

Implementation fidelity: the experience of the Adolescent Substance Abuse Prevention Study

Zili Sloboda*, Peggy Stephens, Amod Pyakuryal, Brent Teasdale, Richard C. Stephens, Richard D. Hawthorne[†], Jesse Marquette[†] and Joseph E. Williams

Abstract

While researchers have developed more effective programs and strategies to prevent the initiation of substance use and increasingly communities are delivering these interventions, determining the degree to which they are delivered as they were designed remains a significant research challenge. In the past several years, more attention has been given to implementation issues during the various stages of program development and diffusion. This paper presents the findings from a substudy of an evaluation of a newly designed middle and high school substance abuse prevention program, Take Charge of Your Life delivered by local Drug Abuse Resistance Education officer instructors. A key aspect of the study was to determine the extent to which implementation fidelity, using the measures of content coverage and appropriate instructional strategy, was associated with improvement in the program mediators of realistic normative beliefs, understanding the harmful effects of substance use and the acquisition of decision-making and resistance skills. Although it was found that higher fidelity was associated with better scores on some of the mediators, this was not a consistent finding. The mixed results are discussed within the context of the lesson activities themselves.

Institute for Health and Social Policy, The University of Akron, Akron, OH 44325-1915, USA

*Correspondence to: Z. Sloboda. E-mail: Zsloboda@aol.com

[†]Retired

Introduction

The March 2004 issue of *Prevention Science* included a number of articles based on presentations given at a conference jointly sponsored by the National Institute on Drug Abuse and the Center for Substance Abuse Prevention entitled, 'What Do Schools Really Think about Prevention Research? Blending Research and Reality' [1]. The authors of these papers listed challenges to the progress of prevention science including dissemination and diffusion of strategies found to be effective through rigorous designs and resultant concerns about implementation fidelity and adaptation [2–4].

The concern about implementation fidelity and adaptation or reinvention is based on the results of national studies which found that evidence- or research-based programs taken 'to scale' at the community level often are not implemented as they were designed and evaluated. Two recent national studies conducted in 1997 and 1999 indicate that although many schools are meeting mandates to implement evidence-based prevention programs, they are not doing so with fidelity. Hallfors and Godette [5] found in their survey of Safe and Drug Free School Coordinators, located in 104 school districts in 12 states, that only 19% were implementing curriculum with fidelity. In another study of staff who taught substance abuse prevention in a nationally representative sample of 1795 public and private schools, Ennett *et al.* [6] found that 62.2% taught effective content which included social influence or comprehensive life skills materials and activities and 17.4% used effective interactive delivery methods. Only 14.2% reported using effective content materials, activities and delivery methods [6].

Although fidelity of implementation is recognized as important [7,8], few studies have examined the relationship between level of fidelity and program outcomes. Tobler and Stratton [9] suggest that decreases in the effect sizes they found in their meta-analysis of school-based substance abuse prevention programs taken to scale may be due to implementation issues. Pentz and Trebow [10] found that children exposed to a program delivered by instructors who maintained high implementation fidelity had better outcomes than those exposed to the program delivered by instructors who implemented the program with low fidelity. Furthermore, children exposed to the program delivered by low implementers had better outcomes than children in the control condition. Botvin *et al.* [11] found that children who experienced 60% of the program had better outcomes compared with the control children. Hansen *et al.* [12] found that program integrity (the term used for fidelity) was significantly related to three of seven program mediators for a resistance skills program. While in a later pilot study of the All Stars Community Program, Hansen *et al.* [13] found that instruction with higher program fidelity had better outcomes.

As the field of prevention continues to disseminate theory- and research-based interventions, the combined issues of fidelity and reinvention will become increasingly important. However, as Dusenbury *et al.* [14] emphasize, although the field has fair agreement on the definition of fidelity ('... the degree to which teachers and other program providers implement programs as intended by the program developers', p. 240), there does not appear to be consensus regarding the specific dimensions of fidelity nor of their measurements. The authors list the five most frequently mentioned definitions as '... (1) strict adherence to methods or implementation that conforms to theoretical guidelines ..., (2) completeness and dosage of implementation, (3) the quality of program delivery (the way a teacher implements a program), (4) the degree to which participants are engaged and (5) program differentiation (the degree to which elements which would distinguish one type of program from another are

present or absent)' (p. 240). As variable as the definitions of fidelity are, their measures remain even more complex [15–22].

It is appropriate to point out that assessing implementation fidelity or adherence to the design and intent of an intervention is important during every phase of development from efficacy trials through to replication studies to assure that the research results can be related to the intervention and not to other factors. The most effective preventive interventions available today are based on strong theoretical or conceptual frameworks. These frameworks specify not only the targets for the intervention such as normative beliefs or skills building but also the processes and mechanisms through which participants of the intervention integrate and internalize desired attitudes and behaviors. Of particular importance is measuring the degree to which a newly developed intervention achieves both short- and long-term objectives as through mediational analyses and to assess adherence to delivery style, coverage of materials and participant involvement in the intervention processes. For this reason, for instance, in efficacy studies, investigators control the training and delivery of the intervention. As one moves along the developmental phases to field trials or replication studies, where delivery is less well controlled, investigators need to assess the extent to which the intervention is delivered as designed, how the deliverer may alter any component of the intervention and what impact revisions may have on the intervention's outcomes.

The purpose of this paper is to present findings regarding the association of implementation fidelity and the targeted mediators of a substance abuse prevention program, Take Charge of Your Life (TCYL) delivered to students when they are in the seventh and ninth grades by Drug Abuse Resistance Education (D.A.R.E.) officers. Implementation fidelity data were collected by specially trained independent observers using standardized forms. Data for the program's targeted mediators come from student surveys completed at seven points in time over the 5-year study period.

Background

The Adolescent Substance Abuse Prevention Study

The evaluation

The Adolescent Substance Abuse Prevention Study (ASAPS) funded by the Robert Wood Johnson Foundation and conducted by a research team at The University of Akron is designed to assess the impact of a substance abuse prevention program with components delivered to students when they are in the seventh and ninth grades (TCYL). The program was delivered by D.A.R.E. officers. The study was conducted in 83 school districts (with 83 clusters of high schools and their 122 feeder middle schools) located in six metropolitan areas (Detroit, Houston, Los Angeles, Newark, New Orleans and St Louis) and began in the 2001–02 school year with a cohort of seventh graders who were surveyed through the 11th grade year school districts in the inner city and other districts within a 50-mile radius around the inner city district were randomly selected and then recruited to participate in the study. Once letters stating agreement to participate in the study were received from district superintendents, school principals and police departments within the respective jurisdiction, school clusters were randomly assigned to receive the TCYL curriculum or to continue with the substance abuse prevention program in place in their school, to provide a new program or not to provide any program at all. Forty-two high schools and their 63 feeder middle schools were assigned to the control condition while 41 high schools and their 59 feeder middle schools were assigned to the treatment condition. At baseline, 19 200 seventh graders with positive parental consents and student assents (56.3% of those considered eligible from school rosters at the beginning of the 2001–02 school year) participated in the study. Of these, 11 118 attended the treatment schools and 8082 attended the control schools. Students completed self-administered surveys in a classroom setting at seven time points: at baseline, after the seventh grade intervention in the treatment schools and ~60 to 90 days after the baseline sur-

vey in the control schools, ~1 year later when students were in the eighth grade, prior to and after the delivery of the ninth grade TCYL curriculum in the treatment schools and about the same time in the control schools, ~1 year later when students were in the 10th grade and, finally, when students were in the 11th grade. All surveys were administered by the research staff; no school personnel or D.A.R.E. officer instructors were present during survey administration.

The curriculum

TCYL draws heavily from both education and substance abuse prevention research and emphasizes normative beliefs about the prevalence and social acceptability of the use of tobacco, alcohol, inhalants, marijuana and other drugs (ATOD); the perceptions of the consequences associated with the use of these substances and the acquisition of social skills including communication skills, decision-making skills and resistance strategies. The program is designed to address ATOD use through a series of scenarios representing problem situations which were drawn from focus groups of middle and high school students. The seventh grade curriculum consists of 10 lessons and the ninth grade curriculum includes seven lessons.

The program content is intertwined throughout the lessons to help students form an understanding of the risks involved when substances are used, particularly by young people whose bodies and brains are still developing. The lesson content also supports students' understanding that most adolescents do not use ATOD on a regular basis and provides opportunities for students to practice the social and communication skills that enable them to resist the use of substances.

The program fosters interaction and active learning through large and small group discussions and role-play to stimulate students to 'try on' and think through problematic situations guided by the D.A.R.E. officer instructors. The program's spiral structure allows concepts and skills to be revisited throughout the lessons such that the skills are introduced, revisited and practiced in increasingly more complex problem situations.

Homework is provided so that students can further refine the skills and understandings being developed in the classroom and to keep parents fully informed about the lessons and their students' work. Follow-up in-class sessions weave the homework and class activities into a meaningful sequence of learning activities.

Training of instructors

The training of instructors to understand, to embrace and to teach the seventh grade curriculum occurred in the fall of 2001 and for the ninth grade curriculum in late August of 2003. Officers who participated in the training consisted of D.A.R.E. officers who were available to serve the study school districts. In a few instances where local law enforcement agencies did not have a D.A.R.E. program, D.A.R.E. officers were recruited from nearby agencies. The study paid for all travel expenses associated with the training. Salaries for the D.A.R.E. officers during training were paid by their local agencies or by D.A.R.E. America. There were a total of six 3-day sessions with 67 officers receiving training for the seventh grade curriculum and three 3-day sessions with 73 officers in training for the ninth grade curriculum. Instructors were given a curriculum manual which included not only the curriculum but also the information and resources covering tobacco, alcohol and other drugs and their effects and describing the purpose, rationale, as well as developmental criteria of the curriculum. In addition, training consisted of teaching each of the lessons, demonstrating how to teach the lessons and practicing the lessons in a simulated setting with fellow officers serving as 'students'. All officers had been trained previously by D.A.R.E. America in the D.A.R.E. elementary curriculum and all officers signed consent forms agreeing to being observed in the classroom.

The principal philosophy of the training was that officers would best understand and be able to teach the program if they had ample opportunities to have significant dialog about the design of the curriculum and content of each lesson, to observe demonstrations by colleagues as to how the lessons are

intended to be taught and to teach each of the lessons and receive instructive feedback about their teaching.

Careful planning and instruction were taken into consideration when training the instructors as the new curriculum, TCYL, employs instructional strategies which are quite different from those emphasized by D.A.R.E. America in prior curricula. The officers were being asked to make a switch from a curriculum which required a more didactic approach to one more highly interactive, from one which was instructor focused to being student focused and to one which allows time to practice and master resistance skills. Moreover, the curriculum team emphasized the importance of understanding the curriculum design and theory to be effective instructors of this new curriculum.

As the curriculum design and structure were key components to the success of the program, a sub-study was incorporated into the evaluation to assess the extent to which the officer instructors implemented the program as they were trained to do. There was particular concern regarding the high reliance on interactive and constructivist instructional style which was counter to the officer instructors' prior training and experience that was less interactive [23]. In addition, the research staff and curriculum developers wanted to know if there was an association between fidelity of implementation and achievement of the aims of each lesson.

Methods

Implementation fidelity study

In the ASAPS, fidelity included multiple dimensions including content coverage, use of appropriate instructional strategy, the expected responsive behaviors of the students and time on task. For this paper, the focus will be on content coverage and use of appropriate instructional strategy. The decision to focus on these measures is in part guided by the findings from studies discussed above by Hallfors and Ennett and their colleagues that addressed content and use of interactive instructional methods.

In this paper, we utilize the observation data from the substudy above and data from student surveys to address the following two questions: (i) to what extent do the officer instructors cover all the material and activities in the program lessons and use the appropriate instructional strategies to deliver the lesson components and (ii) do students exposed to officer instructors who deliver the lesson content with fidelity and who adhere to the appropriate instructional strategies have better scores on the lesson mediators than those who are exposed to officer instructors with lower levels of fidelity and adherence.

Sample

Although 140 officers were trained to deliver the new curricula, 86 officers actually taught TCYL. Of these, 58 taught the seventh grade curriculum and 61 taught the ninth grade curriculum. Thirty-three officers taught both the seventh and ninth grade curricula. At the end of training, officers were asked to complete an anonymous training assessment form which measured the extent to which they were ready to teach the curriculum. The latter form consisted of statements regarding the quality of their training and level of preparedness to teach the new curricula. Each statement was followed by a five-point Likert scale. The statements were (i) I understand the purpose of the curriculum, (ii) I feel ready to use the teaching strategies, (iii) I believe that the content of the lessons will help prevent substance abuse among adolescents, (iv) I believe that most students will become involved in the activities of the lessons, (v) I believe the training was effective, (vi) I am enthusiastic about this curriculum and (vii) the trainers were knowledgeable.

Officer instructors who taught the curriculum ($n = 86$) had an average age of 40 and were more likely to be male (73.3%), white (73.3%) with some college (33.7%) or higher education (39.6%). The mean number of years as a police officer was 14.6 years and as a D.A.R.E. officer was 6.0 years. These findings are consistent with the profile of D.A.R.E. officers reported in a paper by Merrill *et al.* [24]. It is interesting to note that 36% of the officers had taught a prevention program other than a D.A.R.E.

curriculum and 27.8% had some graduate level courses in education.

After completing the training to deliver the seventh grade curriculum, 96% of the officer instructors agreed that they understood the purpose of the curriculum, and the same percentage indicated that they felt ready to use the teaching strategies presented during training. Ninety-two percent of them believed that the content of the curriculum would prevent/reduce substance use among middle and high school students, and 96% indicated that they believed most students would get involved in the activities. Ninety-four percent agreed that the training was effective, and 96% agreed that they were enthusiastic about teaching the new curriculum. Similarly, 94% thought the trainers were prepared and knowledgeable.

After the training for the ninth grade curriculum, 97% of officer instructors agreed that they understood the purpose of the curriculum, and 93% felt they were ready to use the new teaching strategies. Ninety-three percent thought that the content covered would prevent/reduce substance use among middle and high school students, and 97% believed that students would get involved in the activities. Further, 93% of the officer instructors agreed that the training was effective, and 97% felt enthusiastic about teaching the curriculum.

Only a few of the officer instructors had prior teaching experiences in high school, and many expressed concerns about management of high school students during the ninth grade training. However, in debriefing interviews with the officer instructors between 6 and 9 months after their experiences teaching high school students, >85% reported that they enjoyed these older students and would teach them again.

Measurement and data collection

Content coverage and instructional strategy

For purposes of this study and for consistency with the curriculum design, the dimensions of implementation fidelity encompassed content coverage and instructional strategy. To measure these components, officer instructors were observed in the

classroom. Two key lessons (second and sixth in sequence) were selected for observation from the seventh grade curriculum because of their focus and the different instructional strategies required. The first lesson observed (Lesson 2) includes the types of activities, didactic instructional strategies and level of student engagement similar to the D.A.R.E. officers' previous experiences. Conversely, the second lesson observed (Lesson 6) on the other hand focuses on skills building and its activities are more student centered and highly interactive requiring a different set of instructional strategies and facilitation techniques. These two lessons also emphasize key elements of the curricula's theoretical foundation with Lesson 2 focusing on the effects of drugs on the body and Lesson 6 on building decision-making skills.

For the ninth grade curriculum, a similar approach was taken. The ninth grade curriculum was designed to actively engage students in group problem solving and discussion and heavily focuses on the current and future consequences of substance use. Again, Lessons 2 and 6 were selected for observation. Lesson 2 addresses the effects of marijuana and ecstasy on the brain and in turn on emotions and behaviors. In addition, normative beliefs about peer use of these drugs are discussed. Lesson 6 examines the social and legal consequences of substance use with a heavy emphasis on alcohol use.

Instructors were observed teaching each of the two selected lessons twice. The lessons chosen for observation were to be sequential and neither the first class of instruction nor the last class of instruction for that day was observed. Each instructor was observed a total of four times in each observation year. Standardized observation coding sheets were created for each of the lessons being observed. Because of the size and scope of the study in terms of school sites, focus on reliability of coding of observations was placed on instrumentation development and training of the observers. Two pilot studies were conducted in a total of 13 Ohio middle schools with nine officer instructors and 13 high schools with 11 officer instructors and included >1200 middle school students and >700 high school students.

The purpose of the pilot studies was to have trained Ohio D.A.R.E. officer instructors deliver the curricula, to test our consenting procedures and to pre-test data collection instruments including the student survey and the observation forms for the two selected lessons. Two observers independently rated officer instructors. On the first round of observations, the ease of using the forms was assessed. On the second round, reliability of the instrument was determined. Interrater agreement with two raters ranged from 82 to 94% for observation forms for the seventh and ninth grade curricula.

During training for the national study, University-employed observers reviewed the lessons which would be observed. Emphasis was placed on noting what the officer instructors and students were doing during the lesson. The trainees then viewed a video of an officer instructor delivering the lessons in a real classroom setting. Their codes were shared and discussed. The training ended with another viewing of the video and codes reviewed. Trainees who were outliers were not used to make the observations.

The observers recorded for each activity whether the activity and its subactivities were covered, the amount of time involved with that activity, the officer's instructional strategy (from a check-off list—non-interactive, lecture/listen, ask questions, discuss, model or facilitate) and the students' behavior (also from a check-off list—non-interactive, lecture, answer questions, discuss, model or problem solving). Observers were provided with a definition sheet for the instructional strategies with examples of behaviors for each instructional strategy/code. The observers also recorded the number of students who were off-task during an activity. In addition, they recorded 'any interesting, unusual or other information which is pertinent to the study'. Lastly, the observers were asked to rate the instructors on the following six questions, using a five-point Likert scale (agree, kind of agree, neutral, kind of disagree and disagree): (i) was the instructor enthusiastic? (ii) did the students appear to be comfortable talking with the officer? (iii) did the students and the officer smile and laugh while staying on task? (iv) did the instructor appear to listen

attentively? (v) did the officer seem well prepared to teach? and (vi) did the students understand the officer's directions, explanations and questions. During training, officers were told about the observations and signed consents agreeing to be observed.

Analyses were conducted to determine the within-observer consistency and between-observer correlation between observations of the lessons for content coverage and officer instructor instructional strategy. Among these analyses were analyses of variances (ANOVAs) with the observer serving as a fixed-effect, paired *t*-tests for the two observations and correlations between all observations. There were no significant findings on the ANOVAs and the *t*-tests, and the correlations were all in the moderate range indicating that the observers showed no systematic bias in scoring.

Content coverage was measured by summing the number of activities and subactivities presented for each lesson observed. Items were coded 0, did not cover, and 1, covered. Content items were added together and divided by the total number of activities to be covered, producing the proportion of activities covered for each lesson. The seventh grade curriculum included 16 content items in Lesson 2 and 21 content items in Lesson 6 while for the ninth grade, the number of content items was 25 for Lesson 2 and 26 for Lesson 6.

Adherence to appropriate instructional strategy was coded for each activity for each of the lessons observed. First, each activity was scored as to whether it was covered or not. Then, each activity was scored 0 if the officer was non-compliant and 1 if the officer was compliant, i.e. the officer's instructional strategy matched the strategy indicated in the teaching manual.

Lesson mediators

The mediator measures came from the self-administered surveys completed by students participating in each of the observed classes. For these analyses, comparisons of measures of the targeted mediators were made between the seventh pre- and post-test surveys for the observed seventh grade lessons and the ninth grade pre- and post-test surveys for the observed ninth grade lessons. Students who were in

classrooms observed for fidelity were selected and matched to the officer instructor who delivered the curriculum in those students' classrooms. The relationship of interest is estimated at the classroom level (seventh grade Lesson 2, $n = 101$ observed classes; seventh grade Lesson 6, $n = 104$ observed classes; ninth grade Lesson 2, $n = 120$ observed classes and ninth grade Lesson 6, $n = 120$ observed classes).

Lesson 2 of the seventh grade curriculum for TCYL was intended to impact perceptions of the consequences of the use of substances and for Lesson 6, decision-making skills while Lesson 2 of the ninth grade curriculum targeted resistance skills, consequences of substance use and normative beliefs regarding substance use among peers, and Lesson 6 focused primarily on the social and legal consequences of alcohol use. The questions used to measure these mediators (described below) were taken from national studies such as the Monitoring the Future Study.

'Normative beliefs' were measured using two scales. The first scale, normative beliefs I, was computed using the mean of five ordinal items asking students to estimate how many eighth graders (or in the case of the ninth grade surveys, 10th graders) used cocaine, alcohol, tobacco, inhalants or marijuana in the prior 30 days ($\alpha = 0.81$). Scores ranged from one (none) to five (most eighth graders used this substance in the prior 30 days). The second scale, normative beliefs II ($\alpha = 0.90$), was constructed using the mean of seven items completing the statement: 'Most students my age think it's okay to...'. Examples of responses to this statement include 'smoke one or more cigarettes a day' and 'smoke marijuana once in a while'. Scores range from one (agreement) to five (disagreement). These scales were reverse coded so that higher scores indicate better program outcomes.

The 'perceived consequences of substance use' measure focused on the negative effects of tobacco, alcohol and other drugs on the functioning of the brain and perceptions of social consequences as a result of the use of alcohol. For the former measure, we used the mean of five ordinal items asking students how much they think using a specific drug

has an effect on the functioning of the brain ($\alpha = 0.86$ in seventh grade and 0.88 in ninth grade). Scores range from one (none) to five (a lot). For the other consequences measure, we asked students to respond to six statements measuring negative alcohol expectancies such as 'If I drank alcohol, I might do things that I wouldn't normally do' and 'If I drank alcohol, I would get into trouble with parents/caregivers'. Response categories were 'disagree, kind of disagree, neither, kind of agree and agree'. These items were combined by taking the mean across the six items.

'Knowledge of resistance skills' was measured by responses given to three hypothetical scenarios involving the opportunity to use tobacco, alcohol or marijuana. Students were asked to read a scenario in which a given substance was offered by a peer and to select the most appropriate refusal response to the offer 'given the person being offered does not want to use'. Each response was weighted according to the level of assertiveness it demonstrated. For example, a response of 'no, maybe later' was assigned a lower score than a response of 'no thanks, I don't want to smoke'. Scores ranged from zero (least appropriate response) to two (best response chosen) for each scenario and then summed for a final score ranging from zero to six.

'Decision-making skills' were measured in the seventh grade by a single item with a choice of five responses weighted to reflect the level of decision-making skill for that response. Scores include 0 (worse response), 1 (second best response) and 2 (best response).

Results

Coverage

The range of scores (representing the proportion) for content coverage for Lesson 2 of the seventh grade curriculum was 0.34 – 1.00 and Lesson 6, 0.36 – 1.00 . The median coverage for instruction of the seventh grade curriculum was 0.81 for Lesson 2 and 0.72 for Lesson 6.

Coverage for the ninth grade curriculum for Lesson 2 ranged from 0.12 to 1.00 and from 0.13 to

1.00 for Lesson 6. The median coverage of activities for the ninth grade curriculum was 0.70 for Lesson 2 and 0.78 for Lesson 6.

Instructional strategy

The median score for the proportion of activities delivered using the appropriate instructional strategy was 0.63 for Lesson 2 and 0.44 for Lesson 6. That is, half of the officers used the correct instructional strategy for about two-thirds of the activities in Lesson 2. In addition, half of the officers used the correct instructional strategy for at least 0.44 of the activities in Lesson 6. However, as seen for content coverage, the range of scores was large (0.31 – 0.94 for Lesson 2 and from 0.10 to 0.83 for Lesson 6).

The median score for use of the appropriate instructional strategy was 0.50 for Lesson 2 and 0.60 for Lesson 6. That is, half of the officers utilized the appropriate instructional strategy for at least half of the activities in Lesson 2 and in at least 0.60 of the activities in Lesson 6. Again, the range of scores for the lessons was similar to those for content coverage, 0.12 – 0.96 for Lesson 2 and 0.12 – 1.00 for Lesson 6.

Higher content coverage was positively correlated with the use of appropriate instructional strategy ($r = 0.93$, $P < 0.001$). There were no significant correlations between officer instructor's age, sex, race or level of education and the mean level of curriculum content coverage or use of the appropriate instructional strategy at either the seventh or ninth grade.

Relationship of content coverage and adherence to appropriate instructional strategy and the targeted lesson mediators

The analyses examining the relationship between content coverage and use of the appropriate instructional strategy and the scores on the targeted lesson mediators were conducted using hierarchical linear modeling (HLM) to estimate two-level random intercept models (individuals nested within observed classrooms). These models are necessary to take into account the nested nature of the data. HLM

models produce unbiased estimates of standard errors and corresponding significance tests [25]. This is accomplished by the inclusion of a second-level error term (μ_0) that allows the individual level errors (r_{ij}) to be uncorrelated. Higher scores on content coverage and adherence to instructional strategy variables represent higher coverage and adherence during the specific lesson being observed. Thus, we would predict positive relationships between the coverage and adherence variables and the mediators. In addition, we control for the pre-test scores on each of the program mediators when estimating the effect of implementation fidelity on the mediators measured at the post-test. The effect of the pre-test on the post-test score is estimated at the individual level and the effect of implementation fidelity (content coverage and adherence to the appropriate instructional style) is estimated at the classroom level. For that reason, student data used in the analyses came from the surveys for students who were in the classroom at the time of observation. The number of observed classrooms was 101 and 104 for Lessons 2 and 6, respectively, of the seventh grade curriculum and 120 for both Lessons 2 and 6 of the ninth grade curriculum. Table I provides summary statistics for these items measured prior to and after the intervention.

The HLM analyses showed that for Lesson 2 of the seventh grade curriculum, there was no relationship between content coverage and perceptions of consequences (see Table II). For Lesson 6, however, we found a statistically significantly better outcome on decision-making skills for students exposed to higher content coverage. The findings for adherence to appropriate instructional strategy paralleled those for content coverage in the seventh grade, with use of appropriate instructional strategy related significantly to improved decision-making skills. Specifically, the odds of giving the second best answer (1) as opposed to giving the best answer (2) are multiplied by 0.51 ($e^{-0.666}$) for every one unit increase in appropriate instructional strategy. Alternatively, a 10% increase in appropriate instructional strategy would decrease the odds of being in the worst answer category (as opposed to

Table I. Summary scores for program mediators at baseline and post-intervention

	Mean	Standard deviation	Median	Range
Baseline scores				
Seventh grade				
Consequences	3.71	1.01	4.0	1–5
Decision-making skills	1.13	0.73	1.0	0–2
Ninth grade				
Resistance skills	2.73	1.38	3.0	1–6
Consequences	3.85	1.00	4.0	1–5
Normative beliefs I	3.06	0.83	3.0	1–5
Normative beliefs II	3.25	1.08	3.14	1–5
Negative alcohol expectancies	3.50	0.98	3.5	1–5
Post-intervention scores				
Seventh grade				
Consequences	3.80	0.97	4.0	1–5
Decision-making skills	1.18	0.77	1.0	0–2
Ninth grade				
Resistance skills	2.73	1.35	3.0	1–6
Consequences	3.81	1.06	4.0	1–5
Normative beliefs I	3.38	0.82	3.4	1–5
Normative beliefs II	3.26	1.05	3.14	1–5
Negative alcohol expectancies	3.43	1.04	3.5	1–5

being in the best answer category) by 0.04 ($0.10 \times e^{-0.986}$).

The analyses for Lesson 2 of the ninth grade curriculum showed that while coverage did not make a difference on resistance skills or the normative belief measures, students exposed to higher content coverage scored significantly higher on the consequences measure. That is, students exposed to a higher proportion of the content had significantly greater perceptions of negative consequences of substance use compared with students exposed to less content. Similar results were found when the measure of appropriate instructional strategy was examined.

Finally, we examined whether content coverage and the use of the appropriate instructional strategy impacted negative alcohol expectancies, as targeted in Lesson 6 of the ninth grade curriculum. Results from hierarchical regression models suggest that

Table II. Two-level regression model of content coverage and use of appropriate instructional strategy on mediators targeted by curriculum (by lesson)

Content coverage		Number of classrooms		Content coverage		
Lesson	Observed	Outcome	<i>b</i>	Standard error	<i>Z</i>	
Seventh, Lesson 2	101	Consequences (high better)	0.114	0.170	0.671	
Seventh, Lesson 6	104	Decision-making skills (0 versus 2) ^a	−1.011	0.506	− 2.000	
	104	Decision-making skills (1 versus 2) ^a	−1.001	0.499	− 2.003	
Ninth, Lesson 2	120	Resistance skills (high better)	0.174	0.148	1.172	
	120	Consequences (high better)	0.359	0.148	2.427	
	120	Normative beliefs I (high better) ^b	−0.153	0.117	−1.311	
	120	Normative beliefs II (high better) ^b	0.012	0.112	0.104	
Ninth, Lesson 6	120	Negative alcohol expectancies (high better)	0.347	0.128	2.708	
Use of appropriate instructional strategy			Appropriate instructional style			
Lesson	Observed	Outcome	<i>b</i>	Standard error	<i>Z</i>	
Seventh, Lesson 2	101	Consequences (high better)	0.134	0.135	0.992	
Seventh, Lesson 6	104	Decision-making skills (0 versus 2) ^a	−0.986	0.443	− 2.221	
	104	Decision-making skills (1 versus 2) ^a	−0.666	0.421	−1.579	
Ninth, Lesson 2	120	Resistance skills (high better)	0.096	0.174	0.554	
	120	Consequences (high better)	0.380	0.165	2.298	
	120	Normative beliefs I (high better) ^b	−0.291	0.150	1.942	
	120	Normative beliefs II (high better) ^b	0.048	0.131	0.367	
Ninth, Lesson 6	120	Negative alcohol expectancies (high better)	0.308	0.137	2.241	

^aFor the decision-making skills outcome, we utilized a multinomial logistic hierarchical regression model. This model compares giving a worst answer (0) or a second best answer (1) to a best answer (2). Negative coefficients in these models indicate that students with exposure to higher fidelity are less likely to be in the 0 or 1 (worst answer or second best answer) categories than in the 2 (best answer) category. For all other outcomes, we utilized a hierarchical linear model. *Z* scores in bold are significant at the $P = 0.05$ level, using a two-tailed test.

^bBriefly, normative beliefs I refers to the objective beliefs students hold about the proportion of students their age using drugs. Normative beliefs II refers to the idea that most adolescents think it is ok to use drugs. See the measurement section for a full description of these two measures.

both content coverage and instructional strategy were significantly related to negative alcohol expectancies, in the expected direction. That is, exposure to more content and appropriate instructional strategy leads students to perceive more negative expectancies for alcohol use.

Summary and discussion

The purpose of this paper is to explore the issue of implementation fidelity within the context of a large study designed to evaluate a substance abuse prevention program, TCYL, delivered by officer instructors (D.A.R.E.). The focus was on two measures of implementation fidelity—content coverage

and adherence to the appropriate instructional strategy. TCYL's programmatic conceptual design was based on principles of substance abuse prevention which include addressing misconceptions held by students regarding the normative nature of the use of substances and helping students understand the negative consequences of substance use. In addition, the program provides opportunities to discuss this decision and to practice skills to resist use in a variety of age-appropriate and relevant situations. Therefore, the emphasis of TCYL, as with other recently designed substance abuse prevention programs, is on specific content and the use of instructional strategies which help students internalize the skills and knowledge that is provided. D.A.R.E. officers have traditionally used didactic teaching

techniques and the use of interactive strategies which more fully engage students in the learning process is a challenge. The substudy allowed us to determine to what extent the officer instructors delivered TCYL as it was intended both in content and instructional strategy and to relate fidelity with targeted program mediators.

We found that the scores for these components of implementation had large ranges, skewed toward the higher implementation levels. Overall, the median scores indicate that the officer instructors were stronger in covering the content of the curricula than in their use of appropriate instructional strategies. This may be due to the officer instructors' prior experience in didactic teaching methods as used with other D.A.R.E. curricula. For the observed Lessons 2 which required more didactic and modeling instructional strategies, the scores were better than those for observed Lessons 6 which required more interactive and guiding instructional strategies.

The relationship between implementation fidelity and the targeted mediators, although not always consistent, showed promising results. Although higher scores for content coverage and use of appropriate instructional strategies for Lesson 2 of the seventh grade curriculum did not show any impact, Lesson 2 of the ninth grade curriculum and Lessons 6 for both the seventh and ninth grade curricula did show statistically significant differences on decision-making skills, perceived negative consequences of substance use and alcohol expectancies. These findings suggest that the greater the adherence both to the curricula as designed and engagement of students as prescribed appear to produce the desired impact on those elements or targeted mediators of the intervention.

The similarities in findings across the two dimensions of implementation fidelity are not surprising given the high correlations between more complete content coverage and use of appropriate instructional style. However, the inconsistency of the relationship between how the program was delivered and the mediator scores are difficult to assess. Lesson 2 focuses heavily on consequences of substance use showing how these substances impact brain

functioning. Students participate in teams as emergency medical technicians and must determine what substance a hypothetical student has taken by reviewing that student's symptoms. It may be that involvement in the activities of the lesson may be sufficient alone and that it does not matter how the officer instructor presents the material. In contrast, Lesson 6 of the seventh grade curriculum emphasizes two aspects of decision making. First is learning the steps involved in making decisions while the second is to apply these steps through group discussion using a scenario as an organizing framework. In this way, decision making is reinforced using different instructional strategies.

The major thrust of Lesson 2 of the ninth grade curriculum is on the consequences of substance use within the context of impeded brain functioning. Resistance skills in this lesson address both external and internal pressures to use alcohol and drugs. The activities for the session were primarily done in teams with 'go around' brainstorming around strategies to resist these pressures. The idea is to have the students talk about the pressures and within the context of several scenarios suggest how to avoid use of alcohol or drugs. Perhaps the options within the question responses are not sufficiently sensitive to reflect changes.

There are several limitations to the substudy which should be mentioned. First, as noted previously, the decision-making skills' outcome is measured by a single item indicator. This item does not have the psychometric properties of a multiple item scale and thus measurement reliability cannot be assessed. In spite of this limitation, we did find significant associations between implementation fidelity and this outcome. It is unlikely that implementation fidelity is correlated with measurement error. It is more likely that this relationship is misestimated due to our inability to estimate a measurement model. Structural equation models, with a multiple item indicator of decision-making skills, would likely show stronger relationships than those estimated here. However, despite of this limitation, the observed association is statistically significant. Also related to this issue is the measure of resistance skills for the ninth grade curriculum. We drew

from other research in developing these measures. However, as we mentioned earlier, our approach needs to be revisited.

A second limitation of the current project is the limited number of lessons observed each year. Because of time and cost, we decided to sample the lessons choosing those lessons which focused on some of the mediators and which represented different instructional strategies. Therefore, we only observed 20% of the lessons (two of 10 lessons) in seventh grade and 28% of the lessons (two of seven lessons) in ninth grade. It would have been interesting to observe a greater number of lessons to see if the results found for these four observed lessons held. We do believe that these lessons do reflect the broader curriculum and, indeed, high correlations were found between lessons for individual officer instructors with r values of 0.535 being the lowest between the two observations of use of appropriate instructional strategy for the ninth grade curriculum to 0.926 for use of appropriate instructional strategy for the ninth grade curriculum. These high values suggest that the number of observed lessons may be sufficient. Future studies may want to assess more program lessons, although we acknowledge that this is a costly endeavor.

This project is unique in that we had a large sample of observed classrooms. Current statistical techniques, such as the hierarchical linear models, utilized in the current project require relatively large sample sizes. Unfortunately, even with such a large sample of observed classrooms, we may not have had sufficient power to detect effects at the 0.05 level. A larger sample of classrooms would have increased power. This is of real concern for future research and smaller projects, where implementation quality will be assessed across fewer classrooms. Techniques, such as the multilevel models, estimated here are not recommended for sample sizes <15.

One explanation of why the D.A.R.E. officers would adhere to the curriculum so closely may be due to the organization's regimented training and heavy reliance on the printed teaching manuals (e.g. officers were taught outlining skills during

training with D.A.R.E. America as part of preparation for teaching). However, of concern are the moderate scores for using the appropriate instructional strategies, particularly those which are highly interactive. It is possible that more positive experience with such strategies will give the officers more confidence. Clearly, the officers need to have feedback to reinforce their good techniques and to correct those which are problematic. It is therefore important to have teachers within the schools or D.A.R.E. mentors available to provide this assistance.

However, the information needed to guide training of deliverers of substance abuse prevention programming requires the field to develop standard definitions and measures of the components which constitute fidelity of implementation and the methodologies best suited to collect these data. The approaches used in this study to examine implementation were drawn heavily from those reported in the literature. Many former studies rely on self-report of the instructors using a checklist. There are obvious limitations of such an approach. Furthermore, few studies include more than one dimension of implementation, primarily examining some aspect of content coverage or exposure. The curriculum design of TCYL included as key features not only content but also instructional strategy using the most appropriate strategy to enhance the learning process.

Although implementation fidelity has been addressed by individual investigators, there are no standardized definitions, measurements or data collection methodologies to guide this research. As programs go from controlled 'laboratory' settings into the 'real world', it will be more important to develop these standards to determine to what extent programs are implemented in the field as they were designed to be delivered under research controlled conditions.

Funding

Robert Wood Johnson Foundation (039223 and 040371) to ASAPS.

Conflict of interest statement

None declared.

References

1. Kaftarian SJ, Robertson E, Compton W *et al.* Blending prevention research and practice in schools: critical issues and suggestions. *Prev Sci.* 2004; **5**: 1.
2. Pentz MA. Form follows function: designs for prevention effectiveness and diffusion research. *Prev Sci.* 2004; **5**: 23.
3. Greenberg MT. Current and future challenges in school-based prevention: the researcher perspective. *Prev Sci.* 2004; **5**: 5.
4. Botvin GJ. Advancing prevention science and practice: challenges, critical issues, and future directions. *Prev Sci.* 2004; **5**: 69.
5. Hallfors D, Godette D. Will the 'principles of effectiveness' improve prevention practice? Early findings from a diffusion study. *Health Educ Res.* 2002; **17**: 461.
6. Ennett ST, Ringwalt CL, Thorne J *et al.* A comparison of current practice in school-based substance use prevention programs with meta-analysis findings. *Prev Sci.* 2003; **4**: 1.
7. Rohrbach LA, Graham JW, Hansen WB. Diffusion of a school-based substance abuse prevention program: predictors of program implementation. *Prev Sci.* 1993; **22**: 237.
8. Pentz MA, Trebow E, Hansen WB *et al.* Effects of program implementation on adolescent drug use behavior: the midwestern prevention project (MPP). *Eval Rev.* 1990; **14**: 264.
9. Tobler NS, Stratton HH. Effectiveness of school-based drug prevention programs: a meta-analysis of the research. *J Prim Prev.* 1997; **18**: 71.
10. Pentz MA, Trebow E. *Implementation Issues in Drug Abuse Prevention Research*. Rockville, MD: U.S. Dept. of Health and Human Services, 1991No. NIDA Research Monograph 107.
11. Botvin GJ, Eli B, Dusenbury L *et al.* Long-term follow-up results of a randomized drug abuse prevention trial in white middle-class population. *J Am Med Assoc.* 1995; **273**: 1106.
12. Hansen WB, Graham JW, Wolkenstein BH *et al.* Program integrity as a moderator of prevention program effectiveness: results for fifth-grade students in the adolescent alcohol prevention trial. *J Stud Alcohol.* 1991; **52**: 568.
13. Hansen WB, McNeal Jr RB, Ralph B. Drug education practice: results of an observational study. *Health Educ Res.* 1999; **14**: 85.
14. Dusenbury L, Brannigan R, Falco M *et al.* A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health Educ Res.* 2003; **18**: 237.
15. Resnicow K, Cross D, Wynder EL. The know your body program: a review of evaluation studies. *Bull NY Acad Med.* 1993; **70**: 188.
16. Abbott RD, O'Donnell J, Hawkins DJ *et al.* Changing teaching practices to promote achievement and bonding to school. *Am J Orthopsychiatry* 1998; **68**: 542.
17. Basen-Engquist K, O'Hara-Tompkins N, Lovato CY *et al.* The effect of two types of teacher training on implementation of smart choices: a tobacco prevention curriculum. *J Sch Health* 1994; **64**: 334.
18. Kam CM, Greenberg MT, Walls CT. Examining the role of implementation quality in school-based prevention using the PATHS curriculum. Promoting alternative thinking skills curriculum. *Prev Sci.* 2003; **4**: 55.
19. Spoth R, Redmond C, Shin C. Direct and indirect latent-variable parenting outcomes of two universal family-focused preventive interventions: extending a public-health oriented research base. *J Consult Clin Psychol.* 1998; **66**: 385.
20. Perry CL, Murray DM, Griffin G. Evaluating the statewide dissemination of smoking prevention curricula: factors in teacher compliance. *J Sch Health* 1990; **60**: 501.
21. Sobol DF, Dent CW, Gleason L *et al.* The integrity of smoking prevention curriculum delivery. *Health Educ Res.* 1989; **4**: 59.
22. Hahn EJ, Noland MP, Rayens MK *et al.* Efficacy of training and fidelity of implementation of the life skills training program. *J Sch Health* 2002; **72**: 282.
23. Ennett SL, Tobler NS, Ringwalt CL *et al.* How effective is drug abuse resistance education? A meta-analysis of Project DARE outcome evaluations. *Am J Public Health* 1994; **84**: 1394–1401.
24. Merrill J, Dilascio T, Pinsky I. Law enforcement and drug prevention: a profile of the D.A.R.E. officer. *Police Chief* 2006; **69**: 81.
25. Bryk A, Raudenbush S, Congdon R. *HLM: hierarchical linear and nonlinear modeling with the HLM/2L and HLM/3L programs*. Chicago: Scientific Software International. 1996.

Received on February 13, 2007; accepted on May 19, 2008