Running Head: IMPACT OF DRUG PREVENTION

# Technical Report:
# One Year Study of the Effects of the *Too Good for Drugs* Prevention Program on Middle School Students

March 2013

Tina P. Bacon
Saint Petersburg, Florida
Bruce W. Hall
University of South Florida
John M. Ferron
University of South Florida

Key Words: Drug Prevention, Risk Level, Risk and Protective Factors, Social and Emotional Learning, Middle School

## Author Note

This research was supported by the C. E. Mendez Foundation, Inc., with the collaboration of professionals from one of the state of Florida's and nation's largest school districts. The authors express their appreciation to the teachers, site administrators, and district administrators whose cooperation made this study possible. Funding to support instructional services and student materials during the 2009-2010 school year was provided by a grant from the Florida Department of Children and Families and a local city's Community Development Block grant.

TINA P. BACON, is a research and evaluation consultant, 5650 Leeland Street South, Saint Petersburg, FL 33715; e-mail tbacon@tampabay.rr.com.

BRUCE W. HALL is the former chair of the Department of Educational Measurement and Research, and Professor Emeritus, College of Education, University of South Florida, 4202 E. Fowler Avenue, EDU 357, Tampa, FL 33620; e-mail bwhall@tampabay.rr.com.

JOHN M. FERRON, is a Professor for the Department of Educational Measurement and Research, College of Education, University of South Florida, 4202 E. Fowler Avenue, EDU 369, Tampa, FL 33620; ferron@tempest.coedu.usf.edu.

## Abstract

*Too Good for Drugs* (TGFD) is a school-based prevention program designed 1) to prevent or diminish cigarette smoking, alcohol consumption and marijuana use among middle school children, and 2) to reduce risk factors and enhance protective factors that strengthen resiliency in middle school adolescents related to alcohol, tobacco and other drug (ATOD) use. To examine the program's effectiveness and to determine if the program's impact is moderated by student risk level, a randomized trial study was implemented with an ethnically mixed sample of 6th graders from a large school district containing urban, suburban and rural areas.

***Study Design and Implementation***

The evaluation study used a stratified randomized treatment-control group design, whereby 40 middle schools were paired on the basis of key school demographics, then randomly assigned to either treatment or control conditions. The students in the 20 treatment schools and 20 control schools were surveyed in their classrooms on an identical schedule: prior to the delivery of treatment, after the delivery of treatment, and six months following treatment. At each assessment point, all participating students completed the *Student Behavior and Risk and Protective Factor Survey* (SBRPFS) which contains items modeled after those of established ATOD measures. The SBRPFS dealt with student ATOD substance use (cigarette, alcohol and marijuana use) within the past 30 days and past year, and student status with respect to key risk and protective (R&P) factors associated with resiliency to ATOD use (intent to use ATOD, goal setting and decision making skills, bonding with prosocial peers, social and peer resistance skills, emotional competency/self-efficacy, awareness of harmful effects of drugs, and attitudes toward drug use). Three levels of student risk were identified based on those reporting that at age 10 or younger they had tried 1) none of the three ATOD substances of cigarettes, alcohol and

marijuana (low risk), 2) one of the three substances (moderate risk), and 3) two or more of the three substances (high risk).

In the 20 treatment schools, the 10-lesson TGFD program was delivered by TGFD instructors trained by staff from the C. E. Mendez Foundation. During the delivery, the fidelity of program implementation was monitored through unannounced in-class observations. Implementation data were also collected from the TGFD instructors themselves, as well as from the 6th grade teachers in whose classrooms the instruction was delivered.

To maximize the quality of the survey responses across the 20 treatment and 20 control schools, the evaluation team trained all survey administrators. Survey administrators followed a prepared script, and were observed in all treatment and control schools by the team during the administration of the survey.

***Results on Major Outcomes***

Across the three survey periods (pre-survey, post-survey, 6-month follow-up) the treatment and control groups remained highly similar to one another with respect to composition by gender, ethnicity, free or reduced lunch program, limited English proficiency services, and exceptional education services. The pre-survey sample included 10,762 students, the post-survey 10,513 students, and the 6-month survey 10,163 students.

Student survey responses were analyzed using a multilevel model estimated by restricted maximum likelihood using the MIXED PROCEDURE in SAS. Results were reported in terms of probability values and standardized effect size (ES). Post-survey results show the TGFD treatment, in comparison to the control, to be effective in diminishing reported 30-day smoking use, alcohol consumption, binge drinking and marijuana use among high risk 6th graders, and in impacting all seven R&P factors to boost these high risk students' resiliency related to drug use. The post-survey ESs for the four 30-day usage outcomes (.56 to 1.03) and the seven R&P factors

(.33 to .76) evidence a short-term impact of the treatment for high risk students that was broad and substantive. The positive effects, though attenuated by time, were still present six months after treatment for the high risk students on all of the 30-day usage outcomes (ESs of .30 to .65) and on five of the seven R&P outcomes (intent to use ATOD, peer resistance, bonding with prosocial peers, harmful effects of drugs, and ATOD attitudes) with ESs of .30 to .63. Also, students' reported use of cigarettes, alcohol and marijuana over the past year showed a diminution favoring the treatment high risk students (ESs of .26 to .57).

The results show that the TGFD treatment had some impact on the low and moderate risk students, but the effects, in comparison to those for high risk students, were more limited in both scope and time. For moderate risk students, the post-survey produced significant treatment results (ESs of .14 to .19) on three of the 30-day ATOD usage outcomes (drinking, binge drinking, and marijuana use), as well as on two of the R&P factors: peer resistance (ES = .31) and self-efficacy (ES = .19). However, these effects did not carry over to six months later. Nor were there any significant effects on moderate risk students' reported ATOD use over the past year. For low risk students, the post-survey produced no significant treatment effects on the ATOD usage outcomes, but did produce significant effects for three of the R&P outcomes: goals and decisions (ES = .20), peer resistance (ES = .23), and self-efficacy (ES = .23). The effects on peer resistance (ES = .17) and self-efficacy (ES = .15) carried over to six months later.

All of the effects that were produced on the post-survey and 6-month follow-up survey favor the treatment students, such that the results show a general *suppression* effect on students' reported ATOD usage, and a general *strengthening* effect on those R&P factors that are considered important in promoting adolescents' resilience to inappropriate drug use. These findings underscore the efficaciousness of the TGFD treatment as a 6$^{th}$ grade intervention, especially for students identified as being at high risk for early experimentation with drugs. The

evidence also suggests that the treatment effects, though attenuated, extend across time for these high risk students to a point six months after treatment.

Finally, an examination of the TGFD treatment in relation to school achievement found an effect between the treatment/control conditions at the lower levels of prior achievement and subsequent student performance on the Florida Comprehensive Assessment Test (FCAT) Mathematics. Specifically, treatment 6th graders with low and below average 5th grade FCAT Mathematics scores performed better on 6th grade FCAT Mathematics as compared to their counterparts among control students.

***Fidelity Results***

The findings of the study are given substance by the fact that SBRPFS was found to have acceptable estimates of internal consistency reliability and test-retest reliability, and all of its subcomponents showed evidence of concurrent validity in relation to popular drug usage and R&P instruments. Also, an investigation of internal factor structure on the R&P items showed that the items of six of the seven R&P subscales loaded in a manner consistent with a meaningful interpretation of those subscales. The unannounced on-site observations gathered on the survey process found that the survey instrument was appropriately administered throughout both treatment and control classrooms. Data on fidelity of implementation of the TGFD lessons gathered from on-site observers, from the TGFD teachers themselves, and from the regular classroom teachers provide collaborative evidence that the TGFD treatment was delivered with consistent quality and completeness, and in such a way that actively and successfully engaged the participating students.

## Table of Contents

# One Year Study of the Effects of the *Too Good for Drugs* Prevention Program on Middle School Students

## Introduction

This report is a summary of a randomized trial study that examined the impact of the *Too Good for Drugs* (TGFD) prevention program on 6th grade students' behaviors, skills and attitudes. Forty middle schools serving students in grades 6-8 were stratified based on matriculating 6th grade student demographics including ethnicity, federal free or reduced lunch program services, limited English proficiency services, exceptional student education services, and prior reading achievement. School size and population density were also considered in the stratification process. One school from each matched pair of schools was randomly assigned to either the treatment or control condition. Twenty schools serving approximately 5,400 sixth graders were assigned to the treatment group, and 20 schools serving approximately the same number of 6th graders were assigned to the control group. Student survey data were collected prior to the delivery of the treatment intervention, shortly after the delivery of the treatment, and 6-months following the delivery of the treatment intervention.

## Prevalence of Drug Use

According to the University of Michigan's Institute for Social Research, in their Monitoring the Future (MTF) study, national trends in drug use among adolescents (8th, 10th and 12th grade) in recent years have become more complex, and accordingly more difficult to describe (Johnston, O'Malley, Bachman, & Schulenberg, 2011). However, they note the following regarding alcohol, cigarette and marijuana use.

Alcohol use, including binge drinking, has shown a long-term pattern of decline since 1980, with the exception of the early 1990s. Despite declining rates in 2010, 71% of the students report having consumed alcohol at some point in their life by the end of high school, and 36%

report having done so by 8th grade. Over half of 12 graders (54%) and 16% of 8th graders report having been drunk at least once in their life (Johnston et al., 2011).

The long-term decline in cigarette use, which began in the 1990s, came to a halt in 2010, and even showed evidence of a slight increase among 8th and 10th graders in 2010. In 12th grade 42% of teens report having tried cigarettes in their lifetime and 19% report being a current smoker. In 8th grade 20% report having tried cigarettes and 7% report being a current smoker (Johnston et al., 2011).

Marijuana use has been rising for the past two years, and continued to rise in 2010 across all three grade groups. In 12th grade 44% of teens report having smoked marijuana in their lifetime and in 8th grade just over 17% report the same. The Institute notes that in 2010 nearly one in 16 high school seniors was a current daily--or near-daily--user of marijuana (Johnston et al., 2011).

The 2010 MTF findings underscore the need for continued efforts to fight drug use among adolescents through educational and related interventions.

## Description of TGFD Program

The *Too Good for Drugs* (TGFD) is a school-based drug intervention program designed to diminish risk factors and enhance protective factors related to alcohol, tobacco and other drug (ATOD) use, and promote prosocial attitudes, skills and behaviors among children and youth, leading to diminished use of ATOD substances. Risk factors are variables that may lead to an increased probability of ATOD use or other antisocial behaviors among adolescents, and protective factors are those variables that may be associated with resilience to such behaviors, i.e., that moderate the effect of exposure to risk factors (Catalano et al., 2003; Hawkins, Catalano, Kosterman, Abbott, & Hill, 1999; Hawkins, Catalano, & Miller, 1992). A study by

Pollard, Hawkins, and Arthur (1999) suggests that to be maximally effective, preventive interventions—including drug interventions—should concentrate on both the reduction of risk and the strengthening of protective influences. The following have been found by one or more authors to be empirically-based factors that may increase the risk or are associated with increased risk for drug abuse: academic failure, lack of commitment to school, high perceived availability of drugs, poor family discipline, high family conflict, poor impulse control, favorable attitudes toward drugs, peer drug use, peer antisocial behavior, early aggressive behavior, and early initiation of substance use (Hawkins et al., 1992; Maguin & Loeber, 1996; Pollard et al., 1999; Rutter, 1990; Yoshikawa, 1994). Among the factors that appear to increase resiliency to drug use are normative beliefs about drugs (realistic perceptions of the prevalence and expectation of drug use), unfavorable attitudes toward drugs, perceptions of the consequences of drug use, decision making skills, bonding with prosocial others, self-efficacy, and peer resistance information and skills (Arthur, Hawkins, Pollard, Catalano, & Baglioni, Jr., 2002; Donaldson, Graham, Piccinin, & Hansen, 1995; Guo, Hawkins, Hill, & Abbott, 2001; Hansen, 1992; Hawkins et al., 1999; Hawkins et al., 1992; Komro et al., 2001). In a review of school-based drug interventions, Hansen (1992) found that among the most potent factors in ATOD substance use were normative beliefs; personal commitments to avoid use; and perceptions that drug use would compromise personal values and lifestyles, and interfere with prosocial bonding.

At each grade level, K through 12, TGFD uses a developmentally appropriate curriculum based on five interwoven components: Goal Setting, Decision Making, Bonding with Prosocial Others, Identifying and Managing Emotions, and Communicating Effectively. The curriculum also focuses on peer pressure, normative behaviors, the negative consequences of drug use, and the benefits of a drug-free life style. Throughout the lessons, attention is given to such social and

emotional learning concepts as self-awareness, regulating and expressing emotions appropriately, empathizing with others, developing and maintaining healthy relationships, managing and resolving interpersonal problems, recognizing the consequences of one's actions, and making healthy choices (Payton et al., 2008). In addition, curricular content is reflective of recent research on the neurocognitive functioning of children that shows the adolescent brain to be a "work in progress", growing and changing throughout early and late adolescence (Giedd et al., 1999; National Institute of Mental Health, 2001; Restak, 2001), and research indicating potential damage to adolescent brain development from alcohol usage (Brown, Tapert, Granholm, & Delis, 2000). For any given grade level, the curriculum builds on the curriculum of earlier grade levels, enabling students to acquire important skills sequentially and retain and strengthen these skills year after year. Training is made available to service providers, whether classroom teachers or other professionals; interactive workbooks are provided to each participating student; home workout activities are offered to enhance parent involvement; and assessment tools are made available to allow users to assess outcomes and check implementation.

The TGFD curriculum is consistent with a social influences-based prevention approach. Donaldson et al. (1996) reports that in the 1980's it became clear that social influences are a major reason why adolescents begin using ATOD substances. The social influences approach attempts to "inoculate" students against direct social pressure (e.g., explicit offers from others) or indirect social pressure (e.g., overestimation of peer use and social modeling) to use alcohol, tobacco and other drug substances (Donaldson et al., 1996). Social influence-based interventions tend to address elements such as the health and social effects of drug use; information about social norms, expectations and media influences; and personal and social skills development

such as self-management, decision making, and peer resistance skills. Instructional techniques include modeling, role-playing, cooperative learning, group interaction and practice (Donaldson, 1995; Donaldson et al., 1996; Hwang, Yeagley, & Petosa, 2004). In a review of what works in school-based drug prevention programs, Cuijpers (2002) identified the social influences approach as a generally effective model for school-based drug prevention. Preventive programs based on one or more elements of the social influences model have been found effective in numerous ATOD reviews and meta-analyses (Donaldson et al., 1996; Flay, 2009; Hansen, 1992; Skara & Sussman, 2003; Tobler & Stratton, 1997) and in individual studies (e.g., Botvin, Griffin, Diaz, & Ifill-Williams, 2001b; Eisen, Zellman, & Murray, 2003; Ellickson, McCaffrey, Ghosh-Dastidar, & Longshore, 2003; Flay, Phil, DuBois, & Ji, 2007; Hecht et al., 2003; Longshore, Ellickson, McCaffrey, & St. Clair, 2007; Perry et al., 2002; Schwinn, Schinke, & Di Noia, 2010; Sussman, Dent, Stacy, & Craig, 1998).

The TGFD curriculum for 6th graders used in this study included ten 50-minute lesson units delivered by TGFD instructors trained by staff from the C. E. Mendez Foundation, Inc. Funding for TGFD instructors to provide prevention services in the school setting was supported in large part by a grant from the Florida Department of Children and Families and a local city's Community Development Block grant. The prevention lessons were delivered during the first three months of the school year, in either 6th grade science or social studies classrooms, one lesson per week. TGFD teaching methods are highly interactive (Tobler, 2000; Tobler & Stratton, 1997), providing students numerous opportunities to be active participants and to receive recognition and reinforcement for their involvement. Instructional strategies include cooperative learning, role playing, small group activities, purposeful games, demonstrations, class discussions and iterative practice; and the instructors model and encourage bonding with

prosocial others. Each lesson provides student workbook activities of both an individual and group nature that enables student participatory learning and the practice of skills. TGFD also includes a family component whereby students are encouraged to share "Home Workouts" with family members to reinforce concepts practiced during the lesson units. The lessons at grade 6 cover 1) goal setting; 2) decision making; 3) identifying and managing emotions; 4) effective communication; 5) bonding and relationships; 6) the effects and consequences of alcohol use; 7) the effects and consequences of tobacco use; 8) the effects and consequences of marijuana use; 9) inhalants and street drugs; and 10) course review.

## Theoretical Base

The TGFD curriculum is a multifaceted prevention program based upon theoretical constructs in social learning theory (Bandura, 1977, 1985), problem behavior theory (Jessor, 1982; Jessor & Jessor, 1977), and the social development model (Catalano & Hawkins, 1996; Hawkins et al., 1999; Hawkins & Weis, 1985)). In addition, strategies are used based on the developmental assets framework (Leffert, Benson, & Roehlkepartain, 1996; Scales & Leffert, 2004), and tenets of social and emotional learning (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011; Greenberg et al., 2003; Payton et al., 2008).

According to social learning theory (Bandura, 1977, 1985), ATOD use can be viewed as a socially learned, purposeful behavior, which is shaped primarily through modeling (observing behaviors) and reinforcement (experiencing positive consequences for behavior). Modeling contributes to the acquisition of both prosocial and antisocial behaviors. Based on a self-efficacy paradigm, social learning theory posits that behavior change and maintenance depend upon: a) expectations about the outcomes of engaging in the behavior, and b) a sense of self-efficacy, or expectations about one's ability to engage in the behavior. From this perspective, alcohol and

other drug use results from the interplay of personal perceptions and social-environmental influences. TGFD reflects social learning theory by addressing the social influences of peers and media advertising; correcting misperceptions about social norms; modeling prosocial skills and promoting the value of prosocial behaviors; emphasizing the development and application of social and personal competencies to resist social and environmental pressures to use drugs; and providing opportunities to perform skills and recognizing and rewarding their use.

From the perspective of problem behavior theory (Jessor, 1982; Jessor & Jessor, 1977), ATOD use and other correlated behaviors form a syndrome of purposive behaviors that are psychologically functional for many adolescents. Problem behavior theory posits that efforts to change behavior may focus on any or all of the following elements: behavior, personality, and environment. An extension of this theory, health behavior theory (Perry and Jessor, 1983), proposes that strategies be used to introduce or strengthen health-enhancing behaviors and simultaneously weaken or eliminate health-compromising behaviors. This approach suggests that prevention efforts should attend to the larger environment, including social norms and social supports regulating the occurrence of the behaviors; and that interventions should focus on multiple behavioral targets. A number of the assumptions underlying TGFD are based on problem behavior theory and health behavior theory. For example, TGFD provides normative education, teacher tips and a parent component for making the school and family environments more supportive of drug-free choices. TGFD also incorporates role-play and decision-making scenarios dealing with both ATOD use and other highly correlated problem behaviors.

The social development model (Catalano & Hawkins, 1996; Hawkins et al., 1999; Hawkins & Weis, 1985) is an integration of social control theory and social learning theory. The social development model emphasizes the importance of protective factors: a) bonding with

prosocial family, school, peers, and community; and b) clear standards or norms of behavior. According to this model, positive socialization is achieved when youths have the opportunity to be involved in conforming activities, when they develop skills necessary for successful involvement, and when those with whom they interact consistently reward desired behaviors. Such conditions increase attachment to others, commitment to conforming behavior, and belief in the conventional order. The TGFD curriculum is reflective of the social development model in that it focuses on the strengthening of protective factors, including bonding and norms. It teaches prosocial skills, provides opportunities for participation, and rewards/recognizes participation. It emphasizes prosocial norms, providing activities and information to counter students' misperceptions regarding levels of drug use, and strongly supporting healthy normative beliefs and unambiguous standards. It sends a clear message for students: You are "Too Good for Drugs".

The developmental assets framework (Leffert et al., 1996; Scales & Leffert, 2004) suggests that positive, healthy development in youth depends on the presence of a number of developmental assets—building blocks that all children and adolescents need in becoming healthy, competent and caring adults. These assets are both internal (educational commitment, values, social competencies and positive identity) and external (support, empowerment, boundaries and expectations, time). The effect of these assets is seen as cumulative. The more assets young people have, the more resilient they will be, and the more engaged in positive behaviors. The fewer assets they have, the less resilient they will be, and the more likely they will become involved with drugs, violence and other anti-social behaviors. The TGFD curriculum is supportive of the developmental assets framework in that it includes a proactive, positive focus and a commitment to the long-term building of both internal and external assets

for all students. The goal of TGFD is to prevent problem behavior, and to promote positive, healthy development in youths.

Social and emotional learning (SEL), as conceptualized by the Collaborative for Academic, Social and Emotional Learning (CASEL), comprises five inter-related core social and emotional competencies: 1) Self-awareness—accurately assessing one's feelings, interests, values and strengths, and maintaining self-confidence; 2) Self-management—regulating one's emotions to handle stress and control impulses, expressing emotions appropriately, and setting and monitoring progress toward personal and academic goals; 3) Social awareness—taking the perspective of and empathizing with others, recognizing and appreciating individual and group similarities and differences, and recognizing and using family, school and community resources; 4) Relationship skills—developing and maintaining healthy relationships based on cooperation, resisting inappropriate social pressure, preventing or managing and resolving interpersonal conflict, and seeking assistance as needed; and 5) Responsible decision-making—applying decisions-making skills to make decisions based on ethical standards, safety concerns, appropriate social norms, respect for others, and likely consequences (Payton et al., 2008). These core competencies are seen as providing a foundation for better adjustment and academic performance as reflected in more positive social behaviors and less emotional distress and misconduct (Durlak et al., 2011).

The acquisition of social and emotional skills is considered similar to the acquisition of academic skills in that the effect of initial learning is enhanced over time to address increasingly complex situations that children face. SEL skills, therefore, must be developmentally appropriate, and require effective classroom instruction as well as positive activities in and out of the classroom. The SEL approach to school-based prevention incorporates health promotion,

competence enhancement, and a framework for strategies that reduce risk factors and strengthen protective factors (Greenberg et al., 2003). The TGFD curriculum is supportive of the tenets of SEL through its emphasis on developmentally appropriate strategies to enhance adolescents' competencies in recognizing and managing their emotions, understanding and appreciating the perspective of others, establishing and pursuing positive goals, making responsible decisions, and handling interpersonal situations effectively. TGFD recognizes that children and adolescents strive for affiliation, responsibility, nurturance and companionship, and want to feel socially competent (Ryan & Shim, 2008). TGFD further recognizes that competency to interact in social contexts requires a range of social, cognitive and verbal skills (Landry, Smith, & Swank, 2009).

## Design of the Study

### Purpose of the Study

The purpose of the study was to determine the effectiveness of the TGFD prevention program in impacting 1) middle school students' use of drugs (cigarettes, alcohol and marijuana), and 2) several key risk and protective factors, including future intentions to use drugs, attitudes toward drugs, perceptions of harmful effects of drugs, emotional competency/self-efficacy skills, social and peer resistance skills, bonding with prosocial peers, and goal setting and decision making skills.

The evaluation study was also used to examine moderating effects of student risk level on ATOD substance use and on certain risk and protective factors (R&P). While all adolescents may be vulnerable to social, psychological and behavioral influences to use drugs, the most vulnerable students may be those who indicate having already tried cigarettes, alcohol, marijuana or other drugs. Such adolescents can be considered to be high risk in that they have indicated an early predisposition to experiment with one or more ATOD substances. Students who are early

ATOD users have been found to be more likely than nonusers to intensify their involvement with drugs and to express other problem behaviors over time (DuRant, Smith, Kreiter, & Krowchuk, 1999; Ellickson, Tucker, & Klein, 2001, 2003; Lynskey et al., 2003). For example, in a 10-year longitudinal study of alcohol use beginning in grade 7, Ellickson, Tucker, & Klein, (2003) found that both early drinkers and experimenters were more likely than nondrinkers to report substance use, delinquent behavior and academic difficulties in high school. Furthermore, early experimentation was associated with substance abuse in young adulthood, and early drinking was associated with substance abuse, employment problems, and violent behavior in young adulthood.

Although results are mixed, a number of studies have shown that drug prevention curricula can be effective in suppressing current or future drug use among high risk students (e.g., Botvin, Griffin, Diaz, & Ifill-Williams, 2001a; Eisen, Zellman, Massett, & Murray, 2002; Ellickson, McCaffrey et al., 2003; Flay, Phil, DuBois, & Ji, 2007; Hawkins et al., 1999; Longshore et al., 2007; Perry et al., 2002; Sussman et al., 1998). Some have shown no differential effect on drug use among high risk students in comparison to lower risk students (e.g., Donaldson, Graham, & Hansen, 1994). In regard to protective factors, Longshore et al. (2007) found that a school-based drug prevention program strengthened attitudes and perceptions against drug use among at-risk adolescent girls. Similarly, in a study of economically disadvantaged, inner-city minority adolescents, Botvin et al. (2001a) reported a positive effect on attitudinal and personality factors thought to be associated with adolescent substance use.

Evaluation of the TGFD intervention focused on the following questions:

a) Do middle school students receiving the TGFD prevention program, in comparison to students *not receiving* the program, report fewer instances of smoking cigarettes, using alcohol, binge drinking, and using marijuana within the previous month and year?

b) Is the impact of the TGFD intervention on students' ATOD substance use moderated by the risk level of the student?

c) Do middle school students receiving the TGFD prevention program, in comparison to students *not receiving* the program, indicate stronger intentions to resist the use of drugs in the future; greater awareness of the harmful effects of drugs; more positive attitudes concerning the inappropriateness of drug use; and higher levels of goal setting and decision making skills, social and peer resistance skills, emotional competency/self-efficacy skills, and bonding with prosocial peers?

d) Is the impact of the TGFD intervention on students' risk and protective factors moderated by the risk level of the student?

e) Do middle school students receiving the TGFD prevention program, in comparison to students *not receiving* the program, evidence more success on school performance indicators such as days absent from school, number of suspensions, and academic achievement?

f) Was the delivery of the TGFD prevention program accomplished in a complete, appropriate, and effective manner?

g) Was there evidence that student survey responses were reliable and valid?

It was expected that the TGFD intervention would have a positive impact in suppressing cigarette, alcohol and marijuana use, and that the impact would be greatest for students considered moderate to high risk. It was also expected that the TGFD intervention would have a positive impact in strengthening protective factors and weakening risk factors related to student cigarette, alcohol and marijuana use, and that the impact would be greatest for students considered moderate to high risk.

## Target Population and Sample

The target population for the study consisted of approximately 13,000 6th grade students in 45 middle schools. Sixth grade enrollment in these 45 schools ranged from 148 to 565, with a median enrollment of 340. The 6th graders in 40 of these middle schools (about 11,000) were selected for the sample. Fifty-nine percent of the student sample received federal free or reduced lunch program services, 14% limited English proficiency services, and 11% exceptional student education services. Approximately 52% of the sample was male. Ethnically, 38% of these 6th graders were white, 19% black, 30% Hispanic, 3% Asian and 10% multiracial.

## Random Assignment of Schools to Treatment/Control Conditions

For the evaluation, we employed a stratified randomized treatment-control design whereby 40 schools serving approximately 11,000 6th grade students were paired on the basis of key school demographic factors. The stratification factors included percent of students receiving free or reduced lunch services, minority representation, limited English proficiency services, exceptional education services; and where possible, location (urban, suburban & rural), school size, and prior academic performance in reading.

After the school pairing process was completed, one member of each school pair was then randomly assigned to the treatment condition and the other assigned to the control condition. Sixth graders in the treatment schools received the TGFD program at the beginning of the school year. No treatment intervention was provided to 6th graders in the control schools.

Once the schools were initially assigned to treatment and control conditions, the principal of each school was notified as to its status and asked if the school was willing to participate as designated. It was made clear to the principals of both treatment and control schools that their schools' participation in the study was voluntary and would require their cooperation and the

willingness of their teachers to allow survey data collection in 6th grade classrooms at designated times during the school year. Administrators for each of the 40 schools agreed to participate in the evaluation study. The careful pairing of schools on key attributes, followed by their random assignment to treatment-control conditions, provided high confidence in the initial equivalence of the sets of 20 treatment schools and 20 control schools.

**School and Parent Cooperation**

Once the 20 treatment and 20 control schools had communicated their willingness to participate in the study, a letter was sent to all parents or guardians of 6th graders scheduled to attend the schools. The letter was to inform them that during the school year the TGFD program would be undergoing an evaluation of its effectiveness and that, as part of this effort, their child may be asked to complete an anonymous survey on drug prevention. They were assured that no one would see their child's responses, and that all responses from the survey would be used only in aggregated form. They were further informed that the survey was voluntary and that if they would like to examine the survey, a copy would be available in the school's main office. Parents were provided school and study contact information if they did not wish their child to take part in the survey.

For each of the 40 participating schools a package was prepared containing a copy of the letter of approval for the study, a copy of the *Student Behavior and Risk and Protective Factor Survey* (SBRPFS), and a cover letter addressed to parents or guardians who wish to examine the SBRPFS. Each school was asked to keep the package so that it was available to any parent requesting to see it. The cover letter explained to the reader the purpose of the survey, and the fact that students who complete the survey will not be permitted to put their names or any ID on either the answer sheet or the survey booklet. The letter also assured the reader that every

question on the survey was carefully selected "to provide information needed to properly and efficiently evaluate the effectiveness of the TGFD prevention program". Those parents or guardians who wished to have further information about the survey, or who preferred that their child not participate in the survey, were referred to school and study contacts.

The 20 schools serving as treatment schools understood that they would receive the TGFD program the beginning of the school year and that the prevention lessons would be delivered by TGFD instructors. They further understood that their schools would be visited by survey administrators periodically to administer the SBRPFS. They also understood that there would be occasional visits from the evaluation team to observe the survey administrators during the giving of the survey, and to observe the TGFD instructors during the delivery of the lessons. Apart from these elements, each of the 20 treatment schools had agreed to allow their 6$^{th}$ grade teachers to complete the *Classroom Teacher Survey* (CTS) on a voluntary basis at the conclusion of the TGFD program. In addition, principals from each of the 20 treatment schools identified a 6$^{th}$ grade School Contact person who would complete the *Within School Activities Form* (WSAF) on a semester basis to help the evaluation team gather information concerning any confounding influences which may have occurred during the study.

The 20 control schools understood that they would serve as concurrent controls. They further understood that their schools would be visited by survey administrators periodically to administer the SBRPFS, prior to the delivery of the prevention program in the treatment schools, after the delivery of the intervention in treatment schools, and 6-months later. They also understood that there would be occasional visits from the evaluation team to observe the survey administrators during the giving of the survey. Apart from these elements, each of the 20 control

schools, like the treatment schools, had identified a 6th grade School Contact person who would complete the WSAF on a semester basis.

**Survey Schedule**

Students in the treatment and control schools were asked to complete assessment surveys on an identical schedule: 1) Pre-Survey—one to two weeks prior to implementing the treatment intervention, 2) Post-Survey—one to two weeks after delivery of the treatment intervention, and 3) 6-Month Follow-Up—six months after the end of treatment.

The need to examine for the effects of ATOD treatment beyond the immediate conclusion of the initial intervention derives from the knowledge that treatment effects tend to diminish over time (Flay, 2009; Gandhi, Murphy-Graham, Petrosino, Chrismer, & Weiss, 2007; Weiss, Murphy-Graham, Petrosino, & Gandhi, 2008). For example, a meta-analysis of smoking prevention programs by Hwang et al. (2004) found that all program effects (whether behavioral, attitudinal or knowledge-based) were smaller at delayed follow-ups. This fact, coupled with the volatile nature of adolescent growth and development, made it imperative that the evaluation design take into account the sustainability of initial treatment effects over time. In another meta-analysis of tobacco and other drug prevention programs, Skara and Sussman (2003) found that while program effects decayed over time, the effects were less likely to decay for programs with extended programming or booster sessions. Although beyond the scope of the present report, booster sessions were planned for the current subjects when they became 7th graders.

The review board for the school district approved the study under conditions of strict student anonymity. It required that student survey data be gathered such that no one—including evaluators, teachers or parents—could know the identity of individual student respondents, nor that individual students could be tracked from one survey assessment period to the next.

Therefore, a longitudinal linking of the individual student's responses was not possible. Linkage relative to treatment vs. control was pursued through questions on the post-survey and 6-month follow-up asking the respondents if they had attended the school from the beginning of the school year, and if they had received TGFD lessons during the school year.

## Instrumentation and Data Collection

### Online Compliance Management System

In collaboration with professionals from the Leap Frog Group, Inc., we developed an Online Compliance Management System (OCMS) utilizing open source survey and calendar software (Baker, Jarman & Bacon, 2009). The OCMS system allowed us to post online, through a secured offsite server, a calendar of each TGFD instructor's weekly scheduled lessons and a link from the calendar for each lesson taught to a survey checklist to provide implementation data. Specifically, an average 285 treatment classrooms were served on a weekly basis and over the course of the 10-week period a total of 2,836 lesson implementation checklists were completed by the TGFD instructors delivering the treatment intervention.

The OCMS was also used to reach 6$^{th}$ grade classroom teachers for end of program feedback about the quality of the delivery of the prevention program by the TGFD instructors serving their classrooms. Feedback from grade-level School Contacts from the 40 study schools about potential confounding influences and activities in their schools was also gathered twice a year using the OCMS.

The OCMS was a critical tool for scheduling, calendaring and linking our access to survey protocols for random observations of the survey administration at the pre-survey, post-survey and 6-month follow-up as well as arranging paired observations. We used the same

process to schedule, calendar, and link our access to observation protocols for each of the 10 TGFD lessons as well as paired observations of treatment delivery.

## Student Behavior and Risk and Protective Factor Survey (SBRPFS)

### SBRPFS Development

Students in both the treatment and control schools were asked to complete the *Student Behavior and Risk and Protective Factor Survey* at each of the assessment points described above. In addition to containing several demographic questions, the SBRPFS is comprised of questions encompassing two major areas: 1) students' use of cigarettes, alcohol and marijuana, and 2) risk and protective factors. All items on the SBRPFS employ a structured-response format, with the number of response options ranging from two to five.

According to Coyle, Russell, Shields, & Tanaka (2007), evidence suggests that children age 9 and over have sufficient cognitive abilities to contribute valid information about their own feelings, experiences, and behaviors through participation in many traditional data collection methods, including self-administered surveys. In the present study, all survey respondents were 11 or over at the start of the study, with the exception of 10 students (0.09%) at the beginning of grade 6. Donaldson, Thomas, Graham, Au, & Hansen (2000) note that the use of self-report measures is common in the substance abuse prevention literature because such measures are fairly easy to acquire and often provide the only practical way to measure the constructs of interest. However, they also note that respondents may tend to under report undesirable behaviors when using the self-report method. These authors sought to corroborate self-report results of a drug abuse prevention program, using reciprocal best friend reports. Their findings suggest that the underreporting of undesirable behavior through self reports may not be as great a problem as some suspect (Donaldson et al., 2000).

Several studies have found support for the validity of self-report measures of substance use among adolescents and young adults in relation to urine analysis, saliva tests or carbon monoxide breath samples (e.g., Botvin et al., 2001a; Donahue, Hill, Azrin, Cross, & Strada, 2007; Ellickson, Tucker & Klein, 2001, 2003; Murphy, Durako, Muenz, & Wilson, 2000; Solbergsdottir, Bjornsson, Gudmundsson, Tyrfingsson, & Kristinsson, 2004). For example, Donahue et al. (2007) found that contemporaneous self-reports of both youth and the parents of the youths' marijuana usage were corroborated by standardized retrospective reports and by urinalysis testing.

Section 1 of the SBRPFS contains several demographic questions covering descriptive elements such as, gender, ethnicity, age, current grade level, grade level in previous school year, and school attended during previous year. The exact number of demographic questions changed across the assessment points of the study (pre-survey, post-survey, 6-month follow-up), making it necessary to have slightly different forms of the SBRPFS over time.

Section 2 of the SBRPFS contains 25 structured-response questions asking whether or not the respondent had ever used cigarettes, alcohol or marijuana (3 questions); the age of onset of cigarette, alcohol and marijuana usage (3 questions); the extent of cigarettes, alcohol and marijuana usage within the last month (6 questions) and within the last year (6 questions); binge drinking (1 question); intent to use cigarettes, alcohol and marijuana over the next year (3 questions); and honesty in responding (3 questions). The questions concerning whether students had ever used cigarettes, alcohol or marijuana were dichotomous (Yes/No). The honesty questions used a 3-point scale (e.g., I said I smoked more than I really do; I was honest and accurate; I said I smoked less than I really do). All other questions in Section II employed a 5-point scale with the content of the options varying depending upon the specific question.

Section 3 of the SBRPFS contains 36 structured-response questions related to risk and protective factors associated with youths' risk taking and drug use (cigarettes, alcohol, marijuana) behaviors. These questions have been used successfully in previous evaluations of the TGFD program (Bacon, 2000). They cover goal setting and decision making skills (6 questions); social and peer resistance skills (6 questions); emotional competency/self-efficacy skills (6 questions); perceptions of harmful effects of drugs (6 questions); bonding with prosocial peers (6 questions); and attitudes toward drugs (6 questions). Each of the 36 questions uses a 5-option Likert format (Strongly agree, Agree, Not sure, Disagree, Strongly disagree). Most of the questions are stated such that "Strongly agree" is the most desirable response, but a few are stated such that "Strongly disagree" is the most desirable response.

The 36 risk and protective factor questions were adapted from items in leading national youth surveys, such as the *Monitoring the Future Surveys* (SAMHSA, Center for Substance Abuse Prevention, 2003), the *American Drug and Alcohol Survey* (Rocky Mountain Behavioral Science Institute, 2003), and the *Middle School Youth Risk Behavior Survey* (Centers for Disease Control and Prevention, 2010). The 25 drug usage questions were adapted from items in the same sources used for the R&P factor questions, as well as from items in the *Florida Youth Tobacco Survey* (Florida Department of Health, 2008).

Due to a variation in the number of demographic questions used from one survey administration time to another, the total number of questions on the SBRPFS ranged between 68 and 72. The SBRPFS was designed to be administered within a single class period, using a machine-scannable answer sheet. The directions to the SBRPFS remind the students that their participation is voluntary, that they will remain strictly anonymous, and that no one—including those at school or home—will see their individual answers. Evaluation studies of school-based

drug prevention programs that employ survey data collection techniques typically assure the respondent of confidentiality or anonymity to minimize problems related to validity—to encourage broad participation and promote accuracy or truthfulness in responding (e.g., see Anderson & Moore, 2009; Botvin, Griffin, Diaz, & Ifill-Williams, 2001b; Hecht et al., 2008; Longshore et al., 2007; Sussman et al., 1998). The SBRPFS was initially piloted in several 6th grade classrooms to gauge the time to administer and complete the survey and to determine any problems with the questions. The pilot information suggested that the survey took on average about 35 minutes to administer and complete, and that the questions were readable and understandable.

**Administration of the SBRPFS**

The SBRPFS was administered in the 6th grade classes in both treatment and control schools by trained survey administrators (SAs) over a two-week period immediately prior to the delivery of the prevention program, over a two-week period immediately after the delivery of the prevention program, and 6-months following the delivery of the program. In both treatment and control schools, the SBRPFS was administered in individual classrooms during regular classroom periods. The SBRPFS was available for students in English, Spanish and Braille.

At the beginning of each survey administration, the SAs checked with the classroom teacher regarding any students who had been prohibited from taking the survey by their parents/guardians. For any such student and for any student who opted not to take the survey, the classroom teacher provided an alternative activity that did not disturb those working on the survey. Prior to administering the survey, the SAs asked that students' desks be cleared and then checked to be sure each student had sufficient space to complete the survey without crowding from other students.

The scripted directions for administering the survey were followed for all survey administrations in both treatment and control schools. The students were informed that their responses to the survey would be anonymous (i.e., "no one will see your answers, not other students, not your teacher, not even me"), and that the survey was voluntary. To facilitate survey administration and scoring, scannable answer sheets were used. Because many of the survey questions used a 5-option Likert-type response scale (Strongly agree, Agree, Not sure, Disagree, Strongly disagree), two example questions with the response scale were posted using newsprint and presented to clarify for students how they might respond depending upon their perspective. Prior to beginning the survey, the students were encouraged to answer each question as honestly and accurately as possible. Johnson and Richter (2004) suggest that, when surveying sensitive behaviors, one should explicitly stress to the respondents the importance of response accuracy. To assure that students would not mistakenly place their name or ID on the answer sheet, the spaces for such information on the answer sheet were blacked out. The students were informed of the rules that would be operative during the survey and the rules were posted for all to see.

While the students were working on the survey, the SAs monitored the classroom--maintaining quiet and order, making sure students were using the answer sheet appropriately and not putting their name or any identifying information on the survey booklet or answer sheet. Students were told to raise their hand if they needed assistance and to turn their answer sheet over to assure the privacy of their responses while the SAs assisted them. If students asked for clarification of some aspect of a survey question, the SAs were allowed to quietly read aloud the word, term or phrase that concerned the student, but were not allowed to paraphrase or define the word, term or phrase. The SAs also watched for students who appeared to be off-task and tried to discretely refocus any such students. In the typical class of about 20 students, the survey took

about 30 minutes to complete, slightly longer in larger classes. As students began to reach the end of the survey, they were reminded by the SAs to check over their work to make sure they answered all questions, bubbled in each answer choice completely, made only one response per question, and left no stray marks on the answer sheet or booklet. All survey booklets and answer sheets were collected individually by the SAs.

As the SAs collected the survey materials, they were directed to discretely tag the answer sheets of any students who persistently and overtly exhibited off-task behavior, who started and then abruptly quit taking the survey, or who finished in 5 minutes or less. We viewed these highly visible behaviors as students' unspoken decision to not voluntary participate in the survey, in comparison to more assertive youth who signaled or told the SAs they opted not to take the survey. Because the survey classrooms seldom had more than 20 students, the SAs had little difficultly carrying out this task. These tagged answer sheets were excluded from the major analyses, as were the blank answer sheets from students who overtly opted not to participate in the survey.

Across the three survey administration periods, the SAs reported that all or virtually all students had sufficient time to complete the survey; these reports were confirmed by on-site observers in those survey administration sessions in which observations took place. Some students in both treatment and control schools missed the survey due to being absent from school. On random occasions a few students were not present in a classroom due to school administrative functions such as discipline problems, testing, and administering flu inoculations.

## Fidelity Instruments

Century, Rudnick, and Freeman (2010) state that there is a growing recognition of the value and necessity of measuring the fidelity of implementation in the evaluation of

interventions. These authors note that without the clear measurement of implementation, it is impossible to identify the cause of disappointing outcomes—Are they the result of an inadequate program model or simply a reflection of poor or incomplete implementation? For these authors, fidelity of implementation is the "black box" in the evaluation of intervention effectiveness. It is for this reason that Reichardt (2011) insists that adequate program descriptions should include a delineation of the treatment both as planned and as implemented, and should note the extent to which the treatment was or was not implemented as planned. In a review of school-based programs that focus on social and emotional learning (SEL) for elementary and middle school students, Payton et al. (2008) reports that 1) in a number of studies implementation was not mentioned, and 2) studies that reported implementation problems tended to have smaller effects than studies that reported no implementation problems. Dusenbury, Brannigan, Falco and Hansen (2003) warn that poor implementation is likely to result in a loss of effectiveness in drug abuse prevention. A meta-analysis by Tobler and Stratton (1997) found that school-based drug prevention programs delivered on a large scale tended to have smaller effect sizes than those delivered on a smaller scale, and surmised that the difference could be explained by implementation factors. They suggest that implementation could be a crucial mediating factor in determining success.

According to Brandon, Taum, Young, Pottenger III, and Speitel (2008), evaluation theorists, methodologists, and practitioners are increasingly asking for evidence of program implementation in the conduct of program evaluations. These authors lament that most articles and reports present only brief descriptions of the development and application of implementation instruments. For example, they note that while the use of observations is necessary for

examining implementation quality, many studies that employ observational methods present inadequate descriptions of the development, validation and use of these important tools.

To ascertain the fidelity of the 10-week TGFD implementation in treatment schools and the quality of the survey administration in both treatment and control schools, and to monitor within-school conditions that might impact the integrity of treatment-control comparisons, we developed the following checklists, surveys and observation protocols.

**Development of the Survey Administration Protocol (SAP)**

In recognition of the importance of consistency in the collection of student survey responses from all schools participating in the study, the *Survey Administration Protocol* (SAP) was developed. The SAP permits observers to monitor the process whereby the SBRPFS is administered to students in both treatment and control classrooms at each assessment point in the study. Using a 3-point scale (Demonstrated Fully, Demonstrated Partially, Not Demonstrated), the SAP contains 16 items by which an observer can rate the extent to which the survey administrator follows proper procedures 1) before distribution of the survey (6 items), 2) before students begin work on the survey (3 items), 3) during student work on the survey (5 items), and 4) after students finish (2 items).

The 16 items of the SAP were designed to directly reflect the elements of a prepared survey script that all survey administrators were trained to follow each time the SBRPFS was administered in treatment and control classrooms.

In addition to the 3-point ratings, the SAP requires the observer to indicate 1) whether all or almost all participating students had sufficient time to complete the survey, 2) whether there were any unusual problems or distractions in the classroom during the administration of the survey (and if so, explain), and 3) whether in the observer's judgment, there was anything about

the survey administration that might have compromised the integrity of the students' responses (and if yes, explain).

**Data Collection for the SAP**

The SAP was used for gathering observations pertinent to the fidelity of the survey administrations in the 6th grade classrooms of both treatment and control schools. The main function of the SAP observations was to spot possible problems that might compromise or contaminate the survey data being collected. A schedule of observations was established based on the premise that all survey administrators would be observed at least twice, once in a treatment classroom and once in a control classroom. Fifty class administrations of the SBRPFS were observed during the pre-survey, 65 class administrations of the SBRPFS were observed during the post-survey, and 66 class administrations of the SBRPFS were observed during a 6-month follow-up. All observations were unannounced such that the observers appeared in the survey setting without prior notice being given the SAs. The observers typically positioned themselves near a rear corner of the classroom, and conducted themselves in as unobtrusive a manner as possible.

To allow for an estimate of the inter-observer agreement of the observations, paired observations were conducted on about 30 of the survey administrations during the post-survey and 6-month follow-up. Approximately half the paired observations occurred in treatment classrooms and half in control classrooms.

As the observers exited the survey setting, critiques of individual SAs were typically avoided. If the observers detected an undesirable pattern emerging after observations in several survey settings, all SAs would be contacted by text message and email to caution them regarding

the matter for future consideration. After completing an observation, each observer entered the information into the OCMS.

**Development of the Lesson Implementation Checklists (LICs)**

To obtain the TGFD instructors' perspective on their delivery of the TGFD program in the 20 treatment schools, a *Lesson Implementation Checklist* (LIC) was developed for each of the 10 TGFD lessons. Each lesson was broken into several statements of 100 words or less, with each statement representing a coherent segment of lesson activity. As the activity statements were completed for each lesson, the activity statements were screened by curriculum personnel for adherence to the lesson and modifications made as needed.

Depending upon the nature of the lesson, the number of statements representing a lesson ranges from 6 to 12. Placed to the right of each statement is a 4-point scale allowing the TGFD instructor to indicate whether he/she completed "All of the Activity", "Most of the Activity", "Some of the Activity", or "None of the Activity". Thus for a given classroom lesson, the TGFD instructors rate the extent to which they perceive that they implemented each of the activities for that lesson. An LIC is completed for each lesson each time the lesson is given.

Following the several implementation statements on the LIC is Section 2, containing a listing of 12 categories of circumstances that might explain a rating of less than "All of the Activity" (e.g., class started late, lesson was interrupted due to unscheduled events, student disruptive behaviors made classroom management challenging). An "Other" category is added to cover eventualities not represented by the 12 listed categories. To create the list, we distributed an initial set of possible categories among a number of experienced TGFD instructors and made adjustments to the list based on their input. If one or more checklist statements is assigned a rating of *less* than "All of the Activity" completed, the TGFD instructor is asked to

identify all of the circumstances from the list of 13 categories that he or she believes contributed to the incompletion. If "Other" is selected, a brief explanation is required.

The last section of the LIC (Section 3) contains a set of 6 questions that the TGFD instructor is to complete regarding the responsiveness of the students in the class to the lesson that was delivered. Each of these items uses a 4-point scale: Excellent, Good, Fair, and Poor. The questions focus on students' on-task behavior; willingness to be active participants; being respectful of others; following rules; expressing themselves in relevant ways; and responsiveness to the various lesson activities.

**Data Collection for the LICs**

The LICs were used by the TGFD instructors to record their perspectives on how completely they had implemented each TGFD lesson in the 6th grade classrooms, and the reasons for lesson elements not being completely implemented. The schedule for the delivery of the lessons was such that, during any given day, the TGFD instructors typically had back-to-back classes in their school, or needed to leave one school after teaching a class and travel to another school for the next class. Logistically, this meant that it was often impractical for the instructors to transmit their online responses immediately upon the conclusion of each class of instruction, although some instructors were able to enter implementation data immediately after the lesson using their cell phone or PDA device. Therefore, they were provided with printed copies of the LIC formatted to allow the recording of ratings and notes over an entire day. They were encouraged to use these to keep an abbreviated record of the day's classes and to refer to these ratings and notes as needed when transmitting their responses online to the OCMS.

**Development of the Treatment Observation Protocols (TOPs)**

The completion of an LIC by the TGFD instructors for each lesson they delivered was important in representing the perceptions of those responsible for delivering the instruction. In addition, evidence of program implementation fidelity came from a separate source—a representative sample of in-class observations conducted by the evaluation team. For these observations, a *Treatment Observation Protocol* (TOP) was created for each lesson. The activity statements on the TOP for a given lesson were duplicates of the activity statements on the corresponding LIC for that same lesson, and the 4-point response scale ("All of the Activity" to "None of the Activity") was the same. Sections 2 and 3 of each TOP were also duplicates of those on the corresponding LIC, permitting the observers' perspective on the reasons for lesson activities not being fully delivered and the responsiveness of the students to the lesson.

Each TOP also contained a set of items which allowed the observer to rate the TGFD instructor in terms of their instructional performance. Using a 4-point scale (Demonstrated Consistently, Demonstrated Occasionally, Demonstrated Rarely, and Not Demonstrated), the TOP allows observers to rate the TGFD instructor's behavior in regard to whether he/she was prepared for instruction; used appropriate classroom management strategies; kept students on task; actively engaged students in the learning process; effectively transitioned between lesson activities; provided clear directions; defined terms, provided explanations and gave examples; provided students with opportunities to participate and practice; recognized and reinforced students for participating; and responded to student input in a receptive and supportive manner.

**Data Collection for the TOPS**

The TOPs were used for gathering observations on the fidelity and completeness of lesson delivery in the 6$^{th}$ grade classrooms of the 20 treatment schools. The observation team

established a schedule of observations based on the premise that all TGFD instructors would be observed at least twice, once during the delivery of one of the first five lessons and once during the delivery of one of the last five lessons. Seventy-one treatment classes were observed during the 10-week period of instruction. All observations were unannounced such that the observers appeared in the classroom treatment setting without prior notice being given the TGFD instructors. The observers typically positioned themselves near a rear corner of the classroom, and conducted themselves in as unobtrusive a manner as possible. After completing each classroom observation, the observers entered the information directly to the OCMS using a PDA device or transferred their information from hardcopy (paper and pencil) to the OCMS at the end of the day.

To allow for an estimate of the inter-observer agreement of the treatment observations, 24 of the visits were paired observations. As with the single observations, approximately half the paired observations were completed during the delivery of lessons 1-5 and half completed during the delivery of lessons 6-10.

**Classroom Teacher Survey (CTS)**

The TGFD lessons were typically delivered to the 6th graders in their science classrooms, with the science teacher always present (excluding occasions when a substitute teacher was required). In a few schools, the lessons were delivered in the social studies classroom, again with the regular teacher always present. The *Classroom Teacher Survey* (CTS) was developed to permit these 6th grade teachers in the 20 treatment schools to provide input at the end of the 10-lesson treatment regarding the TGFD instructor's overall performance and the perceived impact of the TGFD program.

Using a 5-point Likert scale (Strongly agree, Agree, Uncertain, Disagree, Strongly disagree), the CTS presents the 6th grade teacher with items similar to those the observer uses to rate the classroom performance of the TGFD instructor. In addition, the CTS permits the 6th grade teacher to indicate whether the students' comments and actions 1) suggest that they enjoyed the TGFD lessons, 2) felt the lessons were relevant to their lives, and 3) suggest that the TGFD program had a positive impact on their behaviors and choices. The CTS is completed online after Lesson 10 and only by those classroom teachers in whose classrooms the TGFD instruction took place.

**Within School Activities Form (WSAF)**

Performance bias due to contamination of the control group is historically a common problem in school-based treatment interventions because most schools already have in place certain prevention programs or activities. These programs or activities may even be based on the same or similar theoretical principles as the program being evaluated (Flay, 2009). For example, in a study of the keepin' it R.E.A.L. substance use curriculum Hecht et al. (2003) reported that the control subjects participated in their schools' existing substance use prevention programs, consisting of other research-based programs as well as programs supported by local professional athletes. In addition, during the implementation period an extensive state-wide anti-tobacco campaign was launched that included celebrity endorsements, television commercials and billboards. These researchers correctly noted that the keepin' it R.E.A.L. program, therefore, was being compared to standard existing interventions. According to Gandhi et al. (2007), in some studies it isn't at all clear what the control subjects are receiving; virtually nothing at all, some form of drug prevention as part of the school's general health curriculum, or something more extensive.

It should also be noted that when schools are randomly assigned to treatment-control conditions, whatever is present in the way of substance use prevention in the control schools must also be assumed to be present in the treatment schools. Few studies attempt to monitor anti-drug use activities in treatment schools beyond the targeted prevention program, or discuss the implications of the possible impact of these activities in relation to the targeted program. There is also the very real possibility that—immediately before, during, or immediately after program implementation—one or more schools in either the treatment group or the control group could experience a dramatic event or series of events that could produce a greater sensitivity to drug use among the student population.

The above concerns about potential contamination led to an effort to monitor substance-related activities in both the treatment schools and control schools. Once each of the 40 schools had agreed to participate in the study either as a treatment school or control school, the school principal was asked to select a School Contact person from among the school's 6$^{th}$ grade teaching staff. The function of the School Contact was to note on a semester basis, any school-wide or grade-wide activities or occurrences, apart from the TGFD, that could possibly impact the attitudes or behavior of the students. Such information was considered important in subsequent data interpretations, specifically with respect to questions of the internal validity of the overall study.

The School Contact person in each of the 40 schools was asked to complete a *Within School Activities Form* (WSAF). The WSAF provides the School Contact with a list of possible planned programs or activities (e.g., Great American Teach In, Red Ribbon activities, SRO class presentations) to be checked if they occurred at that school. The School Contact is asked in subsequent questions on the form whether any of the checked programs or activities could have

impacted the students' drug-related attitudes, perceptions or behaviors, or their skills in goal-setting, decision-making, managing emotions or peer pressure refusal. The School Contact is also asked about any unplanned or unscheduled events or occurrences at the school that may have affected the attitudes, perceptions or behaviors of the students.

The School Contact person was asked to complete the WSAF form online at the end of each semester during the school year, consulting with fellow teachers prior to completing the form.

## Training and Orientation

The appropriate training of those who are to implement school-based drug prevention programs can be a key to the ultimate impact of those programs. Ringwalt et al. (2007) notes that the inadequacy of training and support received by teachers may be one of the challenges to successful implementation of research-based prevention programs. This is especially the case if the prevention curriculum requires innovative instructional strategies with which teachers are unfamiliar or uncomfortable (Ringwalt et al., 2007). One may presume that some type of training occurs prior to the initiation of an intervention, whether implementation is by regular classroom teachers or special instructors or service providers external to the school system. However, a number of studies of school-based interventions fail to mention training at all, and those that do typically provide only limited information about the nature and extent of the training provided those responsible for implementation.

Quite apart from program implementation training, there is the important matter of training for the proper collection and transmission of data. In the current study, training was provided by the evaluators for the following data collection activities: 1) the completion of the LIC by the TGFD instructors, 2) the completion of in-class observations of the treatment by the

observation team using the TOP, 3) the student survey administration by the SAs, and 4) the completion of in-class observations of the student survey administration by the observation team using the SAP.

**Training for the TGFD Prevention Program**

Each of the instructors responsible for delivery of the TGFD curriculum holds a Bachelor's or Master's Degree and has completed a background security screening. Prior to the delivery of the middle school curriculum program, the TGFD instructors underwent a three-day workshop provided by trainers from the C. E. Mendez Foundation, that focused primarily on the 10 TGFD lessons, and also included attention to social and emotional learning (SEL) standards, cooperative learning and instructional strategies. During the curriculum training, much emphasis was given to group discussion and techniques of modeling, observation and practice. One or more members of the evaluation team were present to observe the curriculum training.

**Training for the OCMS and Completion of LICs**

Following the 3-day training for the prevention program, we provided the TGFD instructors with a full day of training and orientation on the purpose of the Lesson Implementation Checklists (LICs), and procedures for completing and electronically transmitting their LIC responses online using the OCMS. For Section 1 of the LIC instrument, printed definitions were given of the meaning of "All of the activity", "Most of the activity", "Some of the activity", and "None of the activity". One of the program lessons was modeled in segments as the TGFD instructors watched and listened. After each modeled segment, the instructors used the LIC to rate the appropriate activity statements for the lesson in terms of whether the lesson presenter had presented all of the activity, most of the activity, some of the activity, or none of the activity. This exercise led to a discussion of the meaning of the ratings, specifically what

constituted “all” of an activity as compared to” most” or “some” of an activity. These exercises also helped provide a peer baseline as a future reference for when the instructors' would rate themselves during the actual treatment delivery.

The TGFD instructors were asked to transmit their online LIC responses for each classroom lesson as soon as possible after the delivery of that lesson. However, it was understood that if their schedule had them delivering a lesson in several classes in quick succession, several hours delay could occur in transmitting the responses for each lesson. In such circumstances, the TGFD instructors were asked to check off and make brief notes on specially prepared paper forms of the LICs (covering an entire day's lessons) after each lesson was delivered in order to assist their memory when completing the online LICs for those lessons.

The TGFD instructors were also given an orientation to the TOP and the circumstances under which they would be observed with that instrument. The TGFD instructors were informed that each would be observed on at least two different occasions and that the observations would be unannounced. They were assured that the purpose of the observations was to gather evidence of the overall fidelity of treatment delivery, and not to critique individual TGFD instructors nor provide disaggregated data to their employers. TGFD instructors/service providers and student program materials were funded in part by a grant from the Florida Department of Children and Families and a local city's Community Development Block grant.

**Training and Orientations for Administration of the SBRPFS**

To insure the quality of the pre-survey data gathered using the SBRPFS, we provided a full day of training for survey administrators (SAs) prior to the start of the delivery of the treatment intervention by the TGFD instructors. Two-hour orientations were also provided by the evaluators for each subsequent administration of the survey (post-survey and 6-month

follow-up) to reacquaint SAs with the survey script and all procedures pertinent to survey data collection. One-on-one and small group trainings were also provided when there were changes or attrition among the SAs. The training involved acquainting the attendees with a detailed script entitled, “Directions for Administering Student Survey”. The script contains directions on what each SA is to do and say in the classroom 1) before distributing the survey, 2) before students begin work on the survey, 3) during student work on the survey, and 4) after students finish work on the survey. Verbatim instructions to the student were represented in red within the script. Emphasis was placed on the importance of communicating to students the need to read each survey question carefully and to respond honestly.

We explained the necessity for strict adherence to the script in all treatment and control classrooms in which the survey was to be administered, and then modeled the administration of the survey, using the script. The attendees were then paired off and given the opportunity to practice using the script, with each partner taking turns critiquing the other. The attendees were also given an orientation on the following issues: 1) handling class disruptions during the survey administration; 2) limits on what could be said to a student who asks for clarification regarding a particular item (e.g., if a student does not recognize a word in a survey item, the SA is allowed to pronounce the word only, not explain or describe it); 3) procedures for packaging the completed survey materials for transmission to the evaluators; and 4) procedures for safeguarding all survey materials as the survey administrator travels in and out of schools and classrooms. Questions raised by the attendees regarding the survey administrative procedures were dealt with throughout the training

Toward the end of each training session, each of the SAs was given a survey administration box that contained all of the survey-related materials that would be needed over

the two week period during which the surveys would be administered. The evaluators acquainted the SAs with the contents of the Survey Administration Box and all procedures concerning the use of the contents. Careful attention was given to procedures for the security of all survey-related materials, and to the proper preparation of the completed surveys prior to their submission to the evaluators.

The survey administrators were also given an orientation to the SAP and the circumstances under which they would be observed with that instrument. They were informed that each would be observed on at least two different occasions and that the observations would be unannounced.

**Data Security Procedures**

Throughout the evaluation effort, various procedural elements were implemented in order to maintain the security of the survey data during data collection and retrieval. Survey administrators were trained in the proper procedures for gathering and organizing all survey materials at the end of each survey administration, and for packaging the machine-readable survey answer sheets for physical transmission to the evaluators. All SAs followed prescribed protocol for assuring that all survey answer sheets and survey booklets were in hand before exiting a classroom, and all were careful to separate usable from non-usable survey answer sheets before transmitting both to the evaluators. We maintained continual contact via text and email and during unannounced observations with all SAs during and after each two-week period of survey data collection, to answer any questions that might arise and to check on the progress the SAs were making in meeting their scheduled survey administration sessions in the schools.

Once all the survey answer sheets were collected, they were checked for stray pencil marks or marks that may not be easily read during the scanning process. After each answer

sheets was checked, the sheets were scanned to generate the databases, and hard copies of the response forms stored for future reference if needed.

All data collected using the OCMS was stored on an off-site redundant server that was hidden from typical internet search engines. Access to online calendars with scheduled events and links to accompanying survey tools for the TGFD instructors were protected with usernames and passwords. Sixth grade classroom teachers received their CTS survey request through email with a direct link to their form, and once completed the survey could not be re-accessed by the user. No respondent using the online compliance management system as part of the evaluation study could access or alter any other part of the system or surveys, having access only to the tools needed to fulfill their individual role.

## Instrument Validation and Reliability Estimation

### Validation/Reliability Pilot Study of the SBRPFS

A critical criterion for the rigorous study of any educational, prevention and intervention programs is the use of valid and reliable instruments. Validity focuses on the appropriateness of the interpretations and uses of test or survey scores. The process of validation involves the collection of empirical evidence to provide stable and acceptable interpretations of scores (De Gruijter & van der Kamp, 2008). Reliability concerns the precision of measurement. Estimates of the reliability of an instrument provide researchers and evaluators with information as to the extent to which observed scores reflect true scores (McDonald, 1999).

Coefficient alpha (Cronbach's alpha) is a popular and extensively used measure for estimating the reliability of survey measures and other instruments commonly used in the evaluation of school-based programs. It usefulness derives in part from its convenience; it has the practical advantage of being estimated from the responses of one offering of the instrument.

This means that there is no need for more than one form of the instrument, nor a need to have respondents complete the instrument on more than one occasion (Furr & Bacharach, 2008).

However, as noted by Ghiselli, Campbell, and Zedeck (1981), the conditions under which reliability is estimated should be similar to the conditions under which scores on an instrument are to be used to make decisions. For example, if an instrument is to be used to predict behavior over time, then the instrument should demonstrate reliability across time. The *Standards for Educational and Psychological Testing* (American Educational Research Association et al., 1999) as well as *The Program Evaluation Standards* (Joint Committee on Standards for Educational Evaluation, 1994) recognize that different data collection procedures are sensitive to different sources of measurement error. The latter publication states, "it is thus important that the form of reliability to be considered takes into account the sources of error present in each procedure. For example, in some instances, the primary concern may be with stability of a measurement from one occasion to another (1994, p. 153)." Similarly, Crocker and Algina (1986) stress the need for identifying the types of measurement errors likely to be of greatest concern to the instrument's users and to estimate reliability in a way that will allow assessment of the effect of these measurement errors on the instrument's reliability. For the pilot study we considered it appropriate to generate internal consistency estimates of the SBRPFS using coefficient alpha, as well as test-retest estimates (coefficients of stability) based on the administration of the survey to the same students on two different occasions.

From a validation perspective, it was considered useful to examine the relationship between items subsets on the SBRPFS and items or item subsets drawn from established student survey instruments measuring the same or similar constructs, with all items being administered at the same time. The resulting concurrent correlations can provide a degree of confidence

regarding the relevance of the SBRPFS in reflecting certain ATOD behaviors and pertinent risk and protective factors.

**Design of Validation/Reliability Pilot Study**

Sixth graders from four middle schools not serving as either a treatment or control school comprised the sample for the pilot study for the specific purpose of providing additional data for estimating the reliability and validity of the SBRPFS. Data collection for the pilot study involved 1) splitting the SBRPFS into two parts, a Student Behavior Survey (SBS) and a Student Risk & Protective Factors Survey (SRPFS), 2) administering both the SBS and the SRPFS—along with similar items from other survey instruments—to a sample of 6th graders in the pilot schools not participating in the main study, and 3) re-administering only the SBS and the SRPFS to the same samples of students one week later.

The design allowed for an investigation of the concurrent validity of the SBRPFS subscales, using item components of other established instruments measuring the same or similar constructs. It also allowed for an investigation of the reliability of the SBRPFS subscales, specifically, the estimation of internal consistency (Cronbach's alpha) and the estimation of stability (test-retest). Several authors (e.g., Alwin, 2007; Crocker & Algina, 1986; Furr & Bacharach, 2008; Meyer, 2010) caution that reliability estimation using a test-retest method requires a careful balancing act to find a time span between testing that is neither too short nor too long. Very short time spans may introduce carryover effects such as memory or practice, and very long time spans may open the door to maturational or educational changes. Either circumstance can work to destabilize an individual respondent's true score. In the present study we considered a span of seven days to be a reasonable compromise for purposes of stability estimation.

**Validity/Reliability Pilot Study Data Collection Procedures**

Orientation for the survey administrators and data gathering within pilot classes occurred during the middle of the school year. To lessen the possibility of order effects within the validation data, the SBS and accompanying validation items given during the first week were printed and administered in two forms. One form (SBS, Form B1) contained the SBS items in the first half of the instrument, and behavior items from other surveys in the last half of the instrument. The second form (SBS, Form B5) contained the same behavior items, but in reverse order, i.e., the items from other surveys were in the first half of the instrument, and the SBS items were in the last half. Each form was given to about 120 6th graders in classrooms from the four non-study schools set aside for validation purposes. Overall, a total of approximately 240 6th graders were surveyed as part of the SBS validation. During week two the SBS (Form B), without the accompanying items from other surveys, was given in each of the 6th grade classes that had received Forms B1 and B5 the week before.

During week one, the same script was followed for both Form B1 and Form B5 of the SBS. The script called for students to place their responses directly on the survey booklet. The script followed for SBS Form B during week two conformed closely to the script used during week one. All surveys (SBS Forms B1, B5 and B) were administered by trained survey administrators who had given the SBRPFS previously, and were administered within intact classrooms.

Identical procedures to those used for the SBS were followed in the administration of the SRPFS (risk and protective factor items). The SRPFS and accompanying validation items given during week one were printed and administered in two forms. One form (SRPFS, Form A1)

contained the SRPFS items in the first half of the instrument, and risk and protective factor items from other surveys in the last half of the instrument. The second form (SRPFS, Form A5) contained the same risk and protective factor items, but in reverse order. As with the SBS, each form of the SRPFS was given to about 170 6th graders in classrooms from the four non-study schools. Overall, a total of approximately 340 6th graders were surveyed as part of the SRPFS validation. During week two the SRPFS (Form A), without the accompanying items from other surveys, was given in each of the 6th grade classes that had received Forms A1 and A5 the week before. It should be noted that the 6th graders receiving the SRPFS, although from the same schools, were not the same students that received the SBS.

As with the SBS, the script used for Form A1 and Form A5 of the SRPFS was the same, and called for students to place their responses directly on the survey booklet. The script followed for the SRPFS during week two conformed closely to the script used during week one. All surveys (SRPFS Forms A1, A5 and A) were administered by trained survey administrators who had given the SBRPFS previously, and were administered within intact classrooms.

No student names or ID number were required in the data collection. To facilitate the pairing of students' survey sheets from week one and week two, several demographic items were included on the survey forms. Students were asked to 1) indicate their month and day of birth, 2) select from sets of letters the set that contained the first letter of their first name, and 3) select from sets of letters the set that contained the first letter of their last name. Gender, ethnicity and age were also collected for each student.

**Source of Validation Items**

The item subsets on the SBS and SRPFS subscales were examined in terms of their congruence with items on other widely used instruments. The items and item subsets used in the validation of the SBS subscales on Forms B1 and B5 were drawn from the *National Parents' Resource Information for Drug Education* (PRIDE) *Risk and Protective Factor Questionnaire* (International Survey Associates, 2009b) and the *Survey of Student Attitudes and Behavior* (Project ALERT, 2009). Several items subsets came from the Core Measures Initiative (SAMSHA, Center for Substance Abuse Prevention, 2003). The items and item subsets used in the validation of the SRPFS subscales on Forms A1 and A5 were drawn from the same sources identified above, as well as from the *Communities That Care Youth Survey* (International Survey Associates, 2009a), the *National Youth Survey Baseline Questionnaire, 12-18 Version* (SAMHSA, Center for Substance Abuse Prevention, 1999), and the Botvin *Life Skills Training Questionnaire, Middle School* (Princeton Health Press, 2004). Finally, a self-efficacy measure by Prothrow-Stith was drawn from a compendium of assessment tools in Dahlberg, Toal, Swahn, & Behrens (2005). The expectation was that the SBS and SRPFS subscales would demonstrate congruence with their corresponding validation measures.

The subscales of the SRPFS and their corresponding validation measures are shown in Table 1; these include Goal Setting and Decision Making Skills, Bonding with Prosocial Peers, Social and Peer Resistance Skills, Emotional Competency/Self-Efficacy Skills, Perceptions of Harmful Effects of Drugs, and Attitudes Toward Drugs. The subscales of the SBS and their corresponding validation measures are shown in Table 2; these include Extent of Cigarette Use in Past Month and in Past Year, Extent of Alcohol Use in Past month and in Past Year, Binge

Drinking in Past Month, Extent of Marijuana Use in Past Month and in Past Year, and Intent to Smoke, Drink and Use Marijuana in Future.

For purposes of scale validation, the congruence of a measure with other measures can be examined from the standpoint of either positive or negative associations expected with those other measures. To investigate the SRPFS subscales in relation to a construct to which they should be inversely correlated, we selected the 6-item Orpinas *Modified Depression Scale* (MDS) from the compendium of assessment tools in Dahlberg et al. (2005). The expectation is that the stronger the protective factors of the 6th graders, the less likely are they to express depressive symptoms, thereby producing an inverse relationship between the responses on the protective factors and the responses on the MDS. In the validation study, each of the SRPFS subscales was examined in relation to the MDS.

Table 1

*Subscales of SRPFS Forms A1 and A5 and Corresponding Validation Measures*

| SRPFS Subscale | No. Items | Validation Measure | No. Items |
|---|---|---|---|
| Goals & Decisions | 6 | Goals & Decisions CSAP | 8 |
| Bond Prosocial Peers | 6 | Bond Prosocial Peers PRIDE RP | 3 |
| | | Bond Prosocial Peers ALERT | 3 |
| Peer Resistance | 6 | Peer Resistance LSTQ | 8 |
| | | Peer Resistance NYS | 3 |
| Self-Efficacy | 6 | Self-Efficacy Prothrow | 7 |
| Harmful Effects | 6 | Harmful Effects PRIDE RP | 4 |
| | | Harmful Effects ALERT | 3 |
| ATOD Attitudes | 6 | ATOD Attitudes CTCYS | 6 |

Table 2

*Subscales of SBS Forms B1 and B5 and Corresponding Validation Measures*

| SBS Subscale | No. Items | Validation Measure | No. Items |
|---|---|---|---|
| Smoke Past 30 Days | 2 | Smoke Past 30 Days PRIDE RP | 2 |
| Smoke Past Year | 2 | Smoke Past Year ALERT | 1 |
| | | Smoke Past Year PRIDE RP | 1 |
| Drink Past 30 Days | 2 | Drink Past 30 Days ALERT | 1 |
| | | Drink Past 30 Days PRIDE RP | 1 |
| Binge Drink | 1 | Binge Drink ALERT | 1 |
| Drink Past Year | 2 | Drink Past Year ALERT | 1 |
| | | Drink Past Year PRIDE RP | 2 |
| Dope Past 30 Days | 2 | Dope Past 30 Days ALERT | 1 |
| | | Dope Past 30 Days PRIDE RP | 1 |
| Dope Past Year | 2 | Dope Past Year ALERT | 1 |
| | | Dope Past Year PRIDE RP | 1 |
| Intent ATOD Use | 3 | Intent ATOD Use ALERT | 3 |
| | | Intent ATOD Use Tangle (CSAP) | 3 |

## Data Analysis Procedures for Validity/Reliability Pilot Study

For purposes of reliability and concurrent validity estimation on the SBS, the responses from Form B1 and Form B5 were first aggregated together. The responses from Form A1 and Form A5 were also aggregated together for purposes of reliability and concurrent validity estimation on the SRPFS. Internal consistency estimates of reliability were computed using Cronbach's alpha. Stability estimates of reliability (test-retest) were computed by correlating SBS and SRPFS subscale responses from Week 1 to participants' responses seven days later during Week 2. Concurrent validity estimates were computed by correlating SBS and SRPFS subscales with their corresponding validation measures.

**Results for SRPFS Pilot Study**

Presented in Table 3 below are the gender, ethnicity and age of the participants in the SRPFS portion of the validity/reliability pilot study.

Table 3

*Gender, Ethnicity, and Age of 6th Grade SRPFS Pilot Study Participants by Number and Percent*

| Characteristic | | *n* | *%* |
|---|---|---|---|
| Gender[a] | | | |
| | Female | 161 | 47.1 |
| | Male | 181 | 52.9 |
| Ethnicity | | | |
| | Black | 71 | 20.7 |
| | Hispanic | 76 | 22.1 |
| | White | 147 | 42.9 |
| | Asian | 10 | 2.9 |
| | Multi-racial | 39 | 11.4 |
| Age[b] | | | |
| | 14 | 3 | .9 |
| | 13 | 32 | 9.4 |
| | 12 | 171 | 50.1 |
| | 11 | 135 | 39.6 |
| Total | | 343 | 100.0 |

a = missing 1 response
b = missing 2 responses

***Results for Reliability Estimation of the SRPFS Subscales***

Estimates of internal consistency (Cronbach's alpha) reliability for the various subscales of the SRPFS are presented in Table 4 below. Table 5 contains estimates of test-retest reliability for the SRPFS. Table 6 contains estimates of internal consistency reliability for the comparison validation measures. As indicated in Table 4, virtually all of the alpha reliabilities of the SRPFS subscales were above .70. Similar findings for the SRPFS subscales are shown in Table 5 for the estimates of stability (test-retest). Overall, these results are supportive of the reliability of the SRPFS and its subscales. Alpha reliabilities of the validation measures that correspond to the SRPFS subscales, shown in Table 6, were likewise virtually all above .70.

Table 4

*Internal Consistency Estimates of Reliability of SRPFS Subscales for the Pilot Study*

| | Week One | | | Week Two | | |
|---|---|---|---|---|---|---|
| SRPFS Measure | *n* | No. Items | α | *n* | No. Items | α |
| Goals & Decisions | 332 | 6 | .771 | 340 | 6 | .829 |
| Bond Prosocial Peers | 327 | 6 | .795 | 338 | 6 | .884 |
| Peer Resistance | 318 | 6 | .796 | 341 | 6 | .841 |
| Self-Efficacy | 317 | 6 | .607 | 339 | 5 | .733 |
| Harmful Effects | 312 | 6 | .908 | 341 | 6 | .891 |
| ATOD Attitudes | 304 | 6 | .747 | 338 | 6 | .783 |

Table 5

*Test-Retest (One Week Delay) Estimates of Reliability of SRPFS Subscales for the Pilot Study*

| SRPFS Measure | *n* | No. Items | $r_{xx}$ |
|---|---|---|---|
| Goa1s & Decisions | 333 | 6 | .756 |
| Bond Prosocial Peers | 330 | 6 | .740 |
| Peer Resistance | 321 | 6 | .636 |
| Self-Efficacy | 317 | 6 | .721 |
| Harmful Effects | 313 | 6 | .689 |
| ATOD Attitudes | 307 | 6 | .719 |

Table 6

*Internal Consistency Estimates of Reliability of Validation Measures for the Pilot Study*

| Measure | *n* | No. Items | α |
|---|---|---|---|
| Goals & Decisions CSAP | 333 | 8 | .782 |
| Bond Prosocial Peers PRIDE RP | 319 | 3 | .670 |
| Bond Prosocial Peers ALERT | 316 | 3 | .839 |
| Peer Resistance LSTQ | 340 | 8 | .816 |
| Peer Resistance NYS | 338 | 3 | .883 |
| Self-Efficacy Prothrow | 330 | 7 | .701 |
| Harmful Effects PRIDE RP | 329 | 4 | .756 |
| Harmful Effects ALERT | 330 | 3 | .767 |
| ATOD Attitudes CTCYS | 324 | 6 | .799 |

***Results for Concurrent Validity Estimation of the SRPFS Subscales***

Concurrent validity coefficients for the SRPFS subscales are presented in Table 7. The SRPFS subscales were correlated with corresponding validation measures. Also, to investigate the relationship between the various protective factor measures and a measure thought to be inversely related to protective factors, the Orpinas *Modified Depression Scale* (MDS) was used. The resulting correlations are reported in Table 8.

As indicated in Table 7, the SRPFS subscales show a notable degree of congruence with the appropriate validation measures; concurrent validity coefficients range from .420 for the Harm subscale to .676 for the Goals subscale. As indicated in Table 8, the correlations show a consistently inverse relationship between the MDS and both the SRPFS subscales and the validation measures. The pattern of negative relationships appears slightly stronger for the SRPFS subscales. Overall, the results on the MDS suggest that the stronger the protective factors of the participating 6th graders, the less likely they are to express depressive symptoms. The results in both Table 7 and Table 8 are supportive of the validity of the SRPFS subscales.

Table 7

*Concurrent Validity Estimates of SRPFS Subscales with Validation Measures for the Pilot Study*

| Measures | *n* | *r* | *p* |
|---|---|---|---|
| Goals & Decisions (SRPFS/CSAP) | 333 | .676 | .0001 |
| Bond Prosocial Peers (SRPFS/PRIDE RP) | 319 | .522 | .0001 |
| Bond Prosocial Peers (SRPFS/ALERT) | 318 | .536 | .0001 |
| Peer Resistance (SRPFS/LSTQ) | 320 | .504 | .0001 |
| Peer Resistance (SRPFS/NYS) | 318 | .454 | .0001 |
| Self-Efficacy (SRPFS/Prothrow) | 318 | .490 | .0001 |
| Harmful Effects (SRPFS/PRIDE RP) | 309 | .420 | .0001 |
| Harmful Effects (SRPFS/ALERT) | 308 | .477 | .0001 |
| ATOD Attitudes (SRPFS/CTCYS) | 300 | .645 | .0001 |

Table 8

*Correlation of Modified Depression Scale (MDS) with SRPFS Subscales and with Validation Measures for the Pilot Study*

| SRPFS Subscale | n | $r^{**}$ | Validation Measure | n | $r^{**}$ |
|---|---|---|---|---|---|
| Goals & Decisions | 307 | - .438 | CSAP | 307 | - .371 |
| Bond Prosocial Peers | 306 | - .341 | PRIDE RP | 296 | - .218 |
| | | | ALERT | 295 | - .252 |
| Peer Resistance | 306 | - .400 | LSTQ | 306 | - .265 |
| | | | NYS | 304 | - .293 |
| Self-Efficacy | 306 | - .572 | Prothrow | 304 | - .336 |
| Harmful Effects | 307 | - .345 | PRIDE RP | 302 | - .257 |
| | | | ALERT | 301 | - .266 |
| ATOD Attitudes | 305 | - .371 | CTCYS | 299 | - .336 |

$^{**}p < .01$

**Results for SBS Pilot Study**

Presented in Table 9 below are the gender, ethnicity and age of the participants in the SBS portion of the validity/reliability pilot study.

Table 9

*Gender, Ethnicity and Age of 6th Grade SBS Pilot Study Participants by Number and Percent*

| Characteristic | | *n* | *%* |
|---|---|---|---|
| Gender | | | |
| | Female | 118 | 47.8 |
| | Male | 129 | 52.2 |
| Ethnicity | | | |
| | Black | 59 | 24.1 |
| | Hispanic | 58 | 22.9 |
| | White | 92 | 37.5 |
| | Asian | 9 | 3.7 |
| | Multi-racial | 29 | 11.8 |
| Age | | | |
| | 13 | 23 | 9.3 |
| | 12 | 135 | 54.7 |
| | 11 | 89 | 36.0 |
| Total | | 247 | 100.0 |

***Results for Reliability Estimation of the SBS Subscales***

Estimates of internal consistency (Cronbach's alpha) reliability for the various subscales of the SBS are presented in Table 10 below. Table 11 contains estimates of test-retest reliability for the SBS subscales. Because most of the validation measures were comprised of one item each, no attempt was made to estimate internal consistency reliability of the validation measures. As indicated in Table 10, all of the alpha reliabilities of the SBS subscales were above .80 and most were above .90. Table 11 shows all but one SBS subscale test-retest reliability estimate over .70, the exception being ATOD Intent.

Table 10

*Internal Consistency Estimates of Reliability of SBS Subscales for the Pilot Study*

| | Week One | | | Week Two | | |
|---|---|---|---|---|---|---|
| SBS Measure | *n* | No. Items | α | *n* | No. Items | α |
| Smoke Past 30 Days | 247 | 2 | .879 | 247 | 2 | .981 |
| Smoke Past Year | 247 | 2 | .940 | 247 | 2 | .983 |
| Drink Past 30 Days | 247 | 2 | .931 | 246 | 2 | .900 |
| Drink Past Year | 247 | 2 | .898 | 246 | 2 | .947 |
| Dope Past 30 Days | 247 | 2 | .948 | 247 | 2 | .961 |
| Dope Past Year | 246 | 2 | .998 | 247 | 2 | .971 |
| Intent ATOD Use | 245 | 3 | .848 | 247 | 3 | .870 |

Table 11

*Test-Retest (One Week Delay) Estimates of Reliability of SBS Subscales for the Pilot Study*

| SBS Measure | *n* | No. Items | $r_{xx}$ |
|---|---|---|---|
| Smoke Past 30 Days | 247 | 2 | .904 |
| Smoke Past Year | 247 | 2 | .880 |
| Drink Past 30 Days | 246 | 2 | .795 |
| Binge Drink | 240 | 1 | .706 |
| Drink Past Year | 246 | 2 | .877 |
| Dope Past 30 Days | 246 | 2 | .921 |
| Dope Past Year | 246 | 2 | .796 |
| Intent ATOD Use | 235 | 3 | .614 |

***Results for Concurrent Validity Estimation of the SBS Subscales***

The SBS subscales were correlated with their corresponding validation measures. The resulting concurrent validity coefficients are presented in Table 12. The high concurrent correlations shown in Table 12 between the SBS subscales and similar measures drawn from other widely used instruments give confidence that the SBS subscales are tapping into the same behavioral constructs as these corresponding measures.

Table 12

*Concurrent Validity Estimates of SBS Subscales with Validation Measures for the Pilot Study*

| Measures | *n* | *r* | *p* |
|---|---|---|---|
| Smoke Past 30 Days (SBS/ PRIDE RP) | 243 | .755 | .0001 |
| Smoke Past Year (SBS/ALERT) | 243 | .803 | .0001 |
| Smoke Past Year (SBS/PRIDE RP) | 243 | .813 | .0001 |
| Drink Past 30 Days (SBS/ALERT) | 238 | .877 | .0001 |
| Drink Past 30 Days (SBS/PRIDE RP) | 238 | .833 | .0001 |
| Binge Drink (SBS/ALERT) | 235 | .749 | .0001 |
| Drink Past Year (SBS/ALERT) | 238 | .893 | .0001 |
| Drink Past Year (SBS/PRIDE RP) | 238 | .852 | .0001 |
| Dope Past 30 Days (SBS/ALERT) | 242 | .873 | .0001 |
| Dope Past 30 Days (SBS/PRIDE RP) | 242 | .885 | .0001 |
| Dope Past Year (SBS/ALERT) | 242 | .851 | .0001 |
| Dope Past Year (SBS/PRIDE RP) | 242 | .704 | .0001 |
| Intent ATOD Use (SBS/ALERT) | 235 | .392 | .0001 |
| Intent ATOD Use (SBS/Tanglewood) | 235 | .650 | .0001 |

## Validity and Reliability Data from Main Study

### Internal Consistency Reliability Estimation of SBRPFS from Main Study

In addition to examining the reliability of the SBRPFS subscales in the separate study described above, we generated internal consistency estimates of reliability (Cronbach's alpha) for

the various subscales of the SBRPFS based on the main data base. Estimates were generated for both the post-survey and the 6-month follow-up and are presented in Table 13. As indicated in the table, all but one of the subscales shows levels of reliability above .75 at the two survey periods. Self-Efficacy, the one exception, has internal consistency estimates in the .60s.

Table 13

*Internal Consistency Estimates of Reliability of SBRPFS Subscales for Post-Survey and 6-Month Follow-Up*

| | Post-Survey | | 6-Month Follow-Up | |
|---|---|---|---|---|
| SBRPFS Measure | n | $\alpha$ | n | $\alpha$ |
| ATOD Use Outcomes | | | | |
| Smoke Past 30 Days | 10,513 | .909 | 10,163 | .927 |
| Smoke Past Year | 10,513 | .915 | 10,163 | .935 |
| Drink Past 30 Days | 10,513 | .882 | 10,163 | .891 |
| Drink Past Year | 10,513 | .885 | 10,163 | .900 |
| Dope Past 30 Days | 10,513 | .930 | 10,163 | .946 |
| Dope Past Year | 10,513 | .941 | 10,163 | .949 |
| R&P Outcomes | | | | |
| Intent ATOD Use | 10,499 | .825 | 10,155 | .810 |
| Goals & Decisions | 10,462 | .782 | 10,109 | .789 |
| Bond Prosocial Peers | 10,440 | .840 | 10,098 | .855 |
| Peer Resistance | 10,407 | .843 | 10,092 | .856 |
| Self-Efficacy | 10,346 | .656 | 10,020 | .684 |
| Harmful Effects | 10,287 | .942 | 10,023 | .935 |
| ATOD Attitudes | 10,230 | .757 | 9,933 | .781 |

## Principal Components Analysis of R&P Items from Main Study

To investigate the internal factor structure of the R&P items together with the three Intent ATOD Use items of the SBRPFS, we performed a factor extraction using the Principal Components Analysis (PCA) procedure from the SAS system's PROC FACTOR. Students'

post-survey responses from all 40 participating schools were used as the data source, representing an N of 10,009. Only student response sheets with no missing responses were included in the analysis. Nine factors produced eigenvalues above 1.0; these eigenvalues were 11.820, 3.106, 2.346, 1.723, 1.368, 1.219, 1.160, 1.032, and 1.011, respectively. However, an examination of the scree plot showed a flat array of eigenvalues after factor 7. Therefore, the first seven factors were retained for further analysis. Cumulatively, these seven factors accounted for approximately 60% of the total variance. In order to sharpen the distinction among the factors, a promax rotation was applied, yielding factor loadings for the 39 variables on the seven factors as shown in Table 14. In displaying the results, we included the factor loadings of variables only if those loadings were .40 or above, following Stevens' (1996) suggestion that factor loadings of .40 or higher connote practical significance.

The factor loadings in Table 14 tend to be distributed in patterns that provide a basis for interpreting most of the seven factors. The factor loadings on factor 1 cluster cleanly on variables a25 to a30, variables representing the six items on the Harmful Effects subscale; no other variables load on factor 1. The factor loadings on factor 6 cluster cleanly on variables a31 to a35, variables representing five of the six items on the ATOD Attitudes subscale; no other variables load on factor 6. The factor loadings on factor 7 cluster cleanly on variables b7, b15 and b22, variables representing the three items on the ATOD Intent subscale; no other variables load on factor 7. The factor loadings on factor 2 cluster primarily on variables a13, a15, a16, a17 and a18, variables representing five of the six items on the Peer Resistance subscale; variable a20, representing an item on the Self-Efficacy subscale, is the only other variable loading on factor 2. The factor loadings on factor 4 cluster on variables a7 to a12, variables representing the six items on the Bonding with Prosocial Peers subscale; no other variables load on factor 4.

However, three of these six variables also have loadings on factor 5. Five of the six variables representing items on the Goals and Decisions subscale (variables a1, a2, a3, a5 and a6) load on factor 3; but variable a4, representing the remaining Goals and Decisions item, loads on factor 5. Also, variables a21 and a23 (representing items on the Self-Efficacy subscale) load on factor 3. Factor 5 appears to be the least interpretable of the seven factors, picking up factor loadings on variables a4 (from the Goals and Decisions subscale); a7, a9 and a11 (from the Bonding with Prosocial Peers subscale); and a19 and a22 (from the Self-Efficacy subscale).

Overall, the PCA analysis provides encouraging results for the behavior of most of the 39 survey items. Factorially, the items on the Harmful Effects subscale (a25 to a30) load together, as do the items on the ATOD Intent subscale (b7,b15 and b22). On both the Peer Resistance subscale (a3 to a18) and the ATOD Attitudes subscale (a31 to a36), all but one of the six items load together. Items on the Bonding with Prosocial Peers subscale (a7 to a12) load together, but three of the six items show a cross-loading effect with Factor 5 the Self-Efficacy subscale. On the Goals and Decisions subscale (a1 to a6), five of the six items load together; only one of these six showed a loading on an additional factor. Although 6 of the 7 factors provided fairly simple solutions, the six items of the Self-Efficacy subscale behaved in a manner that provides the least clarity. In the PCA analysis the variables representing self-efficacy items (a19 to a24) had loadings scattered across factors 2, 3, and 5, and one that showed no loading on any of the seven factors. Self-efficacy as measured in the current study appears to be influenced, in part, by participants' perceptions of their skills in goal setting and decision making, peer resistance, and their bonding with prosocial peers.

Table 14

*Factor Loadings for PCA 7-Factor Rotated Factor Pattern (N = 10,009)*

| Variable | Harmful Effects Factor 1 | Peer Resistance Factor 2 | Goals & Decisions Factor 3 | Prosocial Peers Factor 4 | Self-Efficacy Factor 5 | ATOD Attitudes Factor 6 | Intent ATOD Factor 7 |
|---|---|---|---|---|---|---|---|
| b7 | | | | | | | 0.85 |
| b15 | | | | | | | 0.67 |
| b22 | | | | | | | 0.86 |
| a1 | | | 0.57 | | | | |
| a2 | | | 0.73 | | | | |
| a3 | | | 0.69 | | | | |
| a4 | | | | | 0.55 | | |
| a5 | | | 0.68 | | | | |
| a6 | | | 0.69 | | | | |
| a7 | | | | 0.48 | 0.50 | | |
| a8 | | | | 0.82 | | | |
| a9 | | | | 0.51 | 0.40 | | |
| a10 | | | | 0.80 | | | |
| a11 | | | | 0.48 | 0.50 | | |
| a12 | | | | 0.81 | | | |
| a13 | | 0.53 | | | | | |
| a14 | | | | | | | |
| a15 | | 0.58 | | | | | |
| a16 | | 0.67 | | | | | |
| a17 | | 0.73 | | | | | |
| a18 | | 0.73 | | | | | |
| a19 | | | | | 0.58 | | |
| a20 | | 0.56 | | | | | |
| a21 | | | 0.45 | | | | |
| a22 | | | | | 0.64 | | |
| a23 | | | 0.53 | | | | |
| a24 | | | | | | | |
| a25 | 0.85 | | | | | | |
| a26 | 0.91 | | | | | | |
| a27 | 0.80 | | | | | | |
| a28 | 0.89 | | | | | | |
| a29 | 0.88 | | | | | | |
| a30 | 0.88 | | | | | | |
| a31 | | | | | | 0.58 | |
| a32 | | | | | | 0.42 | |
| a33 | | | | | | 0.81 | |
| a34 | | | | | | 0.58 | |
| a35 | | | | | | 0.41 | |
| a36 | | | | | | | |

*Note*. All factor loadings shown are positive. Factor loadings below 0.40 are not shown.

## Data Analysis Procedures for the Study

### Data Analysis Procedures for Study Demographics

Descriptive statistical analyses were performed on the demographic data provided by the survey (i.e., student ethnicity, gender, age) and data provided by the district (i.e., free and reduced lunch program services, limited English proficiency services, exceptional education services) in order to provide a description of the participating students in the treatment and control schools. Data provided by the district contained unique identification codes, permitting individual student matching while allowing students to remain strictly anonymous. In addition to describing the demographics of the study sample, we tested for potential differences between the treatment/control conditions using the *t*-test procedure to see 1) if the two groups were equivalent before the delivery of the TGFD program, and 2) if the demographics for the treatment and control schools remained similar across time (post-survey & 6-month follow-up).

### Data Analysis Procedures for Attrition and Missing Data

#### Survey Data Cleanup

The first important step in survey data cleanup was instituted at the actual point of data collection. During the classroom administration of the SBRPFS, the Survey Administrators (SAs) were directed to watch for students who were persistently and overtly off-task, who started and then abruptly quit taking the survey, or who finished in five minutes or less. Because the surveyed classes were small in size (seldom more than 20 students), the SAs were able to carry out this task with little difficulty. The SAs discretely turned down a corner of the answer sheet from any such student as the answer sheets were being collected. The SAs then tagged these answer sheets and kept the tagged answer sheets separate from the other student answer sheets.

The evaluation team reviewed the individually tagged answer sheets and the SAs' written reasons for tagging any forms. If we agreed with the SAs' decision to tag an answer sheet, it was excluded from the major analyses.

Following each of the three survey administrations, a team was brought in to check and clean-up the answer sheets in preparation for scanning such that all 1) selected responses were completely filled in, 2) old answers were thoroughly erased, and 3) stray marks were removed.

Descriptive and inferential analyses were performed on the number of answer sheets excluded from the study sample due to student off-task or unattending behavior; comparisons were made by treatment and control conditions for each of the three administration periods. Also, an attempt was made to gauge student attrition in the treatment and control schools over the several assessment periods; only rough estimates were possible because of the anonymity of student responses.

**Procedures for Handling Missing Data**

Decision rules for handling missing data were set prior to implementing major analyses on the survey data. No attempt was made to impute missing values for any subscale score if that subscale had 20% or more of the responses missing. In practice that meant that no values were imputed for missing responses for subscale scores based on less than six items. If a student's score on a 2-, 3-, 4- or 5-item subscale contained one or more missing values, the student's score on that particular subscale was excluded from the analyses. For 6-item subscales (e.g., goals and decision making) that contained one missing response, a missing value was imputed. In such cases, the decision was to impute missing values based on the student's mean response to the remaining five items (Steven, 1996). If more than one response was missing on a 6-item subscale, the student's score on that particular subscale was excluded from the analyses.

Descriptive and inferential statistical analyses were performed on the number of survey responses that were excluded from the analyses due to missing data that did not meet the decision rules for imputing missing values by the treatment and control conditions for each of the three administration periods.

**Data Analysis Procedures for the SBRPFS**

The data analyses on the SBRPFS focused on cross comparisons of the student responses from the treatment and control schools at each of the three assessment points in the evaluation: pre-survey, post-survey, and 6-month follow-up survey. Pre-survey responses were examined to determine if the treatment and control students were essentially equivalent on the major outcomes. Subsequent comparative analyses examined for differences between the treatment and control responses for the post-survey and for the 6-month follow-up survey. The focus of the analyses was on the major outcomes of the TGFD program: 1) student ATOD usage outcomes, including frequency of use of cigarettes, alcohol and marijuana over the past month and over the past year, and 2) R&P outcomes, including intentions to use drugs in the future, goal setting and decision-making skills, social and peer resistance skills, emotional competency/self-efficacy skills, bonding with prosocial peers, perceptions of harmful effects of drugs, and attitudes toward drugs. A two-level restricted maximum likelihood model using the MIXED PROCEDURE in SAS 9.2 was employed to examine for differences in treatment and control responses at each of the three time periods.

**Data Analysis Procedures for School Performance Indicators**

Four school performance indicators were examined for the 6$^{th}$ grade study—the number of days absent, the number of student suspensions assigned (in- and out-of-school), and reading and mathematics test performance. Mean school scores for the treatment/control conditions for

the number of days absent and the number of suspensions were computed and then examined using the *t*-test procedure. A two-level restricted maximum likelihood model using the MIXED PROCEDURE in SAS was employed to examine for differences in treatment and control school students' 6th grade reading and mathematics performance on the Florida Comprehensive Assessment Test (FCAT) as a function of prior achievement.

## Results on Fidelity of Program Implementation

Instruments designed for use in ascertaining the fidelity of the program interventions and assessments, and in monitoring within-school conditions, served both *formative* and *summative* functions. The data from these instruments were useful formatively as a basis for modifications or adjustments early on in the treatment process and the evaluation process. They were also useful as a basis for summative judgments on the completeness and quality of the treatment delivery. Additionally, they served to allay concerns about conditions within and across schools that could contaminate the treatment-control comparisons and impact the interpretation of the survey results.

From a formative perspective, observations from the *Survey Administration Protocol* (SAP) and the *Treatment Observation Protocols* (TOPs) were examined as early in the evaluation process as feasible in order to spot potential problems. Immediately after the administration of the pre-survey, we were in contact with all observation team members who observed using the SAP to determine if problems were spotted during the administration of the SBRPFS, and if so, whether further training of the team's SAs on survey delivery was warranted. Similarly, the data from early observations of treatment lessons using the TOPs were examined to determine if any adjustments were needed relative to the TGFD instructors' delivery of the treatment.

From a summative perspective, the TOP observations and responses to the *Lesson Implementation Checklist* (LIC) and *Classroom Teacher Survey* (CTS) were used to provide a data-based picture of the quality of the 10-week treatment delivery. The three sources—classroom observations, 6th grade teachers' ratings of the treatment delivery, and ratings by the TGFD instructors themselves—provided a triangulated basis for gauging the completeness and appropriateness of the intervention. Descriptive statistics are used to report the results from these three sources. Descriptive statistics are also used to report the results from the observations on the SAP regarding the integrity of the SBRPFS data collection, and from the *Within School Activities Form* (WSAF) regarding the presence or absence of possibly contaminating school-based or grade-based activities. Treatment and control schools are compared on both the SAP and WSAF responses.

**Results on the Survey Adminstration Protocol (SAP)**

The *Survey Administration Protocol* (SAP) was used to make a minimum of two unannounced observations of each of the trained Survey Administrators (SAs) during the three two-week periods of survey data collection: pre-survey, post-survey, and 6-month follow-up. During these visits, each of the 16 observational items on the SAP was marked Demonstrated Fully (3), Demonstrated Partially (2) or Not Demonstrated (1).

On the pre-survey, results based on 50 classroom observations using the SAP (divided equally between treatment and control classrooms) showed that 87% of the items were fully demonstrated (86% treatment, 87% control), 13% partially demonstrated (14% treatment, 13% control), and 0% not demonstrated. The typical discrepancy noted by observers was a minor thing, such as the SA failing to remind students to raise their hand if they had received a pencil with a broken pencil point. As to whether all or almost all participating students had time to

attempt all questions on the survey, all 50 observations recorded an affirmative response. As to whether there was anything observed during the survey period that might have compromised the integrity of the students' responses, all 50 observations recorded a negative response.

On the post-survey, results based on 64 classroom observations (divided equally between treatment and control classrooms) showed that 91% of the items were fully demonstrated (92% treatment, 90% control), 9% partially demonstrated (8% treatment, 10% control), and 0% not demonstrated. As to whether all or almost all students had time to attempt all questions, 97% of the 64 observations recorded an affirmative response. As to whether there was anything observed during the survey period that might have compromised the integrity of the students' response, all 64 observations recorded a negative response.

On the 6-month follow-up, results based on 66 classroom observations (divided equally between treatment and control classrooms) showed that 91% of the items were fully demonstrated (90% treatment, 92% control), 8% partially demonstrated (9% treatment, 7% control), and 1% not demonstrated. All 66 of the observations recorded an affirmative response as to whether all or almost all students had time to attempt all questions. All of the 66 observations recorded a negative response as to whether there was anything observed during the survey period that might have compromised the integrity of the students' responses.

**Percent of Inter-Observer Agreement for the SAP**

Paired observations using the SAP were conducted during the post-survey and 6-month follow-up phase of survey administration, with about half being in treatment schools and half being in control schools. These paired observations were used to compute the percent of inter-observer agreement for the 16 items on the SAP. High inter-observer agreement gives

confidence that the observers of survey data collection are consistent in their observations, i.e., that they are seeing the same thing and marking it in the same way.

Using procedures similar to those used by Metze (2000), item inter-observer agreement was represented by computing the percent of exact agreement for the 16 observational items on each paired observation, and then computing the overall mean percent across all paired observations. Because each item on the SAP could be marked Demonstrated Fully (3), Demonstrated Partially (2) or Not Demonstrated (1), two observers could have exact agreement on an individual item (3 and 3, 2 and 2, or 1 and 1), could differ by one position (e.g., 3 and 2, 2 and 1), or could differ by two positions (e.g., 3 and 1). Ideally, the percent of exact agreement across all items for all paired observations would equal 100.

For the post-survey, the overall percent of exact agreement on the SAP items for 14 paired observations was 95%; only 5% of the items differed by one position (e.g., 3 and 2) and none differed by two positions. For the 6-month follow-up, the overall percent of exact agreement on the items for 15 paired observations was 97%; only 3% of the items differed by one position and none differed by two positions.

## Fidelity of Treatment Delivery

### Results on the Lesson Implementation Checklists (LICs)

The completeness of implementation of the 10 TGFD lessons was determined in part through the TGFD instructors' responses to the Lesson Implementation Checklists (LICs). On the LICs the lesson content was broken into several activity statements of 100 words or less, with each statement representing a coherent segment of lesson activity. Because the content and the expected instructor and student behaviors varied across the 10 lessons, the number of LIC activity statements also varied across lessons, ranging from 6 to 12 statements. After a class

lesson was delivered, the TGFD instructor marked each activity statement on the LIC for that lesson in terms of whether all, most, some, or none of the activity statement was completed.

The term "marked" is used liberally, as some TGFD instructors immediately following the delivery of a lesson would use their cell phones or PDAs to access their online calendars to link to a particular class's LIC and key in their information. Other instructors would mark their responses on specially provided paper forms for a given lesson representing up to seven class periods and at the end of the day, use that specific form to guide them in entering their information online through the OCMS. Across the 10-week treatment period, the TGFD instructors completed a total of 2,836 LICs, representing the degree of implementation for every lesson delivered and class served.

Each lesson, with the exception of lesson 9, was given to approximately 285 classes across the 20 treatment schools; slight variations in the number of classes occurred during the first four lessons due to class size adjustments being made during the first month of the school year. Due to schedule changes made by the schools, the classroom teachers or the TGFD instructors, 17 classes across the first eight lessons were delayed one-week and could not be rescheduled. In order to get everyone back on track for the final lesson, we decided to have the TGFD instructors make up these 17 class sessions by skipping lesson 9. Therefore, 94% of the treatment students received all 10 lessons, and 6% received 9 of the 10 lessons. The TGFD teachers' implementation responses for the 10 lessons are summarized in Table 15. Because all lessons were considered of equal importance—regardless of the number of activity statements needed to represent them—the averages across the 10 lessons are unweighted means.

As indicated in Table 15, the TGFD instructors reported completing on average all of the activity for 90.1% of the activity statements across the 10 lessons, and most of the activity for

6.8% of the activity statements. Together, these values represent a high rate of lesson implementation. No activity completed was reported for only 1.1% of the activity statements. Across the 10 lessons, the implementation percentages range from a low of 94.4% completed/mostly completed for lesson 5 to a high of 99.3% completed/mostly completed for lesson 4, suggesting a high degree of consistency in the instructional delivery process throughout the 10-week treatment.

Table 15

*Percent Completion of Lesson Activity Statements on TGFD Instructors' Lesson Implementation Checklists (LICs) Across Treatment Classrooms by Lesson*

| Lesson | No. of Classes | % All of the Activity Completed | % Most of the Activity Completed | % Some of the Activity Completed | % None of the Activity Completed |
|---|---|---|---|---|---|
| Lesson 1 | 285 | 89.7 | 7.0 | 2.6 | 0.7 |
| Lesson 2 | 284 | 89.7 | 7.5 | 1.7 | 1.1 |
| Lesson 3 | 284 | 85.1 | 8.0 | 2.8 | 4.1 |
| Lesson 4 | 285 | 96.4 | 2.9 | 0.6 | 0.1 |
| Lesson 5 | 286 | 79.4 | 15.0 | 4.2 | 1.5 |
| Lesson 6 | 286 | 95.3 | 3.4 | 1.0 | 0.2 |
| Lesson 7 | 286 | 89.6 | 7.1 | 2.4 | 0.9 |
| Lesson 8 | 286 | 91.9 | 5.2 | 2.5 | 0.4 |
| Lesson 9 | 269[a] | 92.6 | 5.2 | 1.5 | 0.8 |
| Lesson 10 | 285[b] | 90.8 | 6.8 | 1.5 | 1.5 |
| Mean | | 90.1 | 6.8 | 2.1 | 1.1 |

a = 17 classes not taught (5.94%)
b = 1 class not taught (0.03%)

For any LIC statement that was marked most, some or none of the activity completed in a lesson, the TGFD instructors checked from a printed list of reasons to explain the incompletion. More than one reason could be selected. A summary of these are presented in order of frequency in Table 16. As indicated in Table 16, the most frequently cited reason was "Some TGFD lesson activities took more time to implement than planned"; this explanation was checked on 44.4% of

the LICs with marks denoting less than complete activity statements. Other reasons checked with high frequency were "Student disruptive behaviors made classroom management challenging" (37.7%); "Student questions or comments delayed start of other lesson activities" (37.1%); and "Class house-keeping, attendance, and bell work delayed start of TGFD lesson" (30.2%). On 21.9% of the LICs the category "Other reasons" was checked, followed by a written explanation composed by the TGFD teacher. Most of these "other" explanations were notes by the instructors that they had inadvertently forgotten to include some small element of the lesson (e.g., " I forgot to ask about emotions", "I inadvertently skipped one part of that section") or that the exclusion was deliberate (e.g., "Skipped … so I could get to other parts of the lesson").

Table 16

*Reasons by TGFD Instructors for Lesson Activities Not Fully Implemented, Expressed as Percent of Incomplete Activity Statements Checked Across Treatment Classrooms*

| Reasons Lesson Activities were Not Fully Implemented | % |
|---|---|
| Some TGFD lesson activities took more time to implement than planned | 44.4 |
| Student disruptive behaviors made classroom management challenging | 37.7 |
| Student questions or comments delayed start of other lesson activities | 37.0 |
| Class house-keeping, attendance, and bell work delayed start of TGFD lesson | 30.2 |
| Other reasons | 21.9 |
| Class started late (students arrived late from prior class) | 9.7 |
| Students distracted by an incident in class (e.g., illness, fight) | 8.5 |
| Large class made distributing materials and grouping time consuming | 8.3 |
| TGFD lesson interrupted or shortened due to scheduled events (e.g., assembly) | 7.4 |
| TGFD lesson interrupted or shortened due to unscheduled events (e.g., fire drill) | 5.4 |
| Class with ESOL students resulting in slower communications or barriers | 4.2 |
| Students distracted by external events (e.g., hallway disruption, outside activity) | 4.0 |
| Class with ESOL students requiring lesson delivery in language other than English | 1.7 |

*Note*. More than one reason could be selected for any given class. ESOL = English for Speakers of Other Languages.

The TGFD instructors were also asked to rate the students in each TGFD class in regard to their response to the lesson. The students of each class session as a group were rated on six aspects of their behavior: 1) on-task behavior, 2) willingness to be active participants, 3) willingness to be respectful of others in the class, 4) willingness to follow the rules for lesson participation, 5) willingness to express themselves in relevant ways, and 6) responsiveness to the various lesson activities. The TGFD instructors rated each of the approximately 2,850 TGFD class sessions on a scale from 1 (poor) to 4 (excellent). The ratings across the six statements were highly homogeneous and highly positive, ranging from 3.61 to 3.74, with an overall mean of 3.66. These ratings suggest that the TGFD instructors perceived that the lessons were well received by the students and that the students were active participants in the lessons.

**Results on the Treatment Observation Protocols (TOPs)**

Completion of the LICs by the TGFD instructors provided important evidence of program implementation from the perspective of those responsible for delivering the instruction. In addition, evidence of program implementation came from a representative sample of in-class observations of instructional delivery conducted by the observation team. A total of 74 observations of TGFD classrooms were conducted on an unannounced basis using the TOP. Observations were conducted on all 10 lessons, though scheduling constraints led to more observations on some lessons than others. The activity statements on the TOP for a given lesson were duplicates of the activity statements on the corresponding LIC for that same lesson and the 4-point scale (All of the Activity Completed to None of the Activity Completed) was the same. The number of activity statements on the TOPs varied from 6 to 12, depending upon the lesson. A summary of results of the observations on the activity statements across the lessons is

presented in Table 17. As with the LIC data, the averages across the 10 lessons are un-weighted means.

As indicated in Table 17, the observations show on average all of the activity completed for 82.9% of the activity statements, and most of the activity completed for 14.7% of the activity statements. Together, these values represent a high rate of lesson implementation observed in the sample of TGFD classrooms. No activity completed was reported for less than 1% of the activity statements. Across the 10 lessons, the implementation percentages range from a low of 90.0% completed/mostly completed for lesson 3 to a high of 100% completed/mostly completed for lessons 1, 4, 6, 7 and 10, suggesting a high degree of consistency in the instructional delivery of the observed lessons. These results by independent observation mirror very closely the results provided by the instructors themselves, and thereby reinforce the assessment that the delivery of the TGFD treatment was thorough and complete.

Table 17

*Percent Completion of Lesson Activity Statements on Observers' Treatment Observation Protocols for a Sample of Treatment Classrooms by Lesson*

| Lesson | No. of Classes | % All of an Activity Completed | % Most of an Activity Completed | % Some of an Activity Completed | % None of an Activity Completed |
|---|---|---|---|---|---|
| Lesson 1 | 6 | 86.1 | 13.9 | 0.0 | 0.0 |
| Lesson 2 | 8 | 89.3 | 8.9 | 1.8 | 0.0 |
| Lesson 3 | 10 | 63.6 | 26.4 | 6.4 | 3.6 |
| Lesson 4 | 10 | 71.4 | 27.9 | 0.0 | 0.0 |
| Lesson 5 | 5 | 85.0 | 10.0 | 5.0 | 0.0 |
| Lesson 6 | 8 | 81.9 | 18.1 | 0.0 | 0.0 |
| Lesson 7 | 2 | 94.4 | 5.6 | 0.0 | 0.0 |
| Lesson 8 | 9 | 82.1 | 12.5 | 4.2 | 1.4 |
| Lesson 9 | 8 | 87.5 | 11.4 | 1.1 | 0.0 |
| Lesson 10 | 8 | 87.5 | 12.5 | 0.0 | 0.0 |
| Mean | 74 | 82.9 | 14.7 | 1.8 | 0.5 |

In regard to activity statements receiving a mark other than fully completed, the observers—like the TGFD instructors—were directed to check from a printed list of reasons to explain the incompletion. More than one reason could be selected. These results in order of frequency are presented in Table 18. As shown in the table, the general category "Other reasons" was checked most frequently by the observers, relating to fully 69.0 % of the incomplete activity statements. The explanatory information supplied by the observers under "Other reasons" typically dealt with details in the lesson that appeared to be inadvertently skipped or omitted by the TGFD instructor. Among the pre-printed reasons listed on the TOP form, the three most frequently cited were "Student disruptive behaviors made classroom management challenging" (19.0%), "Some TGFD lesson activities took more time to implement than planned" (14.0%), and "Student questions or comments delayed start of other lesson activities" (9.2%). These three categories were also among those most frequently cited by the instructors themselves.

Table 18

*Reasons by Observers for Lesson Activities Not Fully Implemented, Expressed as Percent of Incomplete Activity Statements Checked Across Observed Treatment Classrooms*

| Reasons Lesson Activities were Not Fully Implemented | % |
|---|---|
| Other reasons | 69.0 |
| Student disruptive behaviors made classroom management challenging | 19.0 |
| Some TGFD lesson activities took more time to implement than planned | 14.0 |
| Student questions or comments delayed start of other lesson activities | 9.2 |
| Class house-keeping, attendance, and bell work delayed start of TGFD lesson | 4.2 |
| Large class made distributing materials and grouping time consuming | 2.9 |
| Students distracted by an incident in class (e.g., illness, fight) | 1.7 |
| Class started late (students arrived late from prior class) | 1.4 |
| Class with ESOL students resulting in slower communications or barriers | 0.0 |
| Class with ESOL students requiring lesson delivery in language other than English | 0.0 |
| TGFD lesson interrupted or shortened due to scheduled events (e.g., assembly) | 0.0 |
| TGFD lesson interrupted or shortened due to unscheduled events (e.g., fire drill) | 0.0 |
| Students distracted by external events (e.g., hallway disruption, outside activity) | 0.0 |

*Note*. More than one reason could be selected for any given class.

The observation team was also asked to rate the students as a group in regard to their response to the lessons that were being observed. The students were rated on the same six items as those rated by the TGFD instructors; i.e., on-task behavior; willingness to be active participants, to be respectful of others, to follow the rules for lesson participation, and to express themselves in relevant ways; and responsiveness to the various lesson activities and to the instructor. The same scale from 1 (Poor) to 4 (Excellent) was used. The ratings across the six statements were highly homogeneous and highly positive, ranging from 3.41 to 3.65, with an overall mean of 3.55. These ratings mirror the ratings by the instructors themselves, and suggest that the observers perceived that the observed lessons were well received by the students.

Toward the end of their observation, the observers rated the TGFD instructors with respect to the demonstration of 12 aspects of instructional behavior. The observers used a 4-point scale: 1 (Not Demonstrated), 2 (Demonstrated Rarely), 3 (Demonstrated Occasionally) and 4 (Demonstrated Consistently). A summary of the observers' ratings is presented in Table 19. As is apparent from the mean ratings in the table, the TGFD instructors were rated very highly by the observers in terms of various critical instructional behaviors, including "was prepared for instruction" (3.96), "actively engaged students in the learning process" (3.82), "provided students with opportunities to participate in class discussions" (3.88), "provided students opportunities to practice lesson skills" (3.91), and "recognized and reinforced students for participating" (3.93). The mean ratings on the individual items and the overall mean of 3.90 suggest that the observers saw competent instruction being demonstrated in the observed classrooms and students actively engaged in learning.

Table 19

*Means and Standard Deviations of Observers' Ratings of TGFD Instructors Across 74 Observed Treatment Classrooms*

| Item | *M* | *SD* |
|---|---|---|
| *The TGFD instructor…* | | |
| was prepared for instruction. | 3.96 | 0.20 |
| used appropriate classroom management strategies. | 3.85 | 0.36 |
| kept students on task. | 3.81 | 0.39 |
| actively engaged students in the learning process. | 3.82 | 0.45 |
| effectively transitioned between lesson activities. | 3.89 | 0.31 |
| provided clear directions for each activity. | 3.95 | 0.23 |
| defined terms, provided explanations and/or gave examples. | 3.97 | 0.16 |
| displayed or modeled lesson skills. | 3.92 | 0.27 |
| provided students with opportunities to participate in class discussions. | 3.88 | 0.33 |
| provided students opportunities to practice lesson skills. | 3.91 | 0.34 |
| recognized and reinforced students for participating. | 3.93 | 0.25 |
| responded to student input in a receptive and supportive manner. | 3.89 | 0.31 |
| Survey Mean | 3.90 | 0.31 |

*Note.* Maximum score = 4.00.

### ***Percent of Inter-Observer Agreement for the TOP***

Twenty-four of the 74 observations of treatment delivery made with the TOP were paired observations. These paired observations were used to compute the percent of inter-observer agreement on the activity statements of the TOP. High inter-observer agreement gives confidence that the observers of lesson delivery are consistent in their observations, i.e., are seeing the same thing and marking it in the same way.

Inter-observer agreement for the TOP was estimated using the same general procedure as used for the SAP (Metze, 2000). However, the TOP varied in the number and content of its activity statements across the 10 lessons of the intervention. The number of activity statements

on the TOP could vary from 6 to 12, depending upon the particular lesson, with each statement using a 4-point scale (All, Most, Some, or None of the Activity implemented). Because all of the 10 lessons were considered equal in importance regardless of the number of statements needed to represent them, the mean percent of exact agreement computed for each lesson-specific TOP was given equal weight in the computation of an overall percent of exact agreement for the TOP across the 10 lessons. In addition to the percent of exact agreement, the percent of minor disagreement (resulting from paired observations that differed by one position) and major disagreement (resulting from paired observations that differed by two or three positions) were also computed.

Across the 24 paired treatment observations, the overall percent of exact agreement on the activity statements across the 10 lessons was 90.8%; minor disagreements (one score point difference) between paired observations was 8.8%; and major disagreements (two or three score point difference) between paired observations was 0.4%. These results suggest that the observers were operating at a high degree of consistency in their observations of treatment delivery, thereby giving credence to the results summarized in Tables 17 and 18 above.

**Results on the Classroom Teacher Survey (CTS)**

Using the OCMS, electronic invitations to complete the CTS and accompanying survey links were sent by the evaluation team to 72 teachers whose 286 classrooms participated in the treatment program. The CTS permitted the classroom teachers to rate the delivery of the TGFD lessons and their students' response to the lessons at the end of the 10-week treatment. For co-teaching classes, the classroom teacher and the FUSE or collaborative teacher were invited to share their observations. Follow-up emails were sent to encourage survey completion, and in a few cases, paper copies of the CTS were provided by facsimile and mail with stamped, pre-

addressed return envelopes. Out of 72 teachers contacted, 71 (98.6%) completed the CTS. The teachers responded on a 5-point Likert scale ranging from Strongly Disagree (1) to Strongly Agree (5). A summary of the classroom teachers' responses to the CTS is presented in Table 20.

Table 20

*Means and Standard Deviations of Ratings on Classroom Teacher Survey (N =71)*

| Item | *M* | *SD* |
|---|---|---|
| *The TGFD instructor…* | | |
| was well prepared for each class. | 4.90 | 0.38 |
| applied appropriate classroom management techniques. | 4.70 | 0.70 |
| effectively transitioned between lesson activities. | 4.89 | 0.32 |
| kept the students on task. | 4.83 | 0.48 |
| gave clear directions to the students. | 4.89 | 0.43 |
| defined terms and concepts clearly. | 4.90 | 0.42 |
| provided clear explanations and appropriate examples. | 4.92 | 0.41 |
| provided the students with opportunities to participate in class. | 4.90 | 0.34 |
| provided the students with opportunities to practice the lesson skills. | 4.92 | 0.28 |
| recognized and reinforced students for participating in class. | 4.86 | 0.49 |
| responded to student input in a receptive and supportive manner. | 4.87 | 0.38 |
| bonded (developed satisfactory rapport) with the students. | 4.75 | 0.63 |
| demonstrated or modeled healthy behavior and positive choices. | 4.90 | 0.30 |
| appeared well informed in the drug prevention field. | 4.93 | 0.39 |
| *My students' comments and actions suggest…* | | |
| they enjoyed the TGFD program. | 4.82 | 0.49 |
| they felt the TGFD lessons were relevant to their lives. | 4.76 | 0.60 |
| the TGFD program had a positive impact on their behaviors and choices. | 4.77 | 0.59 |
| Survey Mean | 4.85 | 0.47 |

*Note.* Maximum score = 5.00.

As indicated in Table 20 above, the TGFD instructors were rated quite highly by the classroom teachers in the treatment schools. Particularly noteworthy were the high ratings given

the TGFD instructors for being prepared for class (4.90) and being well informed in the drug prevention field (4.93), and for providing students with opportunities to participate in class (4.90) and to practice the lesson skills (4.92). The classroom teachers' ratings of the students' response to the TGFD program were also quite positive, specifically that their students' comments and actions suggested that the students enjoyed the TGFD program (4.82), that the students felt the TGFD lessons were relevant to their lives (4.76), and that the TGFD program had a positive impact on the students' behaviors and choices (4.77).

**Summary of Data on Fidelity of Treatment Delivery**

Century et al. (2010) refers to fidelity of implementation as the "black box" in the evaluation of intervention effectiveness, and Brandon et al. (2008) notes that evaluation theorists, methodologists and practitioners are increasingly asking for evidence of program implementation in the conduct of program evaluations. Recognizing the criticality of lesson implementation to the meaningful evaluation of school-based drug intervention programs, we endeavored to include several checks on the quality of implementation in the delivery of the TGFD program. The TGFD instructors provided their perspective on implementation for every treatment session delivered across the 20 treatment schools, trained observers visited on an unannounced basis a representative sample of treatment classes, and the regular classroom teachers of participating students completed end-of-treatment ratings on the quality of TGFD lesson delivery. The triangulating data from these three sources provide collaborative evidence that the TGFD treatment was delivered as intended, was delivered with consistent quality and completeness, and in such a way that actively engaged the participating students. The insights derived from the implementation data described above provide a firm underpinning in interpreting the results on the major outcomes of the TGFD evaluation study.

**Results on the Within School Activities Form (WSAF)**

Each of the treatment and control school principals assigned a 6th grade School Contact person to complete the WSAF on a semester basis to help the evaluation team gather information concerning any potentially confounding influences which may have occurred during the study. The majority of the assigned school contact persons were classroom teachers (32), six were assistant principals, and two were student intervention specialists. Information regarding school or grade level activities was entered through the OCMS with a survey link sent to each contact person at the end of the first and second semester. Information entered by the school contacts at the mid-point of the year could be adjusted, modified or added to in the subsequent survey in an effort to reduce redundancy and professionals' time.

None of the contact persons reported their school having a drug prevention program implemented with 6th graders, with the exception of the TGFD program in the 20 treatment schools. Some academic course content related to drugs was part of all of the schools' science and physical education/health curricula. Drug-related topics were discussed in a few instances through presentations offered by the school security officer or guidance counselor, and in some instances during assembly. Across both treatment and control schools, the main focus for most of the classroom presentations or assemblies was related to reducing bullying, cyber bullying, gangs and violence; and increasing personal safety and gun safety. These presentations and assemblies were found to be fairly equally distributed between the treatment/control conditions.

**Results on the Student Behavior and Risk and Protective Factor Survey**

The potential attenuation of the distinction between treatment and control conditions due to the physical movement of students during the school year from one district to another, or one school to another, is always a concern when conducting school-based evaluations of

interventions across time. Just as troublesome is the loss of students during survey administration due to their choice to opt out before starting the survey as well as their more subtle decision to opt out after starting the survey (e.g., not reading the survey questions, off-task behavior). In the present study we endeavored to monitor the effect of such factors on the makeup of the study sample, and followed decision rules for minimizing their impact on treatment and control data.

## Attrition and Missing Data Over Time

### Attrition and Missing Data at the Pre-Survey

Administration of the pre-survey resulted in 10,968 scannable or usable answer sheets. A total of 147 answer sheets were tagged by the SAs during the administration of the pre-survey and later approved for exclusion from the analyses by the evaluation team (147 out of 11,115 or 1.3% attrition). The number of tagged answer sheets excluded from the data analyses was 73 for the 20 treatment schools ($M = 3.65$, $SD = 3.12$) and 74 for the 20 control schools ($M = 3.70$, $SD = 2.87$). We found no significant difference between the average number of tagged answer sheets excluded from the sample for treatment and control schools ($t(38) = -0.05$, $p = .9582$) on the pre-survey.

Following a priori decision rules, no values were imputed for survey subscales missing more than 20% of the responses. Two-item ATOD subscales for 30 day and year use with missing values or with extreme differential values from one item to the other item were excluded; i.e., a total of 206 answer sheets were excluded from the pre-survey sample (1.9%). The number of answer sheets with missing or extreme values excluded from the data analyses was 105 for the 20 treatment schools ($M = 5.25$, $SD = 2.73$) and 101 for the 20 control schools ($M = 5.05$, $SD = 2.54$). No significant difference was found between the average treatment and

average control school for answer sheets with missing or extreme values ($t(38) = 0.24$, $p = .8119$). The combined attrition rate for both tagged answer sheets and answer sheets with missing or extreme paired values was 3.2% (i.e., 11,115 - 147 - 206 = 10,762) for the pre-survey administration.

Six of the seven R&P subscales (i.e., goal setting and decision making, peer resistance, bonding with prosocial peers, self-efficacy, harmful effects, ATOD attitudes) were each comprised of 6 items. Mean values were imputed for these six R&P subscales if they were missing no more than 20% of the responses. Across these six R&P subscales, the number of imputed scores ranged from 45 to 115. If students had more than 20% of their answers missing on a subscale, their response to that subscale was excluded from individual analyses. However, respondents' answers to other ATOD and R&P subscales meeting the decision rules were included in other analyses. These exclusions ranged from 28 to 476 across the six R&P subscales. The numbers of responses with more than 20% missing data for the six R&P subscales and excluded from analysis were examined for potential bias between groups using the *t*-test procedure. No significant differences for exclusions due to missing responses were found between the treatment/control conditions across the six R&P subscales ($p$ values ranged from .1884 to .7849).

**Attrition and Missing Data at the Post-Survey**

A total of 233 fewer answer sheets were collected at the time of the post-survey (10,882) in comparison to the 11,115 answer sheets collected at the pre-survey (2.1%). The treatment and control schools maintained their approximate 50% representation, 5,251 and 5,262, respectively. This suggests no attrition bias was present between the two groups due to absenteeism, opting out of the survey, or other random occurrences and events across schools.

Administration of the post-survey resulted in 10,691 usable answer sheets. A total of 191 answer sheets were tagged by the SAs during the administration of the post-survey (191 out of 10,882). The number of tagged answer sheets excluded from the data analyses was 102 for the 20 treatment schools ($M = 5.10$, $SD = 3.13$) and 89 for the 20 control schools ($M = 4.45$, $SD = 4.80$). We found no significant difference between the average number of tagged answer sheets excluded from the sample for treatment and control schools ($t(38) = 0.51$, $p = .6146$) on the post-survey.

Following the decision rules as discussed above, no values were imputed for survey subscales missing more than 20% of the responses. Two-item ATOD subscales for 30 day and year use with missing values or with extreme differential values from one item to the other item were excluded; i.e., a total of 178 answer sheets were excluded from the post-survey sample. The number of answer sheets with missing or extreme values excluded from the data analyses was 91 for the 20 treatment schools ($M = 4.55$, $SD = 3.27$) and 87 for the 20 control schools ($M = 4.35$, $SD = 2.70$). No significant difference was found between the average treatment and average control school for answer sheets with missing or extreme values ($t(38) = 0.21$, $p = .8340$).

The combined attrition rate from the pre-survey to the post-survey for reduced number of respondents ($n = 233$), tagged answer sheets ($n = 191$), and answer sheets with missing or extreme paired values ($n = 178$) was 5.4% (i.e., 11,115 - 233 - 191 - 178 = 10,513).

Mean values were imputed for six of the R&P subscale scores missing no more than 20% of the responses (i.e., scores based on 6-items). Across the six R&P subscales, the number of imputed scores ranged from 37 to 121. If students had more than 20% of their answers missing on a subscale, their response to that subscale was excluded from individual analyses. These

exclusions ranged from 11 to 162 across the six R&P subscales. The numbers of responses with more than 20% missing data for the six R&P subscales and excluded from analysis were examined for potential bias between groups using the *t*-test procedure. No significant differences for exclusions due to missing responses were found between the treatment/control conditions at the post-survey across the six R&P subscales (*p* values ranged from .2493 to .9465).

**Attrition and Missing Data at the 6-Month Follow-Up**

A total of 477 fewer answer sheets were collected at the time of the 6-month follow-up (10,638) in comparison to the 11,115 answer sheets collected at the pre-survey (4.3%). The treatment and control schools maintained their approximate 50% representation, 5,060 and 5,097, respectively. This suggests no attrition bias was present between the two groups with fewer respondents due to absenteeism, opting out of the survey, or other random transitions and occurrences across schools.

Administration of the 6-month follow-up survey resulted in 10,427 scannable answer sheets. A total of 211 answer sheets were tagged by the SAs during the administration of the 6-month follow-up (211 out of 10,638). The number of tagged answer sheets excluded from the data analyses was 117 for the 20 treatment schools ($M = 5.58$, $SD = 5.50$) and 94 for the 20 control schools ($M = 4.70$, $SD = 3.48$). We found no significant difference between the average number of answer sheets excluded from the sample for treatment and control schools ($t(38) = 0.79$, $p = .4337$) on the 6-month follow-up.

No values were imputed for survey subscales missing more than 20% of the responses. Two-item ATOD subscales for 30 day and year use with missing values or with extreme differential values from one item to the other item were excluded; i.e., a total of 264 answer sheets were excluded from the 6-month follow-up sample. The number of answer sheets with

missing or extreme values excluded from the data analyses was 140 for the 20 treatment schools ($M$ = 7.00, $SD$ = 4.03) and 124 for the 20 control schools ($M$ = 6.20, $SD$ = 2.26). No significant difference was found between the average treatment and average control school for answer sheets with missing or extreme values ($t(38) = 0.77$, $p = .4433$).

The combined attrition rate from pre-survey to 6-month follow-up for reduced number of respondents ($n$ = 477), tagged answer sheets ($n$ = 211), and answer sheets with missing or extreme paired values ($n$ = 264) was 8.6% (i.e., 11,115 - 477 - 211 - 264 = 10,163).

Mean values were imputed for six of the R&P subscale scores missing no more than 20% of the responses. Across the six R&P subscales, the number of imputed scores ranged from 31 to 102. If students had more than 20% of their answers missing on a subscale, their response to that subscale was excluded from individual analyses. These exclusions ranged from 16 to 128 across the six R&P subscales. No significant differences for exclusions due to missing responses were found between the treatment/control conditions at the 6-month follow-up across the six R&P subscales ($p$ values ranged from .3640 to .9999).

**Demographics**

As indicated in Table 21, the treatment and control groups across the three survey periods (pre-survey, post-survey, and 6-month follow-up) are highly similar to one another with respect to group size and to composition by gender, ethnicity and age. Gender and ethnicity percentages remain fairly constant within the two groups across the three survey periods, despite the occurrence of a small amount of attrition between the pre-survey ($n$ = 10,762) and the 6-month follow-up ($n$ = 10,163). No significant differences were found between the treatment/control conditions for gender and ethnicity at each of the three survey periods ($p > .05$). Age shows an expected, consistent increase within the two groups across the three periods.

The treatment and control groups are also highly similar in percentage of students receiving federal free or reduced lunch program services (SES), limited English proficiency services (LEP), and exceptional education services (ESE). The demographic data for SES, LEP and ESE were provided by the district at the beginning and at the end of the school year (i.e., unique identification codes allowing students to remain anonymous). No significant differences were found between the treatment/control conditions for SES, LEP and ESE at the beginning and the end of the school year ($p > .05$). Overall, the treatment and control groups were highly similar across the pre-survey, post-survey, and 6-month follow-up with respect to demographic makeup.

Table 21

*Participant Demographics by Group and Time*

| | Pre-Survey % (*N* = 10,762) | | Post-Survey % (*N* = 10,513) | | 6-Month Follow-Up % (*N* = 10,163) | |
|---|---|---|---|---|---|---|
| | Treatment | Control | Treatment | Control | Treatment | Control |
| Demographics | (n = 5,389) | (n = 5,373) | (n = 5,251) | (n = 5,262) | (n = 5,066) | (n = 5,097) |
| Male | 51.7 | 53.0 | 50.5 | 51.6 | 50.4 | 51.4 |
| White | 38.0 | 38.3 | 37.6 | 38.5 | 36.8 | 36.9 |
| Black | 19.7 | 18.7 | 18.6 | 17.5 | 18.6 | 17.0 |
| Hispanic | 28.8 | 30.6 | 29.2 | 31.0 | 29.3 | 32.0 |
| Asian | 2.8 | 2.9 | 3.1 | 3.0 | 3.5 | 3.0 |
| Multiracial | 10.7 | 9.5 | 11.5 | 10.0 | 11.8 | 11.1 |
| Age[a] | 11.32 yrs | 11.34 yrs | 11.56 yrs | 11.60 yrs | 11.99 yrs | 12.04 yrs |
| At Grade Level[b] | | | | | | |
| SES | 59.4 | 59.9 | | | 59.4 | 59.9 |
| LEP | 13.2 | 14.4 | | | 13.2 | 14.5 |
| ESE | 11.3 | 11.1 | | | 11.2 | 11.0 |

a = average across age categories
b = data were provided by the district/schools for classes participating in the study at the beginning and at the end of the school year

## Comparison of Treatment to Control Schools on Student Survey Outcomes

Sixth grade students in the treatment and control schools completed the SBRPFS survey on an identical schedule: 1) before the delivery of the intervention in the treatment schools, 2) after the delivery of the treatment, and 3) 6-months following the delivery of the treatment intervention. Specific to the ATOD drug usage outcomes, analyses were completed using summated scores from each of the three survey administrations for tobacco, alcohol, and marijuana use in the past 30-days. Analyses were also completed across the three survey periods for binge drinking, a one-item variable. Summated scores for smoking, drinking, and marijuana use in the past year were examined for the pre-survey and the 6-month follow-up. Specific to the R&P factors listed above, analyses were completed on summated scores for 1) intentions to use drugs in the future, 2) goal setting and decision making skills, 3) social and peer resistance skills, 4) emotional competency/self-efficacy skills, 5) bonding with prosocial peers, 6) perceptions of the harmful effects of drugs, and 7) attitudes towards drug use.

### ICCs for Pre-Survey

Each of the outcomes was examined prior to the beginning of treatment to assess the degree to which these outcomes were clustered within schools. More specifically, unconditional multilevel models were estimated by restricted maximum likelihood using the MIXED PROCEDURE in SAS. The unconditional model can be represented as a two-level model. For example, using smoking in the past 30-days (Smoke30) as the outcome of interest, the first level of the model could be written as

$$Smoke30_{ij} = \beta_{0j} + e_{ij}$$

where $Smoke30_{ij}$ is the 30-day tobacco use for the $i^{th}$ student in the $j^{th}$ school, $\beta_{0j}$ is the average of Smoke30 for the $j^{th}$ school, and $e_{ij}$ is error accounting for the deviation from the school mean of

the $i^{th}$ individual at the $j^{th}$ school. These individual errors are assumed to be normally distributed with common variance ( $\sigma^2$ ).

At the second level of the model, the school means ( $\beta_{0j}$ ) are modeled to vary. More specifically

$$\beta_{0j} = \gamma_{00} + r_{0j}$$

where $\gamma_{00}$ is the average 30-day tobacco use for an average school, and $r_{0j}$ is an error term that accounts for the deviation of the $j^{th}$ school's mean from the average mean. These school level errors are assumed to be normally distributed with common variance ( $\tau_{00}$ ).

By estimating the variance at each level (i.e., $\sigma^2$ and $\tau_{00}$ ), an intraclass correlation coefficient (ICC) can be computed which indicates the proportion of variance that is between schools. For the 30-day tobacco use variable, the within school variance estimate was 0.0710, while the between school variance estimate was much smaller (0.000293). The ICC was .004, indicating that less than one percent of the variance was between schools.

The estimates of the variance components along with the ICCs for each of the use outcomes and each of the R&P outcomes are shown in Table 22. Scanning these ICC values, we see that almost all the variance is within schools, ranging from 0.3% of the variance between schools for "Drinking in the Past Year" to 4.2% of the variance for "Bonding with Prosocial Peers". Given these results, we would anticipate that the estimates and inferences from the two-level models that we present would not differ drastically from the results one would obtain using more traditional single-level regression models.

Table 22

*Variance Components and ICC Values for ATOD Use Outcomes and R&P Outcomes Measured at the Pre-Survey*

| Outcome | $\sigma^2$ | $\tau_{00}$ | ICC |
|---|---|---|---|
| ATOD Use Outcomes | | | |
| Smoking Past 30 Days | 0.07100 | 0.000293 | .004 |
| Smoke Past Year | 0.10180 | 0.000779 | .008 |
| Drink Past 30 Days | 0.11440 | 0.000892 | .008 |
| Drink Past Year | 0.18530 | 0.000645 | .003 |
| Binge Drink | 0.09420 | 0.000528 | .006 |
| Dope Past 30 Days | 0.06765 | 0.000521 | .008 |
| Dope Past Year | 0.09163 | 0.000730 | .008 |
| R&P Outcomes | | | |
| Intent ATOD Use | 0.42710 | 0.008934 | .021 |
| Goals & Decisions | 0.40640 | 0.002408 | .006 |
| Bond Prosocial Peers | 0.58990 | 0.026090 | .042 |
| Peer Resistance | 0.38920 | 0.007947 | .020 |
| Self-Efficacy | 0.40110 | 0.006246 | .015 |
| Harmful Effects | 0.81520 | 0.012840 | .016 |
| ATOD Attitudes | 0.46290 | 0.008877 | .019 |

## Pre-Survey Equivalence between Treatment and Control Schools

A two-level model was used to examine for possible differences in the treatment and control school students prior to the delivery of the treatment intervention (pre-survey). We specified the model so that it would be consistent with the model we intended to use when examining for post-survey differences. In the post-survey analyses it was important to examine for different levels of student risk, because our review of the intervention literature and our understanding of the treatment both suggested that the treatment effect may differ by risk level. Some studies have indicated that drug intervention curricula may have a greater impact on those adolescents at higher risk for escalating drug use (e.g., Botvin et al., 2001a; Eisen et al., 2002;

Ellickson et al., 2003; Longshore et al., 2007; Lynsky et al., 2003; Sussman et al., 1998). More specifically, we anticipated that those students at higher risk for substance use would benefit more from the treatment. For example, for students who don't smoke and have little to no chance of starting to smoke, the treatment is not expected to meaningfully lessen their smoking behavior. But for students who have either tried or are currently smoking, the treatment is expected to meaningfully lessen their smoking behavior. To account for this anticipated interaction effect, a risk variable was created and added to the statistical models.

The students in the study responded to three survey items asking how old they were when they, respectively, first smoked a cigarette, first drank any alcohol, and first tried marijuana. On the basis of their response to these items, they were categorized in terms of three levels of risk: 1) low risk students are those who indicate that they had tried none of the three ATOD substances (cigarettes, alcohol or marijuana) at 10 years of age or younger; 2) moderate risk students are those who indicate having tried one of the three ATOD substances at 10 years of age or younger; and 3) high risk students are those who indicate having tried two or more or the three ATOD substances at 10 years of age or younger. As would be expected for most sixth graders between the age of 11 and 12 (average age range 11.32 to 12.04), substance use rates are low. Approximately 77% in each of the study groups (treatment and control) were in the low risk category, 18% were in the moderate risk category, and 5% in the high risk category.

The multilevel model was parameterized such that coefficients would be obtained that yielded differences between treatment and control high risk students, between treatment and control low risk students, and between treatment and control moderate risk students. More specifically, the 3-category risk variable was dummy coded and then the dummy coded variables

were included in the first level of the model. The first level of the model for the 30-day tobacco use outcome could be written as

$$Smoke30_{ij} = \beta_{0j} + \beta_{1j} RiskHigh + \beta_{2j} RiskLow + \beta_{3j} RiskMod + e_{ij}$$

where $Smoke30_{ij}$ is the 30-day tobacco use for the $i^{th}$ student in the $j^{th}$ school, $\beta_{0j} + \beta_{1j}$ is the average of Smoke30 for high risk students at the $j^{th}$ school, $\beta_{0j} + \beta_{2j}$ is the average of Smoke30 for low risk students at the $j^{th}$ school, $\beta_{0j} + \beta_{3j}$ is the average of Smoke30 for moderate risk students at the $j^{th}$ school, and $e_{ij}$ is error that is assumed to be normally distributed with common variance ($\sigma^2$).

At the second level of the model, the regression coefficients from the first level are modeled as a function of treatment group (Treatment = 0, Control = 1). More specifically

$$\beta_{0j} = \gamma_{00} + r_{0j}$$
$$\beta_{1j} = \gamma_{11} Tx$$
$$\beta_{2j} = \gamma_{20} + \gamma_{21} Tx$$
$$\beta_{3j} = \gamma_{30} + \gamma_{31} Tx$$

where $\gamma_{00}$ is the average 30-day tobacco use for a high risk student at an average treatment school, $\gamma_{20}$ is the difference in average 30-day tobacco use between an average treatment school low risk student and an average treatment school high risk student, $\gamma_{30}$ is the difference in average 30-day tobacco use between an average treatment school moderate risk student and an average treatment school high risk student, $\gamma_{11}$ is the difference in average 30-day tobacco use between an average control school high risk student and an average treatment school high risk student, $\gamma_{21}$ is the difference in average 30-day tobacco use between an average control school low risk student and an average treatment school low risk student, $\gamma_{31}$ is the difference in average

30-day tobacco use between an average control school moderate risk student and an average treatment school moderate risk student, and $r_{0j}$ is an error term that is assumed to be normally distributed with common variance ($\tau_{00}$).

The results from the two-level model examining ATOD use outcomes on the pre-survey are shown in Table 23. The average 30-day tobacco use for a high risk student at an average treatment school is 1.502; for a low risk student at an average treatment school the average 30-day tobacco use is lower by 0.486, and for a moderate risk student at an average treatment school the average 30-day tobacco use compared to a high risk student is lower by 0.445. The difference between high and low risk students, as well as the difference between high and moderate risk students is statistically significant. Turning to the difference between treatment and control students, we find that among high risk students those in control schools have average 30-day tobacco use values that are 0.019 lower than those in treatment schools, but that this difference is not statistically significant. For low risk students, control students have an average that is 0.002 lower than treatment students, and for moderate risk students, control students have an average that is 0.005 higher than treatment students. As with high risk students, neither of these differences is statistically significant. To account for sequentially testing multiple outcomes we used the modified Bonferroni adjustment of Holm (1979), where the lowest *p*-value is compared to $\alpha/n$, where $n$ is the number of tests, the second lowest *p*-value is compared to $\alpha/(n-1)$, and so on.

Setting $\alpha = .05$ for the set of comparisons and looking across the differences between high risk treatment school students and high risk control school students ($\gamma_{11}$ in Table 23), we see that none of the ATOD use outcomes was statistically significant, with the exception of smoking tobacco during the past year ($p = .0054$). High risk students in control schools had an

average smoking tobacco during the past year value that was 0.112 lower than high risk students in treatment schools.

Using the same modified Bonferroni procedure to control the Type I error rate for the set of comparisons between low risk students ($\gamma_{21}$) we found no statistically significant differences. Finally this process was repeated for the set of moderate risk comparisons ($\gamma_{31}$) and again no significant differences were found. These results suggest that prior to the delivery of the treatment intervention; the treatment and control students were similar on each of the ATOD use outcomes for each level of risk, with the exception of high risk control students smoking less during the past year than their counterparts in the treatment schools.

Table 23

*Two-level Model Results Comparing the Treatment Schools to the Control Schools on ATOD Use Outcomes at Pre-Survey*

| Outcome | $N$ | $\sigma^2$ | $\tau_{00}$ | $\gamma_{00}$ | $\gamma_{20}$ | $\gamma_{30}$ | $\gamma_{11}$ (High) | $\gamma_{21}$ (Low) | $\gamma_{31}$ (Moderate) |
|---|---|---|---|---|---|---|---|---|---|
| Smoke Past | 10,762 | 0.0661 | 0.0001 | 1.502 | -0.486 | -0.445 | -0.019 | -0.002 | 0.005 |
| 30 Days | | | | (.022) | (.023) | (.024) | (.034) | (.007) | (.013) |
| Smoke Past | 10,762 | 0.0906 | 0.0004 | 1.785 | -0.760 | -0.685 | -0.112* | -0.007 | 0.000 |
| Year | | | | (.026) | (.026) | (.028) | (.040) | (.009) | (.009) |
| Drink Past | 10,762 | 0.1039 | 0.0006 | 1.650 | -0.611 | -0.449 | -0.047 | -0.004 | 0.007 |
| 30 Days | | | | (.028) | (.028) | (.030) | (.043) | (.011) | (.017) |
| Drink Past | 10,762 | 0.1582 | 0.0004 | 1.914 | -0.849 | -0.536 | -0.041 | -0.007 | 0.008 |
| Year | | | | (.035) | (.035) | (.037) | (.053) | (.011) | (.020) |
| Binge | 10,756 | 0.0906 | 0.0003 | 1.529 | -0.500 | -0.423 | -0.008 | -0.008 | 0.003 |
| Drinking | | | | (.026) | (.026) | (.028) | (.040) | (.009) | (.015) |
| Dope Past | 10,762 | 0.0635 | 0.0003 | 1.467 | -0.451 | -0.422 | -0.021 | -0.006 | -0.001 |
| 30 Days | | | | (.022) | (.022) | (.023) | (.034) | (.008) | (.013) |
| Dope Past | 10,762 | 0.0847 | 0.0004 | 1.626 | -0.603 | -0.567 | -0.077 | -0.008 | -0.000 |
| Year | | | | (.026) | (.026) | (.027) | (.039) | (.009) | (.015) |

*Note*. The standard errors of the parameter estimate are enclosed in parentheses.
*$p < .05$, controlling for Type I error using Holm-Bonferroni method

The results from the two-level model examining R&P outcomes are shown in Table 24. For example, the average rating for goal setting and decision making skills for a high risk student at an average treatment school is 3.682; for a low risk student at an average treatment school the average rating for goal setting and decision making is higher by 0.550, and for a moderate risk student at an average treatment school the average rating for goal setting and decision making compared to a high risk student is higher by 0.277. The difference between high and low risk students, as well as the difference between high and moderate risk students is statistically significant. Turning to the difference between treatment and control students, we find that among high risk students those in control schools have an average goal setting and decision making value that is 0.109 higher than those in treatment schools, but that this difference is not statistically significant. For low risk students, control students have an average that is 0.004 higher than treatment students, and for moderate risk students, control students have an average that is 0.051 higher than treatment students. As with high risk students, neither of these differences is statistically significant.

Looking across the differences between high risk treatment school students and high risk control school students ( $\gamma_{11}$ in Table 24) we see that none of the differences were statistically significant. Using the same modified Bonferroni procedure to control the Type I error rate for the set of comparisons between low risk students ( $\gamma_{21}$ ), we again find no statistically significant differences. Finally this process was repeated for the set of moderate risk comparisons ( $\gamma_{31}$ ) and again no significant differences were found. These results suggest that prior to delivering the treatment intervention; the treatment and control students were similar on each of the risk and protective outcomes for each level of risk.

Table 24

*Two-level Model Results Comparing the Treatment Schools to the Control Schools on R&P Outcomes at Pre-Survey*

| Outcome | $N$ | $\sigma^2$ | $\tau_{00}$ | $\gamma_{00}$ | $\gamma_{20}$ | $\gamma_{30}$ | $\gamma_{11}$ (High) | $\gamma_{21}$ (Low) | $\gamma_{31}$ (Moderate) |
|---|---|---|---|---|---|---|---|---|---|
| Intent ATOD | 10,728 | 0.0480 | 0.0070 | 3.988 | 0.790 | 0.558 | 0.134 | 0.056 | 0.020 |
| Use | | | | (.058) | (.056) | (.059) | (.089) | (.030) | (.040) |
| Goals & | 10,734 | 0.3934 | 0.0021 | 3.682 | 0.550 | 0.277 | 0.109 | 0.004 | 0.051 |
| Decisions | | | | (.055) | (.055) | (.058) | (.084) | (.020) | (.033) |
| Peer Resistance | 10,633 | 0.3643 | 0.0064 | 3.680 | 0.949 | 0.636 | 0.225 | 0.023 | 0.091 |
| | | | | (.056) | (.054) | (.056) | (.084) | (.029) | (.039) |
| Self-Efficacy | 10,532 | 0.3917 | 0.0059 | 3.558 | 0.439 | 0.210 | 0.058 | 0.010 | 0.033 |
| | | | | (.058) | (.056) | (.059) | (.087) | (.028) | (.039) |
| Bond Prosocial | 10,696 | 0.5489 | 0.0231 | 3.324 | 1.128 | 0.719 | 0.248 | 0.032 | 0.065 |
| Peers | | | | (.073) | (.066) | (.069) | (.110) | (.051) | (.060) |
| Harmful Effects | 10,418 | 0.8081 | 0.0125 | 3.968 | 0.536 | 0.365 | 0.050 | -0.034 | 0.043 |
| | | | | (.084) | (.082) | (.086) | (.126) | (.041) | (.056) |
| ATOD | 10,286 | 0.4253 | 0.0074 | 3.449 | 1.013 | 0.597 | 0.207 | 0.017 | 0.046 |
| Attitudes | | | | (.062) | (.060) | (.063) | (.093) | (.031) | (.042) |

*Note*. The standard errors of the parameter estimate are enclosed in parentheses.

## Comparison of Treatment to Control Schools at Post-Survey

A two-level model was also used to examine for differences in the treatment and control school students' outcomes shortly after the delivery of the treatment intervention. The model specification matched the model specification at the pre-survey (described above), and thus for Smoke30 the model was:

$$Smoke30_{ij} = \beta_{0j} + \beta_{1j} RiskHigh + \beta_{2j} RiskLow + \beta_{2j} RiskMod + e_{ij}$$

$$\beta_{0j} = \gamma_{00} + r_{0j}$$

$$\beta_{1j} = \gamma_{11} Tx$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21} Tx$$

$$\beta_{3j} = \gamma_{30} + \gamma_{31} Tx$$

The results from estimating this model with the 30-day tobacco use outcome is illustrated in Figure 1, which shows the interaction between student risk level and treatment. Initial inspection of the graph reveals differences between the experimental and control group that become more pronounced with the rise in risk. The average value of 30-day tobacco use for high risk students in a treatment school is 1.44. Moving to moderate risk students drops this average .334 and moving to low risk students drops the average .417. If we focus on the gap between students at the control and treatment schools, we see that on average, high risk students at control schools have a higher 30-day tobacco use than high risk students in treatment schools ( $\gamma_{11}$ = 0.297), a difference that is statistically significant ($p$ <.0001). The difference in 30-day tobacco use of students at control and treatment schools is much smaller for moderate risk students (.017) and low risk students (.004).

The estimated parameter values from the two-level model for each of the ATOD use outcomes on the post-survey are provided in Table 25. Looking across the outcomes in Table 25, we see similar patterns emerge. On the usage outcomes, *lower values* indicate less frequent ATOD usage, a desirable effect. For all 30-day usage outcomes (smoking, drinking, binge drinking, and marijuana use), values are higher for the high risk students as evidenced by the negative coefficients for $\gamma_{20}$ and $\gamma_{30}$, but the degree to which they are elevated depends on whether the student is in a treatment or control school. In each case the treatment school students show lower average values than control students as evidenced by the positive signs for $\gamma_{11}$. Overall, these results suggest that the treatment is having a *suppression* effect on students' reported 30-day use of tobacco, alcohol, binge drinking and marijuana, particularly for students defined as high risk.

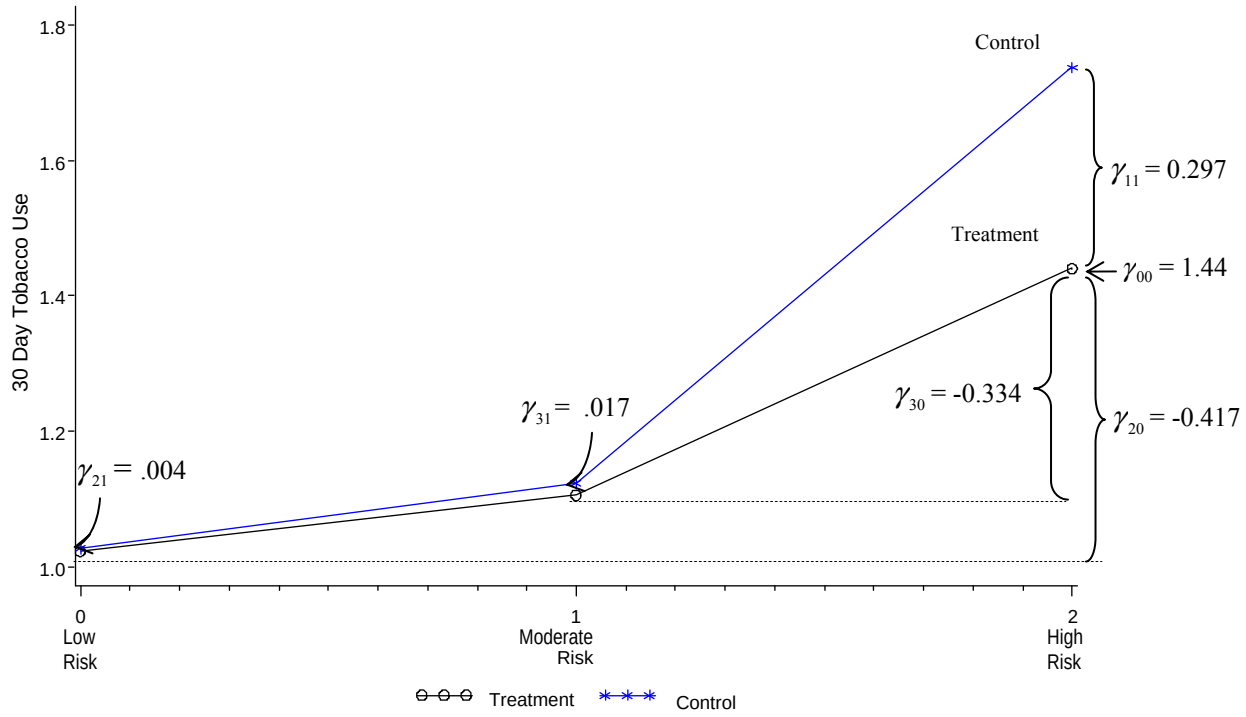



*Figure 1.* Smoking in the Past 30 Days by Group and Risk Level on the Post-Survey

The significance tests for $\gamma_{11}$, which indexes the difference between treatment and control conditions for the high risk students, produced *p*-values of .0001. Each difference was found to be statistically significant using the Holm modified Bonferroni approach to control the Type I error rate to 0.05 for the set of comparisons. These differences are judged to be of practical importance based on an examination of the raw coefficients (see Table 25, or Figure 1 for illustration of the effect on frequency of smoking during the past 30 days), and also based on measures of standardized effect size. These standardized effect size measures were obtained by dividing $\gamma_{11}$ by the square root of $(\sigma^2 + \tau_{00})$. The standardized effect size for high risk treatment students in comparison to high risk control students for smoking during the past 30-days was

0.82. The standardized effect sizes for the other 30-day usage measures for high risk students for drinking, binge drinking, and marijuana use were 0.56, 0.63, and 1.03, respectively.

The significance tests for $\gamma_{31}$, which indexes the difference between treatment and control conditions for the moderate risk students, produced *p*-values ranging from .0001 to .3460. Three of the ATOD use outcomes were found to be statistically significant using the Holm modified Bonferroni approach to control the Type I error rate to 0.05. Standardized effect size measures for moderate risk students were obtained by dividing $\gamma_{31}$ by the square root of $(\sigma^2 + \tau_{00})$. The standardized effect size for 30-day usage measures for moderate risk students was 0.14 for drinking ($p = .0031$), 0.19 for binge drinking ($p = .0001$), and 0.15 for marijuana use ($p = .0046$). No statistically significant difference between moderate risk students in the treatment and control conditions was observed for smoking tobacco in the past 30-days. Using the same modified Bonferroni procedure to control the Type I error rate for the set of ATOD use comparisons between low risk students ($\gamma_{21}$) we found no statistically significant differences.

Table 25

*Two-level Model Results Comparing the Treatment Schools to the Control Schools on ATOD Use Outcomes at Post-Survey*

| Outcome | *N* | $\sigma^2$ | $\tau_{00}$ | $\gamma_{00}$ | $\gamma_{20}$ | $\gamma_{30}$ | $\gamma_{11}$ (High) | $\gamma_{21}$ (Low) | $\gamma_{31}$ (Moderate) |
|---|---|---|---|---|---|---|---|---|---|
| Smoke Past 30 Days | 10,513 | 0.1294 | 0.0005 | 1.440 (.026) | -0.417 (.026) | -0.334 (.028) | 0.297* (.038) d = 0.82 | 0.004 (.011) | 0.017 (.018) |
| Drink Past 30 Days | 10,513 | 0.2406 | 0.0010 | 1.743 (.032) | -0.681 (.032) | -0.469 (.035) | 0.275* (.048) d = 0.56 | 0.012 (.014) | 0.068* (.023) d = 0.14 |
| Binge Drinking | 10,510 | 0.1759 | 0.0006 | 1.525 (.030) | -0.487 (.030) | -0.375 (.034) | 0.263* (.045) d = 0.63 | 0.008 (.012) | 0.081* (.021) d = 0.19 |
| Dope Past 30 Days | 10,513 | 0.1322 | 0.0008 | 1.400 (.026) | -0.375 (.026) | -0.327 (.028) | 0.371* (.039) d = 1.03 | 0.003 (.012) | 0.053* (.019) d = 0.15 |

*Note*. The standard errors of the parameter estimate are enclosed in parentheses, and d = standardized effect size.
*$p < .05$, controlling for Type I error using Holm-Bonferroni method

The estimated parameter values from the two-level model for each of the R&P outcomes on the post-survey are provided in Table 26. On the risk and protective outcomes, higher values indicate greater skill or more positive attitudes, a desirable effect. Looking at Table 26, we see that for all R&P outcomes (intentions to use drugs, goal setting and decision making, social and peer resistance skills, emotional competence/self-efficacy, bonding with prosocial peers, harmful effects, and attitudes toward drugs), high risk students had the lowest values, as evidenced by the positive values for $\gamma_{20}$ and $\gamma_{30}$. The degree to which the values were lower for the high risk students depended on group assignment. For each outcome, the treatment students' average for high risk students was higher than the control students' average for high risk students. This can be seen in the negative signs for $\gamma_{11}$. These results suggest that the treatment is having a *strengthening* effect on students' risk and protective skills and attitudes, particularly for students defined as high risk.

The significance tests for $\gamma_{11}$, which indexes the difference between treatment and control conditions for the high risk students, produced *p*-values that ranged from $<.0001$ for the majority of the R&P outcomes to $<.0069$ for bonding with prosocial peers. Each difference was found to be statistically significant using the Holm modified Bonferroni approach to control the Type I error rate to 0.05. Standardized effect size measures were obtained by dividing $\gamma_{11}$ by the square root of $(\sigma^2 + \tau_{00})$. The standardized effect size measures for high risk treatment students in comparison to high risk control students for the R&P outcomes were 0.67 for Intentions to Use Drugs, 0.33 for Goal Setting & Decision Making, 0.76 for Social and Peer Resistance Skills, 0.38 for Emotional Competency/Self-Efficacy, 0.33 for Bonding with Prosocial Peers, 0.40 for Harmful Effects of Drugs, and 0.43 for ATOD Attitudes.

Standardized effect size measures for moderate risk students were obtained by dividing $\gamma_{31}$ by the square root of $(\sigma^2 + \tau_{00})$. Statistically significant effects were observed for two R&P

outcomes for moderate risk students, favoring the treatment group, with effect sizes of 0.31 for Peer Resistance Skills ($p$ = .0001) and 0.19 for Emotional Competency/Self-Efficacy ($p$ = .0022).

Standardized effect size measures for low risk students were obtained by dividing $\gamma_{21}$ by the square root of ($\sigma^2 + \tau_{00}$). Statistically significant standardized effect sizes for low risk treatment students in comparison to low risk control students for the R&P outcomes were 0.20 for Goal Setting & Decision Making ($p$ = .0001), 0.23 for Social and Peer Resistance ($p$ = .0001), and 0.23 for Emotional Competency/Self-Efficacy ($p$ = .0001).

Table 26

*Two-level Model Results Comparing the Treatment Schools to the Control Schools on R&P Outcomes at Post-Survey*

| Outcome | $N$ | $\sigma^2$ | $\tau_{00}$ | $\gamma_{00}$ | $\gamma_{20}$ | $\gamma_{30}$ | $\gamma_{11}$ (High) | $\gamma_{21}$ (Low) | $\gamma_{31}$ (Moderate) |
|---|---|---|---|---|---|---|---|---|---|
| Intent ATOD Use | 10,499 | 0.4319 | 0.0090 | 4.002 (.051) | 0.783 (.047) | 0.444 (.051) | -0.444* (.075) d = 0.67 | 0.012 (.034) | -0.020 (.043) |
| Goals & Decisions | 10,502 | 0.4589 | 0.0039 | 3.806 (.050) | 0.550 (.049) | 0.177 (.052) | -0.223* (.074) d = 0.33 | -0.134* (.025) d = 0.20 | -0.078 (.037) |
| Peer Resistance | 10,444 | 0.3977 | 0.0083 | 3.989 (.049) | 0.741 (.045) | 0.421 (.049) | -0.487* (.072) d = 0.76 | -0.145* (.032) d = 0.23 | -0.194* (.041) d = 0.31 |
| Self-Efficacy | 10,411 | 0.4383 | 0.0077 | 3.654 (.051) | 0.537 (.048) | 0.203 (.051) | -0.256* (.075) d = 0.38 | -0.155* (.032) d = 0.23 | -0.126* (.041) d = 0.19 |
| Bond Prosocial Peers | 10,480 | 0.6093 | 0.0307 | 3.532 (.068) | 0.903 (.056) | 0.467 (.060) | -0.267* (.099) d = 0.33 | -0.033 (.058) | -0.062 (.066) |
| Harmful Effects | 10,370 | 0.6364 | 0.0123 | 4.077 (.062) | 0.538 (.058) | 0.421 (.062) | -0.321* (.091) d = 0.40 | -0.065 (.040) | -0.096 (.051) |
| ATOD Attitudes | 10,351 | 0.4659 | 0.0121 | 3.614 (.054) | 0.872 (.049) | 0.449 (.053) | -0.298* (.080) d = 0.43 | -0.041 (.038) | -0.077 (.047) |

*Note*. The standard errors of the parameter estimate are enclosed in parentheses, and d = standardized effect size.
*$p$ < .05, controlling for Type I error using Holm-Bonferroni method

**Comparison of Treatment to Control Schools at 6-Month Follow-up**

A two-level model was again used to examine for differences in the treatment and control school students' outcomes 6-months after the delivery of the treatment intervention. The model specification matched the model specification at the pre-survey and post-survey, and thus for Smoke30 the model was:

$$Smoke30_{ij} = \beta_{0j} + \beta_{1j} RiskHigh + \beta_{2j} RiskLow + \beta_{2j} RiskMod + e_{ij}$$

$$\beta_{0j} = \gamma_{00} + r_{0j}$$

$$\beta_{1j} = \gamma_{11} Tx$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21} Tx$$

$$\beta_{3j} = \gamma_{30} + \gamma_{31} Tx$$

The estimated parameter values from this model are provided in Table 27 for the ATOD use outcomes. We included students' reported use of tobacco, alcohol and marijuana over the past year at this stage of the analyses since it had been almost 10 months since their reported use on the pre-survey. The patterns seen in these estimates are highly similar to those from the post-survey. Outcomes were again lower for the high risk students, but this was moderated by treatment group such that high risk students in treatment schools had more successful outcomes than high risk students at control schools. Again the difference between high risk students at treatment schools and high risk students at control schools is indexed by $\gamma_{11}$, and the test of this parameter revealed statistically significant results for all ATOD outcomes when using the Holm modified Bonferroni procedure to hold the α to .05 for the set of seven comparisons ($p$ = .0001 to $p$ = .0115).

Standardized effect sizes were also computed to gauge the size of the difference between high risk students at treatment schools and high risk students at control schools. The standardized effect sizes for the 30-day usage measures for high risk students for smoking,

drinking, binge drinking, and marijuana use, 6-months after treatment intervention, were 0.61, 0.65, 0.49, and 0.30, respectively. The standardized effect sizes for usage measures over the past year for high risk students for smoking, drinking, and marijuana use were 0.57, 0.51, and 0.26, respectively. Overall, these results suggest that, among high risk students, the *suppression* effect on ATOD usage outcomes was still in play six months after the treatment intervention.

Using the same modified Bonferroni procedure to control the Type I error rate for the set of ATOD use comparisons between moderate risk students ($\gamma_{31}$), and again for low risk students ($\gamma_{21}$), we found no statistically significant differences.

Table 27

*Two-level Model Results Comparing the Treatment Schools to the Control Schools on ATOD Use Outcomes at 6-Month Follow-up*

| Outcome | $N$ | $\sigma^2$ | $\tau_{00}$ | $\gamma_{00}$ | $\gamma_{20}$ | $\gamma_{30}$ | $\gamma_{11}$ (High) | $\gamma_{21}$ (Low) | $\gamma_{31}$ (Moderate) |
|---|---|---|---|---|---|---|---|---|---|
| Smoke Past 30 Days | 10,163 | 0.1580 | 0.0003 | 1.585 (.030) | -0.555 (.030) | -0.488 (.032) | 0.243* (.040) d = 0.61 | 0.010 (.011) | 0.029 (.019) |
| Smoke Past Year | 10,163 | 0.2235 | 0.0005 | 1.870 (.035) | -0.821 (.036) | -0.705 (.038) | 0.272* (.048) d = 0.57 | 0.009 (.013) | 0.057 (.023) |
| Drink Past 30 Days | 10,163 | 0.2583 | 0.0017 | 1.720 (.039) | -0.640 (.038) | -0.364 (.041) | 0.333* (.053) d = 0.65 | 0.016 (.018) | 0.015 (.018) |
| Drink Past Year | 10,163 | 0.3309 | 0.0011 | 1.973 (.043) | -0.842 (.043) | -0.433 (.046) | 0.293* (.059) d = 0.51 | 0.011 (.017) | 0.031 (.029) |
| Binge Drinking | 10,163 | 0.2204 | 0.0012 | 1.583 (.036) | -0.532 (.036) | -0.375 (.038) | 0.233* (.049) d = 0.49 | 0.009 (.015) | 0.026 (.024) |
| Dope Past 30 Days | 10,163 | 0.2039 | 0.0002 | 1.721 (.033) | -0.684 (.034) | -0.582 (.036) | 0.134* (.046) d = 0.30 | 0.012 (.011) | -0.002 (.021) |
| Dope Past Year | 10,163 | 0.2579 | 0.0005 | 1.872 (.038) | -0.822 (.038) | -0.682 (.041) | 0.131* (.052) d = 0.26 | 0.013 (.014) | -0.011 (.025) |

*Note*. The standard errors of the parameter estimate are enclosed in parentheses, and d = standardized effect size.
*$p < .05$, controlling for Type I error using Holm-Bonferroni method

On the R&P outcomes the significance tests for $\gamma_{11}$, which indexes the difference between treatment and control conditions for the high risk students on the R&P outcomes, produced *p*-values that ranged from <.0001 to <.1269 (see Table 28). More specifically, all results were statistically significant except those for Goal Setting & Decision Making ($p = .0379$) and Emotional Competency/Self-Efficacy ($p = .1269$). The standardized effect sizes for high risk treatment students in comparison to high risk control students for significant R&P outcomes were 0.35 for Intentions to Use Drugs, 0.63 for Social and Peer Resistance Skills, 0.30 for Bonding with Prosocial Peers, 0.32 for Harmful Effects of Drug Use, and 0.38 for ATOD Attitudes. Overall, these results suggest that, among high risk students, the *strengthening* effect on R&P outcomes were still in play six months after the treatment intervention.

Table 28

*Two-level Model Results Comparing the Treatment Schools to the Control Schools on R&P Outcomes at 6-Month Follow-up*

| Outcome | $N$ | $\sigma^2$ | $\tau_{00}$ | $\gamma_{00}$ | $\gamma_{20}$ | $\gamma_{30}$ | $\gamma_{11}$ (High) | $\gamma_{21}$ (Low) | $\gamma_{31}$ (Moderate) |
|---|---|---|---|---|---|---|---|---|---|
| Intent ATOD Use | 10,155 | 0.4578 | 0.0080 | 3.809 (.054) | 0.977 (.051) | 0.550 (.055) | -0.240* (.074) d = 0.35 | -0.037 (.032) | -0.005 (.042) |
| Goals & Decisions | 10,147 | 0.4947 | 0.0042 | 3.661 (.054) | 0.602 (.053) | 0.241 (.057) | -0.154 (.074) | -0.065 (.026) | -0.034 (.039) |
| Peer Resistance | 10,123 | 0.4564 | 0.0068 | 3.847 (.054) | 0.826 (.051) | 0.428 (.055) | -0.431* (.074) d = 0.63 | -0.113* (.031) d = 0.17 | -0.080 (.041) |
| Self-Efficacy | 10,098 | 0.4748 | 0.0090 | 3.528 (.056) | 0.609 (.053) | 0.272 (.056) | -0.117 (.076) | -0.107* (.034) d = 0.15 | -0.052 (.044) |
| Bond Prosocial Peers | 10,137 | 0.6733 | 0.0395 | 3.317 (.075) | 1.052 (.062) | 0.568 (.066) | -0.257* (.104) d = 0.30 | -0.041 (.066) | -0.068 (.074) |
| Harmful Effects | 10,073 | 0.6076 | 0.0136 | 3.968 (.064) | 0.645 (.060) | 0.388 (.064) | -0.250* (.088) d = 0.32 | -0.026 (.041) | 0.003 (.052) |
| ATOD Attitudes | 10,035 | 0.5386 | 0.0106 | 3.473 (.060) | 0.941 (.056) | 0.465 (.060) | -0.282* (.082) d = 0.38 | -0.045 (.037) | -0.045 (.048) |

*Note*. The standard errors of the parameter estimate are enclosed in parentheses, and d = standardized effect size.
*$p < .05$, controlling for Type I error using Holm-Bonferroni method

Statistically significant effects were observed for two R&P outcomes for low risk students, favoring the treatment group, with effect sizes of 0.17 for Social and Peer Resistance ($p$ = .0002) and 0.15 for Emotional Competency/Self-Efficacy ($p$ = .0017). No statistically significant R&P outcomes, while controlling for the Type I error rate, were observed between moderate risk students.

Cohen (1988) provides a widely accepted perspective on the interpretation of an effect size (d). He presents the operational definition of a small difference between means as d = .20, a medium difference between means as d = .50, and a large difference between means as d = .80. However, he suggests that the expectation of a particular effect size must be considered within the context of the area of behavior being studied, i.e. the terms "small", "medium" and "large" take their meaning relative to the field of research in which one is engaged and even relative to the specific content and research methods employed. Using medical research results as a frame of reference, Lipsey and Wilson (1993) argue that, in examining the efficacy of educational, psychological and behavioral treatments, one should not arbitrarily dismiss as trivial statistically significant results with effects sizes as modest as .10 to .20. For example, Tobler and Stratton (1997) show that a set of interactive school-based drug prevention programs they studied produced overall a modest effect size of .20 in comparison to an effect size of .02 for non-interactive programs. Using Rosenthal and Rubin's (1982) binomial effect size display (BESD), they found that these effect sizes were equal to a success rate of 9.5% and 1.0%, respectively, and suggest that these differences are clearly a clinically significant finding. Within the perspective of Lipsey and Wilson (1993) and Tobler and Stratton (1997), the effect sizes produced by the TGFD treatment are substantive.

## Results on the School Performance Indicators

### Comparison of Treatment to Control Schools on Attendance and Suspensions

The *t*-test procedure was used on the school means to examine for differences between the treatment and control group for the number of student absences and the number of suspensions (i.e., in- and out-of-school). The average number of days students were absent during the year for the 20 treatment schools was 3.94 (*SD* = 0.86) and the average days absent for the 20 control schools was 4.10 (*SD* = 0.93). We found no significant difference between the treatment and control schools for the average number of days students were absent ($t(38) = -0.59$, $p = .5614$). Similarly, no effect or difference between the treatment and control group was observed for the average number of suspensions assigned to students during the year ($t(38) = 1.38$, $p = .1751$). The average number of suspensions assigned during the school year per student for the 20 treatment schools was 0.59 (*SD* = 0.39) and the average number of suspensions for the 20 control schools was 0.46 (*SD* = 0.29).

### Comparison of Treatment to Control Schools on FCAT Mathematics and Reading

A two-level model was used to examine for differences in scale scores of the treatment and control school students on the FCAT Mathematics and on the FCAT Reading. The FCAT Mathematics and FCAT Reading are state-wide standardized achievement measures given each year to Florida public school students in grades 3-10. The 6$^{th}$ graders completed the FCAT in the spring of the year 2010, following the delivery of TGFD treatment during the fall of the previous year. Current and prior year achievement scores on the FCAT Mathematics and FCAT Reading were provided by the district for students in schools participating in the study. While students remained anonymous, the school district provided unique identification codes which allowed

matching between students' 5th and 6th grade reading and mathematics achievement scores. The model specification for FCAT Mathematics was:

$$FCATMath_{ij} = \beta_{0j} + \beta_{1j} PastMath + e_{ij}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} Tx + r_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} Tx$$

where PastMath was the previous year FCAT Mathematics scale score (grade 5), which was rescaled from 0 to 400, as opposed to 100 to 500, so that it would be centered such that zero corresponded to the lowest level of past achievement, and *Tx* was the school level treatment variable (0=Treatment School, 1=Control School). In this model $\gamma_{01}$ provides an estimate of the treatment effect for the lowest level of past mathematics achievement (*PastMath* = 0), and $\gamma_{11}$ is an interaction effect that provides an estimate of the amount the treatment effect changes with increases in past mathematics achievement.

The estimated parameter values are provided in Table 29. We see that at the lowest level of past mathematics achievement, those in the treatment schools were predicted to have a higher mathematics score by about 10 points, which is statistically significant ( $\gamma_{01}$ = -9.98, $t$(11000) = -2.85, $p$ = .0044). This effect, however, is moderated by past achievement such that the beneficial effect of being in a treatment school gets smaller by .032 for each scale score increase in past mathematics achievement ( $\gamma_{11}$ = 0.032, $t$(11000) = 2.47, $p$ = .0136). Thus, if we consider students with a past mathematics achievement score in the middle of the distribution (i.e., a rescaled past FCAT mathematics score of 200, which corresponds to a score of 300 on the original scale), we see a smaller difference of only about 3.6 points scale score points between treatment and control school students (-3.58 = -9.98 + 0.032*200).

Table 29

*Two-level Model Results Comparing the Treatment Schools to the Control Schools on FCAT Mathematics and FCAT Reading Achievement*

| Outcome | $N$ | $\sigma^2$ | $\tau_{00}$ | $\gamma_{00}$ | $\gamma_{10}$ | $\gamma_{01}$ | $\gamma_{11}$ |
|---|---|---|---|---|---|---|---|
| FCAT Mathematics | 10,580 | 1038.99 | 41.29 | 95.33 | 0.959 | -9.98* | 0.032* |
| | | | | (2.50) | (.009) | (3.50) | (.013) |
| FCAT Reading | 10,571 | 1280.99 | 58.93 | 142.7 | 0.816 | -1.99 | 0.013 |
| | | | | (2.48) | (.009) | (3.41) | (.013) |

*Note*. The standard errors of the parameter estimate are enclosed in parentheses
*$p < .05$, controlling for Type I error across the two outcomes using Holm-Bonferroni method

To provide a better sense of the treatment effect across the continuum of past mathematics achievement, standardized effect sizes were computed and plotted in Figure 2. These standardized effect size measures were obtained by dividing ( $\gamma_{01} + \gamma_{11} * PastMath$) by the square root of ( $\sigma^2 + \tau_{00}$ ). From this graph, we can see that the effects on 6th graders' math achievement are moderate to small at the lower levels of past performance, with effect sizes ranging from 0.30 to 0.19 for scores ranging from 100 to 220 for treatment school students in comparison to control school students. The average 6th grade control students' math scores ranging from 100 to 220 were approximately 10 to 6 score points lower than the average 6th grade treatment students (see Figure 3). Smaller effect sizes ranging from 0.18 to 0.11 were observed for treatment students in comparison to control students for math scores ranging from 230 to 300. Sixth graders in the control schools with math scores ranging from 230 to 300 were approximately 6 to 4 points lower than their counterparts in the treatment schools. For students with above average and higher levels of past math achievement (i.e., scale scores above 300), the effects on treatment and control school 6th graders' math performance tended to be negligible, with some advantage in favor of the control schools at the very highest level of past math achievement.

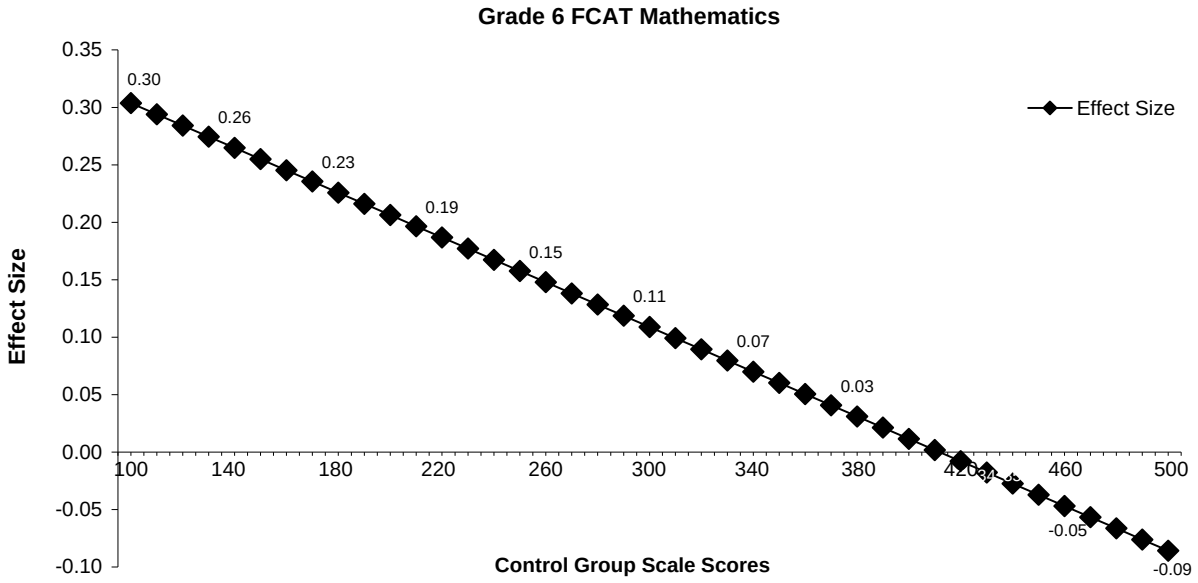
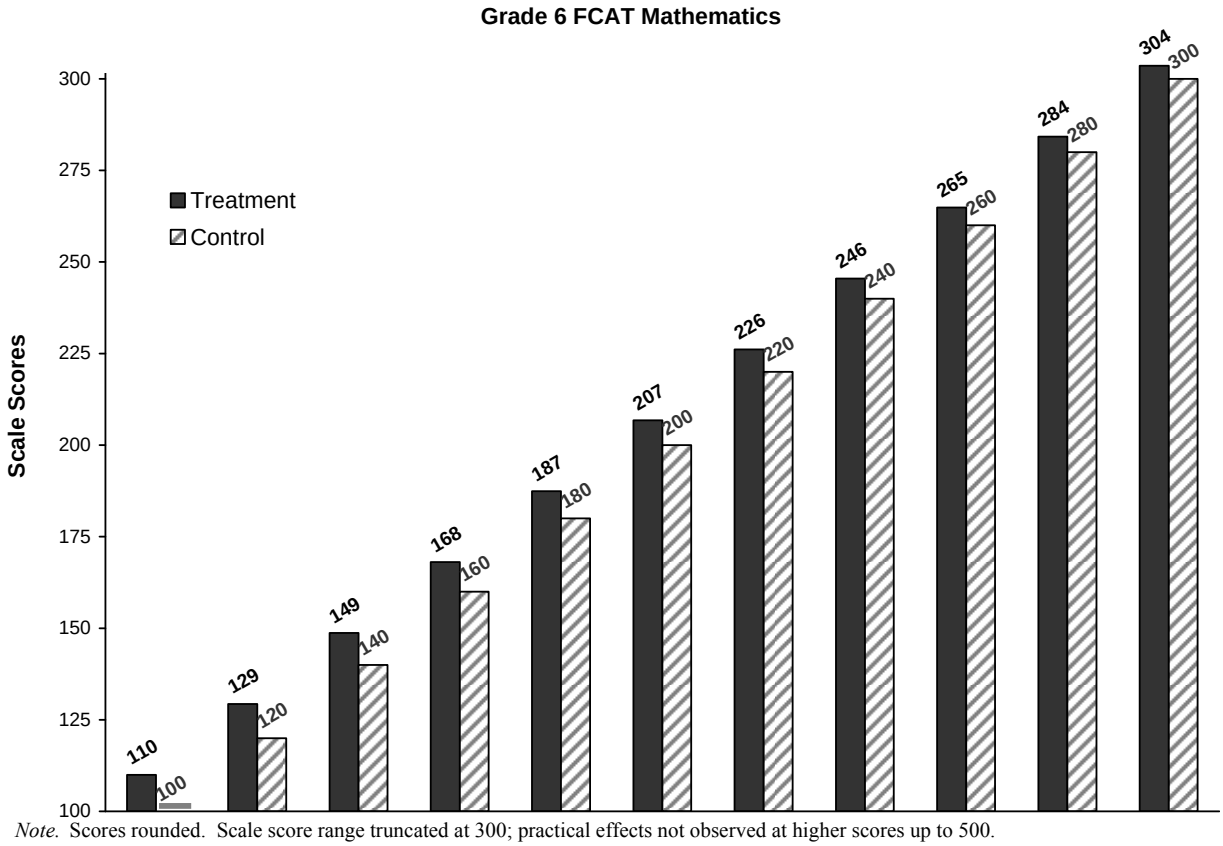



*Figure 2.* Effect Sizes between Treatment and Control for Grade 6 FCAT Mathematics as a Function of Prior Performance




*Note.* Scores rounded. Scale score range truncated at 300; practical effects not observed at higher scores up to 500.

*Figure 3.* Grade 6 FCAT Mathematics Scores for Lower Performing Treatment and Control School Students as a Function of Prior Performance

FCAT Reading scale scores were examined with a two-level model with the same form as the one used for mathematics. The past achievement variable was defined as the previous year's FCAT Reading scale score (grade 5), again centered so that zero corresponded to the lowest level of past achievement. The results for this analysis are also provided in Table 29. With FCAT Reading, the difference between treatment and control school students at the lowest level of past achievement was not statistically significant ( $\gamma_{01}$ = -1.99, $t$(11000) = -0.59, $p$ = .5558, nor was there evidence that the difference between treatment and control school students was moderated by past achievement ( $\gamma_{11}$ = 0.0134, $t$(11000) = 1.05, $p$ = .2952).

The above data show that the TGFD treatment intervention was associated with FCAT Mathematics achievement. An interaction or moderating effect of prior FCAT Mathematics performance was evident, such that those treatment students with below average 5th grade FCAT Mathematics scores performed better on 6th grade FCAT Mathematics as compared to their counterparts among the control students. However among 6th graders who performed above average on 5th grade FCAT Mathematics, little if any difference between treatment and control students was discernible. The data for FCAT Reading show no association between treatment and control school students at the lowest level of past achievement, nor any interaction effects for 6th graders' reading performance between the treatment/control conditions as a function of prior achievement.

## Discussion and Conclusions

The *Too Good for Drugs* (TGFD) is a school-based prevention program designed 1) to prevent or diminish cigarette smoking, alcohol consumption and marijuana use among school age children, and 2) to reduce risk factors and enhance protective factors that strengthen resiliency behavior in children and adolescents related to alcohol, tobacco and other drug use.

An evaluation of the TGFD was undertaken to determine the program's effectiveness on the major usage outcomes and R&P outcomes; and to determine if the program's impact is moderated by student risk level. The association between the TGFD treatment and certain school performance indicators, including 6th grade mathematics and reading performance, was also investigated. A randomized trial study was implemented with a sample of 6th graders from a large Florida school district. The evaluation study used a stratified randomized treatment-control group design, whereby 40 middle schools were paired on the basis of key school demographics, then randomly assigned to either the treatment or control conditions. The 6th graders in the 20 treatment schools and the 20 control schools were surveyed in their classrooms on an identical schedule: prior to implementing the treatment intervention, after delivery of the treatment intervention, and six months after the end of treatment. Efforts were made to determine the extent to which the TGFD program was implemented as planned, and to gather evidence on the reliability and validity of the survey responses. Efforts were also made to check for possible contamination of the treatment/control conditions through an accounting of within school activities at each participating school.

Across the three survey periods, the treatment and control groups remained highly similar in terms of key demographic characteristics, including gender, ethnicity, free or reduced lunch program services, limited English proficiency services, and exceptional education services. The number of 6th graders surveyed included 10,762 on the Pre-Survey, 10,513 on the Post-Survey, and 10,163 on the 6-month Follow-up Survey.

While all adolescents to a degree may be vulnerable to certain social, psychological and behavioral influences to use drugs, most vulnerable may be those who indicate an early predisposition to experiment with one or more ATOD substances. Students who are early

ATOD users have been found to be more likely than nonusers to intensify their drug involvement and to express other problem behaviors over time (DuRant et al., 1999; Ellickson, Tucker, & Klein, 2001, 2003; Hawkins et al., 1992; Lynsky et al., 2003). Some researchers, therefore, have posited that student risk level may be a critical moderator in determining the efficacy of school-based drug prevention programs; i.e., that such programs may have their greatest impact on those students at elevated risk for using and abusing drugs (Botvin et al., 2001a; Eisen et al., 2002; Ellickson, Tucker, and Klein, 2003; Hawkins et al., 1999; Longshore et al., 2007; Lynsky et al., 2003; Sussman et al., 1998). In this study we identified three levels of risk among the participating 6th graders—those who reported that at age 10 or younger they had tried 1) none of the three ATOD substances of cigarettes, alcohol, and marijuana (low risk), 2) one of the three ATOD substances (moderate risk), and 3) two or more of the three ATOD substances (high risk).

The results we found show the effectiveness of the TGFD school-based program in diminishing reported smoking behavior, alcohol consumption, binge drinking and marijuana use among high risk 6th graders, and impacting those R&P factors to boost student resiliency related to drug use (Arthur et al., 2002; Donaldson et al., 1995; Hawkins et al., 1999; Komro et al., 2001; Longshore et al., 2007). Based on the post-survey effect sizes (ESs) for the 30-day usage outcomes (.56 to 1.03) and for the R&P factors (.33 to .76), the short-term impact of the program for the high risk students was broad and substantive. The positive effects, though attenuated by time, were still present six months after treatment for the 30-day usage outcomes (ESs of .30 to .65), and for most of the R&P outcomes (intent to use ATOD, peer resistance, bonding with prosocial peers, harmful effects of drugs, and ATOD attitudes), with ESs of .30 to .63. Also, the reported use of cigarettes, alcohol and marijuana over the past year showed a diminution (ESs of .26 to .57) favoring high risk takers receiving the program. These findings are noteworthy not

only for their strength and relative stability over time, but also because they are based on an ethnically diverse sample from a large district containing a mix of urban, suburban and rural schools.

The results show that the TGFD program had some impact on the low and moderate risk students, but the effects, in comparison to those for high risk students, were more limited in both scope and time. For moderate risk students, the post-survey produced small but significant ESs (.14 to .19) on three of the 30-day ATOD outcomes (drinking, binge drinking, and marijuana use), as well as on two of the R&P factors: peer resistance (.31) and self-efficacy (.19). However, these effects did not carry over to six months later, nor were there significant effects for moderate risk students' reported use of cigarettes, alcohol or marijuana over the past year. For low risk students, the post-survey produced no significant treatment effects on the ATOD usage outcomes, but did produce significant effects on three of the R&P outcomes: goals and decisions (ES = .20), peer resistance (ES = .23), and self-efficacy (ES = .23). Significant effects on peer resistance and self-efficacy carried over to six months later; but again, no effects were found on the ATOD usage outcomes.

All of the effects that were produced over the two survey periods (post survey and 6-month follow-up) favor the treatment students, such that the results show a general *suppression* effect on students' reported ATOD usage, and a general *strengthening* effect on those R&P factors that are considered important in promoting adolescents' resilience to inappropriate drug use. These findings underscore the efficaciousness of the TGFD curriculum as a 6$^{th}$ grade intervention, especially for students identified as being at risk for early experimentation with drugs. Across the three survey periods (pre-survey, post-survey, 6-month follow-up), the treatment and control groups remained highly similar in terms of demographic characteristics.

Although the students categorized as "high risk" constitute a minority of the study sample, they represent the early experimenters in drug use within the sample, and therefore, those adolescents within the sample that may be most vulnerable to the influence of alcohol, tobacco or other drugs. According to a number of sources, early ATOD users over time are more likely than nonusers to intensify their drug involvement (DuRant et al., 1999; Eisen et al., 2002; Ellickson, Tucker & Klein, 2001, 2003; Flay et al., 2007; Hawkins et al., 1999; Longshore et al., 2007; Lynsky et al., 2003; Perry et al., 2002). Therefore, repressing their willingness to experiment during the treatment and up to six months afterward can be considered a critical step in the effort that Donaldson et al. (1996) termed "inoculation" against direct and indirect social pressure to use tobacco, alcohol and other drug substances.

The bulk of the 6th grade sample indicated no experimentation with cigarettes, alcohol or marijuana at age 10 or younger (prior to entering grade 6), but this may be because they have had relatively few opportunities to experiment. As they continue to make their way through the middle school years and into high school, they could begin to avail themselves of such opportunities. According to the Monitoring the Future study by the University of Michigan's Institute for Social Research (Johnston et al., 2011), use of alcohol, cigarettes, marijuana and other drugs among young people show steady increases as they pass through the middle school years and into high school. For example, the study reports that 36% indicated having consumed alcohol by 8th grade and fully 71% by the end of high school. Such trends suggest that appropriate interventions later in middle school might produce a suppression effect even for "low risk" students. As a case in point, Heyne and Bogner (2009) found that an intervention designed to strengthen resistance self-efficacy among 8th graders was effective for the 78% of students that were in the lowest of their four risk groupings. Later middle school interventions might also

become increasingly effective for the portion of our sample that we consider "moderate risk". In a study of problem behaviors associated with early smoking, Ellickson et al. (2001) found that, as compared to nonsmokers, their moderate risk takers (those who had tried cigarettes only once or twice by grade 7) were more likely to engage in relatively severe types of drug use, including binge drinking, at grade 7.

In an investigation of the TGFD program in relation to school performance indicators, the treatment and control schools were compared in terms of student absences from school, number of suspensions (inclusive of in-school and out-of-school), and academic achievement. No differences between treatment and control schools were found for average number of days students were absent, nor for average number of suspensions assigned to students during the school year. However, in an examination of the TGFD program in terms of school achievement, an effect was found between treatment/control conditions for students with below average performance on the Florida Comprehensive Assessment Test (FCAT) Mathematics. Specifically, treatment 6$^{th}$ graders with below average 5$^{th}$ grade FCAT Mathematics scores performed significantly better on 6$^{th}$ grade FCAT Mathematics as compared to their counterparts among control students; those with above average scores showed no difference. No differences were found on FCAT Reading.

Poor school achievement has been identified in previous research to be one of the major risk factors for drug use among adolescents (Hawkins et al., 1992; Pollard et al., 1999). The findings in this study related to low FCAT mathematics achievement are encouraging, for they suggest that developmentally appropriate interventions early in middle school can have an impact in boosting achievement in performance areas critical to overall academic and personal success. Future research is warranted to investigate more specifically the source of the

facilitative FCAT mathematics effect. However, in a scientific review of SEL interventions (those containing core social-emotional competencies of self-awareness, self-management, social awareness, relationship skills, and responsible decision making), Payton et al. (2008) found that SEL programs produced significant gains in academic performance as represented by standardized test scores and school grades. In the TGFD intervention, the curricular emphasis on SEL-related competencies (e.g., goal setting, decision making, identifying and managing emotions, effective communication, self-efficacy, prosocial bonding, and peer resistance), together with the positive results on the R&P outcomes particularly for high risk students, suggest these elements as a likely source of the positive effect on low FCAT mathematics performance. Payton et al. (2008) also noted that positive academic performance effects were associated with those SEL interventions that used active forms of learning and that reported no implementation problems, two more characteristics of the present TGFD intervention.

The cogency of the findings in the present study is buttressed by several factors. First, the pairing of the 40 middle schools on key demographic variables and the subsequent random assignment of the schools within pairs to treatment and control conditions helped assure a level playing field for examining program effects. No statistically significant difference was observed on the pre-survey ATOD and R&P outcomes between the treatment and control school students, with the exception of high risk control students reporting lower rates of smoking cigarettes in the past year in comparison to high risk treatment students before program delivery. Second, potential contamination of the treatment and control groups was checked via the Within School Activities Form (WSAF) completed by school contacts from the 40 participating schools. None of the schools were found to have had in place anti-drug programs apart from the TGFD, and

specific presentations—on such topics as bully prevention, cyber awareness or smoking—were found about equally in both treatment and control schools.

Third, the findings of the study are given substance by the fact that the subcomponents of the SBRPFS, the source of the survey responses, were found to have acceptable estimates of internal consistency reliability based on data from the main study, and both internal consistency and test-retest reliability based on data from a pilot study with a sample of 6th graders ethnically comparable to those in the main study. Furthermore, these SBRPFS subcomponents showed evidence of concurrent validity in relation to the subcomponents of other widely used drug usage and R&P instruments. Also, an investigation of internal factor structure on the R&P items showed that the items of six of the seven R&P subscales loaded in a manner consistent with a meaningful interpretation of those subscales.

In addition to investigating the reliability and validity of the SBRPFS, we gave careful attention to the manner in which the survey instrument was administered in both treatment and control classrooms. An orientation was given to the survey administrators (SAs) of the SBRPFS before each of the three major data gathering periods in the study. Also, the SAs were required to follow a script during survey administration to assure that proper procedures were followed in all classrooms. Finally, observations were carried out in both treatment and control schools to document that the SAs remained on script and to determine whether there were occurrences during the survey administration that might have compromised the integrity of the students' responses. A number of these observations were paired and these paired observations were found to be highly consistent. The results of the observations show that the survey instrument was appropriately administered throughout both treatment and control classrooms, and that in no

treatment or control classroom was there an occurrence that the observers felt could have compromised the integrity of the survey data.

Fidelity of implementation has become a major topic of concern in the evaluation of drug intervention programs (Brandon et al., 2008; Century et al., 2010; Dusenbury et al., 2003; Reichardt, 2011). Century et al. (2010) calls fidelity of implementation the "black box" in the evaluation of intervention effectiveness, and Reichardt (2011) insists that adequate program descriptions include a delineation of the extent to which the treatment was implemented as planned. Pertinent to the fidelity of implementation of the 10 TGFD lessons, the TGFD instructors completed the appropriate Lesson Implementation Checklist (LIC) to give their perspective on the extent to which the lessons had been delivered. For each lesson, LICs were completed for about 285 classes. Evidence of lesson implementation was also gathered through a representative sample of unannounced in-class observations conducted by the observation team, using an observation checklist that paralleled the LIC. A number of these observations were paired and these paired observations were found to be highly consistent. In addition, both the observers and the regular teachers in whose classes the program was implemented rated the TGFD teachers on their instructional delivery and the responsiveness of the students.

The data from the three sources described above provide collaborative evidence that the TGFD treatment was delivered with consistent quality and completeness, and in such a way that actively and successfully engaged the participating students. Both the TGFD teachers and the on-site observers were in agreement that over 95% of the lesson activity statements were completed or mostly completed during lesson delivery. Also, the TGFD instructors were rated highly by both the observers and the regular classroom teachers for their instructional skills and their involvement of students in the learning process. The insights derived from the

implementation data provide a firm basis of confidence for interpreting the results on the major outcomes of the TGFD evaluation study.

In summary, the results of the study present compelling evidence of TGFD program effectiveness for high risk students across all ATOD use outcomes and across almost all R&P outcomes. The evidence to date suggests that the program is having a *suppression* effect on students' reported use of ATOD substances and a *strengthening* effect on R&P factors, particularly among high risk students. The evidence also suggests that these program effects extend across time to a point six months after treatment. The results show more limited effects for low and moderate risk students. Positive effects on 6$^{th}$ grade FCAT mathematics achievement were found for treatment students, as compared to control students, who had performed below average on the previous year's FCAT mathematics. The survey measures used in the study evidenced acceptable reliability and validity, and data from TGFD teachers, on-site observers and regular classroom teachers indicate that the TGFD program actively engaged students in the learning process and that the program was well implemented.

## References

Alwin, D. F. (2007). *Margins of error: A study of reliability in survey measurement*. Hoboken, NJ: John Wiley & Sons.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Anderson, S. W., & Moore, P. A. (2009). The impact of education and school-based counseling on children's and adolescents' views of substance abuse. *Journal of Child & Adolescent Substance Abuse*, *18*, 16-23. doi:10.1080/15470650802526526

Arthur, M. W., Hawkins, J. D., Pollard, J. A., Catalano, R. F., & Baglioni, Jr., A. J. (2002). Measuring risk and protective factors for substance use, delinquency, and other adolescent problem behaviors: The communities that care youth survey. *Evaluation Review*, *26*, 575-601. doi:10.1177/0193841X0202600601

Bacon, T. P. (2000). The effects of the too good for drugs II drug prevention program on students' substance use intentions and risk and protective factors. *Research Bulletin of the Florida Educational Research Council*, *31* (3, 4).

Baker T. L., Jarman T. R. & Bacon T. P. (2009). Online Compliance Management System. St. Petersburg, FL: Leap Frog Group, Inc.

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1985). *Social foundations of thought and action*. Englewood Cliffs, NJ: Prentice-Hall.

Botvin, G. J., Griffin, K. W., Diaz, T., & Ifill-Williams, M. (2001a). Drug abuse prevention among minority adolescents: Posttest and one-year follow-up of a school-based preventive intervention. *Prevention Science*, *2*(1), 1-13. doi:10.1023/A:1010025311161

Botvin, G. J., Griffin, K. W., Diaz, T., & Ifill-Williams, M. (2001b). Preventing binge drinking during early adolescence: One- and two-year follow-up of a school-based preventive intervention. *Psychology of Addictive Behavior*, *15*, 360-365. doi:10.1037//0893-164X.15.4.360

Brandon, P. R., Taum, A. K. H., Young, D. B., Pottenger III, F. M., & Speitel, T. W. (2008). The complexity of measuring the quality of program implementation with observations: The case of middle school inquiry-based science. *American Journal of Evaluation*, *29*, 235-250. doi:10.1177/1098214008319175

Brown, S. A., Tapert, S. F., Granholm, E., & Delis, D. C. (2000). Neurocognitive functioning of adolescents: Effects of protracted alcohol use. *Alcoholism: Clinical and Experimental Research*, *24*, 164-171.

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, NY: Cambridge University Press.

Catalano, R. F., Mazza, J. J., Harachi, J. W., Abbott, R. D., Haggerty, K. P., & Fleming, C. B. (2003). Raising healthy children through enhancing social development in elementary school: Results after 1.5 years. *Journal of School Psychology*, *41*, 143-164. doi:10.1016/S0022-4405(03)00031-1

Centers for Disease Control and Prevention (2010, June). *Youth risk behavior surveillance—United States, 2009*. Surveillance Summaries, MMWR 2010;59(No. SS-5). Retrieved from http://www.cdc.gov/yrbss.

Century, J., Rudnick, M., & Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evaluation*, *31*, 199-218. doi:10.1177/1098214010366173

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd Ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Coyle, K. K., Russell, L. A., Shields, J. P., & Tanaka, B. A. (2007, October). *Summary report: Collecting data from children ages 9-13*. ETR Associates. Retrieved from http://www.preteenalliance.org/research.html

Crocker, L. & Algina, J. (1986). *Introduction to classical & modern test theory*. Orlando, FL: Harcourt Brace Jovanovich.

Cuijpers, P. (2002). Effective ingredients of school-based drug prevention programs: A systematic review. *Addictive Behaviors*, *27*, 1009-1023. doi:10.1016/S0306-4603(02)00295-2

Dahlberg, L., Toal, S., Swahn, M., & Behrens, C. (Eds.). (2005). *Measuring violence-related attitudes, behaviors, and influences among youths: A compendium of assessment tools* (2nd Ed.). Altanta, GA: National Center for Injury Prevention, Centers for Disease Control and Prevention. Retrieved from http://www.cdc.gov/measuringviolence-relatedattitudes

De Gruijter, D. N. M., & van der Kamp, L. J. T. (2008). *Statistical test theory for the behavioral sciences*. Boca Raton, FL: Taylor & Francis.

Donaldson, S. I. (1995). Peer influence on adolescent drug use: A perspective from the trenches of experimental evaluation research. *American Psychologist*, *50*, 801-802. Retrieved from http://web.ebscohost.com

Donaldson, S. I., Graham, J. W. & Hansen, W. B. (1994). Testing the generalizability of intervening mechanism theories: Understanding the effects of adolescent drug use prevention interventions. *Journal of Behavioral Medicine*, *17*, 195-216. doi:10.1007/BF01858105

Donaldson, S. I., Graham, J. W., Piccinin, A. M., & Hansen, W. B. (1995). Resistance-skills training and onset of alcohol use: Evidence for beneficial and potentially harmful effects in public schools and in private catholic schools. *Health Psychology*, *14*, 291-300. Retrieved from http://www.web.ebscohost.com

Donaldson, S. I., Sussman S., Mackinnon, D. P., Severson, H. H., Glynn, T., Murray, D. M., & Stone, E. J. (1996). Drug abuse prevention programming: Do we know what content works? *American Behavioral Scientist*, 39, 868-883. doi:10.1177/0002764296039007008

Donaldson, S. I., Thomas, C. W., Graham, J. W., Au, J. G., & Hansen, W. B. (2000). Verifying drug abuse prevention program effects using reciprocal best friend reports. *Journal of Behavioral Medicine*, *23*, 585-601. doi:10.1023/A:1005559620739

Donohue, B., Hill, H. H., Azrin, N. H., Cross, C., & Strada, M. J. (2007). Psychometric support for contemporaneous and retrospective youth and parent reports of adolescent marijuana use frequency in an adolescent outpatient treatment population. *Addictive Behaviors*, *32*, 1787-1797. doi:10.1016/j.addbeh.2006.12.005

DuRant, R. H., Smith, J. A., Kreiter, S. R., & Krowchuk, D. P. (1999). The relationship between early age of onset of initial substance use and engaging in multiple health risk behaviors among young adolescents. *Archives of Pediatrics & Adolescent Medicine*, *153*, 286-291. Retrieved from http://ovidsp.tx.ovid.com

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, *82*, 405-432. doi:10.1111/j.1467-8624.2010.01564.x

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research*, *18*, 237-256. doi:10.1093/her/18.2.237

Eisen, M., Zellman, G. L., Massett, H. A., & Murray, D. M. (2002). Evaluating the Lions-Quest "skills for adolescence" drug education program: First-year behavior outcomes. *Addictive Behaviors*, *27*, 619-632. doi:1016/S0306-4603(01)00197-6

Eisen, M., Zellman, G. L., & Murray, D. M. (2003). Evaluating the Lions-Quest "skills for adolescence" drug education program: Second-year behavior outcomes. *Addictive Behaviors*, *28*, 883-897. doi:10.1016/S0306-4603(01)00292-1

Ellickson, P. L., McCaffrey, D. F., Ghosh-Dastidar, B., & Longshore, D. L. (2003). New inroads in preventing adolescent drug use: Results from a large-scale trial of Project ALERT in middle schools. *American Journal of Public Health*, *93*, 1830-1836. Retrieved from http://ajph.aphapublications.org/cgi/content/full/93/11

Ellickson, P. L., Tucker, J. S., & Klein, D. J. (2001). High-risk behaviors associated with early smoking: Results from a 5-year follow-up. *Journal of Adolescent Health*, *28*, 465-473. doi:10.1016/S1054-139X(00)00202-0

Ellickson, P. L., Tucker, J. S., & Klein, D. J. (2003). Ten-year prospective study of public health problems associated with early drinking. *Pediatrics*, *111*, 949-955. Retrieved from http://pediatrics.aappublications.org

Flay, B. R. (2009). The promise of long-term effectiveness of school-based smoking prevention programs: A critical review of reviews. *Tobacco Induced Diseases*, 5(7). doi:10.1186/1617-9625-5-7

Flay, B. R., Phil, D., DuBois, D. L., & Ji, P. (2007). *Progress report of the randomized trial of Positive Action in Chicago: End of third year intervention (Grade 5, Spring, 2007)*. Institute for Education Sciences, U. S. Department of Education. Retrieved from http://ies.ed.gov/pubsearch/

Florida Department of Health (2008). *Florida youth tobacco survey*. Retrieved from http://www.floridayouthtobaccosurvey.gov

Furr, R. M., & Bacharach, V. R. (2008). *Psychometrics: An introduction*. Los Angeles, CA: Sage.

Gandhi, A. G., Murphy-Graham, E., Petrosino, A., Chrismer, S. S., & Weiss, C. H. (2007). The devil is in the details: Examining the evidence for "proven" school-based drug abuse prevention programs. *Evaluation Review*, *31*, 43-74. doi:10.1177/0193841X06287188

Ghiselli, E. E., Campbell, J. P., & Zedeck, S. (1981). *Measurement theory for the behavioral sciences*. San Francisco, CA: W. H. Freeman.

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., Paus, T., Evans, A. C., & Rapoport, J. L. (1999). Brain development during childhood and adolescence: A longitudinal mri study. *Nature Neuroscience*, *2*, 861-863. Retrieved from http://neurosci.nature.com

Greenberg, M. T., Weissberg, R. P., O'Brien, M. U., Zins, J. E., Frederick, L., Resnik, H., & Elias, M. J. (2003). Enhancing school-based prevention and youth development through coordinated social, emotional, and academic learning. *American Psychologist*, *58*, 466-474. doi:10.1037/0003-066X.58.6-7.466

Guo, J., Hawkins, J. D., Hill, K. G., & Abbott, R. D. (2001). Childhood and adolescent predictors of alcohol abuse and dependence in young adulthood. *Journal of Studies on Alcohol*, *62*, 754-762. Retrieved from http://find.galegroup.com

Hansen, W. B. (1992). School based substance abuse prevention: A review of the state of the art in curriculum, 1980-1990. *Health Education Research*, *7*, 403-430. Retrieved from http://her.oxfordjournals.org

Hawkins, J. D., Catalano, R. F., Kosterman, R., Abbott, R., & Hill. K. G. (1999). Preventing adolescent health-risk behaviors by strengthening protection during childhood. *Archives of Pediatrics & Adolescent Medicine*, *153*, 226-234. Retrieved from http://ovidsp.tx.ovid.com

Hawkins, J. D., Catalano, R. F., & Miller, J. (1992). Risk and protective factors for alcohol and other drug problems in adolescence and early adulthood: Implication for substance abuse prevention. *Psychological Bulletin*, *112*, 64-105. Retrieved from http://web.ebscohost.com

Hawkins, J. D. & Weis, J. G. (1985). The social development model: An integrated approach to delinquency prevention. *The Journal of Primary Prevention*, *6*, 73-97. doi:10.1007/BF01325432

Hecht, M. L., Elek, E., Wagstaff, D. A., Kam, J. A., Marsiglia, F., Dustman, P., Reeves, L., & Harthun, M. (2008). Immediate and short-term effects of the 5th grade version of the keepin' it real substance use prevention intervention. *Journal of Drug Education*, *38*, 225-251. doi:10.2190/DE.38.3.c

Hecht, M. L., Marsiglia, F. F., Elek, E., Wagstaff, D. A., Kulis, S., Dustman, P., & Miller-Day, M. (2003). Culturally grounded substance use prevention: An evaluation of the keepin' it R.E.A.L. curriculum. *Prevention Science*, *4*, 233-248. doi:10.1023/A:1026016131401

Heyne, T., & Bogner, F. X. (2009). Strengthening resistance self-efficacy: Influence of teaching approaches and gender on different consumption groups. *Journal of Drug Education*, *39*, 439-457. doi:10.2190/DE.39.4.f

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65-70.

Hwang, M. S., Yeagley, K. L., & Petosa, R. (2004). A meta-analysis of adolescent psychosocial smoking prevention programs published between 1978 and 1997 in the United States. *Health Education & Behavior*, *31*, 702-719. doi:10.1177/1090198104263361

International Survey Associates (2009a). *Communities that care youth survey*. Retrieved from http://www.pridesurveys.com.

International Survey Associates (2009b). *PRIDE risk & protective factor questionnaire*. Retrieved from http://www.pridesurveys.com

Jessor, R. (1982). Problem behavior and developmental transition in adolescence. *Journal of School Health*, *52*, 295-300. Retrieved from http://onlinelibrary.wiley.com

Jessor, R., & Jessor, S. L. (1977). *Problem behavior and psychosocial development: A longitudinal study of youth*. NY: Academic Press.

Johnson, P. B., & Richter, L. (2004). Research note. What if we're wrong? Some possible implications of systematic distortions in adolescents' self-reports of sensitive behaviors. *Journal of Drug Issues*, *34*, 951-970. Retrieved from http://vnweb.hwwilsonweb.com

Johnston, L. D., O'Malley, P. M., Bachman, J. G., & Schulenberg, J. E. (2011). National results on adolescent drug use: Overview of key findings, 2010. University of Michigan: Institute of Social Research. Retrieved from http://www.monitoringthefuture.org/pubs.html

The Joint Committee on Standards for Educational Evaluation (1994). *The program evaluation standards: How to assess evaluations of educational programs* (2nd. ed.). Thousand Oaks, CA: Sage.

Komro, K. A., Perry, C. L., Williams, C. L., Stigler, M. H., Farbakhsh, K., & Veblen-Mortenson, S. (2001). How did Project Northland reduce alcohol use among young adolescents? Analysis of mediating variables. *Health Education Research*, *16*, 59-70. doi:10.1093/her/16.1.59

Landry, S. H., Smith, K. E., & Swank, P. R. (2009). New directions in evaluating social problem solving in childhood: Early precursors and links to adolescent social competence. *New Directions in Child and Adolescent Development*, *123*, 51-68. doi:10.1002/ed.235

Leffert, N., Benson, P. L., & Roehlkepartain, J. L. (1996). *Starting out right: Developmental assets for children*. Search Institute. Retrieved from http://www.search-institute.org/research/assets

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *The American Psychologist*, *48*, 1181-1209. Retrieved from http://web.ebscohost.com

Longshore, D., Ellickson, P. L., McCaffrey, D. F., & St. Clair, P. A. (2007). School-based drug prevention among at-risk adolescents: Effects of ALERT Plus. *Health Education and Behavior*, *34*, 651-668. doi:10.1177/1090198106294895

Lynskey, M. T., Heath, A. C., Bucholz, K. K., Slutske, W. S., Madden, P. A., Nelson, E. C., Statham, D. J., & Martin, N. G. (2003). Escalation of drug use in early-onset cannabis users vs co-twin controls. *Journal of the American Medical Association*, *289*, 427-433. Retrieved from http://ovidsp.tx.ovid.com

Maguin, E., & Loeber, R. (1996). Academic performance and delinquency. In M. Tonry (Ed.), *Crime and justice: A review of research* (Vol. 20, pp.145-264). Chicago: University of Chicago Press.

McDonald, R. P. (1999). *Test Theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum Associates.

Metze, L. (October, 2000). *The Pride questionnaire for grades 6 – 12: Validity and reliability Study*. Retrieved from http://www.pridesurveys.com/supportfiles/tr99612.pdf

Meyer, P. (2010). *Reliability*. New York, NY: Oxford University Press.

Murphy, D. A., Durako, S., Muenz, L. R., & Wilson, C. M. (2000). Marijuana use among HIV-positive and high-risk adolescents: A comparison of self-report through audio computer-assisted self-administered interviewing and urinalysis. *American Journal of Epidemiology*, *152*, 805-813. doi:10.1093/aje/152.9.805

National Institute of Mental Health (2001). Teenage brain: A work in progress (fact sheet). Retrieved from www.nimh.nih.gov/publicat/teenbrain.cfm

Payton, J., Weissberg, R. P., Durlak, J. A., Dymnicki, A. B., Taylor, R. D., Schellinger, K. B., & Pachan, M. (2008). *The positive impact of social and emotional learning for kindergarten to eighth-grade students: Findings from three scientific reviews*. Chicago, IL: Collaborative for Academic, Social and Emotional Learning. Retrieved from http://www.casel.org/downloads/PackardTR.pdf

Perry, C. L, & Jessor, R. (1983). Doing the cube: Preventing drug abuse through adolescent health promotion. In J. P. Ludford (Ed.), *Preventing adolescent drug abuse: Intervention strategies* (Research Monograph No. 47). Rockville, MD: National Institute on Drug Abuse.

Perry, C. L, Williams, C. L., Komro, K. A. Veblen-Mortenson, S., Stigler, M. H., Munson, K. A., & Forster, J. L. (2002). Project Northland: Long-term outcomes of community action to reduce adolescent alcohol use. *Health Education Research*, *17*, 117-132. doi:10.1093/her/17.1.117

Pollard, J. A., Hawkins, J. D., & Arthur, M. W. (1999). Risk and protection: Are both necessary to understand diverse behavioral outcomes in adolescence? *Social Work Research*, *23*, 145-158. Retrieved from http://web.ebscohost.com

Princeton Health Press (2004). Botvin *Life skills training questionnaire, middle school*. Retrieved from http://www.lifeskillstraining.com/

Project ALERT (2009). *Survey of student attitudes and behavior*. Retrieved from http://www.projectalert.com/

Reichardt, C. S. (2011). Evaluating methods for estimating program effects. *American Journal of Evaluation*, *32*, 246-272. doi:10.1177/1098214011398954h

Restak, R. (2001). *The Secret Life of the Brain*. Washington, DC: National Academies Press.

Ringwalt, C. L., Pankratz, M. M., Hansen, W. B., Dusenbury, L., Jackson-Newsom, J., Giles, S. M., & Brodish, P. H. (2007). The potential of coaching as a strategy to improve the effectiveness of school-based substance use prevention curricula. *Health Education & Behavior*, *36*, 696-710. doi:10.1177/1090198107303311

Rocky Mountain Behavioral Science Institute (2003). *America drug and alcohol survey*. Retrieved from http://www.rmbsi.com

Rosenthal, R. & Rubin, D. B. (1982). A simple general purpose display of magnitude of experimental effect. *Journal of Educational Psychology*, *74*, 166-169.

Rutter, M. (1990). Psychosocial resilience and protective mechanisms. In J. Rolf, A. S. Matsen, D. Cicchetti, K. H. Nucchterlein, U S. Weintraub (Eds.), *Risk and protective factors in the development of psychopathology* (pp. 181-214). Cambridge, UK: Cambridge University Press.

Ryan, A. M., & Shim, S. S. (2008). An exploration of young adolescents' social achievement goals and social adjustment in middle school. *Journal of Educational Psychology*, *100*, 672-687. doi:10.1037/0022-0663.100.3.672

Scales, P. C. & Leffert, N. (2004). *Developmental assets: A synthesis of the scientific research on adolescent development* (2nd ed.). Minneapolis: Search Institute. Retrieved from http://www.search-institute.org/research/assets

Schwinn, T. M., Schinke, S. P., & Di Noia, J. (2010). Preventing drug abuse among adolescent girls: Outcome data from an internet-based intervention. *Prevention Science*, *11*, 24-32. doi:10.1007/s11121-009-0146-9

Skara, S., & Sussman, S. (2003). A review of 25 long-term adolescent tobacco and other drug use prevention program evaluations. *Preventive Medicine*, *37*, 451-474. doi:10.1016/S0091-7435(03)00166-X

Solbergsdottir, E., Bjornsson, G., Gudmundsson, L. S., Tyrfingsson, T., & Kristinsson, J. (2004). Validity of self-reports and drug use among young people seeking treatment for substance abuse or dependence. *Journal of Addictive Diseases*, *23*, 29-38. doi:10.1300/J069v23n01_03

Steven, J. (1996). *Applied multivariate statistics for the social sciences* (3rd Ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Substance Abuse and Mental Health Services Administration (SAMHSA), Center for Substance Abuse Prevention (2003, February). *Core measures initiative: Phase I recommendations*. Retrieved from http://addictionstudies.dec.uwi.edu/Documents/ep

Substance Abuse and Mental Health Services Administration (SAMHSA), Center for Substance Abuse Prevention (1999). *National youth survey baseline questionnaire, 12-18 version*. Retrieved from http://emt.org/userfiles/NYS_Baseline_12-18Version

Sussman, S., Dent, C. W., Stacy, A. W., & Craig, S. (1998). One-year outcomes of Project Towards No Drug Abuse. *Preventive Medicine*, *27*, 632-642. doi:10.1006/pmed.1998.0338

Tobler, N. S. (2000). Lessons learned. *The Journal of Primary Prevention*, *20*, 261-274. doi:10.1023/A:10213626207401

Tobler, N. S., & Stratton, H. H. (1997). Effectiveness of school-based drug prevention programs: A meta-analysis of the research. *The Journal of Primary Prevention*, *18*, 71-128. doi:10.1023/A:1024630205999

Weiss, C. H., Murphy-Graham, E., Petrosino, A, & Gandhi, A. G. (2008). The fairy godmother—and her warts: Making the dream of evidence-based policy come true. *American Journal of Evaluation*, *29*, 29-47. doi:10.1177/1098214007313742

Yoshikawa, H. (1994). Prevention as cumulative protection: Effects of early family support and education on chronic delinquency and its risks. *Psychological Bulletin*, *115*, 28-54.