WWC confirmed, four positive and statistically significant differences between treatment and comparison groups on academic engaged time, the Child Behavior Checklist–Teacher Report Forms (CBCL-TRF) Aggression Subscale, the Early Screening Project (ESP) Adaptive Behavior Subscale, and the ESP Maladaptive Behavior Subscale.

Walker et al. (2009) found, and the WWC confirmed, four positive and statistically significant differences between treatment and comparison groups on academic engaged time, the Social Skills Rating System (SSRS) Problem Behavior Subscale for Parents, the SSRS Problem Behavior Subscale for Teachers, and the SSBD Maladaptive Behavior Index. Although the overall design of the Walker et al. (2009) study meets evidence standards, there was high attrition on one outcome: the SSRS Problem Behavior Subscale for Parents outcome. The authors established equivalence for the analytic sample for this outcome; thus, this finding meets evidence standards with reservations.

The mean effect size from the four outcomes in Walker et al. (1998) and the mean effect size from the four outcomes in Walker et al. (2009) were both statistically significant. Thus, for the external behavior domain, two studies with strong designs showed statistically significant positive effects. This results in an intervention rating of positive effects for the domain, with a small extent of evidence.

Table 3. Rating of effectiveness and extent of evidence for the external behavior domain

| Rating of effectiveness | Criteria met |
|--|--|
| Positive effects <i>Strong evidence of a positive effect with no overriding contrary evidence.</i> | The review of <i>First Step to Success</i> had two studies that met WWC evidence standards for a strong design showing statistically significant positive effects, and no studies showing statistically significant or substantively important negative effects. |
| Extent of evidence | Criteria met |
| Small | The review of <i>First Step to Success</i> in the external behavior domain was based on two studies that included at least 34 schools and 243 students. ⁵ |

Summary of effectiveness for the emotional/internal behavior domain

One study reported findings in the emotional/internal behavior domain.

Walker et al. (1998) found, and the WWC confirmed, no statistically significant difference between treatment and comparison groups on the CBCL-TRF Withdrawn Subscale. However, the effect was positive and large enough to be substantively important according to WWC criteria (that is, at least 0.25 standard deviations).

Thus, for the emotional/internal behavior domain, one study with a strong design showed substantively important positive effects. This results in an intervention rating of potentially positive effects for the domain, with a small extent of evidence.

Table 4. Rating of effectiveness and extent of evidence for the emotional/internal behavior domain

| Rating of effectiveness | Criteria met |
|---|--|
| Potentially positive effects <i>Evidence of a positive effect with no overriding contrary evidence.</i> | The review of <i>First Step to Success</i> had one study showing a substantively important positive effect and no studies showing a statistically significant or substantively important negative effect or indeterminate effects. |
| Extent of evidence | Criteria met |
| Small | The review of <i>First Step to Success</i> in the emotional/internal behavior domain was based on one study that included an unknown number of schools and 46 students. |

Summary of effectiveness for the social outcomes domain

One study reported findings in the social outcomes domain.

Walker et al. (2009) found, and the WWC confirmed, three positive and statistically significant differences between treatment and comparison groups on the SSRS Social Skills Subscale for Parents, the SSRS Social Skills Subscale for Teachers, and the SSBD Adaptive Behavior Index. Although the overall design of the Walker et al. (2009) study meets evidence standards, there was high attrition on two outcomes in this domain: SSRS Social Skills Subscale for Parents, and SSRS Social Skills Subscale for Teachers. The authors established equivalence for the analytic sample for these outcomes; thus, these findings meet evidence standards with reservations.

The mean effect size from the three outcomes of the single study in this domain was statistically significant. Thus, for the social outcomes domain, one study with a strong design showed statistically significant positive effects.

This results in an intervention rating of potentially positive effects for the domain, with a small extent of evidence.

Table 5. Rating of effectiveness and extent of evidence for the social outcomes domain

| Rating of effectiveness | Criteria met |
|---|--|
| Potentially positive effects <i>Evidence of a positive effect with no overriding contrary evidence.</i> | The review of <i>First Step to Success</i> had one study showing a statistically significant positive effect and no studies showing a statistically significant or substantively important negative effect or indeterminate effects. |
| Extent of evidence | Criteria met |
| Small | The review of <i>First Step to Success</i> in the social outcomes domain was based on one study that included 34 schools and 197 students. ⁵ |

Summary of effectiveness for the reading achievement/literacy domain

One study reported findings in the reading achievement/literacy domain.

Walker et al. (2009) found a statistically significant difference between treatment and comparison groups on the Woodcock-Johnson III Diagnostic Reading Battery (WJ-III DRB) Letter-Word Identification. Based on WWC calculations, the effect was neither statistically significant nor large enough to be substantively important according to the WWC criteria. Walker et al. (2009) found, and the WWC confirmed, no statistically significant difference between treatment and comparison groups on oral reading fluency. Although the overall design of the Walker et al. (2009) study meets evidence standards, there was high attrition on both of the outcomes reported in this domain. The authors established equivalence for the analytic sample for these outcomes; thus, the findings for this domain meet evidence standards with reservations.

Thus, for the reading achievement/literacy domain, no studies showed statistically significant or substantively important effects. This results in an intervention rating of no discernible effects for the domain, with a small extent of evidence.

Table 6. Rating of effectiveness and extent of evidence for the reading achievement/literacy domain

| Rating of effectiveness | Criteria met |
|--|--|
| No discernible effects <i>There is no affirmative evidence of effects.</i> | The review of <i>First Step to Success</i> had no studies showing a statistically significant or substantively important effect, either positive or negative. |
| Extent of evidence | Criteria met |
| Small | The review of <i>First Step to Success</i> in the reading achievement/literacy domain was based on one study that included 34 schools and 193 students. ⁵ |

Summary of effectiveness for the other academic performance domain

One study reported findings in the other academic performance domain.

Walker et al. (2009) found, and the WWC confirmed, a positive and statistically significant difference between treatment and comparison groups on the SSRS Academic Competence Subscale.

Thus, for the other academic performance domain, one study with a strong design showed statistically significant positive effects. This results in an intervention rating of potentially positive effects for the domain, with a small extent of evidence.

Table 7. Rating of effectiveness and extent of evidence for the other academic performance domain

| Rating of effectiveness | Criteria met |
|---|--|
| Potentially positive effects <i>Evidence of a positive effect with no overriding contrary evidence.</i> | The review of <i>First Step to Success</i> had one study showing a statistically significant positive effect and no studies showing a statistically significant or substantively important negative effect or indeterminate effects. |
| Extent of evidence | Criteria met |
| Small | The review of <i>First Step to Success</i> in the other academic performance domain was based on one study that included 34 schools and 197 students. ⁵ |

References

Studies that meet WWC evidence standards without reservations

- Walker, H., Kavanagh, K., Stiller, B., Golly, A., Severson, H., & Feil, E. (1998). First Step to Success. An early intervention approach for preventing school antisocial behavior. *Journal of Emotional and Behavioral Disorders, 6*(2), 66–80.
- Walker, H. M., Seeley, J. R., Small, J., Severson, H. H., Graham, B. A., Feil, E. G., . . . Forness, S. R. (2009). A randomized controlled trial of the First Step to Success early intervention: Demonstration of program efficacy outcomes in a diverse, urban school district. *Journal of Emotional and Behavioral Disorders, 17*(4), 197–212.
- Additional source:**
- Walker, H. M. (2010). *Evidence-based interventions for severe behavior problems, final report* (IES Special Education Annual Performance Report No. CFDA #84324P). Washington, DC: Institute of Education Sciences.

Studies that do not meet WWC evidence standards

- 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. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Zolna, J., Kimmich, M., & Hawkinson, L. (2001). *Final report: Evaluation of the First Step to Success replication*. Unpublished manuscript. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Additional source:**
- Walker, H. M., Golly, A., McLane, J. Z., & Kimmich, M. (2005). The Oregon First Step to Success replication initiative: Statewide results of an evaluation of the program's impact. *Journal of Emotional & Behavioral Disorders, 13*(3), 163–172.

Studies that are ineligible for review using the Children Classified as Having an Emotional Disturbance Evidence Review Protocol

- Epstein, M. H., & Walker, H. M. (2002). Special education: Best practices and First Step to Success. In B. J. Burns & K. Hoagwood (Eds.), *Community treatment for youth: Evidence-based interventions for severe emotional and behavioral disorders* (pp. 179–197). New York: Oxford University Press. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Feil, E. G., Severson, H. H., & Walker, H. M. (2002). Early screening and intervention to prevent the development of aggressive, destructive behavior patterns among at-risk children. In M. R. Shinn, H. M. Walker, & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Preventive and remedial approaches* (pp.143–166). Washington, DC: National Association of School Psychologists. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Golly, A. M., Stiller, B., & Walker, H. M. (1998). First Step to Success: Replication and social validation of an early intervention program. *Journal of Emotional & Behavioral Disorders, 6*(4), 243. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Gunn, B., Feil, E., Seeley, J., Severson, H., & Walker, H. (2006). Promoting school success: Developing social skills and early literacy in Head Start classrooms. *NHSA Dialog: A Research-to-Practice Journal for the Early Intervention Field, 9*(1), 1–11. The study is ineligible for review because it does not use a sample aligned with the protocol—the sample is not within the specified age or grade range.
- Lumsden, L. (2000). Early intervention to prevent violence. *Research Roundup, 17*(1). The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

- Overton, S., McKenzie, L., King, K., & Osborne, J. (2002). Replication of the First Step to Success model: A multiple-case study of implementation effectiveness. *Behavioral Disorders, 28*(1), 40–56. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Additional source:**
- Overton, S., McKenzie, L., King, K., & Osborne, J. (2003). Correction to “Replication of the First Step to Success model: A multiple-case study of implementation effectiveness.” *Behavioral Disorders, 29*(1), 80–81.
- Richards, S. (2008). *Efficacy of First Step to Success and Second Step programs: A review of the literature*. Unpublished master’s thesis, St. Cloud State University, St. Cloud, MN. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Rutherford, R. B. J., & Mathur, S. R. (2005). Introduction to special issue: Severe behavior disorders of children and youth. *Education and Treatment of Children, 28*(4), 325–327. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Seeley, J., Small, J., Walker, H., Feil, E., Severson, H., Golly, A., & Forness, S. (2009). Efficacy of the First Step to Success intervention for students with attention-deficit/hyperactivity disorder. *School Mental Health, 1*(1), 37–48. The study is ineligible for review because it does not use a sample aligned with the protocol—the sample includes less than 50% children described as at risk for or classified as having an emotional disturbance.
- Varghese, R. (2000). Evaluation of a pilot project to identify at-risk children in Head Start. *Dissertation Abstracts International, 61*(5A), 108-1738. The study is ineligible for review because it does not use a sample aligned with the protocol—the sample is not within the specified age or grade range.
- Walker, H., Seeley, J., Golly, A., Severson, H., & Feil, E. (2009). The First Step to Success program for preventing antisocial behavior in young child: Update on past, current and planned research. *Report on Emotional & Behavior Disorders in Children, 8*(1), 17–23. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Walker, H., Stiller, B., & Golly, A. (1998). First Step to Success: A collaborative home-school Intervention for preventing antisocial behavior at the point of school entry. *Young Exceptional Children, 1*(2), 2–6. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Walker, H. M. (2001). Preventing mental disorders in school-aged children: Current state of the field: Commentary. *Prevention & Treatment, 4*(1). The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Walker, H. M. (2002). The First Step to Success program: Preventing destructive social outcomes at the point of school entry. In P. S. Jensen (Ed.), *Report on Emotional & Behavior Disorders in Youth, 3*(1), 3–6, 22–23. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Walker, H. M., & Epstein, M. H. (2001). *Making schools safer and violence free: Critical issues, solutions, and recommended practices*. Austin, TX: PRO-ED. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Walker, H. M., Severson, H. H., Feil, E. G., Stiller, B., & Golly, A. (1998). First Step to Success: Intervening at the point of school entry to prevent antisocial behavior patterns. *Psychology in the Schools, 35*(3), 259–269. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Walker, H. M., Sprague, J. R., Perkins-Rowe, K. A., Beard-Jordan, K. Y., Seibert, B. M., Golly, A. M., . . . Feil, E. G. (2005). The First Step to Success program: Achieving secondary prevention outcomes for behaviorally at-risk children through early intervention. In M. H. Epstein, K. Kutash, & A. J. Duchnowski (Eds.), *Outcomes for children and youth with emotional and behavioral disorders and their families: Programs and evaluation best practices* (2nd ed., pp. 501–523). Austin, TX: PRO-ED. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Appendix A.1: Research details for Walker et al., 1998

Walker, H. M., Kavanagh, K., Stiller, B., Golly, A., Severson, H., & Feil, E. (1998). *First Step to Success. An early intervention approach for preventing school antisocial behavior. Journal of Emotional and Behavioral Disorders, 6(2), 66–80.*

Table A1. Summary of findings

Meets WWC evidence standards

| Outcome domain | Sample size | Study findings | |
|-----------------------------|-------------|--|---------------------------|
| | | Average improvement index (percentile points) | Statistically significant |
| External behavior | 46 students | +34 | Yes |
| Emotional/internal behavior | 46 students | +10 | No |

Setting Study schools were located in the Eugene, Oregon school district.

Study sample Forty-six kindergarten children were randomly assigned to the intervention group (n = 25) or to a wait-list control condition (n = 21). A table of random numbers was used to assign each pool of participants comprising Cohorts 1 and 2 to either an intervention or wait-list control condition. The study included two cohorts of students; Cohort 1 included 24 students who were in kindergarten during the 1993–94 academic year, and Cohort 2 included 22 students who were in kindergarten during the 1994–95 academic year. Participants were 26% female, 7% were of racial/ethnic minorities, and 37% were classified as low income. Students were described as exhibiting antisocial behaviors, including victimizing others, severe tantrums, and aggression.

Intervention group Intervention students were exposed to both the CLASS and HomeBase components of the program. The intervention was delivered by eight trained consultants, in conjunction with the classroom teachers and parents or primary caregivers. HomeBase consisted of six lessons for parents or caregivers to help increase their child’s performance. The consultant visited the home weekly after the 10th day of the CLASS program to conduct the one-hour lesson, which also included parent-child games. All children received the *First Step to Success* intervention over a course of three months.

Comparison group The control condition did not utilize *First Step to Success*. Students assigned to the control group were put on a waiting list and received *First Step to Success* following its termination for participants in the treatment group.

Outcomes and measurement⁹ Four measures of external behavior were assessed immediately following completion of *First Step to Success* in kindergarten. These measures included teacher ratings on the Early Screening Project (ESP) Adaptive and Maladaptive Behavior scales, which are adaptations of the Systematic Screening for Behavior Disorders (SSBD), as well as the Child Behavior Checklist–Teacher Report Form (CBCL-TRF) Aggression Subscale, and a measure of academic engaged time (AET). This study also included the Child Behavior Checklist–Teacher Report Form (CBCL-TRF) Withdrawn Subscale as a measure of emotional/internal behavior. For a more detailed description of these outcome measures, see Appendix B.

Support for implementation

Eight program consultants (graduate students, teachers, school counselors, and teacher aides) were recruited and trained by *First Step to Success* developers to implement the intervention. Each consultant was assigned to two or three children. Training procedures included lectures, videotaped demonstrations, role playing, feedback, and self-evaluation. In the second year, those consultants who chose to participate again were given a refresher training course. New second-year consultants were given intensive training that included using the returning consultants as peer coaches.

Appendix A.2: Research details for Walker et al., 2009

Walker, H. M., Seeley, J. R., Small, J., Severson, H. H, Graham, B. A., Feil, E. G., . . . Forness, S. R. (2009). A randomized controlled trial of the *First Step to Success* early intervention: Demonstration of program efficacy outcomes in a diverse, urban school district. *Journal of Emotional and Behavioral Disorders, 17*(4), 197–212.

Table A2. Summary of findings

Meets WWC evidence standards

| Outcome domain | Sample size ⁵ | Study findings | |
|------------------------------|--------------------------|---|---------------------------|
| | | Average improvement index (percentile points) | Statistically significant |
| External behavior | 197 students | +19 | Yes |
| Social outcomes | 197 students | +23 | Yes |
| Reading achievement/literacy | 193 students | -2 | No |
| Other academic performance | 194 students | +12 | Yes |

Setting

Teachers and students were drawn from 34 elementary schools in Albuquerque Public Schools, New Mexico.

Study sample

A sample of 260 teachers from grades 1–3 in 34 elementary schools were randomly assigned to an intervention or usual care control condition across two cohorts (one for the 2005–06 academic year and the other for 2006–07). Random assignment occurred at classroom level within cohorts. Prior to random assignment, the SSBD was used to identify students who were exhibiting the most severe behavioral concerns within each classroom. The student with the highest average ranking across the SSBD Stage 2 measures was targeted for inclusion in the study; these students were described as exhibiting antisocial behaviors, including victimizing others, severe tantrums, and aggression. Parental consent was obtained for students in 210 of the 260 recruited teachers/classrooms (81%). In cohort 1, parents were more likely to decline participation in the study if their child had been randomized to the comparison condition; thus, the authors randomized a larger proportion of classrooms to the comparison condition in cohort 2 to achieve a balanced design across conditions. Of the 210 consenting students across the two cohorts, approximately half were in classrooms that were randomly assigned to the experimental condition (n = 107) and half were in classrooms that were randomly assigned to the control condition (n = 103). The analysis sample consisted of 101 treatment and 97 control students, although specific sample sizes varied by outcome.¹⁰ Participants were predominantly Hispanic (57%) or Caucasian (24.5%), 73% were males, 70% were eligible for free or reduced-price lunches, and roughly 16% were English language learners.

Intervention group

Intervention students were exposed to both the CLASS and HomeBase components of the program. The HomeBase component was started by the behavioral coach on the 10th day of the intervention and consisted of 6 one-hour home visits by the behavioral coach. The study assessed intervention fidelity, teacher-coach alliance, and student and parent program compliance. Implementation was assessed via expert raters four times, focusing on behavioral coach tasks and the beginning, middle, and end of the teacher phase using a *First Step to Success* checklist. Additional post-intervention fidelity scales and assessment of the alliance among teachers, coaches, and parents also were used. Student compliance was measured by the number of times students successfully completed an intervention session without having to repeat it. Authors did not report concerns pertaining to intervention fidelity.

Comparison group

Control classrooms were described as usual care comparisons.

Outcomes and measurement

The study included a measure of academic engaged time (AET), teacher and parent ratings on the Social Skills Rating System (SSRS) Problem Behavior and Social Skills Subscales, teacher ratings on the Social Skills Rating System (SSRS) Academic Competence Subscale, teacher ratings on the Systematic Screening for Behavior Disorders (SSBD) Adaptive and Maladaptive Behavior Indexes, the Woodcock-Johnson III Diagnostic Reading Battery (WJ-III DRB) Letter-Word Identification Subset and a series of oral reading fluency passages (i.e., average correct words read per minute from a set of passages). For a more detailed description of these outcome measures, see Appendix B.

Support for implementation

Behavior coaches, who implemented the first five days of the classroom portion and all six home visits, attended a two-day training institute. The coaches remained in close contact with supervisory staff and were scheduled for fidelity monitoring checks regularly. The trainer held weekly videoconferences to answer questions and address problems. Parents were trained by the behavioral coaches during the home visits. Teachers were taught how to monitor child behavior, give praise, and provide feedback to parents.

Appendix B: Outcome measures for each domain

| External behavior | |
|---|--|
| <i>Academic Engaged Time (AET)</i> | This outcome assesses the amount of time a student spends engaged in academic activities. In Walker et al. (1998) and Walker et al. (2009), academic engaged time (AET) was collected for each participant during 15-minute classroom observations according to procedures outlined in Stage 3 of the Systematic Screening for Behavior Disorders (SSBD) protocol. The SSBD is a nationally normed, multistage assessment approach developed by Walker and Severson (1990). ¹¹ AET observations measured how often the child was (a) attending to the teacher, (b) making appropriate motor responses (e.g., following directions), (c) asking for assistance in an appropriate manner, (d) cooperating with others, and (e) being appropriately involved in teacher-assigned tasks and activities. A stopwatch was allowed to run as long as the child was academically engaged and stopped whenever he or she was not engaged during the observation period. The time on the stopwatch was divided by the amount of time observed and then multiplied by 100 to derive an AET percentage score (as cited in Walker et al., 1998). |
| <i>Child Behavior Checklist Teacher Report Forms (CBCL-TRF): Aggression Subscale</i> | The Child Behavior Checklist (CBCL) is a nationally normed measure, developed by Achenbach (1991). ¹² The Teacher Report Form (TRF), in conjunction with the CBCL, is one component of the Achenbach System of Empirically Based Assessment (ASEBA). The Aggression Subscale consists of 33 items on which teachers can rate a child’s aggressive problem behaviors (as cited in Walker et al., 1998). |
| <i>Early Screening Project (ESP): Adaptive Behavior Subscale</i> | This is a multi-method screening procedure that integrates teacher ratings and behavioral observations. It was designed by Walker, Severson, and Feil (1995) ¹³ and is described as a downward extension and adaptation of the SSBD. This subscale consists of eight items that measure teachers’ ratings of adaptive behaviors (as cited in Walker et al., 1998). |
| <i>Early Screening Project (ESP): Maladaptive Behavior Subscale</i> | This is a multi-method screening procedure that integrates teacher ratings and behavioral observations. It was designed by Walker et al. (1995) ¹³ and is described as a downward extension and adaptation of the SSBD. This subscale consists of nine items that measure teachers’ ratings of maladaptive behaviors (as cited in Walker et al., 1998). |
| <i>Social Skills Rating System: Problem Behavior Subscale for Parents</i> | The 17-item Problem Behavior Subscale (developed by Gresham & Elliot, 1990) ¹⁴ is a standardized, nationally normed measure that assesses parents’ perceived frequency of internalizing and externalizing problem behavior that may interfere with social skills performance. The problem behavior items are assessed on a three-point scale (as cited in Walker et al., 2009). |
| <i>Social Skills Rating System: Problem Behavior Subscale for Teachers</i> | The 18-item Problem Behavior Subscale (developed by Gresham & Elliot, 1990) ¹⁴ is a standardized, nationally normed measure that assesses teachers’ perceived frequency of internalizing and externalizing problem behavior that may interfere with social skills performance. The problem behavior items are assessed on a three-point scale (as cited in Walker et al., 2009). |
| <i>Systematic Screening for Behavior Disorders (SSBD): Maladaptive Behavior Index (MBI)</i> | The SSBD is a nationally normed, multistage assessment approach developed by Walker and Severson (1990). ¹¹ In Stage 1, teachers select a group of five students who exhibit externalizing behaviors and rank order them in terms of severity. In Stage 2, teachers complete three rating scales, including the Maladaptive Behavior Index. This 11-item subscale assesses student’s teacher-related and peer-to-peer problem behavior symptoms during the past month, measured on a five-point Likert scale (as cited in Walker et al., 2009). |
| Emotional/internal behavior | |
| <i>Child Behavior Checklist–Teacher Report Forms (CBCL-TRF): Withdrawn Subscale</i> | The Child Behavior Checklist (CBCL) is a nationally normed measure, developed by Achenbach (1991). ¹² The Teacher Report Form (TRF), in conjunction with the CBCL, is one component of the Achenbach System of Empirically Based Assessment (ASEBA). The Withdrawn Subscale asks teachers to use a seven-point scale to compare the child to typical students on factors such as the degree to which they feel sad, lack energy, prefer to be alone, and seem withdrawn (as cited in Walker et al., 1998). |
| Social outcomes | |
| <i>Social Skills Rating System (SSRS): Social Skills Subscale for Parents</i> | This standardized, nationally normed 38-item subscale assesses parents’ perceived frequency of child’s social competence in day-to-day activities and interactions at home (as cited in Walker et al., 2009). This subscale was developed by Gresham and Elliot (1990). ¹⁴ |
| <i>Social Skills Rating System (SSRS): Social Skills Subscale for Teachers</i> | This standardized, nationally normed 30-item subscale assesses cooperation, assertion, and self-control as reported by teachers on a three-point scale (as cited in Walker et al., 2009). This subscale was developed by Gresham and Elliot (1990). ¹⁴ |

| | |
|--|---|
| <i>Systematic Screening for Behavior Disorders (SSBD): Adaptive Behavior Index</i> | The SSBD is a nationally normed, multistage assessment approach developed by Walker and Severson (1990). ¹¹ This 12-item subscale assesses functional social impairment (as cited in Walker et al., 2009). |
|--|---|

Reading achievement/literacy

| | |
|-----------------------------|--|
| <i>Oral Reading Fluency</i> | This measure assesses students' oral reading fluency while reading series of written passages. The number of correct words read per minute is calculated and averaged across passages to obtain a total score for a given measurement time point (as cited in Walker et al., 2009). In Walker et al. (2009), two passages were administered by the trained assessor. |
|-----------------------------|--|

| | |
|--|---|
| <i>Woodcock-Johnson III Diagnostic Reading Battery (WJ-III DRB): Letter-Word Identification Subset</i> | This is a standardized subtest from the Woodcock-Johnson Tests of Achievement that assesses a student's word reading skills. This subtest measures a student's ability to identify isolated letters and words (Woodcock, Mather, & Schrank, 2004 ¹⁵) (as cited in Walker et al., 2009). |
|--|---|

Other academic performance

| | |
|---|--|
| <i>Social Skills Rating System (SSRS): Academic Competence Subscale</i> | The SSRS is a standardized, nationally normed measure of social skills. The nine-item Academic Competence Subscale assesses reading and math performance, motivation, intellectual functioning, and parental support as estimated by the teacher on a five-point percentage cluster scale, from the lowest 10% to the highest 10% (as cited in Walker et al., 2009). |
|---|--|

Appendix C.1: Findings included in the rating for the external behavior domain

| Outcome measure | Study sample | Sample size | Mean (standard deviation) | | WWC calculations | | | p-value |
|---|---------------|-------------|---------------------------|------------------|------------------|-------------|-------------------|----------------------------------|
| | | | Intervention group | Comparison group | Mean difference | Effect size | Improvement index | |
| Walker et al., 1998^a | | | | | | | | |
| <i>Academic Engaged Time Kindergarten</i> | Kindergarten | 46 | 87.32 (12.54) | 69.05 (20.44) | 18.27 | 0.97 | +33 | < 0.05 |
| <i>CBCL-TRF Aggression</i> | Kindergarten | 46 | 13.08 (9.42) | 23.71 (9.35) | 10.63 | 0.99 | +34 | < 0.001 |
| <i>ESP Adaptive Behavior</i> | Kindergarten | 46 | 28.8 (4.19) | 22.24 (5.00) | 6.56 | 1.17 | +38 | < 0.001 |
| <i>ESP Maladaptive Behavior</i> | Kindergarten | 46 | 23.52 (8.70) | 31.86 (7.13) | 8.34 | 0.93 | +32 | < 0.001 |
| Domain average for external behavior (Walker et al., 1998) | | | | | | 1.02 | +34 | Statistically significant |
| Walker et al., 2009^b | | | | | | | | |
| <i>Academic Engaged Time</i> | Grades 1 to 3 | 196 | 56.8 (19.6) | 48.6 (22.4) | 8.2 | 0.39 | +15 | 0.01 |
| <i>SSRS Problem Behavior: Parent</i> | Grades 1 to 3 | 186 | 103.5 (13.8) | 110.3 (13.3) | 6.8 | 0.50 | +19 | 0.01 |
| <i>SSRS Problem Behavior: Teacher</i> | Grades 1 to 3 | 194 | 112.6 (12.6) | 119.8 (10.9) | 7.2 | 0.61 | +23 | 0.001 |
| <i>SSBD Maladaptive Behavior</i> | Grades 1 to 3 | 197 | 25.7 (9.4) | 30.4 (9.3) | 4.7 | 0.50 | +19 | 0.001 |
| Domain average for external behavior (Walker et al., 2009) | | | | | | 0.50 | +19 | Statistically significant |
| Domain average for external behavior across all studies | | | | | | 0.76 | +28 | na |

Table Notes: This appendix reports findings considered for the effectiveness rating and the average improvement indices for the external behavior domain. Positive results for mean difference, effect size, and improvement index favor the intervention group; negative results favor the comparison group. For the CBCL, ESP-Maladaptive, SSRS, and SSBD outcomes, signs were reversed on the mean difference, effect size, and improvement index to demonstrate that the treatment group was favored when negative differences were reported (to clarify, lower scores on these measures indicated fewer problems). The effect size is a standardized measure of the effect of an intervention on student outcomes, representing the change (measured in standard deviations) in an average student's outcome that can be expected if that student is given the intervention. The improvement index is an alternate presentation of the effect size, reflecting the change in an average student's percentile rank that can be expected if the student is given the intervention. The WWC-computed average effect size is a simple average rounded to two decimal places; the average improvement index is calculated from the average effect size. The statistical significance of each study's domain average was determined by the WWC; a study is characterized as having a statistically significant positive effect when univariate statistical tests are reported for each outcome measure, the effect for at least one measure within the domain is positive and statistically significant, and no effects are negative and statistically significant. CBCL-TRF Aggression = Child Behavior Checklist-Teacher Report Forms, Aggression Subscale. ESP Adaptive Behavior = Early Screening Project: Adaptive Behavior Subscale. ESP Maladaptive Behavior = Early Screening Project: Maladaptive Behavior Subscale. SSRS Problem Behavior: Parent = Social Skills Rating System: Problem Behavior Subscale for Parents. SSRS Problem Behavior: Teacher = Social Skills Rating System: Problem Behavior Subscale for Teachers. SSBD Maladaptive Behavior = Systematic Screening for Behavior Disorders: Maladaptive Behavior Index. na = not applicable.

^a The intervention and control group means from Walker et al. (1998) are ANCOVA-adjusted posttest scores reported by the authors in the article. The effect sizes presented here were reported by the authors in the paper and were calculated using the pooled standard deviation in the denominator. A correction for multiple comparisons was needed but did not affect significance levels. The p-values presented here were reported in the original study.

^b Walker et al. (2009) imputed missing values of the outcome measures using the expectation-maximization method; missing data were imputed for 40 cases (20% of the sample). The intervention and control group means reported here are ANCOVA-adjusted posttest scores for the non-imputed sample, based on information obtained via an author query. The effect sizes were calculated by the WWC, based on data from the non-imputed sample provided by the authors. The magnitude and statistical significance of the impact estimates were similar to the findings based on the imputed data that were reported in the original study. A correction for multiple comparisons was needed but did not affect significance levels. The p-values presented here were calculated by the WWC. The results for the SSRS Problem Behavior: Parent measure meet evidence standards with reservations, due to high attrition.

Appendix C.2: Findings included in the rating for the emotional/internal behavior domain

| Outcome measure | Study sample | Sample size | Mean (standard deviation) | | WWC calculations | | | p-value |
|---|--------------|-------------|---------------------------|------------------|------------------|-------------|-------------------|--------------------------------------|
| | | | Intervention group | Comparison group | Mean difference | Effect size | Improvement index | |
| Walker et al., 1998^a | | | | | | | | |
| <i>CBCL-TRF Withdrawn</i> | Kindergarten | 46 | 3.08 (3.39) | 4.09 (4.32) | 1.01 | 0.26 | +10 | 0.63 |
| Domain average for emotional/internal behavior (Walker et al., 1998) | | | | | | 0.26 | +10 | Not statistically significant |
| Domain average for emotional/internal behavior across all studies | | | | | | 0.26 | +10 | na |

Table Notes: This appendix reports findings considered for the effectiveness rating and the average improvement indices for the emotional/internal behavior domain. Positive results for mean difference, effect size, and improvement index favor the intervention group; negative results favor the comparison group. Signs were reversed for CLCL-TRF: Withdrawn on the mean difference, effect size, and improvement index to demonstrate that the treatment group was favored when a negative difference was reported. The effect size is a standardized measure of the effect of an intervention on student outcomes, representing the change (measured in standard deviations) in an average student’s outcome that can be expected if that student is given the intervention. The improvement index is an alternate presentation of the effect size, reflecting the change in an average student’s percentile rank that can be expected if the student is given the intervention. CBCL-TRF Withdrawn = Child Behavior Checklist–Teacher Report Forms: Withdrawn Subscale. na = not applicable.

^a The intervention and control group means from Walker et al. (1998) are ANCOVA-adjusted posttest scores reported by the authors in the article. The effect sizes were reported by the authors in the paper and were calculated using the pooled standard deviation in the denominator. No corrections for clustering or multiple comparisons were needed. The p-values presented here were reported in the original study.

Appendix C.3: Findings included in the rating for the social outcomes domain

| Outcome measure | Study sample | Sample size | Mean (standard deviation) | | WWC calculations | | | p-value |
|---|---------------|-------------|---------------------------|------------------|------------------|-------------|-------------------|----------------------------------|
| | | | Intervention group | Comparison group | Mean difference | Effect size | Improvement index | |
| Walker et al., 2009^a | | | | | | | | |
| <i>SSRS Social Skills: Parent</i> | Grades 1 to 3 | 186 | 97.6 (15.9) | 91.8 (15.3) | 5.8 | 0.37 | +14 | 0.01 |
| <i>SSRS Social Skills: Teacher</i> | Grades 1 to 3 | 189 | 95.0 (14.3) | 85.6 (8.8) | 9.4 | 0.78 | +28 | 0.01 |
| <i>SSBD Adaptive Behavior</i> | Grades 1 to 3 | 197 | 41.0 (9.0) | 35.0 (7.5) | 6.0 | 0.72 | +26 | 0.001 |
| Domain average for social outcomes (Walker et al., 2009) | | | | | | 0.62 | +23 | Statistically significant |
| Domain average for social outcomes across all studies | | | | | | 0.62 | +23 | na |

Table Notes: This appendix reports findings considered for the effectiveness rating and the average improvement indices for the social outcomes domain. Positive results for mean difference, effect size, and improvement index favor the intervention group; negative results favor the comparison group. The effect size is a standardized measure of the effect of an intervention on student outcomes, representing the change (measured in standard deviations) in an average student’s outcome that can be expected if that student is given the intervention. The improvement index is an alternate presentation of the effect size, reflecting the change in an average student’s percentile rank that can be expected if the student is given the intervention. The WWC-computed average effect size is a simple average rounded to two decimal places; the average improvement index is calculated from the average effect size. The statistical significance of the study’s domain average was determined by the WWC; a study is characterized as having a statistically significant positive effect when univariate statistical tests are reported for each outcome measure, the effect for at least one measure within the domain is positive and statistically significant, and no effects are negative and statistically significant. SSRS Social Skills: Parent = Social Skills Rating System: Social Skills Subscale for Parents. SSRS Social Skills: Teacher = Social Skills Rating System: Social Skills Subscale for Teachers. SSBD Adaptive Behavior = Systematic Screening for Behavior Disorders: Adaptive Behavior Inventory. na = not applicable.

^a Walker et al. (2009) imputed missing values of the outcome measures using the expectation-maximization method; missing data were imputed for 40 cases (20% of the sample). The intervention and control group means reported here are ANCOVA-adjusted posttest scores for the non-imputed sample, based on information obtained via an author query. The effect sizes were calculated by the WWC, based on data from the non-imputed sample provided by the authors. The magnitude and statistical significance of the impact estimates were similar to the findings based on the imputed data that were reported in the original study. A correction for multiple comparisons was needed but did not affect significance levels. The p-values presented here were calculated by the WWC. Results for the SSRS Social Skills Subscale for Parents and SSRS Social Skills Subscale for Teachers meet evidence standards with reservations, due to high attrition.

Appendix C.4: Findings included in the rating for the reading achievement/literacy domain

| Outcome measure | Study sample | Sample size | Mean (standard deviation) | | WWC calculations | | | p-value |
|--|---------------|-------------|---------------------------|------------------|------------------|--------------|-------------------|--------------------------------------|
| | | | Intervention group | Comparison group | Mean difference | Effect size | Improvement index | |
| Walker et al., 2009^a | | | | | | | | |
| <i>Oral Reading Fluency</i> | Grades 1 to 3 | 190 | 60.5 (43.5) | 58.9 (38.9) | 1.6 | 0.04 | +2 | 0.79 |
| <i>WJ-III DRB Letter-Word Identification</i> | Grades 1 to 3 | 193 | 99.5 (12.8) | 101.2 (16.4) | -1.7 | -0.12 | -5 | 0.42 |
| Domain average for reading achievement/literacy (Walker et al., 2009) | | | | | | -0.04 | -2 | Not statistically significant |
| Domain average for reading achievement/literacy across all studies | | | | | | -0.04 | -2 | na |

Table Notes: This appendix reports findings considered for the effectiveness rating and the average improvement indices for the reading achievement/literacy domain. Positive results for mean difference, effect size, and improvement index favor the intervention group; negative results favor the comparison group. The effect size is a standardized measure of the effect of an intervention on student outcomes, representing the change (measured in standard deviations) in an average student’s outcome that can be expected if that student is given the intervention. The improvement index is an alternate presentation of the effect size, reflecting the change in an average student’s percentile rank that can be expected if the student is given the intervention. The WWC-computed average effect size is a simple average rounded to two decimal places; the average improvement index is calculated from the average effect size. The statistical significance of the study’s domain average was determined by the WWC; a study’s effect is characterized as not statistically significant when univariate statistical tests are reported for each outcome measure and each of the effects within the domain are not statistically significant. WJ-III DRB Letter-Word Identification = Woodcock-Johnson III Diagnostic Reading Battery: Letter-Word Identification Subset. na = not applicable.

^a Walker et al. (2009) imputed missing values of the outcome measures using the expectation-maximization method; missing data were imputed for 40 cases (20% of the sample). The intervention and control group means reported here are ANCOVA-adjusted posttest scores for the non-imputed sample, based on information obtained via an author query. The effect sizes were calculated by the WWC, based on data from the non-imputed sample provided by the authors. The magnitude and statistical significance of the impact estimates were similar to the findings based on the imputed data that were reported in the original study. A correction for multiple comparisons was needed but did not affect significance levels. The p-values presented here were calculated by the WWC. Results for both outcomes in this domain meet evidence standards with reservations, due to high attrition.

Appendix C.5: Findings included in the rating for the other academic performance domain

| Outcome measure | Study sample | Sample size | Mean (standard deviation) | | WWC calculations | | | p-value |
|--|---------------|-------------|---------------------------|------------------|------------------|-------------|-------------------|----------------------------------|
| | | | Intervention group | Comparison group | Mean difference | Effect size | Improvement index | |
| Walker et al., 2009^a | | | | | | | | |
| <i>SSRS Academic Competence</i> | Grades 1 to 3 | 194 | 91.1 (10.5) | 91.1 (10.5) | 3.4 | 0.32 | +12 | .03 |
| Domain average for other academic performance (Walker et al., 2009) | | | | | | 0.32 | +12 | Statistically significant |
| Domain average for other academic performance across all studies | | | | | | 0.32 | +12 | na |

Table Notes: This appendix reports findings considered for the effectiveness rating and the average improvement indices for the other academic performance domain. Positive results for mean difference, effect size, and improvement index favor the intervention group; negative results favor the comparison group. The effect size is a standardized measure of the effect of an intervention on student outcomes, representing the change (measured in standard deviations) in an average student’s outcome that can be expected if that student is given the intervention. The improvement index is an alternate presentation of the effect size, reflecting the change in an average student’s percentile rank that can be expected if the student is given the intervention. SSRS Academic Competence = Social Skills Rating System: Academic Competence Subscale. na = not applicable.

^a Walker et al. (2009) imputed missing values of the outcome measures using the expectation-maximization method; missing data were imputed for 40 cases (20% of the sample). The intervention and control group means reported here are ANCOVA-adjusted posttest scores for the non-imputed sample, based on information obtained via an author query. The effect size was calculated by the WWC, based on data from the non-imputed sample provided by the authors. The magnitude and statistical significance of the impact estimate was similar to the finding based on the imputed data that was reported in the original study. No corrections for clustering or multiple comparisons were needed. The p-value presented here was calculated by the WWC.

Appendix D: Single-Case Design studies reviewed for this intervention

| Study citation | Study disposition |
|--|---|
| <p>Beard, K. Y., & Sugai, G. (2004). First Step to Success: An early intervention for elementary children at risk for antisocial behavior. <i>Behavioral Disorders, 29</i>(4), 396–409.</p> | <p>Meets WWC pilot Single-Case Design standards</p> |
| <p>Lien-Thorne, S., & Kamps, D. (2005). Replication study of the First Step to Success early intervention program. <i>Behavioral Disorders, 31</i>(1), 18–32.</p> | <p>Meets WWC pilot Single-Case Design standards</p> |
| <p>Sprague, J., & Perkins, K. (2009). Direct and collateral effects of the First Step to Success program. <i>Journal of Positive Behavior Interventions, 11</i>(4), 208–221. doi:10.1177/1098300708330935. Additional source: Perkins Rowe, K. A. (2002). Direct and collateral effects of the First Step to Success program: Replication and extension of findings. <i>Dissertation Abstracts International, 62</i>(12-A), 4058.</p> | <p>Meets WWC pilot Single-Case Design standards</p> |
| <p>Diken, I. H., & Rutherford, R. B. (2005). First Step to Success early intervention program: A study of effectiveness with Native-American children. <i>Education & Treatment of Children, 28</i>(4), 444–465. Additional source: Diken, I. H. (2004). First Step to Success early intervention program: A study of effectiveness with children in an American-Indian nation community school. <i>Dissertation Abstracts International, 65</i>(2A), 158–464.</p> | <p>The study does not meet WWC pilot Single-Case Design standards because it does not have at least three attempts to demonstrate an intervention effect at three different points in time.</p> |
| <p>Golly, A., Sprague, J., Walker, H., Beard, K., & Gorham, G. (2000). The First Step to Success program: An analysis of outcomes with identical twins across multiple baselines. <i>Behavioral Disorders, 25</i>(3), 170.</p> | <p>The study does not meet WWC pilot Single-Case Design standards because it does not have at least three attempts to demonstrate an intervention effect at three different points in time.</p> |

Table Notes: The supplemental studies presented in this table do not factor into the determination of the intervention rating.

Endnotes

- ¹ The descriptive information for this program was obtained from a publicly available source: the program's website (<http://www.firststeptosuccess.org>, downloaded March 2010). The WWC requests that developers review the program description sections for accuracy from their perspective. The program description was provided to the developer in March 2010 and we incorporated feedback from the developer. Further verification of the accuracy of the descriptive information for this program is beyond the scope of this review. The literature search for this report includes studies publicly available by August 2011.
- ² The studies in this report were reviewed using WWC Evidence Standards, Version 2.1, as described in protocol Version 2.0. The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.
- ³ Results for both outcomes in the reading achievement/literacy domain meet evidence standards with reservations, due to high attrition.
- ⁴ For criteria used in the determination of the rating of effectiveness and extent of evidence, see the WWC Rating Criteria on p. 19 of this report. These improvement index numbers show the average and range of student-level improvement indices for all findings across the studies. Domains covered in the protocol that are not examined by the studies that meet standards are math achievement and school attendance.
- ⁵ Walker et al. (2009) included 34 schools and Walker et al. (1998) included an unknown number of schools. One study (Walker et al., 2009) reported imputed values for outcome measures, using the expectation-maximization method; missing data were imputed for 40 cases (20% of the sample). Non-imputed sample sizes are reported here; this information was obtained via the author. These non-imputed sample sizes varied across outcomes within each domain; the maximum sample size for each domain is reported here.
- ⁶ Cost information is based on findings from Walker, H. M., Golly, A., McLane, J. Z., & Kimmich, M. (2005). The Oregon First Step to Success replication initiative: Statewide results of an evaluation of program's impact. *Journal of Emotional and Behavioral Disorders*, 13(3), 163–172.
- ⁷ Both cohorts of students were assessed again in grade 1, and Cohort 1 students were assessed again in grade 2. Walker et al. (1998) did not report treatment and control comparisons for the grade 1 and grade 2 follow-up data points, so these outcomes are not included in this review.
- ⁸ Posttest sample sizes were not provided for some outcomes, and the WWC obtained this information directly from the study authors. Attrition varied across outcomes, but high levels of attrition occurred for five outcomes: one outcome in the external behavior domain (SSRS Problem Behavior Subscale for Parents), two outcomes in the social outcomes domain (SSRS Social Skills Subscale for Parents and SSRS Social Skills Subscale for Teachers), and both outcomes in the reading achievement/literacy domain (Oral Reading Fluency and WJ-III DRB Letter-Word Identification Subset). The authors also provided information that demonstrated equivalence of the analytic samples based on baseline measures of these outcomes; thus, results for these outcomes meet evidence standards with reservations.
- ⁹ Both cohorts of students were assessed immediately following completion of *First Step to Success* in kindergarten and again in grade 1. Cohort 1 students were assessed again in grade 2. Walker et al. (1998) did not report treatment and control comparisons for the grade 1 and grade 2 follow-up data points.
- ¹⁰ In Walker et al. (2009), missing values of the outcome measures were imputed using the expectation-maximization method; missing data were imputed for 40 cases (20% of the sample). We report on the non-imputed sample here, based on tables obtained via an author query.
- ¹¹ Walker, H. M., & Severson, H. H. (1990). *Systematic Screening for Behavior Disorders (SSBD): Users guide and technical manual*. Longmont, CO: Sopris West.
- ¹² Achenbach, T. (1991). *The Child Behavior Checklist: Manual for the teacher's report form*. Burlington: University of Vermont, Department of Psychiatry.
- ¹³ Walker, H. M., Severson, H., & Feil, E. (1995). *The Early Screening Project: A proven child-find process*. Longmont, CO: Sopris West.
- ¹⁴ Gresham, F. M., & Elliott, S. N. (1990). *Social Skills Rating System Manual*. Circle Pines, MN: American Guidance Service.
- ¹⁵ Woodcock, R. W., Mather, N., & Schrank, F. (2004). *WJ III Diagnostic Reading Battery*. Rolling Meadows, IL: Riverside.

Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2012, March). *Children Classified as Having an Emotional Disturbance intervention report: First Step to Success*. Retrieved from <http://whatworks.ed.gov>.

WWC Rating Criteria

Criteria used to determine the rating of a study

| Study rating | Criteria |
|---|--|
| Meets evidence standards | A study that provides strong evidence for an intervention's effectiveness, such as a well-implemented RCT. |
| Meets evidence standards with reservations | A study that provides weaker evidence for an intervention's effectiveness, such as a QED or an RCT with high attrition that has established equivalence of the analytic samples. |

Criteria used to determine the rating of effectiveness for an intervention

| Rating of effectiveness | Criteria |
|-------------------------------------|--|
| Positive effects | Two or more studies show statistically significant positive effects, at least one of which met WWC evidence standards for a strong design, AND No studies show statistically significant or substantively important negative effects. |
| Potentially positive effects | At least one study shows a statistically significant or substantively important positive effect, AND No studies show a statistically significant or substantively important negative effect AND fewer or the same number of studies show indeterminate effects than show statistically significant or substantively important positive effects. |
| Mixed effects | At least one study shows a statistically significant or substantively important positive effect AND at least one study shows a statistically significant or substantively important negative effect, but no more such studies than the number showing a statistically significant or substantively important positive effect, OR At least one study shows a statistically significant or substantively important effect AND more studies show an indeterminate effect than show a statistically significant or substantively important effect. |
| Potentially negative effects | One study shows a statistically significant or substantively important negative effect and no studies show a statistically significant or substantively important positive effect, OR Two or more studies show statistically significant or substantively important negative effects, at least one study shows a statistically significant or substantively important positive effect, and more studies show statistically significant or substantively important negative effects than show statistically significant or substantively important positive effects. |
| Negative effects | Two or more studies show statistically significant negative effects, at least one of which met WWC evidence standards for a strong design, AND No studies show statistically significant or substantively important positive effects. |
| No discernible effects | None of the studies shows a statistically significant or substantively important effect, either positive or negative. |

Criteria used to determine the extent of evidence for an intervention

| Extent of evidence | Criteria |
|------------------------|--|
| Medium to large | The domain includes more than one study, AND The domain includes more than one school, AND The domain findings are based on a total sample size of at least 350 students, OR, assuming 25 students in a class, a total of at least 14 classrooms across studies. |
| Small | The domain includes only one study, OR The domain includes only one school, OR The domain findings are based on a total sample size of fewer than 350 students, AND, assuming 25 students in a class, a total of fewer than 14 classrooms across studies. |

Glossary of Terms

| | |
|--|--|
| Attrition | Attrition occurs when an outcome variable is not available for all participants initially assigned to the intervention and comparison groups. The WWC considers the total attrition rate and the difference in attrition rates across groups within a study. |
| Clustering adjustment | If treatment assignment is made at a cluster level and the analysis is conducted at the student level, the WWC will adjust the statistical significance to account for this mismatch, if necessary. |
| Confounding factor | A confounding factor is a component of a study that is completely aligned with one of the study conditions, making it impossible to separate how much of the observed effect was due to the intervention and how much was due to the factor. |
| Design | The design of a study is the method by which intervention and comparison groups were assigned. |
| Domain | A domain is a group of closely related outcomes. |
| Effect size | The effect size is a measure of the magnitude of an effect. The WWC uses a standardized measure to facilitate comparisons across studies and outcomes. |
| Eligibility | A study is eligible for review and inclusion in this report if it falls within the scope of the review protocol and uses either an experimental or matched comparison group design. |
| Equivalence | A demonstration that the analysis sample groups are similar on observed characteristics defined in the review area protocol. |
| Extent of evidence | An indication of how much evidence supports the findings. The criteria for the extent of evidence levels are given in the WWC Rating Criteria earlier in this report. |
| Improvement index | Along a percentile distribution of students, the improvement index represents the gain or loss of the average student due to the intervention. As the average student starts at the 50th percentile, the measure ranges from -50 to +50. |
| Multiple comparison adjustment | When a study includes multiple outcomes or comparison groups, the WWC will adjust the statistical significance to account for the multiple comparisons, if necessary. |
| Quasi-experimental design (QED) | A quasi-experimental design (QED) is a research design in which subjects are assigned to treatment and comparison groups through a process that is not random. |
| Randomized controlled trial (RCT) | A randomized controlled trial (RCT) is an experiment in which investigators randomly assign eligible participants into treatment and comparison groups. |
| Rating of effectiveness | The WWC rates the effects of an intervention in each domain based on the quality of the research design and the magnitude, statistical significance, and consistency in findings. The criteria for the ratings of effectiveness are given in the WWC Rating Criteria earlier in this report. |
| Single-case design | A research approach in which an outcome variable is measured repeatedly within and across different conditions that are defined by the presence or absence of an intervention. |
| Standard deviation | The standard deviation of a measure shows how much variation exists across observations in the sample. A low standard deviation indicates that the observations in the sample tend to be very close to the mean; a high standard deviation indicates that the observations in the sample tend to be spread out over a large range of values. |
| Statistical significance | Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The WWC labels a finding statistically significant if the likelihood that the difference is due to chance is less than 5% ($p < 0.05$). |
| Substantively important | A substantively important finding is one that has an effect size of 0.25 or greater, regardless of statistical significance. |

Please see the [WWC Procedures and Standards Handbook \(version 2.1\)](#) for additional details.