Three-Year Trajectory of Teachers' Fidelity to a Drug Prevention Curriculum

By: Christopher L. Ringwalt, Melinda M. Pankratz, Julia Jackson-Newsom, Nisha C. Gottfredson, William B. Hansen, Steven M. Giles, and Linda Dusenbury

Ringwalt, C.L., Pankratz, M.M., Jackson-Newsom, J., Gottfredson, N.C., Hansen, W.B., Giles, S.M., & Dusenbury, L. (2010). Three year trajectory of teachers' fidelity to a drug prevention curriculum. Prevention Science, 11(1), 67-76. DOI:10.1007/s11121-009-0150-0.

Abstract:

Little is known about the trajectories over time of classroom teachers' fidelity to drug prevention curricula. Using the "Concerns-Based Adoption Model" (C-BAM) as a theoretical framework, we hypothesized that teachers' fidelity would improve with repetition. Participants comprised 23 middle school teachers who videotaped their administration of three entire iterations of the All Stars curriculum. Investigators coded two key curriculum lessons, specifically assessing the proportion of activities of each lesson teachers attempted and whether they omitted, added, or changed prescribed content, or delivered it using new methods. Study findings provided only partial support for the C-BAM model. Considerable variability in teachers' performance over time was noted, suggesting that their progression over time may be nonlinear and dynamic, and quite possibly a function of their classroom and school contexts. There was also evidence that, by their third iteration of All Stars, teachers tended to regress toward the baseline mean. That is, the implementation quality of those that started out with high levels of fidelity tended to degrade, while those that started out with very low fidelity to the curriculum tended to improve. Study findings suggest the need for ongoing training and technical assistance, as well as "just in time" messages delivered electronically; but it is also possible that some prevention curricula may impose unrealistic expectations or burdens on teachers' abilities and classroom time.

Article:

INTRODUCTION

After a lengthy and repetitive process of program development and evaluation, the field of adolescent substance abuse prevention has identified a number of school-based curricula that have demonstrated positive effects over time, and are so recognized on both federally (Elliot 1998; NIDA 2003; SAMHSA 2008) and privately (Drug Strategies 1998) sponsored registries. The most recent prevalence estimates available suggest that 34.6% of the nation's public middle schools had adopted one of these curricula as of 1999 (Ringwalt et al. 2002) and it is likely that more have done so in the interim, spurred by mandates from funding agencies (Hansen 2001) and from No Child Left Behind legislation (Hallfors et al. 2007). However, drug prevention curricula are unlikely to achieve their objectives unless they are taught with fidelity. Fidelity, sometimes referred to as integrity, adherence, or quality of program delivery, is an important characteristic of program implementation (Dane and Schneider 1998; Domitrovich and

Greenberg 2000; Dusenbury et al. 2003; Higgins and Hogan 1999). It includes both the proportion of lesson content attempted, as well as modifications to that content (Jackson-Newsom, J., Giles, S. M., Pankratz, M. M., Ringwalt, C. L., Briola, B., Dusenbury, L., et al. (2009). Measuring adherence to a prevention program curriculum: Findings and Challenges. Manuscript submitted for review). Hill (2007) classified modifications as omissions, additions, and changes, of which omissions were the most prevalent modification type while changes were the least commonly observed.

There is now abundant evidence from a variety of sources that the quality of program implementation differs considerably across teachers (Dusenbury et al. 2003, 2005; Melde et al. 2006; Pankratz et al. 2006; Ringwalt et al. 2003; Rohrbach et al. 1993; Tappe et al. 1995; Tortu and Botvin 1989), which is hardly surprising given their orientation towards creativity and flexibility in the classroom (St. Pierre and Kaltreider 2004). The literature is equally clear that the effects of prevention curricula are likely to suffer as a result of adaptation (Botvin et al. 1995; Dane and Schneider 1998; Drake et al. 1996; Dusenbury et al. 2003). Further, effect sizes tend to attenuate as evaluations move from conditions of efficacy to program implementation in more naturalistic settings (Connell and Turner 1985; Dodge 2001; Glasgow et al. 2003; Tobler et al. 2000). However, there is a dearth of research concerning factors that generate successful program implementation (Dane and Schneider 1998; Fagan and Mihalic 2003; Wandersman et al. 1998).

Diffusion of innovation theory suggests that teachers' fidelity to the prevention curricula they administer will deteriorate with repetition because implementers tend to make novel products their own through "reinvention" (Rogers 2003). However, very little is actually known about the natural progression of fidelity over time within the context of teachers' administration of classroom-based curricula. Indeed, most evaluations are based on initial implementations, and few studies actually track teachers past their first cycle of program delivery. McCormick and her colleagues (1995) found a decline in the proportion of curricula activities that teachers implement from one school year to the next, a process that Bond and colleagues (2000) characterized as "program drift." Similarly, Connell and Turner (1985) found that in their second year of implementation teachers administered fewer components of their curriculum, adapted it more, and taught it for fewer hours. Gingiss and Hamilton (1989) reported that their teachers' initial implementation of a curriculum tended to be mechanical, while in later administrations the teachers relied less on curriculum guides and became more spontaneous. While the limited evidence available suggests that fidelity decreases over time, Rogers and others (e.g., Berman and McLaughlin 1976) assert that reinvention is positively associated with program sustainability, though not necessarily effectiveness, as practitioners use and modify the innovations they have adopted. To date, no one has linked different types of curriculum modifications to the program's outcomes.

Hall and Hord (1987) have developed a framework titled the "Concerns-Based Adoption Model" (C-BAM) that provides specific guidance as to changes that teachers and other practitioners are likely to make over time when they administer novel curricula or programs. C-BAM is now widely accepted in the field of education (Bailey and Palsha 1992; Cheung 2002) and has been called the most "robust" and "empirically grounded" theory to understand the process of implementation of educational innovations (Anderson 1997; Hall et al. 1975). The model's

developers suggest that teachers' "Level of Use" of an educational innovation—that is, their behaviors and skills in regard to the curriculum (Hall and Hord 1987; Hord 1987)—will evolve or develop systematically with each iteration through a hierarchy from initial orientation (Level I) to mastery (Level VI) (Shotsberger and Crawford 1999). The first two stages, orientation (Level I) and preparation (Level II), are those at which a teacher takes the initiative to learn about the innovation and develops definite plans to use the innovation. Level III, "Mechanical use," is the behavioral stage of most pertinence to teachers who are administering a curriculum for the first several iterations (Hall and Hord 1987), when their use is likely to be disjointed, uneven, superficial, and characterized by managerial problems (Cheung 2002; Hall et al. 1975). These problems may relate to the organizational, administrative, or logistical mechanics of curriculum implementation, namely its content, materials, lesson planning, and any new teaching skills or classroom instructional or management behaviors required (Anderson 1997). At this level, teachers may feel awkward and inadequate; like cooks confronted with a new recipe, they are focused on simply coping with demands of the new material (Huberman and Miles 1984). Insofar as they make any changes to the curriculum, these changes are likely to reflect the teacher's logistical or organizational requirements, as opposed to their students' learning needs. That is, the purpose of their adaptations will be to decrease the challenges they face as a result of curriculum delivery (Hall and Hord 1987).

As teachers progress to the next level of familiarity with the curriculum, they establish a routine pattern of behaviors that they find satisfactory (Level IV(a)), which is characterized by "equilibrium," "stability" and full fidelity to the curriculum. Many teachers never advance beyond this level, while others may then begin to adapt it (Level IV(b)), based on their understanding of its effects on their students (Level IV(b)) (Hord 1987; Horsley and Loucks-Horsley 1998). Of particular note, the developers' experience with the model has suggested that at least three iterations over as many years are required before teachers gain sufficient mastery of the curriculum that its implementation is both efficient and effective, although experience per se does not necessarily lead to the competence characterized by this level (Hall et al. 1975; Loucks-Horsely 1996).

The purpose of this study is to examine changes to the fidelity with which teachers implemented a novel drug prevention curriculum three times over the course of three successive years. We hypothesized that we would see considerable variation in fidelity in teachers' performance relative to the content specified by the curriculum guide. We hypothesized further that teachers' fidelity to the curriculum would improve during this period, as they became more familiar with the requirements of its administration and made the transition from Level of Use Stage III "Mechanical" to Stage IV(a) "Routine." We believed it was unlikely that teachers would progress beyond Stage IV(a) over the course of three implementations; however, to the extent that they did, the C-BAM model suggests that their fidelity would then begin to deteriorate.

METHODS

The All Stars Curriculum

All Stars is an evidence-based prevention program designed to reduce adolescent substance use, sexual behavior, and violence through changes in specific mediating variables (Harrington et al. 2002; McNeal et al. 2004). The mediating variables targeted by All Stars include normative beliefs, personal commitments not to use substances, perceptions that substance use may

interfere with personal values and lifestyles, bonding to school, and perceptions of positive parental attentiveness. The curriculum consists of 24 sessions, of which 13 are required and are administered to the entire class during classroom time. The program includes interactive and cooperative learning activities such as debates, games, and general discussion. Each session is designed to affect at least one of the curriculum's five mediating variables specified above. All Stars has been extensively implemented and evaluated, and has repeatedly yielded evidence of effectiveness (Hansen and Dusenbury 2004; Harrington et al. 2002).

Study Participants

Forty-four schools participated in our randomized trial testing the effect of personal coaching on the fidelity with which teachers implemented the All Stars curriculum (Ringwalt et al. 2009). Within each school, one teacher per year implemented All Stars to a class of 7th graders for up to 3 years. In total, 12 teachers implemented it once, 13 implemented it twice, and 23 implemented it all three times. Most teachers who only taught the curriculum once or twice did so because they either left their school (44%) or took a new position within it that did not allow them to continue teaching All Stars (20%). Two teachers (8%) were resistant to completing a second implementation and three (12%) faced significant personal and professional challenges that did not allow them to teach a third time. The remaining four (16%) replaced an earlier teacher who left the study and only had the opportunity to teach one or two iterations of All Stars.

To examine if there were any differences between those teachers that completed only one or two iterations of All Stars versus those that were able to complete all three, we regressed teacher characteristics and baseline fidelity on the number of All Stars iterations completed (one or two versus three). Those completing all three iterations of All Stars were less likely to be classroom teachers than their counterparts (57% vs. 80%; see Table 1). They also omitted more content (33.4% of steps attempted versus 26.1%) and used new methods less frequently (12.2% of steps attempted versus 18.5%) in Lesson 4 than their counterparts (see Table 2). There were no differences between those who completed one or two iterations of All Stars and those who completed three for any other characteristic or fidelity measure assessed.

Teachers that implemented all three times were predominately female (87%), classroom teachers (57%), and master's level professionals (57%; see Table 1). On average, they were 39.3 years old and had 10.5 years experience in education. They predominately identified themselves as White (52%) or African American (44%). Nearly half had experience in substance use prevention (48%).

All teachers were exposed to the standard All Stars two-day in-person training, received lesson-specific teaching tips via email, and had access to follow-up consultation from the lead All Stars trainer upon request. Follow-up consultation was rarely used. Intervention teachers were also assigned a personal coach who provided specific feedback on their implementation. Because the intervention was not effective in changing either teachers' fidelity or student behaviors (Ringwalt et al. 2009), it is not further considered in this manuscript. Most teachers completed their first implementation during the spring or fall of 2004, their second implementation in 2005, and their third implementation in 2006.

Table 1: Teacher demographics and baseline adherence by number of times the teacher implemented All Stars

		Implemented one or two times ($N = 25$)	Implemented three times ($N = 23$)
Teacher demographics		%	%
Age, Mean (SD)		41 (10)	39 (10)
Gender (male)		28	13
Race/ethnicity			
African–American		52	44
White		36	52
Other		12	4
Hispanic/Latino		16	0
Graduate degree		48	57
Classroom teacher		80*	57
Years experience in education, Mean (SD)		9 (7)	11 (7)
Experienced in teaching substance use prevention		28.0	47.8
Baseline adherence		Mean (SD)	Mean (SD
Attempts	Lesson 4	74 (13.4)	75 (19)
	Lesson 11	77 (16.5)	76 (16)
Omissions	Lesson 4	26 (12.4)*	33 (14)
	Lesson 11	35 (18.1)	42 (11)
Additions	Lesson 4	32 (14.5)	25 (15)
	Lesson 11	69 (16.9)	70 (15)
Changes	Lesson 4	15 (12.6)	13 (18)
	Lesson 11	15 (13.1)	26 (15)
New methods	Lesson 4	19 (9.3)**	12 (9)
	Lesson 11	10 (8.9)	12 (7)

*p < .10 **p < .05

Table 2: Teacher fidelity by implementation year ($N = 23$)

	Lesson	Baseline Fidelity Mean (SD), Median, Range	Year 2 Fidelity Mean (SD), Median, Range	Year 3 Fidelity Mean (SD), Median, Range
Attempts	4	75 (19), 75, 31–100	68 (20), 69, 25–94	66 (25), 75, 6–94
	11	75 (16), 77, 38–92	71 (20), 77, 23–92	80 (15), 85, 38–100
Omissions	4	33 (14), 33, 8–67	25 (21), 23, 0–100	28 (19), 23, 7–100
	11	42 (11), 42, 18–60	51 (16), 50, 27–100	50 (15), 50, 25–80
Additions	4	25 (15), 21, 7–60	26 (12), 23, 0–50	25 (21), 21, 0–100
	11	70 (15), 71, 44–100	50 (18), 55, 17–80	42 (21), 39, 0–80
Changes	4	13 (18), 8, 0–80	12 (16), 7, 0–50	15 (22), 11, 0–100
	11	22 (15), 18, 0–60	16 (15), 13, 0–60	13 (14), 9–40
New methods	4	12 (9), 13, 0–33	13 (9), 11, 0–33	11 (8), 9, 0–25
	11	12 (7), 10, 0–30	13 (20), 9, 0–100	10 (5), 10, 0–20

As a criterion for recruitment into the parent study, teachers agreed to videotape each lesson they implemented by placing the camcorder we provided in the back of their classrooms and focusing it towards the front. We identified videotapes only by unique teacher identification numbers and the study year. The videotapes, which were mailed directly to study staff and were not viewed by anyone in the school system, were then rated by trained coders using the observation form described below. All study procedures were reviewed and approved by PIRE's Institutional Review Board for the Protection of Human Subjects.

Coder Observation Form

For the purposes of this paper, we coded Lessons 4 and 11, which we selected because they required extensive interactivity between teachers and students, and thus were expected to demonstrate variability among teachers' ability to successfully administer the program. Given the importance of interactivity to effective prevention program, we also thought that coding a highly interactive lesson would produce the most useful information to advance the field. The

sequence of these lessons in the curriculum also provided an indication of fidelity of lessons located toward the beginning and end of curriculum delivery. The primary purpose of Lesson 4 is to motivate students to set goals for their future, while the function of Lesson 11 is to encourage students to commit to developing good habits. Lesson 4 contains a wealth of instructional detail and requires individual, small group, and then large group work, strategies that are collectively designed to lead students through a process of thinking about their wanted (and unwanted) futures. The bulk of Lesson 11 constitutes a discussion and summary steps that pertain to understanding commitments and good habits. It also includes a worksheet exercise designed to help students share their understanding of good habits to develop.

The coder observation form was created through a highly iterative process of reviewing and coding numerous videotapes of teachers' delivery of All Stars lessons, and evaluating the extent to which our instrument and coding instructions adequately documented the complex nature of our observations. Coders checked whether teachers attempted each step in each lesson. Steps constituted the smallest unit of instruction that teachers completed, for which there were specific goals, concrete instructions for teachers to follow, and specific prompts for student-centered questions. Coders rated a step as attempted if any of the material presented in the step was delivered by the teacher, regardless of the amount. For each step attempted, coders then checked whether there were adaptations to content (the substantive component of the material) and methods (the instructional strategies by which the content was delivered to students). In total there were 16 possible steps in Lesson 4, and 13 in Lesson 11.

We coded three mutually exclusive categories of adaptations to curriculum content: changes, omissions, and additions. Changes included any rewording of material as written, including modifications to statements, questions, or instructions (beyond simple rephrasing). This category of modifications was limited to alterations to content but excluded alterations involving teaching method or strategies. Omissions included any deletion of content within a step, and additions included any new material presented that was not specified in the curriculum. To assess modifications to teaching methods, coders indicated whether or not any new strategies were used during the delivery of the material within each step. New methods were defined as any change in teaching strategy as prescribed by the curriculum.

Coding Process

Coders were graduate students who received the standard training in the All Stars curriculum, which was followed by extensive training in the coding process. All tapes of teachers' initial implementation of both Lessons 4 and 11 were rated by two coders to establish inter-rater reliability (Jackson-Newsom, J., Giles, S. M., Pankratz, M. M., Ringwalt, C. L., Briola, B., Dusenbury, L., et al. (2009). Measuring adherence to a prevention program curriculum: Findings and Challenges. Manuscript submitted for review), and 20% of subsequent tapes were double coded to facilitate continuing assessment. A codebook was created to capture decision rules that we made as common modifications were identified and codified; and the third author reviewed ratings throughout the process, provided detailed feedback to each coder, and served as a "gold standard" when there was a discrepancy between coders. We then used a multi-pronged approach to assess level of agreement between the two raters in which we assessed concordance with log odds ratios, evaluated marginal homogeneity with McNemar's tests, and reported proportions of agreement. Percentage agreement was above 80% for all categories.

Measurement

We began the creation of fidelity variables by calculating the percentage of steps attempted. We then summed the number of steps with omissions, additions, changes, or new methods, respectively, and then divided this by the total number of steps attempted. The resulting fraction created a variable that indicated the percentage of steps attempted that contained modifications. For each of the two lessons coded, this procedure yielded five ratings.

Data Analyses and Results

We were interested in modeling teachers' trajectories of fidelity to the curriculum for all those who implemented the curriculum three times. Ideally we would have liked to use a multilevel model with random intercepts and slopes to address this research question. However, the sample size was too small to support a multilevel model that included both random intercepts and slopes. As an alternative, we computed fidelity means, standard deviations, medians, and ranges by year and then implemented a "brute force" multilevel model. To do so, we estimated a simple linear regression of fidelity on year of implementation for each teacher in our sample:

$$Adherence_{ij} = \beta_{0j} + \beta_{1j} Year_{ij} \qquad (1)$$

In the above equation, the subscript i represents year of implementation. There are three observed years of implementation, j ($I = 3$) for 23 teachers ($J = 23$). Instead of using a multilevel modeling statistical package (e.g., SAS Proc Mixed or HLM) that would have utilized all of the available data from the sample to estimate the trajectories for each individual teacher, we used SAS Proc GLM to estimate simple linear regressions. We selected this procedure because our data were slightly skewed and Proc GLM does not require conditional normality of the dependent variable to yield unbiased parameter estimates. Standard error estimates are biased if normality conditions are not met when using GLM; however, our "brute force" model does not make use of the standard error estimate, so normality is not essential. Individual estimates for the intercepts (β_{0j}) and slopes (β_{1j}) were outputted for each type of fidelity, and we then calculated the mean and standard deviation of each parameter and their inter-correlations. These results are approximately equivalent to the results that would have been obtained had we been able to employ a random intercept, random coefficient multilevel model. That is, we were not interested in each teacher's parameter estimates, but in both 'fixed effects' (the mean of the intercepts, $\overline{\beta}_{0j}$ and the mean of the slopes, $\overline{\beta}_{1j}$, and 'random effects' (the variance of the intercepts and slopes).

Table 2 displays the fidelity means, standard deviations, medians, and ranges by study year. Table 3 then summarizes the results from simple linear regressions of each type of fidelity on time for the 23 teachers who implemented All Stars three times. The first column indicates the mean baseline fidelity (i.e., the fidelity intercept), and its standard deviation shows individual teachers' variation around the mean. The second column indicates the degree of change in each type of fidelity per year (i.e., the fidelity slope), and its standard deviation shows the individual variation in fidelity trajectories. For this, a negative number for attempts means that fidelity declined over time, while a positive number means that fidelity increased. For all other fidelity categories (i.e., omissions, additions, changes, new methods), a negative number indicates greater fidelity over time and a positive number indicates less fidelity. The correlation between the intercept and the slope indicates the relationship between teachers' baseline fidelity levels and their trajectories in regards to fidelity over time. For this, a negative number indicates that

teachers who started out with above average fidelity tend to decline over time, and that teachers who started out with below average fidelity tend to improve over time. The results presented in this table constitute a small-sample alternative to a multilevel model with random slopes.

Table 3: Longitudinal trajectories of fidelity for teachers administering All Stars three times over 3 years (N = 23)

Type of fidelity	Lesson	Baseline fidelity (fidelity intercept)		Change over time (fidelity slope)		Correlation between baseline fidelity and change over time
		Mean	SD	Mean	SD	
Attempts	4	.77***	.22	−.03	.09	−.63**
	11	.70***	.22	.02	.06	−.50*
Omissions	4	.42***	.26	−.08*	.14	−.94***
	11	.39***	.21	.05*	.09	−.91***
Additions	4	.29***	.21	−.02	.08	−.85***
	11	.82***	.30	−.14***	.15	−.93***
Changes	4	.17*	.27	−.02	.10	−.93***
	11	.24***	.23	−.04	.09	−.84***
New methods	4	.14***	.13	−.01	.06	−.84***
	11	.16***	.14	−.02	.06	−.72***

*p < .05, **p < .01, ***p < .001

Table 2 shows that there is substantial variation in fidelity by teacher. A comparison of mean and median fidelity scores reveals that, while there is some skewness to our data, there are no obvious outliers that would lead to parameter bias. Table 3 shows that baseline fidelity rates are all significantly different from zero, with teachers most likely to attempt steps and least likely to use new methods within the steps they attempt. While few of the time trends were significantly different from baseline, teachers did make fewer omissions within the Lesson 4 steps attempted over the 3-year period, and more omissions (but fewer additions) within the Lesson 11 steps attempted. The standard deviations suggested that this degree of change varied from one teacher to another. The most interesting finding from our analyses was that, for all types of fidelity and for both lessons, intercepts and slopes were strongly negatively correlated. That is, teachers who started out making more modifications decreased the number of these over time, whereas teachers who initially made fewer modifications tended to make more of these over time. We found a similar pattern for teachers' lesson attempts over the 3-year period. Clearly, teachers tended to regress to the mean on all measures.

DISCUSSION

This study is the first to track changes in the fidelity with which middle school teachers administer substance use prevention curricula over time. We were not surprised to find a high degree of variation in the fidelity with which teachers initially implemented the All Stars curriculum. However, we did not expect to find that, regardless of their initial level of proficiency, all teachers regressed toward a mean level of fidelity in subsequent iterations of the curriculum. As such, our findings provided only partial support for the hypotheses we derived from the C-BAM model, which led us to believe that curriculum fidelity would generally improve over the period that we observed teachers' implementation. Like many theories of behavior change, the C-BAM model may thus be somewhat reductionistic and need further elaboration.

Although the nature of the curriculum content varied across the two lessons we examined, the overall pattern for both suggests that teachers struggled to deliver all of the prescribed content in

each lesson. For instance, teachers implemented only three-fourths of the curriculum steps for both All Stars Lessons 4 and 11 the first time they taught the curriculum. In regards to Lesson 4, which included a substantial amount of instructional detail, teachers tended to cover *fewer* steps over time, but omitted less material in those steps they did cover. For Lesson 11, which has a higher proportion of discussion-oriented steps, teachers tended to cover *more* steps over time, but may have compensated for increases in the time required to do so by omitting more material in the steps they did deliver. In other words, for Lesson 4, teachers did less over time, but they improved the quality of what they attempted to do. In Lesson 11, teachers tried to do more, but the quality of what they attempted seems to have suffered as a result. In each case, fidelity might be enhanced by allowing more time for instruction. However, given the many competing demands for instructional time in the nation's schools, and the generally fixed length of class periods, any expansion of the time allotted for teaching drug prevention curricula seems unlikely.

A *post hoc* review revealed that difference noted between the two lessons in the precise pattern of attempts and modifications over time can be partially attributed to differences in the nature of the content they contained. Similar to what is found with other curricula (Botvin et al. 2001; McCormick et al. 1995; Rohrbach et al. 1993; Tappe et al. 1995; Tortu and Botvin 1989; Tricker and Davis 1988), it appears that teachers may have struggled initially to deliver all of the instructions specified in Lesson 4, but that in subsequent iterations they became increasingly comfortable with these instructions. However, even as they omitted less material within the steps they completed, their fidelity of implementation may have suffered because they still ran out of time to cover the lesson's summary steps.

In Lesson 11, it appears that teachers made more of an effort over time to provide complete instructions for the lesson's worksheet exercise. As teachers included more of this exercise, they tended to omit some of the summary points and questions specified for the lesson's didactic and interactive steps. Many of these steps comprised multiple questions or summary points, and teachers may have come to believe that they did not need to cover each point in order to convey the desired message to their students.

While differences in teachers' fidelity to the prescribed content of the two All Stars lessons are interesting, we believe that our study's most important finding is that regardless of their initial proficiency in teaching the curriculum, all teachers tended to regress to the baseline mean fidelity score over time. That is, teachers who initially demonstrated a high degree of fidelity seemed to learn what content they could successfully modify in their second or third iterations of the curriculum. The C-BAM model suggests teachers can be expected to deliver a curriculum that is new to them in a mechanical fashion, in that they will follow the curriculum guide closely. Then, as they gain experience with the curriculum, they should feel more comfortable with the material and begin to make changes that they believe to be constructive. Conversely, those teachers who started out with below average fidelity may have realized that their modifications were unsuccessful, and they may then have sought to implement the curriculum more conscientiously in subsequent iterations. In both cases, study findings suggest that all teachers need continual support and reinforcement if they are to implement evidence-based curricula with fidelity. One cannot assume that teachers who perform well in their initial implementation year will continue to do so.

The study has several limitations. First, our 'brute force' estimates of each teacher's individual fidelity trajectory in an OLS regression did not take the standard errors of the parameter estimates into account. Multilevel modeling, using empirical Bayes estimation, draws individual trajectory estimates closer to the sample mean in a manner that is proportional to the precision of the estimates. Because shrinkage from empirically derived Bayes estimates is greatest for small samples, our individual intercept and slope estimates are probably further from the mean than they would have been had we been able to estimate a multilevel model (Raudenbush and Byrk 2002). Second, since we examined a relatively large number of outcomes—10 in all—the findings we reported may have capitalized on chance. Third, we were able to code, with existing resources, only 2 of the All Stars curriculum's total of 14 lessons; an examination of the remaining lessons might have revealed additional patterns of fidelity trajectories, and validated (or refuted) what we reported. Data coding, we found, constituted a highly repetitive and resource intensive process; we estimate that a complete coding of each of the two lessons for the entire set of teachers required 750 hours of raters' and supervisors' time.

Of further concern, it is possible that teacher attrition from the study's first year to its third may have biased our results. To examine this potential confound, we compared the demographic characteristics and baseline fidelity scores of teachers with only one or two iterations of All Stars to those with all three. Overall the groups were similar, which is not surprising given that most of the attrition was due to position changes and not to any lack of desire to continue teaching the curriculum. Lastly, the schools that participated in the study were selected following a focused recruitment effort on the part of a community-based organization with which we contracted for this purpose. In that sense, the process was not "natural" as the C-BAM model typically presumes: That is, the impetus for and process of program adoption did not come from inside the school itself. However, the coaching intervention we were testing did not affect either teacher fidelity or student outcomes (Ringwalt et al. 2009); as such, it is likely that fidelity trajectories are similar to those that could be expected in a more naturalistic context.

Study findings have implications for the ongoing support of teachers confronted with a novel curriculum, and particularly one that demands interactive skills and classroom management practices with which they are unfamiliar or uncomfortable. There is now a considerable literature that addresses the need for coaching to support teachers charged with the administration of drug prevention curricula (Ringwalt et al. 2009). However, many observers have expressed concern that in-service training alone may not result in positive and lasting changes in teaching practice (Bailey and Palsha 1992). C-BAM offers both a conceptual and pragmatic framework by which to match teachers' developmental level of concerns and competence in regards to a novel curriculum with the timing and content of the coaching they receive. Of particular importance is the need for technical assistance to continue over multiple cycles of use (Hall et al. 1975).

Unfortunately, this recommendation is at odds with the common practice of teacher preparation in the field of school-based drug prevention curricula, for which teachers are fortunate if they receive a 1- or 2-day training. This practice is entirely understandable, given both the financial and temporal challenges involved in bringing teachers together periodically for further training, and the likelihood that teachers' needs for ongoing assistance are idiosyncratic. Personalized coaching or mentoring has met with only modest success (Ringwalt et al. 2009; Rohrbach et al. 2008) and is also beyond the means of many schools. However, investigators are studying the

effects of a variety of other strategies, including providing teachers with access to a trainer by telephone, and sending "just in time" tips by e-mail that arrive immediately preceding the implementation of a particular lesson. There are now a number of studies devoted to developing an instrument to assess the level of use that individual teachers have achieved, as conceptualized by the C-BAM model (e.g., Bailey and Palsha 1992); the results of these studies could potentially be used to tailor messages to teachers' Level of Use.

The results of this study also highlight issues pertaining to the linearity of change, one of the lingering areas of confusion in the C-BAM model. The model suggests that practitioners will progress in incremental fashion through a set of clearly definable stages, although their exact number is still unresolved (Bailey and Palsha 1992). On the other hand, it may be equally as appropriate to consider change to be nonlinear and dynamic, since teachers will begin at different starting points (Anderson 1997) and may then diverge along a variety of pathways. It may thus be unrealistic to expect to find much consistency in trajectories of fidelity over time. Indeed, our study's findings suggest that the process by which teachers master an unfamiliar curriculum is personal and idiosyncratic (Anderson 1997).

In conclusion, we strongly recommend that further attention should be paid to how teacher fidelity to prevention curricula unfolds over time, and the implications to the prevention field of any trajectories found. Systematic changes in fidelity in an undesirable direction may suggest areas of improvement for developers in crafting both initial teacher training and follow-up reminders. Such changes may also suggest the need to modify curricula or their associated guides, especially if the demands placed on the teachers come to be recognized as unrealistic.

The implications for evaluators of changes in fidelity over time are equally salient. To the extent that fidelity improves over time, as the C-BAM model suggests, it would seem inappropriate to evaluate curricula that are administered by novice teachers. Regardless, an understanding of what teachers are—and are not—doing in the classroom, and how their implementation changes as a function of program repetition, is essential to efforts to disseminate and sustain drug prevention curricula in a manner that ensures their continued effectiveness.

ACKNOWLEDGEMENT FOR RESEARCH SUPPORT

This study was supported by a grant from the National Institute on Drug Abuse, Grant # 5 R01 DA016098, "Promoting Fidelity Using Remote and Onsite Support." We thank our coders, Dana Bishop, Bianca Briola, Sara Massie, and Kia Thacker for their painstaking and detailed observations of teachers' curriculum delivery and dedication to the project. 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.

REFERENCES

Anderson, S. (1997). Understanding teacher change: Revisiting the Concerns Based AdoptionModel. *Curriculum Enquiry, 27*, 331–367.

Bailey, D. B., & Palsha, S. A. (1992). Qualities of the Stages of Concern Questionnaire and implications for educational innovations. *Journal of Educational Research, 85*, 226–232.

Berman, P., & McLaughlin, M. W. (1976). Implementation of educational innovations. *The Educational Forum*, 345–370.

Bond, G., Evans, G., Salyers, M. P., Williams, J., & Kim, H. W. (2000). Measurement of fidelity in psychiatric rehabilitation. *Mental Health Services Research, 2*, 75–87.

Botvin, G. J., Baker, E. A., Dusenbury, L., Botvin, E. M., & Diaz, T. (1995). Long-term follow-up results of a randomized drug abuse prevention trial in a White middle-class population. *Journal of the American Medical Association, 273*, 1106–1112.

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

Cheung, D. (2002). Refining a stage model for studying teacher concerns about educational innovations. *Australian Journal of Education, 46*, 305–322.

Connell, D. B., & Turner, R. R. (1985). The impact of instructional experience and the effects of cumulative instruction. *Journal of School Health, 55*, 324–331.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18*, 23–45.

Dodge, K. A. (2001). The science of youth violence prevention: Progressing from developmental epidemiology to efficacy to effectiveness to public policy. *American Journal of Preventive Medicine, 20*, 63–70.

Domitrovich, C. E., & Greenberg, M. T. (2000). The study of implementation: Current findings from effective programs that prevent mental disorders in school-aged children. *Journal of Educational and Psychological Consultation, 11*, 193–221.

Drake, R. E., McHugo, G. J., Becker, D. R., Anthony, W. A., & Clark, R. E. (1996). The New Hampshire Study of Supported Employment for People with Severe Mental Illness. *Journal of Consulting and Clinical Psychology, 64*, 391–399.

Drug Strategies (1998). *Making the Grade: A guide to school drug prevention programs*. Washington, DC: Drug Strategies.

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

Dusenbury, L., Brannigan, R., Hansen, W. B., & Walsh, J. (2005). Quality of implementation: Developing measures to crucial to understanding the diffusion of preventive interventions. *Health Education Research, 20*, 308–313.

Elliot, D. S. (1998). *Blueprints for violence prevention*. Boulder: University of Colorado, Center for the Study and Prevention of Violence.

Fagan, A. A., & Mihalic, S. (2003). Strategies for enhancing the adoption of school-based prevention programs: Lessons learned from the Blueprints for Violence Prevention replications of the Life Skills Training program. *Journal of Community Psychology, 31*, 235–253.

Glasgow, R. E., Lichtenstein, E., & Marcus, A. C. (2003). Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. *American Journal of Public Health, 93*, 1261–1267.

Hall, G., & Hord, S. (1987). *Change in schools: Facilitating the process*. New York: SUNY.

Hall, G., Loucks, S., Rutherford, W., & Newlove, B. (1975). Levels of use of the innovation: A framework for analyzing innovation adoption. *Journal of Teacher Education, 26*, 52.

Hallfors, D. D., Pankratz, M. M., & Hartman, S. (2007). Does federal policy support the use of scientific evidence in school-based prevention programs? *Prevention Science, 8*, 75–81.

Hansen, W. B. (2001). Fidelity: Snatching victory from the jaws of defeat. InW. B. Hansen, S. M. Giles,&M. Fearnow-Kenney (Eds.), *Improving prevention effectiveness* pp. 293–301. Greensboro, NC: Tanglewood Research.

Hansen, W. B., & Dusenbury, L. (2004). All Stars Plus: A competence and motivation enhancement approach to prevention. *Health Education, 104*, 371–381.

Harrington, N. C., Cheah, W. H., Norling, G. R., Hoyle, R. H., & Duvall, J. (2002). *An evaluation of the community-based All Stars character education and problematic behavior prevention program*. Paper presented at the Society for Prevention Research Annual Meeting, Seattle, WA.

Higgins, S. H., & Hogan, P. T. (1999). Internal diffusion of high technology industrial innovations: An empirical study. *Journal of Business & Industrial Marketing, 14*, 61–75.

Hill, L. G. (2007). A focused approach to assessing program fidelity. *Prevention Science, 8*, 25–34.

Hord, S. (1987). *Evaluating educational innovation*. New York: Croom Helm.

Horsley, D., & Loucks-Horsley, S. (1998). CBAM brings order to the tornado of change. *Journal of Staff Development, 19*, 17–20.

Huberman, A. M., & Miles, M. B. (1984). *Innovations up close*. New York: Plenum.

Levenson-Gingiss, P., & Hamilton, R. (1989). Determinants of teachers' plans to continue teaching a sexuality education course. *Family and Community Health, 12*, 40–53.

Loucks-Horsely, S. (1996). Professional development for science education: A critical and immediate challenge. In R. W. Bybee (Ed.), *National standards and the science curriculum: Challenges, opportunities, and recommendations*. Dubuque, IA: Kendall/Hung.

McCormick, L. K., Steckler, A. B., & McLeroy, K. R. (1995). Diffusion of innovations in schools: A study of adoption and implementation of school-based tobacco prevention curricula. *American Journal of Health Promotion, 9*, 210–219.

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.

Melde, C., Esbensen, F. A., & Tusinski, K. (2006). Addressing program fidelity using onsite observations and program provider descriptions of program delivery. *Evaluation Review, 30*, 714–740.

NIDA. (2003). *Preventing drug use among children: A research-based guide for parents, educators, and community leaders*. Bethesda, MD: National Institute on Drug Abuse, National Institutes of Health.

Pankratz, M. M., Jackson-Newsom, J., Giles, S. M., Ringwalt, C. L., Bliss, K., & Bell, M. L. (2006). Implementation fidelity in a teacher-led alcohol use prevention curriculum. *Journal of Drug Education, 36*, 317–333.

Raudenbush, S., & Byrk, A. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks: Sage Publications.

Ringwalt, C. L., Ennett, S., Johnson, R., Rohrbach, L. A., Simons-Rudolph, A., Vincus, A., et al. (2003). Factors associated with fidelity to substance use prevention curriculum guides in the nation's middle schools. *Health Education & Behavior, 30*, 375–391.

Ringwalt, C. L., Ennett, S., Vincus, A., Thorne, J., Rohrbach, L. A., & Simons-Rudolph, A. (2002). The prevalence of effective substance use prevention curricula in U.S. middle schools. *Prevention Science, 3*, 257–265.

Ringwalt, C. L., Pankratz, M. M., Hansen, W. B., Dusenbury, L., Jackson-Newsom, J., Giles, S. M., et al. (2009). The potential of coaching as a strategy to improve the effectiveness of school-based substance use prevention curricula. *Health Education Quarterly, 36*, 695–710.

Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). New York: Free.

Rohrbach, L. A., Graham, J. W., & Hansen, W. B. (1993). Diffusion of a school-based substance abuse prevention program: Predictors of program implementation. *Preventive Medicine, 22*, 237–260.

Rohrbach, L. A., Gunning, M., Sussman, S., Sun, P. (2008). *Predictors of implementation fidelity in the Project Towards No Drug Abuse (TND) Dissemination Trial*. Society for Prevention Research, San Francisco, CA.

SAMHSA. (2008). *SAMHSA's National Registry of Evidence-based Programs and Practices*. Retrieved January 28, 2008, from http://nrepp.samhsa.gov/

Shotsberger, P., & Crawford, A. (1999). On the elusive nature of measuring teacher change: An examination of the Stages of Concern Questionnaire. *Evaluation and Research in Education, 13*, 3–17.

St. Pierre, T. L., & Kaltreider, D. L. (2004). Tales of refusal, adoption, and maintenance: Evidence-based substance abuse prevention via school-extension collaborations. *American Journal of Evaluation, 25*, 479–491.

Tappe, M. K., Galer-Unti, R. A., & Bailey, K. C. (1995). Long-term implementation of the teenage health teaching modules by trained teachers: A case study. *Journal of School Health, 65*, 411–415.

Tobler, N. S., Roona, M. R., Ochshorn, P., Marshall, D. G., Streke, A. V., & Stackpole, K. M. (2000). School-based adolescent drug prevention programs: 1998 meta-analysis. *Journal of Primary Prevention, 20*, 275–336.

Tortu, S., & Botvin, G. J. (1989). School-based smoking prevention: The teacher training process. *Preventive Medicine, 18*, 280–289.

Tricker, R., & Davis, L. G. (1988). Implementing drug education in schools: An analysis of the costs and teacher perceptions. *Journal of School Health, 58*, 181–185.

Wandersman, A., Morrissey, E., Davino, K., Seybolt, D., Crusto, C., Nation, M., et al. (1998). Comprehensive quality programming and accountability: Eight essential strategies for implementing successful prevention programs. *Journal of Primary Prevention, 19*, 3–30.