Effect of Tools for Getting Along on Student Risk for Emotional and Behavioral Problems in Upper Elementary Classrooms: A Replication Study

Article *in* School psychology review · March 2016

DOI: 10.17105/SPR45-1.73-92

Ann P Daunic
University of Florida

Burak Aydin
Ege University

Christopher L Van Loan
Appalachian State University

Brian R. Barber
Kent State University

School Psychology Review,
2016, Volume 45, No. 1, pp. 73–92

Effect of Tools for Getting Along on Student Risk for Emotional and Behavioral Problems in Upper Elementary Classrooms: A Replication Study

Stephen W. Smith and Ann P. Daunic
University of Florida

Burak Aydin
Recep Tayyip Erdoğan University

Christopher L. Van Loan
Appalachian State University

Brian R. Barber
Kent State University

Gregory G. Taylor
University of Illinois

Abstract. Social–emotional learning curricula to prevent student problematic behaviors should play a prominent role in public school instruction. While social–emotional curricula have been shown to be effective, there are few replication studies that substantiate their capacity to improve outcomes for students who exhibit problem behaviors. Thus, we conducted a partial replication of a randomized controlled field trial of the Tools for Getting Along curriculum designed to increase self-regulatory functions of upper elementary school students. We found main effects on social problem-solving and significant pretest-by-condition interaction effects on teacher-reported executive function, behavioral adjustment, and aggression. We also found interaction effects on student-reported anger. The current findings are similar to the results reported for the initial study. We also specified high-risk subsamples of students for each of eight outcome factors and found main effects on emotion regulation and positive social problem-solving and a marginally significant effect on metacognition.

There is an increasing body of research to indicate that cognitive–behavioral interventions (CBIs) implemented in school settings can help students regulate their own behavior. Students can learn how to use verbal mediation, or verbal self-regulation skill develop-

This study was supported by a four-year grant from the U.S. Department of Education, Institute of Education Sciences (No. R324B060029), submitted to CFDA 84.324A, National Center for Special Education Research.

Please address correspondence regarding this article to Stephen W. Smith, University of Florida, Special Education, 1403 Norman Hall, Gainesville, FL 32611; e-mail: swsmith@coe.ufl.edu

ment, to improve their reactions to social events, which increases their use of appropriate presocial skills and reduces classroom disruption (e.g., McCart, Priester, Davies, & Azen, 2006; Robinson, Smith, Miller, & Brownell, 1999; Sukhodolsky, Kassinove, & Gorman, 2004). Teaching students how to regulate their own behavior and providing them with the requisite knowledge and skills to develop self-management skills comprise a fundamental goal of social–emotional learning. For students who are at risk for or who already exhibit behavioral problems, acquiring effective self-regulatory processes can reduce the need for external behavioral control and increase students' positive choice making, self-determination, and self-esteem (cf. Donnellan, Trzesniewski, Robins, Moffitt, & Caspi, 2005; Polsgrove & Smith, 2004; Wehmeyer, Palmer, Shogren, Williams-Diehm, & Soukup, 2013). Self-regulation becomes increasingly important as students approach adolescence, when they rely less on adult supervision and are subjected to pressure from peers (Lerner & Steinberg, 2004). CBIs have been established as a viable, research-based approach appropriate for use in school settings to assist students to learn self-regulatory skills (e.g., Mayer, Van Acker, Lochman, & Gresham, 2009; Robinson et al., 1999).

CBIs combine cognitive mediation with behavioral principles such as contingent praise, point systems, and behavioral contracting for motivating students to build new ways of coping (Kendall & Braswell, 1985; Smith & Daunic, 2006). Researchers have examined CBIs through large-scale multi-informant studies and found that CBIs prevented or mollified the effects of disruptive behavior disorders and positioned students for more appropriate and long-term behavioral change through better self-regulation (e.g., Daunic, Smith, Brank, & Penfield, 2006; Daunic et al., 2012; Smith et al., 2014). Two well-researched CBI programs, Promoting Alternative Thinking Strategies (PATHS; Conduct Problems Prevention Research Group, 1999, 2004) and Coping Power (e.g., Lochman & Wells, 2004), address emotional and behavioral risk factors such as disruptive behaviors and aggression, and each has been shown to reduce negative behaviors (Smith, Lochman, & Daunic, 2005) when compared with a high-risk control condition. Although each is a multicomponent program that includes instruction in areas such as goal setting, anger management, emotion awareness, friendships, and perspective taking, both include instruction on strategies to solve social problems.

By almost any measure, the ability to self-regulate behavior successfully is critical to developing and maintaining positive social relations. There is a growing consensus that self-regulatory processes are dependent in part on development of executive function (EF; Hofmann, Schmeichel, & Baddeley, 2012; Lyons & Zelazo, 2011; Rueda, Acosta, & Santonja, 2007), which is defined as the active manipulation, sequencing, and monitoring of information for the purpose of producing goal-oriented actions during novel situations (Lewis & Carpendale, 2009). Individuals with proficient EF are able to set goals competently, plan activities, and monitor their performance and thus self-regulate successfully (Rueda, Posner, & Rothbart, 2005; Zelazo, Carlson, & Kesek, 2008). More precisely, when individuals are regulating their own behavior by engaging in deliberate, goal-directed problem solving to achieve goals and regulate emotions, they are effectively recruiting EF. Conversely, EF deficiencies that affect a person's self-regulatory functioning can contribute to social–emotional and behavioral difficulties (Hughes, 2002; Olson, Sameroff, Kerr, Lopez, & Wellman, 2005). As such, self-regulation skills that contribute to effective social problem-solving, as taught in some CBIs, play an important role in social–emotional functioning.

Despite increasing evidence that CBIs can positively affect overall student behavior and the ability to self-regulate (e.g., McCart et al., 2006; Robinson et al., 1999; Smith, Graber, & Daunic, 2009), more studies are needed to examine further their efficacy and effectiveness. Moreover, school psychologists are in a prime position within multitiered systems of support (MTSS) to help implement promising interventions and monitor and evaluate associated student progress (Cantin, Mann, & Hund, 2012; National Association of School Psychologists, 2006). Replication studies would serve to substantiate CBIs' capacity to improve outcomes for students with problem behavior and further justify the collaborative involvement of school personnel. To that end, the current study is a partial replication (see Darley, 2000; Hendrick, 1991) of the Daunic et al. (2012) randomized controlled field trial of Tools for Getting Along (TFGA), a classroom-based preventive intervention for fourth- and fifth-grade students implemented at the universal level and designed to ameliorate risk for significant emotional and behavioral problems.

IMPORTANCE OF REPLICATION

For several decades, behavioral, sociological, educational, and economic fields have debated the characteristics that comprise replication, particularly for randomized controlled trials (e.g., Burman, Reed, & Alm, 2010; Rosenthal, 1991). Controlled trial replication has been described along a continuum from meticulously following previous study methods to involving slight alterations of specific study characteristics such as context and participants. Bahr, Caplow, and Chadwick (1983) suggested that so long as a replication study revisits several aspects of any category (time, space, subjects, method), it qualifies as a repeat study. Moreover, van IJzendoorn (1994) argued that a partial or varied replication could help determine whether a study's outcomes remain stable or change in a predictable way. The present study aligns with this notion of replication as we increased the sample size by 60% (we added to the sample by collecting pre–post data for an additional year) and used a different statistical estimation technique based on a verified measurement model to investigate the hypotheses described by Daunic et al. (2012).

Daunic et al. (2012) found, with a sample of 1,296 students, that intervention group students with relatively higher baseline risk on measures of proactive aggression, contextualized EF, and self-reports of trait anger (degree of anger habitually experienced) and anger out (open expression of angry feelings) had lower posttest risk on these outcomes compared with control students. Their results also indicated that students taught TFGA, on average, had higher knowledge of curriculum-related concepts and a more positive orientation to social problem-solving. For the current study and similar to Daunic et al., our research hypotheses were that TFGA would reduce risk for emotional and behavioral difficulties by (a) improving teacher-reported student behaviors associated with EF skills, (b) improving teacher-rated student social–emotional adjustment and aggression, (c) positively affecting students' self-rated anger control/expression, and (d) positively affecting students' self-ratings of social problem-solving. Furthermore, we sought to determine whether TFGA improved curriculum-related knowledge about social problem-solving. Extending this partial replication, we also wanted to determine TFGA outcomes for subgroups of students with the poorest scores on each outcome measure to follow up on findings related to pretest-by-condition interaction effects and to examine social validity of the intervention.

METHOD

The findings we report in this study are based on data aggregated across 3 years of implementation, with a distinct group of schools participating each year. All recruitment and study procedures met university institutional review board standards.

Setting and Participants

From 54 schools in North Central Florida, we rank ordered the schools, prior to first-year recruitment, first by percentage of students qualifying for a free or reduced-price lunch (FRL; 55% or more) and second by distance from the research site. After contacting targeted school district personnel during the summer preceding each school year, we telephoned school principals to schedule face-to-face meetings. We contacted a total of 43 principals across 3 years; 31 agreed to meet with project staff, and 20 agreed to participate in the study. We met directly with fourth- and fifth-grade teachers in each participating school to explain our research design, how and when assignment to the TFGA or control condition would occur, and the requirement that at least 75% of the teachers agree to participate prior to random assignment.

Random Assignment and Sample Characteristics

After recruitment, we matched schools on percent of students who received FRL and randomly assigned members of each matched pair to TFGA or a business-as-usual control condition using Maple (Maplesoft, 2005–2015). School principals were informed of their participation group after all schools were assigned each year. Control group schools were told they would be offered TFGA training and materials after participation. The 20 participating schools in the study included 135 classrooms (68 intervention, 67 control) and 2,079 student participants; 55% to 100% of students across schools qualified for FRL.

Student Demographic Characteristics

Because of a predominance of African American and White students and the low number of Hispanics and students of other ethnicities, we designated race as a binary variable allowing us to examine the impact of African American status on outcome measures. Soliciting active parental or guardian consent for 2,777 students, we obtained consent from 2,131, or 76.7% (intervention, 81.1%; control, 71.9%), across 3 years. Racial information was available for 2,006; 30% of these students were African American, and 70% were White or other. Of 2,003 students for whom gender was available, 51% were girls. Of 2,072 with grade information available, 49% were fifth graders and 51% fourth graders. Of 1,932 with socioeconomic information (i.e., FRL) available, 81% qualified for FRL whereas 19% did not. The Florida Comprehensive Assessment Test (FCAT) math score ($M = 2.81$, $SD = 1.10$) was obtained for 1,933 students, and the FCAT reading score ($M = 2.82$, $SD = 1.11$) for 1,936 students. The TFGA and control groups differed on race ($\chi^2 = 13.56$, $p < .01$), grade ($\chi^2 = 9.87$, $p < .01$), and FRL ($\chi^2 = 40.85$, $p < .01$), as well as reading, $t(1934) = 2.36$, $p = .02$.

Measures

The Behavior Rating Inventory of Executive Function–Teacher Form (BRIEF) is a standardized instrument consisting of 86 items that comprise eight clinical scales. Respondents use a Likert scale to indicate *never*, *sometimes*, or *often* for each item. The scales form two broad indices, the Behavioral Regulation Index and the Metacognition Index, and a global executive composite score (see Gioia et al., 2002). The Behavioral Regulation Index is composed of Inhibit, Shift, and Emotional Control scales and relates to the ability to use inhibitory control to shift cognitive set and manage emotions and behavior. The Metacognition Index is composed of Initiate, Working Memory, Plan/Organize, Organization of Materials, and Monitor scales that collectively reflect the ability to cognitively self-manage tasks and monitor performance. The BRIEF has demonstrated adequate reliability and construct validity (Gioia, Isquith, Kenworthy, & Barton, 2002). Full-sample Cronbach's α estimates for the eight individual scales ranged from .92 to .96 at pretest and from .93 to .97 at posttest.

The Reactive–Proactive Aggression Scale includes 19 questions to teachers about student behavior, including six embedded items that assess aggression. Three comprise the Proactive Aggression subscale (e.g., "This child gets other children to gang up on a peer that he/she does not like") and three comprise the Reactive Aggression subscale (e.g., "When this child has been teased or threatened he/she gets angry easily and strikes back"), but aggressive subtypes are not mutually exclusive. All item responses range from 1 (*never true*) to 5 (*always true*) on a Likert scale, such that total scores for each subscale range from 3 to 15. The Reactive–Proactive Aggression Scale has demonstrated moderate construct validity for the two personality-related constructs and high internal consistency (see Dodge & Coie, 1987). For our sample, Cronbach's αs for Reactive Aggression and Proactive Aggression were .90 and .91, respectively, at pretest and .92 and .92, respectively, at posttest.

The Anger Expression Scale for Children (AESC) is modeled after the well-validated State-Trait Anger Expression Inventory (Speilberger, 1988) and is a 30-item self-report measure with Likert scale responses ranging from 1 (*almost never*) to 4 (*almost always*). Four subscales were derived from factor analyses: (a) Anger Control, 6 items; (b) Anger-Out, 6 items; (c) Anger-In, 6 items; and (d) Trait Anger, 12 items. Initial factor analyses indicated that Anger Control and Anger-In clustered together, with higher scores indicating more effort to prevent outward displays of anger and/or to minimize the experience of anger (e.g., "I feel [anger] inside but I don't show it"). Similarly, Anger-Out and Trait Anger clustered together, with higher scores indicating more generalized feelings of anger and a greater tendency to display it outwardly (e.g., "I get in a bad mood when things don't go my way"; Phipps & Steele, 2002). We used the four-factor model as recommended by recent confirmatory factor analyses (Steele, Legerski, Nelson, & Phipps, 2009). Correlational analyses indicate that measure subscales demonstrate good convergent validity with other measures of anger and hostility. Reliability estimates from sample data using Cronbach's α were .84, .78, .40, and .80 for Trait Anger, Anger-Out, Anger-In, and Anger Control, respectively, at pretest and .86, .80, .47, and .81, respectively, at posttest. We excluded Anger-In from statistical analyses because of the low α that may reflect variability in individual children's responses to questions about internalizing characteristics (see Phipps & Steele, 2002).

The Social Problem-Solving Inventory–Revised (SPSI-R) is based on a two-component model of problem solving: *Problem orientation* focuses on metacognitive processes that reflect general awareness and appraisals of problems encountered in everyday life, and *problem-solving style* focuses on four complex skills necessary to solve a problem successfully—(a) problem definition and formulation, (b) alternative solution generation, (c) decision making, and (d) implementing a solution and evaluating its outcome (see D'Zurilla, Nezu, & Maydeu-Olivares, 2004). These skills closely parallel those taught explicitly in TFGA and addressed by Dodge's social information processing model (Dodge, 1986). The SPSI-R includes 52 Likert scale self-report items that comprise two problem orientation scales, Positive Problem Orientation (PPO, 5 items) and Negative Problem Orientation (10 items), and three problem-solving style scales, Rational/Adaptive Problem Solving (RPS, 20 items), Impulsive/Careless Style (10 items), and Avoidance Style (7 items). Item responses range from 1 (*not at all true*) to 5 (*extremely true*). Among diverse populations, the SPSI-R has shown strong internal consistency and stability over time and evidenced strong structural, concurrent, predictive, convergent, and discriminant validity (D'Zurilla, Nezu, & Maydeu-Olivares, 2002). Cronbach's αs obtained from the current sample ranged from .61 (PPO) to .90 (RPS) at pretest and from .66 (Avoidance Style and PPO) to .92 (RPS) at posttest.

The Clinical Assessment of Behavior Teacher-Rating Form (CAB-T) is a standardized behavior scale consisting of 70 questions that assess three clinical scales (Internalizing Behavior, Critical Behavior, Externalizing Behavior), three adaptive scales (Social Skills, Competence, Adaptive Behavior), and educationally related clusters (Executive Function, Gifted and Talented). Likert scale item responses range from 1 (*always or very frequently*) to 5 (*never*). CAB-T subscale scores from previous studies have yielded internal reliability coefficients ranging from .94 to .97 for clinical scales and .80 to .99 for adaptive scales (Bracken & Keith, 2004). The CAB-T also demonstrates evidence of validity based on test content; factor analytic studies; and convergent, discriminant, and concurrent validity studies across clinical groups such as conduct or disruptive behavioral disorders (Bracken & Keith, 2004). Cronbach's α estimates for the four subscales used in our analyses at pretest and at posttest were .93 and .92, respectively, for Internalizing Behavior; .97 and .97, respectively, for Externalizing Behavior; .93 and .95, respectively, for Social Skills; and .95 and .95, respectively, for Competence.

The Problem-Solving Knowledge Questionnaire (KQ) was developed to assess student knowledge of concepts and information taught explicitly in TFGA and serves as an indirect check on curriculum exposure. For the first 11 items, only one answer among several alternatives is appropriate; Items 12–14 require students to check all that apply (e.g., “Check all the ways your body may feel when you are angry”). Two additional items require students to supply curriculum-specific information (e.g., “What are three levels of anger, from lowest to highest?”). The maximum total scale score is 24. Daunic et al. (2006) conducted item analyses and pilot administration prior to its use in the current study. Reliability estimates from the current sample yielded a total score Cronbach’s α of .46 at pretest and .74 at posttest. (Note that low reliability at pretest is expected because of restricted variability in knowledge scores prior to TFGA implementation.)

For all student self-report measures, teachers were instructed to read items aloud to students as a group or to closely monitor students for whom readability might be an issue. We were attentive to this issue because our participating schools had high proportions of students at behavioral and academic risk.

Finally, the TFGA Teacher Questionnaire assessed social validity in three categories. Eight items related to ease of use (e.g., “The curriculum was easy to use; I completed each lesson in the time allotted”), six to appeal or utility to students (e.g., “The curriculum concepts were age-appropriate for my students”), and eight to effectiveness for reducing negative behaviors (e.g., “The curriculum improved my students’ behavior”). Each Likert scale response ranged from 1 (*strongly disagree*) to 5 (*strongly agree*).

TFGA Curriculum

TFGA is a universally delivered CBI designed to prevent or ameliorate emotional and behavioral problems by teaching students to use social problem-solving when experiencing interpersonal conflict (Daunic et al., 2006; Smith et al., 2009). It provides a structure for learning, rehearsing, reviewing, and practicing cumulative steps in a problem-solving sequence. Through direct instruction, teacher modeling, and role-playing with explicit self-talk, TFGA is delivered class-wide so that students at risk for emotional or behavioral difficulties benefit from participating in discussions and activities with typical peers.

TFGA is based on Crick and Dodge’s (1994) social information processing model to account for social cognition’s role in the development of aggression. The model includes six steps: Step 1, encoding of external and internal cues; Step 2, cue interpretation and mental representation; Step 3, goal clarification or selection; Step 4, response access or construction; Step 5, response decision; and Step 6, behavioral enactment. According to Crick and Dodge, children who lack or do not execute the skills required to understand others’ intentions and/or generate and select appropriate social responses are more likely to exhibit aggression (Dodge et al., 2003; Dodge & Somberg, 1987) and tend to attribute hostile intent to others even with evidence to the contrary while displaying limited social problem-solving ability (Lochman & Lampron, 1986; Rubin, Bream, & Rose-Krasnor, 1991). Thus, the problem-solving sequence embedded in TFGA lessons includes six steps that parallel Crick and Dodge’s model. The first steps involve recognizing a social problem situation and calming down to engage cognition (Steps 1 and 2), similar to encoding, interpretation, and mental representation of environmental cues. Next is defining a social problem in terms of goals and barriers (Step 3), parallel to goal clarification or selection, and brainstorming possible solutions (Step 4), corresponding to response access or construction. Finally, selecting, enacting, and evaluating a response choice (Steps 5 and 6) represent the response decision and behavioral enactment steps in the Crick and Dodge model. Instructional strategies include cognitive modeling, strategically placed role-playing for practice opportunities, small-group activities, and explicit application of strategies to real-life social scenarios. Following the 20-lesson core, six booster lessons provide review, practice, and opportunities to generalize learned skills through student-constructed role-playing and real-life problem solving.

Implementation Procedures

After informing school personnel about the results of random assignment, we trained TFGA teachers and guidance counselors in CBI strategies and TFGA procedures for 10 hr over 2 days during the second full month of each implementation year. We administered pretest measures within 2 weeks of training and prior to TFGA implementation, allowing teachers in both conditions to become familiar with their students. We instructed TFGA teachers to begin teaching lessons one to two times per week following pretest data collection. We conducted follow-up meetings at each TFGA school in January to reorient teachers to curricular goals, answer questions, and solicit feedback. Teachers in both conditions completed posttest measures within 2 weeks after the end of instruction (mid-April and early May).

Intervention Fidelity

We observed 13% of lessons taught across teachers and years (477 out of 3,645 total lessons), coordinating researchers' availability with teacher schedules to determine specific teachers and lessons observed. To evaluate fidelity of implementation, observers used individual lesson checklists of curricular components; pairs of observers conducted 20% of these observations to obtain interobserver agreement. Observers offered teachers no formal or systematic instructional support.

The mean observer-rated treatment fidelity from nine observers across 68 teachers was 85.8%, *SD* = 17.8%, range = 14%–100%, and only 15% of the fidelity scores were lower than 70%. Thus, most of the teachers observed followed the curriculum as intended. The mean percent agreement within 98 pairs of observers was 96.4%, *SD* = 7.1%, range = 71.4%–100%, with a single outlier of 57.1%, indicating that observers were able to use fidelity checklists with adequate reliability.

Teachers were also asked to complete the TFGA Curriculum Checks form as lessons were taught. They rated, on a scale of 1 (*not at all*) to 4 (*all*), the amount of total lesson content they included, such as small-group arrangement, role-playing, student worksheets, and other lesson-specific questions, and indicated how long lessons lasted, whether they were teaching one or more lessons each week, and whether they were including their own cognitive models, as instructed. Data from TFGA Curriculum Checks forms indicated that respondents covered most or all lesson content, and most respondents indicated that they included all associated components, such as group activities and worksheets. Lessons averaged 30 min including completion of activities.

Measurement Model

With the exception of the KQ, we hypothesized that the subscales for each measure (e.g., CAB-T, AESC) would cluster together because of (a) their underlying theoretical constructs and (b) type of respondent (i.e., student or teacher). For example, even though items on the Externalizing Behavior subscale of the CAB-T include behaviors similar to those in items on the Anger-Out subscale of the AESC, the former are from the classroom teacher's perspective and the latter are from students' perceptions of their own behavior. Consequently, these instruments are not necessarily measuring the same theoretical construct. To test this hypothesis, we designed and tested a model based on factors associated with the 22 subscales used in the Daunic et al. (2012) study. The model included a cluster for each instrument, with the exception of the BRIEF, where we included three separate indices based on the results of prior factor analyses (see Peters, Algina, Smith, & Daunic, 2012). If the measurement model is a good fit and acceptable statistically, it would reduce the number of multiple related outcome variables and the possibility of an inflated Type I error rate.

We used Mplus 7.11 (Muthén & Muthén, 1998–2012) to test whether the measurement model reflected the underlying theory and fit the data adequately. On the basis of the results, eight factors were defined for each pretest and posttest score, as follows: *Behavior Regulation* is the sum of Inhibit and Monitor; *Emotion Regulation* is the sum of Shift and Emotional Control; *Metacognition* is the sum of Initiate, Working Memory, Organize, Plan, and Monitor (as in Peters et al., 2012); *Aggression* is the sum of Proactive Aggression and Reactive Aggression; *Anger* is the sum of Anger-Out, Trait Anger, and Anger Control (Anger-In was eliminated from the model because of low reliability); *Social Problem-Solving–Negative* is the sum of Negative Problem Orientation, Avoidance Style, and Impulsive/Careless Style; *Social Problem-Solving–Positive* (SPS-P) is the sum of PPO and Rational Style; and *Behavioral Adjustment* is the sum of Externalizing Behavior, Internalizing Behavior, Social Skills, and Competence. (Note that, on the basis of model fit and theoretical constructs, we divided the SPSI-R measure into a positive problem-solving factor and a negative problem-solving factor.)

We allowed residuals to be correlated for each pretest and posttest pair. Figure 1 depicts the model, including standardized factor loadings and significant residual correlations. For simplicity, correlations between latent factors are not included in Figure 1. With a sample size of 2,079, the χ^2 test statistic for the model fit was significant, $\chi^2(758) = 7{,}856.5$, $p < .001$; the Tucker–Lewis index (TLI) was .91; the comparative fit index (CFI) was .93; the root mean square error of approximation (RMSEA) was .067; and the standardized root mean residual (SRMR) was .045. Hu and Bentler (1999) suggested that values smaller than .06 for RMSEA, smaller than .08 for SRMR, and larger than .95 for CFI and TLI indicate adequate model fit. Thus, the factor loadings and associated statistics supported using the model in further analyses.

Statistical Analyses

We created eight factor scores by summing subscale scores for each factor designated by the measurement model. We controlled for student characteristics associated with risk for emotional and behavioral difficulties including gender, race or ethnicity, and socioeconomic status as indicated by FRL (see Daunic et al., 2012). We used the FCAT verbal and math scores to control for academic competence. Dependent variables were posttest scores for the eight factors, and independent variables were pretest scores, FRL, gender, race, and FCAT scores at Level 1 and condition and grade at Level 2. We also included the pretest-by-condition cross-level interaction term as an independent variable, hypothesizing that at-risk students would be likely to evidence different responses to intervention at posttest than those who did not exhibit baseline risk. We grand mean centered all continuous independent variables. The Level 1 equation reads as follows:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(\text{PRE SCORE})_{ij} + \beta_{2j}(\text{RACE})_{ij} + \beta_{3j}(\text{GENDER})_{ij} + \beta_{4j}(\text{FRL})_{ij} + \beta_{5j}(\text{FCAT READING})_{ij} + \beta_{6j}(\text{FCAT MATH})_{ij} + e_{ij} \quad (1)$$

The Level 2 equations read as follows:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\text{CONDITION}_j + \gamma_{02}\text{GRADE}_j + U_{0j} \quad (2)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}\text{CONDITION}_j \quad (3)$$

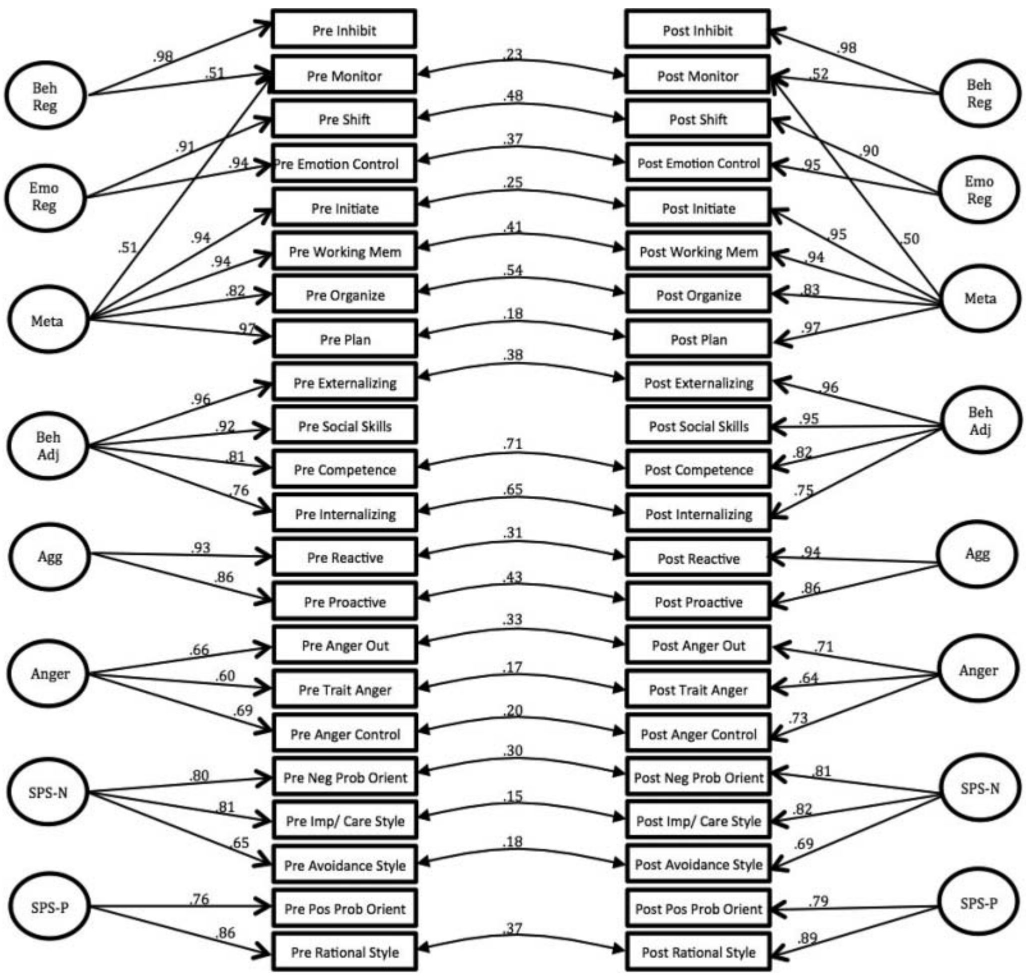
$$\beta_{2j} = \gamma_{20} \quad (4)$$

$$\beta_{3j} = \gamma_{30} \quad (5)$$

$$\beta_{4j} = \gamma_{40} \quad (6)$$

$$\beta_{5j} = \gamma_{50} \quad (7)$$

Figure 1. Standardized Factor Loadings and Significant Residual Correlations of Measurement Model



Note. Agg = Aggression; Beh Adj = Behavioral Adjustment; Beh Reg = Behavior Regulation; Emo Reg = Emotion Regulation; Imp/Care = Impulsive/Careless; Mem = Memory; Meta = Metacognition; Neg Prob Orient = Negative Problem Orientation; Pos Prob Orient = Positive Problem Orientation; SPS-N = Social Problem-Solving–Negative; SPS-P = Social Problem-Solving–Positive.

$$\beta_{6j} = \gamma_{60} \tag{8}$$

Y_{ij} represents the posttest score; β_{0j} represents the intercept; GENDER is the dichotomous variable equal to 1 if the child was a girl and 0 if the child was a boy; RACE is the dichotomous variable equal to 1 if the child was African American and 0 if the child was White or other; FRL is the dichotomous variable equal to 1 if the child received FRL and 0 if the child paid full price for lunch; FCAT READING is the variable equal to the grand mean-centered FCAT reading score; FCAT MATH is the variable equal to the grand mean-centered FCAT math score; GRADE is the dichotomous variable equal to 1 if the child was a fifth grader and 0 if the child was a fourth grader; PRE SCORE represents the pretest grand mean-centered factor score; and CONDITION is the effect coded condition variable equal to –1 for the control group and 1 for the TFGA group. We assumed β_{1j} is fixed, indicating that pretest and condition-by-pretest interaction effects do not vary across clusters.

We used full information maximum likelihood as the estimation approach to handle missing data in Mplus 7.11. Using R software (R Core Team, 2012), we obtained intraclass correlation coefficients (ICCs) to estimate the proportion of variance in outcomes resulting from classroom cluster effects. Per Snijders and Bosker (2012), we computed variance reduction (R^2) and investigated Level 1 and Level 2 residuals, and we created interaction graphs using the coefficients from Mplus outputs.

RESULTS

We analyzed the KQ outcome separately because it served primarily as a check on implementation fidelity (i.e., "treatment received") rather than an indication of curriculum efficacy. With higher scores indicating more knowledge, the control group pretest scores based on 913 observations yielded a mean of 12.23 (SD = 3.02), and posttest scores based on 824 observations yielded a mean of 13.11 (SD = 3.27). For the TFGA group, pretest scores based on 1,085 observations yielded a mean of 12.33 (SD = 3.19), and posttest scores based on 1,001 observations yielded a mean of 21.88 (SD = 5.17). The clustering effect was relatively larger at posttest (ICC = .56) compared with pretest (ICC = .13).

Using a total of 1,790 observations and the same multilevel model, we found a significant pretest effect ($\gamma_{10} = 0.312$, $p < .001$), condition effect ($\gamma_{01} = 4.307$, $p < .001$), and pretest-by-condition interaction effect ($\gamma_{11} = 0.065$, $p = .038$). The R^2_{model} value was .61 and the $R^2_{\text{condition}}$ value was .55, indicating that most of the variance in KQ posttest scores was explained by the condition and condition-by-pretest interaction, as found by Daunic et al. (2012). Furthermore, gender and FCAT scores were significant predictors; girls and higher FCAT scores were associated with higher posttest KQ scores. The FRL variable was also significant, indicating that students who received FRL had lower posttest scores.

TFGA Efficacy for Full Sample

Prior to conducting analyses, we reversed scores where a high score indicated low risk. Thus, in all cluster outcomes, a high score indicates high risk. Pretest and posttest means, standard deviations, and sample sizes are shown in Table 1.

Multilevel model findings are presented in Table 2; statistical significance is based on two-tailed p values. Significant pretest-by-condition interaction effects found for five outcomes, as well as a marginally significant ($p = .058$) interaction effect for one outcome, indicated that TFGA group students with relatively high pretest risk had lower risk after TFGA than comparable controls. Specifically, significant pretest-by-condition interaction effects were evidenced for the Behavior Regulation, Emotion Regulation, Metacognition, Aggression, and Behavioral Adjustment, with a marginally significant interaction effect for the Anger factor. A significant main effect was found for the SPS-P outcome factor, indicating that TFGA group students had a more positive approach to problem solving than control students at posttest. Graphs for significant findings depicted in Figure 2 were created in R using coefficients computed by Mplus.

Table 2 also shows unconditional pretest and posttest ICCs and effect size calculations using R^2 statistics. As expected, teacher-report measures had higher posttest ICC values (.17-.25), as student self-reports (.04-.08) were more likely to be independent of teacher and classroom influences. R^2_{model} values, indicating the total variance in an outcome measure that can be explained by the model (i.e., condition and covariates), were .56 on average for all teacher-report measures and .28 on average for student self-report measures. $R^2_{\text{condition}}$ values indicate how much variation was explained by condition and condition-by-pretest interaction when they were added into the model that included pretest and demographic characteristics (i.e., all remaining predictors).

Table 1. Sample Sizes, Unadjusted Means, and Standard Deviations for Eight Factors

	Control				TFGA			
	Pretest		Posttest		Pretest		Posttest	
Factor	*n*	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*
Beh Reg	927	30.73 (11.12)	852	31.92 (12.13)	1,108	31.56 (11.41)	1,019	31.99 (11.63)
Emo Reg	928	26.25 (9.35)	853	27.31 (10.07)	1,110	27.15 (9.77)	1,019	27.53 (9.79)
Meta	927	67.93 (22.87)	852	70.19 (24.94)	1,110	69.17 (22.60)	1,019	68.86 (22.92)
Beh Adj	920	148.38 (50.05)	857	151.58 (53.37)	1,117	152.07 (50.63)	1,023	151.02 (53.31)
Agg	922	11.22 (5.55)	848	11.98 (5.88)	1,116	11.22 (5.35)	1,026	11.67 (5.56)
Anger	912	51.20 (12.22)	819	51.52 (12.09)	1,094	50.57 (11.76)	993	51.69 (12.11)
SPS-N	913	43.16 (12.11)	844	42.36 (11.40)	1,093	43.79 (11.88)	998	43.47 (11.68)
SPS-P	913	72.64 (19.92)	844	76.64 (19.68)	1,091	71.64 (19.44)	999	72.34 (19.66)

Note. Agg = Aggression; Beh Adj = Behavioral Adjustment; Beh Reg = Behavior Regulation; Emo Reg = Emotion Regulation; Meta = Metacognition; SPS-N = Social Problem-Solving–Negative; SPS-P = Social Problem-Solving–Positive; TFGA = Tools for Getting Along.

We investigated Level 1 and Level 2 standardized residuals and did not detect a substantial threat to the assumption of normal distribution or possible outliers.

Effects of Student Characteristics

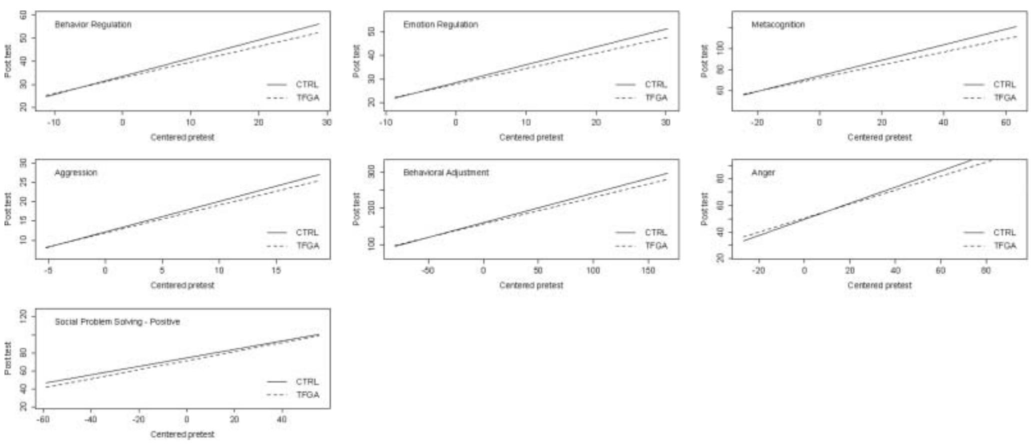
We controlled for effects of individual student characteristics on outcome variables in the hierarchical linear models. Results showed that gender predicted factor scores on Behavior Regulation, Emotion Regulation, Metacognition, Aggression, and Behavioral Adjustment, such that girls evidenced lower posttest risk. Race predicted Behavioral Adjustment, Aggression, Anger, and SPS, such that students who were White or other races evi-

Table 2. Sample Sizes, ICCs, and Estimates for Multilevel Modeling

				Estimates				
Factor	*n*	*ICC* Pretest	*ICC* Posttest	Pretest	Condition	Pretest by Condition	R^2_{model}	$R^2_{\text{condition}}$
Beh Reg	1,824	.21	.17	.732*	–0.322	–0.053*	.56	.01
Emo Reg	1,827	.27	.24	.701*	–0.292	–0.051*	.49	.01
Meta	1,826	.24	.22	.680*	–1.086	–0.058*	.55	.02
Beh Adj	1,826	.25	.25	.771*	–1.895	–0.039*	.65	.01
Agg	1,822	.19	.19	.755*	–0.135	–0.035*	.54	.01
Anger	1,795	.06	.08	.570*	0.452	–0.040†	.32	.01
SPS-N	1,793	.08	.07	.448*	0.378	0.009	.27	.01
SPS-P	1,792	.03	.04	.486*	–1.545*	0.016	.25	.01

Note. The R^2_{model} explained variance by all predictors. The $R^2_{\text{condition}}$ explained variance after adding condition and interaction into the model. Agg = Aggression; Beh Adj = Behavioral Adjustment; Beh Reg = Behavior Regulation; Emo Reg = Emotion Regulation; ICC = unconditional intraclass coefficient; Meta = Metacognition; *n* = number of observations used in Mplus; SPS-N = Social Problem-Solving–Negative; SPS-P = Social Problem-Solving–Positive.
*$p < .05$. †$p = .058$.

Figure 2. Graphs Comparing TFGA and Control Conditions on Variables With Significant Differences



Note. TFGA = Tools for Getting Along.

denced lower posttest risk. FCAT reading predicted Behavior Regulation, Emotion Regulation, Metacognition, and Behavioral Adjustment, indicating that students with higher reading scores showed less posttest risk, and FCAT math predicted scores on the same four factors plus Aggression, indicating that higher math scores were associated with lower risk. FRL predicted Behavior Regulation, with lower posttest risk for students not receiving FRL. Grade predicted Behavior Regulation, Emotion Regulation, Metacognition, Aggression, and Behavioral Adjustment, with fifth-grade students evidencing lower posttest risk. Although we were not interested in exploring the relation between demographic characteristics and outcome variables, these findings justified our inclusion of these variables in our analytic models.

TFGA Efficacy for High-Risk Sample

To create a high-risk sample for each outcome variable, we determined the proportion of our sample that would fall into the "clinical" or at-risk category on the CAB-T by (a) using a linear equation for rescaling CAB-T pretest scores to have a mean of 50 and a standard deviation of 10 and (b) using a score cutoff of 59 (Peters, Kranzler, Algina, Smith, & Daunic, 2014). This proportion was approximately 20%; thus, we created a high-risk subsample for each outcome factor to include students with pretest scores in the top 20% (highest risk). Pretest and posttest means, standard deviations, and sample sizes for the high-risk sample are shown in Table 3. The maximum likelihood estimation procedure with the *nlme* package (Pinheiro, Bates, DebRoy, Sarkar, & R Development Core Team, 2013) in R was used to investigate the following multilevel model:

$$\begin{aligned} Y_{ij} = \gamma_{00} &+ \gamma_{01}(\text{CONDITION})_j \\ &+ \gamma_{02}(\text{GRADE})_j + \gamma_{10}(\text{PRE SCORE})_{ij} \\ &+ \gamma_{20}(\text{RACE})_{ij} + \gamma_{30}(\text{GENDER})_{ij} \\ &+ \gamma_{40}(\text{FRL})_{ij} + \gamma_{50}(\text{FCAT READING})_{Ij} \\ &+ \gamma_{60}(\text{FCAT MATH})_{Ij} + e_{ij}u_{0j} \quad (9) \end{aligned}$$

Y_{ij} represents the posttest score; GRADE, RACE, GENDER, and FRL are dichotomous variables equal to 1 for fifth grader, African American, female, and receiving FRL, respectively. The CONDITION variable is equal to 1 for the TFGA group and 0 for the control group. Pretest and FCAT scores represent grand mean-centered continuous variables. The number of students, number of class-

Table 3. Sample Sizes, Unadjusted Means, and Standard Deviations for High-Risk Sample

	Control			TFGA		
Factor	*n*	Pretest *M (SD)*	Posttest *M (SD)*	*n*	Pretest *M (SD)*	Posttest *M (SD)*
Beh Reg	137	49.48 (5.79)	47.50 (10.36)	199	49.16 (5.99)	45.16 (10.47)
Emo Reg	131	42.43 (6.54)	40.36 (10.60)	210	42.19 (6.39)	37.37 (10.60)
Meta	138	104.13 (12.02)	100.97 (19.48)	184	103.80 (12.08)	95.33 (21.83)
Beh Adj	141	225.15 (24.55)	219.82 (43.99)	200	225.02 (26.70)	210.27 (43.44)
Agg	159	20.21 (3.67)	19.24 (5.54)	191	19.67 (3.23)	18.56 (5.00)
Anger	145	68.71 (6.54)	62.43 (10.93)	177	67.67 (5.39)	60.67 (11.87)
SPS-N	140	61.42 (5.95)	49.91 (10.63)	186	61.11 (6.44)	51.69 (11.28)
SPS-P	144	100.98 (9.37)	90.10 (16.85)	187	99.46 (8.85)	85.04 (18.24)

Note. Agg = Aggression; Beh Adj = Behavioral Adjustment; Beh Reg = Behavior Regulation; Emo Reg = Emotion Regulation; Meta = Metacognition; SPS-N = Social Problem-Solving–Negative; SPS-P = Social Problem-Solving–Positive; TFGA = Tools for Getting Along.

rooms, percentage of singletons, coefficients for pretest and condition, explained variance, and multilevel model-adjusted Hedges's *g* for the condition effect are shown in Table 4. *Singleton* refers to classrooms with one observation. Given that the smallest number of clusters was 91 and the largest proportion of singletons was 32% in our high-risk samples, fixed effect estimates were expected to have no statistical bias (Bell, Ferron, & Kromrey, 2008). The adjusted Hedges's *g* formula was borrowed from the *What Works Clearinghouse: Procedures and Standards Handbook* (Version 2.1; U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse, 2008).

Significant condition effects were found for the Emotion Regulation and SPS-P factors and a marginally significant (p = .10) condition effect was found for the Metacognition

Table 4 Multilevel Model Results for High-Risk Sample

Factor	*n*	*j*	Singleton, %	Pretest	Condition	R^2_{model}	Hedges's *g*
Beh Reg	336	103	28	.64*	–1.42	.18	–0.14
Emo Reg	341	91	23	.67*	–3.04*	.22	–0.27
Meta	322	99	29	.49*	–4.67†	.17	–0.22
Beh Adj	341	99	17	.91*	–6.91	.34	–0.16
Agg	350	106	24	.66*	–0.29	.25	–0.05
Anger	322	111	27	.42*	–2.05	.07	–0.18
SPS-N	326	113	32	.47*	1.97	.09	0.17
SPS-P	331	110	24	.44*	–4.08*	.09	–0.23

Note. *Singleton* refers to the proportion of clusters with one observation. The R^2_{model} explained variance by all predictors. Hedges's *g* indicates the effect size for condition, adjusted for multilevel data structure. Agg = Aggression; Beh Adj = Behavioral Adjustment; Beh Reg = Behavior Regulation; Emo Reg = Emotion Regulation; *j* = number of classrooms; Meta = Metacognition; SPS-N = Social Problem-Solving–Negative; SPS-P = Social Problem-Solving–Positive.
*$p < .05$. $^{\dagger}p = .10$.

factor, each indicating that TFGA group students had lower posttest risk than controls on these outcomes.

Social Validity

As reported in the study by Daunic et al. (2012), we examined social validity with the TFGA Teacher Questionnaire. The teachers found TFGA relatively easy to implement for the Appeal, Ease of Use, and Efficacy subscales and reported moderately positive effects on student behavior.

DISCUSSION

The purpose of this study was to examine further the effects of TFGA, reported by Daunic et al. (2012), on the social and behavioral outcomes of upper elementary students using an extended participant sample and a different analytic procedure. Given that our study is a partial replication, our hypotheses remained the same about TFGA-related improvements for teacher reports of executive functioning, behavior problems, and aggression, as well as student self-reported problem solving and anger. We also wanted to investigate whether TFGA would have effects specifically for students with the highest risk (top 20%) on a given outcome measure. Results of multilevel analyses on the total sample indicated positive effects of TFGA on seven of eight measurement factors, consistent with results obtained by Daunic et al. (2006, 2012). We also found main effects for the high-risk subsamples on the Emotion Regulation and SPS-P factors and a marginal effect on the Metacognition factor. Collectively, the results of the two Daunic et al. studies and the present investigation serve to extend the evidence base for universally delivered CBIs for improving social–emotional and cognitive functioning and reducing maladaptive behavior patterns among students at risk for emotional and behavioral disorders. (Note that although we found that demographic characteristics were associated with outcome variables, an exploration of the relation of these variables to intervention effects was beyond the scope of this study.)

Effects on Social Problem-Solving

Consistent with the previous positive findings for the individual SPSI-R subscales of PPO and Rational Problem Solving (Daunic et al., 2012), we found a significant main effect of TFGA for the SPS-P factor. The effect of TFGA on a positive approach to problem solving is logical, given the program's explicit focus on developing students' abilities to appraise social problems as challenges and to maintain self-efficacy about their ability to solve them. This finding is also consistent with the literature suggesting that orientation influences responsiveness to problem situations (Allen, Leadbeater, & Aber, 1994; Catalano & Hawkins, 1996; D'Zurilla, Maydeu-Olivares, & Gallardo-Pujol, 2011; D'Zurilla & Nezu, 2010; Kuperminc & Allen, 2001). Regarding a rational problem-solving style, the systematic application of a solution sequence (e.g., problem definition, solution generation, decision making, solution implementation, and verification) is at the core of TFGA lessons and serves as the basis for practice activities and role-playing. Thus, gathering facts and information about a problem, identifying obstacles, setting realistic goals, generating a variety of realistic solutions, comparing alternatives, choosing a "best" solution, and monitoring and evaluating the outcome are specific skills covered by TFGA lessons.

Effects on EF

Our findings also indicate positive effects of TFGA for students with relatively poor pretest scores on Behavior Regulation, Emotion Regulation, and Metacognition, which are factors that reflect refinements made to the original BRIEF instrument, such that emotion regulatory processes are distinct from Behavior Regulation (see, e.g., Gioia, Isquith, Retzlaff, & Espy, 2002; Peters et al., 2012). The pretest-by-condition interaction for Behavior Regulation indicates that TFGA served to reduce teachers' perceptions of impulsivity and improve their perceptions of students' ability to self-monitor their behavior. This finding helps substantiate the significant pretest-by-condition interaction effect found by Daunic et al. (2012) for the inhibition subscale specifically.

A significant pretest-by-condition interaction effect was also found for Emotion Regulation (Shift, Emotion Control), indicating that students with less initial teacher-perceived ability to regulate their emotions benefited from exposure to TFGA instruction relative to comparable peers in the control condition. Thus, TFGA students improved their ability to respond flexibly to problems and to avoid outbursts related to negative emotions. Explicit TFGA instruction related to developing multiple solutions to problems, recognizing the physiological signs of emotions, and using strategies for calming down is consistent with these improvements.

Finally, we also found a significant, positive pretest-by-condition interaction effect for Metacognition, which includes Initiate, Working Memory, Organize, Plan, and Monitor subscales. Whereas Daunic et al. (2012) reported positive treatment results for the Initiate, Organize, and Monitor subscales, the Metacognition factor in the current study also includes Working Memory and Plan, thus representing a broad conceptualization of metacognition as higher order processes governing the deployment of self-control (see Duckworth, Gendler, & Gross, 2014). As such, current findings indicate treatment-induced improvements beyond the age-related changes in metacognition noted by other authors (e.g., Cunningham, Zelazo, Packer, & Van Bavel, 2007; Dimmitt & McCormick, 2012). The modeling and practice opportunities in TFGA lessons apparently helped students with poorer pretest scores develop critical metacognitive skills. Given the importance of developing skills such as self-talk for improving behavioral control (Carlson & Beck, 2009; Diamond & Lee, 2011), the Metacognition finding indicates that TFGA has promise for improving foundational cognitive processes that facilitate behavioral self-control.

Effects on Aggression

We found a significant pretest-by-condition interaction on the Aggression factor, indicating that TFGA participants with relatively higher teacher-rated aggression (combined proactive and reactive) showed a significant decrease in level of aggression as compared with comparable controls. This finding partially replicates the Daunic et al. (2012) finding that TFGA had positive effects on proactive, but not reactive, aggression for students with high pretest scores. The increased sample size may have improved the ability to detect change in total aggression because effects on the Reactive Aggression subscale in the 2012 study approached significance. In addition, the current findings support and are consistent with those reported by Daunic et al. (2006), in that positive effects on both reactive and proactive aggression were evidenced for students in the highest risk quartile.

Effects on Social–Emotional Behavior

We found a significant pretest-by-condition interaction effect for Behavioral Adjustment, also indicating that TFGA group students with higher pretest scores on this factor showed lower risk at posttest relative to comparable controls. This finding extends the results of Daunic et al. (2012), in that there were no significant positive TFGA-by-pretest interaction effects on either teacher-rated Internalizing Behavior or Externalizing Behavior in their report, although the findings for the pretest-by-condition interaction on Internalizing Behavior approached significance. The Behavioral Adjustment factor in the current study included both of these subscales.

Internal reliabilities for the Social Skills and Competence subscales were also acceptable in the current study and thus included in the measurement model, whereas these subscales were eliminated from the study of Daunic et al. (2012) because of low reliability estimates. Consequently, the Behavioral Adjustment factor consisted of Internalizing Behavior, Externalizing Behavior, Social Skills, and Competence subscales and thereby constitutes a teacher-report measure of students' overall adjustment, including two types of problem behavior: interpersonal interaction (social) skills and ability to function competently in the school environment (Bracken & Keith, 2004). The positive finding that TFGA affected the behavioral adjustment of students with higher pretest risk is consistent with evidence from other large-scale studies indicating that CBIs can prevent or ameliorate behavioral problems and strengthen self-regulatory functioning in the school environment (e.g., Daunic et al., 2006; Mayer et al., 2009; McCart et al., 2006).

Effects on Anger

The significant pretest-by-condition interaction effect found for the Anger factor, indicating that students with higher baseline risk presented lower risk at posttest relative to comparable controls, partially replicates the positive effect on student self-reports of anger reported by Daunic et al. (2012). Whereas Daunic et al. found effects on the Trait Anger (general predisposition or temperament to become angry) and Anger-Out (expression of anger overtly in a negative and poorly controlled manner) subscales, the Anger factor in the current measurement model included Trait Anger, Anger-Out, and Anger Control (monitoring and controlling overt expressions of anger). The finding related to anger is not surprising, considering that TFGA instruction includes an attendant focus on anger management, a necessary component of emotion regulation in highly emotional social conflicts. Throughout TFGA, there is continual attention to anger awareness and levels or degrees of anger, techniques for calming down and managing anger, and opportunities through numerous role-playing events to practice skills using social scenarios in which anger is evident.

Effects for Students With Highest Risk

In the current study, we also analyzed TFGA effects for the 20% of students at highest risk on each of the eight factors, thus creating factor-specific high-risk subsamples. Using these subsamples, we found main effects of TFGA on the Emotion Regulation and SPS-P factors. High-risk groups for these factors evidenced significant improvements in their ability to control emotions and shift between response options (regulate emotions) and to appraise social problems appropriately, maintain efficacy about their ability to solve problems, and use a rational problem-solving style, after exposure to TFGA. In addition, there was a marginally significant effect on Metacognition, suggesting that TFGA also may have the potential to improve critical metacognitive skills that support behavioral self-control for students at highest risk. The fact that we did not see evidence that the intervention was effective in reducing risk on several other factors (e.g., self-reported anger management and aggression) suggests that students with higher levels of risk may need more intensive interventions, such as those provided at a Tier 2 (secondary intervention) level in an MTSS model. Further research could include investigating how the involvement of school psychologists in delivering CBIs such as TFGA to small groups of students might strengthen their preventive effects. According to Splett, Fowler, Weist, McDaniel, and Dvorsky (2013), school psychologists are “positioned to be in critical leadership roles” (p. 255) under MTSS by ensuring that effective and targeted school-based intervention services are implemented, supported, and monitored.

Limitations and Future Research

Although an additional year of posttest outcome data and a different estimation model have enhanced the interpretation and application of current findings, this partial replication study has some of the same limitations noted by Daunic et al. (2012) in their randomized controlled study. First, the differential consent rates for TFGA and control groups introduces the potential for bias (see *What Works Clearinghouse: Procedures and Standards Handbook*; U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse, 2008), which we could not assess given the lack of baseline data from nonconsenting students. A higher consent rate would have diminished the threat of bias, even though we have no reason to believe that attrited students differed systematically from study participants.

Second, the issue of expectancy effects on the part of teachers who implemented the intervention and completed outcome measures is a legitimate concern. Although this is no small matter, using an assessment to query teacher perceptions about curriculum effects seems reasonable and perhaps unavoidable. Moreover, we agree with Daunic et al. (2012) that the impressions of teachers who implement an intervention are still valuable, as they are authentically contextualized in the classroom and occur over an extensive period. A change in teacher perception of student behaviors is thus noteworthy, especially considering that regular-education teachers are often progenitors of referral for special-education services (Daunic et al., 2006; Ollendick & King, 1999; Sutherland & Oswald, 2005).

We also relied on student self-reports that are subject to bias because responses can be influenced by individual respondent characteristics, as suggested by Phipps and Steele (2002). Although combined teacher- and student-report data strengthen our results, future studies would benefit from additional informants such as parents and particularly from direct observations of student behavior. Furthermore, investigations of the relations among outcome measures associated with risk for emotional and behavioral difficulties using statistical methods such as path analysis could help illuminate the role of mediating (e.g., EF) and moderating (e.g., gender, socioeconomic status) variables.

On the positive side, the increase in sample size and use of a different estimation model satisfied several limitations, advancing our interpretation and application of findings. First, whereas Daunic et al. (2012) noted low reliability estimates for the problem-solving knowledge measure, the current study yielded a Cronbach's α of .74 at posttest, thus enhancing the utility of the findings. Second, we reduced the number of multiple related outcome variables and the possibility of an inflated Type I error rate by using the eight-factor measurement model.

We note the need for future studies. Although there were significant positive outcomes associated with TFGA for students with scores in the top 20% risk category, we did not find positive TFGA effects on more than three outcome factors, despite significant pretest-by-condition interaction effects on seven of these factors for the full-sample data. Researchers, therefore, might focus on how additional social problem-solving skill development and substantive practice opportunities that could be provided through small-group settings might strengthen results for students with more pronounced risk. It may be that students at risk for significant behavior problems need more focused instruction for skill acquisition through practice, teacher modeling, and reinforcement to enhance TFGA effectiveness and promote skill generalization.

Finally, Daunic et al. (2012) suggested that studies are needed to determine whether TFGA effects can be sustained after treatment. In response, Smith et al. (2014) collected follow-up data 1 year after the conclusion of TFGA instruction and determined that several TFGA effects were evident, including positive effects on curricular knowledge for students taught TFGA relative to controls and pretest-by-condition interaction effects on teacher reports of skills associated with several EF skills and on teacher-reported internalizing and externalizing behavior. More follow-up studies would continue to explore the effects of CBI instruction over extended periods without periodic skill reinforcement.

REFERENCES

Allen, J., Leadbeater, A., & Aber, J. L. (1994). The development of problem behavior syndromes in at-risk adolescents. *Development and Psychopathology, 6*, 323–342. doi:10.1017/S0954579400004612

Bahr, H. M., Caplow, T., & Chadwick, B. A. (1983). Middletown III: Problems of replication, longitudinal measurement, and triangulation. *Annual Review of Sociology, 9*, 243–264. doi:10.1146/annurev.so.09.080183.001331

Bell, B. A., Ferron, J. M., & Kromrey, J. D. (2008). Cluster size in multilevel models: The impact of sparse data structures on point and interval estimates in two-level models. In: *JSM proceedings, section on survey research methods* (pp. 1122–1129). Alexandria, VA: American Statistical Association. http://www.amstat.org/Sections/Srms/Proceedings/y2008/Files/300933.pdf

Bracken, B. A., & Keith, L. K. (2004). *Clinical assessment of behavior*. Lutz, FL: Psychological Assessment Resources.

Burman, L. E., Reed, W. R., & Alm, J. (2010). A call for replication studies. *Public Finance Review, 38*, 787–793. doi:10.1177/1091142110385210

Cantin, R. H., Mann, T. D., & Hund, A. M. (2012). Executive functioning predicts school readiness and success: Implications for assessment and intervention. *Communiqué, 41*, 1, 20–21.

Carlson, S. M., & Beck, D. M. (2009). Symbols as tools in the development of executive functioning. In A. Winsler, C. Fernyhough, & I. Montero (Eds.), *Private speech, executive functioning, and the development of verbal self-regulation* (pp. 163–175). Cambridge, UK: Cambridge University Press.

Catalano, R. F., & Hawkins, J. D. (1996). The social development model: A theory of antisocial behavior. In J. D. Hawkins (Ed.), *Delinquency and crime: Current theories* (pp. 149–197). New York: Cambridge University Press.

Conduct Problems Prevention Research Group. (1999). Initial impact of the Fast Track prevention trial for conduct problems: II. Classroom effects. *Journal of Consulting and Clinical Psychology, 67*, 648–657.

Conduct Problems Prevention Research Group. (2004). The effects of the Fast Track program on serious problem outcomes at the end of elementary school. *Journal of Clinical Child and Adolescent Psychology, 33*, 650–661.

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin, 115*, 74–101. doi:10.1037/0033-2909.115.1.74

Cunningham, W. A., Zelazo, P. D., Packer, D. J., & Van Bavel, J. J. (2007). The iterative reprocessing model: A multi-level framework for attitudes and evaluation. *Social Cognition, 25*, 736–760. doi:10.1521/soco.2007.25.5.736

Darley, W. K. (2000). Status of replication studies in marketing: A validation and extension. *Marketing Management Journal, 10*, 121–132.

Daunic, A. P., Smith. S. W., Brank, E. M., & Penfield, R. D. (2006). Classroom based cognitive–behavioral intervention to prevent aggression: Efficacy and social validity. *Journal of School Psychology, 44*, 123–139. doi:10.1016/j.jsp.2006.01.005

Daunic, A. P., Smith, S. W., Garvan, C. W., Barber, B. R., Becker, M. K., Peters, C. D., . . . Naranjo, A. H. (2012). Reducing developmental risk for emotional/behavioral problems: A randomized controlled trial examining the Tools for Getting Along curriculum. *Journal of School Psychology, 50*, 149–166. doi:10.1016/j.jsp.2011.09.003

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4–12 years old. *Science, 333*, 959–964. doi:10.1126/science.1204529

Dimmitt, C., & McCormick, C. B. (2012). Metacognition in education. In K. R. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook* (Vol. 1, pp. 157–187). Washington, DC: American Psychological Association. doi:10.1037/13273-007

Dodge, K. A. (1986). A social information processing model of social competence in children. In M. Perlmutter (Ed.), *Minnesota symposium on child psychology* (Vol. 18, pp. 77–125). Hillsdale, NJ: Erlbaum.

Dodge, K. A., & Coie, J. D. (1987). Social-information-processing factors in reactive and proactive aggression in children's peer groups. *Journal of Personality and Social Psychology, 53*, 1146–1158. doi:10.1037/0022-3514.53.6.1146

Dodge, K. A., Lansford, J. E., Burks, V. S., Bates, J. E., Pettit, G. S., Fontaine, R., & Price, J. M. (2003). Peer rejection and social information processing factors in the development of aggressive behavior problems in children. *Child Development, 74*, 374–393. doi:10.1111/1467-8624.7402004

Dodge, K. A., & Somberg, D. (1987). Hostile attributional biases among aggressive boys as exacerbated under conditions of threat to self. *Child Development, 58*, 213–224. doi:10.2307/1130303

Donnellan, M. B., Trzesniewski, K. H., Robins, R. W., Moffitt, T. E., & Caspi, A. (2005). Low self-esteem is related to aggression, antisocial behavior, and delinquency. *Psychological Science, 16*, 328–335. doi:10.1111/j.0956-7976.2005.01535.x

Duckworth, A. L., Gendler, T. S., & Gross, J. J. (2014). Self-control in school-age children. *Educational Psychologist, 49*, 199–217. doi:10.1080/00461520.2014.926225

D'Zurilla, T. J., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2011). Predicting social problem solving using personality traits. *Personality and Individual Differences, 50*, 142–147. doi:10.1016/j.paid.2010.09.015

D'Zurilla, T. J., & Nezu, A. M. (2010). Problem-solving therapy. In K. S. Dobson (Ed.), *Handbook of cognitive-behavioral therapies* (3rd ed., pp. 197–225). New York, NY: Guilford Press.

D'Zurilla, T. J., Nezu, A. M., & Maydeu-Olivares, A. (2002). *Social Problem-Solving Inventory-Revised (SPSI-R): Technical manual*. North Tonawanda, NY: Multi-Health Systems. doi:10.1037/t05068-000

D'Zurilla, T. J., Nezu, A. M., & Maydeu-Olivares, A. (2004). Social problem solving: Theory and assessment. In E. C. Chang, T. J. D'Zurilla, & L. J. Sanna (Eds.), *Social problem solving: Theory, research, and training* (pp. 11–27). Washington, DC: American Psychological Association. doi:10.1037/10805-001

Gioia, G. A., Isquith, P. K., Kenworthy, L., & Barton, R. M. (2002). Profiles of everyday executive function in acquired and developmental disorders. *Child Neuropsychology, 8*, 121–137. doi:10.1076/chin.8.2.121.8727

Gioia, G. A., Isquith, P. K., Retzlaff, P. D., & Espy, K. A. (2002). Confirmatory factor analysis of the Behavior Rating Inventory of Executive Function (BRIEF) in a clinical sample. *Child Neuropsychology, 8*, 249–257. doi:10.1076/chin.8.4.249.13513

Hendrick, C. (1991). Replication, strict replications, and conceptual replications: Are they important? In J. W. Neuliep (Ed.), *Replication research in the social sciences* (pp. 41–49). Newbury Park, CA: Sage.

Hofmann, W., Schmeichel, B. J., & Baddeley, A. D. (2012). Executive functions and self regulation. *Trends in Cognitive Sciences, 16*, 174–180. doi:10.1016/j.tics.2012.01.006

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55. doi:10.1080/10705519909540118

Hughes, C. (2002). Executive functions and development: Why the interest? *Infant and Child Development, 11*, 69–71. doi:10.1002/icd.296

Kendall, P. C., & Braswell, L. (1985). *Cognitive-behavioral therapy for impulsive children.* New York, NY: Guilford Press.

Kuperminc, G. P., & Allen, J. P. (2001). Social orientation: Problem behavior and motivations toward interpersonal problem solving among high-risk adolescents. *Journal of Youth and Adolescence, 30*, 597–622. doi:10.1023/A:1010452705746

Lerner, R. M., & Steinberg, L. (Eds.). (2004). *Handbook of adolescent psychology* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Lewis, C., & Carpendale, J. I. M. (2009). Introduction: Links between social interaction and executive function. *New Directions in Child and Adolescent Development,* 123, 1–15. doi:10.1002/cd.232

Lochman, J. E., & Lampron, L. B. (1986). Situational problem-solving and self-esteem of aggressive and nonaggressive boys. *Journal of Abnormal Child Psychology, 14*, 605–617. doi:10.1007/BF01260527

Lochman, J. E., & Wells, K. C. (2004). The Coping Power program for preadolescent aggressive boys and their parents: Outcome effects at the one-year follow-up. *Journal of Consulting and Clinical Psychology, 72*, 571–578.

Lyons, K. E., & Zelazo, P. D. (2011). Monitoring, metacognition, and executive function: Elucidating the role of self-reflection in the development of self-regulation. *Advances in Child Development and Behavior, 40*, 379–412. doi:10.1016/B978-0-12-386491-8.00010-4

Maplesoft [Computer software]. (2005–2015). Maple. Retrieved from http://www.maplesoft.com/

Mayer, M., Van Acker, R., Lochman, J., & Gresham, F. (2009). *Cognitive–behavioral interventions for emotional and behavioral disorders: School-based practice.* New York, NY: Guilford Press.

McCart, M. R., Priester, P. E., Davies, W. H., & Azen, R. (2006). Differential effectiveness of behavioral parent-training and cognitive-behavioral therapy for antisocial youth: A meta-analysis. *Journal of Abnormal Child Psychology, 34*, 527–543. doi:10.1007/s10802-006-9031-1

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Author.

National Association of School Psychologists. (2006). *Social/emotional development: School-based mental health services and school psychologists.* Retrieved from http://www.nasponline.org/resources/handouts/sbmhservices.pdf

Ollendick, T. H., & King, N. (1999). Child behavioral assessment and cognitive–behavioral interventions in schools. *Psychology in the Schools, 36*, 427–436. doi:10.1002/(SICI)1520-6807(199909)36:5<427::AID-PITS6>3.0.CO;2-H

Olson, S. L., Sameroff, A. J., Kerr, D. C. R., Lopez, N. L., & Wellman, H. M. (2005). Developmental foundations of externalizing problems in young children: The role of effortful control. *Development and Psychopathology, 17*, 25–45. doi:10.1017/S0954579405050029

Peters, C., Algina, J., Smith, S. W., & Daunic, A. (2012). Factorial validity of the Behavior Rating Inventory of Executive Function (BRIEF) teacher form. *Child Neuropsychology, 18*, 168–181. doi:10.1080/09297049.2011.594427

Peters, C. D., Kranzler, J., Algina, J., Smith, S. W., & Daunic, A. P. (2014). Understanding disproportionate representation in special education by examining group differences in behavior ratings. *Psychology in the Schools, 51*, 452–465. doi:10.1002/pits.21761

Phipps, S., & Steele, R. (2002). Repressive adaptive style in children with chronic illness. *Psychosomatic Medicine, 64*, 34–42. doi:10.1097/00006842-200201000-00006

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Development Core Team. (2013). *nlme: Linear and nonlinear mixed effects models.* R package version 3.1–111.

Polsgrove, L., & Smith, S. W. (2004). Informed practice in teaching self-control to children with emotional and behavioral disorders. In R. B. Rutherford, M. M. Quinn, & S. R. Mathur (Eds.) *Handbook of research in emotional and behavioral disorders* (pp. 399–425). New York, NY: Guilford Press.

R Core Team. (2012). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Robinson, T. R., Smith, S. W., Miller, M. D., & Brownell, M. T. (1999). Cognitive behavior modification of hyperactivity/impulsivity and aggression: A meta-analysis of school-based studies. *Journal of Educational Psychology, 91*, 195–203. doi:10.1037/0022-0663.91.2.195

Rubin, K. H., Bream, L., & Rose-Krasnor, L. (1991). Social problem solving and aggression in childhood. In D. J. Pepler & K. H. Rubin (Eds.), *The development and treatment of childhood aggression* (pp. 219–248). Hillsdale, NJ: Erlbaum.

Rueda, M. R., Acosta, A., & Santonja, M. (2007). *Attention and motivation interdependence in self-regulation: A neurocognitive approach.* Hauppauge, NY: Nova Science.

Rueda, M. R., Posner, M. I., & Rothbart, M. K. (2005). The development of executive attention: Contributions to the emergence of self-regulation. *Developmental Neuropsychology, 28*, 573–594. doi:10.1207/s15326942dn2802_2

Smith, S. W., & Daunic, A. P. (2006). *Managing difficult behavior through problem solving instruction: Strategies for the elementary classroom.* Boston, MA: Allyn & Bacon.

Smith, S. W., Daunic, A. P., Barber, B. R., Aydin, B., Van Loan, C. L., & Taylor, G. G. (2014). Preventing risk for significant behavior problems through a cognitive-behavioral intervention: Effects of the Tools for Getting Along curriculum at one-year follow-up. *Journal of Primary Prevention, 35*, 371–387. doi:10.1007/s10935-014-0357-0

Smith, S. W., Graber, J., & Daunic, A. P. (2009). Cognitive–behavioral interventions for anger/aggression: Review of research and research-to-practice issues. In M. Mayer, R. Van Acker, J. Lochman, & F. Gresham (Eds.), *Cognitive–behavioral interventions for emotional and behavioral disorders: School-based practice* (pp. 111–142). New York, NY: Guilford Press.

Smith, S. W., Lochman, J. E., & Daunic, A. P. (2005). Managing aggression using cognitive–behavioral interventions: State of the practice and future directions. *Behavioral Disorders, 30*, 227–240.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* Los Angeles, CA: Sage.
Speilberger, C. D. (1988). *State-trait anger expression inventory: Professional manual.* Odessa, FL: Psychological Assessment Resources.
Splett, J. W., Fowler, J., Weist, M. D., McDaniel, H., & Dvorsky, M. (2013). The critical role of school psychology in the school mental health movement. *Psychology in the Schools, 50*, 245–258. doi:10.1002/pits.21677
Steele, R. G., Legerski, J. P., Nelson, T. D., & Phipps, S. (2009). The Anger Expression Scale for Children: Initial validation among healthy children and children with cancer. *Journal of Pediatric Psychology, 34*, 51–62. doi:10.1093/jpepsy/jsn054
Sukhodolsky, D. G., Kassinove, H., & Gorman, B. S. (2004). Cognitive-behavioral therapy for anger in children and adolescents: A meta-analysis. *Aggression and Violent Behavior, 9*, 247–269. doi:10.1016/j.avb.2003.08.005
Sutherland, K. S., & Oswald, D. P. (2005). The relationship between teacher and student behavior in classrooms for students with emotional and behavioral disorders: Transactional processes. *Journal of Child and Family Studies, 14*, 1–14. doi:10.1007/s10826-005-1106-z
U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2008). *What Works Clearinghouse: Procedures and standards handbook (version 2.1).* Washington, DC: Author.
van IJzendoorn, M. H. (1994). Process model of replication studies: On the relations between different types of replication. In R. van der Veer, M. H. van IJzendoorn, & J. Valsiner (Eds.), *On reconstructing the mind. Replicability in research on human development* (pp. 57–70). Norwood, NJ: Ablex.
Wehmeyer, M. L., Palmer, S., Shogren, K., Williams-Diehm, K., & Soukup, J. (2013). Establishing a causal relationship between interventions to promote self-determination and enhanced student self-determination. *Journal of Special Education, 46*, 195–210. doi:10.1177/0022466910392377
Zelazo, P. D., Carlson, S. M., & Kesek, A. (2008). Development of executive function in childhood. In C. A. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (2nd ed., pp. 553–574). Cambridge, MA: MIT Press.

Date Received: September 25, 2014
Date Accepted: June 6, 2015
Associate Editor: Lisa Bowman-Perrott
Article accepted by previous Editor ■

Stephen W. Smith is a professor in the Special Education Program, College of Education, University of Florida, Gainesville, Florida.

Ann P. Daunic is a scholar in the Special Education Program, College of Education, University of Florida, Gainesville, Florida.

Burak Aydin is an assistant professor, Department of Educational Sciences, College of Education, Recep Tayyip Erdoğan University, Rize, Turkey.

Christopher L. Van Loan is an associate professor in the Special Education Program, Reich College of Education, Appalachian State University, Boone, North Carolina.

Brian R. Barber is an assistant professor of special education, School of Lifespan Development and Educational Sciences, Kent State University, Kent, Ohio.

Gregory G. Taylor is Teaching Faculty I, School of Teacher Education, Florida State University, Tallahassee, Florida.