# What Works Clearinghouse™

**:ies** INSTITUTE OF EDUCATION SCIENCES

## WWC Review of the Report "Reducing Developmental Risk for Emotional/Behavioral Problems: A Randomized Controlled Trial Examining the Tools for Getting Along Curriculum"[1]

**The findings from this review do not reflect the full body of research evidence on *Tools for Getting Along.***

### What is this study about?

The study examined the impact of *Tools for Getting Along (TFGA)* on fourth- and fifth-grade students' behavioral outcomes in 14 Florida elementary schools. The curriculum aims to improve students' social problem-solving skills.

Researchers randomly assigned seven schools enrolling 898 fourth- and fifth-grade students to receive the *TFGA* intervention and seven other schools enrolling 877 fourth- and fifth-grade students to serve as a comparison condition. The final analysis sample contained between 1,119 and 1,174 students total (sample sizes varied by the outcome).[2]

Schools in the intervention condition received *TFGA* materials and training to implement the program immediately, and schools in the comparison condition received *TFGA* materials and training after the conclusion of the study. Teachers in the intervention condition taught 27 *TFGA* lessons, including six booster lessons. On average, teachers covered one to two lessons per week.

Researchers assessed the effectiveness of the *TFGA* curriculum by comparing student performance on nine main outcomes and seven subscale outcomes.[3] The outcomes assessed dimensions of student behavior and were based on student self-reporting and teacher assessments of behavior.

### Features of *Tools for Getting Along*

*Tools for Getting Along (TFGA)* is a classroom curriculum that aims to help students develop social problem-solving skills by teaching them a sequence of steps used to address problem situations. The problem-solving sequence uses a six-step process. Steps 1 and 2 focus on recognizing the problem and calming down to be able to address the problem. Step 3 involves defining the goals and barriers of the social problem. Step 4 incorporates brainstorming possible solutions to the problem, and Steps 5 and 6 include selecting and evaluating a response.

Students are taught this process through direct instruction, teacher modeling, and student role-playing that occurs across a sequence of 27 lessons (one to two lessons were taught each week). The first 21 lessons provide the instructional core of the program, and the final six booster lessons allow students to generalize the skills that they learned through additional role-playing and real-life problem solving.

## What did the study find?

Study authors reported that students who received *TFGA* had statistically significantly greater use of rational problem-solving styles than students who did not receive *TFGA*. The WWC did not confirm this finding to be statistically significant. The study authors' analyses did not appropriately adjust for the clustering of students within schools as the unit of assignment, and as such, the WWC's calculations of the statistical significance differ from what was presented in the article.

Study authors also reported, and the WWC confirmed, no statistically significant differences between students who received *TFGA* and students who did not receive *TFGA* for eight other main behavioral and emotional outcomes and seven reported subscale results.

### WWC Rating

***The research described in this report for the overall sample meets WWC evidence standards without reservations***

**Strengths:** The study is a well-implemented randomized controlled trial with low levels of sample attrition for three outcomes (rational problem solving, impulsiveness/carelessness, and avoidance). The research on these three outcomes with low sample attrition meets WWC evidence standards without reservations.

**Cautions:** The analysis samples for six outcomes (Metacognition Index scale, Internalizing scale, Externalizing scale, Negative Problem Orientation, Behavior Regulation Index, Proactive Aggression scale) showed high levels of attrition, but the samples were demonstrated to be equivalent at baseline on pre-intervention assessments of the outcome variables. As a result, the research on these six outcomes meets WWC evidence standards with reservations.

## Appendix A: Study details

Daunic, A., Smith, S., Garvan, C., Barber, B., Becker, M., Peters, C., Taylor, G., Van Loan, C., Li, W., & Naranjo, A. (2012). Reducing developmental risk for emotional/behavioral problems: A randomized controlled trial examining the Tools for Getting Along curriculum. *Journal of School Psychology, 50*(2), 149–166.

| | |
|---|---|
| **Setting** | The study was conducted in 14 elementary schools with fourth- and fifth-grade classrooms in Florida. |
| **Study sample** | Researchers pooled information from two separate samples of schools across a 2-year period. Over the two years, the study authors contacted 32 elementary schools that were selected based on their proximity to the research site and having a high (> 60%) portion of students receiving free/reduced-price lunch. Fourteen schools agreed to participate in the study. Next, schools were matched on the percentage of students receiving free/reduced-price lunch. Random assignment of schools to the intervention or comparison condition was conducted within matched pairs. There was no attrition of schools in the full analysis sample. Information obtained from the authors indicated that there were 898 students in the intervention group and 877 students in the comparison group at baseline. The final analysis sample contained between 1,119 and 1,174 students (sample sizes varied by the outcome). |
| | About 70% of students in the intervention group received free/reduced-price lunch, compared to 87% of students in the comparison group. Approximately half of the intervention and comparison groups were female. About 54% of participating students were White, and 35% were African American. |
| **Intervention group** | *TFGA* is a classroom curriculum that aims to help students develop social problem-solving skills by teaching them a sequence of six steps to address problem situations. Steps 1 and 2 focus on recognizing the problem and calming down to be able to address the problem. The third step involves defining the goals and barriers of the problem. Step 4 incorporates brainstorming possible solutions to the problem, and Steps 5 and 6 include selecting and evaluating a response. Students are taught this process through direct instruction, teacher modeling, and student role-playing that occurs across a sequence of 27 lessons (one to two lessons were taught each week). The first 21 lessons provide the instructional core of the program, and the final six booster lessons allow students to generalize the skills that they learned through additional role-playing and real-life problem solving. |
| **Comparison group** | The comparison condition received the regular "business-as-usual" curriculum. Comparison schools were informed that they would be given the training and materials for *TFGA* after the study concluded. |

| | |
|---|---|
| **Outcomes and measurement** | This WWC report examines nine main outcomes that assess student behavior. |

- Two outcomes, the Behavior Regulation Index (BRI) and the Metacognition Index (MI) are from the Behavior Rating Inventory of Executive Function (BRIEF) Teacher Form, a teacher-reported measure. These two main indices are further reported as seven sub-scales: The BRI in this WWC report is composed of two subscales, and the MI is composed of five subscales, each of which is an eligible outcome.[4]
- The third and fourth main outcome measures are from the Clinical Assessment of Behavior Teacher Rating (CAB-T) form: the Internalizing scale and the Externalizing scale.
- The fifth main outcome measure is from the Proactive Aggression scale (part of the teacher-reported Reactive-Proactive Aggression scale).
- Finally, the study examines four outcomes from the Social Problem-Solving Inventory– Revised (SPSI-R), which is a student-reported measure: Negative Problem Orientation, Rational/Adaptive Style, Impulsive/Careless Style, and Avoidance Style.

For a more detailed description of these main outcome measures and their respective subscales, see Appendix B.

| | |
|---|---|
| **Support for implementation** | Teachers and guidance counselors at the intervention schools were trained in *TFGA* in the second full month of each study year. The training took 10 hours. In January of each study year, the researchers visited the intervention schools to provide refresher information and answer questions about *TFGA*. |
| **Reason for review** | This study was identified for review by the WWC in response to a request by the Institute of Education Sciences. |

## Appendix B: Outcome measures for each domain

| **Emotional/internal behavior** | |
|---|---|
| *Avoidance Style* | This scale is from the Social Problem-Solving Inventory–Revised (SPSI-R) and consists of seven items. A lower score for this subscale indicates a more positive outcome. |
| *Externalizing scale* | This scale is from the Clinical Assessment of Behavior Teacher Rating Form (CAB-T). |
| *Impulsive/Careless Style* | This scale is from the SPSI-R and consists of 10 items. A lower score for this subscale indicates a more positive outcome. |
| *Initiate subscale* | This subscale is from the BRIEF and is a component of the MI. It measures the tendency of individuals to initiate tasks and strategies. |
| *Internalizing scale* | This scale is from the CAB-T. The CAB-T consists of 70 questions and uses Likert scales ranging from 1 (always or very frequently) to 5 (never). Lower scores in the CAB-T indicate more positive outcomes. |
| *Metacognition Index (MI) scale* | This scale is from the Behavior Rating Inventory of Executive Function (BRIEF) Teacher Form, an assessment with 86 items. Each item is a Likert scale ranging from 1 (never) to 3 (often), and higher scores indicate more executive function. The MI measures the ability of individuals to plan and manage tasks. The MI is formed from the Organize, Initiate, Monitor, Working Memory, and Plan subscales in the BRIEF. |
| *Monitor subscale* | This subscale is from the BRIEF and is a component of the MI. It measures the tendency of individuals to monitor performance. |
| *Negative Problem Orientation* | This scale is from the SPSI-R and consists of 10 items. It measures whether individuals approach problems in a negative way. A lower score for this subscale indicates a more positive outcome. |
| *Organize subscale* | This subscale is from the BRIEF and is a component of the MI. It measures the tendency of individuals to organize materials and plans. |
| *Plan subscale* | This subscale is from the BRIEF and is a component of the MI. |
| *Rational Problem-Solving Style* | This scale is from the SPSI-R and consists of 20 items. The rational style aligns closely with the method taught in the *TFGA* curriculum. A higher score for this subscale indicates a more positive outcome. |
| *Working Memory subscale* | This subscale is from the BRIEF and is a component of the MI. |
| **External behavior** | |
| *Behavior Regulation Index (BRI)* | This scale is from the BRIEF. The BRI measures the ability of individuals to inhibit and manage emotions and behavior. The BRI is formed in part by the Inhibit and Emotional Control subscales in the BRIEF. |
| *Emotional Control subscale* | This subscale is from the BRIEF and is a component of the BRI. |
| *Inhibit subscale* | This subscale is from the BRIEF and is a component of the BRI. |
| *Proactive Aggression scale* | This scale is from the Reactive/Proactive Aggression scale (R/P), a teacher form where a lower score indicates a more positive outcome. All items are on a Likert scale from 1 (never true) to 5 (always true). The Proactive Aggression scale is composed of three items. A sample item from the Proactive Aggression scale is, "This child gets other children to gang up on a peer that he/she does not like." |

**Table Notes:** Lower scores indicate improved outcomes for some measures (Internalizing and Externalizing scales on the CAB-T, Negative Problem Orientation, Impulsive/Careless Style, and Avoidance Style on the SPSI-R, and the Proactive Aggression scale). We have reversed the signs of the WWC calculations for these outcomes in Appendices C and D so that positive mean differences, effect sizes, and improvement indices always correspond to improvements for the intervention group. The study also examined the following main outcomes that are not included here: Problem-Solving Knowledge Questionnaire, Reactive Aggression scale from R/P, Positive Problem Orientation scale from SPSI-R, and scales from the Anger Expression scale for Children (AESC). The AESC includes the Trait Anger, Anger Out, and Anger Control scales. The Problem-Solving Knowledge Questionnaire outcome does not meet requirements because it is overaligned with the intervention. The other outcomes are excluded because the initial differences on these measures were too large for statistical adjustment. In addition to these six main outcomes, there was one subscale, the Shift Subscale from the BRIEF assessment that did not meet standards due to high levels of sample attrition and inequivalence of the analytic samples at baseline.

## Appendix C: Study findings for each domain

| Domain and outcome measure | Study sample | Sample size | Mean (standard deviation) | | WWC calculations | | | |
|---|---|---|---|---|---|---|---|---|
| | | | Intervention group | Comparison group | Mean difference | Effect size | Improvement index | *p*-value |
| **Emotional/internal behavior** | | | | | | | | |
| *Avoidance Style* | Full sample | 14 schools/ 1,124 students | −16.96 (4.99) | −15.90 (4.94) | −1.06 | −0.21 | −8 | > 0.05 |
| *Externalizing scale* | Full sample | 14 schools/ 1,171 students | −73.13 (16.22) | −72.45 (18.02) | −0.68 | −0.04 | −2 | > 0.05 |
| *Impulsive/Careless Style* | Full sample | 14 schools/ 1,123 students | −24.31 (7.74) | −23.99 (7.29) | −0.32 | −0.04 | −2 | > 0.05 |
| *Internalizing scale* | Full sample | 14 schools/ 1,174 students | −63.92 (11.39) | −64.07 (12.26) | 0.15 | 0.01 | +1 | > 0.05 |
| *Metacognition Index scale* | Full sample | 14 schools/ 1,158 students | 68.10 (22.62) | 71.00 (25.77) | −2.90 | −0.12 | −5 | > 0.05 |
| *Negative Problem Orientation* | Full sample | 14 schools/ 1,127 students | −24.64 (7.44) | 23.61 (7.29) | −1.03 | −0.14 | −6 | > 0.05 |
| *Rational Problem-Solving Style* | Full sample | 14 schools/ 1,119 students | 59.56 (16.26) | 57.33 (16.06) | 2.23 | 0.14 | +5 | < 0.05 |
| **Domain average for emotional/internal behavior** | | | | | | −0.06 | −2 | **Not statistically significant** |
| **External behavior** | | | | | | | | |
| *Behavior Regulation Index* | Full sample | 14 schools/ 1,158 students | 42.37 (14.84) | 43.26 (16.43) | −0.89 | −0.06 | −2 | > 0.05 |
| *Proactive Aggression scale* | Full sample | 14 schools/ 1,163 students | −5.01 (2.63) | −5.18 (3.03) | 0.17 | +0.06 | +2 | > 0.05 |
| **Domain average for external behavior** | | | | | | 0 | 0 | **Not statistically significant** |

**Table Notes:** Because lower raw scores on the following measures indicate a positive outcome, we have reversed the signs for the following outcomes: Internalizing scale, Externalizing scale, Negative Problem Orientation, Impulsive/Careless Style, Avoidance Style, and Proactive Aggression scale. As a result, across all outcomes, positive results for mean difference, effect size, and improvement indices 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. The statistical significance of the study's domain average was determined by the WWC; the study is characterized as having an indeterminate effect because none of the individual effects or the mean effect was found to be statistically significant or substantively important.

**Study Notes:** Study Notes: The *p*-values presented here were reported in the original study. Corrections for clustering and multiple comparisons were needed and resulted in significance levels that differ from those in the original study. The WWC finds that the result for the Rational Problem-Solving style outcome is not statistically significant after correcting for clustering. The analyses reported in the study did not appropriately adjust for schools as the unit of assignment in the study (the authors only included a clustering adjustment at the classroom level, not the school level).

The WWC calculated the intervention group mean (based on information on sample sizes, means, and standard deviations provided by the author in email correspondence) by adding the difference-in-differences adjusted estimate of the average impact of the program (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means. Please see the WWC Procedures and Standards Handbook version 2.1 for more information.

Out of all of the comparisons presented in Appendix C, the only contrasts that meet WWC standards without reservations are those for the Rational Problem-Solving Style, Impulsive/Careless Style, and Avoidance Style. The analysis samples for six outcomes (Metacognition Index scale, Internalizing scale, Externalizing scale, Negative Problem Orientation, Behavior Regulation Index, Proactive Aggression scale) showed high levels of attrition, but the samples were demonstrated to be equivalent at baseline on pre-intervention assessments of the outcome variables. As a result, the research on these six outcomes meets WWC evidence standards with reservations.

Results for subscales are presented in Appendix D.

## Appendix D: Supplemental subscale findings for each domain

| Domain and outcome measure | Study sample | Sample size | Mean (standard deviation) | | WWC calculations | | | |
|---|---|---|---|---|---|---|---|---|
| | | | Intervention group | Comparison group | Mean difference | Effect size | Improvement index | *p*-value |
| **Emotional/internal behavior** | | | | | | | | |
| *Initiate subscale* | Full sample | 14 schools/ 1,160 students | 11.00 (3.99) | 11.75 (4.48) | −0.75 | −0.18 | −7 | > 0.05 |
| *Monitor subscale* | Full sample | 14 schools/ 1,160 students | 16.02 (5.48) | 16.45 (6.06) | −0.43 | −0.07 | −3 | > 0.05 |
| *Organize subscale* | Full sample | 14 schools/ 1,160 students | 10.20 (3.72) | 10.61 (4.48) | −0.41 | −0.10 | −4 | > 0.05 |
| *Plan subscale* | Full sample | 14 schools/ 1,160 students | 15.76 (5.38) | 16.39 (6.21) | −0.63 | −0.11 | −4 | > 0.05 |
| *Working Memory subscale* | Full sample | 14 schools/ 1,159 students | 15.14 (5.65) | 15.80 (6.08) | −0.66 | −0.11 | −4 | > 0.05 |
| **External behavior** | | | | | | | | |
| *Emotional Control subscale* | Full sample | 14 schools/ 1,160 students | 13.03 (4.97) | 13.18 (5.39) | −0.15 | −0.03 | −1 | > 0.05 |
| *Inhibit subscale* | Full sample | 14 schools/ 1,160 students | 15.41 (6.19) | 15.76 (6.69) | −0.35 | −0.05 | −2 | > 0.05 |

**Table Notes:** Higher scores indicate a beneficial outcome for the intervention group for all measures in Appendix D. For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors 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.

**Study Notes:** Corrections for clustering and multiple comparisons were needed; however, these corrections do not result in significance levels that differ from those in the original study. The *p*-values presented here were reported in the original study. The analyses reported in the study did not appropriately adjust for schools as the unit of assignment in the study (the authors only included a clustering adjustment at the classroom level, not the school level).

The WWC calculated the intervention group mean (based on information on sample sizes, means, and standard deviations provided by the author in email correspondence) by adding the difference-in-differences adjusted estimate of the average impact of the program (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means. Please see the WWC Procedures and Standards Handbook version 2.1 for more information. All contrasts presented in Appendix D meet WWC standards with reservations.

Results for the main outcomes that combine these subscale findings are presented in Appendix C.

## 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 single study review protocol, 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] Sample sizes for each outcome were obtained from the author in email correspondence.

[3] Six additional scale outcomes were examined in this study, but are not included in this report: Problem-Solving Knowledge Questionnaire, Reactive Aggression scale from R/P, Positive Problem Orientation scale from SPSI-R, and the three scales from the Anger Expression scale for Children (AESC). The AESC includes the Trait Anger, Anger Out, and Anger Control scales. The Problem-Solving Knowledge Questionnaire outcome does not meet requirements because it is overaligned with the intervention. The other five main outcomes are excluded from the report because there were high levels of attrition demonstrated for the outcomes and the analytic samples were not shown to be equivalent on the pre-intervention (baseline) assessments of these measures. In addition to these six main outcomes, the Shift Subscale from the BRIEF assessment that did not meet standards due to high levels of sample attrition and inequivalence of the analytic samples at baseline. Furthermore, the study compared outcomes for fourth- and fifth-graders who scored below the 25th percentile (referred to as the "highest-risk quartile") on each outcome measure prior to *TFGA*. There was insufficient information available for this sample to meet the WWC attrition standards, and there were large differences (i.e., greater than 0.25 standard deviations) at baseline shown in each of the two domains assessed across the eligible outcomes. As a result, all comparisons that examine effects for the "highest-risk quartile" do not meet WWC standards and are not included in this WWC report.

[4] As indicated in Endnote 3, the results that examine effects on the Shift Subscale from the BRIEF measure do not meet WWC standards; however, the main result for the BRI (which include this subscale) meets WWC standards with reservations and is included in this WWC report. The seven subscales reported in the study that meet WWC standards with reservations are shown in Appendix D.

## Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2013, March). *WWC review of the report: Reducing developmental risk for emotional/behavioral problems: A randomized controlled trial examining the Tools for Getting Along curriculum*. 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.