

WWC Review of the Report “Assessing the Effectiveness of First Step to Success: Are Short-term Results the First Step to Long-term Behavioral Improvements?”¹

The findings from this review do not reflect the full body of research evidence on *First Step to Success*.

What is this study about?

The study examined the effects of *First Step to Success (First Step)*, a school- and home-based program intended to improve outcomes for students with moderate to severe behavior problems who may be at risk for academic failure. The study took place in six locations across five states: San Jose, California; Tampa, Florida; Cook County, Illinois; Eugene and Springfield, Oregon; and Huntington, West Virginia.

Researchers randomly assigned 48 elementary schools either to receive the *First Step* program or to continue implementing regular instruction. Within each of the 48 schools, researchers used teacher-administered behavioral assessments to identify students who were eligible for the study. The three students with the highest scores on a systematic screening procedure used to measure externalizing behavior were identified for inclusion in the study, and additional high-ranked students were approached if any of the top three students did not provide consent.² The final analysis sample contained between 117 and 134 students in the intervention condition and between 125 and 139 students in the comparison condition, depending on the outcome.

Students in the intervention condition received the *First Step* intervention, a 3-month program that consists of a universal screening, classroom-based

Features of *First Step to Success (First Step)*

First Step is a 3-month school- and home-based program for elementary school students who have moderate to severe behavior problems. This is a secondary level intervention, one that is used when students do not respond to primary, schoolwide universal behavior strategies. The program consists of a universal screening process, a classroom-based intervention, and an in-home parent education program called homeBase. Trained behavior coaches model the program in classrooms for teachers and work with parents to support the lessons at home. The teacher provides daily progress reports to the parents on their student's behavior, and parents are expected to reinforce positive behavior at home with rewarding activities for the student.

behavior coaching, and an at-home parent education program. Students in the comparison group received business-as-usual services.

Study authors measured the effects of *First Step* by comparing parent, teacher, and researcher assessments of student behavior for students in the intervention and comparison groups. Results for three measures are presented in this WWC report: (a) academic engaged time, defined as the proportion of time a student is academically involved, (b) problem behavior, and (c) academic competence.³

What did the study find?

The study authors reported that *First Step* increased student academic engaged time, increased teacher assessment of academic competence, and had no impact on parental assessment of problem behavior. Using unimputed data provided by the authors in response to a query, the WWC determined that none of the effects from analyses that met standards were statistically significant.⁴ However, the effect size for academic engaged time was determined to be substantively important (greater than 0.25 standard deviations).

WWC Rating

The research described in this report meets WWC evidence standards with reservations

Strengths: The study is a randomized controlled trial.

Cautions: While schools were randomized to the intervention and comparison conditions, the students who were selected to participate in the study may have differed systematically across the intervention and comparison schools. Teachers' selection of students for the study and parents' consent for the study both occurred after randomization and, therefore, teachers' selection and parental consent could have been affected by knowledge of the school's research condition. Because of these selection and consent issues, the study was reviewed as a quasi-experimental design by the WWC. The study demonstrated baseline equivalence of the analysis samples for the three outcomes presented in this WWC report and, therefore, this evidence meets WWC standards with reservations. There were seven additional outcomes that did not meet WWC standards.

Appendix A: Study details

Sumi, W. C., Woodbridge, M. W., Javitz, H. S., Thornton, S. P., Wagner, M., Rouspil, K., Yu, J. W., Seeley, J. R., Walker, H. M., Golly, A., Small, J. W., Feil, H. G., & Severson, H. H. (2013). Assessing the effectiveness of First Step to Success: Are short-term results the first step to long-term behavioral improvements? *Journal of Emotional and Behavioral Disorders, 21*(1), 66–78. Retrieved from <http://ebx.sagepub.com/content/early/2012/01/18/1063426611429571>

Setting The study was conducted in six locations across five states: San Jose, California; Tampa, Florida; Cook County, Illinois; Eugene and Springfield, Oregon; and Huntington, West Virginia.

Study sample A total of 48 elementary schools were randomized either to receive *First Step* or to continue regular instruction. Twenty schools in two sites entered the study in the first cohort (2006–07 school year), and 28 schools from the remaining three sites entered in a second cohort (2007–08 school year). After schools were randomly assigned to a research condition, three teachers at each school (one in each of grades 1–3) identified the top three to five students in their classrooms who had the highest levels of externalizing behavior. The teachers then used Stage 2 of the Systematic Screening for Behavior Disorders (SSBD) process to rank-order the three students with the highest levels of externalizing behavior. In Stage 2 of the SSBD, teachers completed the Adaptive Behavior Index (ABI), the Maladaptive Behavior Index (MBI), and the Critical Events Index (CEI). The study requested parental consent for the student in each grade with the highest ranking on the MBI, ABI, and CEI, and if consent was not provided, the research team would request consent from the next rank-ordered student. The final sample consisted of students who had parental consent to participate in the study. Overall, 77% of the study participants were male, and 73% of the students were eligible to receive free or reduced-price lunch. Forty-five percent of the students were White, 27% were Hispanic, and 23% were African American.

There is evidence that the selection of students for participation in the study may have compromised the initial random assignment procedure. First, teachers knew of their research condition before identifying students, and this may have influenced how they rated students. This is especially a concern for the second round of implementation that occurred in the spring semester in each school because teachers may have developed a perception about the effectiveness of *First Step* based on the fall semester, which may have influenced how they rated students. Second, 22 teachers refused to complete the CEI after identifying potentially eligible students (these students were excluded from the analysis sample). Third, once students were identified as eligible, parents were asked to consent knowing their child's treatment status, and there were differential consent rates across intervention and comparison students. For all of these reasons, this study is reviewed by the WWC as a quasi-experimental design.

Intervention group

At intervention schools, students received *First Step*, a 3-month intervention that consists of three linked components: a universal screening process, a classroom-based intervention, and an in-home parent education program called homeBase. Trained behavior coaches provide coaching and modeling support to teachers and parents to teach students replacement behaviors and to distribute rewards when students behave appropriately and consistently. The coach works with the student in the classroom for the first week, after which the teacher takes over. Coaches provide students with visual cues in class (red or green cards) that indicate whether the student is on-task or not. If students gain enough points, they are able to select an enjoyable activity for the class. Coaches also provide parents with six weekly lessons on parenting skills, with a focus on encouraging school-home problem solving. The teacher provides daily progress reports to the parents on their student's behavior, and parents are expected to reinforce positive behavior at home with rewarding activities for the student.

Comparison group

Students in comparison schools received their business-as-usual instruction and services.

Outcomes and measurement

This WWC report focuses on three outcomes examined in the study across two domains. Each of these outcomes had different analysis samples because of differences in the completion rates for each assessment at follow-up, so baseline equivalence was assessed separately for each outcome. The three outcomes described in this WWC report include AET, the Problem Behavior (PB) subscale of the Social Skills Rating System (SSRS) for Parents, and the Academic Competence (AC) subscale of the SSRS for Teachers. For a more detailed description of these outcome measures, see Appendix B.

Support for implementation

The intervention coaches who supported teachers and implemented the intervention in children's homes received a 2-day training provided by a program developer for *First Step*. These coaches worked directly with the targeted students for the first 5 days of the intervention and provided modeling and consultation to classroom teachers and peers during the remaining 8 weeks of the classroom-based component of the intervention. Technical assistance support was offered by email or a conference call to teachers and coaches.

Reason for review

This study was identified for review by the WWC because it was supported by a grant to SRI International (Principal Investigator: Mary Wagner) from the National Center for Special Education Research (NCSEER) at the Institute of Education Sciences (IES).

Appendix B: Outcome measures for each domain

External behavior	
<i>Academic engaged time (AET)</i>	This measure is based on the proportion of time that a student is academically involved during two 15-minute observation periods. Academic involvement consists of (a) paying attention to the material or task, (b) acting appropriately, (c) requesting assistance appropriately, (d) interacting with the teacher or classmates about the material or task, and (e) listening to teacher instructions. The inter-rater reliability for this measure was 0.80.
<i>Problem Behavior (PB) subscale of the Social Skills Rating System (SSRS) for Parents</i>	This 18-item subscale measures internalizing and externalizing behaviors. The internal consistency of this measure was 0.88.
Other academic performance	
<i>Academic Competence (AC) subscale for the SSRS for Teachers</i>	This 9-item subscale measures a teacher's perceptions of a student's academic competence. The items ask about a student's academic performance in reading and math, motivation and intellectual functioning, and parental support in comparison to other students in the classroom. The internal consistency of this measure was 0.91.

Table Notes: Seven additional outcomes from four domains are not included in this SSR because they did not meet WWC evidence standards with or without reservations (the analysis samples for these seven outcomes were not shown to be equivalent at baseline). Two of these outcomes were in the external behavior domain: the measure of maladaptive behavior from the SSBD and the Problem Behavior subscale of the SSRS for Teachers; three outcomes were in the social outcomes domain: the Social Skills subscale of the SSRS for Parents and for Teachers and the adaptive behavior index from the SSBD; and two outcomes were in the reading achievement/literacy domain: the Woodcock-Johnson III Letter-Word Identification and the Oral Reading Fluency test.

Appendix C: Study findings for each domain

Domain and outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
External behavior								
<i>Academic engaged time (AET)</i>	Full sample	48 schools/ 262 students	0.72 (0.18)	0.67 (0.19)	0.05	0.28	+11	0.02
<i>Problem Behavior (PB) subscale of the Social Skills Rating System (SSRS) for Parents</i>	Full sample	48 schools/ 242 students	-109.10 (14.86)	-111.98 (13.16)	2.88	0.20	+8	> 0.05
Domain average for external behavior						0.24	+10	Not statistically significant
Other academic performance								
<i>Academic Competence (AC) subscale of the SSRS for Teachers</i>	Full sample	48 schools/ 273 students	88.31 (10.88)	86.16 (11.47)	2.15	0.19	+8	< 0.05
Domain average for other academic performance						0.19	+8	Not statistically significant

Table Notes: 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 the 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.

Study Notes: The statistics reported in the table are based on means, standard deviations, and sample sizes that were provided to the WWC by the authors and do not use imputed data. Specifically, the WWC used the (non-imputed) statistics provided by the authors to calculate the intervention group mean, which equals the unadjusted comparison group posttest mean plus the difference in mean gains between the intervention and comparison groups. Please see the WWC Procedures and Standards Handbook version 2.1 for more information. The p-values presented here were reported in the original study based on an analysis that uses imputed data. WWC calculations using unimputed data required a correction for clustering and resulted in WWC-computed p-values of 0.06, 0.19, and 0.18 for AET, Problem Behavior (PB) SSRS subscale, and Academic Competence (AC) SSRS subscale, respectively; therefore, the WWC does not find any of these results to be statistically significant.

The study is characterized as having an indeterminate effect on both external behavior and other academic performance because the average effect in each domain is neither statistically significant nor substantively important, accounting for multiple comparisons.

Endnotes

¹ Single study reviews examine evidence published in a study (supplemented, if necessary, by information obtained directly from the author[s]) to assess whether the study design meets WWC evidence standards. The review reports the WWC's assessment of whether the study meets WWC evidence standards and summarizes the study findings following WWC conventions for reporting evidence on effectiveness. This study was reviewed using the topic area review protocol for Interventions for Children Classified as Having an Emotional Disturbance, version 2.0. The WWC rating applies only to the results that were eligible under this topic area and met WWC standards without reservations or met WWC standards with reservations, and not necessarily to all results presented in the study.

² Externalizing behaviors are those that are directed outward and affect others (e.g., arguing, disturbing others, fighting). Internalizing behaviors are those that are directed inward (e.g., becoming withdrawn or depressed).

³ There were seven other outcomes included in the study that are not described in this WWC report. See the table notes in Appendix B for more information.

⁴ The WWC will only review analyses that use imputed data from RCTs that have a low level of sample attrition. Analyses of imputed data obtained from RCTs with high attrition or from QEDs cannot meet WWC evidence standards.

Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2013, September). *WWC review of the report: Assessing the effectiveness of First Step to Success: Are short-term results the first step to long-term behavioral improvements?* Retrieved from <http://whatworks.ed.gov>

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 intervention 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 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.
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 intervention 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 intervention and comparison groups.
Single-case design (SCD)	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 are 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.