

2006

Classroom-Based Cognitive-Behavioral Intervention to Prevent Aggression: Efficacy and Social Validity

Ann P. Daunic

University of Florida, adaunic@coe.ufl.edu

Stephen W. Smith

University of Florida, swsmith@coe.ufl.edu

Eve M. Brank

University of Nebraska-Lincoln, ebrank2@unl.edu

Randall D. Penfield

University of Miami, penfield@miami.edu

Follow this and additional works at: <http://digitalcommons.unl.edu/psychfacpub>



Part of the [Psychiatry and Psychology Commons](#)

Daunic, Ann P.; Smith, Stephen W.; Brank, Eve M.; and Penfield, Randall D., "Classroom-Based Cognitive-Behavioral Intervention to Prevent Aggression: Efficacy and Social Validity" (2006). *Faculty Publications, Department of Psychology*. 553.
<http://digitalcommons.unl.edu/psychfacpub/553>

This Article is brought to you for free and open access by the Psychology, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications, Department of Psychology by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Classroom-Based Cognitive-Behavioral Intervention to Prevent Aggression: Efficacy and Social Validity

Ann P. Daunic^{a,*}, Stephen W. Smith^a, Eve M. Brank^{a,§}, Randall D. Penfield^b

^aUniversity of Florida, Gainesville, Florida, USA; ^bUniversity of Miami, Coral Gables, Florida, USA. E-mail addresses: adaunic@coe.ufl.edu; swsmith@coe.ufl.edu; ebrank2@unl.edu; penfield@miami.edu. *Corresponding author; [§]Affiliation 2009-: University of Nebraska-Lincoln (UNL), Lincoln, Nebraska, USA.

Received June 1, 2005; received in revised form January 16, 2006; accepted January 20, 2006; posted in the UNL repository March 2012.

Abstract

Classroom teachers need effective, efficient strategies to prevent and/or ameliorate destructive student behaviors and increase socially appropriate ones. During the past two decades, researchers have found that cognitive strategies can decrease student disruption/aggression and strengthen pro-social behavior. Following preliminary pilot work, we conducted a study to determine whether a class-wide, social problem-solving curriculum affected measures of knowledge and behavior for 165 4th and 5th grade students at risk for behavior problems. We found significant positive treatment effects on knowledge of problem-solving concepts and teacher ratings of aggression. Outcomes differed across teachers/classrooms, and there was no evidence that booster lessons affected treatment efficacy. Teacher ratings of social validity were generally positive. We discuss issues about classroom-based prevention research and future research directions.

Keywords: At risk populations; Cognitive-behavioral intervention; Externalization; Aggressive behavior problems; Elementary school students; Prevention.

The proactive assurance of safe and productive school environments is a critical objective for education policy makers. Implementing preventive interventions may be particularly challenging in schools with a high proportion of children at risk for academic failure and conduct problems (Bierman, Greenberg, & CPPRG, 1996). The pressure to improve student academic performance and standardized test scores creates significant demands that compete with social programs for instructional time. Moreover, school administrators and classroom teachers need access to evidenced-based practices to avoid adopting prevention programs that are intuitively appealing but unsubstantiated by empirical research (see e.g., Vaughn & Dammann, 2001).

Cullinan (2002) describes universal prevention as the application of interventions to a broadly defined group (e.g., classroom) to reduce risk and maintain student health and safety. Selective prevention involves activities designed for groups of students who may share characteristics that put them at risk for developing mental disorders or a school diagnosis of emotional or behavioral disorders (EBD). Thus, classroom-based interventions can act as both universal and selective prevention. Walker, Colvin, and Ramsey (1995) maintain that universally implemented (e.g., classwide) interventions are especially effective for students who are "on the margins," or beginning to behave in ways that compromise their future school success, noting that such interventions allow children with or at risk for EBD to learn effective coping strategies with support from socially appropriate peers.

One approach to the prevention of behavior disorders is the classwide application of a cognitive-behavioral intervention (CBI). A research-based approach to teaching students positive coping strategies (e.g., Kendall & Braswell, 1985; Lochman, Whidby, & FitzGerald, 2000; Robinson, Smith, Miller, & Brownell, 1999), CBIs incorporate behavior therapy techniques such as modeling, feedback, and reinforcement, and cognitive mediation techniques such as "think-alouds" to build what Kendall (1993) called a new coping template. The underlying assumptions are that overt behavior is mediated by cognitive events and that people can learn to influence cognitive events to change their own behavior. Literature reviews and meta-analyses (Abikoff, 1991; Dush, Hirt, & Schroeder, 1989; Robinson et al., 1999; Smith, Lochman, & Daunic, 2005) have substantiated CBI's usefulness for the prevention and remediation of specific behavioral deficits and the maintenance of appropriate behavior for mainstream students. Teaching students cognitive strategies has been found to decrease hyperactivity/impulsivity and disruption/aggression, strengthen pro-social behavior, increase social cognition, and improve peer relations (cf. Ager & Cole, 1991; CPPRG, 2002a,b; Dodge, 1986; Lochman, Coie, Underwood, & Terry, 1993; Robinson, Smith, & Miller, 2002; Smith, Siegel, O'Connor, & Thomas, 1994).

In addition to efficacy, researchers are necessarily concerned with intervention efficiency and sustainability. Although the degree of exposure needed to achieve and maintain a desired behavioral effect has been difficult to specify because of the variability in intervention packages, longer treatments generally tend to result in better outcomes (e.g., Heinicke, 1988; Waltman & Zimpfer, 1988; Whalen, Henker, & Hinshaw, 1985). For example, Lochman (1985) manipulated duration of exposure to an anger coping program and found that students who received one lesson per week for 18 as opposed to 12 weeks achieved greater gains. More recently, Larson and Lochman (2002) noted that booster sessions designed to supplement and reinforce initial instructional content helped sustain student learning and improvements in pro-social behavior.

Research Purpose

In light of the CBI literature and the need for evidenced-based practice as mandated by the No Child Left Behind Act (2001), we developed, piloted, and investigated the efficacy of *Tools for Getting Along (TFGA): Teaching Students to Problem Solve*, a cognitive-behavioral social problem-solving curriculum. Our purpose was to determine whether a CBI implemented by classroom teachers could engender

and sustain positive outcomes for selected students at risk for disruptive and/or aggressive behavior. Thus, the intervention was implemented in a classroom setting where students with emerging behavior problems could be taught social problem-solving skills alongside typical peers.

Prior to this investigation, we conducted a preliminary study that involved 76 target students in 17 classrooms who were (a) nominated by teachers as having, or being at risk for developing, disruptive or aggressive behavior patterns and (b) differed from typical peers on pre-treatment measures of anger and externalizing behaviors (Daunic, Smith, & Miller, in press). Using a treatment/wait-list control group design (Cook & Campbell, 1979), we found that exposure to TFGA significantly increased knowledge of problem-solving concepts for both target students and their typical peers and decreased teacher ratings of target students' reactive and proactive aggression; that is, target students' ratings on teacher reports of aggression were more similar to those of non-targets after intervention.

Although these changes in knowledge and behavior following exposure to TFGA were positive, we did not find change on target student self-reports of anger expression. Lack of consistent findings across measures may have resulted from limited treatment efficacy, but it could also have been a consequence of several design and/or methodological issues. Students were taught only 15 lessons in a 5-8 week period, and the experimental groups were self-selecting, in that teachers chose to implement the curriculum in the fall or spring, placing them in either the treatment or wait-list/control group. Moreover, there were considerable between-group demographic differences that could have differentially affected response to treatment. Finally, the treatment/wait-list control design precluded replication of a treatment vs. control comparison, and the sample size prohibited using statistical techniques that appropriately modeled student data as nested within classrooms and schools.

To address these issues in the present study, we used a more robust treatment, a larger sample, random assignment of schools to conditions, and more appropriate statistical techniques, as described in the following section. We designed the study to answer these research questions:

- (a) Does TFGA increase target students' knowledge of social problem-solving strategies?
- (b) Does TFGA improve ratings of student aggression and anger expression/control?
- (c) Does degree of treatment exposure affect treatment efficacy?

We also examined TFGA's social validity for regular elementary school teachers, a factor critical to the feasibility and sustainability of universal and selective prevention.

Method

Intervention

We designed TFGA to help 4th- and 5th-grade students develop positive solutions to social problems, particularly in anger-provoking situations. We selected

grades 4 and 5 because students at this level are cognitively ready to accommodate more sophisticated problem-solving concepts and are approaching the transition to middle school, which requires increasing independence and the ability to resist negative peer influence. Moreover, Lochman, Dunn, and Klimes-Dougan (1993) suggested that teachers introduce CBI skills within targeted domains. The focus of the problem-solving framework in TFGA is, therefore, understanding and dealing with frustration and anger, frequent correlates of disruptive and aggressive behavior (Averill, 1982).

Content

TFGA presents six social problem-solving steps that guide personal decision-making. The ultimate goal is that students will use the steps automatically in challenging social situations at school and elsewhere. Following an introductory lesson, three lessons are devoted to Step 1: Recognize that a problem exists. This includes recognizing anger in oneself and others and understanding how emotional reactions can create problems or exacerbate existing ones. The next two lessons detail Step 2: Calm down and think. These lessons focus on emotion regulation and provide specific strategies to use in anger-provoking situations. The remaining lessons cover Step 3: Define the problem in terms of goals and barriers, Step 4: Generate solutions through brainstorming, Step 5: Select a strategy by considering possible outcomes, and Step 6: Evaluate the outcome after a solution is carried out. For a more complete description of TFGA and its development, see Daunic and Smith (2003).

Each of the 15 lessons devoted to a problem-solving step begins with a cumulative review and ends with an opportunity to practice learned skills through an associated activity. In addition, we added five role-play lessons placed strategically throughout the curriculum to allow students to practice each step after they have learned the relevant skills. Assuming that repeated behavioral practice enhances learning (Bandura, 1986), we also developed six booster lessons to be taught at 1-2 week intervals during the second half of the school year. The first consists of a general review of the problem-solving steps and rationale taught in lessons 1-20. In the second, students act out scripted role-plays demonstrating the six problem-solving steps. In booster lessons 3-4, teachers divide students into small groups in which they design their own role-plays and act them out for the class. In lesson 5, the class as a whole uses problem solving with real life problems shared by volunteers, who then report back to the class during the final lesson about how the chosen strategy worked.

Teacher Training

We conducted a daylong training for all treatment group teachers to present the curriculum's conceptual foundation, rationale, organization, essential approaches and procedures, and an overview of each lesson's content. We encouraged teachers to use self-instructional training as they taught students to manage their own behavior through manipulating their overt and eventually covert speech, answered questions about TFGA implementation, and discussed data collection protocols. We met separately with teachers in the control condition to discuss data collection.

Lesson Format and Features

We designed TFGA lessons to last 30 min each and be taught at a rate of two per week. Each lesson includes a specific step with objectives, a cumulative review, teacher presentation of new material, and opportunities for guided and independent practice. Lessons are scripted to promote treatment fidelity, consistency across sites, and to provide teachers with ideas for student-teacher dialogue. We recommended that teachers frequently pair or group students, and we included an optional point system at the end of periodic student-guided practice activities to allow students to reward themselves for self-reflection and appropriate participation in lesson activities.

Setting and Sampling

We solicited participation through area school district personnel and school principals, explaining prior to their commitment that a particular school might fall into one of three conditions: 20 lessons, 20 lessons plus boosters, or control. The schools were in varied environments from rural to small city, and the percent of students eligible for free or reduced-price lunch at a given school ranged from 40 to 86. After obtaining a commitment from seven schools in three districts, we matched schools on (a) state assigned school grade based on student achievement, (b) SES as determined by percent of students receiving free or reduced price lunch, and (c) school size. Bloom, Bos, and Lee (1999) recommend ranking groups from highest to lowest according to baseline characteristics (e.g., SES) and, starting with the group with the highest values, randomly assigning members to treatment and control conditions. Matching thus ensures that treatment and control groups are as similar as possible on baseline characteristics. Using size as a matching variable also increases the likelihood that an approximately equal number of classrooms and students are assigned to each condition. After following this procedure to assign schools randomly to condition, we recruited 4th and 5th grade teachers at each school. The resulting sample consisted of 2 schools (10 teachers) in the 20-lesson condition, 3 schools (17 teachers) in the 20-lesson plus booster condition, and 2 schools (8 teachers) in the control condition. At any given school, from 1 to 8 teachers agreed to participate.

Participants

We asked each teacher to complete a target student nomination form to identify the most disruptive and/or aggressive students relative to others in their class. We solicited parent permission from approximately 800 students across the 7 schools; a total of 525 (66%) returned signed parental consent forms. Of the 210 nominated as target students, 165 (79%) returned parental consent forms. The number of targets in the 20-lesson, 20lesson plus booster, and control groups consisted of 42, 86, and 37, respectively. Group size differences resulted from the number of teachers who volunteered within each school and the number of target students each one nominated. Target students' gender, race, SES, and educational program by condition are shown in Table 1. We only included typical student data prior to treatment to determine whether students targeted by teachers were at greater risk for behavior problems. The target and non-target student comparisons on pretreatment measures are presented in Table 2.

Table 1
Target student gender, race, SES, and education program for three conditions

Group	Gender		Race			Lunch status Free/reduced	Program Special ^a
	Male	Female	W	B	H		
20 Lessons (<i>n</i> =38 ^b)	82%	18%	58%	32%	10%	13%	8%
20+Boosters (<i>n</i> =82)	57%	43%	37%	60%	2%	43%	17%
Control (<i>n</i> =27)	63%	37%	63%	33%	4%	26%	15%

^aThe majority of special education students received services for learning disabilities and were mainstreamed for most or all of the school day.

^bThe sample size for each condition is the number of target students with complete data across assessments, i.e., the sample used in statistical analyses.

Measures

Problem-Solving Questionnaire

The *Problem-Solving Questionnaire* consists of 14 questions developed from TFGA by project staff, for a total possible score of 24. For some questions, only one answer among several alternatives is appropriate; other items 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 information (e.g., *What are three levels of anger, from lowest to highest?*). Scale development included (a) a pilot test with pre-and post-treatment administrations to 35 students taught TFGA and (b) subsequent item revisions. Reliability estimates using post-test data yielded a total scale Cronbach's alpha of 0.72.

Pediatric Personality (PPS-1 and 2) and Anger Expression PAES-3) Scales

The PPS and PAES, derived by Jacobs, Phelps, & Rohrs (1989) from the *State-Trait Anger Expression Inventory* (see Spielberger, Sydeman, Owen, & Marsh, 1999) include

Table 2
Pretest means, standard deviations, sample sizes, and significance tests for target vs. non-target students

Measure	Target			Non-target			t	p
	M	S.D.	n	M	S.D.	n		
State anger	22.55	5.08	126	24.08	4.33	314	-3.17	0.002*
Trait anger	19.19	4.59	125	17.50	4.69	312	3.43	0.001*
Anger out	9.66	2.69	121	8.70	2.61	311	3.38	0.001*
Anger suppression	9.22	2.57	121	9.67	2.51	311	-1.69	0.092
Anger control	9.91	2.59	123	10.68	2.42	311	-2.93	0.004*
Reactive aggression	11.18	2.67	133	4.67	2.13	95	-19.63	0.000*
Proactive aggression	8.71	3.12	133	3.51	1.32	95	-15.26	0.000*
Self control (SSRS)	7.76	3.49	127	15.58	4.38	89	-14.56	0.000*
Externalizing (SSRS)	7.54	3.16	127	3.07	2.39	41	-8.29	0.000*

* Statistically significant.

five subscales: State and trait anger (PPS-1 and 2), and anger control, suppression, and expression (PAES-3). (The PPS-1 and 2 also include anxiety subscales, which we did not analyze.) Items associated with state anger ask respondents to describe how they currently feel (e.g., *I feel like yelling at somebody; I feel grouchy*) using three response options: very much so, somewhat, or not at all. Trait anger items ask respondents to describe how they *usually* feel (e.g., *I have a bad temper; I get angry quickly*) with response options of hardly ever, sometimes, or often. The last three subscale items ask respondents to select how frequently they feel or act a particular way when they are angry or very angry and include anger control (e.g., *I do something totally different until I calm down; I control my temper*), anger suppression (e.g., *I hold my anger in; I get mad inside but I don't show it*), and anger expression (e.g., *I say mean things; I do things like slam doors*). Cronbach's alphas yielded post treatment internal subscale reliabilities of 0.91 for state anger, 0.86 for trait anger, 0.74 for anger suppression, 0.77 for anger control, and 0.81 for anger out.¹

Reactive-Proactive Aggression Scale (R/P)

Developed by Dodge and Coie (1987), the reactive and proactive aggression subscales include three items each, imbedded in 19 total questions about behavior. Respondents are asked to indicate how true a statement is, from never to always, for a particular child (e.g., *When this child has been teased or threatened he/she gets angry easily and strikes back* and *This child threatens or bullies others in order to get his/her own way*). Cronbach's alphas using post treatment teacher responses yielded internal reliabilities of 0.92 for reactive aggression and 0.90 for proactive aggression.

Social Skills Rating System (SSRS)

The SSRS is a broad assessment of student social behaviors that includes five social skills subscales, three problem behaviors subscales, and an academic competence subscale. Originally standardized and norm-referenced with a national sample of over 4,000 students, the SSRS is a valid and reliable social skills assessment that can be used for intervention planning and is appropriate for elementary students. More complete technical information can be found in Gresham and Elliott (1990). We used the self-control and externalizing behavior subscales only, with Cronbach alphas of 0.91 and 0.88, in the analyses for this study.

TFGA Teacher Questionnaire

To solicit feedback about the appeal and utility of TFGA for teachers, the teacher questionnaire was developed by project staff and includes eight items that focus on ease of use (e.g., *The curriculum was easy to use; I completed each lesson in the time allotted*), six that focus on appeal/utility to students (e.g., *The curriculum concepts*

¹A substantial number of students lacked full PPS and PAES subscale data at Assessment 1 due to a printing error. We used students with complete data (n =approximately 180) to compare truncated subscales that contained items completed by all participants at Assessment 1 with the corresponding full-scale data; the correlations for individual subscales ranged from 0.92 to 0.97. For individuals with missing data, we imputed subscale scores from truncated scores using a linear regression model.

were age-appropriate for my students), and eight that ask about TFGA's effectiveness for reducing disruption and aggression (e.g., *The curriculum improved my students' behavior*). Each item in the three subscales has a likert-type response format, with responses ranging from 1 (strongly disagree) to 5 (strongly agree). We asked teachers and university faculty to review all scale items during its development, and we made further adjustments after it was administered to TFGA pilot teachers.

Design and Analyses

Since random assignment was at the school level, all classrooms within each school were in the same condition. We followed this procedure because teachers are often grouped in collaborative grade-level teams, and random assignment at the classroom level would risk contamination across treatment and control classrooms in the same school as students and teachers interacted throughout the school day. When the possibility of treatment contamination across classrooms exists, schools should be the unit of assignment (Cook, 2005).

Three assessments were conducted in all schools regardless of condition: Prior to fall treatment (Assessment 1), after the two treatment groups completed the core 20 lessons (Assessment 2), and at the end of the academic year, following the implementation of booster lessons in the booster condition (Assessment 3). Using data from target students only, we used hierarchical linear modeling (HLM) of residual change scores to determine intervention effects at Assessment 2 and booster effects only at Assessment 3, with experimental condition as a between-subjects factor and Assessment as a within-subjects factor. HLM is the recommended procedure when the experimental design violates the independence of observations assumption required by traditional analysis of variance (ANOVA) designs (Raudenbush & Bryk, 2002). In our study, F tests based on ANOVA would not provide accurate estimates of the Type I error rate (Weinfurt, 2000) because students are nested within classrooms where the intervention occurred. In addition to a residual change score model, we used a simple difference score (SDS) model, a procedure recommended when a high rank-order stability in outcomes is likely over time and when a powerful relation between pre and post treatment scores could obscure the detection of a treatment effect (Stoolmiller & Bank, 1995).

Procedures

Although only data for students with parental consent were analyzed, all students in participating classrooms completed self-report assessments as part of classroom activities, and all students in treatment classrooms received instruction in TFGA. We asked teachers to complete the R/P and SSRS on target students and an approximately equal number of randomly selected non-target students at Assessment 1 only, so that we could compare pretreatment scores between target and typical students on all measures. Teachers in both treatment conditions taught core lessons during the first half of the school year approximately 2-3 times per week for a total of 7-10 weeks. Following Assessment 2, students in the 20-lesson plus booster group received 6 additional lessons, approximately one every other week, prior to Assessment 3. Although TFGA lessons and all student assessments were given classwide, only data from target students were compiled and analyzed

to determine treatment effects.

One project staff member was assigned to each treatment school to address ongoing questions or concerns. To monitor treatment fidelity, we asked teachers to complete feedback forms containing checklists about TFGA content covered and questions about lesson duration, strengths, weaknesses, and student responses. Each form covered four consecutive lessons and was to be completed as soon after teaching the lessons as possible. Project staff also observed 15 lessons using a checklist to note whether each teacher followed the lesson plan, included all lesson concepts and activities, and provided student feedback. Following the intervention, treatment group teachers completed the *TFGA Teacher Questionnaire* and were debriefed about implementation and social validity.

Results

Target vs. Non-Target Comparisons

As shown in Table 2, we found target students across the three groups to report more Anger Out, State Anger, and Trait Anger and less Anger Control at Assessment 1 than their typical peers. Teachers also rated target students higher on proactive and reactive aggression, lower on self-control, and higher on externalizing behavior.

Group Equivalence

To test for pre-treatment group equivalence for target students on outcome measures across experimental conditions, we compared subscale means and standard deviations at Assessment 1, as shown in Table 3. Groups differed significantly on reactive aggression only. Post-hoc *t*-tests using a Bonferroni correction for Type I error indicated that only the 20-lesson and the no treatment group differed ($t = 1.79, p = 0.040$). Taken in sum, these comparisons indicated that the experimental groups were similar on almost all subscales prior to treatment, despite demographic differences.

Table 3

Pretest means, standard deviations, sample sizes, and significance tests for target students in three experimental conditions

Measure	20 lessons			20 lessons+			Control			F	p
	M	S.D.	n	M	S.D.	n	M	S.D.	n		
State anger	22.99	5.32	34	22.06	5.03	73	23.66	4.82	19	0.916	0.403
Trait anger	18.84	4.80	33	19.49	4.59	73	18.63	4.36	19	0.388	0.679
Anger out	9.00	2.61	34	9.93	2.65	68	9.85	2.90	19	1.44	0.240
Anger suppression	8.99	2.74	34	9.41	2.45	68	8.92	2.79	19	0.457	0.634
Anger control	10.02	2.63	34	9.71	2.56	70	10.44	2.66	19	0.628	0.535
React aggression	11.66	2.29	35	11.34	2.70	76	9.86	2.87	22	3.47	0.034*
Proactive aggression	9.17	2.83	35	8.55	3.23	76	8.54	3.25	22	0.50	0.605
Self control (SSRS)	7.60	2.58	35	7.51	3.58	68	8.71	4.29	24	1.09	0.339
External (SSRS)	7.54	3.17	37	7.77	3.07	66	6.92	3.47	24	0.64	0.530

* Statistically significant.

Treatment Efficacy

Subscale means, standard deviations, and sample sizes across three assessments are shown for each condition in Table 4. To determine treatment effects, we asked two primary questions: (a) After controlling for pretest scores, were outcomes related to experimental condition? (b) After controlling for pretest scores and experimental condition, was there significant between-classroom variability on outcome scores? The first question addressed whether treatment induced a change in the outcome of interest, and the second addressed whether changes in outcome variables could be attributed to classroom-level factors *other* than experimental condition.

Table 4
Means, standard deviations, and sample sizes across measures and conditions for three assessments

Measure	Assess	20 lessons			20+ boosters			No treatment control		
		M	S.D.	<i>n</i>	M	S.D.	<i>n</i>	M	S.D.	<i>n</i>
Knowledge	1	6.21	2.41	37	5.92	2.39	80	6.48	2.13	23
	2	13.07	5.14	38	13.24	5.38	70	6.81	2.76	16
	3	10.90	3.85	33	11.01	5.13	66	7.13	2.91	23
State anger	1	22.98	5.32	34	22.06	5.03	73	23.66	4.82	19
	2	23.69	4.50	36	21.67	5.62	62	25.08	3.06	23
	3	23.65	4.50	30	21.14	6.17	62	23.79	4.01	24
Trait anger	1	18.84	4.80	33	19.49	4.59	73	18.63	4.36	19
	2	18.99	4.84	36	20.06	4.57	61	18.13	5.96	23
	3	18.88	5.73	29	19.58	5.23	62	19.50	4.54	24
Anger out	1	8.90	2.61	34	9.93	2.65	68	9.85	2.90	19
	2	9.23	2.93	34	9.55	2.81	60	9.00	3.19	22
	3	9.32	3.01	28	9.33	2.87	60	10.06	2.71	24
Anger suppression	1	8.99	2.74	34	9.41	2.45	68	8.92	2.79	19
	2	9.44	2.36	34	8.84	2.78	60	8.77	2.41	22
	3	9.18	2.45	28	9.21	2.46	60	9.33	2.35	24
Anger control	1	10.02	2.63	34	9.71	2.56	70	10.44	2.66	19
	2	10.11	2.80	35	9.51	2.58	65	9.60	2.51	23
	3	9.50	2.59	29	9.07	2.35	60	9.58	2.44	24
Reactive aggression	1	11.65	2.28	35	11.34	2.70	76	9.86	2.86	22
	2	9.43	2.96	37	9.70	2.96	68	9.89	2.61	28
	3	9.00	2.94	34	9.81	2.94	67	9.52	2.13	21
Proactive aggression	1	9.17	2.83	35	8.55	3.23	76	8.54	3.24	22
	2	6.70	2.17	37	6.92	2.92	68	8.25	2.70	28
	3	7.03	3.22	34	7.73	3.41	67	8.14	1.90	21
Self control	1	7.60	2.58	35	7.51	3.58	68	8.71	4.29	24
	2	8.85	3.83	34	9.59	4.37	66	9.39	5.20	18
	3	9.78	3.77	32	10.09	4.08	66	10.14	4.11	21
External	1	7.54	3.17	37	7.77	3.07	66	6.92	3.47	24
	2	6.37	3.01	32	6.32	3.40	63	6.06	3.32	17
	3	6.17	3.33	30	6.77	3.51	63	6.62	3.46	21

To answer these questions, we employed HLM 5 to fit a series of nested linear models, as follows.

$$\text{Level 1 (individual)} : Y_{ij} = b_{0j} + b_{1j} (\text{covariate}) + r_{ij}$$

$$\text{Level 2 (classroom)} : b_{0j} = g_{00} + g_{01} (\text{intervention}) + u_{0j}$$

$$b_{1j} = g_{10}$$

In the level-1 model, Y_{ij} corresponds to the outcome variable score of the i th individual in the j th classroom, b_{0j} corresponds to the outcome mean of the j th classroom, b_{1j} corresponds to the coefficient associated with the pretest score, and r_{ij} represents the error in the level-1 model for the individual in question. In the level-2 model, the outcome variable mean for the j th classroom is expressed as a function of the total outcome mean (g_{00}), the intervention effect (g_{01}), and error (u_{0j}). The second level-2 model for b_{1j} , the within-classroom slope associated with the regression of posttest on pretest, is set as a fixed effect model because prior analyses showed that the error term variance did not differ significantly from zero.

For HLM analyses, we selected only those measures on which descriptive data indicated possible treatment effects. In total, six versions of the level-1 and level-2 models were fit, differing according to outcome variable and comparison of interest. Thus, three versions compared reactive aggression (RA), proactive aggression (PA), and Knowledge outcomes at Assessment 2, with scores at Assessment 1 as a covariate, to compare the control group to the two treatment groups combined, since they had both received the same number of lessons at Assessment 2. The other three versions compared RA, PA, and Knowledge outcomes at Assessment 3, with scores at Assessment 2 as a covariate, to compare outcomes for the 20-lesson versus the 20-lesson plus booster group.

Target students in treatment classrooms scored higher on Knowledge at Assessment 2 than did their control counterparts. Table 5 shows that intervention classrooms had an adjusted mean Knowledge score at Assessment 2 that was 6.842 units higher than that of control classrooms. In addition, the variance of level 2 errors differed significantly from zero ($X^2=120.470, df=29, p < 0.001$), indicating that after adjusting for Assessment 1 scores and treatment effect, there was significant variation in classroom mean Knowledge scores at Assessment 2. Analyses using simple difference scores replicated these findings.

Table 5
Summary of HLM analyses for reactive aggression, proactive aggression, and knowledge

Outcome variable	Covariate	Fixed effects		Random effects	
		g_{01}	p	$\text{Var}(u_{0j})$	p
Reactive aggression (Assessment 2)	Reactive aggression (Assessment 1)	-1.583	0.002	1.423	0.000
Proactive aggression (Assessment 2)	Proactive aggression (Assessment 1)	-1.647	0.002	0.873	0.000
Knowledge (Assessment 2)	Knowledge (Assessment 1)	6.842	0.000	10.852	0.000
Reactive aggression (Assessment 3)	Reactive aggression (Assessment 2)	0.556	0.365	1.403	0.000
Proactive aggression (Assessment 2)	Proactive aggression (Assessment 3)	0.182	0.745	0.441	0.137
Knowledge (Assessment 3)	Knowledge (Assessment 2)	-0.040	0.972	6.728	0.000

To investigate whether the addition of booster lessons affected Knowledge at Assessment 3, HLM models used Assessment 3 Knowledge as the outcome variable and Assessment 2 Knowledge as the covariate. The results indicated that participation in the booster lesson group was not significantly related to adjusted Assessment 3 Knowledge scores ($t=0.036$, $df=23$, $p=0.972$), suggesting that booster lessons did not add to treatment effects. Nevertheless, the adjusted mean Assessment 3 Knowledge scores displayed a significant nonzero variance ($X^2=6.728$, $df=23$, $p < 0.001$), indicating that after adjusting for Assessment 2 Knowledge scores and condition, there was still significant variation in mean Assessment 3 Knowledge scores across classrooms. This finding probably resulted from the significant Assessment 2 classroom variance.

The findings for RA and PA outcomes were similar to those observed for Knowledge (see Table 5). Experimental condition was significantly related to the adjusted classroom mean RA score ($t=3.441$, $df=26$, $p=0.002$) and PA score ($t=3.490$, $df=26$, $p=0.002$) at Assessment 2. These results indicate that after controlling for pre-treatment scores, Assessment 2 mean RA and PA scores were 1.583 and 1.647 units lower for treatment than for control classrooms. After adjusting for pre-treatment score and condition, there also was significant variation among classrooms in Assessment 2 mean scores for RA ($X^2=106.913$, $df=26$, $p < 0.001$) and PA ($X^2=66.163$, $df=26$, $p < 0.001$). As with Knowledge, we did not find a significant relation between participation in booster lessons and adjusted RA and PA scores at Assessment 3. Again, analyses using simple difference scores replicated these findings.

There was little or no variation between mean scores at Assessments 1 and 2 for SSRS externalizing and self-control subscales or for self-reports of anger disposition and control on PPS and PAES subscales and thus no indication of a treatment effect on these outcomes.

Effect Size

Estimating effect sizes can be problematic in a multi-level design (Raudenbush & Bryk, 2002). We used a "proportion of variance explained" approach, computing the proportion of between-group variance explained by the intervention effect at Assessment 2, covarying out Assessment 1, for measures on which we found significant effects. With this procedure, we found that 0.418 of the between-group variance in mean Knowledge scores, 0.169 of the between-group variance in mean RA scores, and 0.353 of the between-group variance in mean PA scores could be explained by the intervention.

Treatment Fidelity and Social Validity

Of the 27 teachers in treatment conditions, the number returning each of the four treatment fidelity feedback forms ranged from 18 to 22. For the 15 core *content* lessons, 98% of teachers who responded reported they had covered all lesson content and that lessons averaged 30 min. Eighty-eight percent of respondents reported that students completed worksheets and activities during class; 61% reported that students completed some activities at home with a family member, as suggested in 8 of the lessons. In the five role-play lessons included in the 20-lesson core, 98% of respondents indicated that students engaged in the role-plays, and 91% indicat-

ed that students participated in associated practice activities. Some teachers paired students or used small group instruction as suggested in lesson directions, and others used a whole-class delivery format. Observational data substantiated that teachers included lesson concepts, activities, and components (e.g., role-plays, discussion questions, worksheets) during the teaching period. One of the 15 observed did not include the review of previous material or use the overhead transparencies, and expressed doubt that TFGA could affect her students' behavior.

We used descriptive data from the *TFGA Teacher Questionnaire* to examine social validity. On subscales of 1 to 5 points, appeal/utility to students had a mean of 4.18 (S.D.=0.77); appropriateness and ease of use 3.99 (S.D.=1.03); and efficacy in addressing problem behavior 3.55 (S.D.=0.95). The overall scale mean was 3.86 (S.D.=0.97).

Discussion and Limitations

Efficacy

The findings from our study showed that following a cognitive-behavioral curricular intervention, target students' knowledge about problem solving increased, teacher ratings of target student reactive and proactive aggression improved, and improvements in knowledge and teacher-reported behavior were maintained over several months. Since the skill of problem solving requires procedural knowledge that is foundational to sufficient performance and subsequent reinforcement (Bandura, 1986), it is not trivial that students learned curricular concepts. We must temper our conclusions about the intervention's efficacy for changing student behavior, however, with the fact that we did not include direct observational measures or peer reports, both of which would have substantially enhanced the supporting evidence. In addition, we did not find that treatment affected teacher-reports on more general measures of self-control and externalizing behavior or student self-reports on subscales designed to assess anger disposition and expression.

When considering these findings, it is important to remember that the students in our sample were not likely to have exhibited behavior classified as chronic or severe, or have scores that differed markedly from the norm on measures of anger and aggression, as compared to students placed in special programs. Inherent in working with this population and in prevention research in general, therefore, is the challenge of ensuring enough statistical power through effect size or sample size to detect relatively small improvements, particularly over the short term (see Muehrer & Koretz, 1992). Thus, we believe that the positive teacher-report data that replicate those from our preliminary work represent a modest indication of treatment efficacy, at least for changing teacher perceptions of aggression. Others who have investigated preventive interventions similar to ours have relied solely or in part on teacher reports (see CPPRG, 1999b, 2002b). Ollendick and King (1999) even argue that changes in teacher perceptions have important consequences that make these measures as informative as direct assessments of behavior, since classroom teachers most often refer students for special services. If program implementation improves a teacher's view of student behavior, the classroom environment may become more conducive to the promotion of prosocial skills (Sutherland & Oswald, 2005).

We did not find that students' perceptions about their own behavior changed significantly. One explanation is that the measures we used may not be sensitive enough to detect modest changes over the short term. We agree with Hinshaw (2002) that when the goal is to prevent serious behavioral issues from occurring, initial findings of minimal treatment effects may belie a delayed emergence of real prevention. Longer-term studies would reveal whether this is indeed the case. It is also feasible that the self-reports of anger expression did not change significantly because target students were not yet confident enough of their own developing skills to alter reports of their "typical" behavior when provoked.

We hypothesized that increasing opportunities for practice would strengthen student learning and generalization (Bandura, 1986) but did not find that students in the booster condition outperformed students in the 20-lesson only condition at Assessment 3. Though all teachers in the booster condition reported that they taught the additional six lessons, booster instruction took place late in the school year following high-stakes testing. Teachers may, therefore, have delivered these lessons less diligently than the 20-lesson core, and we may have unintentionally conveyed the message during training that boosters were less important. Six additional lessons may also not make enough of a difference, although again, longer-term studies would help determine potential sleeper effects (Hinshaw, 2002).

Fidelity and Social Validity

As the emphasis on state and national accountability continues to escalate, teachers are increasingly preoccupied with meeting performance standards and often view nonacademic curricula as a distraction from more pressing concerns. Such attitudes could affect student responses to social skills intervention (Polsgrove & Smith, 2004). Although all participating teachers in this study volunteered, it was difficult to discern the priority they placed on teaching TFGA. The fact that classroom means on outcome measures differed after controlling for pre-treatment scores and experimental condition indicates that some teachers were more successful than others in effecting positive changes in student knowledge and behavior, regardless of intervention.

To assess treatment fidelity, we relied on teacher reports about instruction time and content completion because of widespread school locations and limited resources. Although a majority of teachers indicated they had covered almost all TFGA lesson content and activities, more frequent classroom observations would have (a) strengthened confidence about treatment fidelity across sites and (b) added qualitative information about how teacher characteristics related to student outcomes across classrooms.

It is noteworthy that teacher responses on the social validity survey were generally positive about TFGA ease of use and value to students. During informal follow-up discussions, most teachers expressed a desire to teach TFGA in the future and said they would recommend its use to other teachers and school personnel.

Implications and Future Research

Notably, the vast majority of teachers with whom we have worked expressed the opinion that students know what to do to control their behavior but often do

not act on what they know because of competing behavioral models at school, at home, or in their communities. Although peer, family, and neighborhood influences, including reinforcement contingencies, may compete strongly with a skill-based, cognitive-behavioral intervention, we believe teachers can increase student resilience in the face of behavioral risk factors through what they teach and model in schools. Our findings, though limited, support the continuing study of such efforts.

Researchers have noted that prevention programs are most powerful when they include family, peer, and community components (see CPPRG, 1999a,b). Although multicomponent programs requiring support staff and specialized materials can be difficult for schools to sustain after external leadership and funding have ended (Sindelar & Brownell, 2001), these approaches are clearly warranted when adequate resources exist. We also recognize that the skills involved in social problem solving are complex and require repeated cognitive and behavioral practice. Nevertheless, we recommend the continued study of factors that contribute to the efficacy of feasible, classroom-based CBIs. These factors include optimal treatment exposure, treatment fidelity, sustainable parent involvement, and efforts to ensure generalization.

In light of the significant teacher-level effect we found across conditions, future studies should also focus on teacher characteristics that affect treatment outcomes. These characteristics might include classroom discipline style and/or attitudes toward students with challenging behavior, issues that could be addressed in teacher training. In addition, longitudinal measures to examine prevention effects over time (see, e.g., CPPRG, 1999b), such as referrals for special education, rates of behavioral incidents and delinquency, and measures of academic performance, would add significant information to findings based on shorter-term outcomes.

By continuing these efforts, researchers can further define the parameters of feasible, sustainable, classwide preventive strategies that can be used in tandem with other approaches to reduce risk and enhance resilience. We concur with Muehrer and Koretz (1992) and CPPRG (2002a) that prevention research is challenging, but it offers the potential of avoiding negative outcomes and the costs associated with specialized treatment. Quality of life for many students and educators may rest in early and effective efforts to prevent anger and aggression. The most promising strategies can be found only through continued sharing of information and effective collaboration between the research community and school practitioners.

References

- Abikoff, H. (1991). Cognitive training in ADHD children: less to it than meets the eye. *Journal of Learning Disabilities* 24: 205-209.
- Ager, C.L., and C.L. Cole (1991). A review of cognitive-behavioral interventions for children and adolescents with behavior disorders. *Behavioral Disorders* 16: 276-277.
- Averill, J.R. (1982). *Anger and Aggression: An Essay on Emotion*. New York: Springer-Verlag.
- Bandura, A. (1969). *Principles of Behavior Modification*. New York: Holt, Rinehart, and Winston.
- Bierman, K.L., M.T. Greenberg and the Conduct Problems Prevention Research Group (CPPRG). (1996). Social skills training in the fast track program. In: R.D. Peters and R.J. McMahon (editors), *Preventing Childhood Disorders, Substance Abuse, and Delinquency* (pp. 65-89). Thousand Oaks, Calif.: Sage.

- Bloom, H.S., J.M. Bos and S.-W. Lee (1999). Using cluster random assignment to measure program impacts: statistical implications for the evaluation of educational programs. *Evaluation Review* 23: 445-469.
- Conduct Problems Prevention Research Group (CPPRG) (1999a). Initial impact of the Fast Track prevention trial for conduct problems: I. The high-risk sample. *Journal of Consulting and Clinical Psychology* 67(5): 631-647.
- Conduct Problems Prevention Research Group (CPPRG) (1999b). Initial impact of the Fast Track prevention trial for conduct problems: II. Classroom effects. *Journal of Consulting and Clinical Psychology* 67(5): 648-657.
- Conduct Problems Prevention Research Group (2002a). The implementation of the Fast Track program: An example of a large-scale prevention science efficacy trial. *Journal of Abnormal Child Psychology* 30(1): 1-17.
- Conduct Problems Prevention Research Group (2002b). Evaluation of the first 3 years of the Fast Track prevention trial with children at high risk for adolescent conduct problems. *Journal of Abnormal Child Psychology* 30(1): 19-35.
- Cook, T.D. (2005). Emergent principles for the design, implementation, and analysis of cluster-based experiments in social science. *Annals of the American Academy of Political and Social Science* 599(1): 176-198.
- Cook, T.D., and D.T. Campbell (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston: Houghton Mifflin.
- Cullinan, D. (2002). Students with emotional and behavior disorders: an introduction for teachers and other helping professionals. Upper Saddle River, N.J.: Pearson.
- Daunic, A.P., and S.W. Smith (2003). Technical Report #9: *Tools for Getting along: Teaching Students to Problem Solve*. University of Florida, Department of Special Education.
- Daunic, A.P., S.W. Smith and M.D. Miller. *Teaching Students Problem-Solving to Counter Aggression: Findings from Two Years of Research*. Unpublished manuscript.
- Dodge, K. A. (1986). A social information-processing model of social competence in children. In: M. Perlmutter (editor), *The Minnesota Symposium on Child Psychology* (pp. 77-126). Hillsdale, N.J.: Lawrence Erlbaum.
- Dodge, K.A., and J.D. Coie (1987). Social-information-processing factors in reactive and proactive aggression in children's peer groups. *Journal of Personality and Social Psychology* 53(6): 1,146-1,158.
- Dush, D.M., M.L. Hirt and H.E. Schroeder (1989). Self-statement modification in the treatment of child behavior disorders: a meta-analysis. *Psychological Bulletin* 106: 97-106.
- Gresham, F.M., and S.N. Elliott (1990). *The Social Skills Rating System*. Circle Pines, Minn.: American Guidance Service.
- Heinicke, C.M. (1988). Psychodynamic psychotherapy with children: current status and guidelines for future research. In: B.B. Lahey and A.E. Kazdin (editors), *Advances in Clinical Child Psychology*, vol. 12. New York: Plenum.
- Hinshaw, S.P. (2002). Prevention/intervention trials and developmental theory: commentary on the Fast Track special section. *Journal of Abnormal Child Psychology* 30(1): 53-59.
- Jacobs, G.A., M. Phelps and B. Rohrs (1989). Assessment of anger expression in children: the pediatric anger expression scale. *Personality and Individual Differences* 10(1): 59-65.
- Kendall, P.C. (1993). Cognitive-behavioral therapies with youth: guiding theory, current status, and emerging developments. *Journal of Consulting and Clinical Psychology* 61: 235-247.
- Kendall, P.C., and L. Braswell (1985). *Cognitive-behavioral therapy for impulsive children*. New York: Guilford.
- Larson, J., and J.E. Lochman (2002). *Helping school-children cope with anger: a cognitive-behavioral intervention*. New York: Guilford.
- Lochman, J.E. (1985). Cognitive-behavioral intervention with aggressive boys: three-year follow-up and preventive effects. *Journal of Consulting and Clinical Psychology* 60: 426-432.
- Lochman, J.E., J.D. Coie, M.K. Underwood and R. Terry (1993). Effectiveness of a social relations intervention program for aggressive and nonaggressive, rejected children. *Journal of Consulting and Clinical Psychology* 61(6): 1,053-1,058.

- Lochman, J.E., S.E. Dunn and B. Klimes-Dougan (1993). An intervention and consultation model from a social cognitive perspective: a description of the Anger Coping Program. *School Psychology Review* 22(3): 458-471.
- Lochman, J., J. Whidby, J. and D. FitzGerald (2000). Cognitive-behavioral assessment and treatment with aggressive children. In: P. C. Kendall (editor), *Child and Adolescent Therapy: Cognitive-Behavioral Procedures* (2nd ed.). New York: Guilford.
- Muehrer, P., and D. Koretz (1992). Issues in preventive intervention research. *Current Directions in Psychological Science* 1(3): 109-112.
- No Child Left Behind Act of 2001, Pub. L. No 107-110, 115 Stat. 1425(2002).
- Ollendick, T.H., and N. King (1999). Child behavioral assessment and cognitive-behavioral interventions in schools. *Psychology in the Schools* 36: 427-436.
- Polsgrove, L., and S.W. Smith (2004). Informed practice in teaching self-control to children with emotional and behavioral disorders. In: R.B. Rutherford, M. Quinn and S.R. Mathur (editors), *Handbook of Research in Emotional and Behavioral Disorders*. New York: Guilford.
- Raudenbush, S.W., and A.S. Bryk (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd ed.). Thousand Oaks, Calif.: Sage.
- Robinson, T.R., S.W. Smith and M.D. Miller (2002). Effect of a cognitive behavioral intervention on responses to anger by middle school students with chronic behavior problems. *Behavioral Disorders* 27: 256-271.
- Robinson, T.R., S.W. Smith, M.D. Miller and M.T. Brownell (1999). Cognitive behavior modification of hyperactivity/impulsivity and aggression: a meta-analysis of school-based studies. *Journal of Educational Psychology* 91: 195-203.
- Sindelar, P.T., and M.T. Brownell (2001). Research to practice dissemination, scale, and context: we can do it, but can we afford it? *Teacher Education and Special Education* 24: 348-355.
- Smith, S.W., J.E. Lochman and A.P. Daunic (2005). Managing aggression using cognitive-behavioral interventions: state of the practice and future directions. *Behavioral Disorders* 30: 227-240.
- Smith, S.W., E.M. Siegel, A.M. O'Connor and S.B. Thomas (1994). Effects of cognitive-behavioral training on aggressive acts and anger behavior of three elementary-aged students. *Behavioral Disorders* 19: 126-135.
- Spielberger, C.D., S.J. Sydeman, A.E. Owen and B.J. Marsh (1999). Measuring anxiety and anger with the State-Trait Anxiety Inventory (STAI) and the State-Trait Anger Expression Inventory (STAXI). In: M. E. Maruish (editor), *The Use of Psychological Testing for Treatment Planning and Outcome Assessment* (2nd ed.). Mahwah, N.J.: Lawrence Erlbaum.
- Stoolmiller, M., and L. Bank (1995). Autoregressive effects in structural equation models: we see some problems. In: J.M. Gottman (editor), *The Analysis of Change*. Mahwah, N.J.: Lawrence Erlbaum.
- Sutherland, K.S., and D.P. Oswald (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): 1-14.
- Vaughn, S., and J.E. Dammann (2001). Science and sanity in special education. *Behavioral Disorders* 27: 21-29.
- Walker, H.M., G. Colvin and E. Ramsey (1995). *Antisocial Behavior in School: Strategies and Best Practices*. Pacific Grove, Calif.: Brookes/Cole.
- Waltman, D.E., and D.G. Zimpfer (1988). Composition, structure, and duration of treatment: interacting variables in counseling groups. *Small Group Behavior* 19: 171-184.
- Whalen, C.K., B. Henker and S.P. Hinshaw (1985). Cognitive-behavioral therapies for hyperactive children: premises, problems, and prospects. *Journal of Abnormal Child Psychology* 13: 391-410.
- Weinfurt, K.P. (2000). Repeated measures analyses: ANOVA, MANOVA, and HLM. In: L.G. Grimm and P.R. Yarnold (editors), *Reading and Understanding More Multivariate Statistics* (pp. 317-362). Washington, D.C.: American Psychological Association.