



Published in final edited form as:

*Prev Sci.* 2015 January ; 16(1): 90–99. doi:10.1007/s11121-014-0459-1.

## Adherence and delivery: Implementation quality and program outcomes for the 7th grade *keepin' it REAL* program

Jonathan Pettigrew<sup>1,2</sup>, John W. Graham<sup>3,4</sup>, Michelle Miller-Day<sup>1,5</sup>, Michael L. Hecht<sup>1,4</sup>, Janice L. Krieger<sup>6</sup>, and Young Ju Shin<sup>1,7</sup>

<sup>1</sup>Department of Communication Arts & Sciences, Penn State

<sup>3</sup>Department of Biobehavioral Health, Penn State

<sup>4</sup>Prevention Research Center, Penn State

<sup>6</sup>School of Communication, The Ohio State University

### Abstract

Poor implementation quality (IQ) is known to reduce program effects making it important to consider IQ for evaluation and dissemination of prevention programs. However, less is known about the ways specific implementation variables relate to outcomes. In this study, two versions of the *keepin' it REAL*, 7<sup>th</sup> grade drug prevention intervention were implemented in 78 classrooms in 25 schools in rural districts in Pennsylvania and Ohio. IQ was measured through observational coding of 276 videos. IQ variables included adherence to the curriculum, teacher engagement (attentiveness, enthusiasm, seriousness, clarity, positivity), student engagement (attention, participation), and a global rating of teacher delivery quality. Factor analysis showed that teacher engagement, student engagement, and delivery quality formed one factor, which was labeled delivery. A second factor was adherence to the curriculum. Self-report student surveys measured substance use, norms (beliefs about prevalence and acceptability of use), and efficacy (beliefs about one's ability to refuse substance offers) at two waves (pretest, immediate posttest). Mixed model regression analysis which accounted for missing data and controlled for pretest levels examined implementation quality's effects on individual level outcomes, statistically controlling for cluster level effects. Results show that when implemented well, students show positive outcomes compared to students receiving a poorly implemented program. Delivery significantly influenced substance use and norms, but not efficacy. Adherence marginally significantly predicted use and significantly predicted norms, but not efficacy. Findings underscore the importance of comprehensively measuring and accounting for IQ, particularly delivery, when evaluating prevention interventions.

---

Correspondence should be addressed to: Jonathan Pettigrew, 293 Communications Building, University of Tennessee, Knoxville, TN 37996. [jpettigrew@utk.edu](mailto:jpettigrew@utk.edu).

<sup>2</sup>Now at School of Communication Studies, University of Tennessee

<sup>5</sup>Now at Department of Communication Studies, Chapman University

<sup>7</sup>Now at Department of Communication Studies, Indiana University Purdue University Indianapolis

Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health. (NIH Manuscript # NIHMS272843).

Portions of this paper were presented at the 2013 annual meeting of the Society of Prevention Research.

## Keywords

Implementation quality; program evaluation; substance use; adolescents

---

## Problem

Despite the fact that almost half of middle schools select evidence-based drug prevention programs (Ringwalt et al., 2011) few implement these programs as intended. For example, a survey of 81 Safe and Drug Free School district coordinators found only 19% of schools implemented evidence-based programs with good fidelity (Hallfors & Godette, 2002). In other studies, implementers adapted more than 50% of the program (Knoche, Sheridan, Edwards & Osborn, 2010; Odom et al., 2010) with Durlak (1998) estimating that as many as 80% of program activities may be omitted during implementation. This is problematic because poorly implemented programs tend not to achieve positive outcomes (for review, see Durlak & DuPre, 2008). Thus, quality implementation is crucial to individual program success.

Given its importance in program success, the study of implementation has implications for prevention science as well as practice. Several scholars recognize the need to investigate implementation issues as part of program evaluation (e.g., Berkel, Mauricio, Schoenfelder, & Sandler, 2011; Domitrovich & Greenberg, 2000). Shortcomings with respect to implementation can be mistaken for shortcomings of the actual curriculum (Domitrovich & Greenberg, 2000; Dusenbury, Brannigan, Falco, & Hansen, 2003), which may lead to an underestimate of the value of some prevention programs. The study of implementation, then, assists in understanding how and why programs work as well as provides a foundation for knowing what programs have the potential to work if well implemented.

The current study adds to this growing body of knowledge by providing a more comprehensive approach to measuring implementation quality (IQ) and examining its effects on program outcomes for the *keepin' it REAL (kiR)* curriculum. *keepin' it REAL* is a 10-lesson curriculum that teaches decision making, general social skills, and “REAL” resistance strategies (refuse, explain, avoid, leave) using videos, presentation of concepts, discussion, activities, and role plays. *kiR* is recognized as evidence-based by SAMHSA’s National Registry of Evidence-based Programs and Practices and California’s Healthy Kids Resource Center based on two at least partially successful clinical trials (Hecht, Graham, & Elek, 2006; Marsiglia, Kulis, Yabiku, Nieri, & Coleman, 2011). Previous studies, however, failed to adequately measure *kiR*’s implementation quality (Marsiglia et al., 2011), something that is particularly important given its widespread dissemination. This study also has the potential to build prevention science by incorporating extensive video observational measures of various dimensions of implementation quality (e.g., adherence, participant responsiveness, and delivery quality) in order to better understand how they relate to specific program outcomes.

## Review of Literature

Implementation, broadly, refers to the “process by which interventions are put into action” (Graczyk, Domitrovich, & Zins, 2002, p.306). Dusenbury et al. (2003) published a conceptual review of implementation and, building on Dane and Schneider (1998), described five key dimensions: adherence to the curriculum (fidelity), quality of delivery (how a teacher presents a curriculum), participant responsiveness (how engaged students are during a lesson), dose (amount of program delivered), and program differentiation (presence/absence of distinguishing features of a particular program). Despite the recognition of various dimensions of implementation, research has focused almost exclusively on adherence and dose. Dane and Schneider (1998), for example, report that in a sample of 162 intervention studies conducted between 1980 and 1994, 11% examined adherence and 13% dose with only 7% measuring quality of delivery, 2% participant responsiveness, and 6% program differentiation. Clearly more attention on various elements of implementation quality is needed.

Early research focused primarily on the amount of a curriculum that was delivered and demonstrated that when more of a program was implemented there were better outcomes (Botvin et al., 1995; Pentz et al., 1990). Botvin et al., (1995), for example, used observer reports to calculate the proportion of program objectives covered by teachers and showed that a subsample receiving at least 60% of program objectives had stronger program effects than the full sample. A more recent evaluation, which measured this type of implementation as a continuous variable, confirms previous findings (Lillehoj, Griffin, & Spoth, 2004). There is a clear relationship between the amount of curriculum covered and student outcomes.

Less, however, is known about how delivery quality, student responsiveness, and other aspects of implementation relate to student outcomes. Fewer studies have examined these variables, but those that have suggest they may be equally important (Dusenbury, Brannigan, Hansen, Walsh, & Falco, 2005; Hansen, Graham, Wolkenstein, & Rohrbach, 1991; Low, Ryzin, Brown, Smith, & Haggerty, 2013). Hansen et al., (1991), for example, had trained observers measure implementation quality as adherence to the curriculum (i.e., fidelity), quality of delivery, student responsiveness, and teacher control of the class. They found that these variables formed a single implementation variable that moderated program outcomes. A more recent study used teacher reports of students’ engagement (i.e., participant responsiveness) and teacher self-reports of adherence and showed that greater levels of student engagement significantly predicted program outcomes whereas adherence was unrelated to outcomes (Low et al., 2013). While null findings for adherence could result from social desirability bias in teacher self-reports (Lillehoj et al., 2004), this study more poignantly directs attention to the role of participant responsiveness in predicting program outcomes. Such studies suggest that more detailed examination of dimensions of implementation is warranted.

The present study focuses on three dimensions of implementation: implementer quality of delivery, participant responsiveness, and adherence to curriculum content (fidelity). These dimensions are common to extant reviews implying they may be essential to understanding

implementation. We utilize extensive observational measures of these dimensions. Thus, this study advances prevention science by offering a detailed examination of quality of delivery, participant responsiveness, and adherence and relationships between these dimensions and outcomes of the *keepin' it REAL* program. The study hypothesizes that higher levels of implementation quality will predict better outcomes (H1). We also seek to understand how specific aspects of implementation quality (e.g., delivery quality, participant responsiveness) relate to outcomes (i.e., norms, efficacy, substance use) (RQ1).

## Methods

### Procedures

Data for this study are part of a larger investigation of adaptation processes and implementation of the *keepin' it REAL* (*kiR*) program (see Colby et al., 2013; Miller-Day et al., 2013; Pettigrew et al., 2013). The larger study examines adaptation for rural schools and implementation processes, seeking to evaluate the relative effectiveness of designer adapted (i.e., curriculum customized by developers to rural students) versus non-adapted (i.e., the original, multicultural/urban curriculum) versus controls (i.e., schools continuing prevention activities as usual and not receiving *keepin' it REAL*). The present study examines implementation of both curriculum versions and explores effects of adherence, teacher delivery quality, and student responsiveness on proximal program outcomes (i.e., norms, efficacy) and youth substance use.

For the present study, both the adapted and non-adapted versions of the *kiR* curriculum were implemented in school districts defined as rural by the National Center for Education Statistics in Pennsylvania (PA) and Ohio (OH) during the 2009–10 school year. Schools ( $n = 39$ ) were randomly assigned to control ( $n = 14$ ), classic ( $n = 11$ ), or rural ( $n = 14$ ) conditions (see Graham et al., 2013). During training, teachers in both treatment conditions were directed to set up a camcorder in the back of the room and record their implementation. Although they were provided different versions of the curriculum (adapted, non-adapted), teachers in both conditions were given the same training in delivery and were both allowed to modify the curriculum, which allowed us to examine implementer adaptation across conditions through video observations. Teachers were provided a \$10 incentive for completing a short on-line evaluation after each lessons and mailing videos of each lesson to project staff. Videos were collected during a single year of implementation and included data from 31 teachers in 25 schools. Treatment schools ranged in size from 194 to 1087 students ( $M = 552$ ,  $SD = 27$ ) and included 7<sup>th</sup> grade classrooms in elementary ( $n = 4$ ), middle ( $n = 7$ ), and high schools ( $n = 14$ ). The number of 7<sup>th</sup> graders in treatment schools ranged from 27 to 226 students ( $M = 99$ ,  $SD = 59$ ) (school data available from NCES Common Core of Data, 2006–2007 school year at <http://nces.ed.gov/ccd>).

Students from all conditions participated in four waves of data collection administered by a university survey research center. These analyses focus on the first two waves of surveys administered in the fall semester of participants' 7<sup>th</sup> grade year and subsequently in the spring semesters of that year. Spacing of surveys provided a baseline and immediate post-test following intervention delivery. For both waves of data collection, lists of students were obtained from participating schools. Passive parental consent was obtained for all

participating students. Project staff coordinated with schools to schedule a survey date on which surveys were administered on site. Student informed assent forms were provided in writing and read aloud to all students. Only assented students, whose parents did not deny permission, participated in the survey. One makeup date was scheduled for each school to collect data from students who were absent during the scheduled survey date. All procedures were approved by a university institutional review board.

### Teachers and Training

Participating teachers averaged about 13 years of teaching experience ( $M = 12.81$ ,  $SD = 9.04$ ) and were recruited by the Principal or Assistant Principal to participate voluntarily in the project. As reported elsewhere (Pettigrew et al., 2013) teachers in treatment conditions participated in a one-day training workshop during which they received copies of the curriculum manual. Manuals included detailed lesson plans, handouts and homework sheets as well as PowerPoint files that corresponded with each lesson. Training included an overview of research on youth drug use, a presentation of the theoretical model undergirding the curriculum, and a report of existing studies of the effectiveness of the *kiR* program. The training also included several activities that familiarized implementers with the design of the *kiR* curriculum as well as instruction on and practice in using the lessons. Finally, teachers received training about how to conduct the research activities related to the project, such as video recording.

### Measures

This study utilized both observational coding of videos to measure implementation and student self-reports of attitudes, beliefs, and behaviors. Implementation was measured by adherence, teacher engagement, student engagement, and teacher quality. Student outcomes include substance use related norms, efficacy, and substance use behaviors.

**Implementation quality**—Observational coding of teacher videos was used to measure implementation. Advantages of observation over self-report have been noted, as teachers often under-report adaptation and over-report fidelity (e.g., Miller-Day et al., 2013) and self-reports of implementation have not been as useful predictors of outcomes as observer reports (Lillehoj et al., 2004). Of the possible 780 videos (78 classes, 10 lessons per class), 688 teacher videos (88%) were returned to project staff. Given their number and length, a subset of videos were randomly selected for coding using a procedure designed for this study that allowed for a balanced selection (e.g., across lessons) of videos (for details, see Pettigrew et al., 2013). This procedure resulted in coding of 276 videos, which were watched and rated by trained coders using both qualitative and quantitative techniques to measure a variety of indicators, including implementation quality. Coder training included self-study of the coding manual, guided coding practice with feedback, and regular participation in coder meetings. An acceptable level of inter-coder reliability (see Hayes & Krippendorff, 2007) was set at .80 prior to the outset of coding, and was checked four times during the coding process by randomly selecting videos for double coding resulting in the following levels of agreement: .94, .93, .84, .92. A detailed coder's manual, standardized training, and coder meetings held at regular intervals were used to help maintain agreement as well as prevent

coder drift (for details, see also Pettigrew et al., 2013; Shin, Miller-Day, Pettigrew, & Hecht, 2011).

Variables used in this study were patterned after prior articulations of implementation quality (e.g., Dusenbury et al., 2005). Based on a review of the extant literature at the outset of the study, these dimensions included adherence, teacher engagement, a global rating of teacher delivery quality, and student engagement/responsiveness.

**Adherence:** Adherence was conceptualized as how closely the implementation matched the prescribed content of the curricula. Closely aligned with fidelity, adherence was assessed by coder ratings on a three point scale measuring how well teachers met the stated objectives of each lesson. Coders indicated very well, adequate, or poor. A rating of “Very Well” indicated that coders were fully confident that the listed objective was accomplished in the lesson. A rating of “Adequate” indicated that some of the listed objectives were accomplished. For example, information pertaining to the objective might have been introduced, but there was no practice, discussion, or indication that the students understood the concepts. A rating of “Poor” indicated that the listed objective was not accomplished. Adherence was rated for each of the selected lessons per class and averaged across lessons to create a mean level of adherence per class.

**Teacher engagement:** Teacher engagement was rated on five dimensions (attentiveness, enthusiasm, seriousness, clarity, positivity) each using a four-point scale. Attentiveness measured the degree to which teachers appeared to be attentive to student needs and nonverbal cues throughout the lesson. Enthusiasm tapped how energetically the teacher delivered lesson content. A third item, seriousness, measured of how seriously the teacher presented the lesson content (e.g., did they make sarcastic or critical jokes about objectives). Clarity rated the degree to which a teacher’s directions, explanations, and lecture were lucid, clear, or easy to understand. Finally, positivity measured a teacher’s verbal and nonverbal feedback provided to the students throughout a lesson (e.g., verbal praises, smiles, backchannel cues of support). Each dimension was averaged across lessons (e.g., mean level of attentiveness in all coded lessons). These were then averaged together to form the mean level of teacher engagement per class.

**Student engagement:** Student engagement was measured by two variables, both rated on a four-point scale. One measured how attentive students were to the lesson while the second measured the degree to which students participated in lesson activities. These two constructs were averaged across lessons to form class levels of attentiveness and participation. These were then averaged together to create a mean level of student engagement per class.

**Global teaching quality:** A single-item, global evaluation of delivery quality was also included, measured on a five-point scale ranging from poor to excellent. The global assessment judged the teachers overall excellence in the quality of delivery of the content and teaching technique in the lesson. To receive an excellent rating, the instructor adhered to the lesson’s content, materials, and objectives, students and teacher were engaged, and teaching was effective. The global teaching quality variable was computed by averaging across lessons.



**Outcome measures**—Student surveys took about 55-minutes to complete and consisted of approximately 146 items. The survey was administered in a 3-form design to maximize the number of items included in the questionnaire within the time allotted by the schools (Graham, Taylor, Olchowski, & Cumsille, 2006). Students completed one of three versions of the survey each consisting of approximately 109 items. One block of items (X) appeared on all the surveys and two additional blocks items (AB, AC, or BC) were given at random to different sets of students. Variables included in this study assess two proximal outcomes (efficacy, norms), which are hypothesized mediators of longer-term program outcomes, and students' substance use behaviors (drugs).

**Efficacy:** One of the goals of *keepin' it REAL* is to teach youth refusal skills to equip them with the confidence to use these skills during offer-response episodes. Using modified items from previous prevention research (Hansen & Graham, 1991), two aspects of efficacy were measured: self and response efficacy (see also Choi, Krieger, & Hecht, 2013). Eight items (four each for alcohol and marijuana) assessed students' self-efficacy, or their belief that they possessed the ability to use the REAL strategies (i.e., refuse, explain, avoid, or leave: see Alberts, Miller-Rassulo, & Hecht, 1991; Pettigrew, Miller-Day, Krieger, & Hecht, 2011) if they received a drug offer. Response efficacy, or the belief that the resistance strategies taught in the lessons will be effective, was assessed by four items. Scales were created for response efficacy and for self-efficacy (averaging within substance, and then averaging together). A summary variable (efficacy) was computed by equally-weighting and averaging self-efficacy and response-efficacy scales.

**Norms:** *kiR* lessons attempt to correct misperceptions about the prevalence middle-school students' substance use. Some content explicitly teaches about the prevalence of peer use while other content models non-use as an appropriate and desirable choice, potentially influencing participants' views on the acceptability of substance use. A norms scale was adapted from previous studies (Hansen & Graham, 1991) with four items that tapped into descriptive norms, or students' perceptions of peer use (e.g., out of 100 students your age, how many do you think drink alcohol), and four items adapted from the Communities that Care Youth Survey (Arthur, Hawkins, Pollard, & Catalano, 2002) that measured peer injunctive norms, or perceptions of the degree to which peers find adolescent substance use acceptable (e.g., do you think it is wrong for someone your age to smoke cigarettes). Scales were computed for both types of norms separately, then equally-weighted and averaged together to form a composite norms variable.

**Substance Use:** Substance use was measured by 13 items assessing use of alcohol, cigarettes, chewing tobacco, and marijuana from Hansen and Graham (1991). While not perfect, previous work has validated self-reported substance use behaviors showing 95–100% agreement with analysis of saliva samples (Ellickson & Bell, 1990; Needle, Jou, & Su, 1989). Thus, self-reported use of each substance was measured by three items that asked about lifetime use (e.g., how many cigarettes have you smoked in your entire life?) and recent use (e.g., how many cigarettes have you smoked in the past 30 days?; how many days in the past 30 days have your smoked cigarettes?). Appropriate responses were given for each item. For example, response choices for the amount of recent use ranged from none to

more than about once a day for alcohol, cigarettes, chewing tobacco, and marijuana. Frequency rates for all substances ranged from none to 16 to 30 days. We included one additional question about alcohol that asked students to report the number of times they had five or more drinks in a row in the past two weeks. Scales were created for each substance by equally-weighting and averaging items related to each substance together. The drugs variable was computed as an equally-weighted average of alcohol, cigarettes, marijuana, and chewing tobacco scales.

## Analytical Plan

**Summary variables**—In order to test the appropriateness of treating each implementation quality indicator separately, an exploratory factor analysis (Principle Axis Factoring with promax rotation) of teacher engagement variables (attentiveness, enthusiasm, seriousness, clarity, positivity), student engagement variables (attentiveness, participation), and global teaching quality was conducted. This analysis showed that the items loaded onto a single factor that explained 57% of the variance. To confirm the factor structure, we reran the factor analysis using the combined variables rather than the dimensions of each construct; that is, we used teacher engagement, student engagement, and global teaching quality rather than, for example, teacher attentiveness, enthusiasm, seriousness, clarity, and positivity. This analysis again showed a single factor that explained 77% of the variance. Thus, a composite delivery variable was computed using a weighted average of student engagement, teacher engagement, and global teaching quality such that teacher engagement was given twice the weight as student engagement and global teaching quality. The weighting scheme was judged appropriate because (a) coded videos primarily focused on teachers' verbal and nonverbal behaviors rather than students' because video cameras were typically set up in the rear of the classroom and captured only the backs of students; and (b) teachers had a greater role in the delivery of the curriculum (e.g., introducing concepts and activities, managing time, etc.) indicated by the increased number of variables used to measure teacher engagement ( $n = 5$ ) compared with student engagement ( $n = 2$ ) (see also Pettigrew et al., 2013). Our weighting scheme doubled the standardized teacher engagement variable and averaged it with standardized student engagement and global teaching quality variables to compute a composite delivery variable.

Owing to the high correlation between the composite delivery variable and the adherence variable ( $r = .73$ ), a second factor analysis (principle axis with promax rotation) was computed. Both items loaded onto a single factor that explained 87% of the variance, so a single implementation quality summary variable (IQsum) was computed by equally weighting and averaging delivery and adherence.

A single summary outcome variable (DVsum) was also computed to allow for an omnibus test of implementation quality on outcomes. The summary outcome variable was the equally-weighted average of efficacy, norms, and drugs and was computed for both waves of data (with efficacy reverse coded for this weighted average).

**Linking data sets**—Prior to analyses, we linked the implementation data (i.e., coded video data) with student survey data. Under ideal circumstances we would be able to match



individual student scores with the specific classrooms in which the program was delivered. However, circumstances precluded a perfect match for all classrooms. To account for the imperfect match, we determined that as long as a match was likely between implementation data and individual student data (e.g., a teacher delivered *kiR* to only one class of students within a school, teachers delivered *kiR* to multiple classes but provided information linking the video and student data, teacher delivered information to multiple classes but information linking video and student data could be inferred from information provided by teachers), we used individual class level data. However, where a match was unlikely (e.g., teachers taught *kiR* to more than one class but provided inconsistent, little or no corroborating information), we used data for the entire school. Following these guidelines, we matched 68% of the cases using individual class level data and for the remaining 32% of cases used school level outcome data.

**Missing data**—Missing data occurred almost exclusively within the data for the individual students. In order to take best advantage of partial data, and to limit biases, we employed multiple imputation (MI) with SAS (version 9.2) Proc MI following guidelines in Graham (2012). In order to take the school structure into account, we created dummy variables for school membership to be included in the imputation analysis. In order to maximize statistical power compared to the equivalent ML analysis, we generated a total of  $m = 200$  data sets (see Graham, Olchowski, & Gilreath, 2007). Following Graham (2012), half of these were imputed including the school-membership dummy codes, and half were imputed omitting the dummy variables. This procedure has been found to produce realistic, but slightly conservative estimates of the intraclass correlation (ICC). For both models, the EM convergence criterion was set to .00001, following Graham (2012). When the dummy variables were included, EM (MLE) converged normally in 37 iterations. When dummy variables were omitted, EM (MLE) converged normally in 31 iterations. For imputation with both models, we set the number of MCMC burn-in iterations to 500, and the number of MCMC iterations between imputed data sets to 100.

## Results

To examine classroom level effects, statistically controlling for school level effects, we used a mixed model design. We used Proc Mixed in SAS (version 9.2) to conduct the analyses. All analyses controlled for initial levels of the dependent variable.

We were interested in exploring how delivery and adherence impacted drugs, norms, and efficacy. However, conducting six different tests would require some adjustment to prevent capitalizing on chance. Rather than altering the significance level (e.g., through a Bonferroni adjustment) we opted to run a single omnibus test using the composite implementation quality measure (IQsum) predicting the composite DV (DVsum). This test controlled for experiment-wise type I error; it helped rule out the possibility that none of the six effects was significant. We reasoned that if the omnibus test was significant we would be justified in exploring the individual relationships among delivery quality and adherence on the three outcomes without further adjustment.

The omnibus analysis showed that the combined implementation quality variable significantly predicted outcomes controlling for pre-intervention levels,  $t(44) = -2.07, p = .044$ . Thus, results support H1 that higher levels of implementation quality predicted better outcomes; better implementation was related in the predicted direction to the outcome variables.

Given the significant relationship between implementation quality and outcomes, we were justified in examining how specific dimensions of implementation quality related to outcomes (RQ1). We computed six additional tests to examine separately the effects of delivery and adherence on each of the three dependent variables (efficacy, norms, and substance use). Statistics associated with these tests are reported in Table 1. Results show that better delivery quality significantly predicted more conservative norms and less substance use among adolescents whereas better adherence to the curriculum significantly predicted more conservative norms and was marginally significantly related to less substance use. Efficacy was not predicted by delivery quality or adherence.

Finally, the findings in response to RQ1 that delivery and adherence predicted drugs and norms justified further decomposing the outcomes to examine what effect, if any, the implementation variables had on the individual variables making up these outcomes. Rather than look at effects on drugs, for example, we looked at effects on alcohol, smoking tobacco, chewing tobacco, and marijuana. We imputed a new data set following the same procedures used for the previous test (a new imputation model was required due to the nested nature of the data; one imputation could not incorporate both the composite and individual levels of variables; Graham, 2012). We included implementation quality variables (i.e., delivery, adherence) as well as both pre-intervention and post-intervention variables for drugs (i.e., alcohol, smoking tobacco, marijuana, chewing tobacco) and norms (i.e., perceptions of peer acceptability, perceptions of peer prevalence). With this data set we computed mixed model regression using Proc Mixed in SAS. Results of these tests are also reported in Table 1.

Delivery significantly predicted less use of alcohol and marijuana and marginally significantly predicted less use of smoking and chewing tobacco. Better delivery also significantly predicted conservative perceptions of peer prevalence but not perceptions of peer acceptability. Adherence was significantly related to less alcohol use but was not significantly related to any other type of drug use. Adherence was also significantly related to perceptions of peer prevalence but not perceptions of peer acceptability.

## Discussion

The current study presents a method for more comprehensively and objectively measuring implementation quality and confirms previous findings that demonstrate that implementation affects outcomes. The study adds to existing knowledge by (a) presenting results of a quasi-experimental test of *keepin' it REAL* and (b) describing procedures for measuring implementation quality (adherence, delivery) as well as its relationships with outcomes.

### Quasi-experimental test of *kiR*

This study presents a quasi-experimental evaluation of the *kiR* curriculum. Findings show that when delivered well, the program has a better effect on proximal outcomes and substance use compared to when the program is delivered poorly. When well implemented, the program showed significant reductions in substance use even at the end of 7<sup>th</sup> grade immediately following curriculum delivery. Upon close examination, findings show that adherence (i.e., meeting curriculum objectives) was significantly related to reduced alcohol use whereas better delivery quality (e.g., higher levels of teacher and student engagement) either significantly or marginally predicted reductions in all types of substance use. This finding underscores the importance of engaging classrooms and aligns with a seminal meta-analysis that showed interactive prevention programs outperformed non-interactive programs (Tobler et al., 2000). Absent engaged classroom behavior (i.e., high delivery quality), high fidelity to the curriculum may only have a limited effect on adolescent substances use. Another study (Low et al., 2013) also demonstrated that engagement with a prevention program trumped adherence in predicting outcomes.

It is possible to interpret these findings as devaluing the importance of prevention curricula and emphasizing individual characteristics of a teacher, however, both are necessary. Without an efficacious curriculum, high implementation quality logically will result in null or negative program outcomes. For example, Sanchez et al. (2007) found that observer reports of adherence, dose, and delivery quality predicted negative outcomes (e.g., anger, alcohol use), which corresponded with overall negative results of the program at a six-month follow-up. Participant responsiveness (measured as peer and teacher support) was the only implementation variable they found to predicted positive outcomes. If a poorly conceived or non-efficacious program is well implemented it will have no effects, or worse, will result in negative effects. Both the program and its implementation are crucial, although engagement with an efficacious curriculum may be more important than mere adherence.

Results also demonstrated significant effects on norms, such that a well delivered program resulted in more conservative beliefs about peer use. Delivery and adherence both affected perceptions of peer prevalence but neither affected perceptions of peer acceptability. Thus, although meeting program objectives and delivering the program well provides youth with a more accurate assessment of how much substance use takes place among peers, neither has an impact on whether or not youth see peer use as right or wrong. Since *kiR* presents substances use as a risk with potentially negative consequences and not as an inherent wrong, it follows that the program would have little effect on this variable. However, it is likely, too, that because the curriculum teaches decision making in light of risks and consequences as well as presenting a positive image of substance-free activities, as youth who receive the curriculum age and gain either personal or vicarious exposure to the consequences of substance use that they will come to see it as unadvised, if not unacceptable.

The study found no effects for efficacy. The program is designed to increase both self and response efficacy as these constructs are associated with decreased substance use (Choi et al., 2013). Counter to expectations, delivery and adherence were not significant predictors of

efficacy in the current study. We speculate that an immediate post-test is too soon to find changes in the aggregate of these variables among young adolescents. Developing efficacy may be a slow moving process that requires students gain exposure to situations demanding them to apply (or not) refusal skills (see Bandura & Wood, 1989; Choi et al., 2013). Using a finer grain measure would track these developmental changes. Since this study reports only two waves of data from youth in the 7<sup>th</sup> grade, it cannot test this hypothesis. Second, since we used an omnibus test to justify decomposition of the variables, we did not look at effects on self and response efficacy separately. It seems reasonable immediately following the program to expect greater effects on self-efficacy (i.e., the ability to use the REAL strategies) with lesser effects on response efficacy (i.e., the effectiveness of the REAL strategies); self-efficacy may be a skill that is learned and practiced in the abstract whereas response efficacy may be a skill that is tested through application to concrete substance offer situations. As youth age and gain exposure to personal and peer use, a well implemented program could show effects on efficacy over a poorly implemented program.

Although only quasi-experimental, together these findings demonstrate effects for the *kiR* program. This study is not a program evaluation since it does not compare treatment conditions to control conditions. Nor does this study perform a mediation analysis of implementation quality (IQ → efficacy, norms → substances use) as it only includes two waves of data. This study lacks a long-term follow-up and two versions of *kiR* were considered together in these analyses. A full examination of effects should account for implementation quality and look at the researcher adapted version (rural) versus the non-adapted version (classic) versus the control condition and should study effects over a longer period of time. Based on this study and others, researchers ought to include implementation as a mediator of program outcomes during intent-to-treat evaluations.

### Studying implementation quality

Our findings, more broadly, add to the emerging body of knowledge about various dimensions of implementation quality. Studies and reviews of implementation quality consistently include, at a minimum, adherence, participant responsiveness, and delivery quality (e.g., Berkel et al., 2011; Dusenbury et al., 2003; Hansen et al., 1991). These dimensions have been treated as separate in models of implementation processes on program evaluation (e.g., Berkel et al., 2011); however, practical experience (Rohrbach, Gunning, Sun, and Sussman, 2010) and empirical studies have not tended to support that these can be treated separately. For example, Hansen et al. (1991) found one factor of implementation quality as did Rohrbach, Graham, and Hansen (1993). In our study we reliably rated dimensions separately. Nevertheless, when subjected to factor analysis participant responsiveness and teacher delivery quality created only a single factor. Delivery was highly correlated with adherence ( $r = .73$ ) indicating that these two dimensions may legitimately be considered part of the same construct.

There are both conceptual and methodological explanations for the single-factor model of implementation quality. Conceptually, even though rated separately, dimensions of implementation quality are expressed simultaneously. Teacher and student engagement are two measures of a single phenomenon unfolding at the same time, or at least in very rapid

sequence. Perhaps this is why empirical studies of responsiveness, delivery quality, and adherence find only one implementation factor. Methodologically, this study averaged across lessons to create variables that characterized a teachers' overall delivery, adherence to the curriculum, and students' overall responsiveness. Averaging tends to obscure differences among the dimensions. Thus, we may find more of a distinction within single lessons, but with the across-lesson averages we would expect less distinction. Future research examining implementation during a single lesson might investigate the distinctiveness of these variables.

Our findings also demonstrates advantages for measuring implementation quality using observations of video data. Observations, compared to self-reports, help decrease bias, as they have been shown to be more accurate and more variable (e.g., Hansen et al., 1991; Lillehoj et al., 2004; Miller-Day et al., 2013). For example, Hansen et al. (1991) reported that in their pilot study of implementation quality, self-reports of the degree to which implementers met program goals and their enthusiasm delivering the program averaged to above 6 on a 7-point scale. Videos may have also contributed to more "natural" teaching behaviors than if an observer was present. Training teachers then releasing them to deliver *kiR* without intensive oversight may have facilitated variability in program delivery. As Dane and Schneider (1998) note, when "investigators relinquish control of the implementation..., inconsistencies in program delivery become more likely" (p. 24). Recordings also allowed us to easily aggregate observations across multiple lessons and multiple coders, which has been shown to enhance reliability of measuring teacher behaviors (Ho & Kane, 2013). Extensive coding of video data to measure implementation quality was, in these ways, beneficial.

Findings illustrate the importance of a well delivered program. When students receive a poorly delivered program it is like they are not receiving all of an intervention while participating in a well delivered program is like receiving the full treatment. For example, if teachers only present half the curriculum material or design (i.e., low adherence), it is as if their students only receive half the program. If students are inattentive or do not participate (i.e., low delivery quality), they are not receiving the entirety of the program. To aggregate data across teachers with highly varied implementation quality will potentially obscure real program effects (type III error, Dusenbury et al., 2003) because it is like looking at effects for patients who only participated in half a treatment alongside those who received a full treatment.

Findings from this study have important implications for program design, training, and implementation processes. First, instructors' manuals need to be clearly written and easily executed. The *keepin' it REAL* manual was developed with teachers and other implementers, but further research should be conducted to check implementer understanding and ability to easily follow the manual. Thought might be given as to whether materials incorporated into the program assure a high level of student engagement (i.e., a poor curriculum cannot be delivered well). Similarly, training should emphasize the importance of delivery quality and student indicators (engagement, etc.) of success. Selected videos collected from this research are being integrated into *kiR* training to accomplish this purpose. Finally, ongoing technical support is needed to reinforce and possibly even reteach the curriculum to implementers.

The support system should include ongoing feedback based on implementer input (i.e., how they implement each lesson) as well as social networking opportunities for implementers to share experiences and obtain and provide reinforcement.

## Conclusion

In summary, this paper contributes to a growing body of research demonstrating the importance of implementation research as a component of prevention science. While clearly developing efficacious and effective curriculum is vital, if not implemented well they are unlikely to achieve their prevention goals. Thus, implementation research, such as the current study, needs to stand alongside curriculum development as two of the pillars of the emerging science of prevention. These findings not only contribute to our understanding of how such research should be conducted, but, as well, the implications of implementation quality for prevention outcomes.

## Acknowledgments

This publication was supported by Grant Number R01DA021670 from the National Institute on Drug Abuse to The Pennsylvania State University (Michael Hecht, Principal Investigator).

## References

- Alberts JK, Miller-Rassulo M, Hecht ML. A typology of drug resistance strategies. *Journal of Applied Communication Research*. 1991; 19:129–151.10.1080/00909889109365299
- Arthur MW, Hawkins JD, Pollard JA, Catalano RF, Baglioni AJ. Measuring risk and protective factors for use, delinquency, and other adolescent problem behaviors the communities that care youth survey. *Evaluation Review*. 2002; 26:575–601.10.1177/0193841X0202600601 [PubMed: 12465571]
- Bandura A, Wood RE. Effect of perceived controllability and performance standards on self-regulation of complex decision making. *Journal of Personality and Social Psychology*. 1989; 56:805–814. [PubMed: 2724068]
- Berkel C, Mauricio A, Schoenfelder E, Sandler I. Putting the pieces together: An integrated model of program implementation. *Prevention Science*. 2011; 12:23–33.10.1007/s11121-010-0186-1 [PubMed: 20890725]
- Botvin GJ, Baker E, Dusenbury L, Botvin EM, Diaz T. Long-term follow-up results of a randomized drug abuse prevention trial in a white middle-class population. *Journal of the American Medical Association*. 1995; 273:1106–1112. [PubMed: 7707598]
- Colby M, Hecht ML, Miller-Day M, Krieger JR, Syvertsen AK, Graham JW, Pettigrew J. Adapting school-based substance use prevention curriculum through cultural grounding: A review and exemplar of adaptation processes for rural schools. *American Journal of Community Psychology*. 2013; 1–2:190–205.10.1007/s10464-012-9524-8
- Choi HJ, Krieger JL, Hecht ML. Reconceptualizing efficacy in substance use prevention research: Refusal response efficacy and drug resistance self-efficacy in adolescent substance use. *Health Communication*. 2013; 28:40–52.10.1080/10410236.2012.720245 [PubMed: 23330857]
- Dane AV, Schneider BH. Program integrity in primary and early secondary prevention: are implementation effects out of control? *Clinical Psychology Review*. 1998; 18:23–45.10.1016/S0272-7358(97)00043-3 [PubMed: 9455622]
- Domitrovich CE, Greenberg MT. The study of implementation: Current findings from effective programs that prevent mental disorders in school-aged children. *Journal of Educational and Psychological Consultation*. 2000; 11:193–221.10.1207/S1532768XJEPC1102\_04
- Durlak JA. Why program implementation is important. *Journal of Prevention & Intervention in the Community*. 1998; 17:5–18.10.1300/J005v17n02\_02



- Durlak J, DuPre E. Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*. 2008; 41:327–350.10.1007/s10464-008-9165-0 [PubMed: 18322790]
- Dusenbury L, Brannigan R, Falco M, Hansen WB. A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research*. 2003; 18:237–256.10.1093/her/18.2.237 [PubMed: 12729182]
- Dusenbury L, Brannigan R, Hansen WB, Walsh J, Falco M. Quality of implementation: Developing measures crucial to understanding the diffusion of preventive interventions. *Health Education Research*. 2005; 20:308–313.10.1093/her/cyg134 [PubMed: 15522898]
- Ellickson PL, Bell RM. Drug prevention in junior high: a multi-site longitudinal test. *Science*. 1990; 247:1299–1305.10.1126/science.2180065 [PubMed: 2180065]
- Graczyk PA, Domitrovich CE, Zins JE. Facilitating the implementation of evidence-based prevention and mental health promotion efforts in schools. *Handbook of School Mental Health Advancing Practice and Research*. 2002:301–318.10.1007/978-0-387-73313-5\_21
- Graham, JW. *Missing data: Analysis and design*. New York, NY: Springer; 2012.
- Graham J, Olchowski A, Gilreath T. How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*. 2007; 8:206–213.10.1007/s11121-007-0070-9 [PubMed: 17549635]
- Graham JW, Pettigrew J, Miller-Day M, Krieger JL, Zhou J, Hecht ML. Random assignment of schools to groups in the drug resistance strategies rural project: Some new methodological twists. *Prevention Science*. 2013.10.1007/s11121-013-0403-9
- Graham JW, Taylor BJ, Olchowski AE, Cumsille PE. Planned missing data designs in psychological research. *Psychological Methods*. 2006; 11:323–343.10.1037/1082-989X.11.4.323 [PubMed: 17154750]
- Hallfors D, Godette D. Will the “Principles of Effectiveness” improve prevention practice? Early findings from a diffusion study. *Health Education Research*. 2002; 17:461–470.10.1093/her/17.4.461 [PubMed: 12197591]
- Hansen WB, Graham JW, Wolkenstein BH, Rohrbach LA. Program integrity as a moderator of prevention program effectiveness: Results for fifth-grade students in the adolescent alcohol prevention trial. *Journal of Studies on Alcohol and Drugs*. 1991; 52:568–579.
- Hansen WB, Graham JW. Preventing alcohol, marijuana, and cigarette use among adolescents: Peer pressure resistance training versus establishing conservative norms. *Preventive Medicine*. 1991; 20:414–430.10.1016/0091-7435(91)90039-7 [PubMed: 1862062]
- Hayes AF, Krippendorff K. Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*. 2007; 1:77–89.10.1080/19312450709336664
- Hecht ML, Graham JW, Elek E. The drug resistance strategies intervention: program effects on substance use. *Health Communication*. 2006; 20:267–276.10.1207/s15327027hc2003\_6 [PubMed: 17137418]
- Ho, AD.; Kane, TJ. *The Reliability of Classroom Observations by School Personnel*. 2013. Retrieved from [http://metproject.org/downloads/MET\\_Reliability\\_of\\_Classroom\\_Observations\\_Research\\_Paper.pdf](http://metproject.org/downloads/MET_Reliability_of_Classroom_Observations_Research_Paper.pdf)
- Knoche LL, Sheridan SM, Edwards CP, Osborn AQ. Implementation of a relationships-based school readiness intervention: A multidimensional approach to fidelity measurement for early childhood. *Early Childhood Research Quarterly*. 2010; 25:299–313.10.1016/j.ecresq.2009.05.003 [PubMed: 20824112]
- Lillehoj CJ, Griffin KW, Spoth R. Program provider and observer ratings of school-based preventive intervention implementation: agreement and relation to youth outcomes. *Health Education & Behavior*. 2004; 31:242–257.10.1177/1090198103260514 [PubMed: 15090124]
- Low S, Ryzin MJV, Brown EC, Smith BH, Haggerty KP. Engagement matters: Lessons from assessing classroom implementation of steps to respect: a bullying prevention program over a one-year period. *Prevention Science*. 2013.10.1007/s11121-012-0359-1
- Marsiglia FF, Kulis, Yabiku S, Nieri ST, Coleman TAE. When to intervene: elementary school, middle school or both? Effects of *keepin' it REAL* on substance use trajectories of mexican

- heritage youth. *Prevention Science*. 2011; 12:48–62.10.1007/s11121-010-0189-y [PubMed: 21128119]
- Miller-Day M, Pettigrew J, Hecht ML, Shin MY, Graham J, Krieger J. How prevention curricula are taught under real-world conditions: Types of and reasons for teacher curriculum adaptations in 7th grade drug prevention curriculum. *Health Education*. 2013; 113:324–344.10.1108/09654281311329259
- Needle RH, Jou SC, Su SS. The impact of changing methods of data collection on the reliability of self-reported drug use of adolescents. *The American Journal of Drug and Alcohol Abuse*. 1989; 15:275–289.10.3109/00952998908993408 [PubMed: 2788363]
- Odom SL, Fleming K, Diamond K, Lieber J, Hanson M, Butera G, Marquis J. Examining different forms of implementation and in early childhood curriculum research. *Early Childhood Research Quarterly*. 2010; 25:314–328.10.1016/j.ecresq.2010.03.001 [PubMed: 21874091]
- Pentz MA, Trebow EA, Hansen WB, Mackinnon DP, Dwyer JH, Johnson CA, Cormack C. Effects of program implementation on adolescent drug use behavior: The Midwestern Prevention Project (MPP). *Evaluation Review*. 1990; 14:264–289.
- Pettigrew J, Miller-Day M, Krieger J, Hecht ML. Alcohol and other drug resistance strategies employed by rural adolescents. *Journal of Applied Communication Research*. 2011; 39:103–122.10.1080/00909882.2011.556139 [PubMed: 21552345]
- Pettigrew J, Miller-Day M, Shin Y, Hecht ML, Krieger JR, Graham JW. Describing teacher-student interactions: A qualitative assessment of teacher implementation of the 7th grade *keepin' it REAL* substance use intervention. *American Journal of Community Psychology*. 2013.10.1007/s10464-012-9539-1
- Rohrbach LA, Graham JW, Hansen WB. Diffusion of a school-based substance abuse prevention program: Predictors of program implementation. *Preventive Medicine*. 1993; 22:237–260.10.1006/pmed.1993.1020 [PubMed: 8483862]
- Rohrbach L, Gunning M, Sun P, Sussman S. The Project Towards No Drug Abuse (TND) dissemination trial: Implementation fidelity and immediate outcomes. *Prevention Science*. 2010; 11:77–88.10.1007/s11121-009-0151-z [PubMed: 19757052]
- Ringwalt C, Vincus A, Hanley S, Ennett S, Bowling J, Haws S. The prevalence of evidence-based drug use prevention curricula in U.S. middle schools in 2008. *Prevention Science*. 2011; 12:63–69.10.1007/s11121-010-0184-3 [PubMed: 20683664]
- Sanchez V, Steckler A, Nitirat P, Hallfors D, Cho H, Brodish P. Fidelity of implementation in a treatment effectiveness trial of Reconnecting Youth. *Health Education Research*. 2007; 22:95–107. [PubMed: 16807378]
- Shin, Y.; Miller-Day, M.; Pettigrew, J.; Hecht, ML. Qualitative approach to implementer typologies: How teachers and students interact in implementation of school based prevention intervention. Paper presented at the meeting of the Society for Prevention Research; Washington D.C. 2011 Jun.
- Tobler NS, Roona MR, Ochshorn P, Marshall DG, Streke AV, Stackpole KM. School-based adolescent drug prevention programs: 1998 meta-analysis. *The Journal of Primary Prevention*. 2000; 20:275–336.10.1023/A:1021314704811

**Table 1**

## Mixed Regression Model of Implementation Quality on Outcomes

| Test  | <i>t</i> | <i>p</i> |
|---|----------|----------|
| IQsum → DVsum                                 | -2.07    | .044     |
| delivery → drugs                              | -2.96    | .005     |
| delivery → norms                              | -2.23    | .031     |
| delivery → efficacy                           | .81      | ns       |
| adherence → drugs                             | -1.65    | .106     |
| adherence → norms                             | -2.44    | .019     |
| adherence → efficacy                          | 1.29     | ns       |
| delivery → alcohol                            | -3.48    | .001     |
| delivery → smoking                            | -1.88    | .067     |
| delivery → marijuana                          | -2.11    | .041     |
| delivery → chewing tobacco                    | -1.85    | .071     |
| delivery → perceptions of peer acceptability  | -1.11    | ns       |
| delivery → perceptions of peer prevalence     | -2.98    | .005     |
| adherence → alcohol                           | -2.66    | .011     |
| adherence → smoking                           | -0.70    | ns       |
| adherence → marijuana                         | -1.19    | ns       |
| adherence → chewing tobacco                   | -0.84    | ns       |
| adherence → perceptions of peer acceptability | -1.32    | ns       |
| adherence → perceptions of peer prevalence    | -3.28    | .002     |

Note. *df* = 44 for all *t* tests shown; ns denotes *p* values greater than .20;

IQsum = summary variable, equal-weighted average of delivery and adherence

DVsum = equally weighted average of efficacy, norms, drugs

delivery = weighted average (2:1:1) of teacher engagement, student engagement, teacher quality

adherence = coder rating of degree to which lesson objective were met

drugs = equally weighted average of alcohol, cigarettes, marijuana, and chewing tobacco scales

norms = equally weighted average of perceptions of peer prevalence and peer acceptability

efficacy = equally weighted average of self-efficacy and response-efficacy scales