

## Measuring Quality of Delivery in a Substance Use Prevention Program

Steven Giles · Julia Jackson-Newsom ·  
Melinda M. Pankratz · William B. Hansen ·  
Christopher L. Ringwalt · Linda Dusenbury

Published online: 22 November 2008  
© Springer Science+Business Media, LLC 2008

**Abstract** The purpose of this study was to develop and validate an observation measure designed to capture teachers' use of interactive teaching skills within the delivery of the All Stars substance use prevention program. Coders counted the number of times teachers praised and encouraged students, accepted and used students' ideas, asked questions, self-disclosed personal anecdotes, and corrected student misbehavior. These teacher behaviors loaded on three factors: classroom management, acknowledgment, and student-centered methods. Classroom management was negatively related to student engagement. Acknowledgment was negatively related to students' normative beliefs. Student-centered methods were positively related to student idealism and normative beliefs, and marginally predicted decreases in student marijuana use. *Editors' Strategic Implications:* The authors provide a promising approach to studying pedagogical prevention approaches, and they also link teaching processes to student outcomes. This study of program delivery should be of general interest (i.e., not limited to substance use prevention) to practitioners and researchers.

**Keywords** Interactivity · Program delivery · Prevention · Substance use

---

S. Giles (✉)  
Wake Forest University, Winston-Salem, NC, USA  
e-mail: gilessm@wfu.edu

J. Jackson-Newsom · W. B. Hansen · L. Dusenbury  
Tanglewood Research, Greensboro, NC, USA

M. M. Pankratz · C. L. Ringwalt  
Pacific Institute for Research and Evaluation, Chapel Hill, NC, USA

## Introduction

In their seminal paper, Tobler and Stratton (1997) identified interactivity as crucial to achieving prevention objectives, including improving drug use attitudes and preventing drug use (Tobler 1986, 1992, 2000; Tobler et al. 1999; Tobler and Stratton 1997). For Tobler and colleagues, interactivity was a function of the activities specified in the curriculum's lessons. From this perspective, the inclusion of methods such as role-plays, games, small group activities, and class discussions qualified a program as interactive. Although there is general agreement that such activities are essential to program effectiveness, little research attention has been paid to the mechanisms by which teachers' delivery of interactive programmatic content may be measured.

The purpose of this study was to develop an observation system to quantify the frequency of specific teaching behaviors within the All Stars substance use prevention program. Our primary objectives were to develop a measure that could (a) achieve adequate inter-rater reliability, (b) assess generally applied teaching skills, not only lesson-specific teaching practices, and (c) help us understand the relationships among particular teaching practices, as well as understand how these practices predict programmatic outcomes.

## Interactive Delivery Skills and Student Outcomes

In the context of school-based substance abuse prevention curricula, interactivity refers to the degree to which the program guides for such curricula specify that teachers should engage students and invite discussion and other types of involvement (Dusenbury and Falco 1995). Program developers have frequently sought to include teaching methods that have the potential to promote interactivity. For example, in a review of ten drug prevention curricula, Bosworth and Sailes (1993) found that three quarters of the activities prescribed by prevention curricula required active student involvement.

Prior attempts to measure interactivity have used both teachers' self-reports and observers' ratings as sources of data. The types of measures used included (a) assessment of whether teaching practices were delivered as intended (Abbott et al. 1998); (b) ratings of teachers' effectiveness and enthusiasm in the classroom (Botvin et al. 1989; Hansen et al. 1991); (c) assessments of the quality of teaching strategies used, such as proactive classroom management (Harachi et al. 1999); (d) ratings of the degree to which the instructor involved the class in discussion (Sobel et al. 1989); and (e) ratings of teaching quality and interactivity (Hansen 1996; Pentz et al. 1990). However, there is little understanding of what specific teacher behaviors enhance interactivity and contribute to program effectiveness. That is, there is no consensus as to standardized ways by which such constructs as "quality," "student engagement," "acceptance of students' ideas," and "enthusiasm" (e.g., Dusenbury et al. 2005) may be effectively and consistently assessed or how these constructs map onto program outcomes. The measures used have tended

to rely upon vague operationalizations of interactivity that are often subjective and difficult to replicate.

Two studies investigated the relationship between program outcomes and interactive teaching practices. Hansen and colleagues (1991) found that observer reports of teachers' control and enthusiasm in teaching the Adolescent Alcohol Prevention Trial curriculum predicted students' involvement in class discussions and improvements in their ability to resist peer pressure. Harachi and her colleagues (1999) examined teaching practices that promoted positive student involvement and classroom management and improved social competency and bonding to school. The authors created positive and negative summary scale scores for six primary categories of interactivity: (a) proactive classroom management by teachers; (b) motivation to teach; (c) students' involvement; (d) cooperative learning; (e) reading; and (f) social skills reinforcement. Summary positive scales consisted of teacher behaviors that were believed to lead to high-quality implementation and the attainment of program goals. Summary negative scales consisted of those behaviors that detracted from high-quality implementation and program goals. Teachers who used positive involvement practices, such as checking for student understanding and assessing student engagement in a task, generated the greatest gains in students' social competency and decreases in antisocial behavior. Teachers' use of positive classroom management strategies, such as articulating clear directions that are easily followed by the class, were also associated with desired changes in students' social competency and bonding to school.

Prior research studies primarily have relied on summary measures of teacher behavior—i.e., single item measures designed to summarize an entire class session based on observers', teachers', or students' overall impressions. These measures tend to be based on subjective data that do not facilitate replication or an understanding of how specific behaviors contribute to or detract from interactivity. Absent from the literature is the assessment of discrete behaviors, such as how teachers perform during question and answer periods and how they respond to students' unsolicited comments. Such behaviors may differentially affect student outcomes.

## All Stars

All Stars is a school-based prevention curriculum that was recognized as “model” by the Substance Abuse and Mental Health Services Administration in 2001 (Substance Abuse and Mental Health Services Administration 2008) and as a promising program by the U.S. Department of Education. The program's goal is to reduce adolescents' participation in problematic health behaviors, including tobacco, alcohol, marijuana, and inhalant use. Program outcomes depend on affecting five mediating variables related to adolescent risk behavior (McNeal et al. 2004), including change in normative beliefs (acceptability and prevalence of problem behaviors among peers), lifestyle incongruence (realization that high risk behavior is incompatible with one's ideals), commitment to avoid high-risk behaviors, bonding to school and other prosocial institutions, and positive parental

attentiveness. All Stars includes 13 classroom sessions that prescribe many interactive activities including games, small group activities, and class discussions. Previously published evaluation results repeatedly have yielded evidence that All Stars affects both the mediating variables it targets and its substance use outcomes (Harrington et al. 2001; McNeal et al. 2004).

## Methods

### Participants

Forty-eight teachers and their respective seventh grade students in the Chicago area participated in this study. Teachers administered the program over the course of the academic year for up to three consecutive years. They received the standard 2 days of training and had access to the master All Stars trainer upon request, as well as web-based support, throughout the study. About half also received onsite, personalized coaching as a part of the parent study (Ringwalt et al. 2007), although this intervention was found to be ineffective. Most were classroom teachers (68.8%), and the remainder included guidance counselors, social workers, physical education teachers, and teaching assistants. They averaged 9.7 years of experience in the education field. A small majority had a graduate degree (52.1%). Teachers were predominately female (76.5%), and primarily White (58.8%). All were inexperienced in All Stars at the time they were recruited into the study, and the majority (62.5%) had not previously taught any substance use prevention program. Teachers videotaped each All Stars lesson they delivered, and their students completed pretest and posttest measures to assess change in the mediators and substance use outcomes targeted by the curriculum. All students were in the 7th grade, averaged 12.7 years of age, and were predominately African American (56.7%) or Hispanic (26.9%).

From this pool, we created two participant samples. Our first cohort of teachers served as a development sample ( $n = 17$ ), for which we used videotapes of Lesson 11, “Defending Commitments,” to develop the measure. We then used Lesson 8, “Norms—Unwritten Rules of Behavior,” to validate the instrument with the entire study sample (48 teachers, 107 implementations). Although teachers videotaped all 13 All Stars classroom sessions, we chose to code only one lesson in each of our samples because we believed that teachers’ general teaching skills would be quite stable across lessons. The first lesson that we coded did not occur until about half of the All Stars program had been delivered. We did this to reduce the likelihood of the Hawthorne effect. Discussions with teachers who participated in a pilot study revealed that they grew increasingly comfortable teaching in front of a camera, such that by Lesson 8 they hardly realized that they were being observed. We also selected both of these lessons for coding because they required more interaction on the part of the teacher than other lessons in the curriculum.

With the exception of teachers’ race/ethnicity, the development and the validation samples were similar. The majority of teachers in the development

sample were White (58.8%), while the teachers in the validation sample were nearly evenly divided between African American (45.8%) and White (41.7%) teachers.

### Interactivity Measurement Development

We designed an interactivity measure based on the early work of Flanders (1970), who developed an “Interaction Analysis” measure to assess teacher–student transactions. Specifically, our initial measure comprised seven categories of teacher-directed behavior: (a) praising and encouraging students; (b) accepting and using ideas of students; (c) asking questions; (d) sharing personal self-disclosures; (e) managing the classroom; (f) lecturing; and (g) giving directions. We trained two communication graduate research assistants as coders. The coders participated in All Stars training and then familiarized themselves with the curriculum, the interactive style of teaching it requires, and the coding form. The coders first used this measure to rate videotape lessons obtained from an earlier pilot study. The coders and first author then discussed their ratings after each session and focused on strategies to resolve discrepancies, which resulted in detailed decision rules and revisions to some categories. We eliminated the latter two categories, lecturing and giving directions, because they proved too difficult to quantify in a consistent and replicable manner. Final measures included the following:

#### *Praising and Encouraging Students*

Coders counted the number of times each teacher complimented a student. We distinguished between genuine praise, which elaborates on how well a student performed (e.g., “Wow, that is a really interesting point”), and instrumental praise, which is limited to one or two word perfunctory answers (e.g., “good” or “excellent”). We did not count as praise or encouragement comments that simply acknowledged the fact that the student had spoken (e.g., “alright” or “okay”).

#### *Accepting or Using Pupils’ Ideas*

Coders counted the number of times each teacher accepted a student’s comment or tried to link it to some aspect of the lesson. Accomplished All Stars teachers often provide a verbal bridge between their students’ comments and a main point of the activity. Examples of this type of interaction include paraphrasing what the students said, or accepting student ideas, (“So being brave means standing up for your commitments”) and linking the comment to one of the lesson’s points or using student ideas (“That’s like being free of bad things like drugs”).

#### *Asking Questions*

Coders tallied teachers’ questions that were relevant to the curriculum and ignored questions about other classroom issues or topics. We distinguished between questions that were “original,” and thus not related to any previous statement or

question, and those that were “repeated” or “probing.” Most of the former were those specified by the curriculum. Repeated questions included questions that teachers asked more than once, and which were often used to solicit additional student responses (e.g., “Who else?”). Probing questions referred directly to a student’s previous comment and served as a strategy for follow-up. We distinguished between original, repeated, and probing questions because some were specified in the manual, whereas others were originated by the teacher. Some teachers, though, asked the same questions multiple times in an apparently mechanical fashion, whereas others asked their students probing or clarifying questions to elicit their response. We distinguished between these follow-up questions by categorizing them into either repeated or probing categories, respectively.

### *Sharing Personal Self-Disclosures*

Coders tallied personal anecdotes, which included teachers’ opinions and stories about the topic under discussion. Many anecdotes involved what a teacher did, or might do, in a given situation. We counted each full story as one anecdote. A true or hypothetical story about a student or someone other than the teacher was not coded as a personal anecdote.

### *Managing the Classroom*

Classroom management included statements from teachers that served to correct students’ behavior and keep students on task. We categorized corrective statements into four sub-categories: student specific appropriate, class specific appropriate, student specific inappropriate, and class specific inappropriate. Our purpose in so doing was to determine whether teachers corrected specific students who were being disruptive (e.g., “Carol, you are being disruptive to your classmates,”) or reprimanded the entire class (“Everyone is being too loud. Please be quiet.”). We further assessed the appropriateness of these corrections. Some teachers demonstrated little respect for students in trying to shape their behavior (e.g., “Shut up!”) while other teachers were more respectful (e.g., “Please be quiet.”). See Appendix for the coding instrument used.

Coders reviewed the videotapes and completed the ratings as described above. We collected paired ratings for the development sample ( $n = 17$ ) and for 30 percent of the validation sample ( $n = 107$ ). The first author, who was primarily responsible for the development of the observation measures, met weekly with both coders to review ratings. We noted discrepancies and used them as examples with which to further refine our coding protocol.

### Interactivity Coding Inter-Rater Agreement

To assess inter-rater agreement, we calculated product-moment correlations between the raters’ counts for each category and the intraclass correlations for each count variable (see Table 1). Intraclass correlations, unlike product-moment correlations, provide an index of agreement that takes similarities in both rank and

**Table 1** Bivariate Pearson product-moment correlations and intraclass correlations for teacher interactivity behaviors

	Development sample ( $n = 17$ )		Validation sample ( $n = 107$ )	
	Pearson correlation	Intra-class correlation	Pearson correlation	Intra-class correlation
Classroom management	.97	.98	.96	.92
Praises students	.87	.86	.95	.94
Accepts ideas	.75	.83	.97	.98
Incorporates ideas	.84	.92	.58	.73
Asks original questions	.77	.86	.94	.97
Repeats questions	.82	.87	.93	.85
Asks probing questions	.78	.88	.86	.91

All relationships are significant at  $p < .05$

mean into account. Because intra-rater variance should be similar across both raters and because we used two independent coders throughout, we used a two-way random effects model for absolute agreement. Good agreement between coders was found for all but three of the categories: praise and encouragement; self-disclosed personal anecdotes; and student behavior corrections. The lower reliability for praise and encouragement appeared, from post-hoc discussions with raters, to result from their difficulty in distinguishing between genuine and non-genuine praise; collapsing the two items into one variable yielded a construct with much improved inter-rater agreement. We also discovered that teachers' self-disclosure of personal feelings or anecdotes was an infrequent event, with only 19.6% of teachers having more than one personal self-disclosure. Therefore, we made the decision to dichotomize this variable, which also yielded a variable with good inter-rater agreement. Finally, post-hoc discussions with coders highlighted their difficulty in distinguishing between student- and class-specific corrections, as well as appropriate and inappropriate corrections. These four variables were therefore collapsed into one variable with considerably improved inter-rater reliability. Our final categories had satisfactory inter-rater agreement. We report values based on the average of the raters' ratings, as has been suggested elsewhere (McGraw and Wong 1996; Shrout and Fleiss 1979).

### Correlations Between Samples

Correlations were computed between delivery skill ratings for the developmental lesson versus the validation lesson (see Table 2). In general, the extent to which school staff provided praise and encouragement, accepted and used pupil ideas, asked probing questions, self disclosed personal anecdotes, and corrected students misbehavior was highly correlated when comparing the ratings from the developmental lesson to the ratings from the validation lesson. Correlations between coder ratings for the development lesson and the validation lesson were moderately high but non-significant for asking original and repeated questions.

**Table 2** Correlations between delivery skill in the developmental sample versus the validation sample ( $n = 17$ )

Praises and encourages	.65***
<i>Pupil ideas</i>	
Accepts	.43*
Uses	.58**
<i>Question asking</i>	
Original	.33
Repeated	.39
Probing	.71***
Self-disclosed personal anecdotes	.45*
Student behavior corrections	.94***

Note: \*  $p < .05$ , \*\*  $p < .01$ ,  
\*\*\*  $p < .10$

### Interactivity Factor Analysis

To further reduce the data for predictive validity analyses, we conducted an exploratory factor analysis using Mplus v. 5 (Muthén and Muthén 2007). A Quartamin rotation, which is appropriate for frequency (count) variables, was used in order to allow factors to correlate. A three-factor solution provided the best fit (Log Likelihood =  $-2620.07$  with 29 degrees of freedom). Classroom management loaded on its own factor (management), praise and use of student ideas loaded on a second factor (acknowledgment), and student idea acceptance, asking original questions, repeating questions, and asking probing questions loaded on the third factor (student centered methods;  $\alpha = .85$ ). Personal self-disclosures did not load on any factor, even when a four-factor solution was specified, and as such, was not used in further analyses. Composites for the acknowledgement and student-centered methods scales were created by summing the raw scores of the variables loading on those factors. Raw scores were used in order to maintain the meaningful nature of the scales. The resulting composites indicate frequencies with which each type of behavior was used. See Table 3 for descriptive statistics for each of the original seven variables and the resulting three factors. Acknowledgment and student-centered methods were moderately positively correlated ( $r = .53$ ,  $p < .001$ ). Classroom management was not correlated with student-centered methods ( $r = .13$ ,  $p = .20$ ) or with acknowledgment ( $r = .03$ ,  $p = .74$ ).

**Table 3** Teacher interactivity descriptive statistics ( $n = 107$ )

	Mean	SD	Range
Acknowledgment	23.05	13.57	0–66
Praise and encouragement	7.24	6.43	0–28
Using student ideas	15.81	10.17	0–40
Participation facilitation	90.02	43.38	14–228
Accepting student ideas	36.03	21.00	1–103
Original questions	35.73	14.91	11–72
Repeated questions	10.70	7.70	0–34
Probing questions	7.21	5.10	0–23
Classroom management	5.87	5.19	0–22



## Predictive Validity

To assess the predictive validity of the teacher interactivity measures, student engagement with the curriculum, change in curriculum mediators, and past 30-day substance use were regressed on teacher interactivity (classroom management, acknowledgment, and student-centered methods). Student engagement was measured at follow-up with a 10-item scale (e.g., I looked forward to the All Stars program) ( $\alpha = .93$ ). Curriculum mediators were measured at pre- and post-test with multi-item scales ranging from zero to 10, with the highest value representing attitudes conducive to not using substances. They included Lifestyle Incongruence (11 items, e.g., Smoking cigarettes fits with the kind of life I would like to live, baseline  $\alpha = .78$ ), Normative Beliefs (12 items, e.g., How many people your age do you think get drunk at least once a month, baseline  $\alpha = .82$ ), and Commitment to Not Use Drugs (11 items, e.g., I have decided that I will smoke cigarettes, baseline  $\alpha = .83$ ). Student substance use also was measured at pre- and post-test and included 30-day use of alcohol, tobacco, and marijuana. Given the low frequency of use in this age group, outcomes were dichotomized (yes/no).

Because students were nested within classrooms, multilevel regression models (MLMs) were used. Each student group (teacher-year combination) counted as an independent sampling unit. The SAS mixed model procedure for continuous outcomes, Proc Mixed, was used to estimate models for student engagement and curriculum mediators. Proc Glimmix, the SAS procedure for generalized mixed models, was used for substance use models. Student demographics (gender and race/ethnicity) were included as controls in all models. The dataset was stacked in long form so that pre- and post-test curriculum mediator scores and substance use were employed as outcomes with the two main effects of time and teacher interactivity (acknowledgment, student-centered methods, classroom management) as covariates, and the interaction between time and teacher interactivity as predictors. Engagement/Enjoyment was only measured at post-test.

Intra-class correlations (ICCs) ranged from .05 for both tobacco and marijuana use to .26 for student engagement, with lower ICCs for substance use measures and higher ICCs for student engagement and mediators. This discrepancy may be attributed to the fact that a quality like student engagement may largely, but not entirely, be a function of the teacher, and normative beliefs are primarily a function of classroom peers.

### *Classroom Management*

As expected, student engagement was negatively related to classroom management ( $\beta = -.081$ ,  $SE = .018$ ,  $p < .001$ ). The variable had no relationship with any other student outcome examined.

### *Acknowledgment*

Counter to expectations, the more teachers praised students and incorporated their ideas, the lower students' scored on the normative beliefs scale ( $\beta = -.007$ ,

SE = .004,  $p = .05$ ). Acknowledgment had no relationship to any other student outcome examined.

### *Student-Centered Methods*

The extent to which teachers utilized techniques such as question-asking and accepting student ideas was positively related to student ideals/lifestyle incongruence ( $\beta = .002$ , SE = .001,  $p = .07$ ) and to normative beliefs ( $\beta = .002$ , SE = .001,  $p < .10$ ). Marijuana use slightly decreased as a function of student-centered methods (OR = .999, SE = .000,  $p = .09$ ). Student-centered methods had no relationship with any other student outcome examined.

## **Discussion**

The purpose of this study was to develop a measure that could be used to assess teachers' interactivity during their implementation of All Stars, a drug prevention curriculum. Our first objective was to create item categories that would yield high levels of inter-rater agreement. We achieved this objective in a measurement development sample and then successfully replicated the process in a validation sample. Indeed, in the validation sample we either maintained or enhanced high levels of inter-rater agreement. The fact that this measure maintained very good inter-rater agreement between development and verification samples, each of which involved coding lessons that targeted two different program objectives, suggests that with proper training and adequate supervision this measure may be of value in investigations of teachers' interactivity in delivering other drug prevention curricula.

Our second objective was to develop a measure of key teacher skills typically required by evidence-based drug prevention curricula in general, rather than of curriculum- or lesson-specific delivery strategies. Certain delivery skills were used frequently, regardless of which All Stars lesson was taught. In both lessons we observed teacher variability in praising and encouraging students, accepting and using student ideas, correcting student misbehavior, asking probing questions, and revealing personal anecdotes. The only delivery skills that differed substantially across lessons were asking original or repeated questions. Question-asking skills were most likely confounded by the curriculum itself—Lesson 11 of All Stars specifies 11 questions for teachers to ask of students, while Lesson 8 specifies 38 questions. With each original question comes the potential for repeated questions. It is therefore not surprising that the correlations for original and repeated questions across these two lessons were not significant.

Our third objective was to determine the extent to which item categories were inter-correlated and also related to proximal program outcomes. It appeared that some teachers were more interactive than others, as was reflected by the factor analysis of the delivery skill items. The skills most associated with interactivity, notably accepting students' ideas and asking original, repeated, and probing questions, loaded on one factor ("student-centered methods"). The use of these skills was associated with improvements in students' idealism and normative beliefs

and was marginally related to decreases in marijuana use. Thus, it appears that student-centered delivery skills may, in part, influence important program objectives as well as behavioral outcomes. The mechanism by which this occurs is unclear, however. Student-centered methods were not associated with student engagement. One plausible explanation is that teachers who ask thoughtful questions and listen to students' ideas may influence student affective (in the form of idealism and normative beliefs) and behavioral (i.e., substance use) learning by demonstrating respect for and interest in their students.

Contrary to our expectations, teacher acknowledgment of students was associated with decreases in normative beliefs and failed to predict any other student outcomes. This factor included two items: praising students and using their ideas. Many of the responses that were coded in these categories were rather standard. Even responses that were coded as "genuine" praise were relatively insincere. It is possible that students are desensitized to teachers praise when it is limited to one or two word responses (e.g., "Good job.").

Lastly, teachers who demonstrated greater use of classroom management techniques (i.e., student- and classroom-specific corrections) had students who were less engaged in the program. One possible explanation is that teachers who engaged in corrective strategies may have been less comfortable with the interactive nature of All Stars. The All Stars manual encourages teachers to keep class discussions on the "edge of chaos," which may be difficult for teachers who prefer to use didactic rather than interactive teaching strategies (Hansen and McNeal 1999). These same teachers may have a more difficult time engaging their students in the curriculum. Of course, this assumes that the nature of influence is directed from teacher to student; it is also possible that more disruptive classrooms lead teachers to engage in greater efforts to manage the classroom. It is important to note, however, that all the schools in this study came from the same inner-city school district and as such had similar resources, student composition, and school organizational structures. All of the teachers received the same pre-service All Stars training and support. So although it is difficult to determine the direction of causality, it is not unreasonable to expect teachers who lack skills for managing classrooms in general to rely more heavily upon using corrective statements in their All Stars classes.

Although the use of videotapes provides advantages over live observations (e.g., use for training, repeat observations, resolving discrepancies in coding), there are limitations that make assessing classroom interactions difficult. For instance, we found a high degree of variability in the quality of the videotapes. Teachers often moved off camera, which we instructed them to set up in the back of the classroom facing forward, and we found that the audio sometimes failed to record fully what students said in the classroom. Our interest in this study was to assess teacher interactivity, or delivery skills believed to promote active student involvement. Future research should consider the interaction, or transactional communication between teachers and students, as the unit of analysis. To accomplish this future studies should develop strategies for streamlining coder training and augmenting the quality of the recorded material given to observers. Until methods are developed that can be routinely implemented and economically replicated with relative ease, a significant effort will be required to ensure an acceptable level of inter-rater agreement.

In sum, the field of prevention is increasingly aware of the importance of high quality, interactive teaching. The measure we developed can serve as a valuable tool in research to examine variations in the quality of interactive teaching, which may help explain curricula's failure to achieve their objectives, particularly in full-scale effectiveness trials in which the curricula are judged by the teachers' initial iteration of the curriculum. This tool may be valuable in diagnosing the quality of teachers' interactions with their students. Future investigations should consider combining assessments of interactivity with innovations such as coaching to improve the quality with which interactive programs are administered. The measure we developed has several immediate benefits, including a high level of specificity, which should lend itself to replication with All Stars and other substance abuse prevention curricula. Ultimately, results may help both program developers and practitioners improve their performance.

**Acknowledgments** This study was supported by a grant from the National Institute on Drug Abuse, Grant No. 5 R01 DA016098, "Promoting Fidelity Using Remote and Onsite Support." We would like to acknowledge the contribution of our coders—Terry Dumansky, Sara Patterson, and Grant Bollmer. The work was tedious but their insights and acumen were invaluable. We are also indebted to Dr. Inez Drummond of Chicago Public Schools for her support and to all the teachers who collaborated with us on this study.

## Appendix

Coding instrument	Category	Tally
	Praises or encourages	Genuine: Non-genuine:
	Accepts or uses ideas of pupils	Accepts: Uses:
	Asks questions	Original: Repeated: Probing:
	Self-discloses personal anecdotes	
	Corrects students misbehavior	Student specific, appropriate: Class specific, appropriate: Student specific, inappropriate: Class specific, inappropriate:
	Comments:	

## References

- Abbott, R. D., O'Donnell, I., Hawkins, D., Hill, K. G., Kosterman, R., & Catalano, R. F. (1998). Changing teaching practices to promote achievement and bonding to school. *American Journal of Orthopsychiatry*, *68*, 542–552. doi:10.1037/h0080363.
- Bosworth, K., & Sailes, J. (1993). Content and teaching strategies in 10 selected drug abuse prevention curricula. *Journal of School Health*, *63*, 247–253.

- Botvin, G. J., Dusenbury, L., Baker, E., James-Ortiz, S., & Kerner, J. (1989). A skills training approach to smoking prevention among Hispanic youth. *Journal of Behavioral Medicine*, *12*, 279–296. doi: [10.1007/BF00844872](https://doi.org/10.1007/BF00844872).
- Dusenbury, L., Brannigan, R., Hansen, W. B., Walsh, J., & Falco, M. (2005). Quality of implementation: Developing measures crucial to understanding the diffusion of preventive interventions. *Health Education Research*, *20*, 308–313. doi: [10.1093/her/cyg134](https://doi.org/10.1093/her/cyg134).
- Dusenbury, L., & Falco, M. (1995). Eleven components of effective drug abuse prevention curricula. *Journal of School Health*, *65*, 420–425.
- Flanders, N. (1970). *Analyzing teacher behavior*. Reading, MA: Addison-Wesley.
- Hansen, W. B. (1996). Pilot test results comparing the All Stars program with seventh grade D.A.R.E.: Program integrity and mediating variable analysis. *Substance Use and Misuse*, *31*, 1359–1377. doi: [10.3109/10826089609063981](https://doi.org/10.3109/10826089609063981).
- Hansen, W. B., Graham, J. W., Wolkenstein, B. H., & Rohrbach, L. A. (1991). 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*, *52*, 568–579.
- Hansen, W. B., & McNeal, R. B. (1999). Drug education practice: Results of an observational study. *Health Education Research*, *14*, 85–97. doi: [10.1093/her/14.1.85](https://doi.org/10.1093/her/14.1.85).
- Harachi, T. W., Abbott, R. D., Catalano, R. F., Haggerty, K. P., & Fleming, C. B. (1999). Opening the black box: Using process evaluation measures to assess implementation and theory building. *American Journal of Community Psychology*, *27*, 711–731. doi: [10.1023/A:1022194005511](https://doi.org/10.1023/A:1022194005511).
- Harrington, N. G., Giles, S. M., Hoyle, R. H., Feeney, G. J., & Yungbluth, S. C. (2001). Evaluation of the All Stars character education and problem behavior prevention program: Pretest-posttest effects on mediator and outcome variables for middle school students. *Health Education & Behavior*, *28*, 533–546. doi: [10.1177/109019810102800502](https://doi.org/10.1177/109019810102800502).
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, *1*, 30–46. doi: [10.1037/1082-989X.1.1.30](https://doi.org/10.1037/1082-989X.1.1.30).
- McNeal, R. B., Hansen, W. B., Harrington, N. G., & Giles, S. M. (2004). How All Stars works: An examination of program effects on mediating variables. *Health Education & Behavior*, *31*, 165–178. doi: [10.1177/1090198103259852](https://doi.org/10.1177/1090198103259852).
- Muthén, L. K., & Muthén, B. O. (2007). *Mplus statistical software (Version 5.0)*. Los Angeles, CA: Muthén & Muthén.
- Pentz, M. A., Trebow, E. A., Hansen, W. B., MacKinnon, D. P., Dwyer, J. H., Johnson, C. A., et al. (1990). Effects of program implementation on adolescent drug use behavior. *Evaluation Review*, *14*, 264–289. doi: [10.1177/0193841X9001400303](https://doi.org/10.1177/0193841X9001400303).
- Ringwalt, C. L., Pankratz, M. M., Hansen, W. B., Dusenbury, L., Jackson-Newsom, J., Giles, S. M., Brodish, P. (2007). The potential of coaching as a strategy to improve the effectiveness of school-based substance use prevention curricula. *Health Education & Behavior*, July 25, 2007. doi: [10.1177/1090198107303311](https://doi.org/10.1177/1090198107303311).
- Shrout, P. E., & Fleiss, J. J. (1979). Intraclass correlations: Uses in assessing reliability. *Psychological Bulletin*, *86*, 420–428. doi: [10.1037/0033-2909.86.2.420](https://doi.org/10.1037/0033-2909.86.2.420).
- Sobol, D. F., Dent, C. W., Gleason, L., Brannon, B. R., Johnson, C. A., & Flay, B. R. (1989). The integrity of smoking prevention curriculum delivery. *Health Education Research*, *4*, 59–67. doi: [10.1093/her/4.1.59](https://doi.org/10.1093/her/4.1.59).
- Substance Abuse and Mental Health Services Administration. (2008). SAMHSA model programs: Effective substance abuse and mental health programs for every community. <http://modelprograms.samhsa.gov/model.htm>. Retrieved 22 July 2008.
- Tobler, N. (1986). Meta-analysis of 143 adolescent drug prevention programs: Quantitative outcome results of program participants compared to a control or comparison group. *Journal of Drug Issues*, *16*, 537–567.
- Tobler, N. S. (1992). Drug prevention programs can work: Research findings. *Journal of Addictive Diseases*, *11*(3), 1–28. doi: [10.1300/J069v11n03\\_01](https://doi.org/10.1300/J069v11n03_01).
- Tobler, N. S. (2000). Lessons learned. *Journal of Primary Prevention*, *20*, 261–274. doi: [10.1023/A:1021362620740](https://doi.org/10.1023/A:1021362620740).
- Tobler, N. S., Lessard, T., Marshall, D., Ochshorn, P., & Roona, M. (1999). Effectiveness of school-based drug prevention programs for marijuana use. *School Psychology International*, *20*, 105–137. doi: [10.1177/0143034399201008](https://doi.org/10.1177/0143034399201008).
- Tobler, N. S., & Stratton, H. (1997). Effectiveness of school-based drug prevention programs: A meta-analysis of the research. *Journal of Primary Prevention*, *18*, 71–128. doi: [10.1023/A:1024630205999](https://doi.org/10.1023/A:1024630205999)